

Calculating Advanced Statistics

By:

John R. Slate

Ana Rojas-LeBouef

Calculating Advanced Statistics

By:

John R. Slate
Ana Rojas-LeBouef

Online:

< <http://cnx.org/content/col11346/1.1/> >

C O N N E X I O N S

Rice University, Houston, Texas

This selection and arrangement of content as a collection is copyrighted by John R. Slate, Ana Rojas-LeBouef. It is licensed under the Creative Commons Attribution 3.0 license (<http://creativecommons.org/licenses/by/3.0/>).

Collection structure revised: August 18, 2011

PDF generated: August 18, 2011

For copyright and attribution information for the modules contained in this collection, see p. 289.

Table of Contents

1	1. Introductory Chapter	1
2	2. Selecting a Single Group for Statistical Analysis	3
3	3. Steps Involved in Recoding Data in SPSS	25
4	4. Steps Involved in Converting Data from An Excel File into SPSS	57
5	5. Typing Survey Data Into SPSS	75
6	6. Internal Consistency Analysis: Part I	101
7	7. Internal Consistency Analysis: Part II	113
8	8. Internal Consistency Analysis: Part III	123
9	9. Factor Analysis: Part I	135
10	10. Factor Analysis: Part II	147
11	11. Multivariate Analysis of Variance: Part I	161
12	12. Multivariate Analysis of Variance: Part II	189
13	13. Multiple Analysis of Variance: Part III	203
14	14. Discriminant Analysis: Assumptions	215
15	15. Discriminant Analysis: Part I	223
16	16. Discriminant Analysis: Part II	233
17	17. Discriminant Analysis: Part III	245
18	18. Multiple Regression: Assumptions	255
19	19. Multiple Regression: Part I	265
20	20. Multiple Regression: Part II	275
	Attributions	289

Chapter 1

1. Introductory Chapter¹



NOTE: This chapter is published by NCPEA Press² and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of education administration.

About the Authors

John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website (Writing and Statistical Help³) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

Ana Rojas-LeBoeuf is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. Dr. LeBoeuf recently completed her doctoral degree in Reading, where she conducted a 16-year analysis of Texas statewide data regarding the achievement gap. Her research interests lie in examining the inequities in achievement among ethnic groups. Dr. Rojas-LeBoeuf also assists students and faculty in their writing and statistical needs on the Writing and Statistical Help website.

As we stated in our first book, *Calculating Basic Statistical Procedures in SPSS: A Self-Help and Practical Guide to Preparing Theses, Dissertations, and Manuscripts*, we have observed student after student who experienced difficulty with using the Statistical Package for the Social Sciences (SPSS) and with interpreting the voluminous output generated by SPSS. These difficulties were present for basic statistical procedures and even more noticeable for more advanced statistical analyses. The problems students experienced in conducting these procedures, along with statistics anxiety they experienced, led the two of us to develop very specific

¹This content is available online at <<http://cnx.org/content/m40743/1.1/>>.

²<http://www.ncpeapublications.org/books.html>

³<http://cnx.org/content/m40743/latest/www.writingandstatisticalhelp>

and very detailed steps for students to follow. Even with the use of these steps, we still had students who managed to experience difficulty in being able to use SPSS successfully. As a result, we generated screenshots for every major point-and-click step. This combination of steps and screenshots has met with excellent student satisfaction and, most importantly for us as instructors, has enhanced their ability to be successful in using SPSS.

We have written this textbook to be a complement to the basic statistics text mentioned above. We believe that using our detailed steps and screenshots will assist readers in conducting the more complicated analyses discussed in this text. Graduate students, especially doctoral students, who take an advanced statistics course in which SPSS is used will find these steps and screenshots to be very practical and very easy to follow. Finally, faculty members who engage in scholarly activities but are years removed from their own statistics courses will find this textbook to be helpful. We hope that you find our materials helpful to you in your use of SPSS and in your interpretation of SPSS output.

John R. Slate, Professor, Sam Houston State University

Ana Rojas-LeBouef, Assistant Professor, Sam Houston State University

Chapter 2

2. Selecting a Single Group for Statistical Analysis¹



NOTE: This chapter is published by NCPEA Press² and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of education administration.

About the Authors

John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website (Writing and Statistical Help³) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

Ana Rojas-LeBouef is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. Dr. LeBouef recently completed her doctoral degree in Reading, where she conducted a 16-year analysis of Texas statewide data regarding the achievement gap. Her research interests lie in examining the inequities in achievement among ethnic groups. Dr. Rojas-LeBouef also assists students and faculty in their writing and statistical needs on the Writing and Statistical Help website.

About the Editors

¹This content is available online at <<http://cnx.org/content/m40632/1.2/>>.

²<http://www.ncpeapublications.org/books.html>

³<http://cnx.org/content/m40632/latest/www.writingandstatisticalhelp>

Theodore B. Creighton, is a Professor at Virginia Tech and the Publications Director for NCPEA Publications⁴, the Founding Editor of Education Leadership Review,⁵ and the Senior Editor of the NCPEA Connexions Project.

Brad E. Bizzell, is a recent graduate of the Virginia Tech Doctoral Program in Educational Leadership and Policy Studies, and is a School Improvement Coordinator for the Virginia Tech Training and Technical Assistance Center. In addition, Dr. Bizzell serves as an Assistant Editor of the NCPEA Connexions Project in charge of technical formatting and design.

Janet Tareilo, is a Professor at Stephen F. Austin State University and serves as the Assistant Director of NCPEA Publications. Dr. Tareilo also serves as an Assistant Editor of the NCPEA Connexions Project and as a editor and reviewer for several national and international journals in educational leadership.

In this chapter, we will provide you with the steps and screenshots to select out a single group for analysis. In this circumstance, your data set consists of data from multiple groups. Your interest, however, is in obtaining statistics on only one specific group. Using SPSS for all students will only provide you with statistical information for all of the individuals, aggregated across all of the groups. It will not give you statistical information on a specific group. If you follow the steps and screenshots we provide, you can have SPSS generate statistics for the group of students in which you are interested.

2.1

This step presupposes that you have already opened up SPSS and your data file. Click on **Variable View** in the data window.

⁴<http://www.ncpeapublications.org>

⁵<http://ncpeapublications.org/about-elr.html>

book practice data set.sav [DataSet1] - SPSS Statistics Data Editor

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	DISTRICT	Numeric	6	0		None	None	12	Right	Scale
2	DFLCHART	String	1	0		None	None	9	Left	Nominal
3	DPFEIERP	Numeric	4	2	District 2009 Finance: Expenditure-Instructional Expenditures Ratio	None	None	12	Right	Scale
4	DA011TA09R	Numeric	4	2	TAKS Test: Grade 11 All Students All Tests Rate	None	None	6	Right	Scale
5	DA011TA10R	Numeric	4	2	TAKS Test: Grade 11 All Students All Tests Rate	None	None	6	Right	Scale
6	DB011TA09R	Numeric	4	2	TAKS Test: Grade 11 African American All Tests Rate	None	None	2	Right	Scale
7	DB011TA10R	Numeric	4	2	TAKS Test: Grade 11 African American All Tests Rate	None	None	2	Right	Scale
8	DE011TA09R	Numeric	4	2	TAKS Test: Grade 11 Econ Disadv All Tests Rate	None	None	2	Right	Scale
9	DE011TA10R	Numeric	4	2	TAKS Test: Grade 11 Econ Disadv All Tests Rate	None	None	2	Right	Scale
10	DF011TA09R	Numeric	3	2	TAKS Test: Grade 11 Female All Tests Rate	None	None	5	Right	Scale
11	DF011TA10R	Numeric	4	2	TAKS Test: Grade 11 Female All Tests Rate	None	None	2	Right	Scale
12	DH011TA09R	Numeric	4	2	TAKS Test: Grade 11 Hispanic All Tests Rate	None	None	4	Right	Scale
13	Recodedins...	Numeric	8	0	Recoded Instructional Expenditures into 3 groups	{1, Poorest ...	None	28	Right	Scale
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										
31										
32										
33										
34										
35										
36										
37										
38										

Data View Variable View

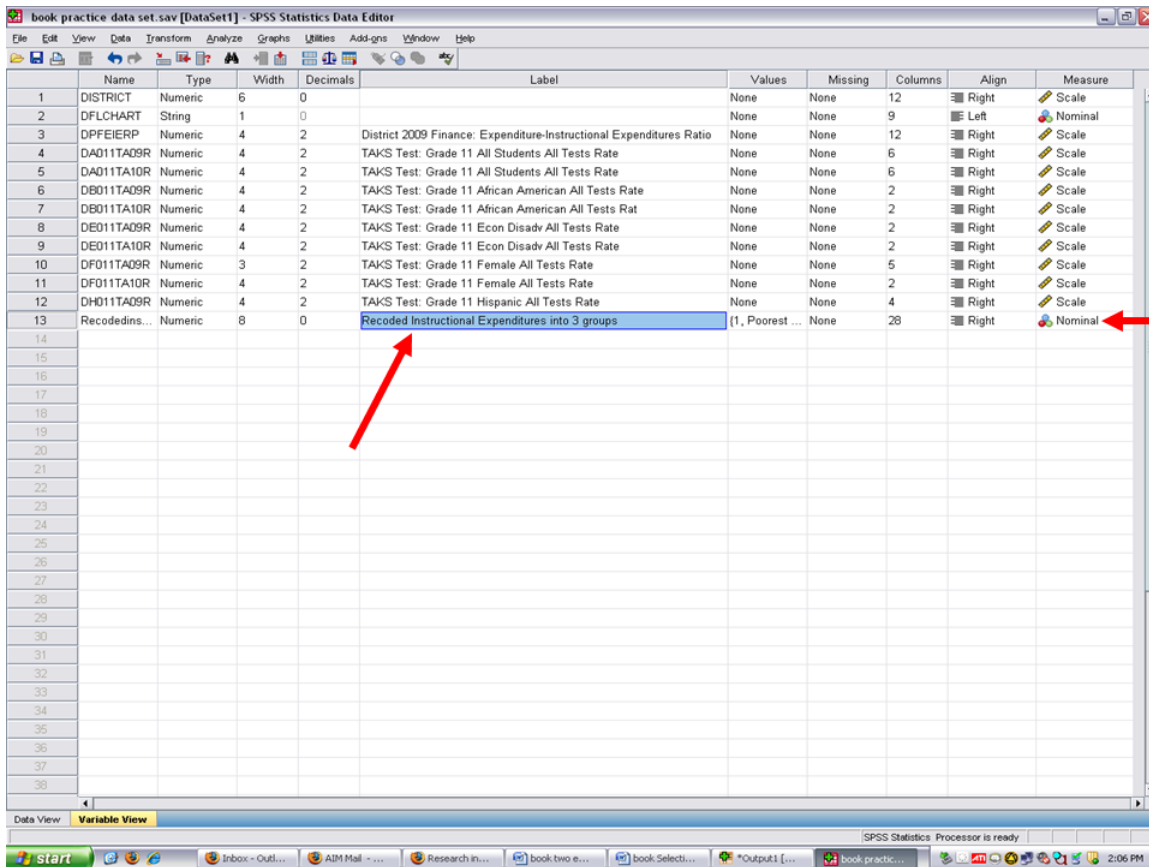
SPSS Statistics Processor is ready

6

Identify your grouping variable. It should be a nominal measure (see the far right column). In the example below, Recoded Instructional Expenditures into 3 groups, is our grouping variable.

Click on the Values cell for this grouping variable. Three dots will appear. These three dots indicate that a screen is hidden underneath this cell.

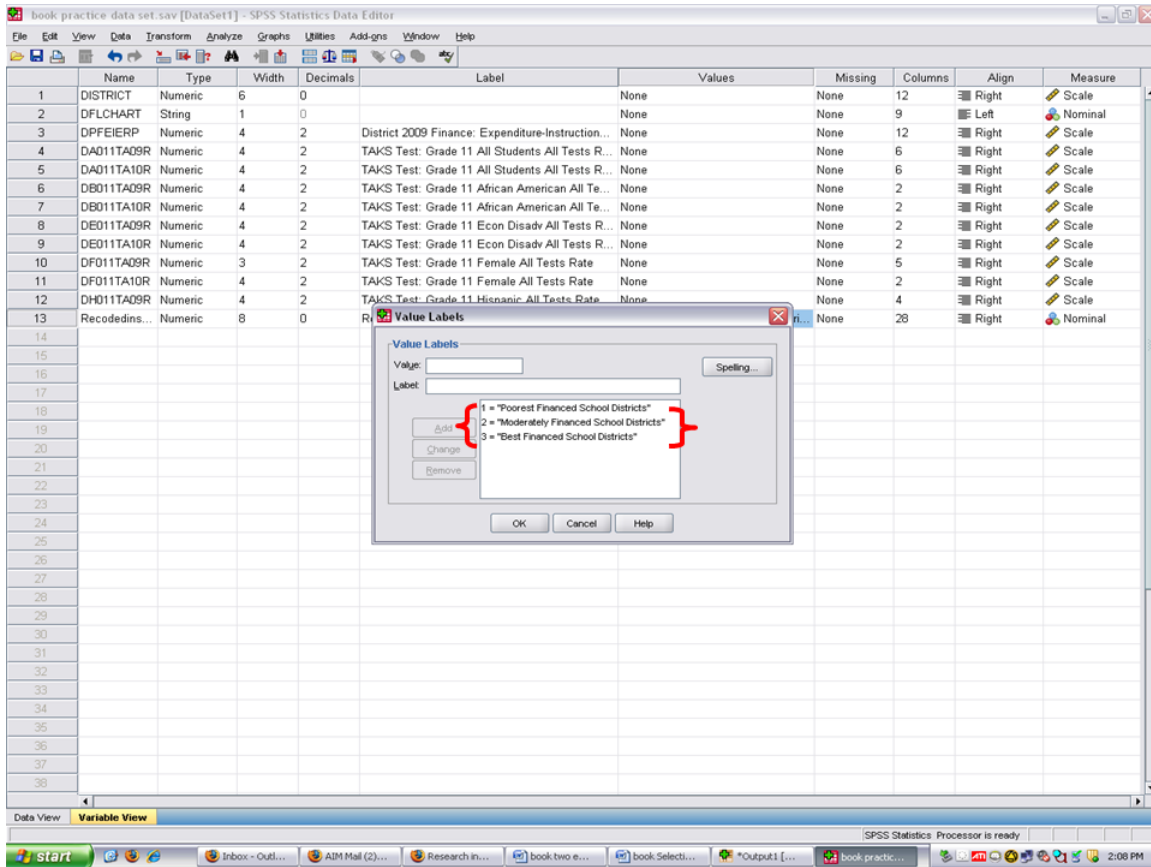
⁶<http://cnx.org/content/m40632/latest/1.1.png/image>



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	DISTRICT	Numeric	6	0		None	None	12	Right	Scale
2	DFLCHART	String	1	0		None	None	9	Left	Nominal
3	DPFEIERP	Numeric	4	2	District 2009 Finance: Expenditure-Instructional Expenditures Ratio	None	None	12	Right	Scale
4	DA011TA09R	Numeric	4	2	TAKS Test: Grade 11 All Students All Tests Rate	None	None	6	Right	Scale
5	DA011TA10R	Numeric	4	2	TAKS Test: Grade 11 All Students All Tests Rate	None	None	6	Right	Scale
6	DB011TA09R	Numeric	4	2	TAKS Test: Grade 11 African American All Tests Rate	None	None	2	Right	Scale
7	DB011TA10R	Numeric	4	2	TAKS Test: Grade 11 African American All Tests Rat	None	None	2	Right	Scale
8	DE011TA09R	Numeric	4	2	TAKS Test: Grade 11 Econ Disadv All Tests Rate	None	None	2	Right	Scale
9	DE011TA10R	Numeric	4	2	TAKS Test: Grade 11 Econ Disadv All Tests Rate	None	None	2	Right	Scale
10	DF011TA09R	Numeric	3	2	TAKS Test: Grade 11 Female All Tests Rate	None	None	5	Right	Scale
11	DF011TA10R	Numeric	4	2	TAKS Test: Grade 11 Female All Tests Rate	None	None	2	Right	Scale
12	DH011TA09R	Numeric	4	2	TAKS Test: Grade 11 Hispanic All Tests Rate	None	None	4	Right	Scale
13	Recodedins...	Numeric	8	0	Recoded Instructional Expenditures into 3 groups	{1, Poorest ...	None	28	Right	Nominal
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										
31										
32										
33										
34										
35										
36										
37										
38										

Clicking on the three dots will reveal the hidden screen. After clicking on the three dots, this screen appears and shows us the names of the three groups.

⁷ <http://cnx.org/content/m40632/latest/1.3.png/image>



For purposes of this chapter, we will have statistics calculated for group 1 which is the Poorest Financed School Districts. Now that we know that this group is numbered as group 1, we will cancel out of this screen.

⁸<http://cnx.org/content/m40632/latest/1.4.png/image>

The screenshot shows the SPSS Statistics Data Editor window with a list of variables. A dialog box titled "Value Labels" is open, showing a list of value labels. The list contains three entries:

- 1 = "Poorest Financed School Districts"
- 2 = "Moderately Financed School Districts"
- 3 = "Best Financed School Districts"

Red arrows point to the "1" in the first entry and the "OK" button at the bottom of the dialog box.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	DISTRICT	Numeric	6	0		None	12	Right	Scale
2	DFLCHART	String	1	0		None	9	Left	Nominal
3	DPFEIERP	Numeric	4	2	District 2009 Finance: Expenditure-Instruction...	None	12	Right	Scale
4	DA011TA09R	Numeric	4	2	TAKS Test: Grade 11 All Students All Tests R...	None	6	Right	Scale
5	DA011TA10R	Numeric	4	2	TAKS Test: Grade 11 All Students All Tests R...	None	6	Right	Scale
6	DB011TA09R	Numeric	4	2	TAKS Test: Grade 11 African American All Te...	None	2	Right	Scale
7	DB011TA10R	Numeric	4	2	TAKS Test: Grade 11 African American All Te...	None	2	Right	Scale
8	DE011TA09R	Numeric	4	2	TAKS Test: Grade 11 Econ Disadv All Tests R...	None	2	Right	Scale
9	DE011TA10R	Numeric	4	2	TAKS Test: Grade 11 Econ Disadv All Tests R...	None	2	Right	Scale
10	DF011TA09R	Numeric	3	2	TAKS Test: Grade 11 Female All Tests Rate	None	5	Right	Scale
11	DF011TA10R	Numeric	4	2	TAKS Test: Grade 11 Female All Tests Rate	None	2	Right	Scale
12	DH011TA09R	Numeric	4	2	TAKS Test: Grade 11 Hispanic All Tests Rate	None	4	Right	Scale
13	Recodedins...	Numeric	8	0		None	28	Right	Nominal

Click on the data view screen.

⁹<http://cnx.org/content/m40632/latest/1.5.png/image>

The screenshot shows the SPSS Statistics Data Editor window with the Variable View tab active. The table below represents the data structure shown in the window. A red arrow points to the 'Data' button in the bottom-left corner of the window.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	DISTRICT	Numeric	6	0		None	None	12	Right	Scale
2	DFLCHART	String	1	0		None	None	9	Left	Nominal
3	DPFEIERP	Numeric	4	2	District 2009 Finance: Expenditure-Instruction...	None	None	12	Right	Scale
4	DA011TA09R	Numeric	4	2	TAKS Test: Grade 11 All Students All Tests R...	None	None	6	Right	Scale
5	DA011TA10R	Numeric	4	2	TAKS Test: Grade 11 All Students All Tests R...	None	None	6	Right	Scale
6	DB011TA09R	Numeric	4	2	TAKS Test: Grade 11 African American All Te...	None	None	2	Right	Scale
7	DB011TA10R	Numeric	4	2	TAKS Test: Grade 11 African American All Te...	None	None	2	Right	Scale
8	DE011TA09R	Numeric	4	2	TAKS Test: Grade 11 Econ Disadv All Tests R...	None	None	2	Right	Scale
9	DE011TA10R	Numeric	4	2	TAKS Test: Grade 11 Econ Disadv All Tests R...	None	None	2	Right	Scale
10	DF011TA09R	Numeric	3	2	TAKS Test: Grade 11 Female All Tests Rate	None	None	5	Right	Scale
11	DF011TA10R	Numeric	4	2	TAKS Test: Grade 11 Female All Tests Rate	None	None	2	Right	Scale
12	DH011TA09R	Numeric	4	2	TAKS Test: Grade 11 Hispanic All Tests Rate	None	None	4	Right	Scale
13	Recodedins...	Numeric	8	0	Recoded Instructional Expenditures into 3 grou... (1, Poorest Financed School Distr...	None	None	28	Right	Nominal
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										
31										
32										
33										
34										
35										
36										
37										
38										

We are now at the data screen. To select out a specific group on which to have statistics calculated, click on Data.

✓ Data

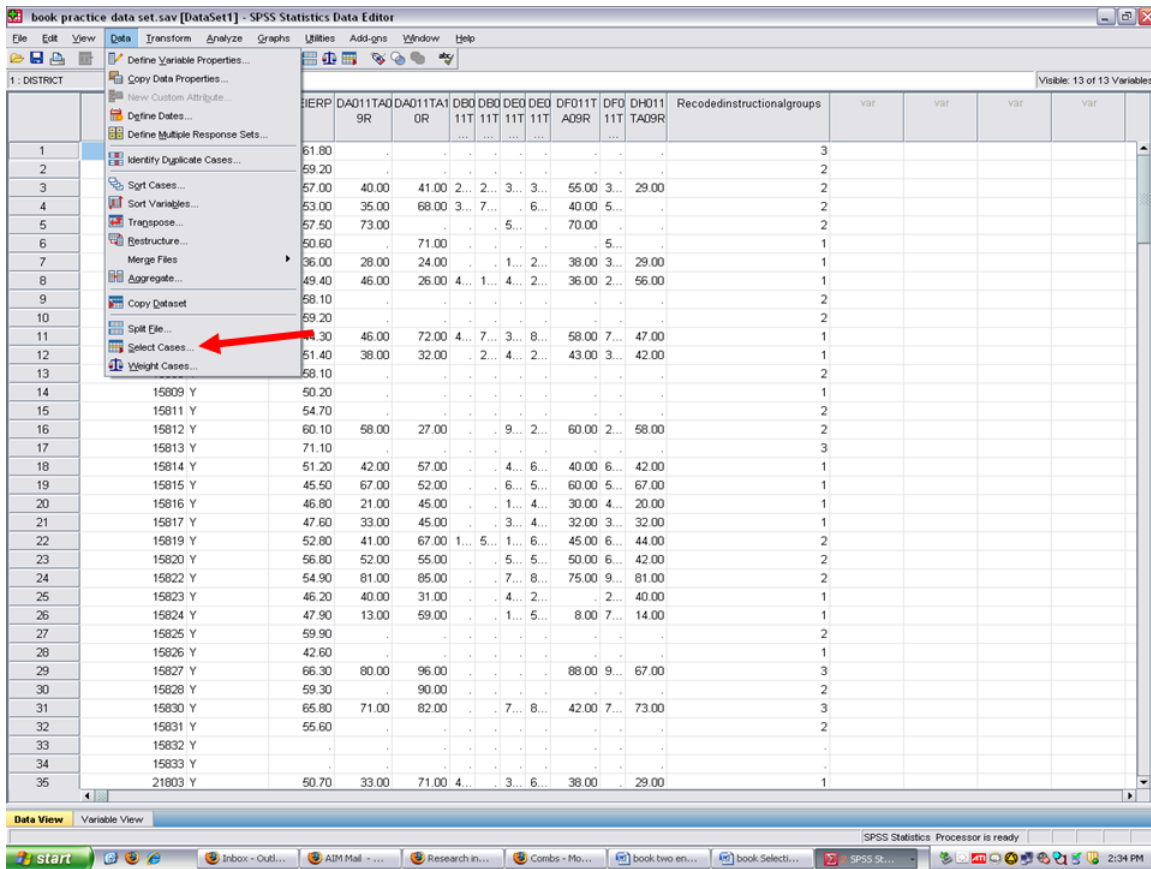
¹⁰<http://cnx.org/content/m40632/latest/1.6.png/image>

The screenshot shows the SPSS Statistics Data Editor interface. The title bar reads "book practice data set.sav [DataSet1] - SPSS Statistics Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, and Help. The toolbar contains icons for file operations, editing, and analysis. A red arrow points to the "Data" menu option. The main window displays a data table with 36 rows and 13 columns. The first column is labeled "DISTRICT" and contains values from 1 to 36. The second column is "DFLCHART" with values "Y". The third column is "DPFEIERP" with values ranging from 46.20 to 66.30. The remaining columns are labeled with "DA011TA" followed by "9R", "OR", "11T", "11T", "11T", "A09R", "11T", and "A09R". The last column is "Recodedinstructionalgroups" with values from 1 to 3. The status bar at the bottom indicates "SPSS Statistics Processor is ready" and the system clock shows "2:33 PM".

	DISTRICT	DFLCHART	DPFEIERP	DA011TA9R	DA011TAOR	DE011TA11T	DE011TA211T	DE011TA311T	DF011TA09R	DF011TA11T	DH011TA09R	Recodedinstructionalgroups	VAR	VAR	VAR	VAR
1	3801	Y	61.80	3				
2	13801	Y	59.20	2				
3	14801	Y	57.00	40.00	41.00	2...	2...	3...	55.00	3...	29.00	2				
4	14802	Y	53.00	35.00	68.00	3...	7...	6...	40.00	5...	.	2				
5	14803	Y	57.50	73.00	.	.	.	5...	70.00	.	.	2				
6	14804	Y	50.60	.	71.00	.	.	.	5...	.	.	1				
7	15801	Y	36.00	28.00	24.00	.	.	1...	2...	38.00	3...	29.00	1			
8	15802	Y	49.40	46.00	26.00	4...	1...	4...	2...	36.00	2...	56.00	1			
9	15803	Y	58.10	2				
10	15805	Y	59.20	2				
11	15806	Y	44.30	46.00	72.00	4...	7...	3...	8...	58.00	7...	47.00	1			
12	15807	Y	51.40	38.00	32.00	.	2...	4...	2...	43.00	3...	42.00	1			
13	15808	Y	58.10	2				
14	15809	Y	50.20	1				
15	15811	Y	54.70	2				
16	15812	Y	60.10	58.00	27.00	.	.	9...	2...	60.00	2...	58.00	2			
17	15813	Y	71.10	3				
18	15814	Y	51.20	42.00	57.00	.	.	4...	6...	40.00	6...	42.00	1			
19	15815	Y	45.50	67.00	52.00	.	.	6...	5...	60.00	5...	67.00	1			
20	15816	Y	46.80	21.00	45.00	.	.	1...	4...	30.00	4...	20.00	1			
21	15817	Y	47.60	33.00	45.00	.	.	3...	4...	32.00	3...	32.00	1			
22	15819	Y	52.80	41.00	67.00	1...	5...	1...	6...	45.00	6...	44.00	2			
23	15820	Y	56.80	52.00	55.00	.	.	5...	5...	50.00	6...	42.00	2			
24	15822	Y	54.90	81.00	85.00	.	.	7...	8...	75.00	9...	81.00	2			
25	15823	Y	46.20	40.00	31.00	.	.	4...	2...	.	2...	40.00	1			
26	15824	Y	47.90	13.00	59.00	.	.	1...	5...	8.00	7...	14.00	1			
27	15825	Y	59.90	2				
28	15826	Y	42.60	1				
29	15827	Y	66.30	80.00	96.00	88.00	9...	67.00	3			
30	15828	Y	59.30	.	90.00	2				
31	15830	Y	65.80	71.00	82.00	.	.	7...	8...	42.00	7...	73.00	3			
32	15831	Y	55.60	2				
33	15832	Y				
34	15833	Y				
35	21803	Y	50.70	33.00	71.00	4...	.	3...	6...	38.00	.	29.00	1			

✓ Select Cases

¹¹<http://cnx.org/content/m40632/latest/1.7.png/image>



You will note that All Cases is checked. You should click on the If Condition is satisfied

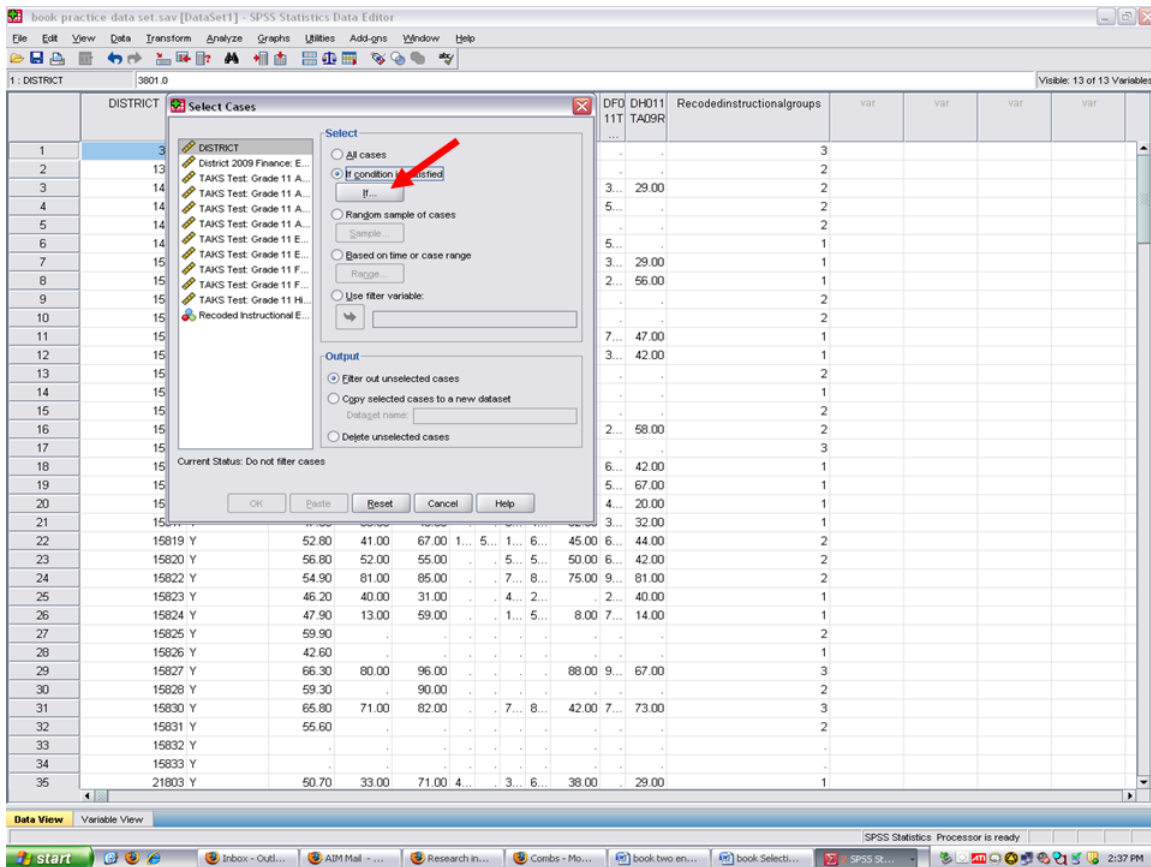
¹²<http://cnx.org/content/m40632/latest/1.9.png/image>

The screenshot displays the SPSS Statistics Data Editor interface. A 'Select Cases' dialog box is open, showing the 'If condition is satisfied' option selected. A red arrow points to the 'If...' button. The background shows a data table with columns for DISTRICT, DFO, DH011, and TA09R. The table contains 36 rows of data, with the first row having a DISTRICT value of 3 and a TA09R value of 3. The dialog box also shows options for 'Random sample of cases', 'Based on time or case range', and 'Use filter variable'. The 'Output' section has 'Filter out unselected cases' selected. The 'Current Status' is 'Do not filter cases'. The 'Data View' tab is active at the bottom.

13

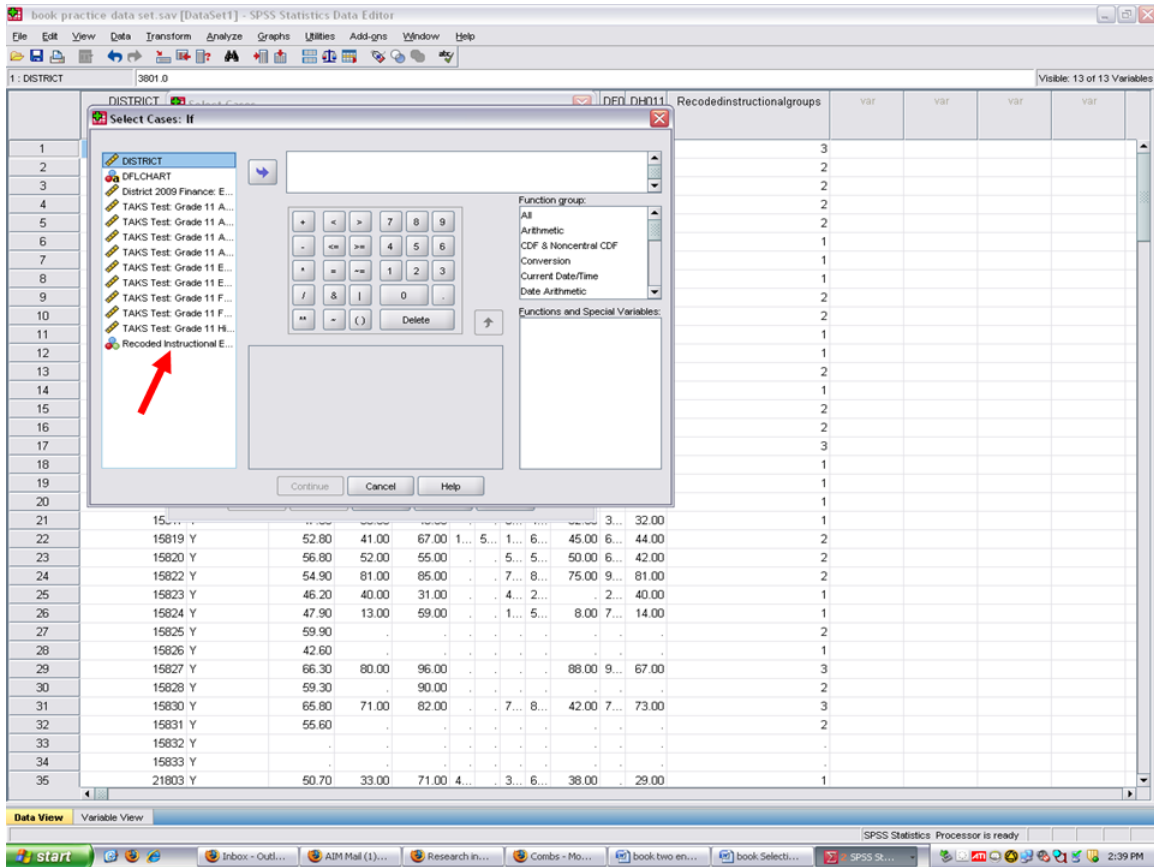
When you click on the If Condition is satisfied, the If button then becomes activated.
 ✓ Click on the If, under the If Condition is satisfied.

¹³<http://cnx.org/content/m40632/latest/1.10.png/image>



Clicking on the If opens up this screen. We will have to send over the name of the grouping variable, Recoded Instructional Expenditures into 3 groups.

¹⁴<http://cnx.org/content/m40632/latest/1.11.png/image>



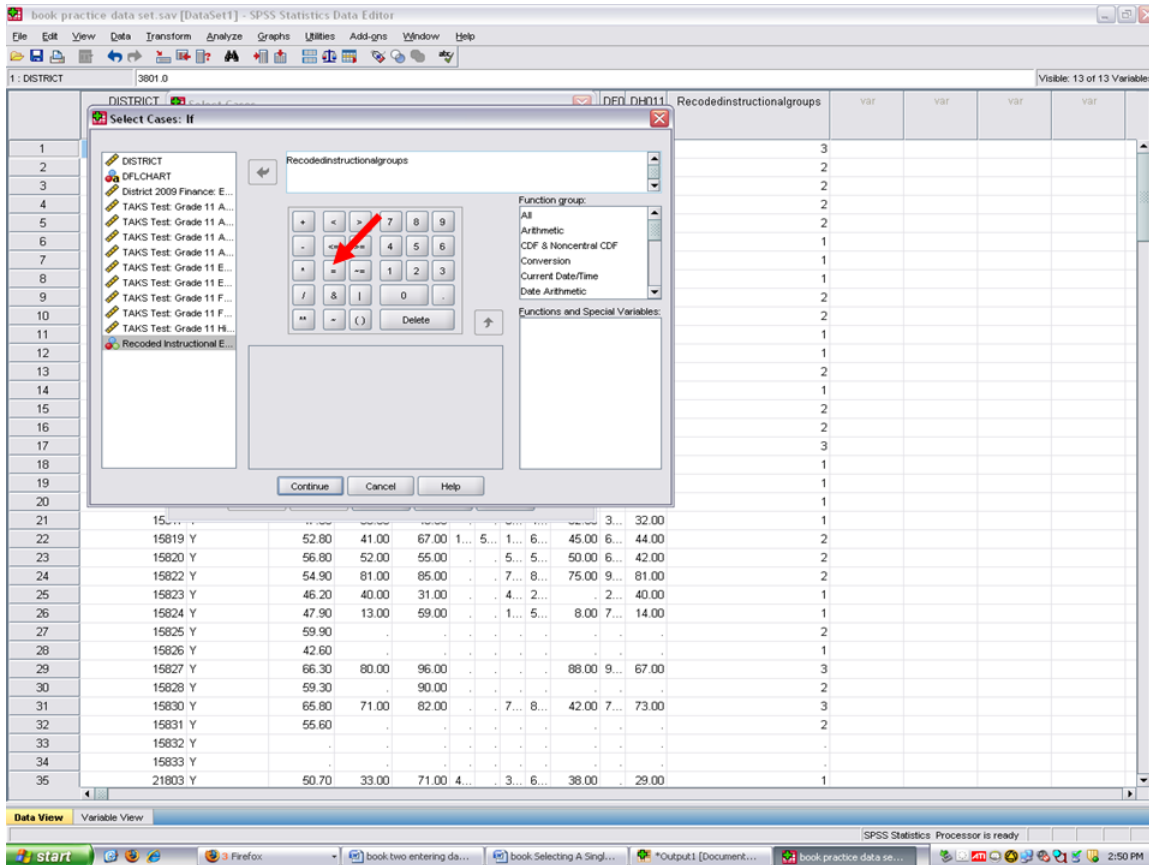
✓ Click on the Recoded Instructional Expenditures into 3 groups variable and then on the right arrow key.

¹⁵<http://cnx.org/content/m40632/latest/1.12.png/image>

Now the grouping variable has been placed into the empty rectangle, we can now begin the process of selecting the specific group of interest.

✓ Click on the = sign

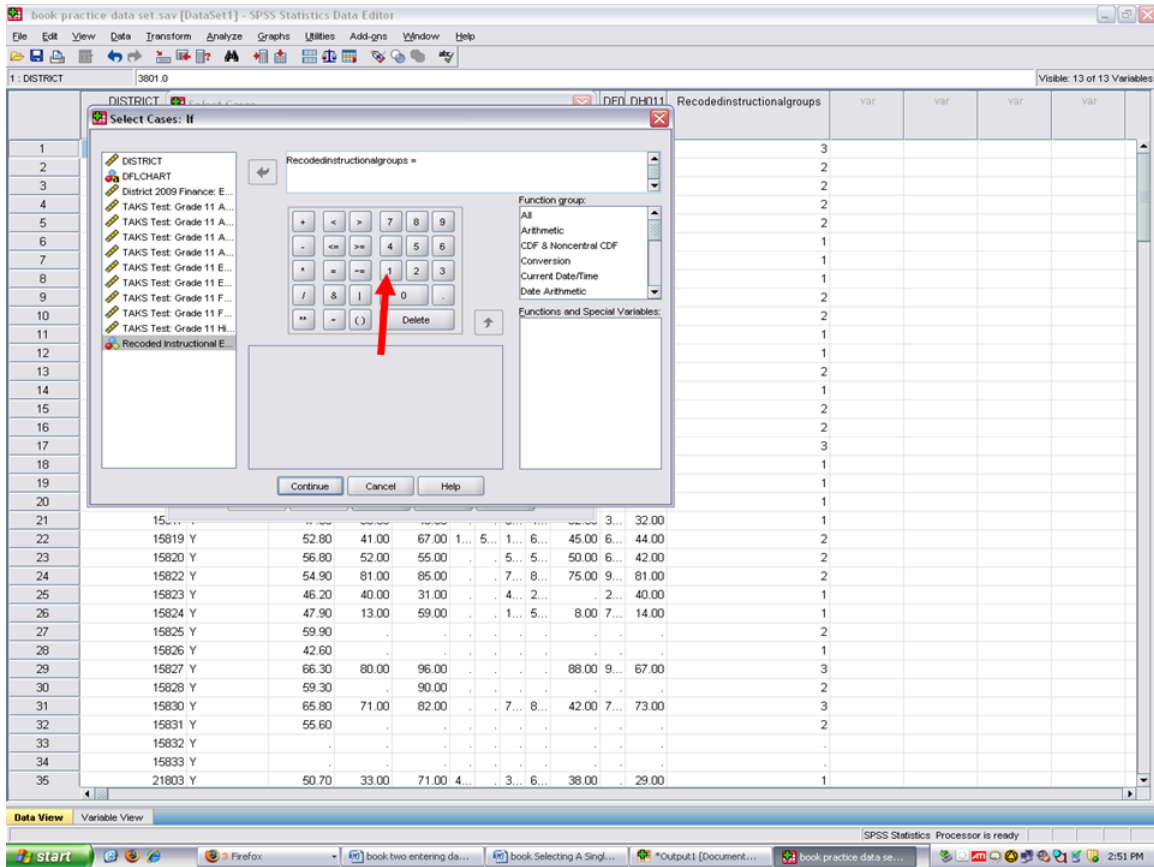
¹⁶<http://cnx.org/content/m40632/latest/1.13.png/image>



After clicking on the = sign, it appears to the right of the grouping variable in the rectangle.
 ✓ Click on the number of the group on which you want to calculate statistics. In our case, we want to use group # 1 (remember our previous steps in identifying the Poorest Financed School Districts).

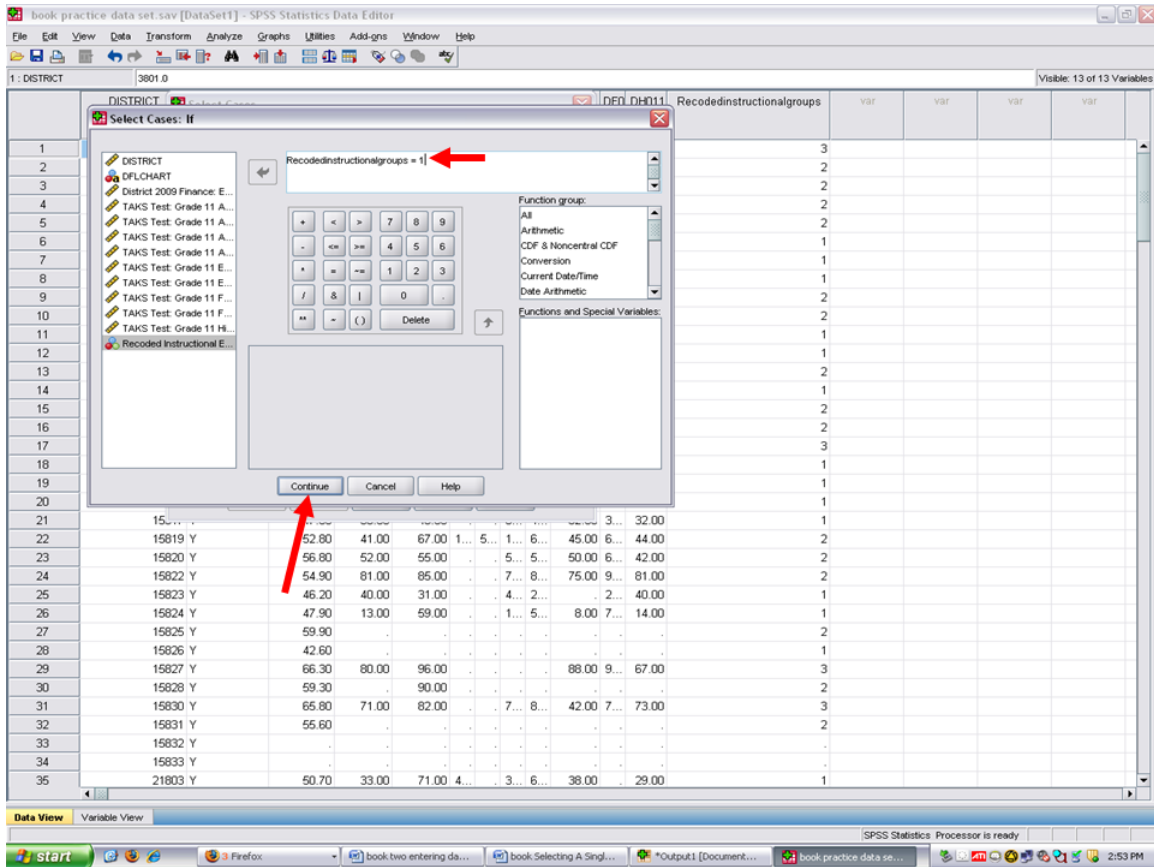
In this case we will click on 1.

¹⁷<http://cnx.org/content/m40632/latest/1.14.png/image>



We have now informed SPSS that we want to select out only those Poorest Financed School Districts for statistical analyses.

¹⁸<http://cnx.org/content/m40632/latest/1.15.png/image>



- ✓ Click on Continue above.
- ✓ After clicking on continue, this screen appears.
- ✓ Click on OK.

¹⁹<http://cnx.org/content/m40632/latest/1.16.png/image>

The screenshot shows the SPSS Statistics Data Editor interface. A 'Select Cases' dialog box is open, allowing the user to filter data based on a condition. The condition is set to 'If condition is satisfied' with the expression 'Recodedinstructionalgroups = 1'. The 'Output' section is set to 'Filter out unselected cases'. A red arrow points to the 'OK' button. The background shows a data table with columns for DISTRICT, DFO, DH011, TA09R, and Recodedinstructionalgroups.

DISTRICT	DFO	DH011	TA09R	Recodedinstructionalgroups				
1				3				
2				2				
3	3...	29.00		2				
4	5...			2				
5				2				
6	5...			1				
7	3...	29.00		1				
8	2...	56.00		1				
9				2				
10				2				
11	7...	47.00		1				
12	3...	42.00		1				
13				2				
14				1				
15				2				
16	2...	58.00		2				
17				3				
18	6...	42.00		1				
19	5...	67.00		1				
20	4...	20.00		1				
21	3...	32.00		1				
22	15819 Y	52.80	41.00	67.00	1... 5... 1... 6... 45.00	6... 44.00	2	
23	15820 Y	56.80	52.00	55.00	5... 5... 50.00	6... 42.00	2	
24	15822 Y	54.90	81.00	85.00	7... 8... 75.00	9... 81.00	2	
25	15823 Y	46.20	40.00	31.00	4... 2... 2... 40.00	1	1	
26	15824 Y	47.90	13.00	59.00	1... 5... 8.00	7... 14.00	1	
27	15825 Y	59.90					2	
28	15826 Y	42.60					1	
29	15827 Y	66.30	80.00	96.00		88.00	9... 67.00	3
30	15828 Y	59.30		90.00				2
31	15830 Y	65.80	71.00	82.00	7... 8... 42.00	7... 73.00	3	
32	15831 Y	55.60						2
33	15832 Y							
34	15833 Y							
35	21803 Y	50.70	33.00	71.00	4... 3... 6... 38.00	29.00	1	

20

You will now be sent to your output file. Click on the data file and data view.

In the data view screen, you will note that the Moderately Financed School Districts and the Best Financed School Districts have been marked through. The only active cases on which SPSS will calculate statistics are the Poorest Financed School Districts (group = 1).

²⁰<http://cnx.org/content/m40632/latest/1.17.png/image>

book practice data set.sav [DataSet1] - SPSS Statistics Data Editor

Visible: 14 of 14 Variables

	DISTRICT	DFLCHART	DPFEIERP	DA011TA09R	DA011TA0R	DE011TA09R	DE011TA0R	DE011TA09R	DE011TA0R	DF011TA09R	DF011TA0R	DH011TA09R	Recodedinstructionalgroups	filter_\$	VAR	VAR	VAR
1	3801	Y	61.00	3	0			
2	13801	Y	59.20	2	0			
3	14801	Y	57.00	40.00	41.00	2...	2...	3...	3...	55.00	3...	29.00	2	0			
4	14802	Y	53.00	35.00	68.00	3...	7...	6...	6...	40.00	5...	.	2	0			
5	14803	Y	57.50	73.00	.	.	.	5...	5...	70.00	.	.	2	0			
6	14804	Y	50.60	.	71.00	5...	.	.	1	1			
7	15801	Y	36.00	28.00	24.00	.	1...	2...	2...	38.00	3...	29.00	1	1			
8	15802	Y	49.40	46.00	26.00	4...	1...	4...	2...	36.00	2...	56.00	1	1			
9	15803	Y	58.10	2	0			
10	15805	Y	59.20	2	0			
11	15806	Y	44.30	46.00	72.00	4...	7...	3...	8...	58.00	7...	47.00	1	1			
12	15807	Y	51.40	38.00	32.00	.	2...	4...	2...	43.00	3...	42.00	1	1			
13	15808	Y	58.10	2	0			
14	15809	Y	50.20	1	1			
15	15811	Y	54.70	2	0			
16	15812	Y	60.10	58.00	27.00	.	.	9...	2...	60.00	2...	58.00	2	0			
17	15813	Y	71.10	3	0			
18	15814	Y	51.20	42.00	57.00	.	4...	6...	6...	40.00	6...	42.00	1	1			
19	15815	Y	45.50	67.00	52.00	.	6...	5...	5...	60.00	5...	67.00	1	1			
20	15816	Y	46.80	21.00	45.00	.	1...	4...	4...	30.00	4...	20.00	1	1			
21	15817	Y	47.60	33.00	45.00	.	3...	4...	4...	32.00	3...	32.00	1	1			
22	15819	Y	52.80	41.00	67.00	1...	5...	1...	6...	45.00	6...	44.00	2	0			
23	15820	Y	56.80	52.00	55.00	.	5...	5...	5...	50.00	6...	42.00	2	0			
24	15822	Y	54.90	81.00	85.00	.	7...	8...	8...	75.00	9...	81.00	2	0			
25	15823	Y	46.20	40.00	31.00	.	4...	2...	2...	40.00	.	.	1	1			
26	15824	Y	47.90	13.00	59.00	.	1...	5...	5...	8.00	7...	14.00	1	1			
27	15825	Y	59.90	2	0			
28	15826	Y	42.60	1	1			
29	15827	Y	66.30	80.00	96.00	88.00	9...	67.00	3	0			
30	15828	Y	59.30	.	90.00	2	0			
31	15830	Y	65.80	71.00	82.00	.	7...	8...	8...	42.00	7...	73.00	3	0			
32	15831	Y	55.60	2	0			
33	15832	Y			
34	15833	Y			
35	21803	Y	50.70	33.00	71.00	4...	.	3...	6...	38.00	.	29.00	1	1			

21

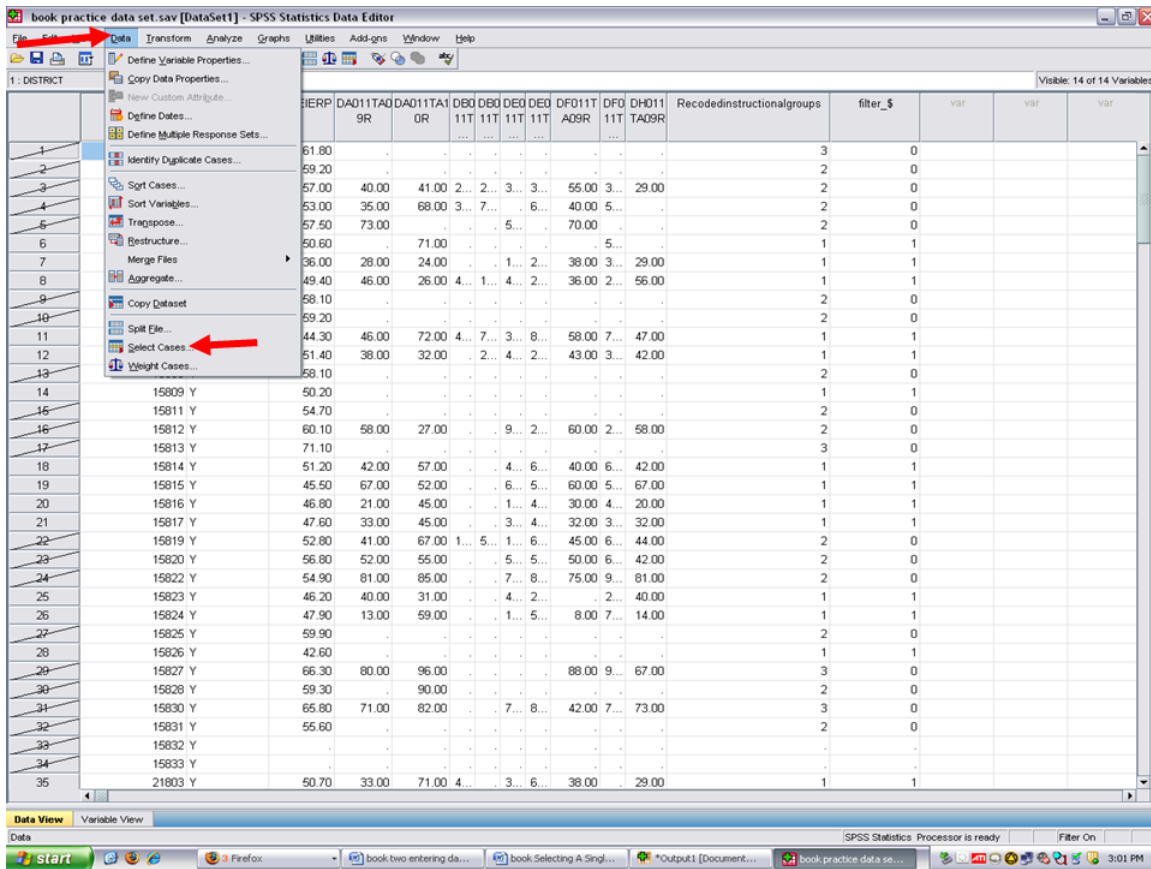
At this point, you can calculate descriptive statistics or other analyses. Such calculations will only have the Poorest Financed School Districts in them.

After obtaining the calculations of interest for this group, if you wish to select the Best Financed School Districts, you would follow the steps above in selecting group 1. Replace group = 1, with group = 2.

If you intend to analyze all cases (i.e., have all three groups back together), follow these steps.

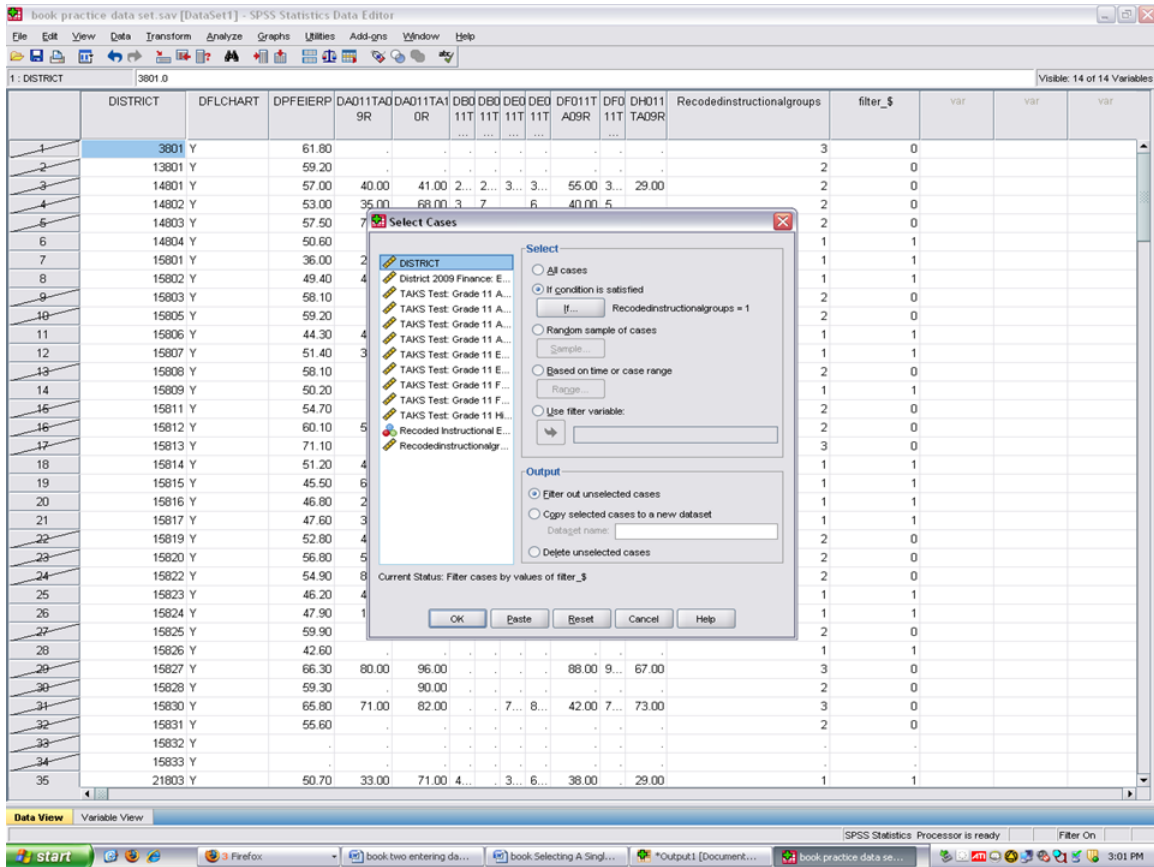
- ✓ Click on Data
- ✓ Click on Select Cases

²¹<http://cnx.org/content/m40632/latest/1.18.png/image>



After clicking on Select Cases, this screen will appear.

²²<http://cnx.org/content/m40632/latest/1.19.png/image>



✓ Click on All Cases

²³<http://cnx.org/content/m40632/latest/1.20.png/image>

The screenshot shows the SPSS Statistics Data Editor interface. A 'Select Cases' dialog box is open, centered over the data table. The 'Select' section of the dialog box has 'All cases' selected with a radio button. A red arrow points to this radio button. Below it, there are options for 'If condition is satisfied', 'Random sample of cases', 'Based on time or case range', and 'Use filter variable'. The 'Output' section has 'Filter out unselected cases' selected. At the bottom of the dialog box, the 'OK' button is highlighted with a red arrow. The background data table shows columns for DISTRICT, DFLCHART, DPFEIERP, and various test scores (DA011TA, DA011TA1, etc.).

24

✓ Click on OK above.
 You will now note that no cases are crossed through. SPSS will now use all three groups in its analyses.

²⁴<http://cnx.org/content/m40632/latest/1.21.png/image>

CHAPTER 2. 2. SELECTING A SINGLE GROUP FOR STATISTICAL ANALYSIS

The screenshot shows the SPSS Statistics Data Editor window with a data table. The table has the following columns: DISTRICT, DFLCHART, DPFEIERP, DA011TA9R, DA011TA0R, DE011TA1, DE011TA2, DE011TA3, DE011TA4, DE011TA5, DE011TA6, DE011TA7, DE011TA8, DE011TA9, DF011TA09R, DF011TA11T, DF011TA09R, Recodedinstructionalgroups, filter_\$, and three unlabeled variables. The data rows are numbered 1 through 35. The first row (row 1) has a DISTRICT value of 3801 and a filter_\$ value of 0. The last row (row 35) has a DISTRICT value of 21803 and a filter_\$ value of 1.

	DISTRICT	DFLCHART	DPFEIERP	DA011TA9R	DA011TA0R	DE011TA1	DE011TA2	DE011TA3	DE011TA4	DE011TA5	DE011TA6	DE011TA7	DE011TA8	DE011TA9	DF011TA09R	DF011TA11T	DF011TA09R	Recodedinstructionalgroups	filter_\$				
1	3801	Y	61.80	3	0				
2	13801	Y	59.20	2	0				
3	14801	Y	57.00	40.00	41.00	2...	2...	3...	3...	55.00	3...	29.00	2	0				
4	14802	Y	53.00	35.00	68.00	3...	7...	6...	6...	40.00	5...	2	0				
5	14803	Y	57.50	73.00	5...	70.00	2	0				
6	14804	Y	50.60	.	71.00	5...	1	1				
7	15801	Y	36.00	28.00	24.00	.	.	1...	2...	38.00	3...	29.00	1	1				
8	15802	Y	49.40	46.00	26.00	4...	1...	4...	2...	36.00	2...	56.00	1	1				
9	15803	Y	58.10	2	0				
10	15805	Y	59.20	2	0				
11	15806	Y	44.30	46.00	72.00	4...	7...	3...	8...	58.00	7...	47.00	1	1				
12	15807	Y	51.40	38.00	32.00	.	2...	4...	2...	43.00	3...	42.00	1	1				
13	15808	Y	58.10	2	0				
14	15809	Y	50.20	1	1				
15	15811	Y	54.70	2	0				
16	15812	Y	60.10	58.00	27.00	.	.	9...	2...	60.00	2...	58.00	2	0				
17	15813	Y	71.10	3	0				
18	15814	Y	51.20	42.00	57.00	.	.	4...	6...	40.00	6...	42.00	1	1				
19	15815	Y	45.50	67.00	52.00	.	.	6...	5...	60.00	5...	67.00	1	1				
20	15816	Y	46.80	21.00	45.00	.	.	1...	4...	30.00	4...	20.00	1	1				
21	15817	Y	47.60	33.00	45.00	.	.	3...	4...	32.00	3...	32.00	1	1				
22	15819	Y	52.80	41.00	67.00	1...	5...	1...	6...	45.00	6...	44.00	2	0				
23	15820	Y	56.80	52.00	55.00	.	.	5...	5...	50.00	6...	42.00	2	0				
24	15822	Y	54.90	81.00	85.00	.	.	7...	8...	75.00	9...	81.00	2	0				
25	15823	Y	46.20	40.00	31.00	.	.	4...	2...	.	2...	40.00	1	1				
26	15824	Y	47.90	13.00	59.00	.	.	1...	5...	8.00	7...	14.00	1	1				
27	15825	Y	59.90	2	0				
28	15826	Y	42.60	1	1				
29	15827	Y	66.30	80.00	96.00	88.00	9...	67.00	3	0				
30	15828	Y	59.30	.	90.00	2	0				
31	15830	Y	65.80	71.00	82.00	.	.	7...	8...	42.00	7...	73.00	3	0				
32	15831	Y	55.60	2	0				
33	15832	Y
34	15833	Y
35	21803	Y	50.70	33.00	71.00	4...	3...	6...	38.00	.	29.00	1	1				

²⁵<http://cnx.org/content/m40632/latest/1.22.png/image>

Chapter 3

3. Steps Involved in Recoding Data in SPSS¹



NOTE: This chapter is published by NCPEA Press² and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of education administration.

About the Authors

John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website (Writing and Statistical Help³) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

Ana Rojas-LeBouef is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. Dr. LeBoeuf recently completed her doctoral degree in Reading, where she conducted a 16-year analysis of Texas statewide data regarding the achievement gap. Her research interests lie in examining the inequities in achievement among ethnic groups. Dr. Rojas-LeBouef also assists students and faculty in their writing and statistical needs on the Writing and Statistical Help website.

About the Editors

¹This content is available online at <<http://cnx.org/content/m40708/1.2/>>.

²<http://www.ncpeapublications.org/books.html>

³<http://cnx.org/content/m40708/latest/www.writingandstatisticalhelp>

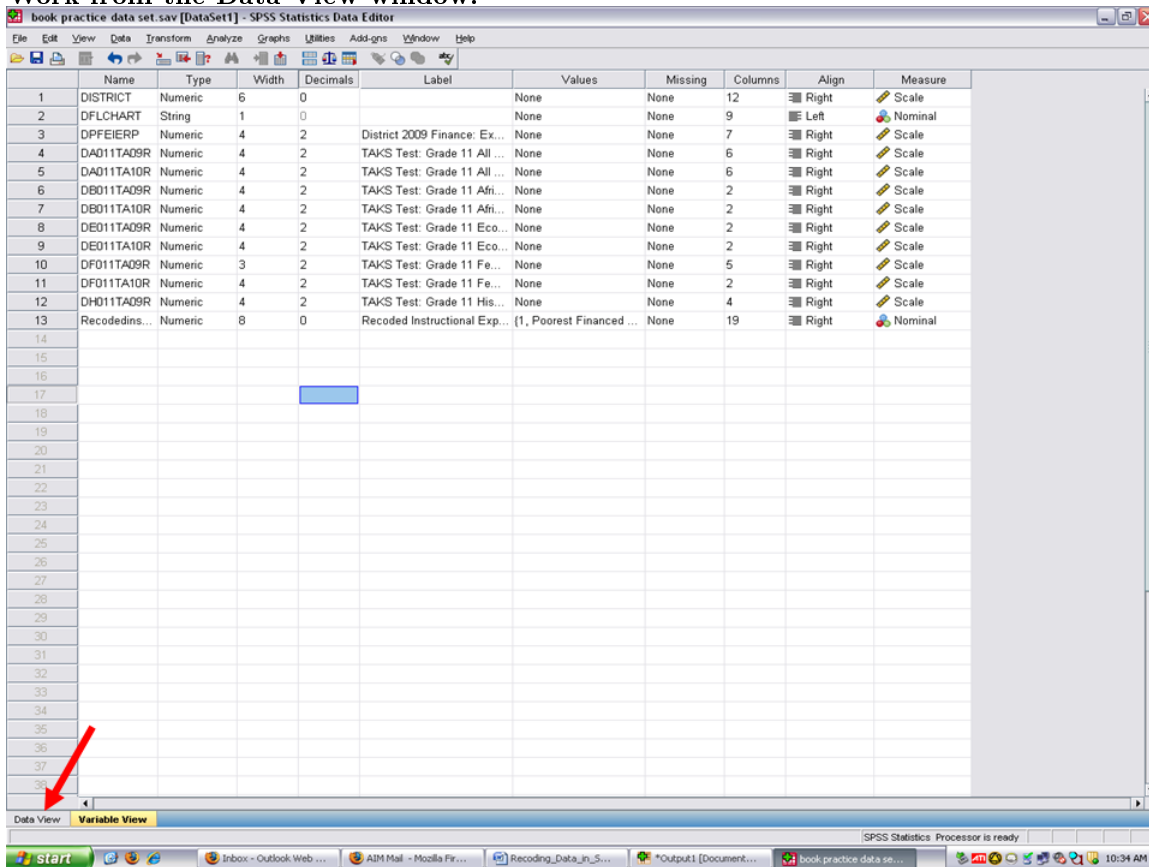
Theodore B. Creighton, is a Professor at Virginia Tech and the Publications Director for NCPEA Publications⁴, the Founding Editor of Education Leadership Review,⁵ and the Senior Editor of the NCPEA Connexions Project.

Brad E. Bizzell, is a recent graduate of the Virginia Tech Doctoral Program in Educational Leadership and Policy Studies, and is a School Improvement Coordinator for the Virginia Tech Training and Technical Assistance Center. In addition, Dr. Bizzell serves as an Assistant Editor of the NCPEA Connexions Project in charge of technical formatting and design.

Janet Tareilo, is a Professor at Stephen F. Austin State University and serves as the Assistant Director of NCPEA Publications. Dr. Tareilo also serves as an Assistant Editor of the NCPEA Connexions Project and as a editor and reviewer for several national and international journals in educational leadership.

These procedures are to be followed when your data are already entered into SPSS and when you have values that you need to either remove or to recode into different values.

Work from the Data View window.



After clicking on the Data View button, your screen should look like the screen below in that you are looking at data.

⁴<http://www.ncpeapublications.org>

⁵<http://ncpeapublications.org/about-elr.html>

⁶<http://cnx.org/content/m40708/latest/2.1.png/image>

The screenshot shows the SPSS Statistics Data Editor window with a data table. The table has the following columns: DISTRICT, DFLCHART, DPFEIERP, DA011TA9R, DA011TA0R, D0011T11T, D0011T21T, D0011T31T, D0011T41T, D0011T51T, D0011T61T, D0011T71T, D0011T81T, D0011T91T, D0011T01T, D0011T10T, D0011T20T, D0011T30T, D0011T40T, D0011T50T, D0011T60T, D0011T70T, D0011T80T, D0011T90T, Recodedinstructionalgroups, VAR, VAR, VAR, VAR, VAR. The 'Recodedinstructionalgroups' column contains values 1, 2, and 3. A red arrow points to the 'Recodedinstructionalgroups' column.

	DISTRICT	DFLCHART	DPFEIERP	DA011TA9R	DA011TA0R	D0011T11T	D0011T21T	D0011T31T	D0011T41T	D0011T51T	D0011T61T	D0011T71T	D0011T81T	D0011T91T	D0011T01T	D0011T10T	D0011T20T	D0011T30T	D0011T40T	D0011T50T	D0011T60T	D0011T70T	D0011T80T	D0011T90T	Recodedinstructionalgroups	VAR	VAR	VAR	VAR	VAR
1	3801	Y	61.80	3					
2	13801	Y	59.20	2					
3	14801	Y	57.00	40.00	41.00	2...	2...	3...	3...	55.00	3...	29.00	2						
4	14802	Y	53.00	35.00	68.00	3...	7...	6...	6...	40.00	5...	2						
5	14803	Y	57.50	73.00	5...	70.00	2						
6	14804	Y	50.60	.	71.00	5...	1						
7	15801	Y	36.00	28.00	24.00	.	.	1...	2...	38.00	3...	29.00	1						
8	15802	Y	49.40	46.00	26.00	4...	1...	4...	2...	36.00	2...	56.00	1						
9	15803	Y	58.10	2						
10	15805	Y	59.20	2						
11	15806	Y	44.30	46.00	72.00	4...	7...	3...	8...	58.00	7...	47.00	1						
12	15807	Y	51.40	38.00	32.00	.	2...	4...	2...	43.00	3...	42.00	1						
13	15808	Y	58.10	2						
14	15809	Y	50.20	1						
15	15811	Y	54.70	2						
16	15812	Y	60.10	58.00	27.00	.	.	9...	2...	60.00	2...	58.00	2						
17	15813	Y	71.10	3						
18	15814	Y	51.20	42.00	57.00	.	.	4...	6...	40.00	6...	42.00	1						
19	15815	Y	45.50	67.00	52.00	.	.	6...	5...	60.00	5...	67.00	1						
20	15816	Y	46.80	21.00	45.00	.	.	1...	4...	30.00	4...	20.00	1						
21	15817	Y	47.60	33.00	45.00	.	.	3...	4...	32.00	3...	32.00	1						
22	15819	Y	52.80	41.00	67.00	1...	5...	1...	6...	45.00	6...	44.00	2						
23	15820	Y	56.80	52.00	55.00	.	.	5...	5...	50.00	6...	42.00	2						
24	15822	Y	54.90	81.00	85.00	.	.	7...	8...	75.00	9...	81.00	2						
25	15823	Y	46.20	40.00	31.00	.	.	4...	2...	.	2...	40.00	1						
26	15824	Y	47.90	13.00	59.00	.	.	1...	5...	8.00	7...	14.00	1						
27	15825	Y	59.90	2						
28	15826	Y	42.60	1						
29	15827	Y	66.30	80.00	96.00	88.00	9...	67.00	3						
30	15828	Y	59.30	.	90.00	2						
31	15830	Y	65.80	71.00	82.00	.	.	7...	8...	42.00	7...	73.00	3						
32	15831	Y	55.60	2						
33	15832	Y					
34	15833	Y					
35	21803	Y	50.70	33.00	71.00	4...	3...	6...	38.00	.	29.00	1						

Identify the variable in which values are present that you want to recode.
 In our case, we want to recode the values for Question One below.

⁷ <http://cnx.org/content/m40708/latest/2.2.png/image>

The screenshot shows the SPSS Statistics Data Editor interface. The main window displays a data table with the following columns: DISTRICT, DPFEIERP, DFO11T, DH011TA09R, Recodedinstructio nalgroups, QuestionOne, and several unlabeled VAR columns. The 'QuestionOne' column header is highlighted with a red arrow. The data rows are numbered 1 through 35. The bottom status bar indicates 'SPSS Statistics Processor is ready' and the system tray shows the time as 10:42 AM.

	DISTRICT	DPFEIERP	DFO11T	DH011TA09R	Recodedinstructio nalgroups	QuestionOne	VAR	VAR	VAR	VAR	VAR	VAR	VAR	VAR
1	3801	61.80	.	.	3	3								
2	13801	59.20	.	.	2	3								
3	14801	57.00	3...	29.00	2	3								
4	14802	53.00	5...	.	2	2								
5	14803	57.50	.	.	2	3								
6	14804	50.60	5...	.	1	2								
7	15801	36.00	3...	29.00	1	1								
8	15802	49.40	2...	56.00	1	2								
9	15803	58.10	.	.	2	3								
10	15805	59.20	.	.	2	3								
11	15806	44.30	7...	47.00	1	2								
12	15807	51.40	3...	42.00	1	2								
13	15808	58.10	.	.	2	3								
14	15809	50.20	.	.	1	2								
15	15811	54.70	.	.	2	2								
16	15812	60.10	2...	58.00	2	3								
17	15813	71.10	.	.	3	5								
18	15814	51.20	6...	42.00	1	2								
19	15815	45.50	5...	67.00	1	2								
20	15816	46.80	4...	20.00	1	2								
21	15817	47.60	3...	32.00	1	2								
22	15819	52.80	6...	44.00	2	2								
23	15820	56.80	6...	42.00	2	3								
24	15822	54.90	9...	81.00	2	2								
25	15823	46.20	2...	40.00	1	2								
26	15824	47.90	7...	14.00	1	2								
27	15825	59.90	.	.	2	3								
28	15826	42.60	.	.	1	2								
29	15827	66.30	9...	67.00	3	4								
30	15828	59.30	.	.	2	3								
31	15830	65.80	7...	73.00	3	4								
32	15831	55.60	.	.	2	3								
33	15832								
34	15833								
35	21803	50.70	.	29.00	1	2								

Identify the values within this variable that you wish to recode. You may need to run the frequencies procedure to list out all of the values present for the variable of interest.

- ✓ Click on Analyze
- ✓ Click on Descriptive Statistics
- ✓ Click on Frequencies

⁸ <http://cnx.org/content/m40708/latest/2.3.png/image>

The screenshot shows the SPSS Statistics Data Editor interface. The 'Analyze' menu is open, and 'Frequencies...' is selected. The data table below shows the following columns: DISTRICT, D, and several variables labeled VAR. The data rows are numbered 1 through 35.

	DISTRICT	D	VAR	VAR	VAR	VAR	VAR	VAR	VAR	VAR	VAR	VAR	VAR
1	3801												
2	13801												
3	14801												
4	14802												
5	14803												
6	14804												
7	15801												
8	15802												
9	15803												
10	15805	59.20											
11	15806	44.30	7...	47.00									
12	15807	51.40	3...	42.00									
13	15808	58.10											
14	15809	50.20											
15	15811	54.70											
16	15812	60.10	2...	58.00									
17	15813	71.10											
18	15814	51.20	6...	42.00									
19	15815	45.50	5...	67.00									
20	15816	46.80	4...	20.00									
21	15817	47.60	3...	32.00									
22	15819	52.80	6...	44.00									
23	15820	56.80	6...	42.00									
24	15822	54.90	9...	81.00									
25	15823	46.20	2...	40.00									
26	15824	47.90	7...	14.00									
27	15825	59.90											
28	15826	42.60											
29	15827	66.30	9...	67.00									
30	15828	59.30											
31	15830	65.80	7...	73.00									
32	15831	55.60											
33	15832												
34	15833												
35	21803	50.70		29.00									

Highlight the variable on which you will request frequencies so that you can then determine the values that you will need to change/recode. In this example, we are using Survey Question One.

⁹<http://cnx.org/content/m40708/latest/2.4.png/image>

The screenshot shows the SPSS Statistics Data Editor interface. The main window displays a data table with columns: DISTRICT, DPFEIERP, DFD111T, DH011TA09R, Recodedinstructionalgroups, and QuestionOne. The 'Frequencies' dialog box is open, showing a list of variables on the left and an empty 'Variable(s):' box on the right. A red arrow points to the 'Survey Question One' variable in the list. The dialog box also includes buttons for 'Statistics...', 'Charts...', and 'Format...', and a checked option for 'Display frequency tables'.

	DISTRICT	DPFEIERP	DFD111T	DH011TA09R	Recodedinstructionalgroups	QuestionOne	VAR	VAR	VAR	VAR	VAR	VAR	VAR	VAR
1	3801	61.80	.	.	.	3	3							
2	13801	59.20	.	.	.	2	3							
3	14801	57.00	3...	29.00	.	2	3							
4	14802	53.00	5...	.	.	2	2							
5	14803	57.50	.	.	.	2	3							
6	14804	50.60	5...	.	.	1	2							
7	15801	36.00	3...	29.00	.	1	1							
8	15802	49.40	2...	56.00	.	1	2							
9	15803	58.10	.	.	.	2	3							
10	15805	59.20							
11	15806	44.30	7...	47.00	.	.	.							
12	15807	51.40	3...	42.00	.	.	.							
13	15808	58.10							
14	15809	50.20							
15	15811	54.70							
16	15812	60.10	2...	58.00	.	.	.							
17	15813	71.10							
18	15814	51.20	6...	42.00	.	.	.							
19	15815	45.50	5...	67.00	.	.	.							
20	15816	46.80	4...	20.00	.	.	.							
21	15817	47.60	3...	32.00	.	.	.							
22	15819	52.80	6...	44.00	.	.	.							
23	15820	56.80	6...	42.00	.	2	3							
24	15822	54.90	9...	81.00	.	2	2							
25	15823	46.20	2...	40.00	.	1	2							
26	15824	47.90	7...	14.00	.	1	2							
27	15825	59.90	.	.	.	2	3							
28	15826	42.60	.	.	.	1	2							
29	15827	66.30	9...	67.00	.	3	4							
30	15828	59.30	.	.	.	2	3							
31	15830	65.80	7...	73.00	.	3	4							
32	15831	55.60	.	.	.	2	3							
33	15832							
34	15833							
35	21803	50.70	.	29.00	.	1	2							

✓ Click on the arrow in the middle to send over the Survey Question One variable.

¹⁰<http://cnx.org/content/m40708/latest/2.5.png/image>

The screenshot shows the SPSS Statistics Data Editor window with a data table. The table has columns for DISTRICT, DPFEIERP, DFO, DH011, Recoded instructional groups, and QuestionOne. The 'Frequencies' dialog box is open, showing 'Survey Question One' selected in the 'Variable(s)' list. The 'Display frequency tables' checkbox is checked. The 'OK' button is highlighted with a red arrow.

	DISTRICT	DPFEIERP	DFO	DH011	Recoded instructional groups	QuestionOne	VAR	VAR	VAR	VAR	VAR	VAR	VAR	VAR
1	3801	61.80	.	.	.	3	3							
2	13801	59.20	.	.	.	2	3							
3	14801	57.00	3...	29.00	.	2	3							
4	14802	53.00	5...	.	.	2	2							
5	14803	57.50	.	.	.	2	3							
6	14804	50.60	5...	.	.	1	2							
7	15801	36.00	3...	29.00	.	1	1							
8	15802	49.40	2...	56.00	.	1	2							
9	15803	58.10	.	.	.	2	3							
10	15805	59.20							
11	15806	44.30	7...	47.00	.	.	.							
12	15807	51.40	3...	42.00	.	.	.							
13	15808	58.10							
14	15809	50.20							
15	15811	54.70							
16	15812	60.10	2...	58.00	.	.	.							
17	15813	71.10							
18	15814	51.20	6...	42.00	.	.	.							
19	15815	45.50	5...	67.00	.	.	.							
20	15816	46.80	4...	20.00	.	.	.							
21	15817	47.60	3...	32.00	.	.	.							
22	15819	52.80	6...	44.00	.	.	.							
23	15820	56.80	6...	42.00	.	2	3							
24	15822	54.90	9...	81.00	.	2	2							
25	15823	46.20	2...	40.00	.	1	2							
26	15824	47.90	7...	14.00	.	1	2							
27	15825	59.90	.	.	.	2	3							
28	15826	42.60	.	.	.	1	2							
29	15827	66.30	9...	67.00	.	3	4							
30	15828	59.30	.	.	.	2	3							
31	15830	65.80	7...	73.00	.	3	4							
32	15831	55.60	.	.	.	2	3							
33	15832							
34	15833							
35	21803	50.70	.	29.00	.	1	2							

Because we want a listing of the values present for this variable, Survey Question One, we will click on OK.

Go to the output screen. For this example, our output looks like the following. Survey Question One has five different values (i.e., 1, 2, 3, 4, 5). The value of 1 only occurred 6 times. These values could represent a 5-point Likert format of 1 = Strongly Disagree, 2 = Disagree, 3 = Neutral/Undecided, 4 = Agree, and 5 = Strongly Agree. [Note. Readers are referred to a separate chapter on typing in survey data with appropriate labeling.]

¹¹<http://cnx.org/content/m40708/latest/2.6.png/image>

Frequencies

[DataSet1] E:\book practice data set.sav

Statistics

Survey Question One

N	Valid	201
	Missing	6

Survey Question One

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1.0	6	2.9	3.0	3.0
2.0	81	39.1	40.3	43.3
3.0	64	30.9	31.8	75.1
4.0	25	12.1	12.4	87.6
5.0	25	12.1	12.4	100.0
Total	201	97.1	100.0	
Missing System	6	2.9		
Total	207	100.0		

Based upon our frequencies, we will combine values 1 and 2 into a Disagreement category and 4 and 5 into an Agreement category.

✓ Click on **Transform**.

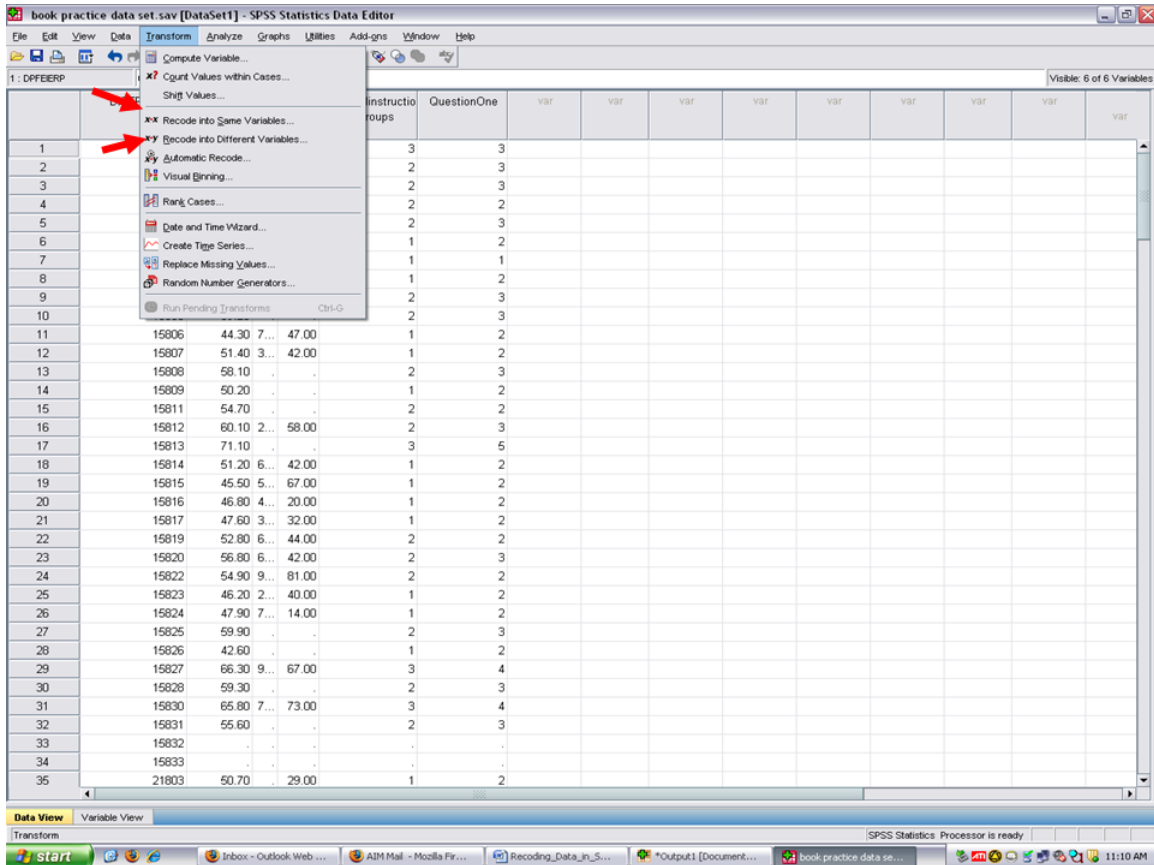
¹²<http://cnx.org/content/m40708/latest/2.7.png/image>

The screenshot shows the SPSS Statistics Data Editor interface. The 'Transform' menu is open, displaying options such as 'Compute Variable...', 'Count Values within Cases...', 'Shift Values...', 'Recode into Same Variables...', 'Recode into Different Variables...', 'Automatic Recode...', 'Visual Binning...', 'Rank Cases...', 'Date and Time Wizard...', 'Create Time Series...', 'Replace Missing Values...', and 'Random Number Generators...'. A red arrow points to the 'Recode into Different Variables...' option. The data table below shows columns for 'instructio', 'QuestionOne', and several 'VAR' columns. The data rows are numbered 1 through 35.

	DISTR	instructio	QuestionOne	VAR	VAR	VAR	VAR	VAR	VAR	VAR	VAR	VAR	VAR
1		3	3										
2		2	3										
3		2	3										
4		2	2										
5		2	3										
6		1	2										
7		1	1										
8		1	2										
9		2	3										
10		2	3										
11		1	2										
12		1	2										
13		2	3										
14		1	2										
15		2	2										
16		2	3										
17		3	5										
18		1	2										
19		1	2										
20		1	2										
21		1	2										
22		2	2										
23		2	3										
24		2	2										
25		1	2										
26		1	2										
27		2	3										
28		1	2										
29		3	4										
30		2	3										
31		3	4										
32		2	3										
33		.	.										
34		.	.										
35		1	2										

You now have an option to Recode Into Same Variables or Recode Into Different Variables. We recommend that you recode into different variables so that the data present in the variable of Question One is left untouched.

¹³<http://cnx.org/content/m40708/latest/2.8.png/image>



When you click on Recode Into Different Variables, you will then see this screen.

¹⁴<http://cnx.org/content/m40708/latest/2.9.png/image>

The screenshot shows the SPSS Statistics Data Editor interface. The main window displays a data view with the following columns: DISTRICT, DPFIEERP, DFO, DH011, Recodedinstructionalgroups, QuestionOne, and several unlabeled variables (VAR). A red arrow points to the 'Recode into Different Variables' dialog box, which is currently open. The dialog box has a list of variables on the left, including DISTRICT, District 2009 Finance: E..., TAKS Test: Grade 11 F..., TAKS Test: Grade 11 H..., Recoded Instructional E..., and Survey Question One [...]. The 'Input Variable -> Output Variable' box is empty. The 'Output Variable' section has fields for Name and Label, and a 'Change' button. The 'If...' button is also visible. The background data view shows rows of data with values for each variable.

Case	DISTRICT	DPFIEERP	DFO	DH011	Recodedinstructionalgroups	QuestionOne	VAR	VAR	VAR	VAR	VAR	VAR	VAR	VAR
1														
2														
3														
4														
5														
6														
7														
8														
9														
10														
11														
12														
13														
14														
15														
16														
17	15813	71.10	.	.		3	5							
18	15814	51.20	6...	42.00		1	2							
19	15815	45.50	5...	67.00		1	2							
20	15816	46.80	4...	20.00		1	2							
21	15817	47.60	3...	32.00		1	2							
22	15819	52.80	6...	44.00		2	2							
23	15820	56.80	6...	42.00		2	3							
24	15822	54.90	9...	81.00		2	2							
25	15823	46.20	2...	40.00		1	2							
26	15824	47.90	7...	14.00		1	2							
27	15825	59.90	.	.		2	3							
28	15826	42.60	.	.		1	2							
29	15827	66.30	9...	67.00		3	4							
30	15828	59.30	.	.		2	3							
31	15830	65.80	7...	73.00		3	4							
32	15831	55.60	.	.		2	3							
33	15832							
34	15833							
35	21803	50.70	.	29.00		1	2							

Highlight the variable on which you will be recoding the values. In this example, our variable is Survey Question One. After highlighting it, click on the arrow to send it to the Input Variable – Output Variable box.

¹⁵<http://cnx.org/content/m40708/latest/2.10.png/image>

The screenshot displays the SPSS Statistics Data Editor interface. A dialog box titled "Recode into Different Variables" is open, showing the process of recoding a variable. The "Input Variable -> Output Variable" list contains "Survey Question One". The "Output Variable" section has "Name:" and "Label:" fields. A red arrow points to the "Name:" field. The background data view shows columns for DISTRICT, DPFIEERP, DFD, DH011, Recodedinstructionalgroups, and QuestionOne. The bottom status bar shows "SPSS Statistics Processor is ready".

	DISTRICT	DPFIEERP	DFD	DH011	Recodedinstructionalgroups	QuestionOne	VAR	VAR	VAR	VAR	VAR	VAR	VAR	VAR
1														
2														
3														
4														
5														
6														
7														
8														
9														
10														
11														
12														
13														
14														
15														
16														
17	15813	71.10	.	.		3	5							
18	15814	51.20	6...	42.00		1	2							
19	15815	45.50	5...	67.00		1	2							
20	15816	46.80	4...	20.00		1	2							
21	15817	47.60	3...	32.00		1	2							
22	15819	52.80	6...	44.00		2	2							
23	15820	56.80	6...	42.00		2	3							
24	15822	54.90	9...	81.00		2	2							
25	15823	46.20	2...	40.00		1	2							
26	15824	47.90	7...	14.00		1	2							
27	15825	59.90	.	.		2	3							
28	15826	42.60	.	.		1	2							
29	15827	66.30	9...	67.00		3	4							
30	15828	59.30	.	.		2	3							
31	15830	65.80	7...	73.00		3	4							
32	15831	55.60	.	.		2	3							
33	15832							
34	15833							
35	21803	50.70	.	29.00		1	2							

Click under the Output Variable where Name is present.

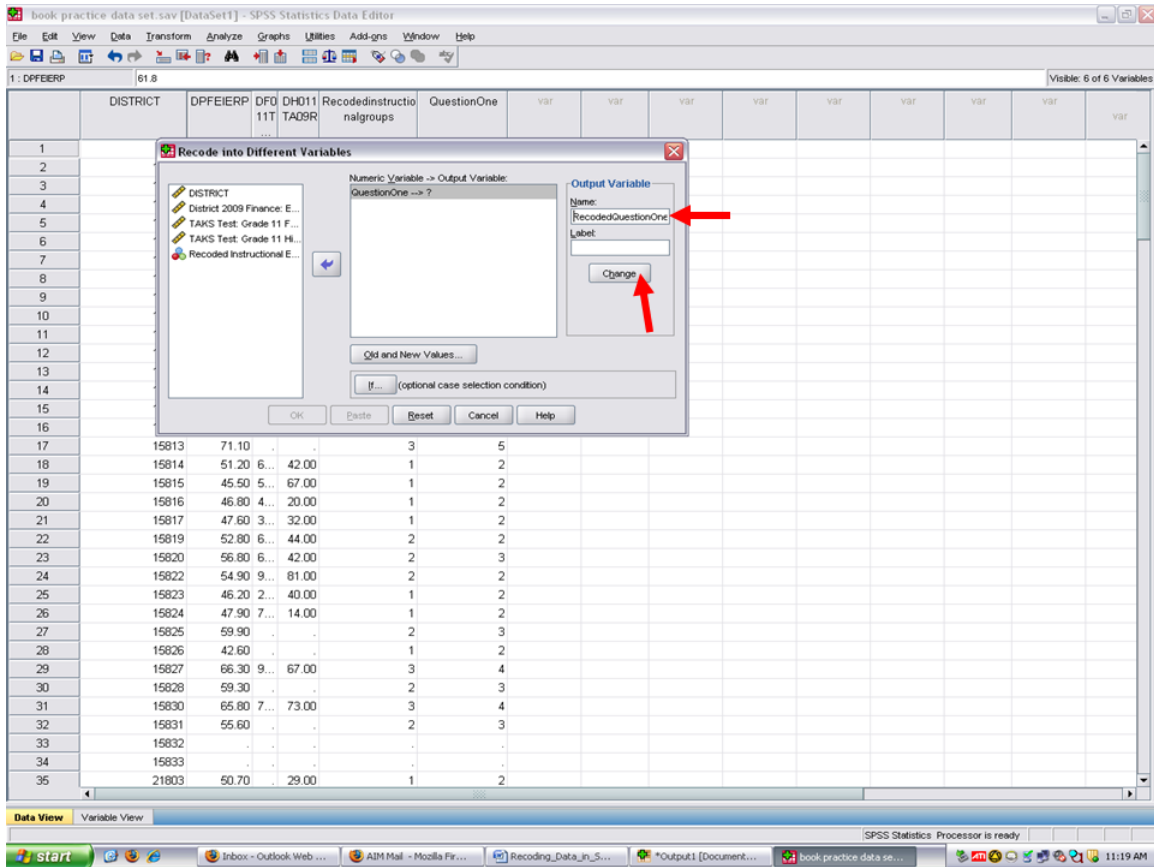
¹⁶<http://cnx.org/content/m40708/latest/2.11.png/image>

The screenshot shows the SPSS Statistics Data Editor interface. A dialog box titled "Recode into Different Variables" is open, centered over a data grid. The dialog box has a "Numeric Variable -> Output Variable" section where "QuestionOne --> ?" is selected. To the right, the "Output Variable" section has a "Name:" field with a red arrow pointing to it. Below the "Name:" field are "Label:" and "Change" fields. At the bottom of the dialog box are "OK", "Paste", "Reset", "Cancel", and "Help" buttons. The data grid in the background shows columns for DISTRICT, DPFIEERP, DFD, DH011, Recodedinstructionalgroups, and QuestionOne. The data rows are numbered 1 to 36.

Now type in a new name for the variable you will be creating. Remember that the information in Survey Question One is being recoded from 5 separate values into 3 values. We encourage you to type Recoded in front of the new variable name.

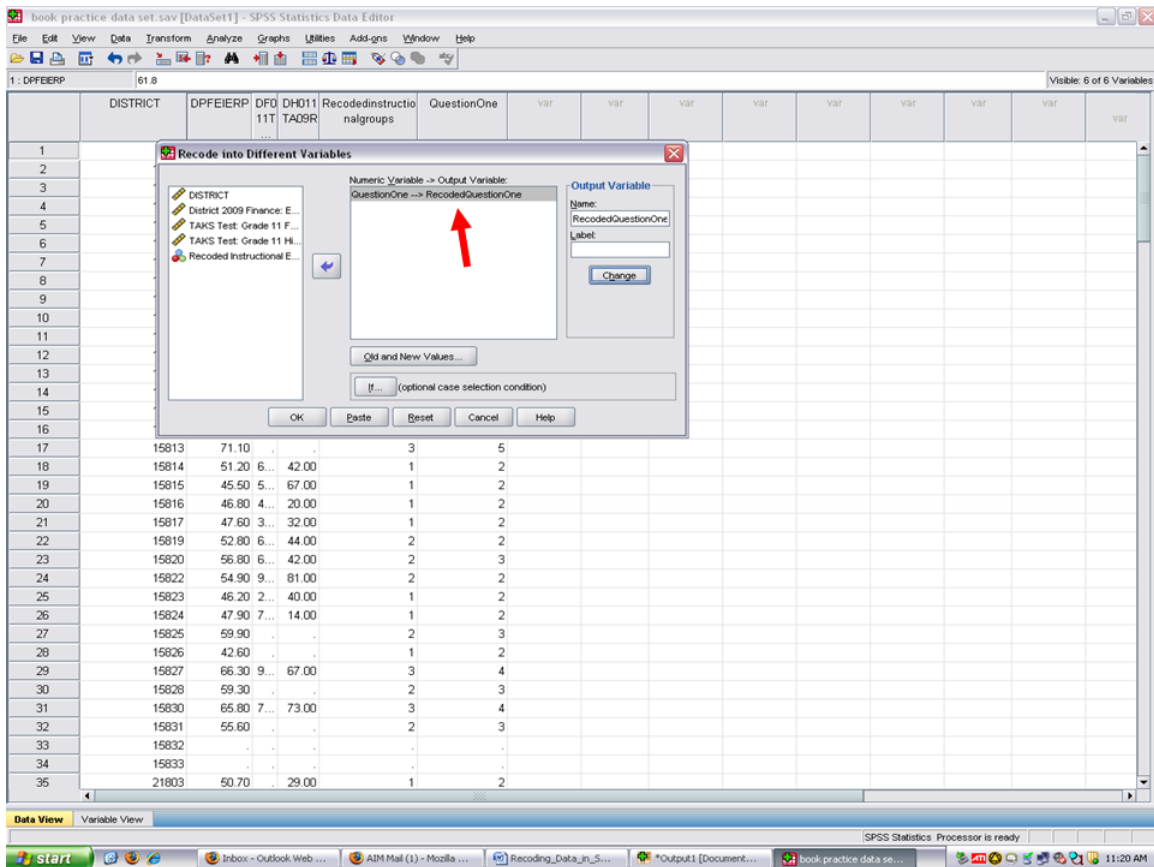
The name, RecodedQuestionOne, will be the name of our new variable. After typing in the name, click on Change.

¹⁷<http://cnx.org/content/m40708/latest/2.12.png/image>



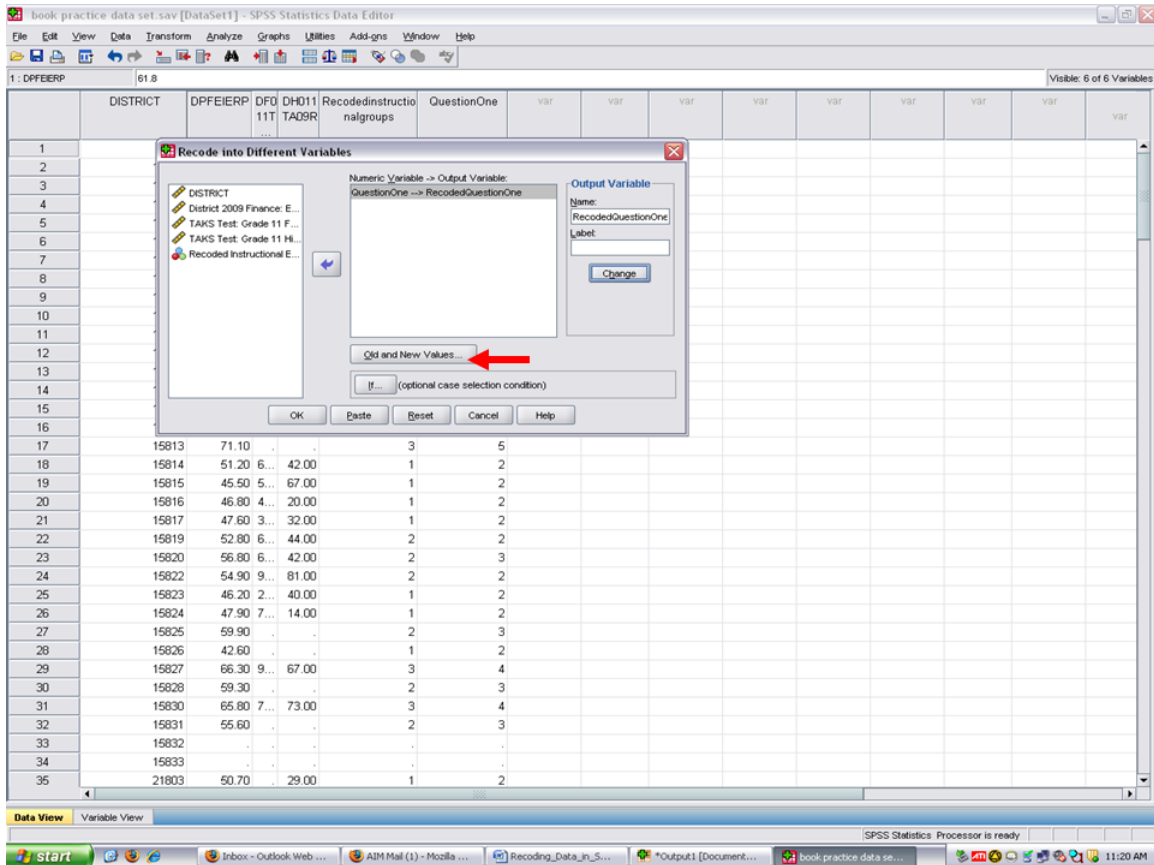
After clicking on Change, the name for the new variable that is being created appears in the Numeric Variable – Output Variable box.

¹⁸<http://cnx.org/content/m40708/latest/2.13.png/image>



Now click on Old and New Values. This is where the values will be recoded.

¹⁹<http://cnx.org/content/m40708/latest/2.14.png/image>



This screen will then appear.

²⁰<http://cnx.org/content/m40708/latest/2.15.png/image>

book.practice.data.set.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1: DPFEERP 61.8 Visible: 6 of 6 Variables

DISTRICT DPFEERP DEF DH01 RecodedInstructio QuestionOne VAR VAR VAR VAR VAR VAR VAR VAR

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36

15813 71.10 . . 3 5
 15814 51.20 6... 42.00 1 2
 15815 45.50 5... 67.00 1 2
 15816 46.80 4... 20.00 1 2
 15817 47.60 3... 32.00 1 2
 15819 52.80 6... 44.00 2 2
 15820 56.80 6... 42.00 2 3
 15822 54.90 9... 81.00 2 2
 15823 46.20 2... 40.00 1 2
 15824 47.90 7... 14.00 1 2
 15825 59.90 . . 2 3
 15826 42.60 . . 1 2
 15827 66.30 9... 67.00 3 4
 15828 59.30 . . 2 3
 15830 65.80 7... 73.00 3 4
 15831 55.60 . . 2 3
 15832
 15833
 21803 50.70 29.00 1 2

Old Value
 Value:
 System-missing
 System- or user-missing
 Range:
 through
 Range, LOWEST through value:
 Range, value through HIGHEST:
 All other values

New Value
 Value:
 System-missing
 Copy old value(s)

Old -> New:

Apply Change Remove

Output variables are strings Width: 8
 Convert numeric strings to numbers (S->N)

Continue Cancel Help

Variable
 dQuestionOne
 Change

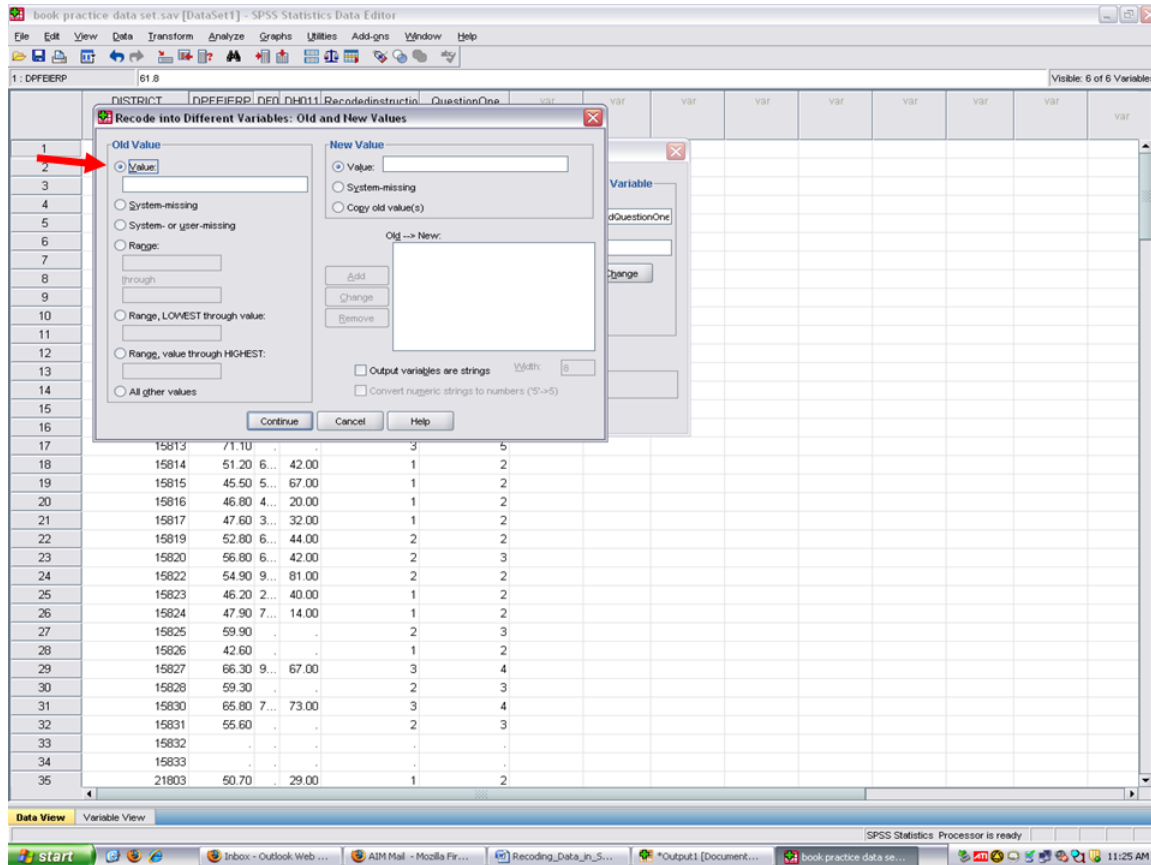
Data View Variable View

SPSS Statistics Processor is ready

start In-box - Outlook Web ... AIM Mail - Mozilla Fir... Recoding_Data_in_S... *Output1 [Document... book.practice.data.se... 11:24 AM

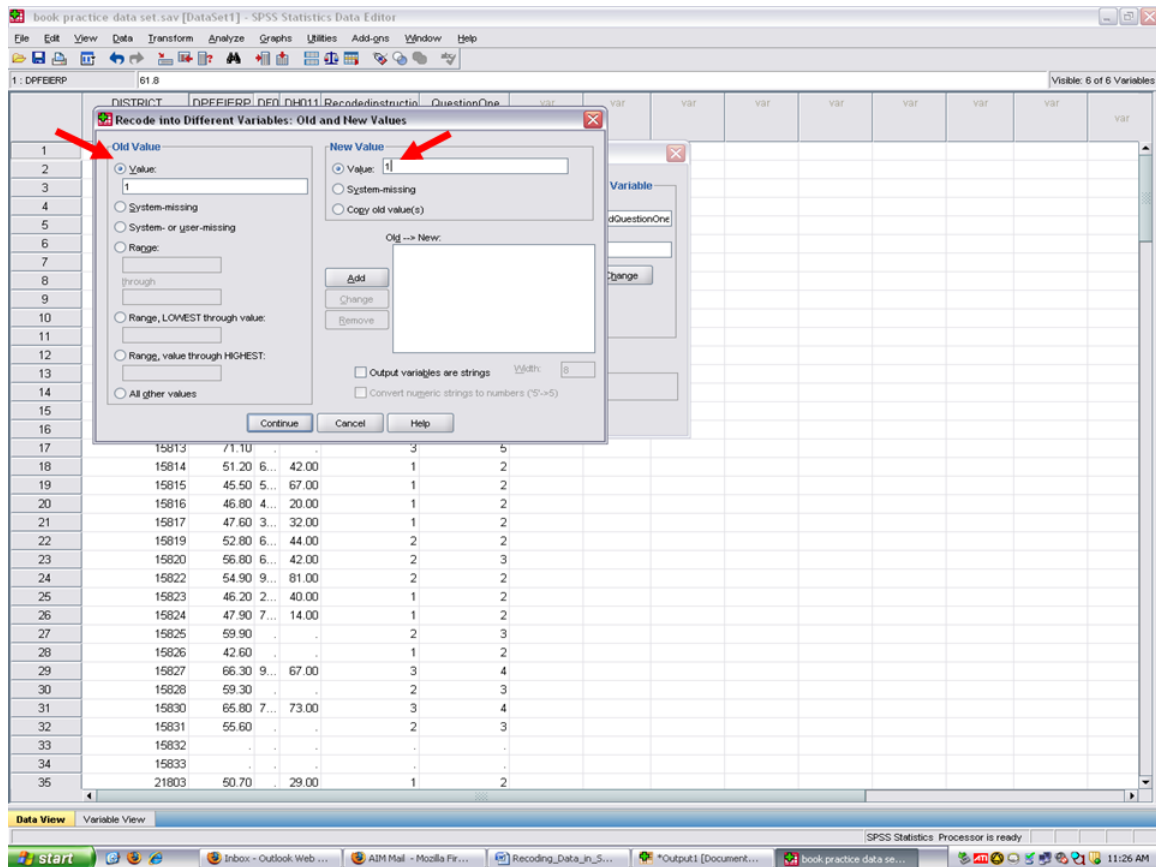
Click on Old Value and the rectangle underneath it becomes activated.

²¹<http://cnx.org/content/m40708/latest/2.16.png/image>



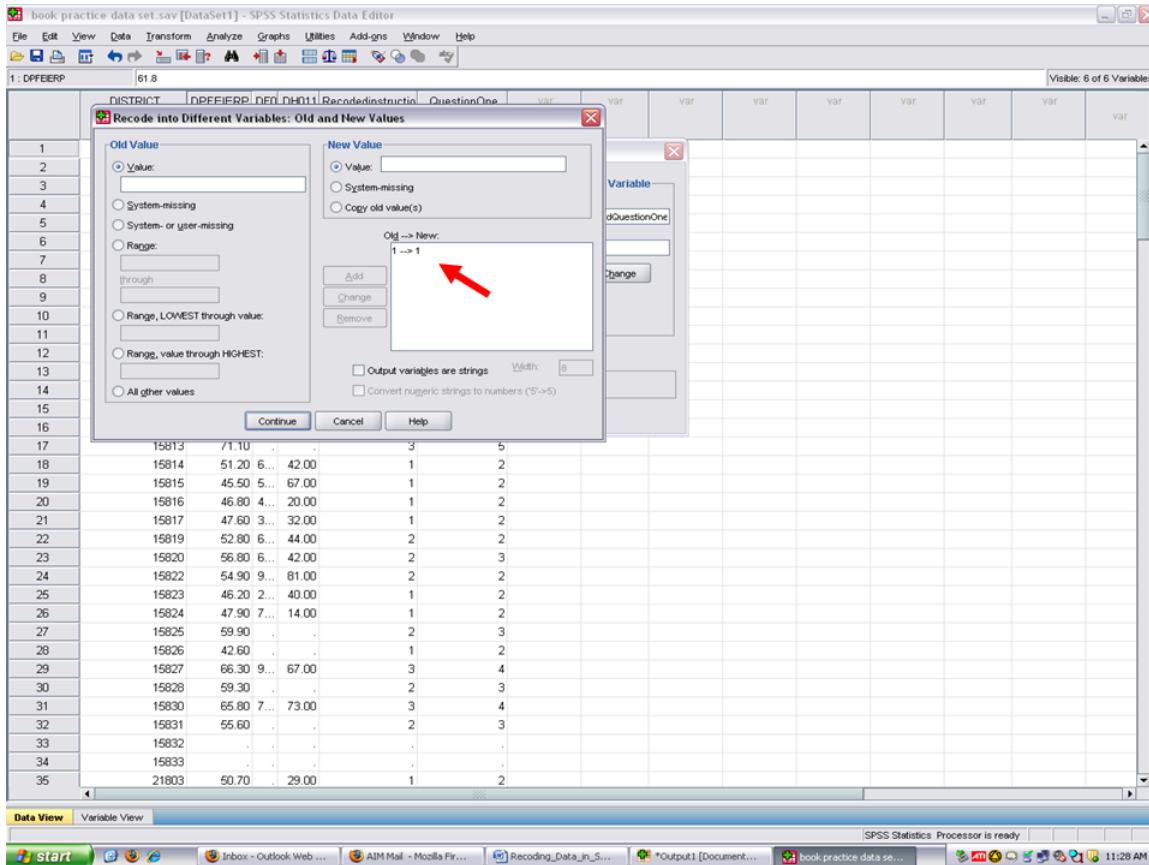
Now we will begin recoding the values for this variable. Type a 1 in the Old Value box and a 1 in the New Value box.

²²<http://cnx.org/content/m40708/latest/2.17.png/image>



Click on Add. This then moves the 1 in the Old Value and 1 in the New Value into the larger box.

²³<http://cnx.org/content/m40708/latest/2.18.png/image>



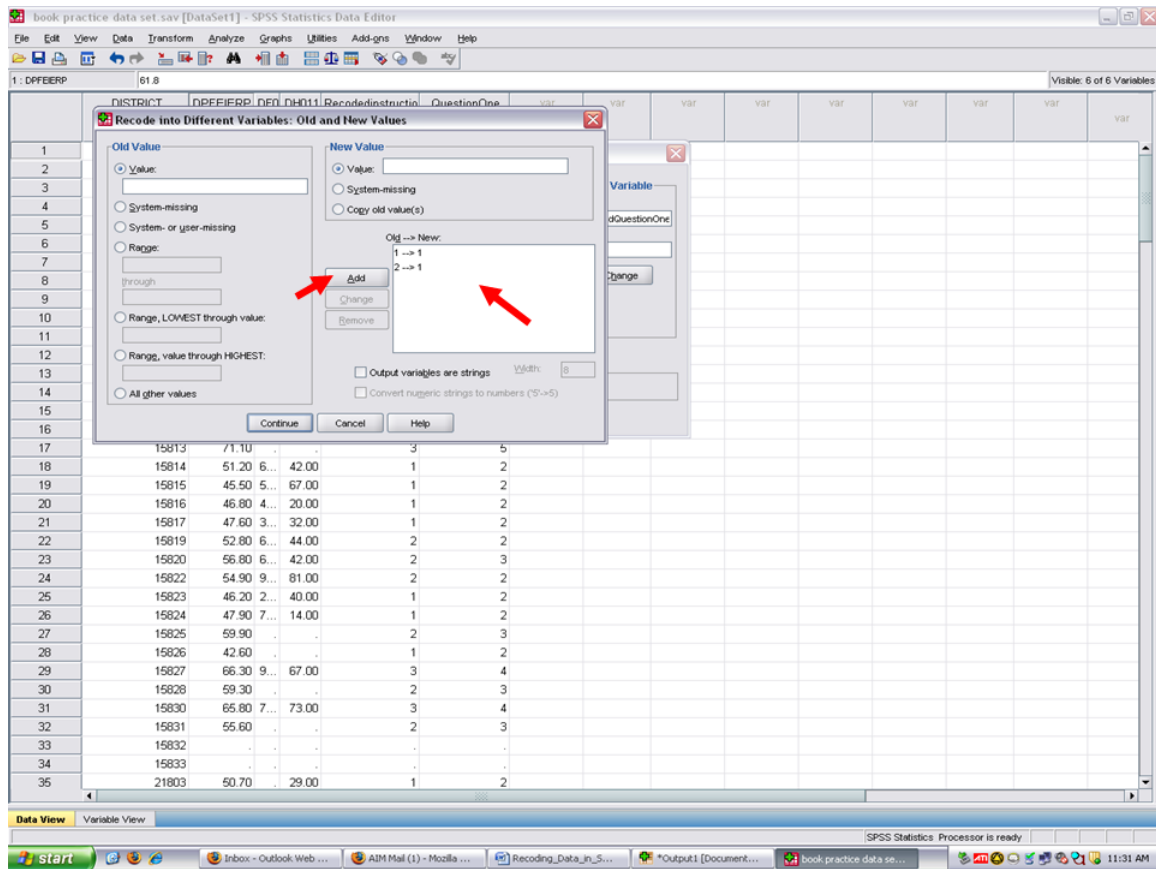
Now type a 2 in Old Value and a 1 in New Value. When this process is completed, SPSS will merge all values of a 1 and all values of a 2 into a 1 category.

²⁴<http://cnx.org/content/m40708/latest/2.19.png/image>

The screenshot shows the SPSS Statistics Data Editor interface. A dialog box titled "Recode into Different Variables: Old and New Values" is open. In the "Old Value" section, the "Value:" radio button is selected, and the value "2" is entered in the text box. In the "New Value" section, the "Value:" radio button is selected, and the value "1" is entered in the text box. A red arrow points from the "2" in the Old Value field to the "1" in the New Value field. The "Add" button is highlighted. The background data grid shows columns for DISTRICT, DPFEERP, DEF, DH011, RecodedInstructio, QuestionOne, and several VAR columns. The data rows are numbered 1 to 35.

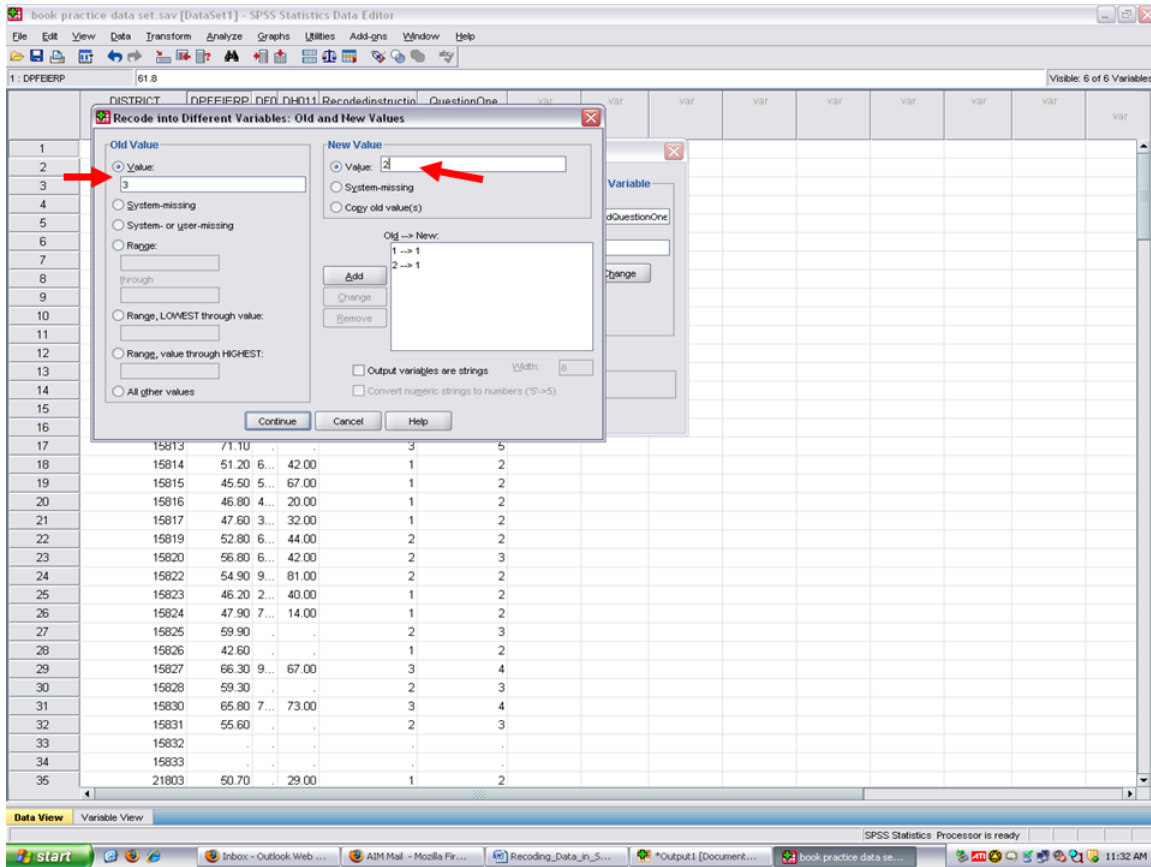
Click on Add. The Old Value of 2 and the New Value of 1 now appear in the larger box. 25

²⁵<http://cnx.org/content/m40708/latest/2.20.png/image>



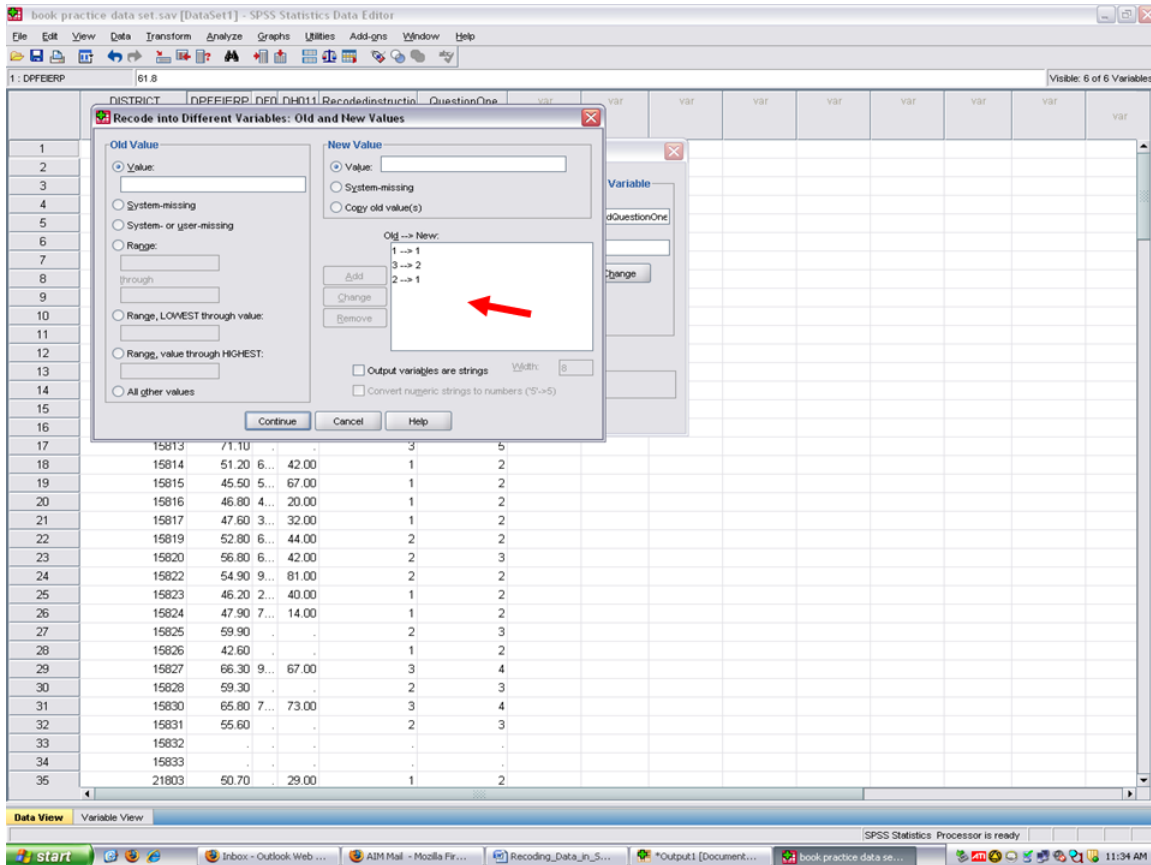
Type a 3 in Old Value and a 2 in New Value.

²⁶<http://cnx.org/content/m40708/latest/2.21.png/image>



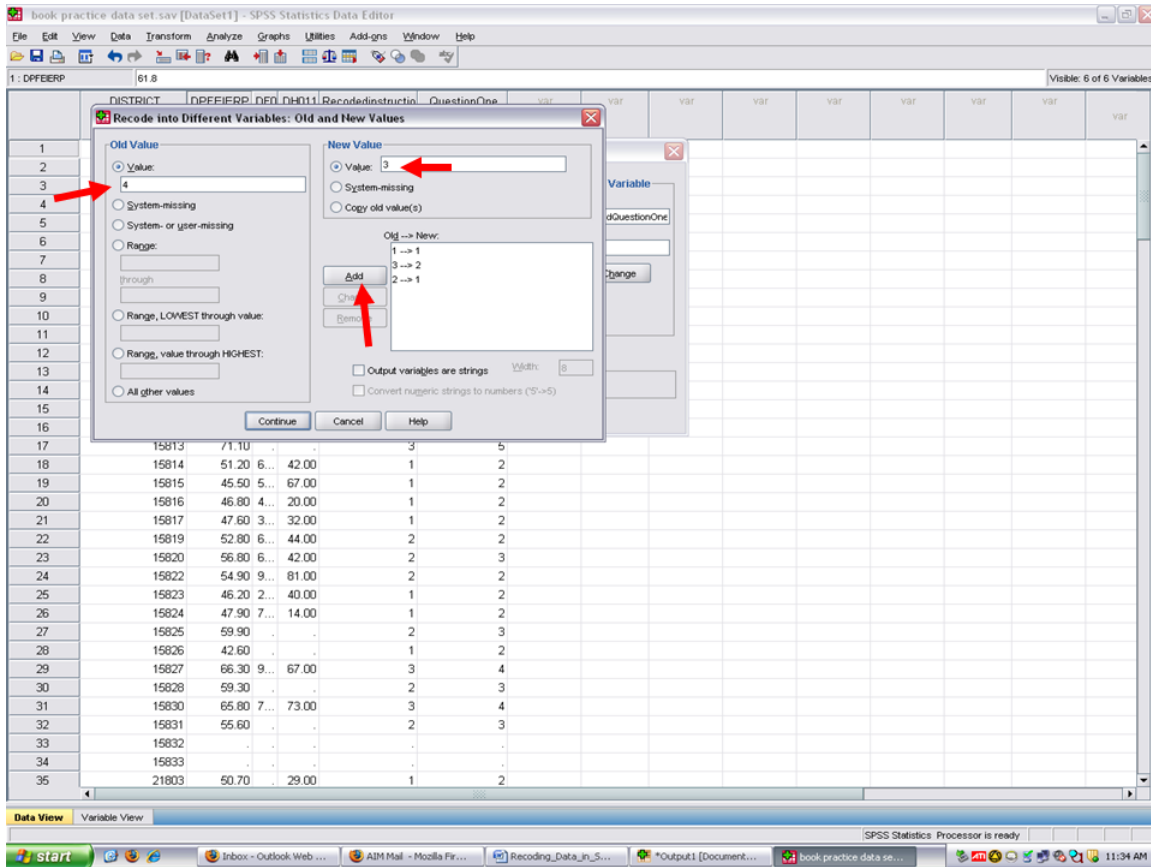
Click on Add. The Old Value of 3 and the New Value of 2 have been added into the larger box.

²⁷ <http://cnx.org/content/m40708/latest/2.22.png/image>



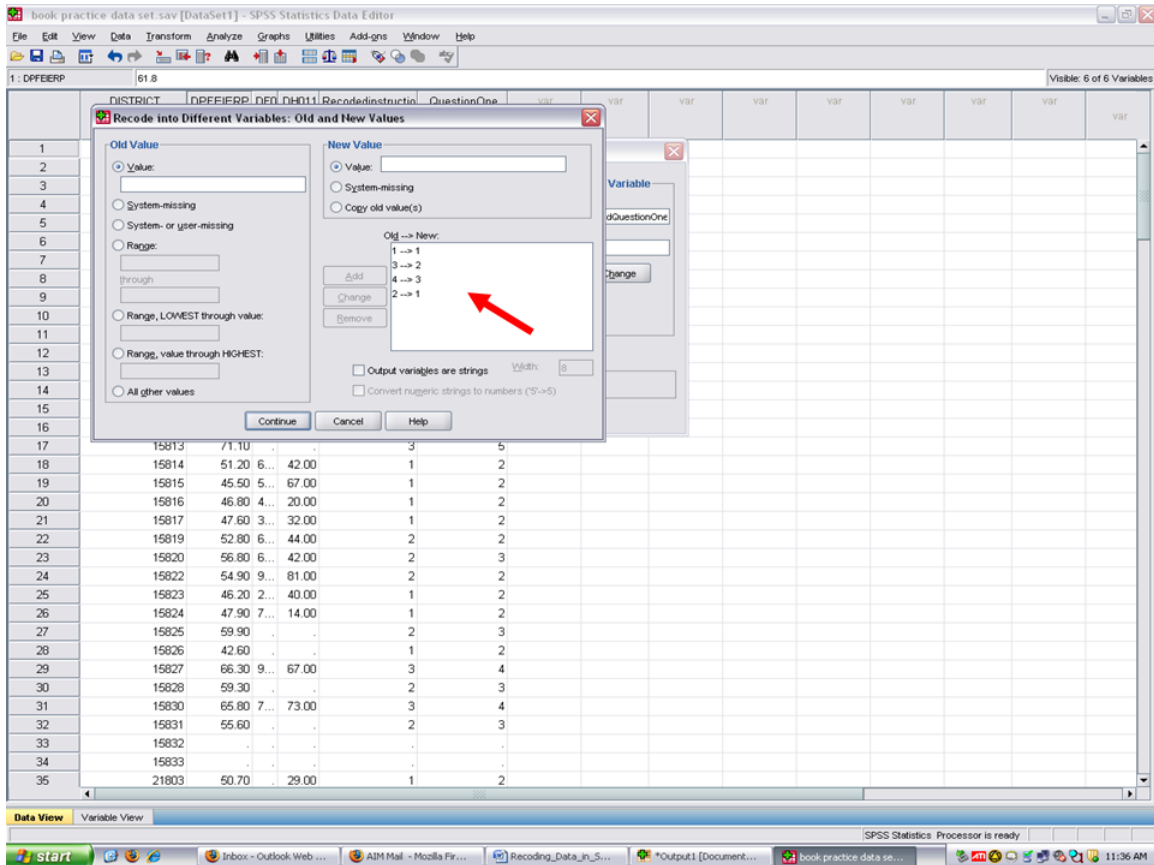
Type a 4 in Old Value and a 3 in New Value. Then click on Add

²⁸<http://cnx.org/content/m40708/latest/2.23.png/image>



The Old Value of 4 and the New Value of 3 have now been added to the larger box.

²⁹<http://cnx.org/content/m40708/latest/2.24.png/image>



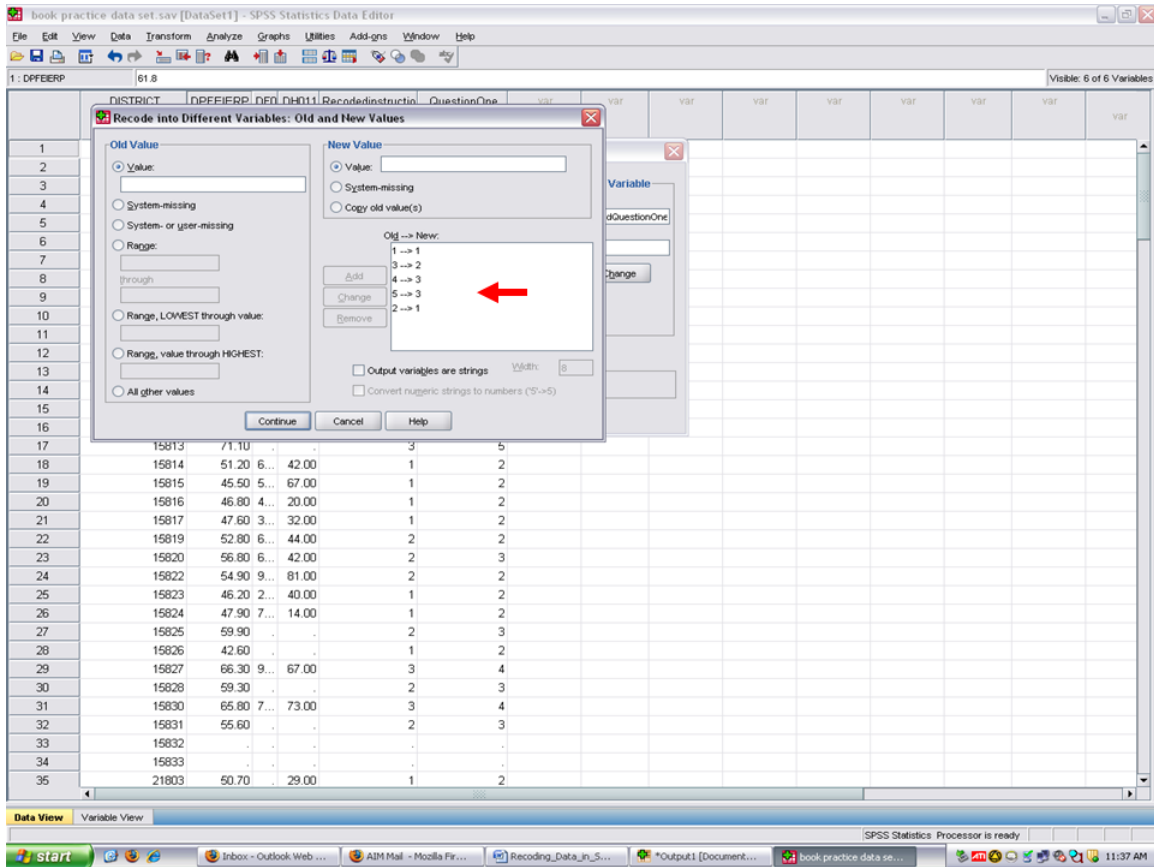
Finally, type 5 in the Old Value box and 3 into the New Value box. Click on Add

³⁰<http://cnx.org/content/m40708/latest/2.25.png/image>

The screenshot shows the SPSS Statistics Data Editor interface. A dialog box titled "Recode into Different Variables: Old and New Values" is open. In the "Old Value" section, the "Value" radio button is selected, and the value "5" is entered in the text box. In the "New Value" section, the "Value" radio button is selected, and the value "3" is entered. Below these sections, the "Old -> New:" list contains the following mappings: 1 -> 1, 3 -> 2, 4 -> 3, and 2 -> 1. The "Add" button is highlighted with a red arrow. The background data view shows a table with columns for DISTRICT, DPFEERP, DEF, DH01, RecodedInstructio, QuestionOne, and several VAR columns. The 'QuestionOne' column contains values 5, 3, 4, 2, 1, 2, 2, 3, 2, 2, 2, 1, 2, 2, 3, 1, 2, 3, 4, 2, 3, 3, 4, 2, 3, 3, 4, 2, 3, 1, 2.

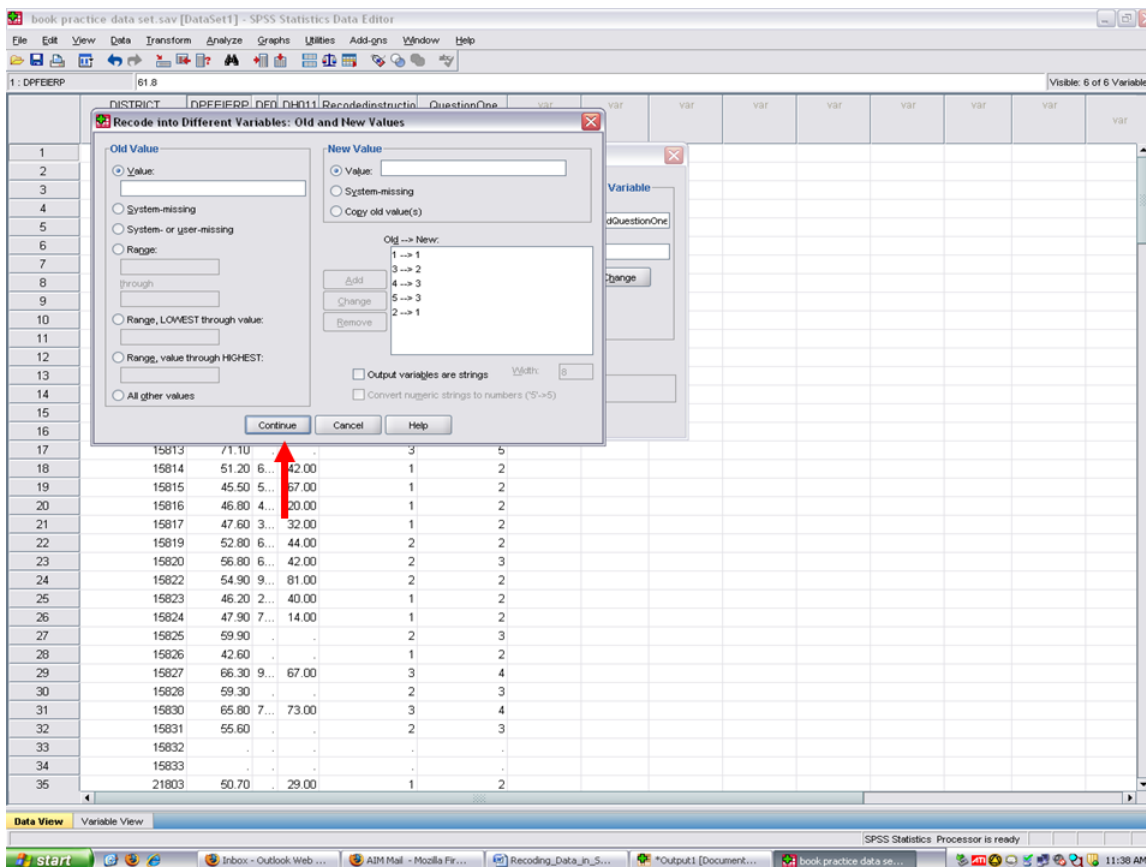
Now all 5 values in the Survey Question One variable are present in the Old – New box below. They will be recoded into 3 values.

³¹ <http://cnx.org/content/m40708/latest/2.26.png/image>



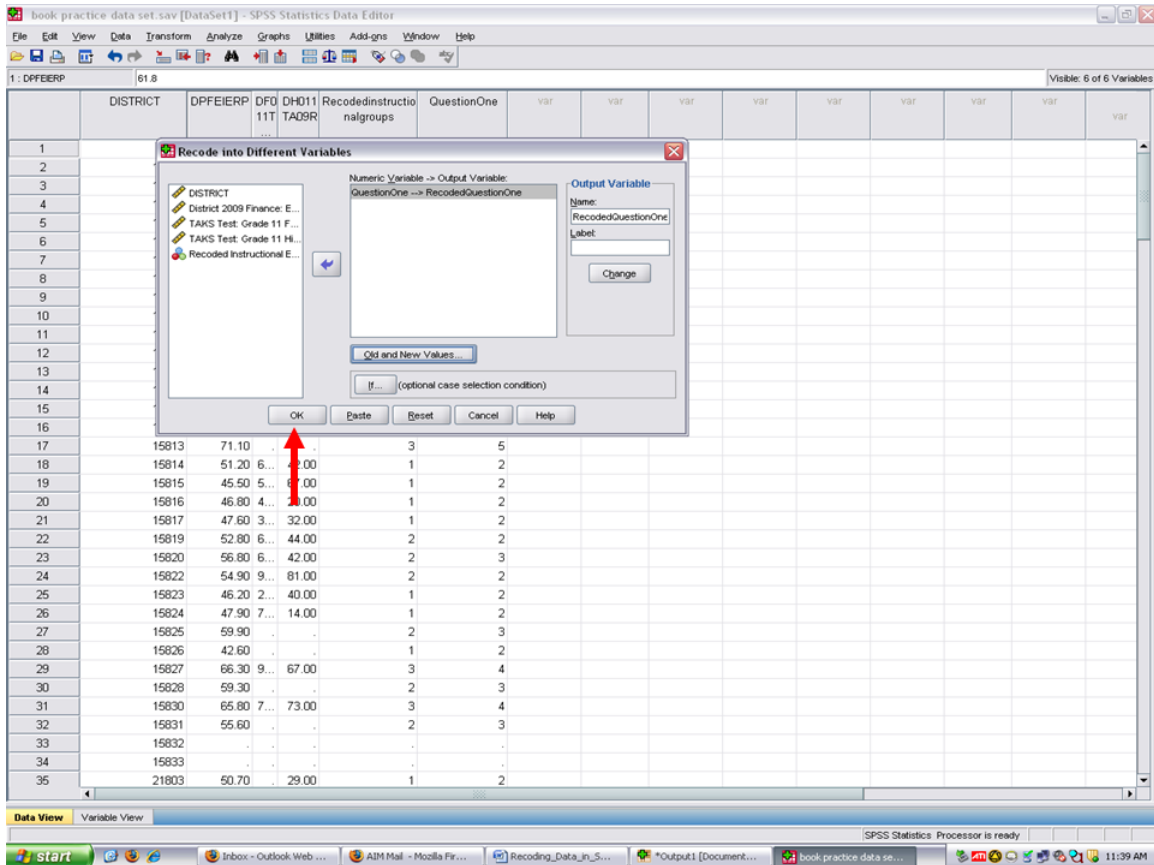
Click on Continue.

³²<http://cnx.org/content/m40708/latest/2.27.png/image>



Click on OK. SPSS will then generate a new variable for you, RecodedQuestionOne.

³³<http://cnx.org/content/m40708/latest/2.28.png/image>



The newly created variable, RecodedQuestionOne, will be added as the last variable in your dataset.

³⁴<http://cnx.org/content/m40708/latest/2.29.png/image>

book practice data set.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1: DPFEERP 61.8 Visible: 7 of 7 Variables

	DISTRICT	DPFEERP	DFD 11T	DH011 TA09R	Recodeddestructio nalgroups	QuestionOne	RecodedQuestionOne	Var	Var	Var	Var	Var	Var
1	3901	61.00	.	.	3	3	2.00						
2	13801	59.20	.	.	2	3	2.00						
3	14801	57.00	3...	29.00	2	3	2.00						
4	14802	53.00	5...	.	2	2	1.00						
5	14803	57.50	.	.	2	3	2.00						
6	14804	50.60	5...	.	1	2	1.00						
7	15801	36.00	3...	29.00	1	1	1.00						
8	15802	49.40	2...	56.00	1	2	1.00						
9	15803	58.10	.	.	2	3	2.00						
10	15805	59.20	.	.	2	3	2.00						
11	15806	44.30	7...	47.00	1	2	1.00						
12	15807	51.40	3...	42.00	1	2	1.00						
13	15808	58.10	.	.	2	3	2.00						
14	15809	50.20	.	.	1	2	1.00						
15	15811	54.70	.	.	2	2	1.00						
16	15812	60.10	2...	58.00	2	3	2.00						
17	15813	71.10	.	.	3	5	3.00						
18	15814	51.20	6...	42.00	1	2	1.00						
19	15815	45.50	5...	67.00	1	2	1.00						
20	15816	46.80	4...	20.00	1	2	1.00						
21	15817	47.60	3...	32.00	1	2	1.00						
22	15819	52.80	6...	44.00	2	2	1.00						
23	15820	56.80	6...	42.00	2	3	2.00						
24	15822	54.90	9...	81.00	2	2	1.00						
25	15823	46.20	2...	40.00	1	2	1.00						
26	15824	47.90	7...	14.00	1	2	1.00						
27	15825	59.90	.	.	2	3	2.00						
28	15826	42.60	.	.	1	2	1.00						
29	15827	66.30	9...	67.00	3	4	3.00						
30	15828	59.30	.	.	2	3	2.00						
31	15830	65.80	7...	73.00	3	4	3.00						
32	15831	55.60	.	.	2	3	2.00						
33	15832						
34	15833						
35	21803	50.70	.	29.00	1	2	1.00						

Data View Variable View

SPSS Statistics Processor is ready

start | Inboxes - Outlook Web ... | AIM Mail - Mozilla Fir... | Recoding_Data_in_S... | *Output1 [Document... | book.practice data se... | 11:40 AM

You have generated a new variable, one not in your original dataset, that contains 3 values, instead of the 5 values in the Question One variable.

³⁵<http://cnx.org/content/m40708/latest/2.30png/image>

Chapter 4

4. Steps Involved in Converting Data from An Excel File into SPSS¹



NOTE: This chapter is published by NCPEA Press² and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of education administration.

About the Authors

John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website (Writing and Statistical Help³) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

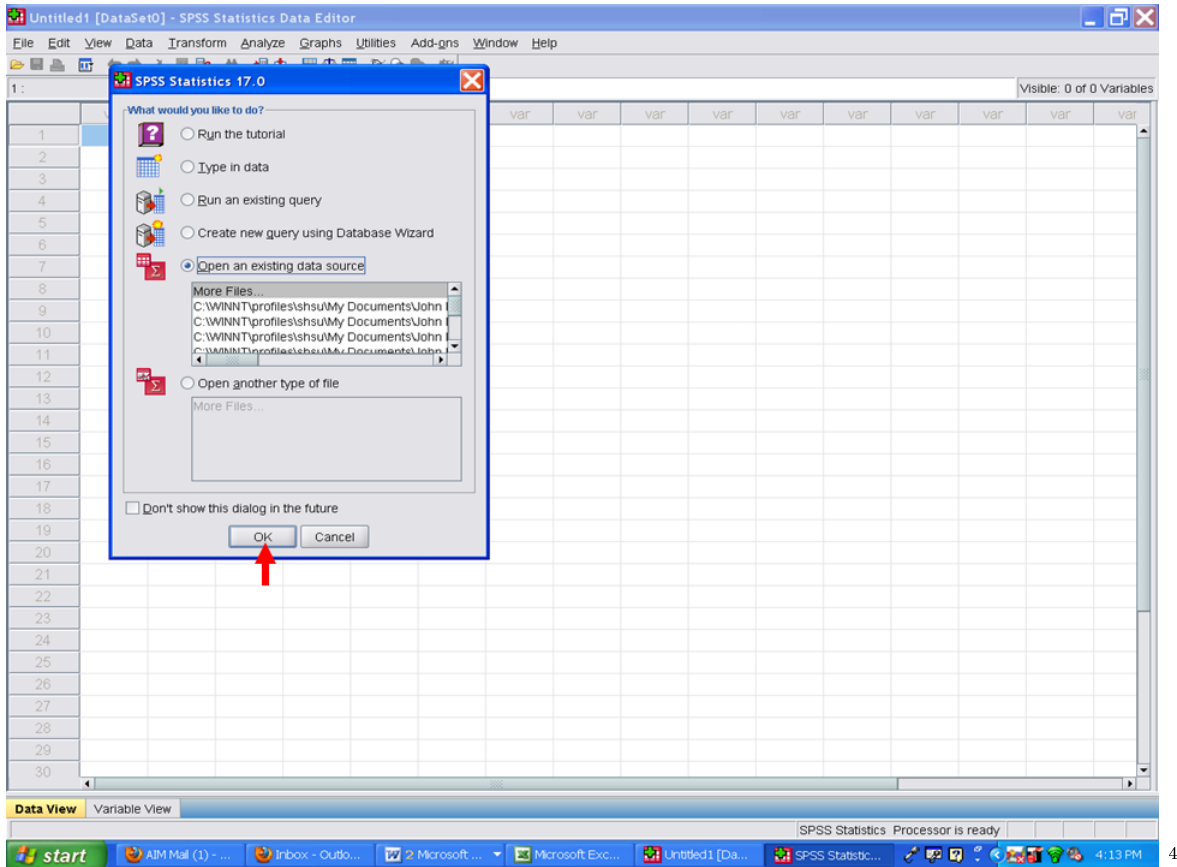
Ana Rojas-LeBouef is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. Dr. LeBoeuf recently completed her doctoral degree in Reading, where she conducted a 16-year analysis of Texas statewide data regarding the achievement gap. Her research interests lie in examining the inequities in achievement among ethnic groups. Dr. Rojas-LeBouef also assists students and faculty in their writing and statistical needs on the Writing and Statistical Help website.

These procedures are to be followed when you have data already entered into an Excel file. In this chapter, you will be provided with steps and screenshots on how to convert the data from an Excel file into an SPSS data file. Open up your SPSS program and this screen will appear. Click on Cancel.

¹This content is available online at <<http://cnx.org/content/m40709/1.2/>>.

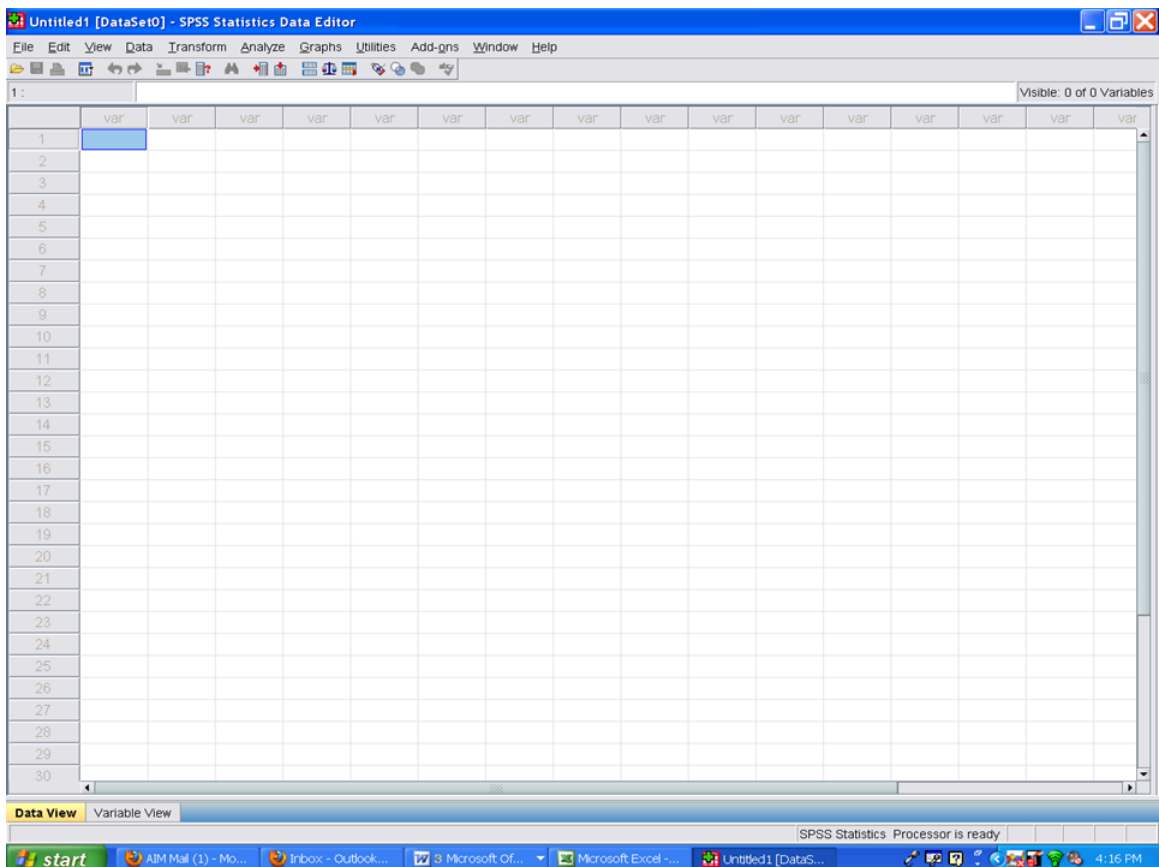
²<http://www.ncpeapublications.org/books.html>

³<http://cnx.org/content/m40709/latest/www.writingandstatisticalhelp>



Your screen will now look like this.

⁴<http://cnx.org/content/m40709/latest/3.1.png/image>



The Excel file we will be converting for this example looks like the following:

⁵<http://cnx.org/content/m40709/latest/3.2.png/image>

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Averaged Writing Score	Gone	Gtwo										
2	4.5	4.5	4.5										
3	3	3	3										
4	3.5	3.5	3.5										
5	4	4	4										
6	4.5	4.5	4.5										
7	4	4	4										
8	4	4	4										
9	3	3	3										
10	2.75	2.5	3										
11	4.5	4.5	4.5										
12	4.25	4	4.5										
13	4.5	4.5	4.5										
14	2	2	2										
15													
16													
17													
18													
19													
20													
21													
22													
23													
24													
25													
26													
27													
28													
29													
30													
31													
32													
33													
34													
35													

In this excel file are three variables: AveragedWritingScore, Gone, and Gtwo. All three are numeric variables. The first row contains the variable names. Rows 2 through 14 contain the data for the variables.

Prior to the next steps, make sure that the Excel file that you intend to convert into SPSS is not open. If it is open and in use, SPSS will not be able to convert the file.

Once you are back in the SPSS data view window,

- ✓ Click on File
- ✓ Click on Open
- ✓ Click on Data

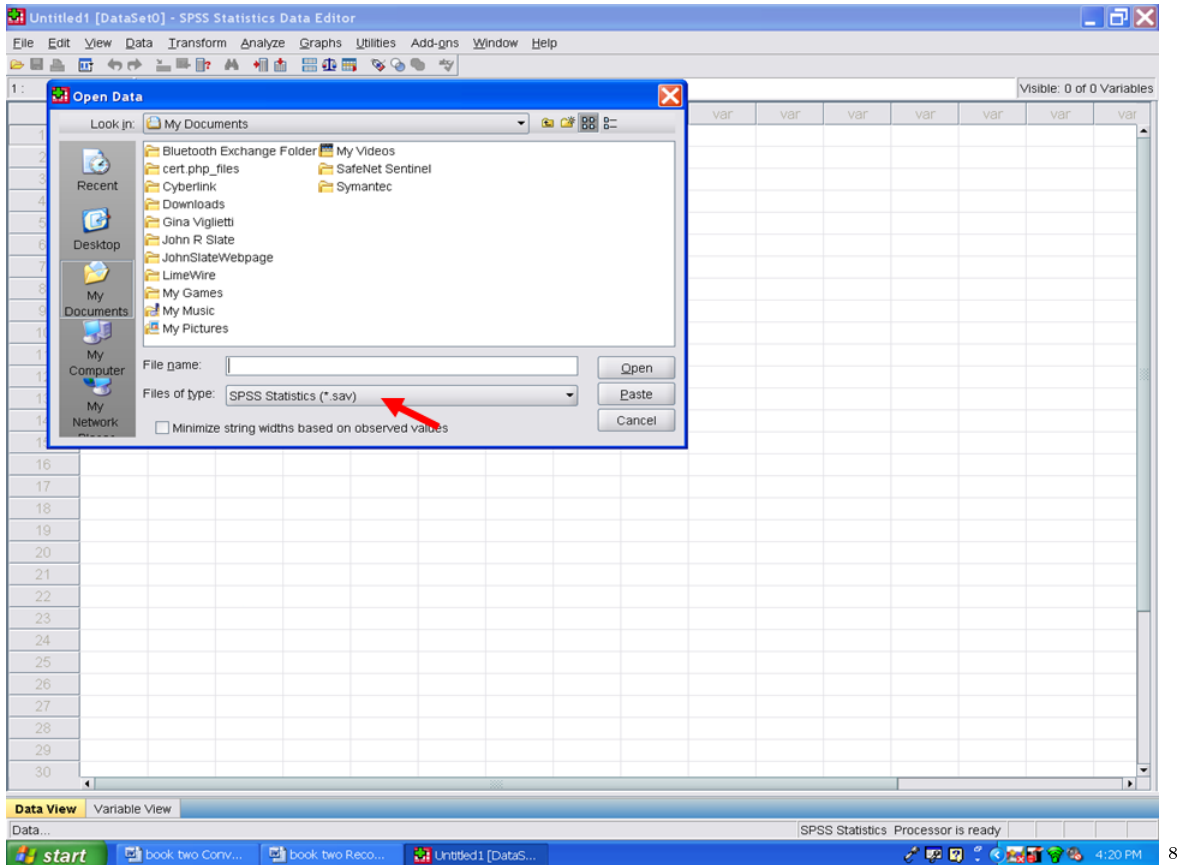
⁶<http://cnx.org/content/m40709/latest/3.3.png/image>

The screenshot shows a Microsoft Excel window titled "Practice Excel File". The spreadsheet has the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Averaged Writing Score	Gone	Gtwo										
2	4.5	4.5	4.5										
3	3	3	3										
4	3.5	3.5	3.5										
5	4	4	4										
6	4.5	4.5	4.5										
7	4	4	4										
8	4	4	4										
9	3	3	3										
10	2.75	2.5	3										
11	4.5	4.5	4.5										
12	4.25	4	4.5										
13	4.5	4.5	4.5										
14	2	2	2										
15													
16													
17													
18													
19													
20													
21													
22													
23													
24													
25													
26													
27													
28													
29													
30													
31													
32													
33													
34													
35													

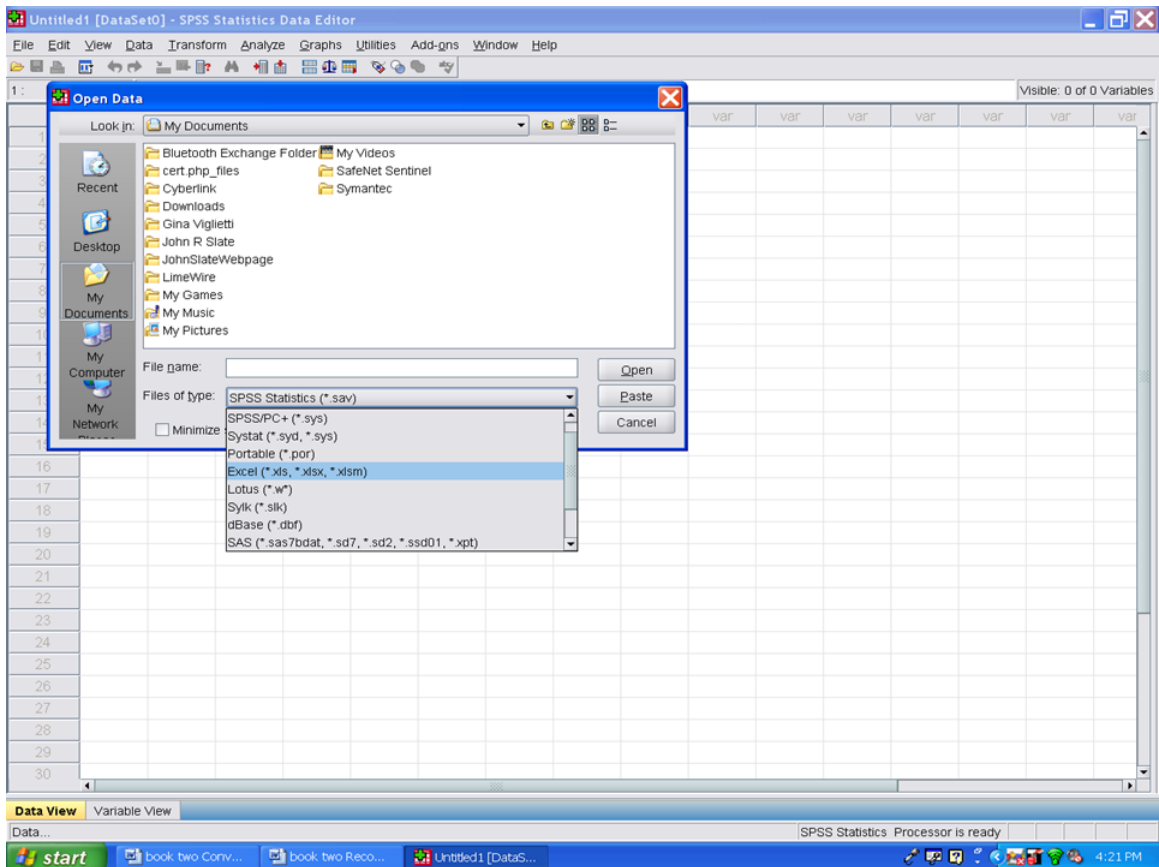
A screen similar to the one below will open up on your computer. You need to locate where your Excel file is located. First, you will need to change the Files of type below.

⁷ <http://cnx.org/content/m40709/latest/3.41.png/image>



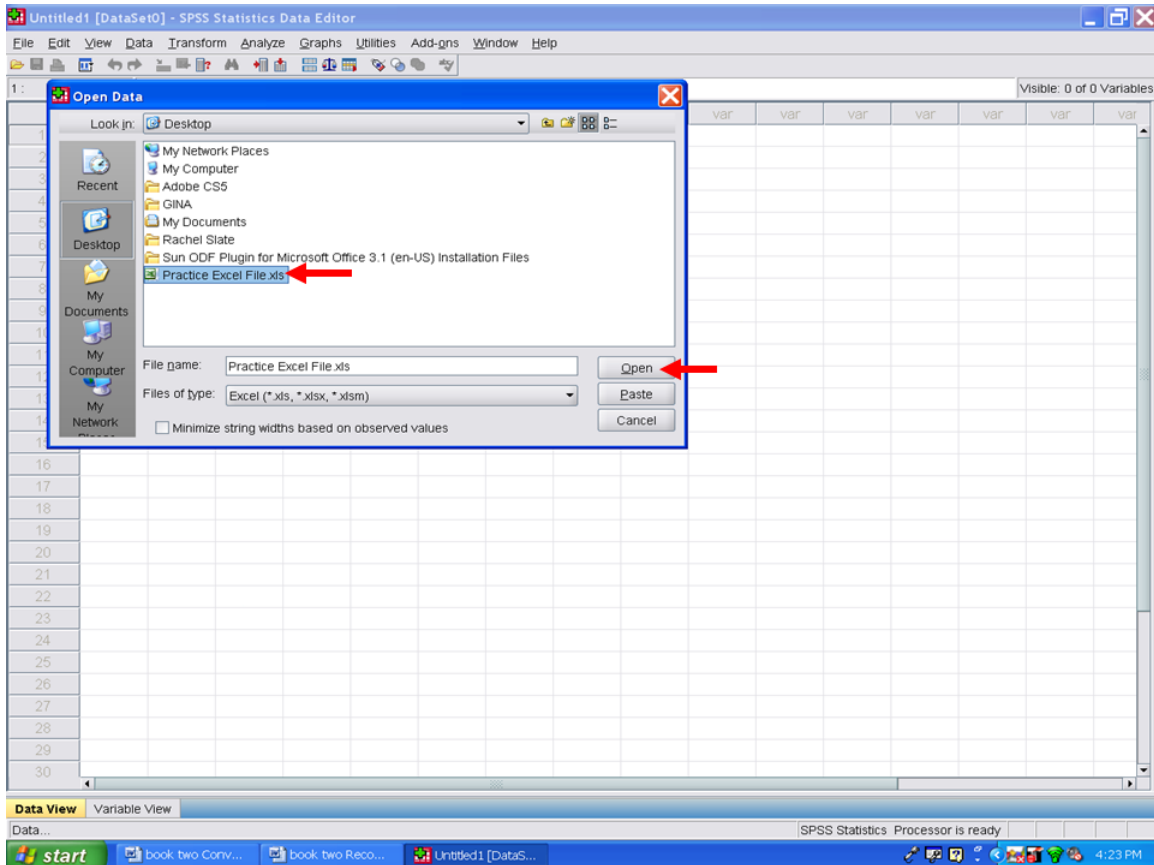
Change from SPSS statistics (*.sav) to Excel (*.xls). Then find where the file is located and its name.

⁸<http://cnx.org/content/m40709/latest/3.5.1.png/image>



In the case of this example, the Excel file was located on the Desktop and is titled, Practice Excel File.xls. Then click on Open.

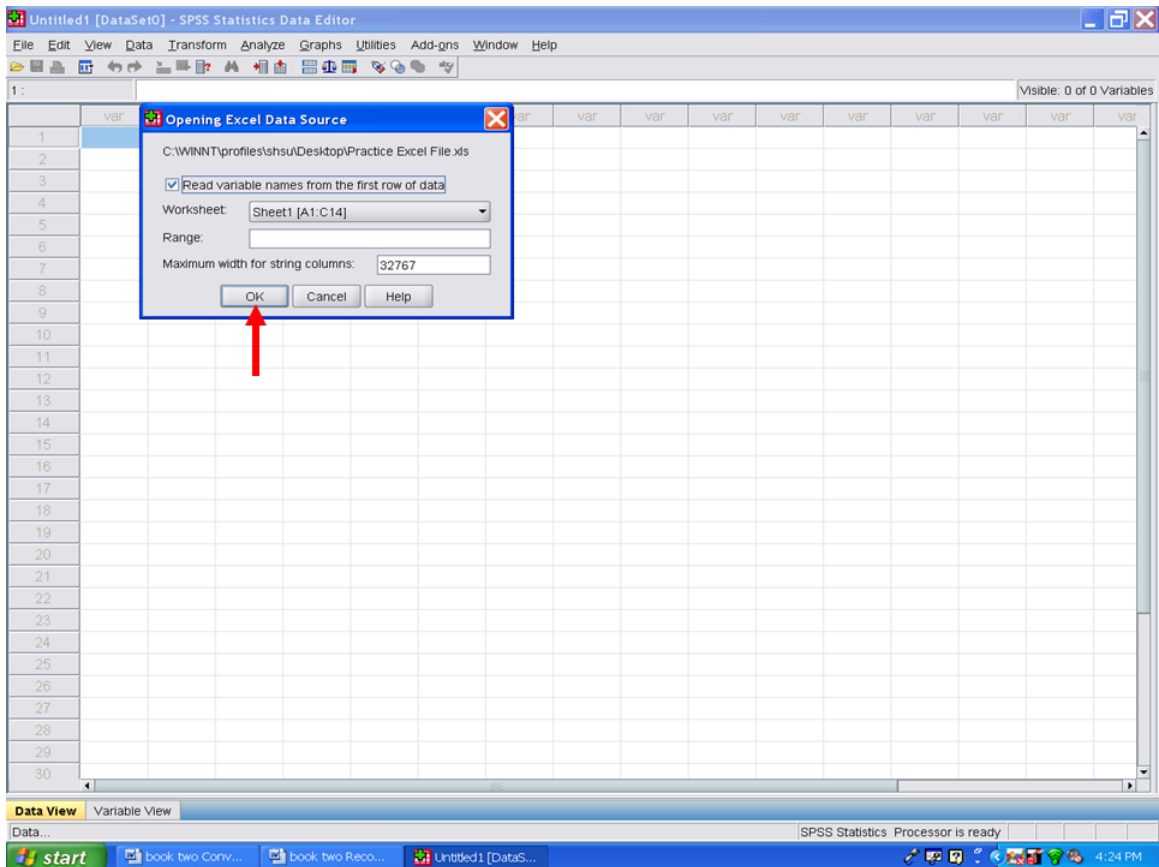
⁹<http://cnx.org/content/m40709/latest/3.6.png/image>



After clicking on open, the screen below will appear. For an Excel file to be converted properly into SPSS, the first row in the Excel file should be the names of the variables. All of the other rows in the Excel file should be the data. If you use more than one row to describe your variables, you will not be able to convert your Excel file into SPSS using these steps.

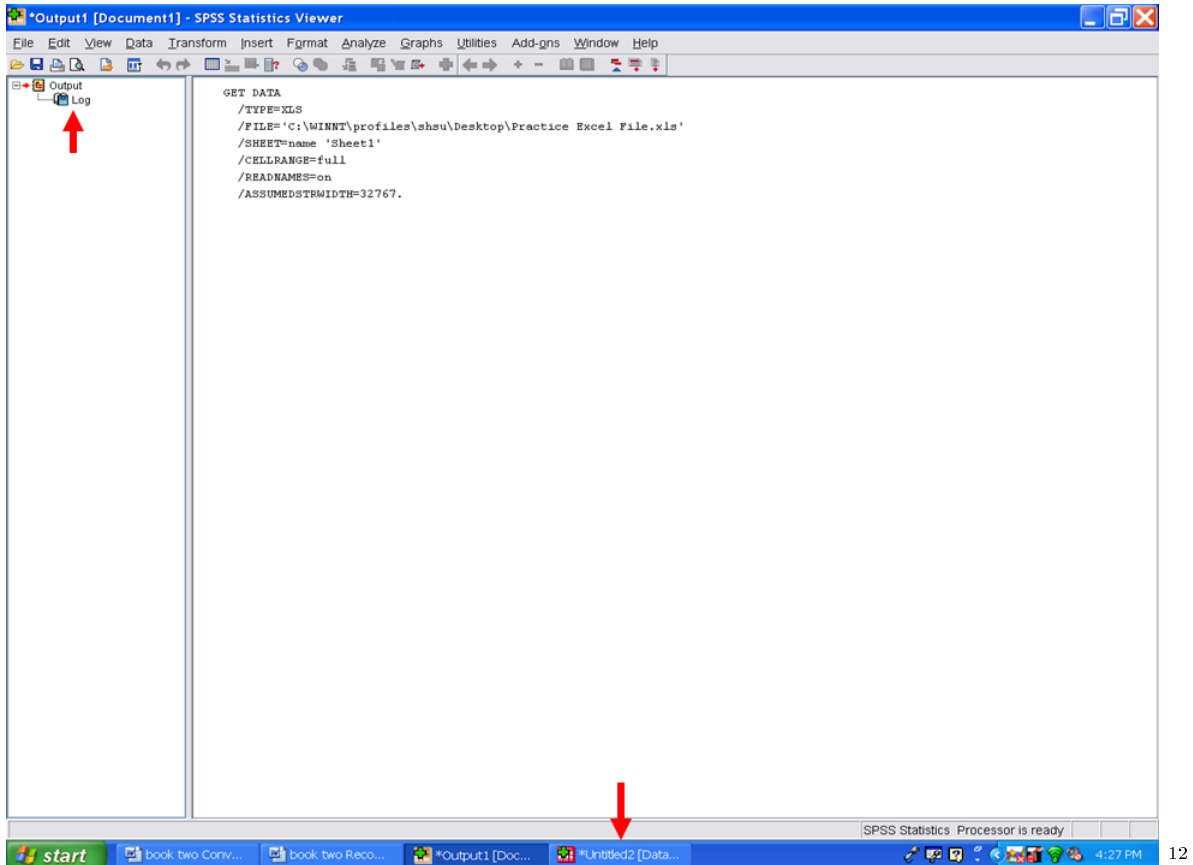
Once this screen has appeared, click on OK.

¹⁰<http://cnx.org/content/m40709/latest/3.7.png/image>



You should then be sent to the Output window.
Click on the SPSS data button below to get back to your SPSS screen.

¹¹ <http://cnx.org/content/m40709/latest/3.8.png/image>



Clicking on the SPSS data icon as depicted above should take you to the Data View screen. In this case, three variables were in the Excel data file: AveragedWritingScore, Gone, and Gtwo. These three variables were directly converted from the Excel file into a SPSS datafile.

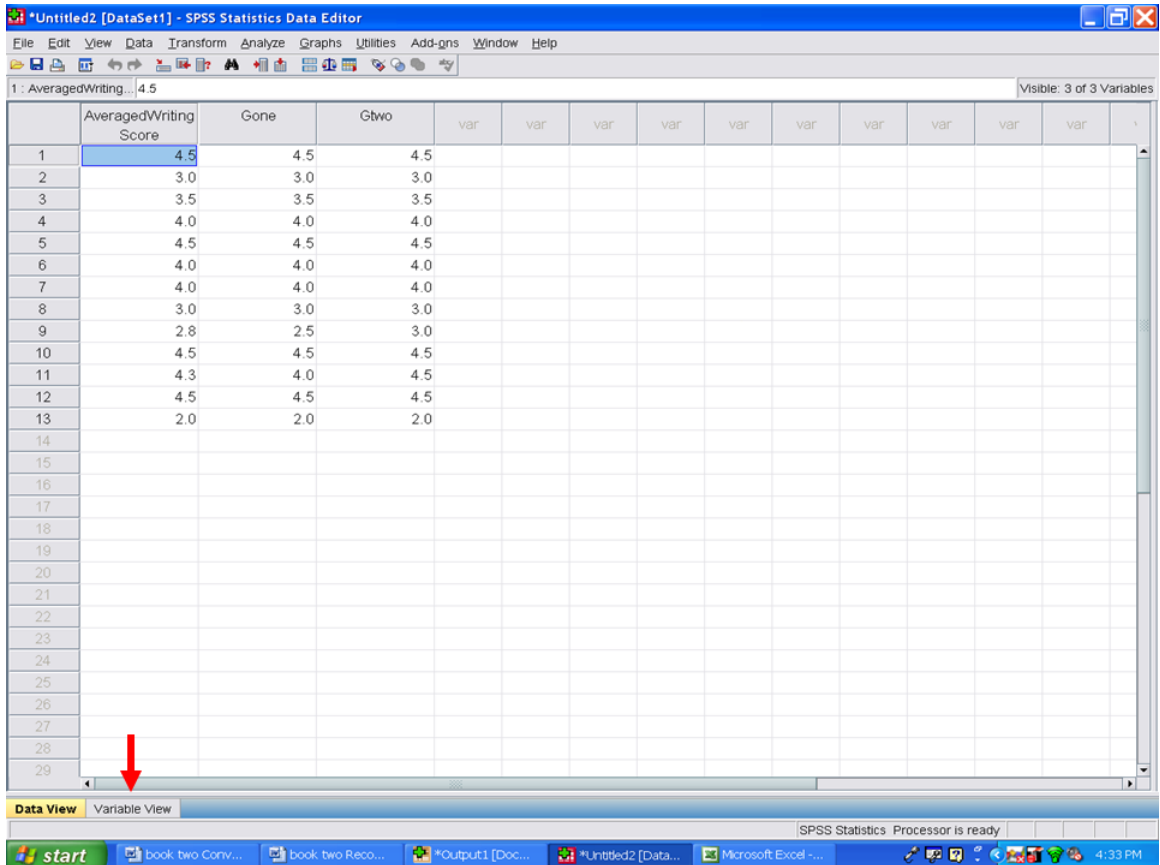
¹²<http://cnx.org/content/m40709/latest/3.9.png/image>

The screenshot displays the SPSS Statistics Data Editor interface. The main window shows a dataset with 13 rows and 14 columns. The first three columns are labeled 'AveragedWriting Score', 'Gone', and 'Gtwo'. The first row has values 4.5, 4.5, and 4.5. Red arrows point to these values. The bottom status bar shows 'SPSS Statistics Processor is ready' and the system tray shows the time as 4:28 PM.

	AveragedWriting Score	Gone	Gtwo	var	var	var	var	var	var	var	var	var	var
1	4.5	4.5	4.5										
2	3.0	3.0	3.0										
3	3.5	3.5	3.5										
4	4.0	4.0	4.0										
5	4.5	4.5	4.5										
6	4.0	4.0	4.0										
7	4.0	4.0	4.0										
8	3.0	3.0	3.0										
9	2.8	2.5	3.0										
10	4.5	4.5	4.5										
11	4.3	4.0	4.5										
12	4.5	4.5	4.5										
13	2.0	2.0	2.0										
14													
15													
16													
17													
18													
19													
20													
21													
22													
23													
24													
25													
26													
27													
28													
29													

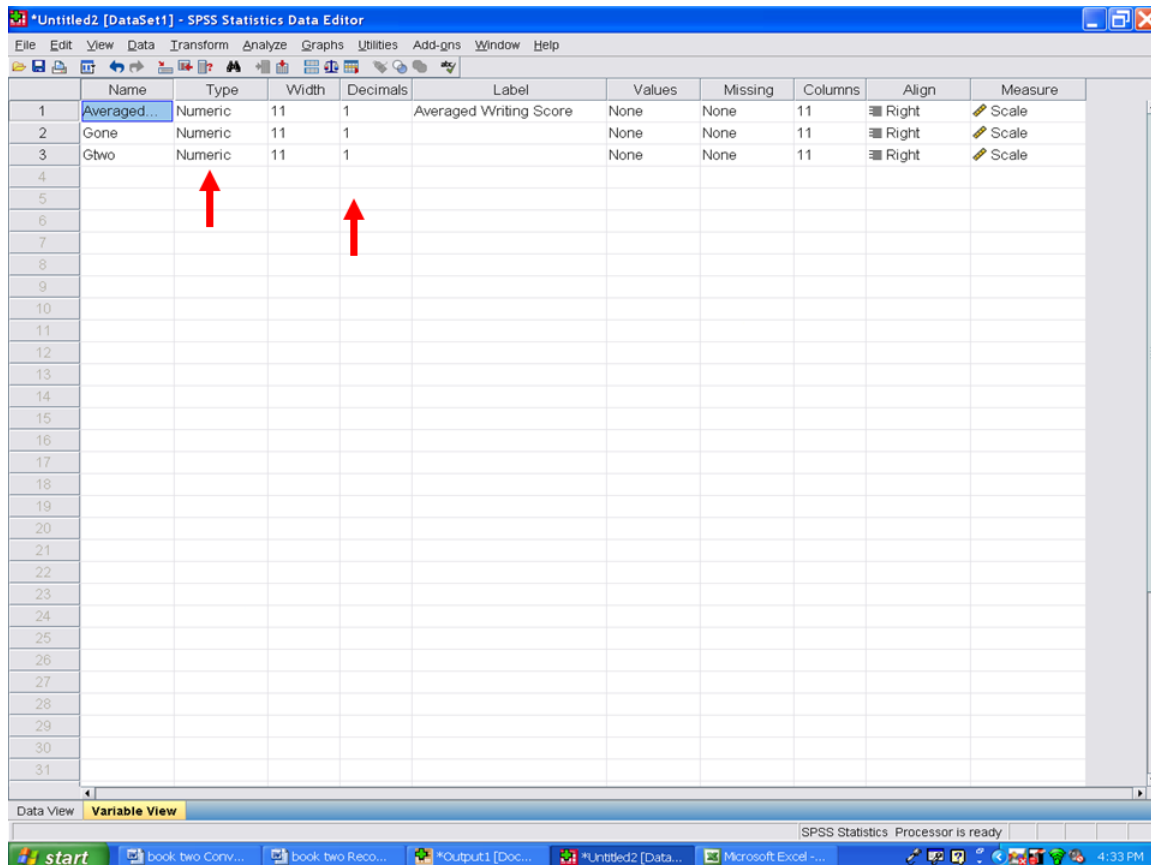
Now we need to make sure that the variables are correctly identified and labeled in this SPSS dataset. Click on Variable View.

¹³<http://cnx.org/content/m40709/latest/3.10.png/image>



In the case of our example, the screen looks like the following:

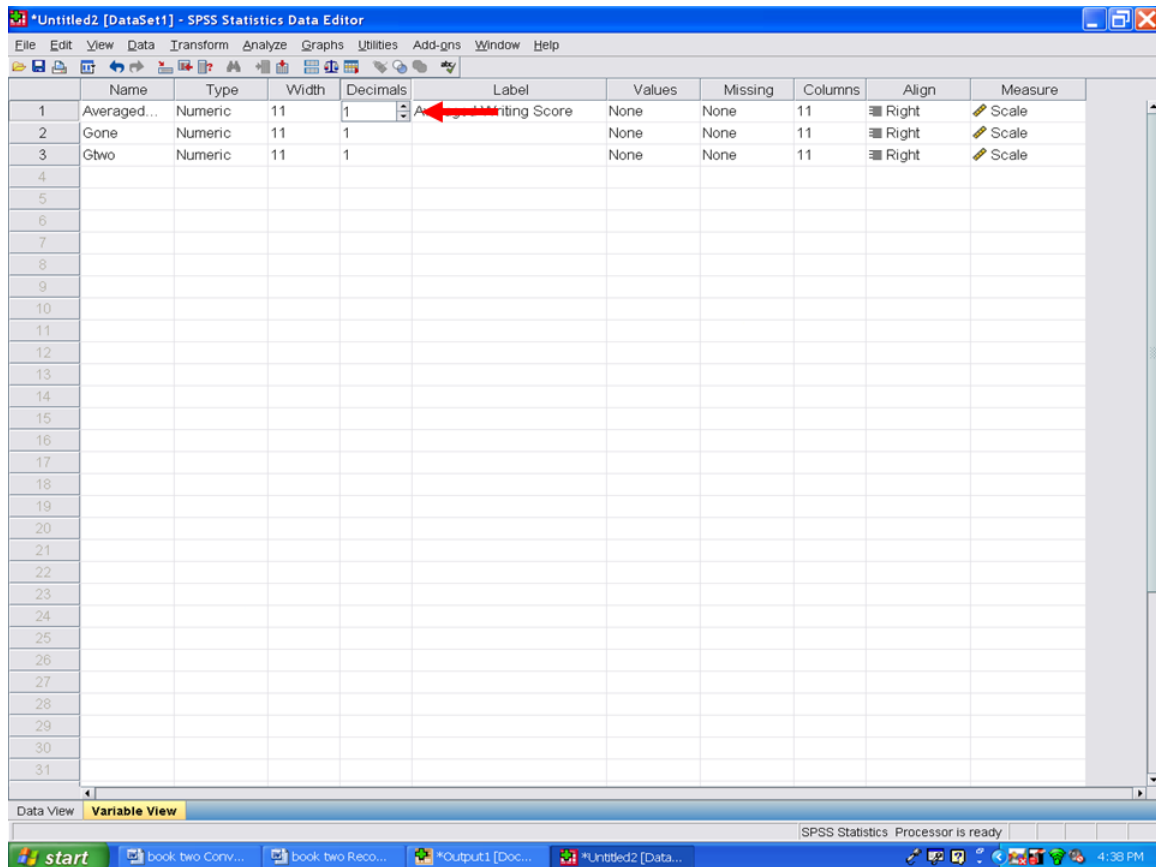
¹⁴<http://cnx.org/content/m40709/latest/3.11.png/image>



All three of the variables have been correctly identified as numeric. The number of decimals needs to be increased to 2, not 1 as shown above.

To change the number of decimal places from 1 to 2, click on one of the decimal cells. In the screen below, the decimal cell for the first row has been clicked on. Once clicked an up and a down arrow will appear.

¹⁵<http://cnx.org/content/m40709/latest/3.12.png/image>



We need to increase the number of decimal places so we will click on the up arrow. Clicking one time on the up arrow will change the number of decimals from 1 to 2. Now we will do the same for row 2 and for row 3.

¹⁶<http://cnx.org/content/m40709/latest/3.13.png/image>

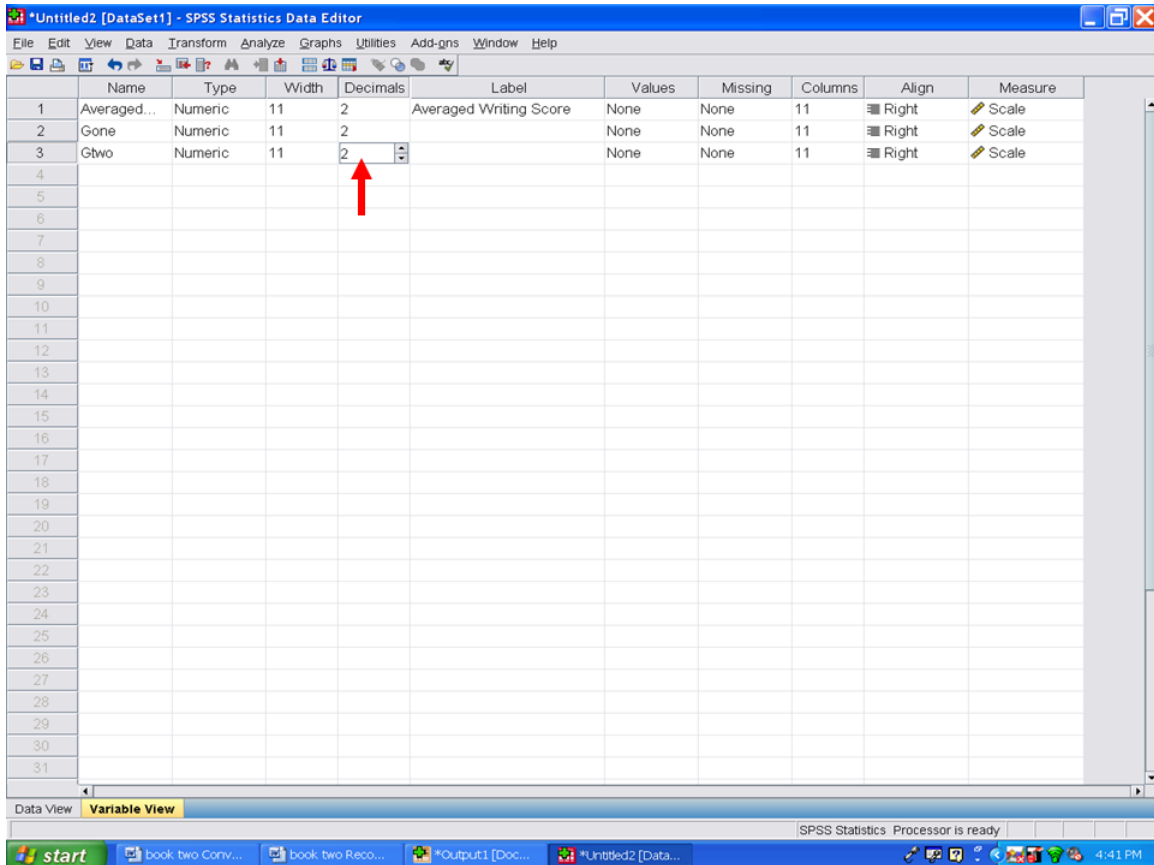
The screenshot shows the SPSS Statistics Data Editor window titled '*Untitled2 [DataSet1] - SPSS Statistics Data Editor'. The interface includes a menu bar (File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, Help) and a toolbar. The main area displays the Variable View for three variables:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	Averaged...	Numeric	11	2	Averaged...	None	None	11	Right	Scale
2	Gone	Numeric	11	1	Gone	None	None	11	Right	Scale
3	Gtwo	Numeric	11	1	Gtwo	None	None	11	Right	Scale
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										
31										

The 'Decimals' column for the first variable is highlighted with a red arrow pointing to the value '2'. The status bar at the bottom indicates 'SPSS Statistics Processor is ready' and the system clock shows '4:39 PM'.

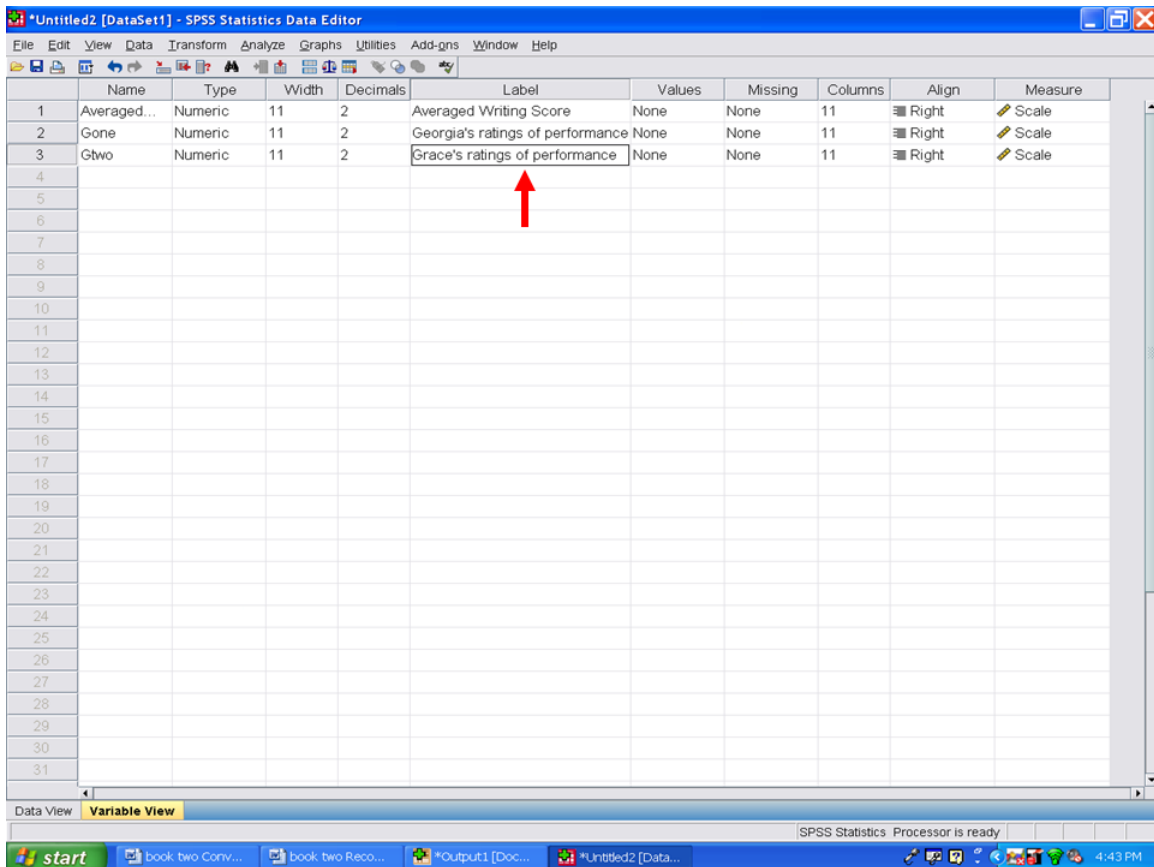
Viewing the screen below will show you that all three variables now have 2 decimal places.

¹⁷ <http://cnx.org/content/m40709/latest/3.14.png/image>



Prior to conducting any statistical analyses, we recommend that you type in identifying information in the Label column for each variable. The information you type in the Label column will be the information printed out in the SPSS output and will make it easier to interpret.

¹⁸<http://cnx.org/content/m40709/latest/3.15.png/image>

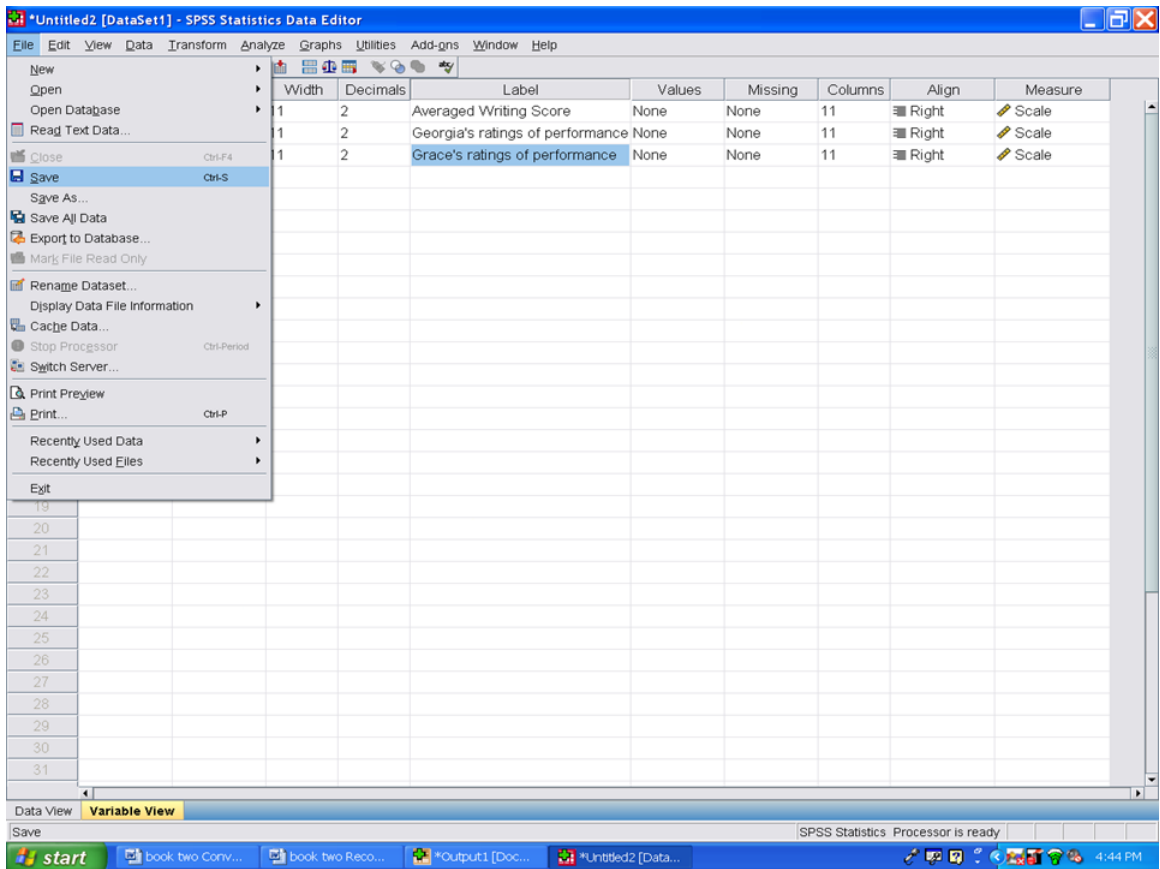


We have converted our Excel file into a SPSS dataset. Please save it prior to conducting any statistical analyses.

- ✓ Click on File
- ✓ Click on Save

Give it a name of your choosing and save it in a location that you will remember later.

¹⁹<http://cnx.org/content/m40709/latest/3.16.png/image>



²⁰<http://cnx.org/content/m40709/latest/3.17.png/image>

Chapter 5

5. Typing Survey Data Into SPSS¹



NOTE: This chapter is published by NCPEA Press² and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of education administration.

About the Authors

John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website (Writing and Statistical Help³) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

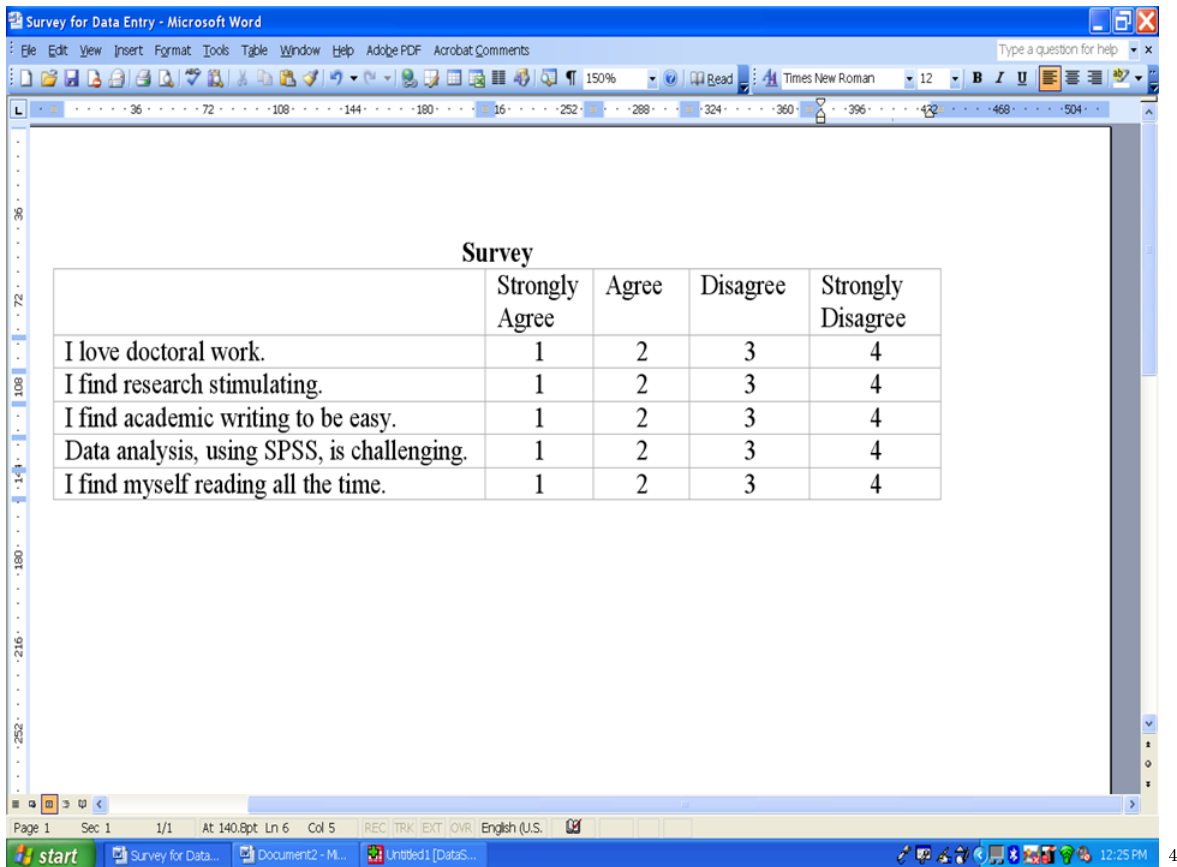
Ana Rojas-LeBouef is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. Dr. LeBoeuf recently completed her doctoral degree in Reading, where she conducted a 16-year analysis of Texas statewide data regarding the achievement gap. Her research interests lie in examining the inequities in achievement among ethnic groups. Dr. Rojas-LeBouef also assists students and faculty in their writing and statistical needs on the Writing and Statistical Help website.

In this chapter, we will show you how to enter data from a survey into SPSS. You can either develop your SPSS data file straight from your survey and then enter in data once your surveys have been completed. Or, you can develop your SPSS data file from your completed surveys. In this set of steps and screenshots, we will assume that you have completed surveys on hand. In our example, we will use the following survey:

¹This content is available online at <<http://cnx.org/content/m40713/1.2/>>.

²<http://www.ncpeapublications.org/books.html>

³<http://cnx.org/content/m40713/latest/www.writingandstatisticalhelp>

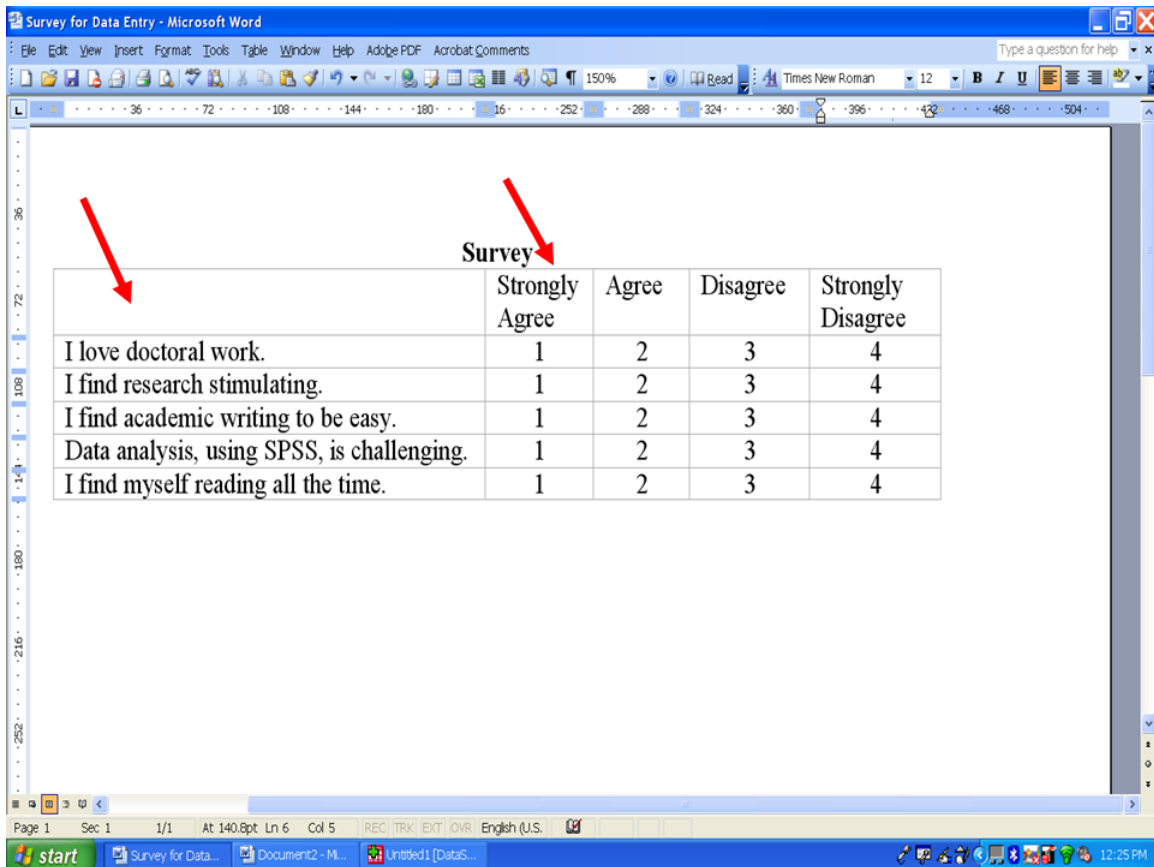


The screenshot shows a Microsoft Word document titled "Survey for Data Entry". The document contains a table with the following data:

Survey				
	Strongly Agree	Agree	Disagree	Strongly Disagree
I love doctoral work.	1	2	3	4
I find research stimulating.	1	2	3	4
I find academic writing to be easy.	1	2	3	4
Data analysis, using SPSS, is challenging.	1	2	3	4
I find myself reading all the time.	1	2	3	4

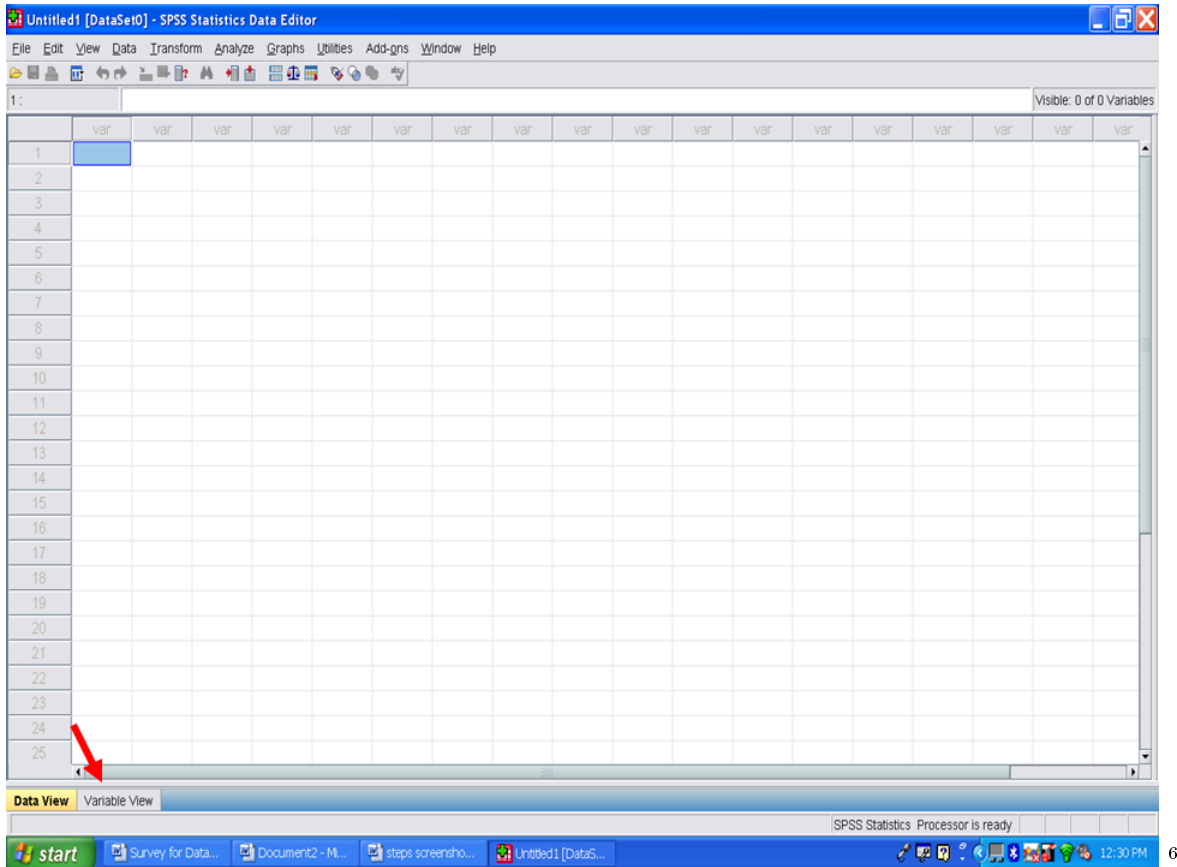
You will note that this survey consists of five items, with a 4-point Likert format. We have assigned a value of 1 to Strongly Agree; a value of 2 to Agree; a value of 3 to Disagree; and a value of 4 to Strongly Disagree. You could, of course, reverse these values so that Strongly Disagree could be a 1.

⁴<http://cnx.org/content/m40713/latest/4.1.png/image>



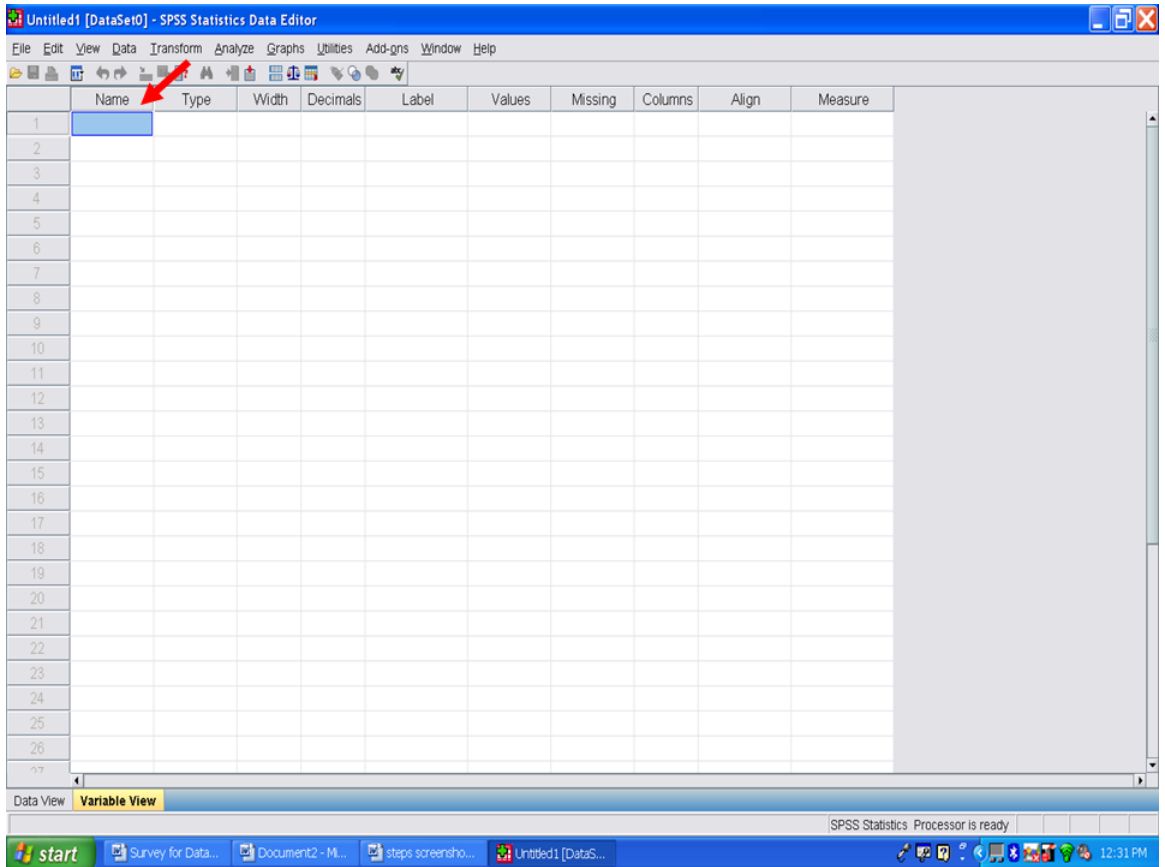
Open up your SPSS so that you have a blank data screen in front of you. Click on the Variable View button at the bottom left of your screen.

⁵ <http://cnx.org/content/m40713/latest/4.2.png/image>



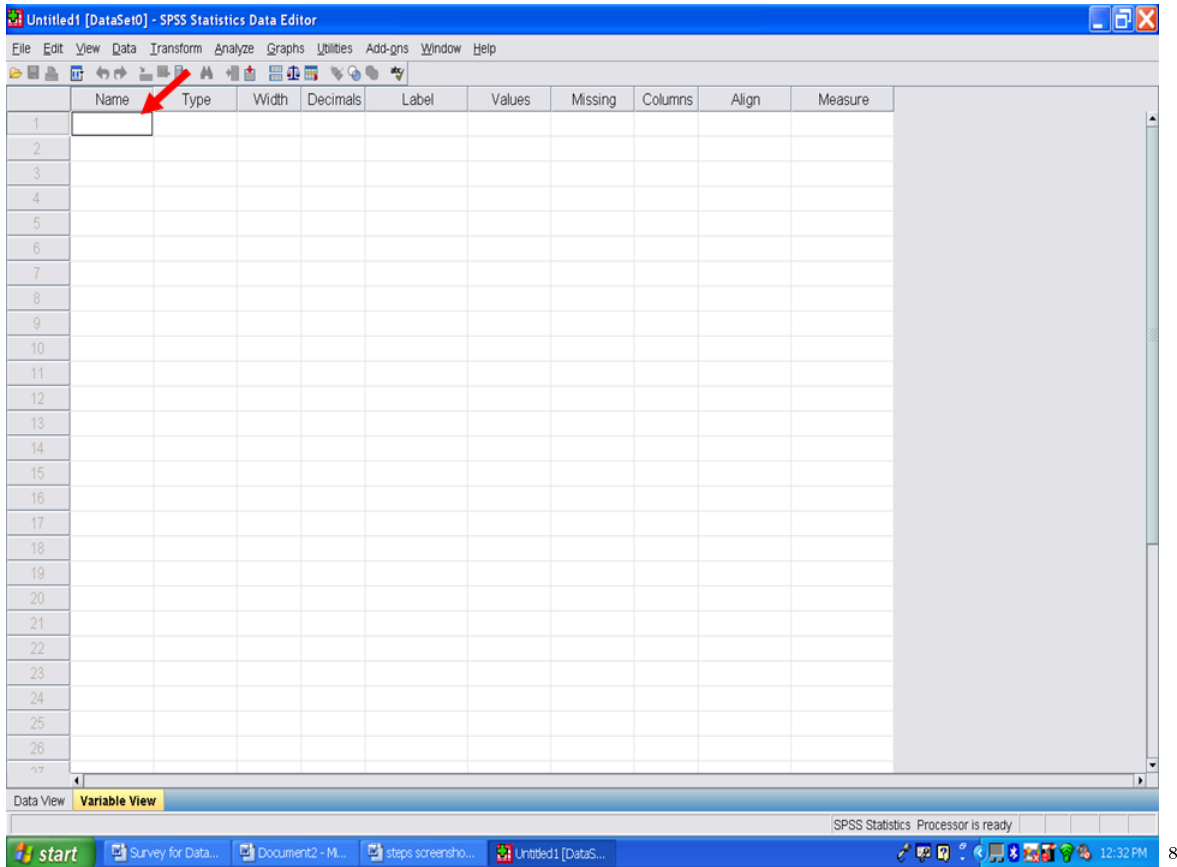
We are now at the screen where we can enter in variable names for our survey items. For simplicity, we will refer to the first Survey Item as S1; the second Survey Item as S2; and so on. We will click in the first cell, row one, under Name.

⁶ <http://cnx.org/content/m40713/latest/4.3.png/image>



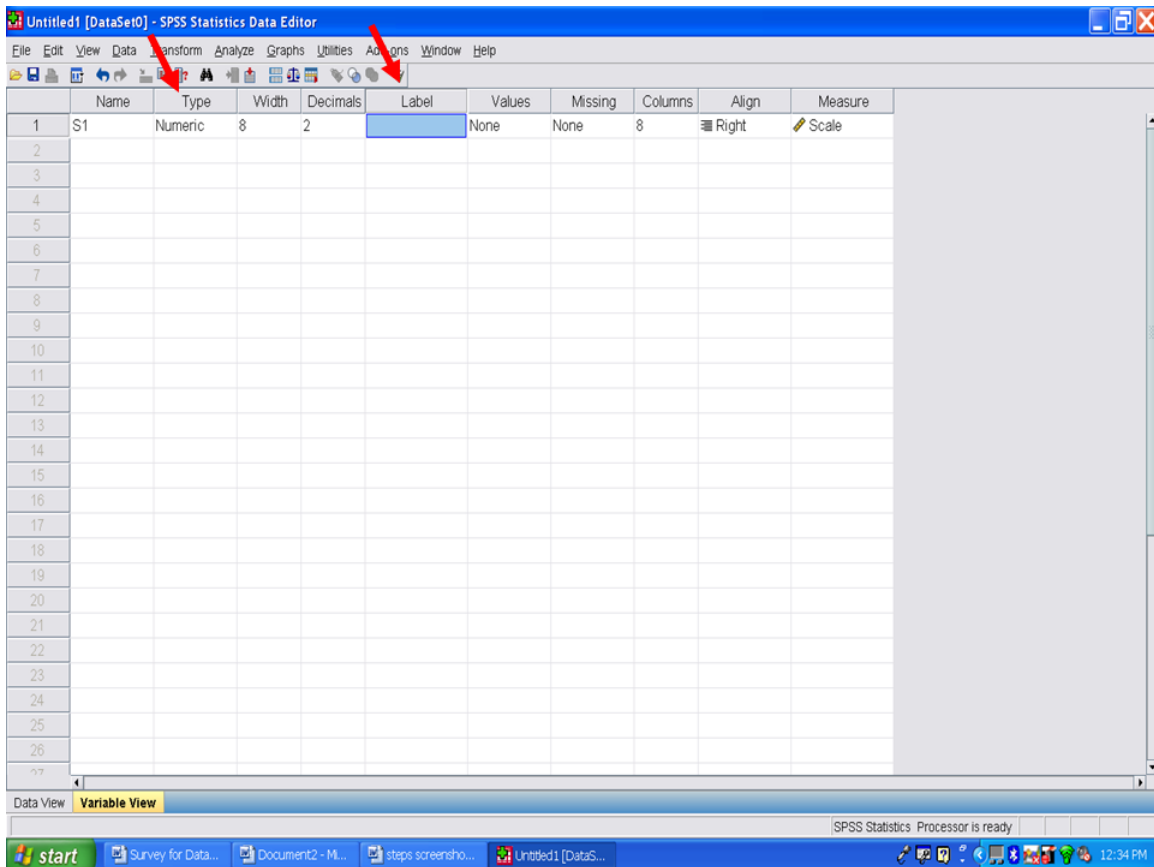
Doing this permits us to type in S1 for our first Survey Item.

⁷ <http://cnx.org/content/m40713/latest/4.4.png/image>



After typing in S1, then move your cursor to the column that reads Label. Once you have typed in S1 and moved your cursor out of that cell, SPSS will fill in the default values for that variable. This variable is a numeric one. In the column marked Label, we will place in the survey item from the Word file.

⁸<http://cnx.org/content/m40713/latest/4.5.png/image>



If you have the survey in Word format, have your Word file opened and copy the first survey item.

⁹<http://cnx.org/content/m40713/latest/4.6.png/image>

Survey

	Strongly Agree	Agree	Disagree	Strongly Disagree
I love doctoral work.	1	2	3	4
I find research stimulating.	1	2	3	4
I find academic writing to be easy.	1	2	3	4
Data analysis, using SPSS, is challenging.	1	2	3	4
I find myself reading all the time.	1	2	3	4

Go back to your SPSS datafile and paste the copied survey item in the Label column, first row.

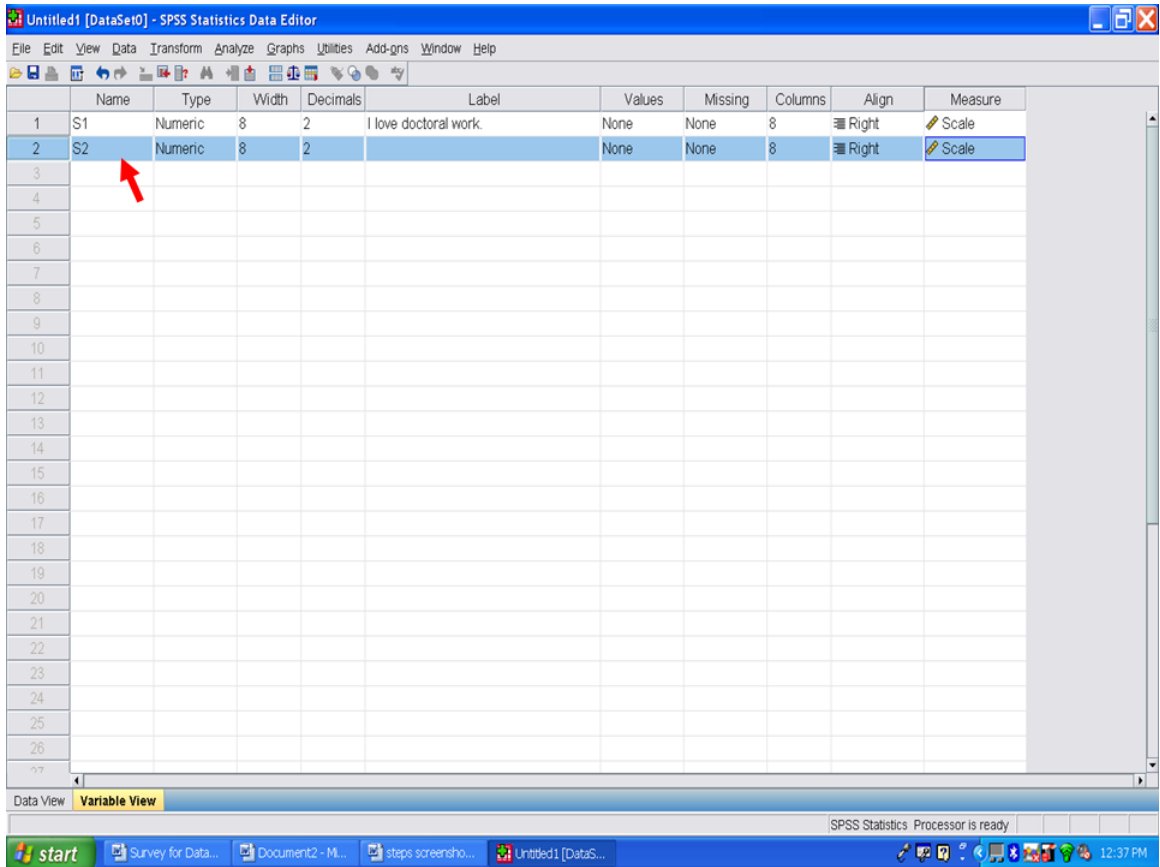
¹⁰<http://cnx.org/content/m40713/latest/4.71.png/image>

The screenshot shows the SPSS Statistics Data Editor window titled 'Untitled1 [DataSet0] - SPSS Statistics Data Editor'. The 'Variable View' tab is active, displaying a table with columns: Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, and Measure. The first row, labeled '1', contains the following data: Name: S1, Type: Numeric, Width: 8, Decimals: 2, Label: love doctoral work., Values: None, Missing: None, Columns: 8, Align: Right, Measure: Scale. A red arrow points to the 'Label' column header. The 'Data View' tab is also visible at the bottom left. The Windows taskbar at the bottom shows the Start button and several open applications, including 'Survey for Data...', 'Document2 - M...', 'steps screenho...', and 'Untitled1 [DataS...'. The system clock shows 12:37 PM.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	S1	Numeric	8	2	love doctoral work.	None	None	8	Right	Scale
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										

Repeat this process for S2.

¹¹ <http://cnx.org/content/m40713/latest/4.8.png/image>



Go to your Word file and copy the second survey item.

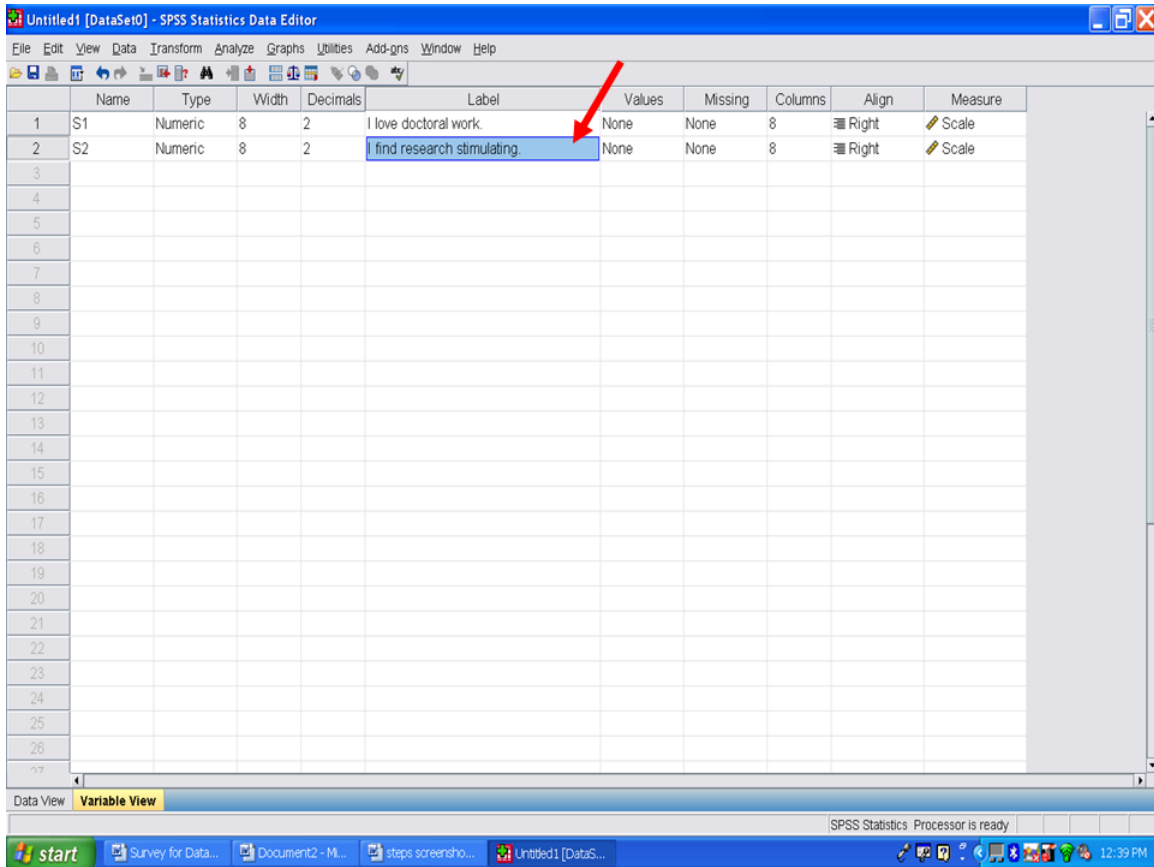
¹²<http://cnx.org/content/m40713/latest/4.9.png/image>

Survey

	Strongly Agree	Agree	Disagree	Strongly Disagree
I love doctoral work.	1	2	3	4
I find research stimulating.	1	2	3	4
I find academic writing to be easy.	1	2	3	4
Data analysis, using SPSS, is challenging.	1	2	3	4
I find myself reading all the time.	1	2	3	4

Paste it in the Label column for the second row.

¹³<http://cnx.org/content/m40713/latest/4.10.png/image>



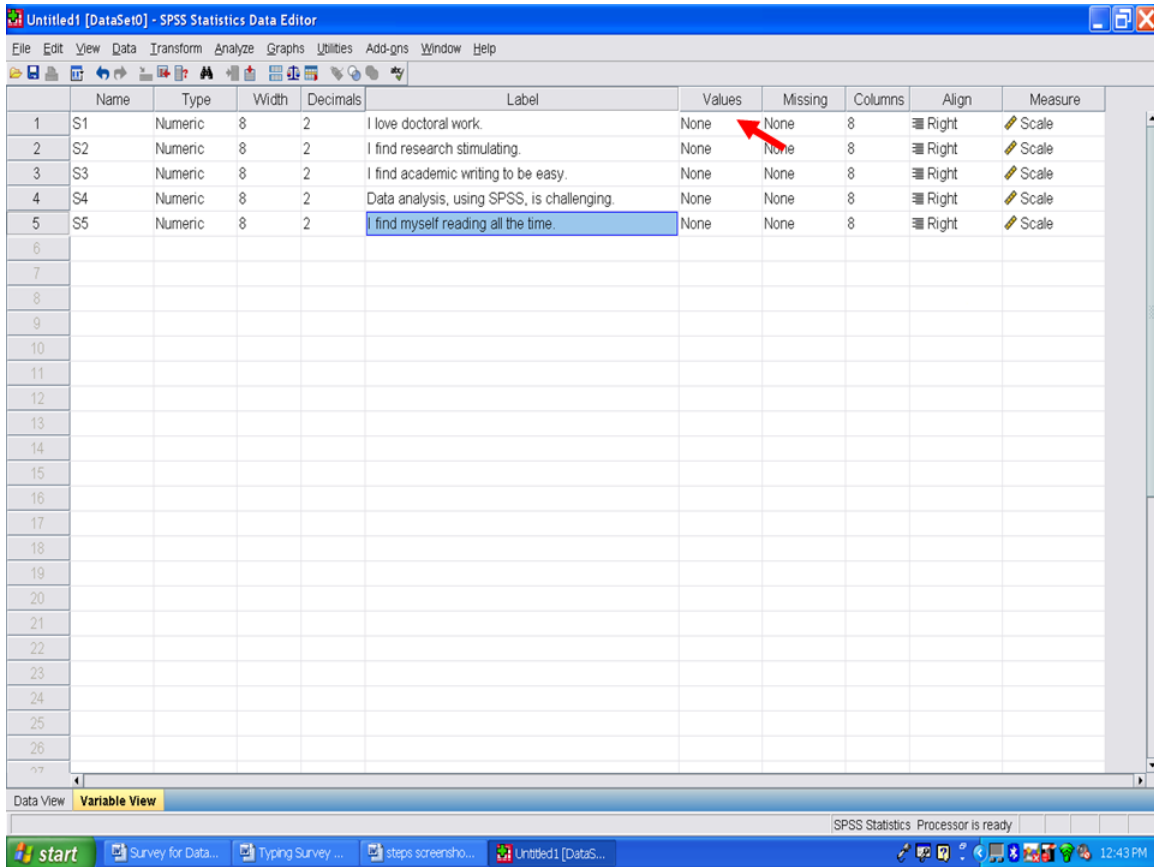
You would repeat the same process for the remaining survey items. After this work, your file would look like the following.

¹⁴<http://cnx.org/content/m40713/latest/4.11.png/image>

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	S1	Numeric	8	2	I love doctoral work.	None	None	8	Right	Scale
2	S2	Numeric	8	2	I find research stimulating.	None	None	8	Right	Scale
3	S3	Numeric	8	2	I find academic writing to be easy.	None	None	8	Right	Scale
4	S4	Numeric	8	2	Data analysis, using SPSS, is challenging.	None	None	8	Right	Scale
5	S5	Numeric	8	2	I find myself reading all the time.	None	None	8	Right	Scale
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										

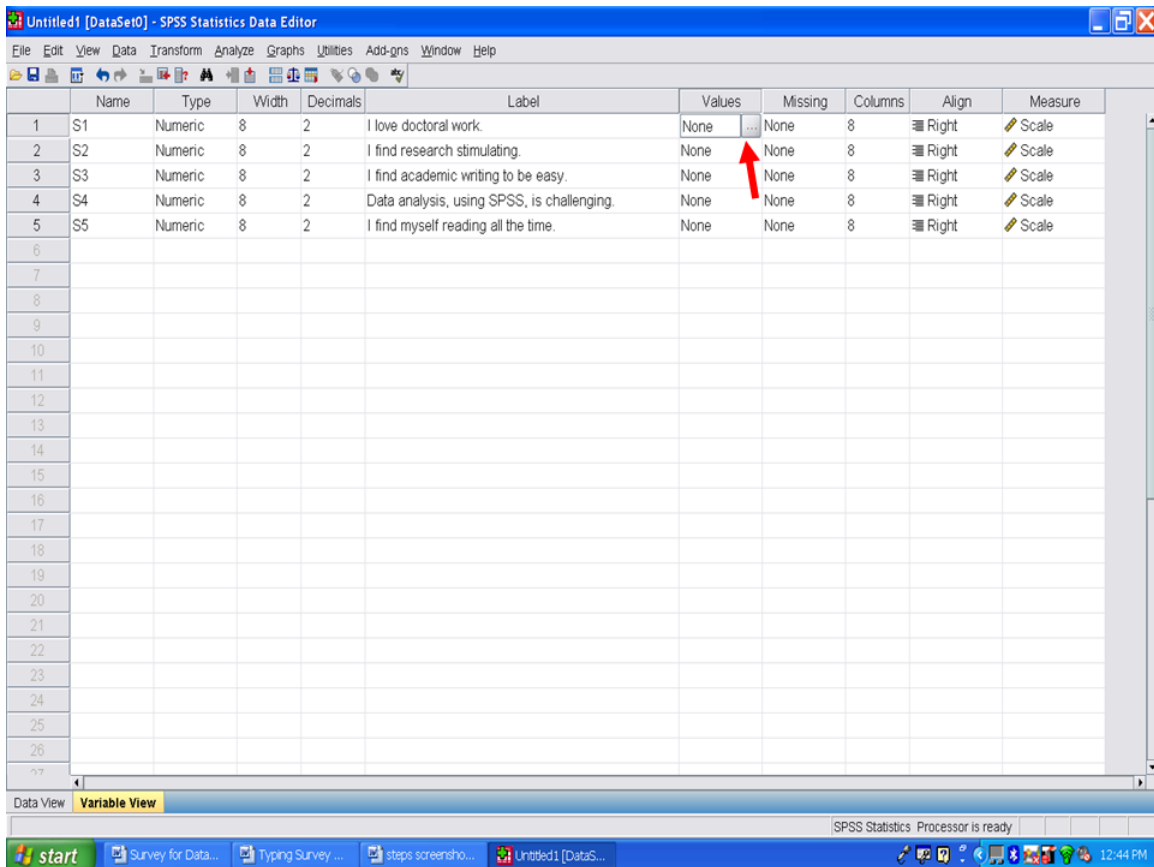
Now we need to type in the values for the 4-point Likert scale. To do so, click on the first cell in row one under the column heading, Values. See the arrow below.

¹⁵<http://cnx.org/content/m40713/latest/4.12.png/image>



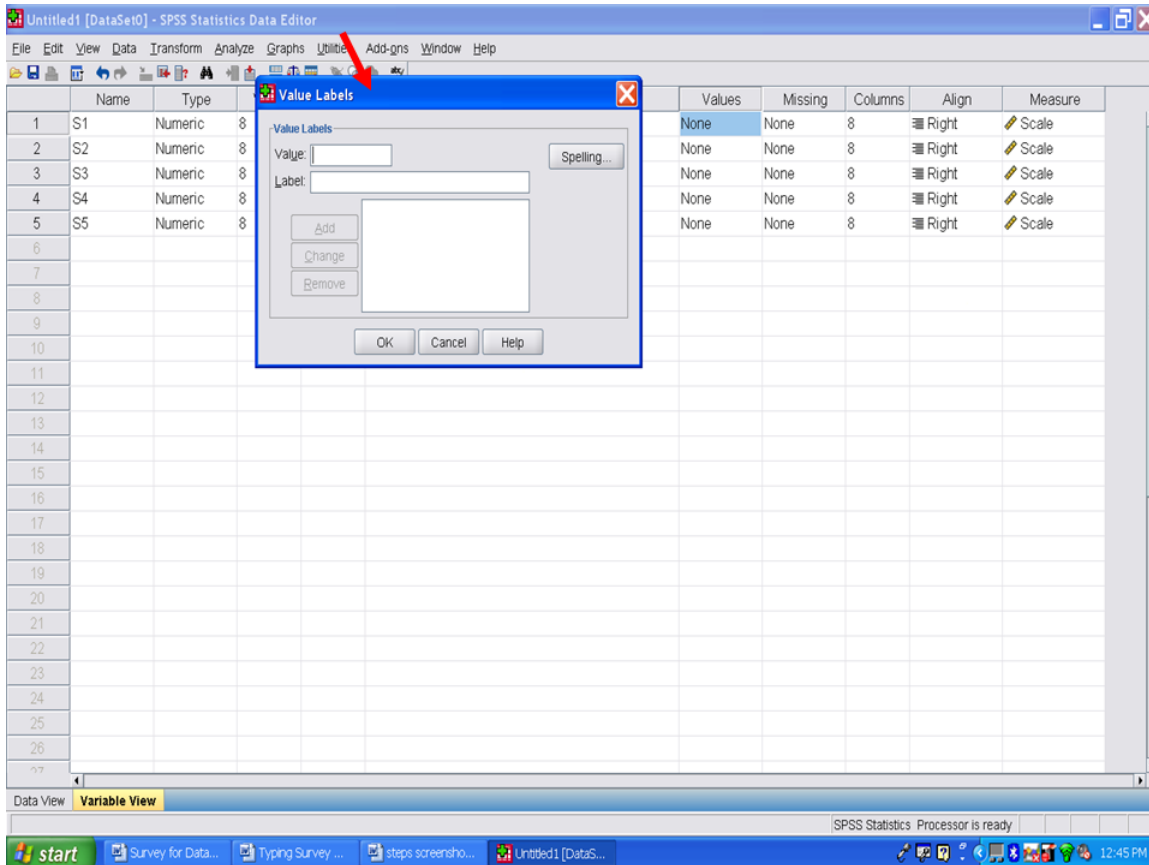
When you click in this cell, three dots appear. These three dots indicate that a screen is hidden beneath this one. Click on the three dots.

¹⁶<http://cnx.org/content/m40713/latest/4.13.png/image>



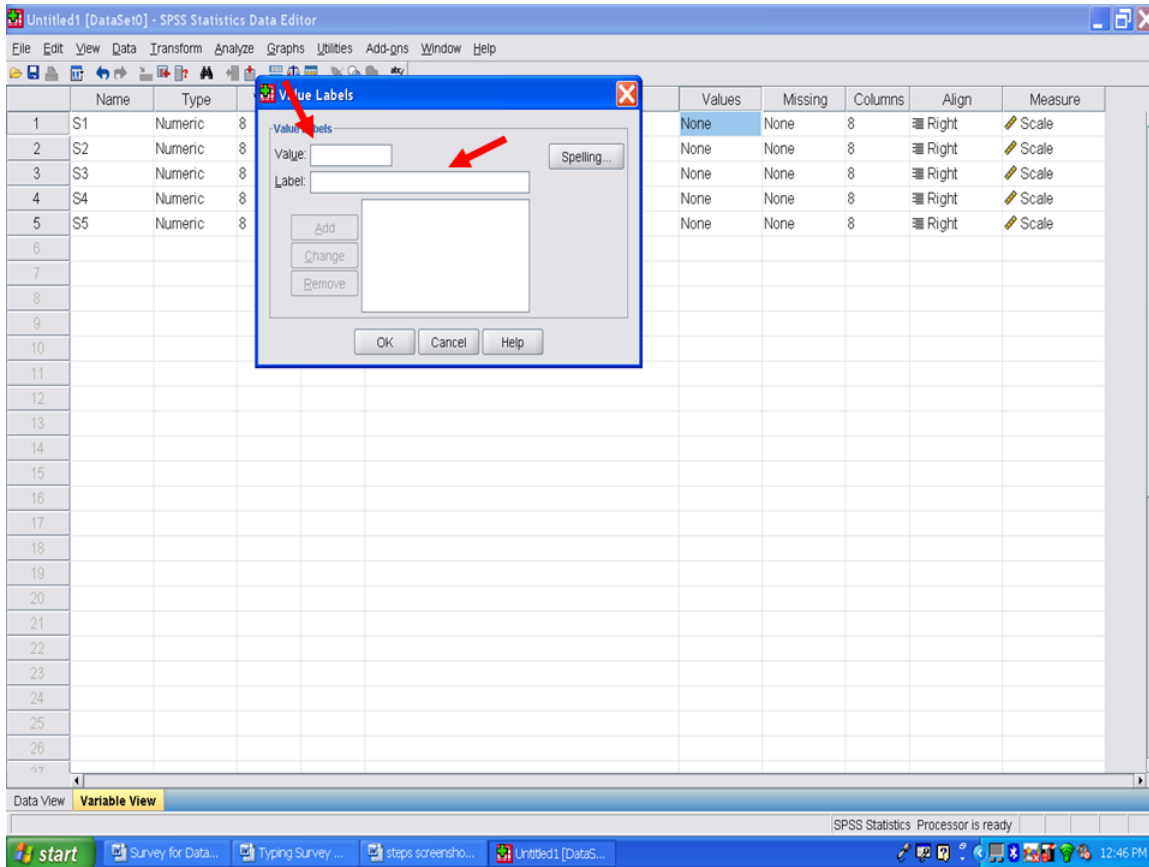
After clicking on the three dots, the following screen will appear.

¹⁷ <http://cnx.org/content/m40713/latest/4.4.png/image>



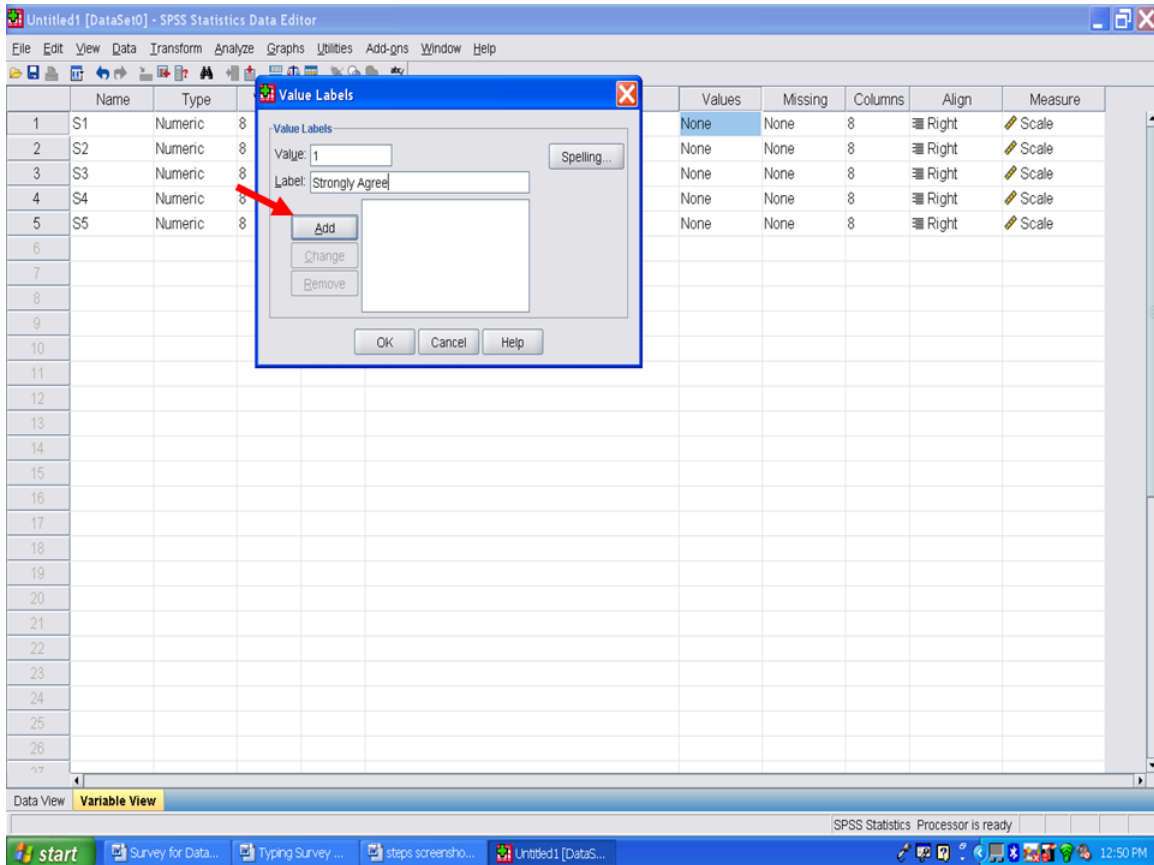
In this screen, we will type in a numeric value in the Value cell and the Likert-format response label in the Label cell. That is, our four values are 1, 2, 3, and 4, which correspond to a Strongly Agree, Agree, Disagree, and Strongly Disagree. Refer back to the survey which we are using for this example.

¹⁸<http://cnx.org/content/m40713/latest/4.15png/image>



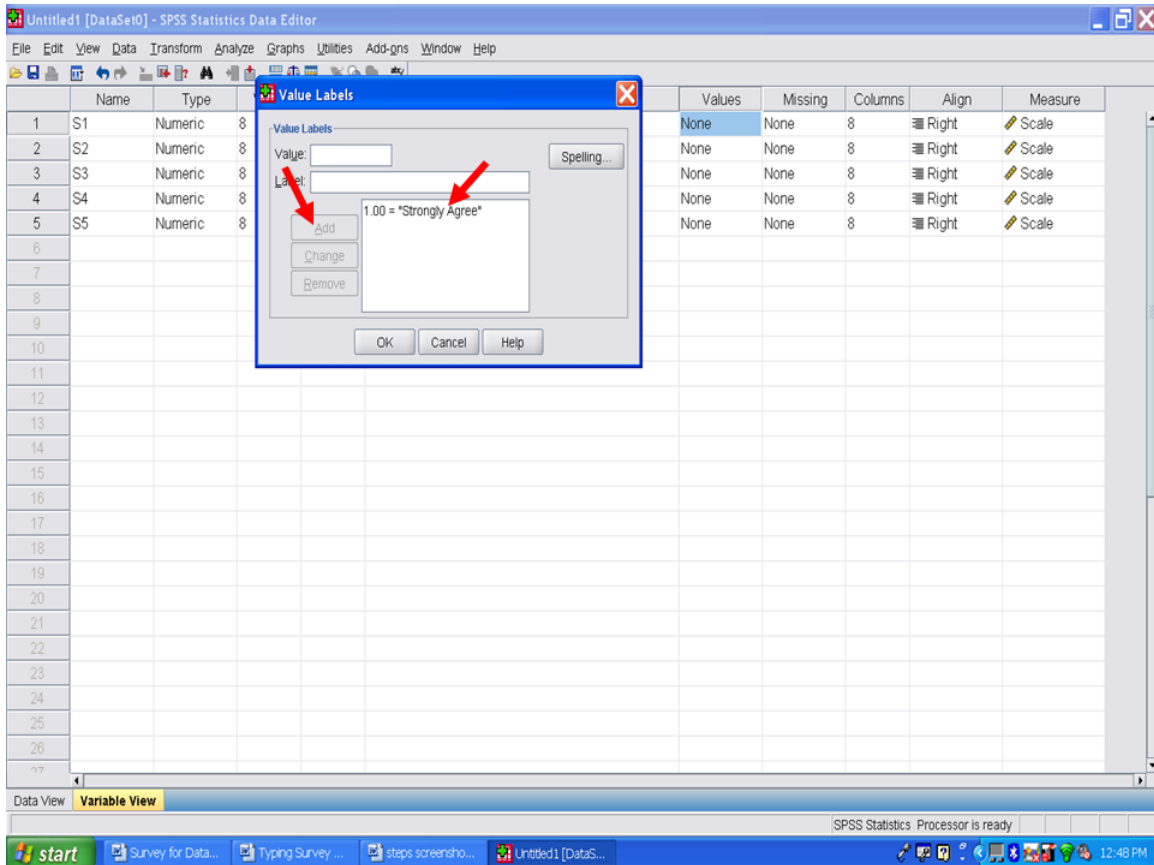
After typing in a 1 in Value and Strongly Agree in Label, we will now click on Add.

¹⁹<http://cnx.org/content/m40713/latest/4.16png/image>



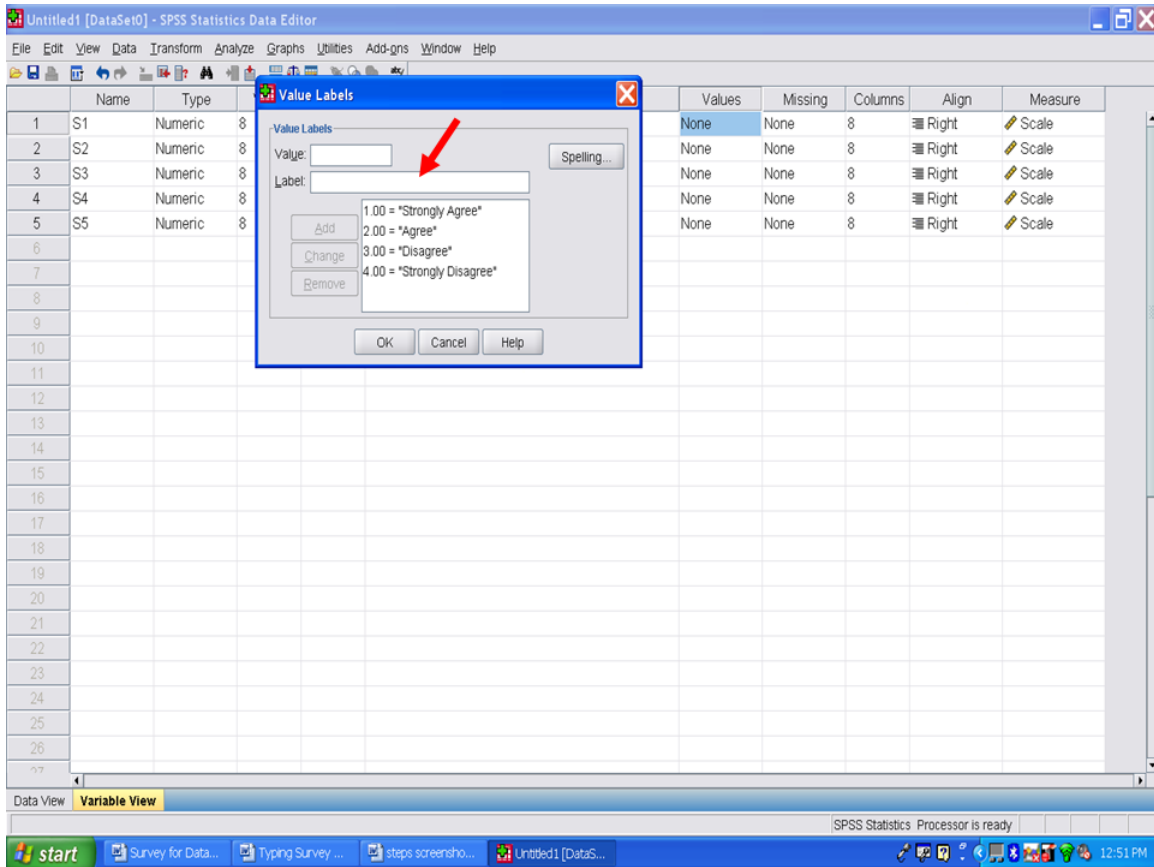
After clicking on Add, your screen should look like the one below.

²⁰<http://cnx.org/content/m40713/latest/4.17.png/image>



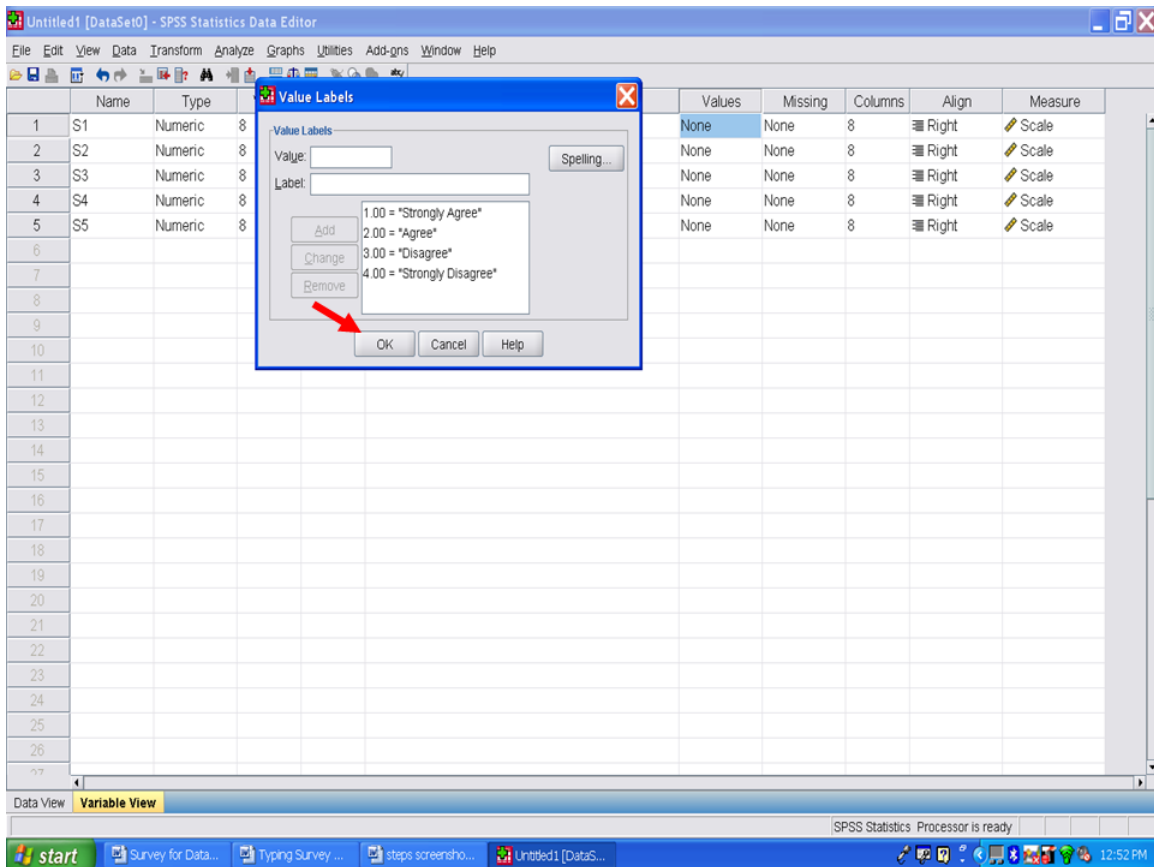
We will repeat this process for 2 Agree; 3 Disagree; and 4 Strongly Disagree. Your screen should look like the one below.

²¹ <http://cnx.org/content/m40713/latest/4.18.png/image>



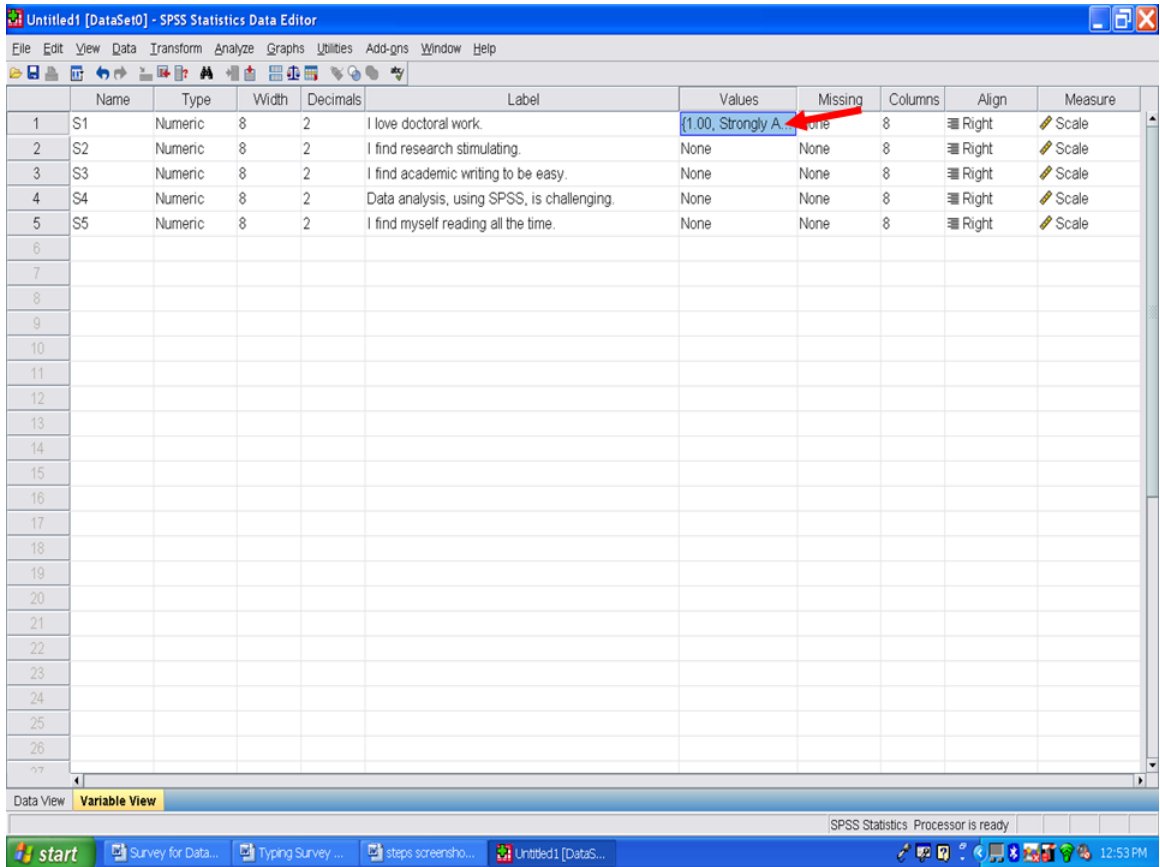
Now click on OK.

²²<http://cnx.org/content/m40713/latest/4.19.png/image>



Your screen should now look like the following one. The 4-point Likert scale has been successfully entered for the first survey item. Now we will copy it and paste it one at a time for the other four survey items.

²³<http://cnx.org/content/m40713/latest/4.20.png/image>



Right mouse click on the cell that you just completed.
Copy.

²⁴<http://cnx.org/content/m40713/latest/4.21.png/image>

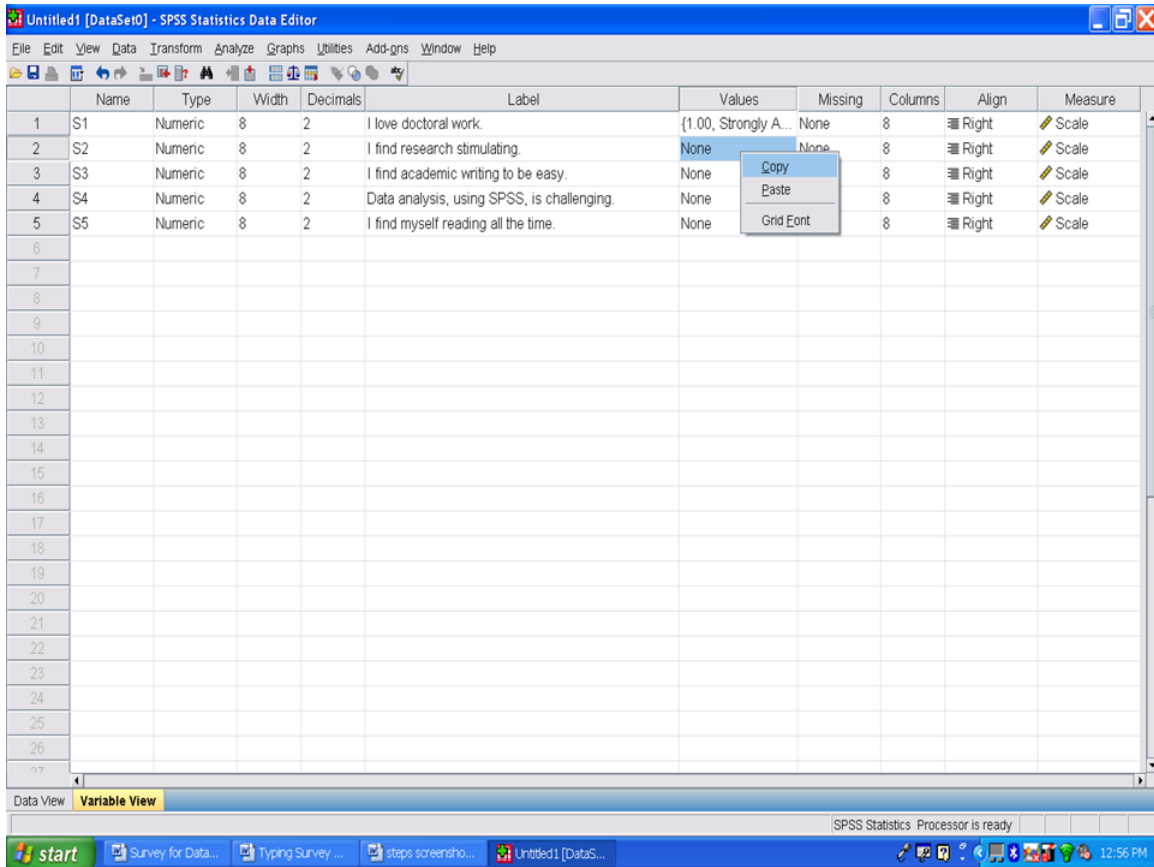
The screenshot shows the SPSS Statistics Data Editor in Variable View. The table below represents the data structure shown in the interface:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	S1	Numeric	8	2	I love doctoral work.	1, 00, Strongly A	None	8	Right	Scale
2	S2	Numeric	8	2	I find research stimulating.	None	None	8	Right	Scale
3	S3	Numeric	8	2	I find academic writing to be easy.	None	None	8	Right	Scale
4	S4	Numeric	8	2	Data analysis, using SPSS, is challenging.	None	None	8	Right	Scale
5	S5	Numeric	8	2	I find myself reading all the time.	None	None	8	Right	Scale
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										

The interface also shows the menu bar (File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, Help), a toolbar, and a taskbar at the bottom with the text 'SPSS Statistics Processor is ready' and a system clock showing 12:55 PM.

**Move to the second cell in the Label column.
Paste.**

²⁵ <http://cnx.org/content/m40713/latest/4.22.png/image>



Your screen will now have that cell filled in with the same information that you typed in for survey item one. Repeat this paste process for survey items 3, 4, and 5.

²⁶<http://cnx.org/content/m40713/latest/4.23.png/image>

The screenshot shows the SPSS Statistics Data Editor window titled 'Untitled1 [DataSet0] - SPSS Statistics Data Editor'. The 'Variable View' tab is active, displaying a table with columns for Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, and Measure. Five variables are defined: S1, S2, S3, S4, and S5, all of type 'Numeric' with a width of 8 and 2 decimal places. The labels for these variables are: 'I love doctoral work.', 'I find research stimulating.', 'I find academic writing to be easy.', 'Data analysis, using SPSS, is challenging.', and 'I find myself reading all the time.' respectively. The 'Values' column shows labels for each variable: S1 and S2 have labels '{1.00, Strongly A...}', while S3, S4, and S5 have 'None'. The 'Missing' column is empty for all variables. The 'Columns' column is set to 8 for all. The 'Align' column is set to 'Right' for all. The 'Measure' column is set to 'Scale' for all. The bottom status bar indicates 'SPSS Statistics Processor is ready' and the system clock shows '12:56 PM'.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	S1	Numeric	8	2	I love doctoral work.	{1.00, Strongly A...	None	8	Right	Scale
2	S2	Numeric	8	2	I find research stimulating.	{1.00, Strongly A...	None	8	Right	Scale
3	S3	Numeric	8	2	I find academic writing to be easy.	None	None	8	Right	Scale
4	S4	Numeric	8	2	Data analysis, using SPSS, is challenging.	None	None	8	Right	Scale
5	S5	Numeric	8	2	I find myself reading all the time.	None	None	8	Right	Scale
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										

Your screen will now look like the following one. All of your survey items are present and typed out in the Label column. All of the numeric values for participants' responses are labeled in the Values column. Remember to save your file.

²⁷<http://cnx.org/content/m40713/latest/4.24.png/image>

The screenshot shows the SPSS Statistics Data Editor window titled 'Untitled1 [DataSet0] - SPSS Statistics Data Editor'. The 'Variable View' is active, displaying a table with the following columns: Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, and Measure. The table contains five rows of variables (S1-S5) and several empty rows (6-27). A red arrow points to the 'Values' column for variable S5.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	S1	Numeric	8	2	I love doctoral work.	{1.00, Strongly A...	None	8	Right	Scale
2	S2	Numeric	8	2	I find research stimulating.	{1.00, Strongly A...	None	8	Right	Scale
3	S3	Numeric	8	2	I find academic writing to be easy.	{1.00, Strongly A...	None	8	Right	Scale
4	S4	Numeric	8	2	Data analysis, using SPSS, is challenging.	{1.00, Strongly A...	None	8	Right	Scale
5	S5	Numeric	8	2	I find myself reading all the time.	{1.00, Strongly A...	None	8	Right	Scale
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										

The taskbar at the bottom shows the Windows Start button, several open applications (Typing Survey..., steps screenho..., Untitled1 [DataS...]), and the system tray with the time 12:58 PM and date 28.

You have now successfully created a SPSS dataset for your survey.

²⁸<http://cnx.org/content/m40713/latest/4.25.png/image>

Chapter 6

6. Internal Consistency Analysis: Part I¹



NOTE: This chapter is published by NCPEA Press² and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of education administration.

About the Authors

John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website (Writing and Statistical Help³) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

Ana Rojas-LeBouef is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. Dr. LeBouef recently completed her doctoral degree in Reading, where she conducted a 16-year analysis of Texas statewide data regarding the achievement gap. Her research interests lie in examining the inequities in achievement among ethnic groups. Dr. Rojas-LeBouef also assists students and faculty in their writing and statistical needs on the Writing and Statistical Help website.

In this set of steps and screenshots, you will be presented with information regarding how to determine the internal consistency (i.e., reliability) of scores derived from a larger set of variables. In particular, this chapter follows up on the factor analysis chapter. The subscales that were determined to constitute Factor 1 and the subscales that were determined to comprise Factor 2 will be the variables used in this chapter to ascertain the extent to which they constitute internally consistent factors. For more detailed information regarding

¹This content is available online at <<http://cnx.org/content/m40717/1.2/>>.

²<http://www.ncpeapublications.org/books.html>

³<http://cnx.org/content/m40717/latest/www.writingandstatisticalhelp>

internal consistency analysis, readers are referred to the *Hyperstats Online Statistics Textbook* at <http://davidmlane.com/hyperstat/>⁴; to the *Electronic Statistics Textbook* (2011) at <http://www.statsoft.com/textbook/>⁵; or to Andy Field's (2009) *Discovering Statistics Using SPSS* at http://www.amazon.com/Discovering-Statistics-Introducing-Statistical-Method/dp/1847879071/ref=sr_1_1?s=books&ie=UTF8&qid=1304967862&sr=1-1⁶

Conducting an internal consistency analysis is usually part of a set of statistical analyses. Its purpose is to determine score reliability and not to determine whether groups differ on some variable. As such, internal consistency analysis precedes the use of inferential statistical procedures.

Have your data set pulled into SPSS-PC. In this data set, we have the same 10 subscales used in the factor analysis chapter.

	pc	inf	cod	sim	pa	ari	bd	voc	oa	comp	var	var	var	var
1	10	7	8	7	10	5	7	7	6	7				
2	8	8	7	6	7	9	9	4	4	7				
3	10	8	2	9	10	11	9	6	11	8				
4	3	4	11	4	7	6	3	4	10	6				
5	5	7	17	6	8	10	7	5	9	1				
6	2	5	8	8	9	7	10	3	5	5				
7	7	4	5	8	7	11	11	2	12	5				
8	7	7	7	7	14	8	10	8	11	3				
9	9	6	10	6	9	8	4	5	8	5				
10	5	9	7	10	9	12	11	7	8	11				
11	6	7	5	8	7	7	7	4	7	5				
12	13	8	5	13	10	5	6	9	9	7				
13	7	8	11	6	8	4	3	3	4	4				
14	6	6	8	12	9	9	9	10	6	8				
15	7	8	3	1	9	8	5	10	3	9				
16	10	7	5	5	9	5	5	8	3	6				
17	12	8	7	8	13	10	11	7	6	5				
18	15	7	5	10	15	8	7	8	7	10				
19	9	6	4	8	3	5	8	6	10	3				
20	6	7	10	6	9	6	8	6	5	2				
21	13	7	8	7	9	6	3	6	6	4				
22	10	7	12	8	8	8	12	5	11	4				
23	10	9	10	10	11	11	14	11	11	10				
24	6	6	6	2	5	7	6	6	11	6				
25	14	7	6	9	10	6	7	6	8	5				

Click on Analyze

Click on Scale

Click on Reliability Analysis

⁴<http://davidmlane.com/hyperstat/>

⁵<http://www.statsoft.com/textbook/>

⁶http://www.amazon.com/Discovering-Statistics-Introducing-Statistical-Method/dp/1847879071/ref=sr_1_1?s=books&ie=UTF8&qid=13049

1

⁷<http://cnx.org/content/m40717/latest/5.1.png/image>

factor analysis dataset for book.sav [Data12] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1: pc 10.0

Visible: 10 of 10 Variables

	pc	inf	pa	ari	bd	voc	oa	comp	var	var	var	val
1	10		10	5	7	7	6	7				
2	8		7	9	9	4	4	7				
3	10		10	11	9	6	11	8				
4	3		7	6	3	4	10	6				
5	5		8	10	7	5	9	1				
6	2		9	7	10	3	5	5				
7	7		7	11	11	2	12	5				
8	7						8	11	3			
9	9						5	8	5			
10	5		9	12	11	7	8	11				
11	6		7	7	7	4	7	5				
12	13		10	5	6	9	9	7				
13	7		8	4	3	3	4	4				
14	6		9	9	9	10	6	8				
15	7	8	3	1	9	8	5	10	3	9		
16	10	7	5	5	9	5	8	3	6			
17	12	8	7	8	13	10	11	7	6	5		
18	15	7	5	10	15	8	7	8	7	10		
19	9	6	4	8	3	5	8	6	10	3		
20	6	7	10	6	9	6	8	6	5	2		
21	13	7	8	7	9	6	3	6	6	4		
22	10	7	12	8	8	8	12	5	11	4		
23	10	9	10	10	11	11	14	11	11	10		
24	6	6	6	2	5	7	6	6	11	6		
25	14	7	6	9	10	6	7	6	8	5		

Data View Variable View

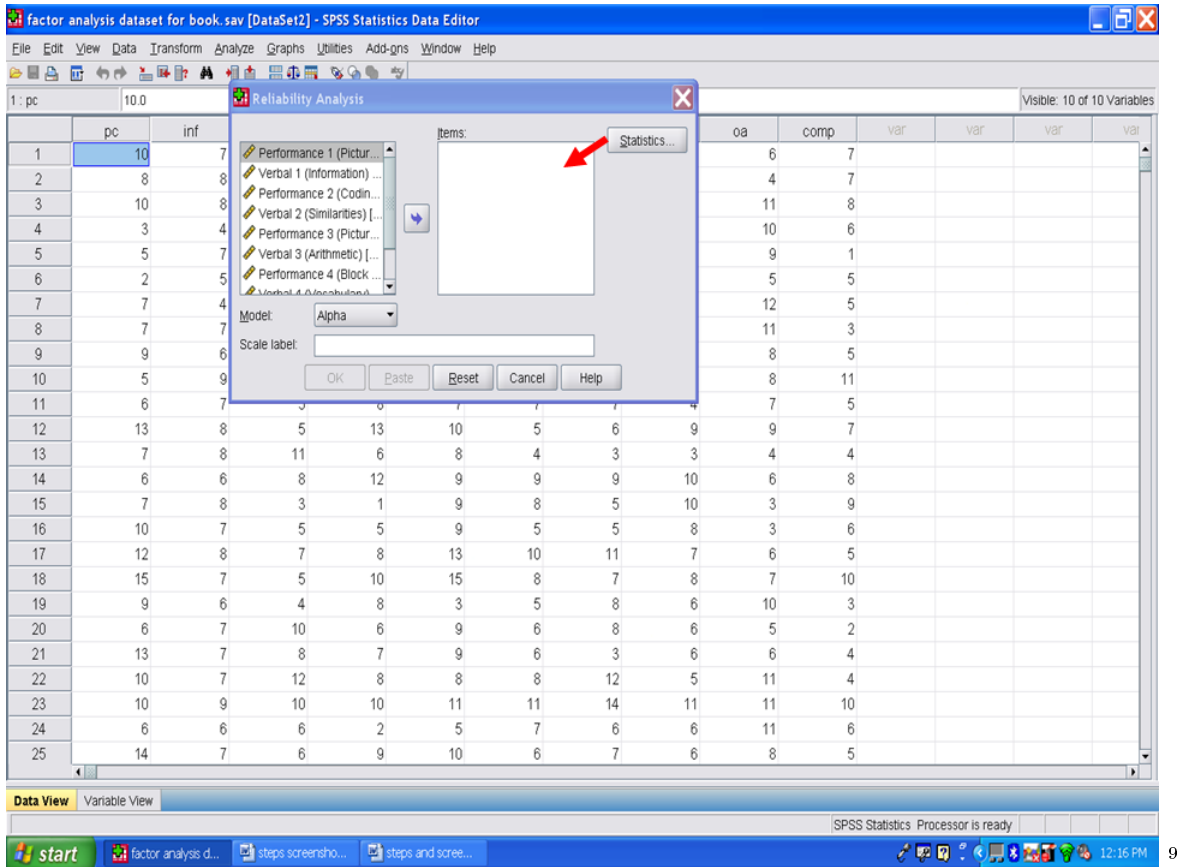
Scale SPSS Statistics Processor is ready

start factor analysis d... steps screeno...

12:08 PM 8

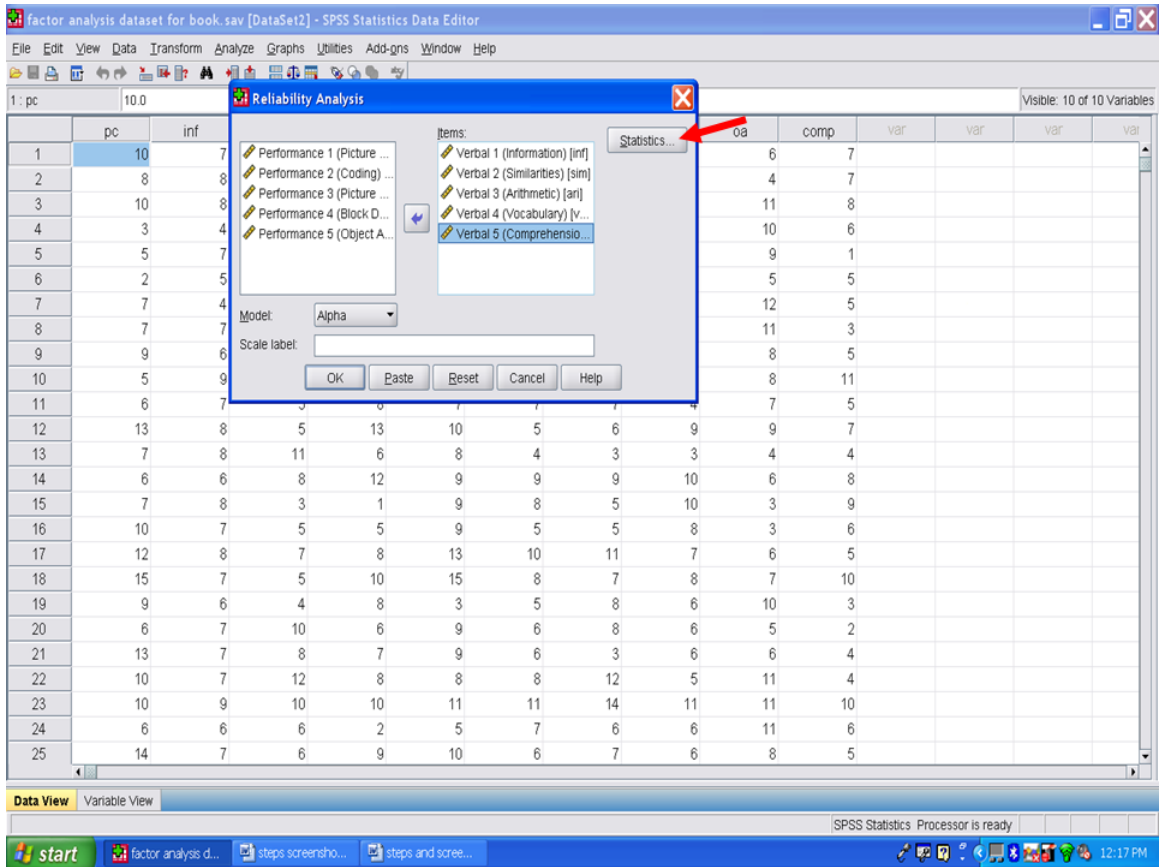
Using the information from the chapter on factor analysis, we want to determine the internal consistency of: Verbal 1 (Information), Verbal 2 (Similarities), Verbal 3 (Arithmetic), Verbal 4 (Vocabulary), and Verbal 5 (Comprehension). Click on each of these variables and send each to the box labeled Items.

⁸ <http://cnx.org/content/m40717/latest/5.2.png/image>



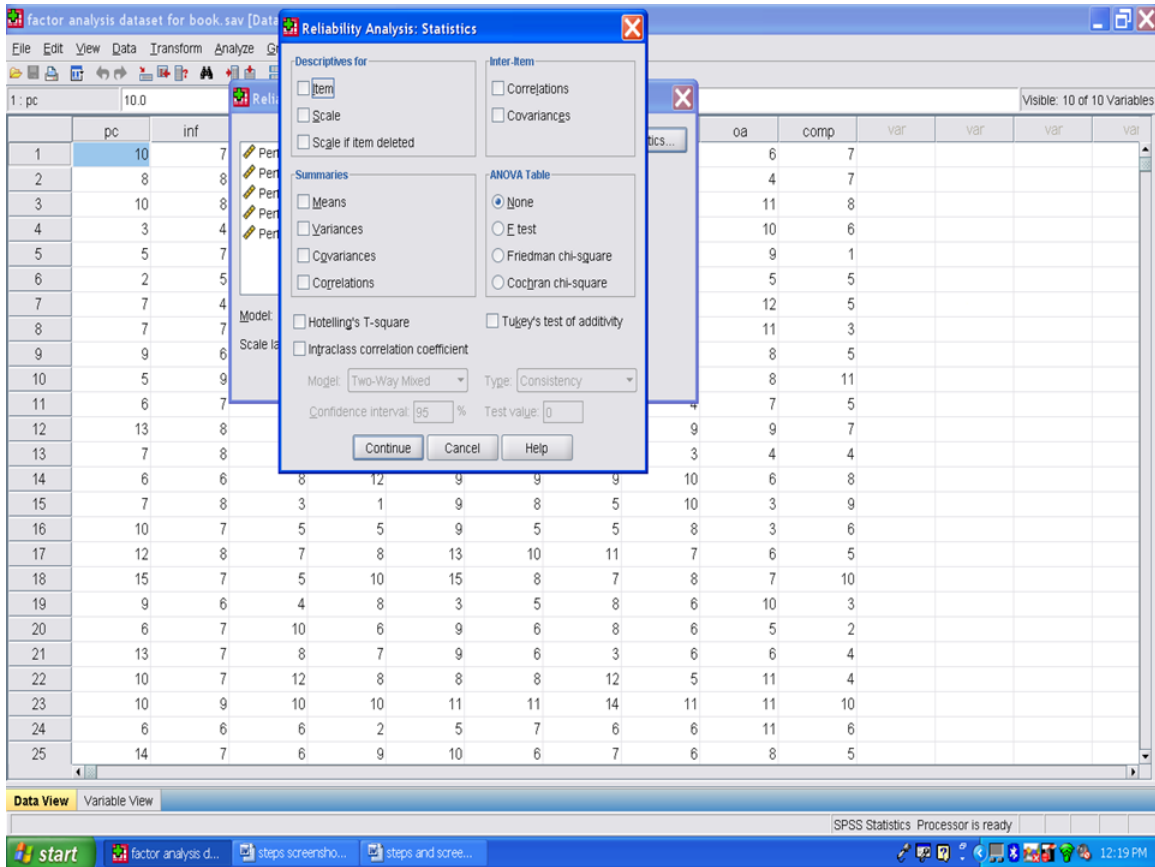
Now your screen should look like the one below.
Click on Statistics.

⁹<http://cnx.org/content/m40717/latest/5.3.png/image>



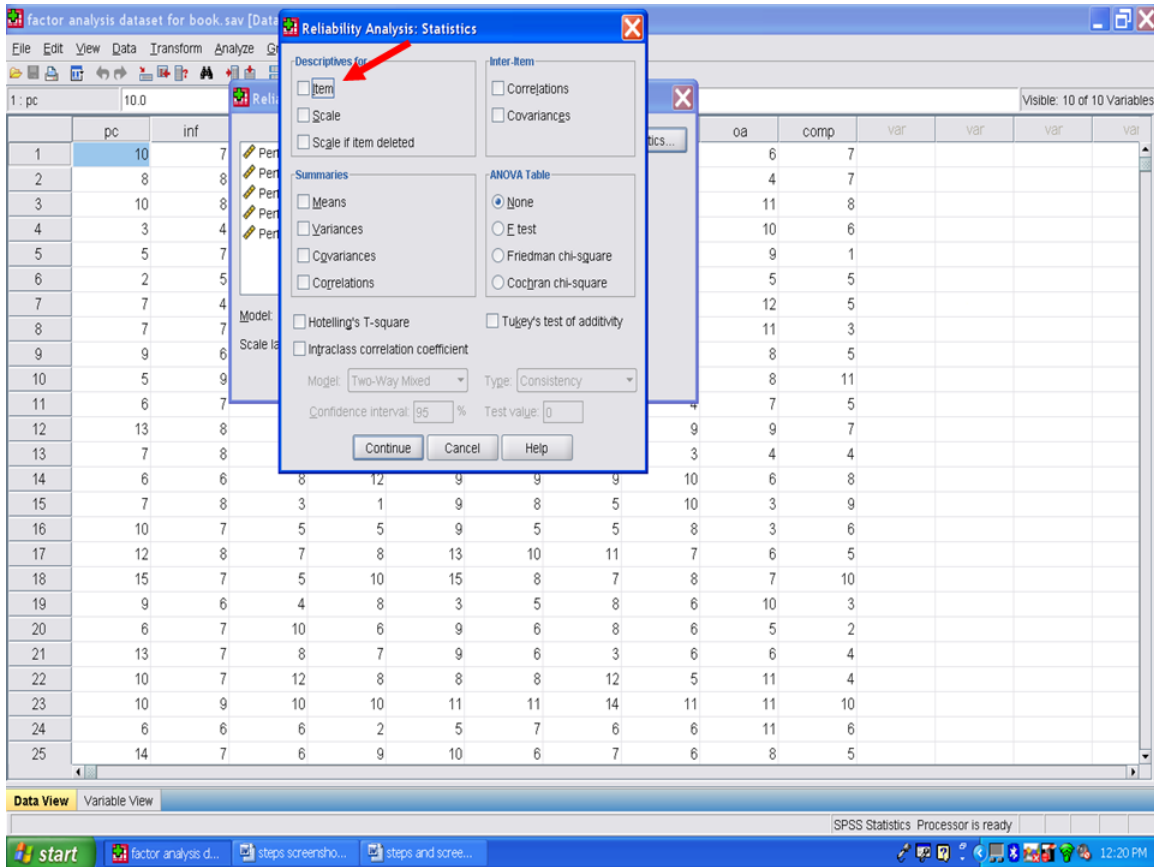
After clicking on Statistics, the following screen should appear.

¹⁰<http://cnx.org/content/m40717/latest/5.4.png/image>



Click under Descriptives on
 Item
 Scale
 Scale if item deleted

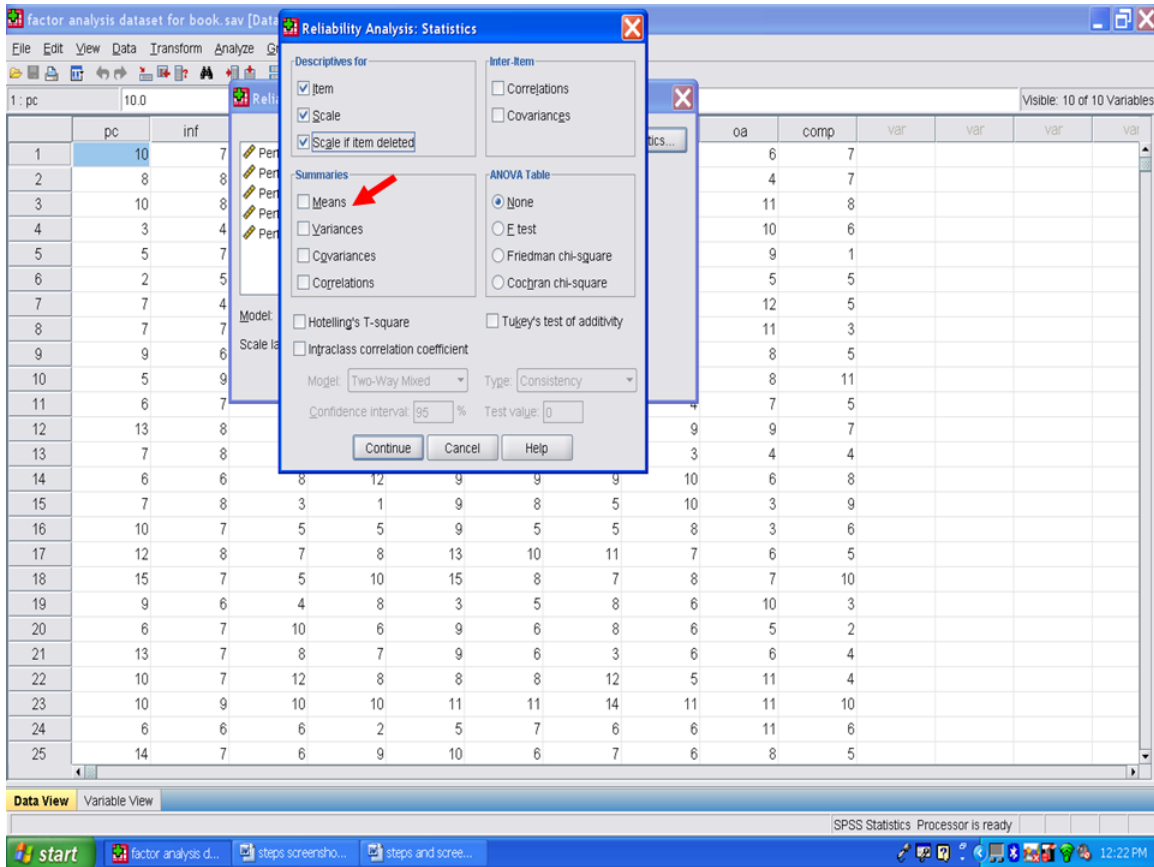
¹¹ <http://cnx.org/content/m40717/latest/5.5.png/image>



After you have clicked on the three items under Descriptives, next click under Summaries on

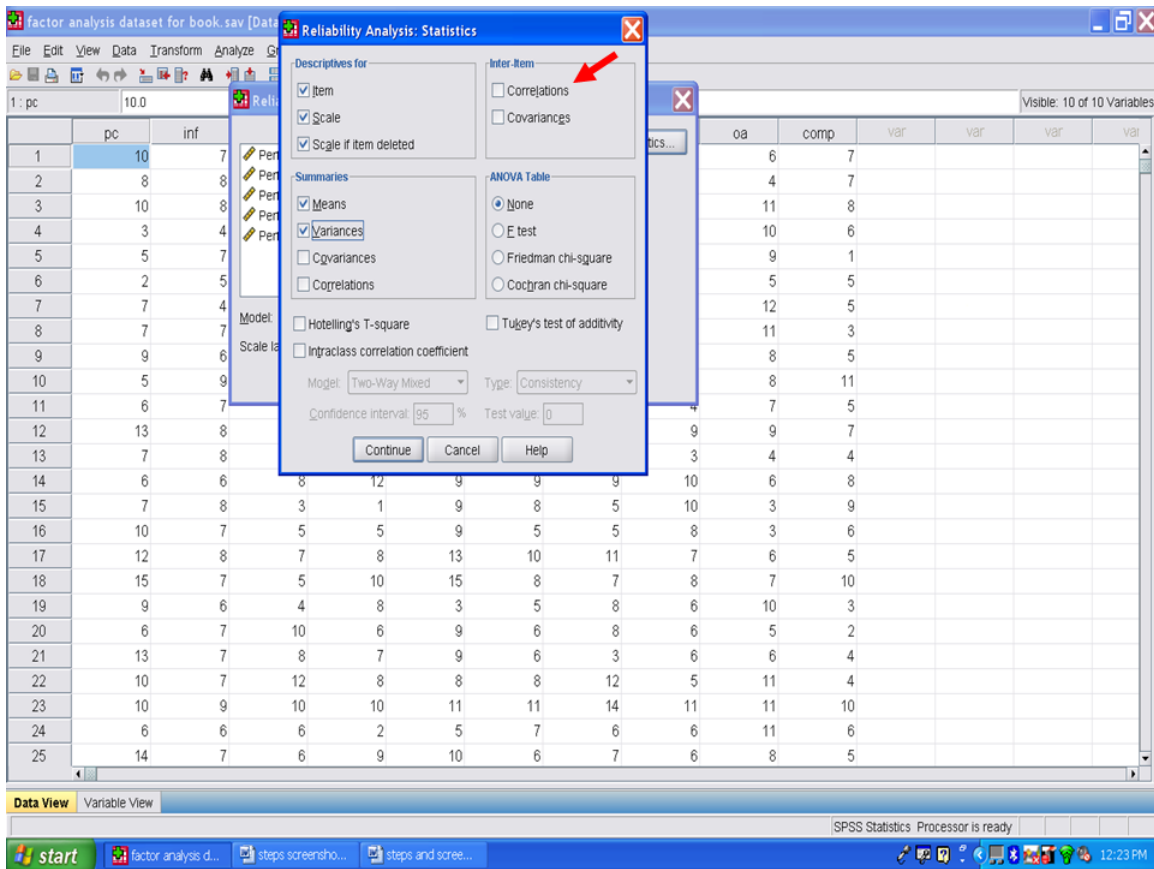
- Means
- Variances

¹²<http://cnx.org/content/m40717/latest/5.6.png/image>



Now that you have clicked on Means and Variances under Summaries, your screen should now look like the screen below. Next, under Inter-Item, click on Correlations

¹³<http://cnx.org/content/m40717/latest/5.7.png/image>



After you check to verify that your screen looks the one below, then you may click on **Continue**.

¹⁴<http://cnx.org/content/m40717/latest/5.8.png/image>

The screenshot shows the SPSS Reliability Analysis: Statistics dialog box. The 'Descriptives for' section has 'Item', 'Scale', and 'Scale if item deleted' checked. The 'Inter-Item' section has 'Correlations' checked and 'Covariances' unchecked. The 'Summaries' section has 'Means' and 'Variances' checked, while 'Covariances' and 'Correlations' are unchecked. The 'ANOVA Table' section has 'None' selected. The 'Model' is set to 'Two-Way Mixed' and the 'Type' is 'Consistency'. The 'Confidence Interval' is 95% and the 'Test Value' is 0. The 'Continue' button is highlighted with a red arrow.

pc	inf	oa	comp	var	var	var	var
10	7	6	7				
8	8	4	7				
10	8	11	8				
3	4	10	6				
5	7	9	1				
2	5	5	5				
7	4	12	5				
7	7	11	3				
9	6	8	5				
5	9	8	11				
6	7	7	5				
13	8	9	7				
7	8	3	4				
6	6	8	8				
7	8	3	9				
10	7	5	8				
12	8	7	8				
15	7	10	15				
8	5	8	7				
9	6	3	5				
6	7	6	8				
13	7	9	6				
8	7	6	3				
10	7	8	8				
12	8	8	12				
10	9	11	11				
10	10	11	14				
6	6	6	6				
14	7	6	8				
6	9	6	7				
10	10	6	6				
7	9	6	8				

Next click on OK.

¹⁵<http://cnx.org/content/m40717/latest/5.9.png/image>

The screenshot shows the SPSS Reliability Analysis dialog box. The 'Items' list includes:

- Performance 1 (Picture ...)
- Performance 2 (Coding) ...
- Performance 3 (Picture ...)
- Performance 4 (Block D...
- Performance 5 (Object A...
- Verbal 1 (Information) [inf]
- Verbal 2 (Similarities) [sim]
- Verbal 3 (Arithmetic) [ari]
- Verbal 4 (Vocabulary) [v...]
- Verbal 5 (Comprehensio...

The 'Model' is set to 'Alpha'. The 'Scale label' field is empty. The 'OK' button is highlighted with a red arrow.

The background data editor window shows a grid of data points. The first two columns are labeled 'pc' and 'inf'. The first 10 rows of data are:

	pc	inf
1	10	7
2	8	8
3	10	8
4	3	4
5	5	7
6	2	5
7	7	4
8	7	7
9	9	6
10	5	9

SPSS should send you to the Output screen. If not, click on the Output icon at the bottom of your screen.

¹⁶<http://cnx.org/content/m40717/latest/5.10.png/image>

factor analysis dataset for book.sav [DataSet2] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : pc 10.0 Visible: 10 of 10 Variables

	pc	inf	cod	sim	pa	ari	bd	voc	oa	comp	var	var	var	var
1	10	7	8	7	10	5	7	7	6	7				
2	8	8	7	6	7	9	9	4	4	7				
3	10	8	2	9	10	11	9	6	11	8				
4	3	4	11	4	7	6	3	4	10	6				
5	5	7	17	6	8	10	7	5	9	1				
6	2	5	8	8	9	7	10	3	5	5				
7	7	4	5	8	7	11	11	2	12	5				
8	7	7	7	7	14	8	10	8	11	3				
9	9	6	10	6	9	8	4	5	8	5				
10	5	9	7	10	9	12	11	7	8	11				
11	6	7	5	8	7	7	7	4	7	5				
12	13	8	5	13	10	5	6	9	9	7				
13	7	8	11	6	8	4	3	3	4	4				
14	6	6	8	12	9	9	9	10	6	8				
15	7	8	3	1	9	8	5	10	3	9				
16	10	7	5	5	9	5	5	8	3	6				
17	12	8	7	8	13	10	11	7	6	5				
18	15	7	5	10	15	8	7	8	7	10				
19	9	6	4	8	3	5	8	6	10	3				
20	6	7	10	6	9	6	8	6	5	2				
21	13	7	8	7	9	6	3	6	6	4				
22	10	7	12	8	8	8	12	5	11	4				
23	10	9	10	10	11	11	14	11	11	10				
24	6	6	6	2	5	7	6	6	11	6				
25	14	7	6	9	10	6	7	6	8	5				

Data View Variable View

SPSS Statistics Processor is ready

start factor analysis d... *Output2 [Doc... steps screencho... steps and scree...

17

¹⁷ <http://cnx.org/content/m40717/latest/5.11.png/image>

Chapter 7

7. Internal Consistency Analysis: Part II¹



NOTE: This chapter is published by NCPEA Press² and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of education administration.

John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website (Writing and Statistical Help³) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

Ana Rojas-LeBouef is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. Dr. LeBouef recently completed her doctoral degree in Reading, where she conducted a 16-year analysis of Texas statewide data regarding the achievement gap. Her research interests lie in examining the inequities in achievement among ethnic groups. Dr. Rojas-LeBouef also assists students and faculty in their writing and statistical needs on the Writing and Statistical Help website.

Now that you are in the SPSS output screen, your screen should resemble the following one.

The first table, labeled Case Processing Summary, indicates how many cases were in the analysis. In this example, 1,136 cases out of the total sample of 1,182 were used in the analysis.

¹This content is available online at <<http://cnx.org/content/m40718/1.2/>>.

²<http://www.ncpeapublications.org/books.html>

³<http://cnx.org/content/m40718/latest/www.writingandstatisticalhelp>

Case Processing Summary

Cases	Valid	N	%
	Valid	1136	96.1
	Excluded ^a	46	3.9
	Total	1182	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.872	.873	5

Item Statistics

	Mean	Std. Deviation	N
Verbal 1 (Information)	6.35	2.811	1136
Verbal 2 (Similarities)	6.21	2.986	1136
Verbal 3 (Arithmetic)	6.56	2.731	1136
Verbal 4 (Vocabulary)	5.63	2.891	1136
Verbal 5 (Comprehension)	5.82	3.219	1136

Inter-Item Correlation Matrix

	Verbal 1 (Information)	Verbal 2 (Similarities)	Verbal 3 (Arithmetic)	Verbal 4 (Vocabulary)	Verbal 5 (Comprehension)
Verbal 1 (Information)	1.000	.624	.534	.663	.580
Verbal 2 (Similarities)	.624	1.000	.504	.663	.573
Verbal 3 (Arithmetic)	.534	.504	1.000	.494	.484
Verbal 4 (Vocabulary)	.663	.663	.494	1.000	.666
Verbal 5 (Comprehension)	.580	.573	.484	.666	1.000

Summary Item Statistics

The next table titled Reliability Statistics indicates the score reliability for these five Verbal subscales. The Cronbach's Coefficient Alpha is .872 for these five Verbal subscales.

⁴<http://cnx.org/content/m40718/latest/6.1.png/image>

The screenshot displays the SPSS Statistics Viewer interface with the following tables:

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.872	.873	5

Item Statistics

	Mean	Std. Deviation	N
Verbal 1 (Information)	6.35	2.811	1136
Verbal 2 (Similarities)	6.21	2.986	1136
Verbal 3 (Arithmetic)	6.56	2.731	1136
Verbal 4 (Vocabulary)	5.63	2.891	1136
Verbal 5 (Comprehension)	5.82	3.219	1136

Inter-Item Correlation Matrix

	Verbal 1 (Information)	Verbal 2 (Similarities)	Verbal 3 (Arithmetic)	Verbal 4 (Vocabulary)	Verbal 5 (Comprehension)
Verbal 1 (Information)	1.000	.624	.534	.663	.580
Verbal 2 (Similarities)	.624	1.000	.504	.663	.573
Verbal 3 (Arithmetic)	.534	.504	1.000	.494	.484
Verbal 4 (Vocabulary)	.663	.663	.494	1.000	.666
Verbal 5 (Comprehension)	.580	.573	.484	.666	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	6.114	5.634	6.556	.924	1.164	.145	5
Item Variances	8.599	7.457	10.360	2.903	1.389	1.262	5

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Verbal 1 (Information)	24.22	94.494	.731	.541	.838

The table, labeled as Item Statistics, below the Reliability Statistics table contains the descriptive statistics for these five Verbal subscales for all participants in the sample. This information may be helpful in your Results section.

⁵ <http://cnx.org/content/m40718/latest/6.2.png/image>

The screenshot displays the SPSS Statistics Viewer interface. The main window shows several statistical tables. A red arrow points to the 'Item Statistics' table. The left sidebar shows a tree view of the output, with 'Item Statistics' selected. The bottom taskbar shows the Windows taskbar with several open applications and the system clock at 12:35 PM.

Item Statistics

	Mean	Std. Deviation	N
Verbal 1 (Information)	6.35	2.811	1136
Verbal 2 (Similarities)	6.21	2.986	1136
Verbal 3 (Arithmetic)	6.56	2.731	1136
Verbal 4 (Vocabulary)	5.63	2.891	1136
Verbal 5 (Comprehension)	5.82	3.219	1136

Inter-Item Correlation Matrix

	Verbal 1 (Information)	Verbal 2 (Similarities)	Verbal 3 (Arithmetic)	Verbal 4 (Vocabulary)	Verbal 5 (Comprehension)
Verbal 1 (Information)	1.000	.624	.534	.663	.580
Verbal 2 (Similarities)	.624	1.000	.504	.663	.573
Verbal 3 (Arithmetic)	.534	.504	1.000	.494	.484
Verbal 4 (Vocabulary)	.663	.663	.494	1.000	.666
Verbal 5 (Comprehension)	.580	.573	.484	.666	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	6.114	5.634	6.558	.924	1.164	.145	5
Item Variances	8.599	7.457	10.360	2.903	1.389	1.262	5

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Verbal 1 (Information)	24.22	94.494	.731	.541	.838
Verbal 2 (Similarities)	24.36	92.250	.718	.527	.841
Verbal 3 (Arithmetic)	24.01	102.171	.593	.358	.870
Verbal 4 (Vocabulary)	24.94	91.554	.767	.608	.829
Verbal 5 (Comprehension)	24.75	89.544	.697	.504	.847

Scale Statistics

The next table of importance is the one labeled Item-Total Statistics. The far right column of this table, Cronbach's Alpha if Item Deleted, indicates how much of a change in internal consistency will occur if you delete that variable from the group. As our Cronbach's Coefficient Alpha for all 5 Verbal subscales was .872, as previously noted, no increase occurs in the internal consistency if any of the 5 Verbal subscales are deleted.

⁶ <http://cnx.org/content/m40718/latest/6.3.png/image>

The screenshot displays the SPSS Statistics Viewer interface. The main window shows the following tables and sections:

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Verbal 1 (Information)	24.22	94.494	.731	.541	.838
Verbal 2 (Similarities)	24.36	92.250	.718	.527	.841
Verbal 3 (Arithmetic)	24.01	102.171	.593	.358	.870
Verbal 4 (Vocabulary)	24.94	91.554	.767	.608	.829
Verbal 5 (Comprehension)	24.75	89.544	.697	.504	.847

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
30.57	142.339	11.931	5

RELIABILITY

```

/VARIABLES=inf sim ari voc comp
/SCALE ('ALL VARIABLES') ALL
/MODEL=ALPHA
/STATISTICS=DESCRIPTIVE SCALE CORR
/SUMMARY=TOTAL MEANS VARIANCE.

```

Reliability

[DataSet2] C:\WINNT\profiles\shsu\My Documents\John R Slate\Scholarly Works\Books\Book Two Advanced Statistics\factor analysis dataset for book.sav

Scale: ALL VARIABLES

Case Processing Summary

	N	%
Cases Valid	1136	96.1
Excluded ^a	46	3.9
Total	1182	100.0

a. Listwise deletion based on all

The next column in this table that we examine is the one labeled Corrected Item-Total Correlation. These values indicate the extent to which each variable is related with the group of variables, in this case with the total of the Verbal subscales. The higher the correlation, the more that item/variable is associated with the total of the group. The values in this column reflect that these five Verbal subscales are consistently contributing to the measurement of the same/similar construct.

⁷ <http://cnx.org/content/m40718/latest/6.4.png/image>

The screenshot displays the SPSS Statistics Viewer interface. The main window shows the following output:

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Verbal 1 (Information)	24.22	94.494	.731	.541	.838
Verbal 2 (Similarities)	24.36	92.250	.718	.527	.841
Verbal 3 (Arithmetic)	24.01	102.171	.593	.358	.870
Verbal 4 (Vocabulary)	24.94	91.554	.767	.608	.829
Verbal 5 (Comprehension)	24.75	89.544	.697	.504	.847

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
30.57	142.339	11.931	5

RELIABILITY

```

/VARIABLES=inf sim ar1 voc comp
/SCALE ('ALL VARIABLES') ALL
/MODEL=ALPHA
/STATISTICS=DESCRIPTIVE SCALE CORR
/SUMMARY=TOTAL MEANS VARIANCE.

```

Reliability

[DataSet2] C:\MINWT\profiles\shsu\My Documents\John R Slate\Scholarly Works\Books\Book Two Advanced Statistics\factor analysis dataset for book.sav

Scale: ALL VARIABLES

Case Processing Summary

Cases	Valid	N	%
	Excluded ^a	46	3.9
	Total	1182	100.0

a. Listwise deletion based on all

The taskbar at the bottom shows the Windows taskbar with the Start button, several open applications (including SPSS), and the system clock showing 12:40 PM on 8/24/2008.

In the case of the five Verbal subscales, all five subscales may be regarded as providing internally consistent scores. Moreover, their scores can be aggregated or added together to create a composite score. See our chapter on data transformation – creating a new variable.

Now that we have determined the internal consistency of the Verbal subscales, we will ascertain the extent to which the Performance subscales are internally consistent.

From the data window, click on Analyze

⁸ <http://cnx.org/content/m40718/latest/6.5.png/image>

factor analysis dataset for book.sav [*.sav] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1: pc 10.0 Visible: 10 of 10 Variables

	pc	inf	cod	sim	pa	ari	bd	voc	oa	comp	var	var	var	var
1	10	7	8	7	10	5	7	7	6	7				
2	8	8	7	6	7	9	9	4	4	7				
3	10	8	2	9	10	11	9	6	11	8				
4	3	4	11	4	7	6	3	4	10	6				
5	5	7	17	6	8	10	7	5	9	1				
6	2	5	8	8	9	7	10	3	5	5				
7	7	4	5	8	7	11	11	2	12	5				
8	7	7	7	7	14	8	10	8	11	3				
9	9	6	10	6	9	8	4	5	8	5				
10	5	9	7	10	9	12	11	7	8	11				
11	6	7	5	8	7	7	7	4	7	5				
12	13	8	5	13	10	5	6	9	9	7				
13	7	8	11	6	8	4	3	3	4	4				
14	6	6	8	12	9	9	9	10	6	8				
15	7	8	3	1	9	8	5	10	3	9				
16	10	7	5	5	9	5	5	8	3	6				
17	12	8	7	8	13	10	11	7	6	5				
18	15	7	5	10	15	8	7	8	7	10				
19	9	6	4	8	3	5	8	6	10	3				
20	6	7	10	6	9	6	8	6	5	2				
21	13	7	8	7	9	6	3	6	6	4				
22	10	7	12	8	8	8	12	5	11	4				
23	10	9	10	10	11	11	14	11	11	10				
24	6	6	6	2	5	7	6	6	11	6				
25	14	7	6	9	10	6	7	6	8	5				

Data View Variable View

SPSS Statistics Processor is ready

start factor analysis d... *Output2 [Doc... steps and scree... steps screenho...

12:44 PM 9

After clicking on Analyze,
Click on Scale.
Click on Reliability Analysis.

⁹<http://cnx.org/content/m40718/latest/6.6.png/image>

The screenshot shows the SPSS Statistics Data Editor window with the following data table:

	pc	inf	pa	ari	bd	voc	oa	comp	var	var	var	var
1	10		10	5	7	7	6	7				
2	8		7	9	9	4	4	7				
3	10		10	11	9	6	11	8				
4	3		7	6	3	4	10	6				
5	5		8	10	7	5	9	1				
6	2		9	7	10	3	5	5				
7	7		7	11	11	2	12	5				
8	7						8	11	3			
9	9						5	8	5			
10	5		9	12	11	7	8	11				
11	6		7	7	7	4	7	5				
12	13		10	5	6	9	9	7				
13	7		8	4	3	3	4	4				
14	6		9	9	9	10	6	8				
15	7	8	3	1	9	8	5	10	3	9		
16	10	7	5	5	9	5	5	8	3	6		
17	12	8	7	8	13	10	11	7	6	5		
18	15	7	5	10	15	8	7	8	7	10		
19	9	6	4	8	3	5	8	6	10	3		
20	6	7	10	6	9	6	8	6	5	2		
21	13	7	8	7	9	6	3	6	6	4		
22	10	7	12	8	8	8	12	5	11	4		
23	10	9	10	10	11	11	14	11	11	10		
24	6	6	6	2	5	7	6	6	11	6		
25	14	7	6	9	10	6	7	6	8	5		

Because this reliability analysis is the second one you are conducting, the screen shows the five Verbal subtests included in the first analysis. Remove these from the Items box.

¹⁰<http://cnx.org/content/m40718/latest/6.7.png/image>

The screenshot shows the SPSS Reliability Analysis dialog box. The 'Items' list contains the following subtests:

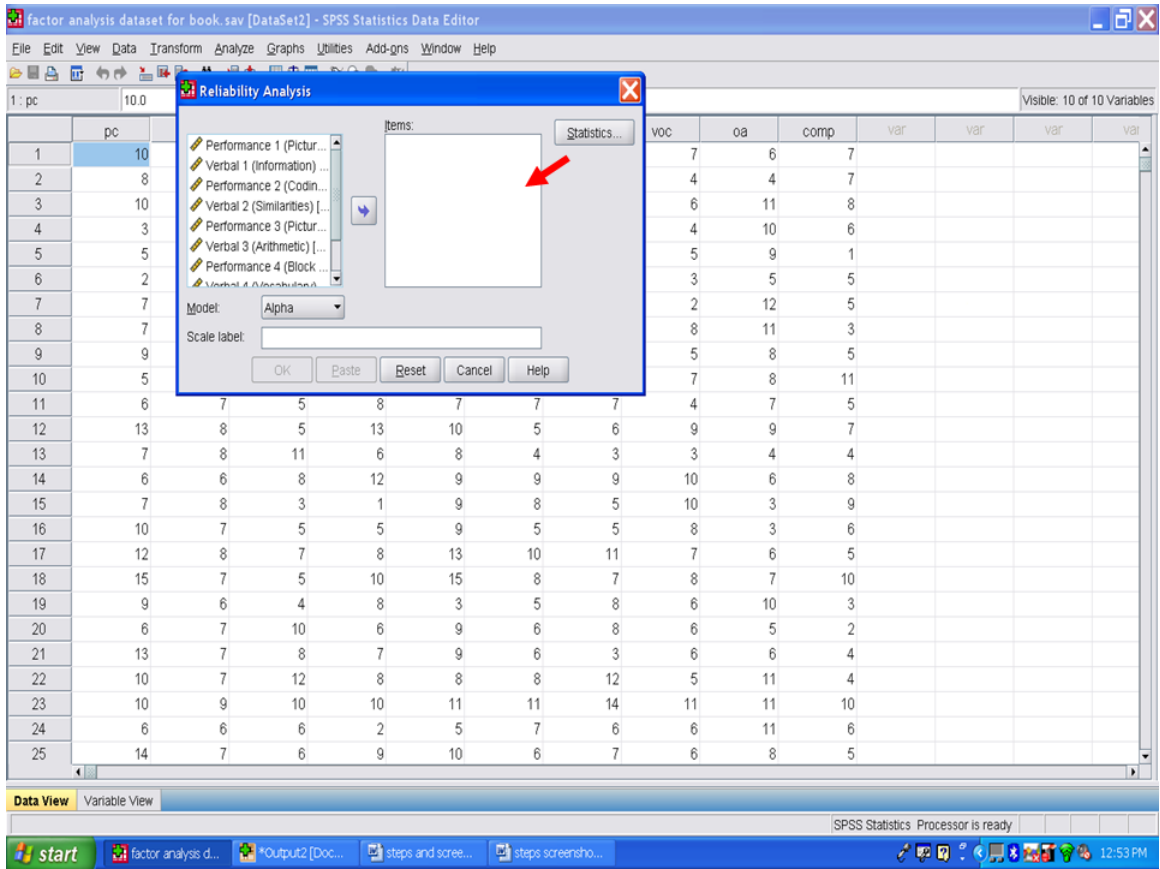
- Performance 1 (Picture...)
- Performance 2 (Coding ...)
- Performance 3 (Picture ...)
- Performance 4 (Block D...)
- Performance 5 (Object A...)
- Verbal 1 (Information) [inf]
- Verbal 2 (Similarities) [sim]
- Verbal 3 (Arithmetic) [ari]
- Verbal 4 (Vocabulary) [v...]
- Verbal 5 (Comprehensio...)

The 'Model' is set to 'Alpha'. The 'Scale label' is empty. The 'Statistics...' button is highlighted with a red arrow. The background data table shows the following variables and their values:

pc	voc	oa	comp	var	var	var	var
10	7	6	7				
8	4	4	7				
10	6	11	8				
3	4	10	6				
5	5	9	1				
2	3	5	5				
7	2	12	5				
7	8	11	3				
9	5	8	5				
5	7	8	11				
6	7	5	7	7	7	7	7
13	8	5	13	10	5	6	6
7	8	11	6	8	4	3	3
6	6	8	12	9	9	9	10
7	8	3	1	9	8	5	10
10	7	5	5	9	5	8	3
12	8	7	8	13	10	11	7
15	7	5	10	15	8	7	8
9	6	4	8	3	5	8	6
6	7	10	6	9	6	8	6
13	7	8	7	9	6	3	6
10	7	12	8	8	8	12	5
10	9	10	10	11	11	14	11
6	6	6	2	5	7	6	6
14	7	6	9	10	6	7	6

Your screen will now look like the following one. Now we will determine the internal consistency of the Performance subscales. From the factor analysis chapter, Performance 2 was determined to not be part of this cluster or group of subscales. Therefore, we will place only Performance 1 (Picture Completion), Performance 3 (Picture Arrangement), Performance 4 (Block Design), and Performance 5 (Object Assembly) in the Items box.

¹¹<http://cnx.org/content/m40718/latest/6.8.png/image>



¹²<http://cnx.org/content/m40718/latest/6.9.png/image>

Chapter 8

8. Internal Consistency Analysis: Part III¹



NOTE: This chapter is published by NCPEA Press² and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of education administration.

About the Authors

John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website (Writing and Statistical Help³) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

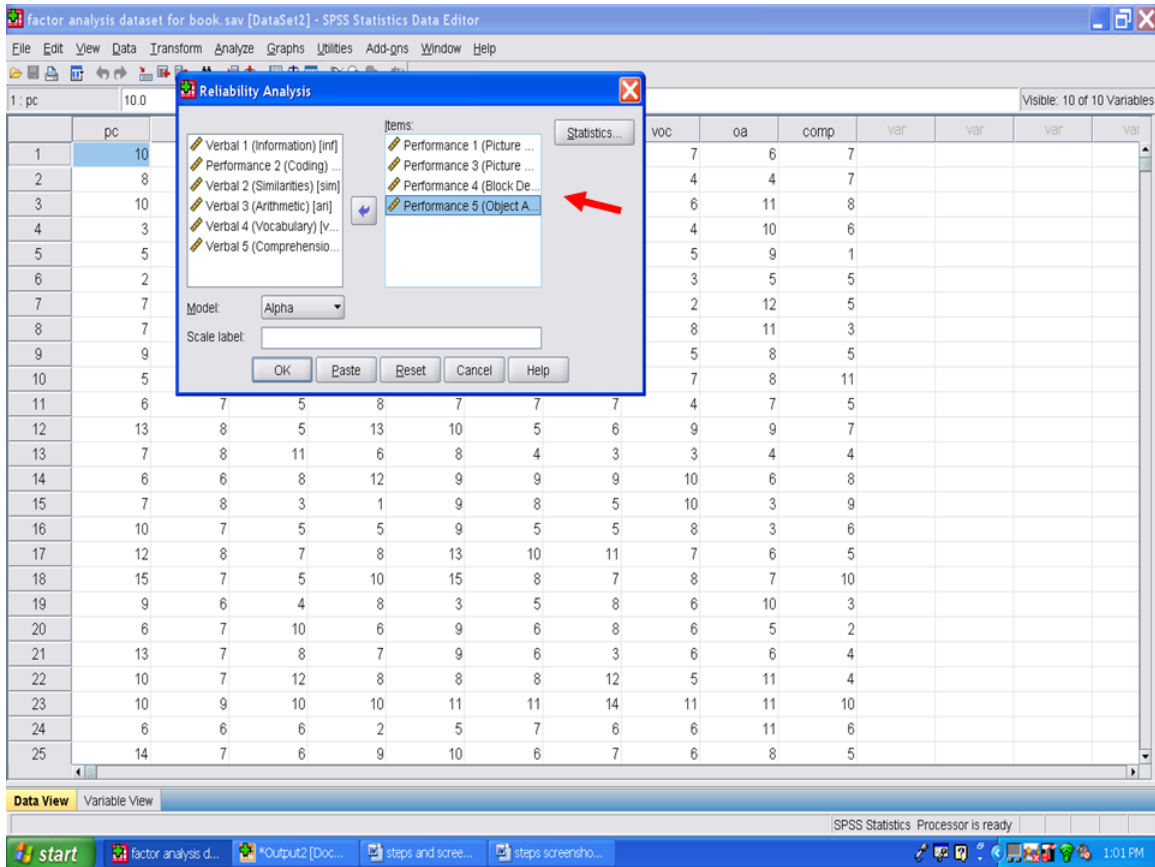
Ana Rojas-LeBouef is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. Dr. LeBoeuf recently completed her doctoral degree in Reading, where she conducted a 16-year analysis of Texas statewide data regarding the achievement gap. Her research interests lie in examining the inequities in achievement among ethnic groups. Dr. Rojas-LeBouef also assists students and faculty in their writing and statistical needs on the Writing and Statistical Help website.

Once the four Performance subscales have been sent to the Items box, your screen would look like the following one:

¹This content is available online at <<http://cnx.org/content/m40721/1.2/>>.

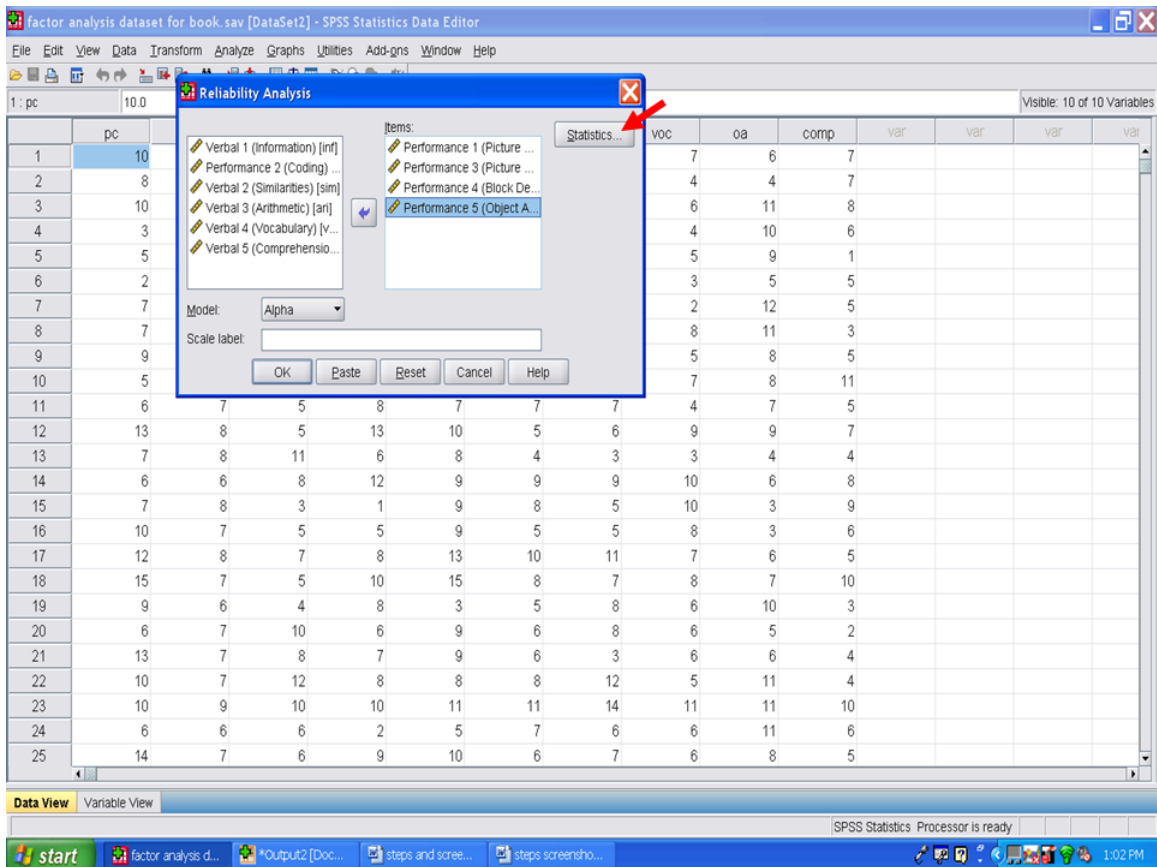
²<http://www.ncpeapublications.org/books.html>

³<http://cnx.org/content/m40721/latest/www.writingandstatisticalhelp>



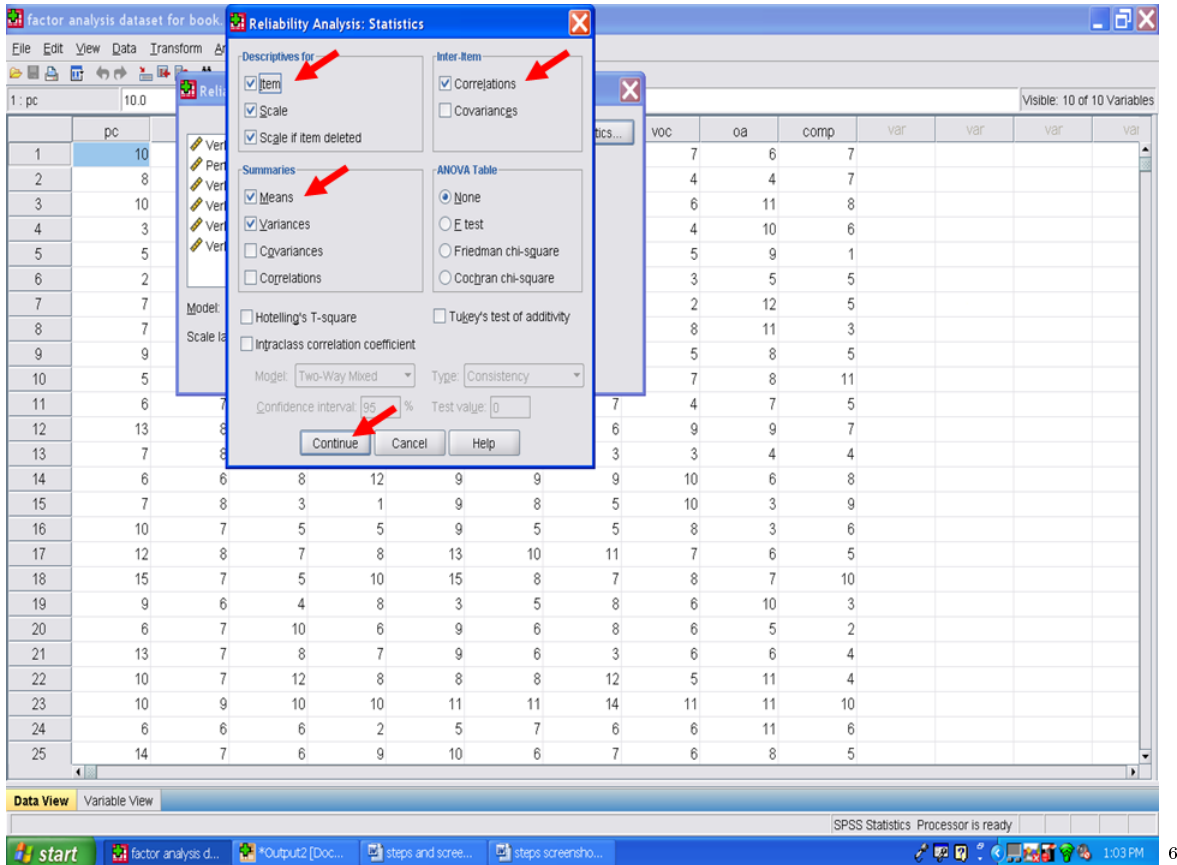
Next, click on Statistics.

⁴<http://cnx.org/content/m40721/latest/7.1.png/image>



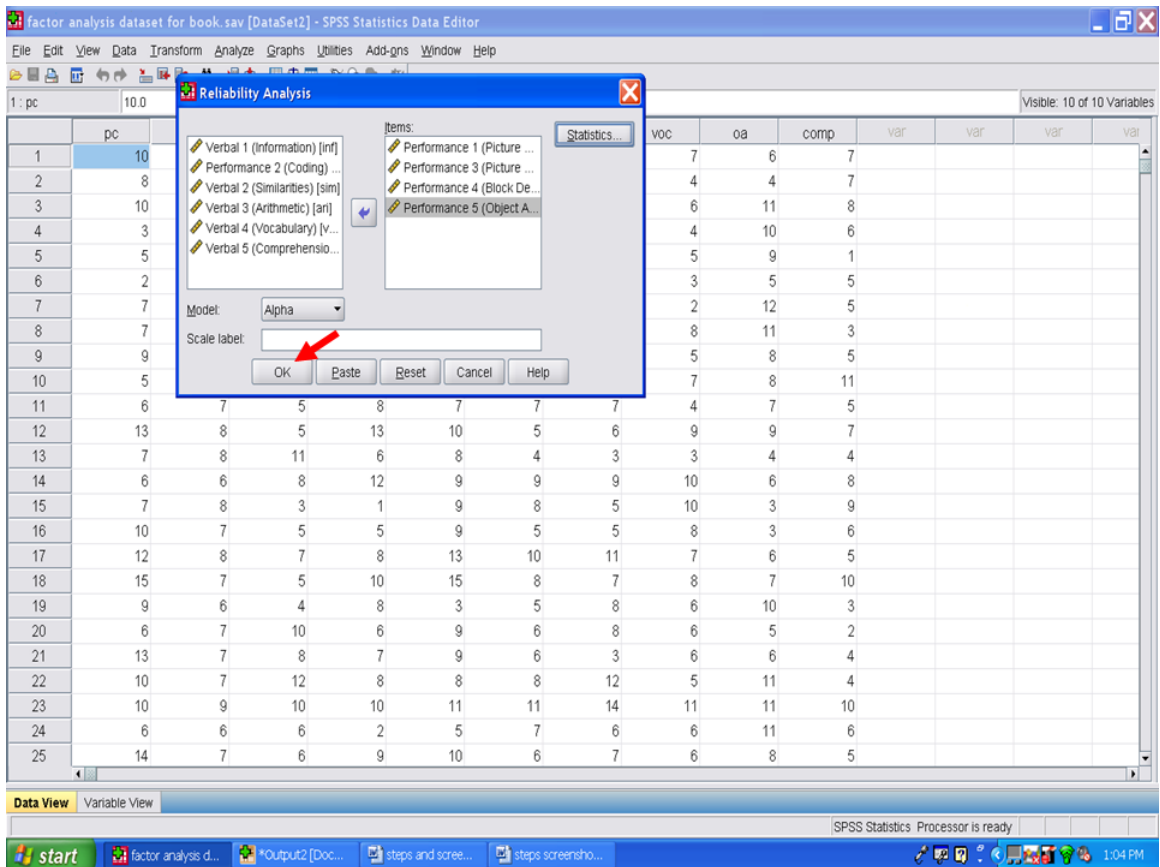
You will note that all of the items you had clicked on when conducting the previous internal consistency analysis remain active.
Click on Continue.

⁵ <http://cnx.org/content/m40721/latest/7.2.png/image>



Click on OK and this internal consistency analysis will be conducted.

⁶ <http://cnx.org/content/m40721/latest/7.3.png/image>



SPSS should place you in the Output window. If not, click on the Output screen icon at the bottom of your screen.

⁷ <http://cnx.org/content/m40721/latest/7.4.png/image>

	pc	inf	cod	sim	pa	ari	bd	voc	oa	comp	var	var	var	var
1	10	7	8	7	10	5	7	7	6	7				
2	8	8	7	6	7	9	9	4	4	7				
3	10	8	2	9	10	11	9	6	11	8				
4	3	4	11	4	7	6	3	4	10	6				
5	5	7	17	6	8	10	7	5	9	1				
6	2	5	8	8	9	7	10	3	5	5				
7	7	4	5	8	7	11	11	2	12	5				
8	7	7	7	7	14	8	10	8	11	3				
9	9	6	10	6	9	8	4	5	8	5				
10	5	9	7	10	9	12	11	7	8	11				
11	6	7	5	8	7	7	7	4	7	5				
12	13	8	5	13	10	5	6	9	9	7				
13	7	8	11	6	8	4	3	3	4	4				
14	6	6	8	12	9	9	9	10	6	8				
15	7	8	3	1	9	8	5	10	3	9				
16	10	7	5	5	9	5	5	8	3	6				
17	12	8	7	8	13	10	11	7	6	5				
18	15	7	5	10	15	8	7	8	7	10				
19	9	6	4	8	3	5	8	6	10	3				
20	6	7	10	6	9	6	8	6	5	2				
21	13	7	8	7	9	6	3	6	6	4				
22	10	7	12	8	8	8	12	5	11	4				
23	10	9	10	10	11	11	14	11	11	10				
24	6	6	6	2	5	7	6	6	11	6				
25	14	7	6	9	10	6	7	6	8	5				

In the output screen, the first table you will see is the Case Processing Summary. In this example, 1,139 cases were used out of the total sample of 1,182 cases.

⁸<http://cnx.org/content/m40721/latest/7.5.png/image>

SPSS Statistics Processor is ready

Scale: ALL VARIABLES

Case Processing Summary

Cases	Valid	N	%
	Valid	1139	96.4
	Excluded ^a	43	3.6
	Total	1182	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.800	.800	4

Item Statistics

	Mean	Std. Deviation	N
Performance 1 (Picture Completion)	6.97	3.109	1139
Performance 3 (Picture Arrangement)	7.27	3.034	1139
Performance 4 (Block Design)	6.51	3.352	1139
Performance 5 (Object Assembly)	6.64	3.247	1139

Inter-Item Correlation Matrix

	Performance 1 (Picture Completion)	Performance 3 (Picture Arrangement)	Performance 4 (Block Design)	Performance 5 (Object Assembly)
Performance 1 (Picture Completion)	1.000	.470	.496	.510
Performance 3 (Picture Arrangement)	.470	1.000	.468	.420
Performance 4 (Block Design)	.496	.468	1.000	.630
Performance 5 (Object Assembly)	.510	.420	.630	1.000

start factor analysis d... *Output2 [Doc... steps and scree... steps screenho... 1:06 PM 9

The overall internal consistency of the scores of these four Performance subscales is located in the Reliability Statistics table. In this example, Cronbach's Coefficient Alpha is .800.

⁹<http://cnx.org/content/m40721/latest/7.6.png/image>

The screenshot displays the SPSS Statistics Viewer interface. The main window shows the following tables:

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.800	.800	4

Item Statistics

	Mean	Std. Deviation	N
Performance 1 (Picture Completion)	6.97	3.109	1139
Performance 3 (Picture Arrangement)	7.27	3.034	1139
Performance 4 (Block Design)	6.51	3.352	1139
Performance 5 (Object Assembly)	6.64	3.247	1139

Inter-Item Correlation Matrix

	Performance 1 (Picture Completion)	Performance 3 (Picture Arrangement)	Performance 4 (Block Design)	Performance 5 (Object Assembly)
Performance 1 (Picture Completion)	1.000	.470	.496	.510
Performance 3 (Picture Arrangement)	.470	1.000	.468	.420
Performance 4 (Block Design)	.496	.468	1.000	.630
Performance 5 (Object Assembly)	.510	.420	.630	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	6.849	6.507	7.274	.766	1.118	.119	4
Item Variances	10.163	9.206	11.238	2.032	1.221	.821	4

Item-Total Statistics

	Scale	Corrected	Squared	Cronbach's

The next table labeled Item Statistics contains the descriptive statistics for these four Performance subscales for the sample of 1,139 persons.

¹⁰<http://cnx.org/content/m40721/latest/7.7.png/image>

The screenshot displays the SPSS Statistics Viewer interface with the following tables:

	Cronbach's Alpha	Standardized Items	N of Items
	.800	.800	4

Item Statistics

	Mean	Std. Deviation	N
Performance 1 (Picture Completion)	6.97	3.109	1139
Performance 3 (Picture Arrangement)	7.27	3.034	1139
Performance 4 (Block Design)	6.51	3.352	1139
Performance 5 (Object Assembly)	6.64	3.247	1139

Inter-Item Correlation Matrix

	Performance 1 (Picture Completion)	Performance 3 (Picture Arrangement)	Performance 4 (Block Design)	Performance 5 (Object Assembly)
Performance 1 (Picture Completion)	1.000	.470	.496	.510
Performance 3 (Picture Arrangement)	.470	1.000	.468	.420
Performance 4 (Block Design)	.496	.468	1.000	.630
Performance 5 (Object Assembly)	.510	.420	.630	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	6.849	6.507	7.274	.766	1.118	.119	4
Item Variances	10.163	9.206	11.238	2.032	1.221	.821	4

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Performance 1 (Picture Completion)	20.42	62.510	.601	.362	.756
Performance 3 (Picture Arrangement)	20.12	65.813	.542	.302	.783

In the next table of importance, labeled Item-Total Statistics, locate the far right column of this table. This column, titled as Cronbach's Alpha if Item Deleted, indicates how much of a change in internal consistency would occur if you deleted that variable from the group. As our Cronbach's Coefficient Alpha for all four Performance subscales was .800, as previously noted, no increase occurs in the internal consistency if any of these subscales are deleted.

¹¹ <http://cnx.org/content/m40721/latest/7.8.png/image>

The screenshot displays the SPSS Statistics Viewer interface. The main window shows the output of a Reliability analysis. A red arrow points to the 'Corrected Item-Total Correlation' column in the 'Item-Total Statistics' table. The 'Scale Statistics' table is also visible, along with the 'Case Processing Summary' table.

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Performance 1 (Picture Completion)	20.42	62.510	.601	.362	.756
Performance 3 (Picture Arrangement)	20.12	65.813	.542	.302	.783
Performance 4 (Block Design)	20.89	56.868	.664	.466	.724
Performance 5 (Object Assembly)	20.76	58.860	.648	.454	.733

Mean	Variance	Std. Deviation	N of Items
27.40	101.698	10.085	4

Cases	Valid	N	%
	Excluded ^a	43	3.6
	Total	1182	100.0

Reliability

```

/VARIABLES=pc pa bd oa
/SCALE ('ALL VARIABLES') ALL
/MODEL=ALPHA
/STATISTICS=DESCRIPTIVE SCALE CORR
/SUMMARY=TOTAL MEANS VARIANCE.

```

[DataSet2] C:\MINWT\profiles\shsu\My Documents\John R Slate\Scholarly Works\Books\Book Two Advanced Statistics\factor analysis dataset for book.sav

Scale: ALL VARIABLES

a. Listwise deletion based on all

The next column in this table that we examine is the one labeled Corrected Item-Total Correlation. These values indicate the extent to which each variable is related with the group of variables, in this case with the total of the Performance subscales. The higher the correlation, the more that item/variable is associated with the total of the group. The values in this column reflect that these four Performance subscales are consistently contributing to the measurement of the same/similar construct.

¹²<http://cnx.org/content/m40721/latest/7.9.png/image>

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Performance 1 (Picture Completion)	20.42	62.510	.601	.362	.756
Performance 3 (Picture Arrangement)	20.12	65.813	.542	.302	.783
Performance 4 (Block Design)	20.89	56.868	.664	.466	.724
Performance 5 (Object Assembly)	20.76	58.860	.648	.454	.733

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
27.40	101.698	10.085	4

RELIABILITY

```

/VARIABLES=pc pa bd oa
/SCALE ('ALL VARIABLES') ALL
/MODEL=ALPHA
/STATISTICS=DESCRIPTIVE SCALE CORR
/SUMMARY=TOTAL MEANS VARIANCE.

```

Reliability

[DataSet2] C:\MINWT\profiles\shsu\My Documents\John R Slate\Scholarly Works\Books\Book Two Advanced Statistics\factor analysis dataset for book.sav

Scale: ALL VARIABLES

Case Processing Summary

Cases	Valid	N	%
	Excluded ^a	43	3.6
	Total	1182	100.0

a. Listwise deletion based on all

In the case of the four Performance subscales, all four subscales may be regarded as providing internally consistent scores. Moreover, their scores can be aggregated or added together to create a composite score. See our chapter on data transformation – creating a new variable.

Now that we have determined the internal consistency of the scores from the Verbal subscales and of the scores from the Performance subscales, a name needs to be assigned to each of these two factors. Naming the factor involves interpreting the pattern/structure coefficients from the factor analysis that was described in the factor analysis chapter. All significant pattern/structure coefficients are typically used in the interpretation process. Variables with higher coefficients influence to a greater extent the name or label selected to represent a factor. Naming of factors is NOT very scientific and is based on the subjective opinion of the analyst—naming is open to criticism.

¹³<http://cnx.org/content/m40721/latest/7.10.png/image>

Chapter 9

9. Factor Analysis: Part I¹



NOTE: This chapter is published by NCPEA Press² and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of education administration.

About the Authors

John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website (Writing and Statistical Help³) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

Ana Rojas-LeBouef is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. Dr. LeBoeuf recently completed her doctoral degree in Reading, where she conducted a 16-year analysis of Texas statewide data regarding the achievement gap. Her research interests lie in examining the inequities in achievement among ethnic groups. Dr. Rojas-LeBouef also assists students and faculty in their writing and statistical needs on the Writing and Statistical Help website.

In this set of steps, readers will conduct a factor analysis procedure, in particular, a Varimax factor analysis. For detailed information regarding the assumptions underlying use of a factor analysis and reasons for using a Varimax factor analysis, readers are referred to the *Hyperstats Online Statistics Textbook* at <http://davidmlane.com/hyperstat/>⁴; to the *Electronic Statistics Textbook* (2011) at

¹This content is available online at <<http://cnx.org/content/m40725/1.2/>>.

²<http://www.ncpeapublications.org/books.html>

³<http://cnx.org/content/m40725/latest/www.writingandstatisticalhelp>

⁴<http://davidmlane.com/hyperstat/>

<http://www.statsoft.com/textbook/>⁵ ; or to Andy Field's (2009) *Discovering Statistics Using SPSS* at http://www.amazon.com/Discovering-Statistics-Introducing-Statistical-Method/dp/1847879071/ref=sr_1_1?s=books&ie=UTF8&qid=1304967862&sr=1-1⁶

Research questions for which a factor analysis procedure is appropriate involve analyzing a survey or an instrument in which multiple scores are present. For example, if a 100-item survey is administered, it is not feasible nor is it desirable to conduct 100 separate *t*-tests or to calculate all of the possible intercorrelations. Rather, the 100 items on the survey should be collapsed or aggregated in some manner. Use of a factor analysis permits the clustering of items into coherent and logical clusters which can then be analyzed through use of multivariate statistical procedures. As such, factor analysis procedures do not directly answer inferential research questions. Instead they serve as precursors to other statistical methods.

For this particular chapter, we will conduct a factor analysis of the 10 subscales of the Wechsler Intelligence Scale for Children-Third Edition. The 10 subscales are shown below: Picture Completion (pc), Information (inf), Coding (cod), Similarities (sim), Picture Arrangement (pa), Arithmetic (ari), Block Design (bd), Vocabulary (voc), Object Assembly (oa), and Comprehension (comp). Prior to conducting a factor analysis, you should have verified the accuracy of your data.

	pc	inf	cod	sim	pa	ari	bd	voc	oa	comp	var	var	var	var
1	10	7	8	7	10	5	7	7	6	7				
2	8	8	7	6	7	9	9	4	4	7				
3	10	8	2	9	10	11	9	6	11	8				
4	3	4	11	4	7	6	3	4	10	6				
5	5	7	17	6	8	10	7	5	9	1				
6	2	5	8	8	9	7	10	3	5	5				
7	7	4	5	8	7	11	11	2	12	5				
8	7	7	7	7	14	8	10	8	11	3				
9	9	6	10	6	9	8	4	5	8	5				
10	5	9	7	10	9	12	11	7	8	11				
11	6	7	5	8	7	7	7	4	7	5				
12	13	8	5	13	10	5	6	9	9	7				
13	7	8	11	6	8	4	3	3	4	4				
14	6	6	8	12	9	9	9	10	6	8				
15	7	8	3	1	9	8	5	10	3	9				
16	10	7	5	5	9	5	5	8	3	6				
17	12	8	7	8	13	10	11	7	6	5				
18	15	7	5	10	15	8	7	8	7	10				
19	9	6	4	8	3	5	8	6	10	3				
20	6	7	10	6	9	6	8	6	5	2				
21	13	7	8	7	9	6	3	6	6	4				
22	10	7	12	8	8	8	12	5	11	4				
23	10	9	10	10	11	11	14	11	11	10				
24	6	6	6	2	5	7	6	6	11	6				
25	14	7	6	9	10	6	7	6	8	5				

In this example, we will click on
Analyze
Dimension Reduction

⁵<http://www.statsoft.com/textbook/>

⁶http://www.amazon.com/Discovering-Statistics-Introducing-Statistical-Method/dp/1847879071/ref=sr_1_1?s=books&ie=UTF8&qid=13049

1

⁷<http://cnx.org/content/m40725/latest/8.1.png/image>

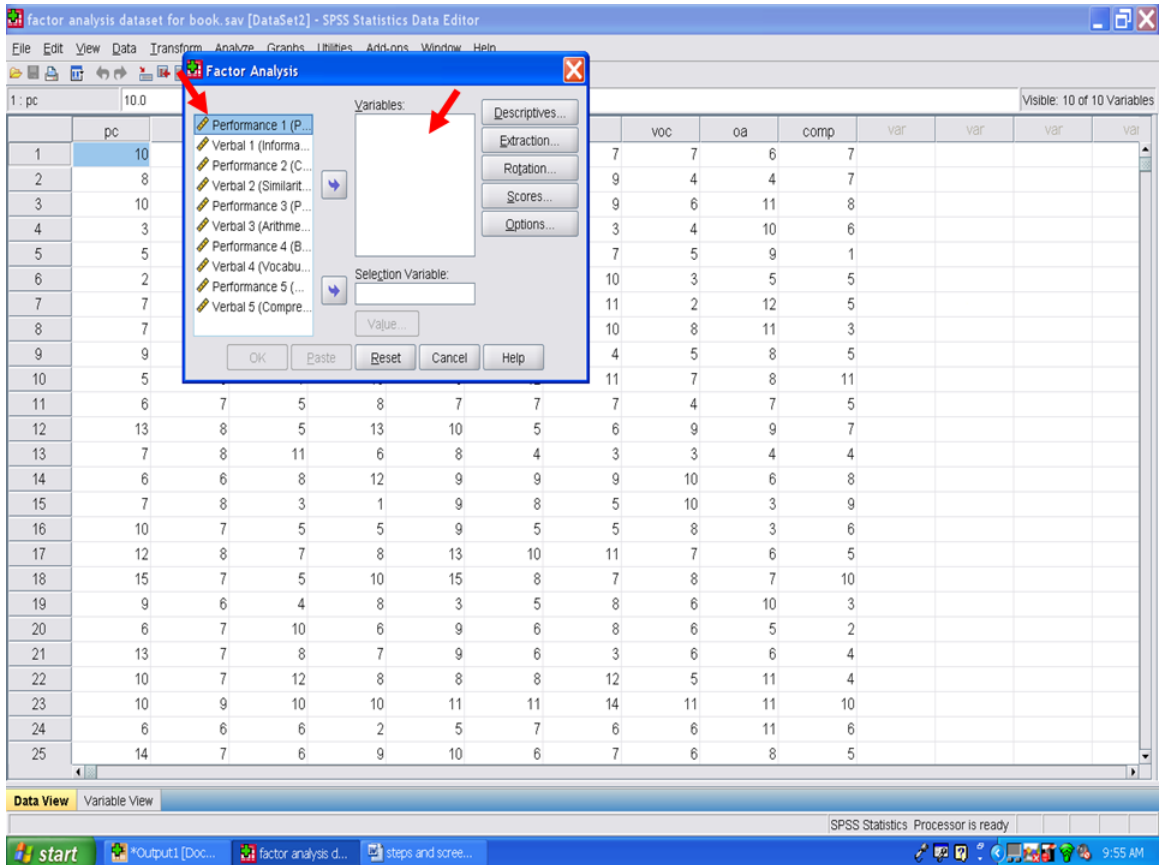
Factor

The screenshot shows the SPSS Statistics Data Editor window titled "Factor analysis dataset for book.sav [DataSet2] - SPSS Statistics Data Editor". The "Analyze" menu is open, and the "Dimension Reduction" option is selected, with the "Factor..." sub-option highlighted. Red arrows indicate the path from the "Analyze" menu to "Dimension Reduction" and then to "Factor...". The data table below shows 25 rows of data with 10 columns of variables.

	pc	inf	pa	ari	bd	voc	oa	comp	var	var	var	var
1	10		10	5	7	7	6	7				
2	8		7	9	9	4	4	7				
3	10		10	11	9	6	11	8				
4	3		7	6	3	4	10	6				
5	5		8	10	7	5	9	1				
6	2		9	7	10	3	5	5				
7	7		11	11	2	12	5					
8	7		14	8	10	8	11	3				
9	9		9	8	4	5	8	5				
10	5		9	12	11	7	8	11				
11	6		7	7	7	4	7	5				
12	13		10	5	6	9	9	7				
13	7		8	4	3	3	4	4				
14	6		9	9	9	10	6	8				
15	7	8	3	1	9	8	5	10	3			
16	10	7	5	5	9	5	8	3	6			
17	12	8	7	8	13	10	11	7	6			
18	15	7	5	10	15	8	7	8	7			
19	9	6	4	8	3	5	8	6	10			
20	6	7	10	6	9	6	8	6	5			
21	13	7	8	7	9	6	3	6	6			
22	10	7	12	8	8	8	12	5	11			
23	10	9	10	10	11	11	14	11	11			
24	6	6	6	2	5	7	6	6	11			
25	14	7	6	9	10	6	7	6	8			

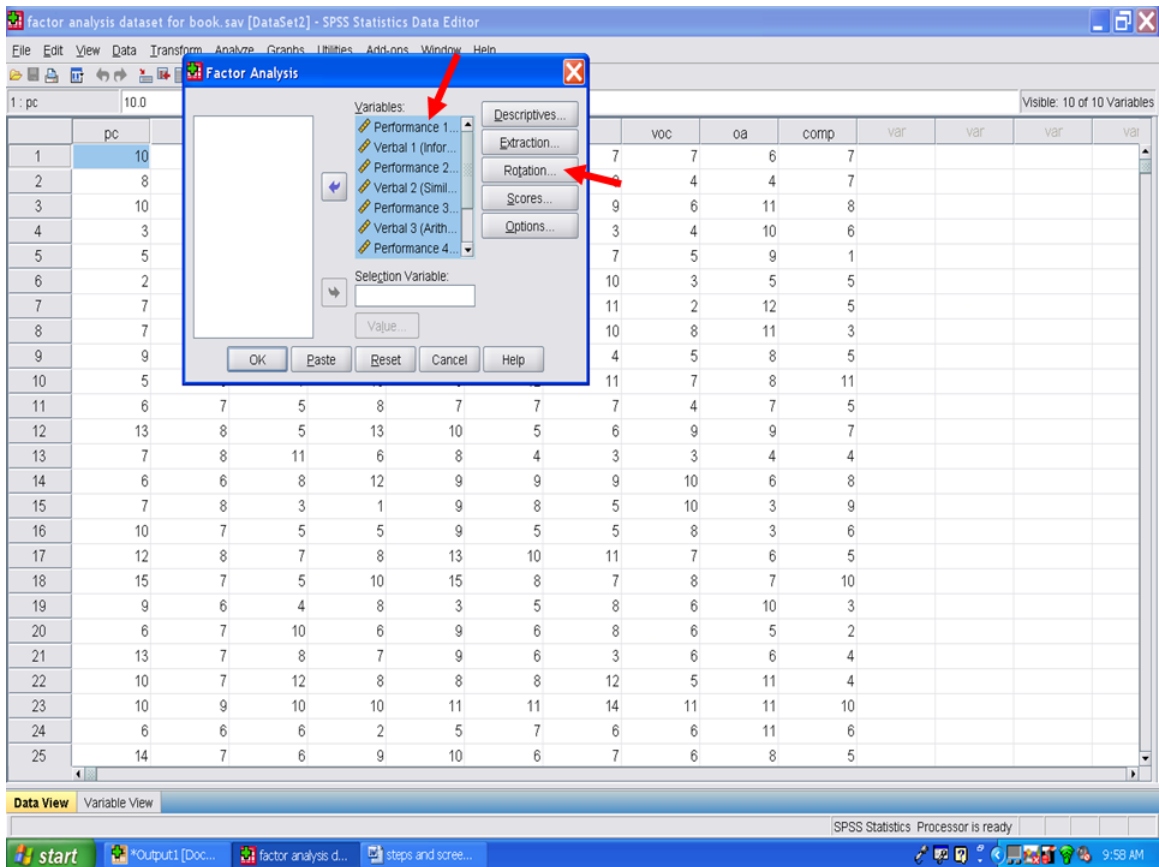
After clicking on factor, the following screen will appear. The variables on which you want a factor analysis conducted should be sent into the box titled Variables. In our example, we will send all 10 of the subscales into the Variables box.

⁸<http://cnx.org/content/m40725/latest/8.2.png/image>



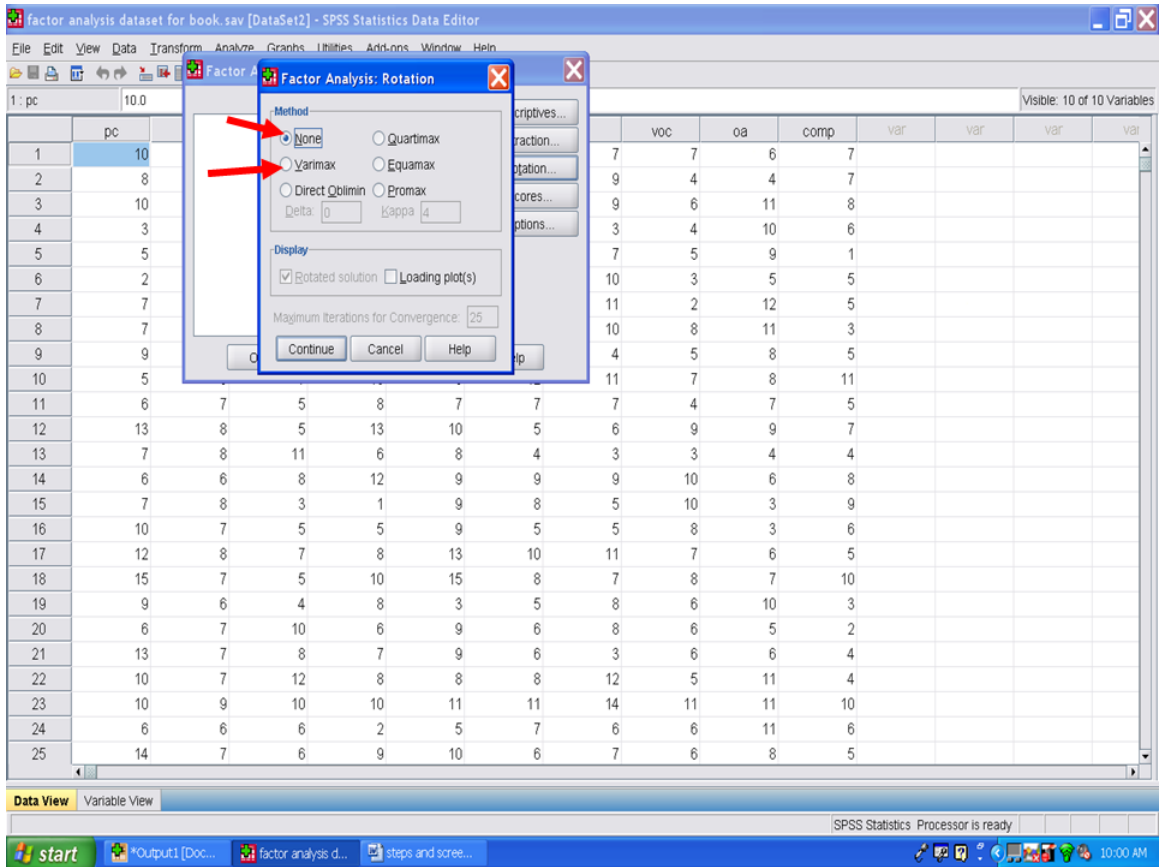
Once all of the subscales on which you want the factor analysis conducted have been placed into the Variables box, click on Rotation.

⁹<http://cnx.org/content/m40725/latest/8.3.png/image>



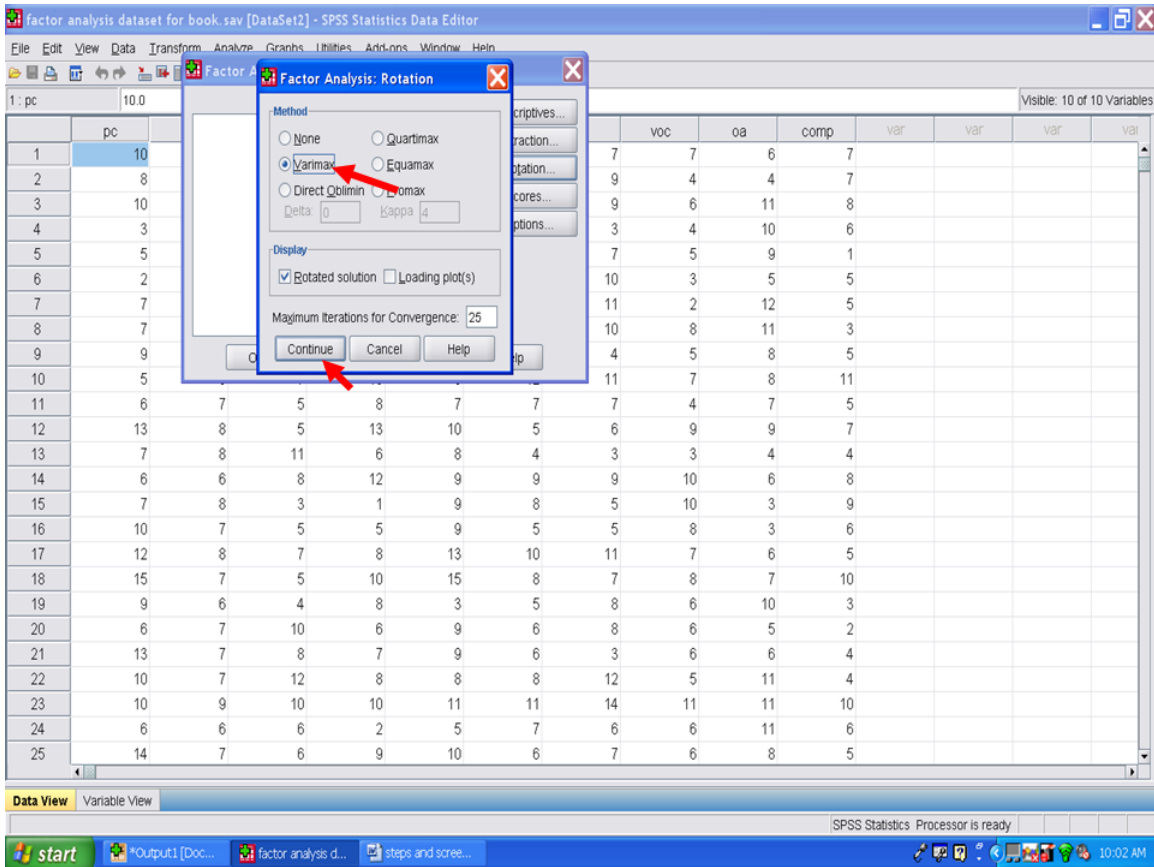
After clicking on Rotation, the following screen will appear. Under Method, the box for None is already checked because it is the SPSS default value. We will conduct a Varimax factor analysis. Click on Varimax.

¹⁰<http://cnx.org/content/m40725/latest/8.4.png/image>



Varimax has been clicked below. Click on Continue.

¹¹ <http://cnx.org/content/m40725/latest/8.5.png/image>



The screen below will appear after you have clicked on Continue.
Now click on Descriptives.

¹²<http://cnx.org/content/m40725/latest/8.6.png/image>

The screenshot shows the SPSS Factor Analysis dialog box overlaid on a data editor window. The dialog box has a 'Variables:' list containing 'Performance 1...', 'Verbal 1 (Infor...', 'Performance 2...', 'Verbal 2 (Similarities) sim', 'Performance 3...', 'Verbal 3 (Arith...', and 'Performance 4...'. The 'Univariate Descriptives' checkbox is checked, and a red arrow points to it. Other options like 'Extraction...', 'Rotation...', 'Scores...', and 'Options...' are also visible. The background data editor shows a table with columns 'pc', 'voc', 'oa', 'comp', and 'var'.

pc	voc	oa	comp	var	var	var	var
1	7	7	6	7			
2	9	4	4	7			
3	9	6	11	8			
4	3	4	10	6			
5	7	5	9	1			
6	10	3	5	5			
7	11	2	12	5			
8	10	8	11	3			
9	4	5	8	5			
10	11	7	8	11			
11	6	7	4	7	5		
12	13	8	5	13	10	5	6
13	7	8	11	6	8	4	3
14	6	6	8	12	9	9	9
15	7	8	3	1	9	8	5
16	10	7	5	5	9	5	8
17	12	8	7	8	13	10	11
18	15	7	5	10	15	8	7
19	9	6	4	8	3	5	8
20	6	7	10	6	9	6	6
21	13	7	8	7	9	6	3
22	10	7	12	8	8	8	12
23	10	9	10	10	11	11	14
24	6	6	6	2	5	7	6
25	14	7	6	9	10	6	7

Your screen should now look like the one below. Click on Univariate Descriptives. Unclick Initial Solution.

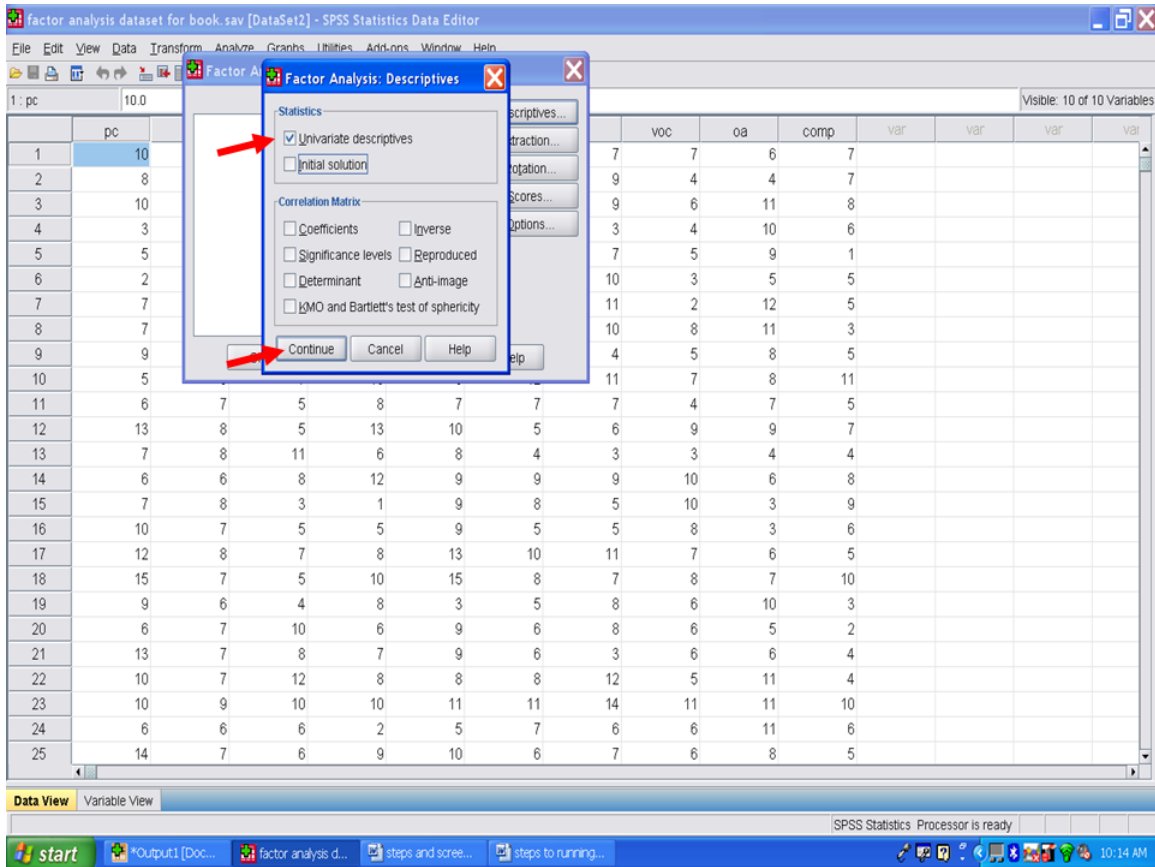
¹³<http://cnx.org/content/m40725/latest/8.7.png/image>

The screenshot shows the SPSS Factor Analysis: Descriptives dialog box. The 'Statistics' section has 'Initial solution' checked. The 'Correlation Matrix' section has several options unchecked. Red arrows point to the 'Initial solution' checkbox and the 'Continue' button. The background shows a data editor with variables 'pc', 'voc', 'oa', 'comp', and 'var'.

pc	voc	oa	comp	var	var	var	var
10	7	7	6	7			
8	9	4	4	7			
10	9	6	11	8			
3	3	4	10	6			
5	7	5	9	1			
2	10	3	5	5			
7	11	2	12	5			
7	10	8	11	3			
9	4	5	8	5			
5	11	7	8	11			
7	7	4	7	5			
13	6	9	9	7			
7	3	3	4	4			
6	9	10	6	8			
8	5	10	3	9			
10	7	5	8	3			
7	12	7	6	5			
8	13	10	7	10			
15	9	8	7	3			
6	6	4	8	5			
7	10	6	9	6			
13	8	7	9	6			
7	9	6	3	6			
10	12	5	11	4			
10	14	11	11	10			
6	6	6	11	6			
14	7	6	8	5			

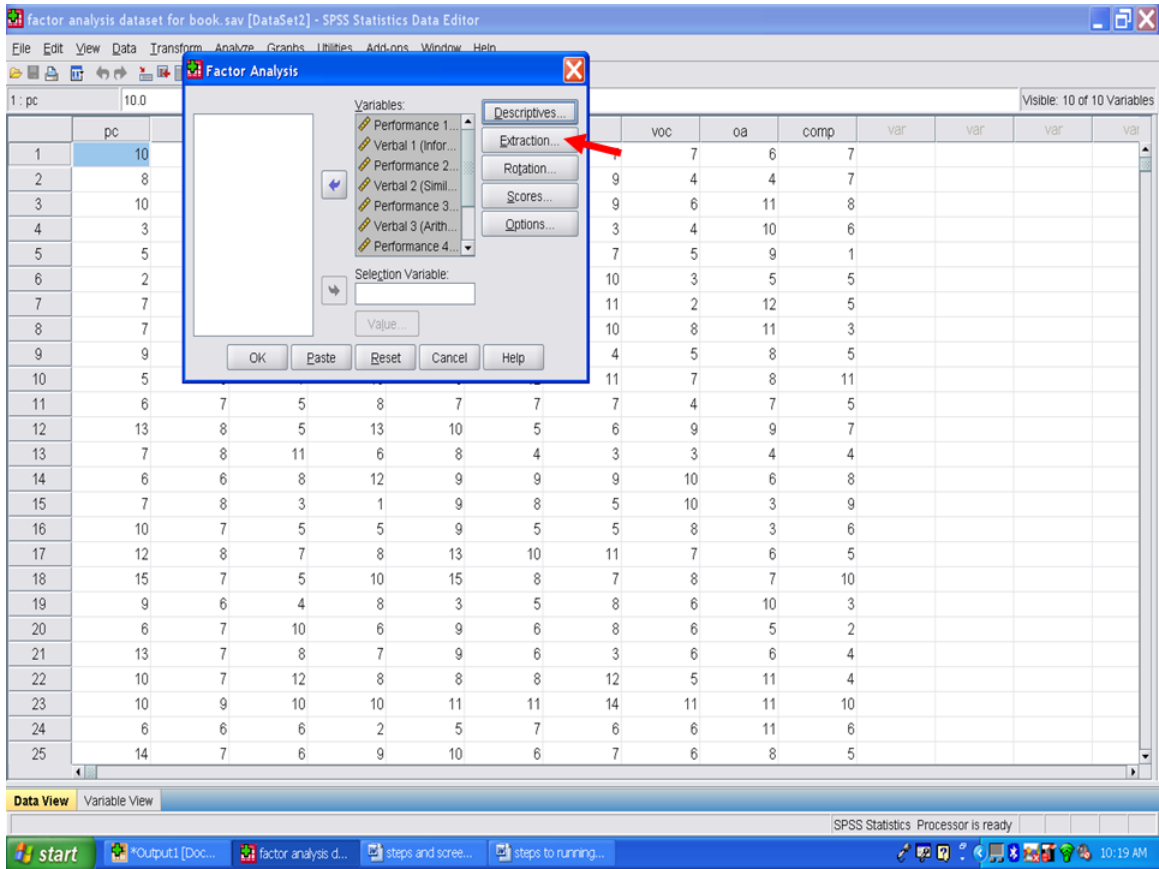
Your screen will now look like the following screenshot.
Click on Continue.

¹⁴<http://cnx.org/content/m40725/latest/8.8.png/image>



The following screen should now be present.
 Click on Extraction.

¹⁵<http://cnx.org/content/m40725/latest/8.9.png/image>



¹⁶<http://cnx.org/content/m40725/latest/8.10.png/image>

Chapter 10

10. Factor Analysis: Part II¹



NOTE: This chapter is published by NCPEA Press² and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of education administration.

About the Authors

John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website (Writing and Statistical Help³) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

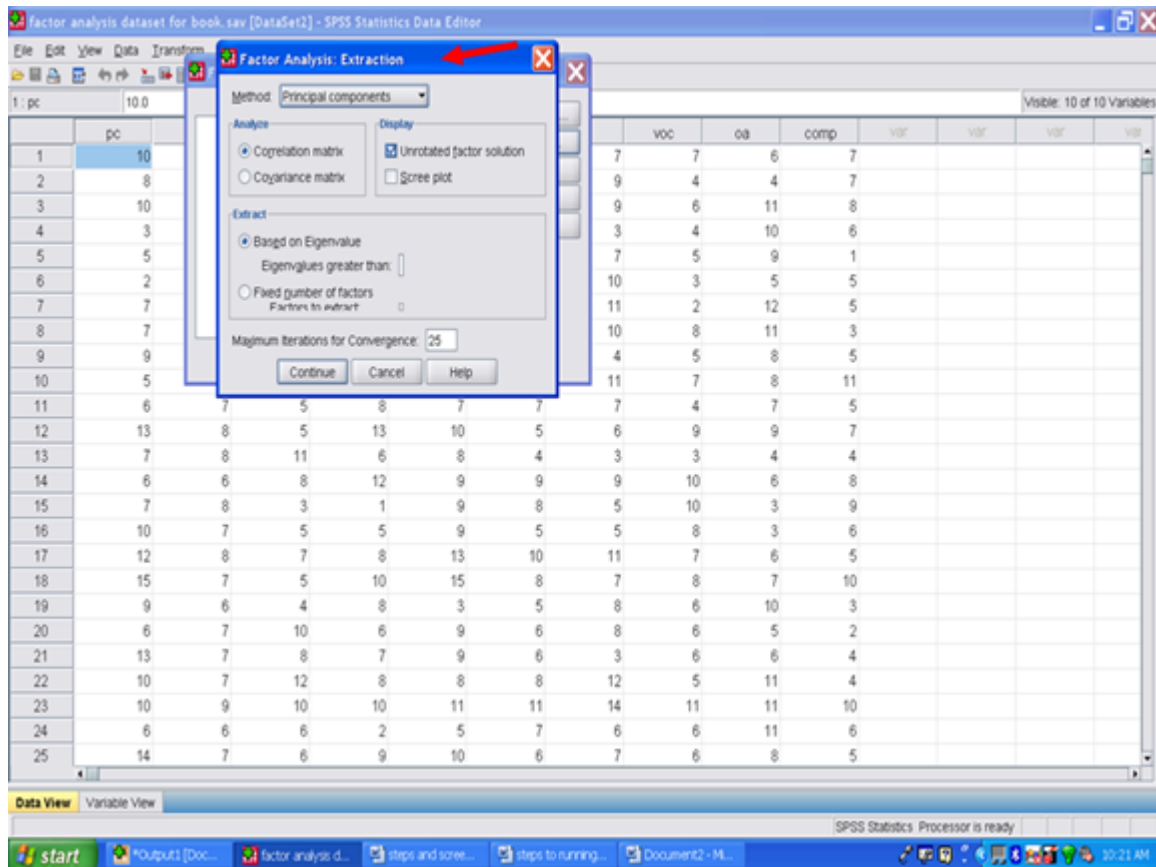
Ana Rojas-LeBouef is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. Dr. LeBoeuf recently completed her doctoral degree in Reading, where she conducted a 16-year analysis of Texas statewide data regarding the achievement gap. Her research interests lie in examining the inequities in achievement among ethnic groups. Dr. Rojas-LeBouef also assists students and faculty in their writing and statistical needs on the Writing and Statistical Help website.

After clicking on Extraction, the following screen should now be present.

¹This content is available online at <<http://cnx.org/content/m40727/1.2/>>.

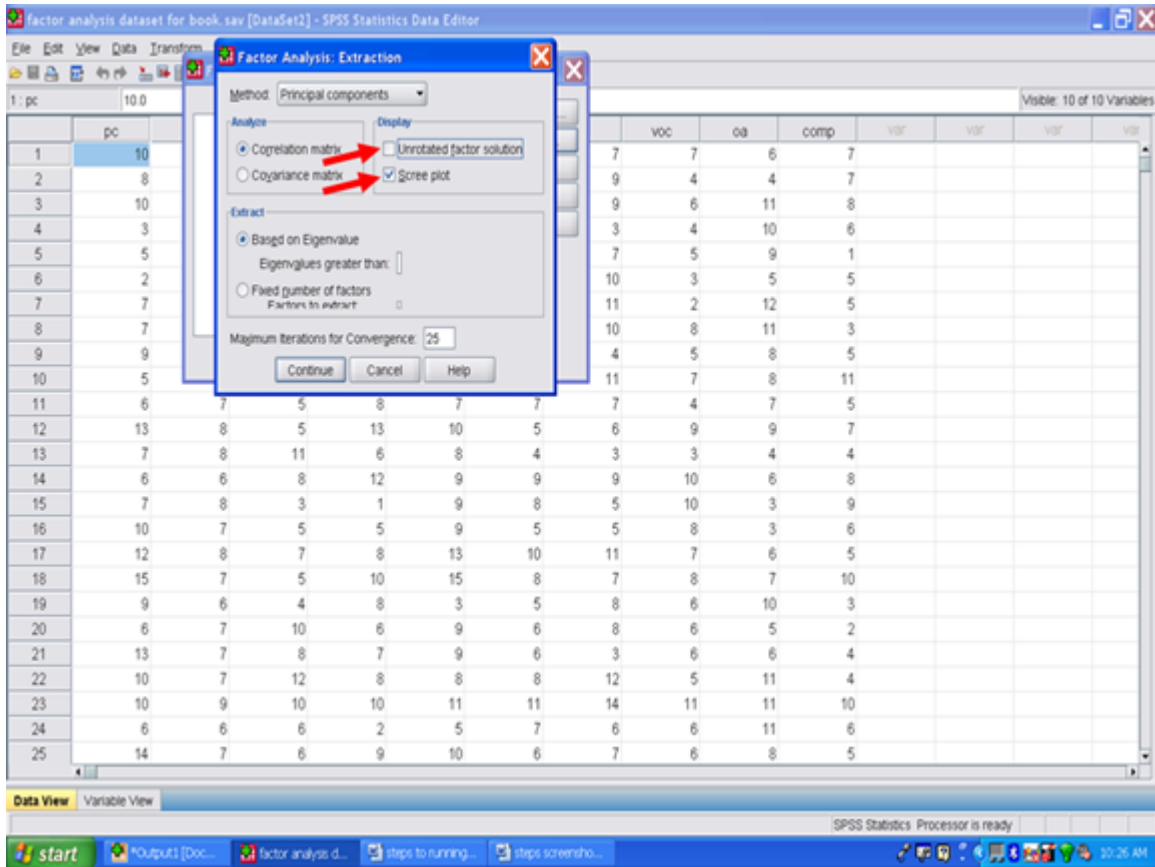
²<http://www.ncpeapublications.org/books.html>

³<http://cnx.org/content/m40727/latest/www.writingandstatisticalhelp>



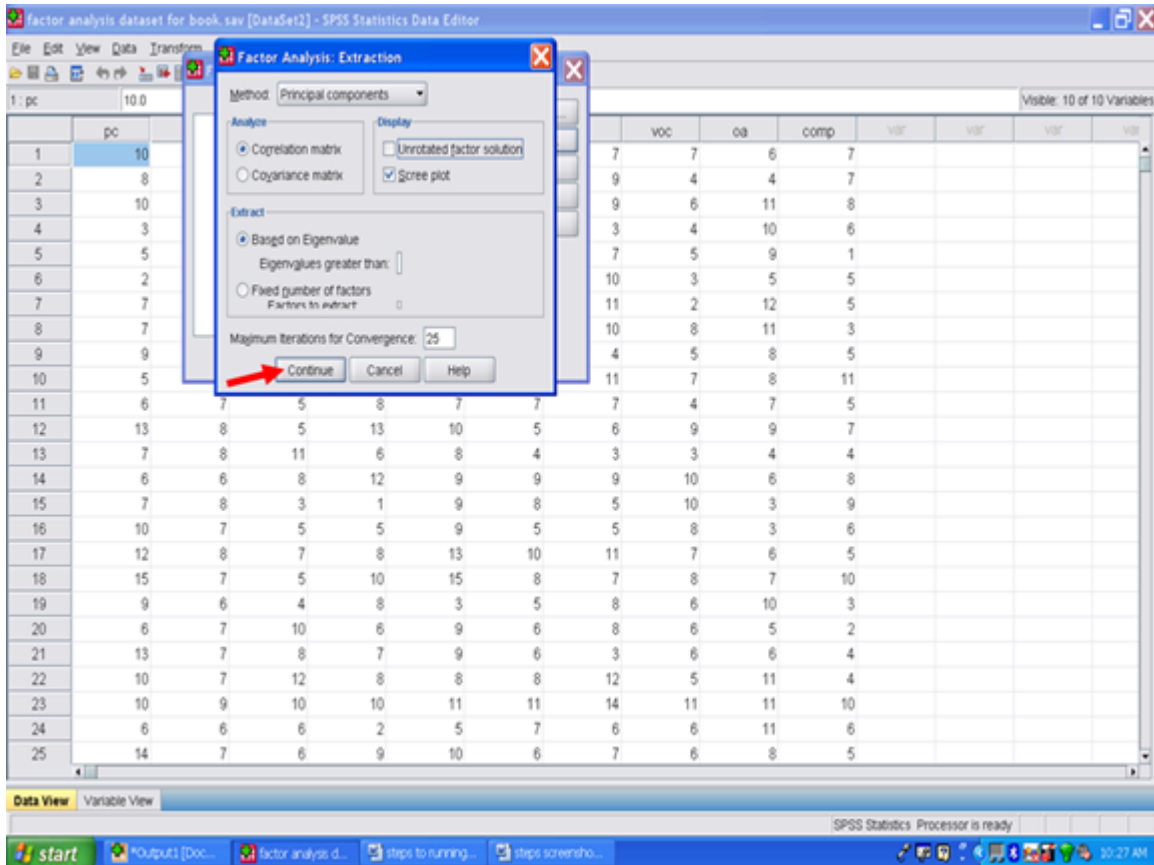
Under Display, click on Screen plot.
 Unclick the Unrotated factor solution.

⁴<http://cnx.org/content/m40727/latest/9.1.png/image>



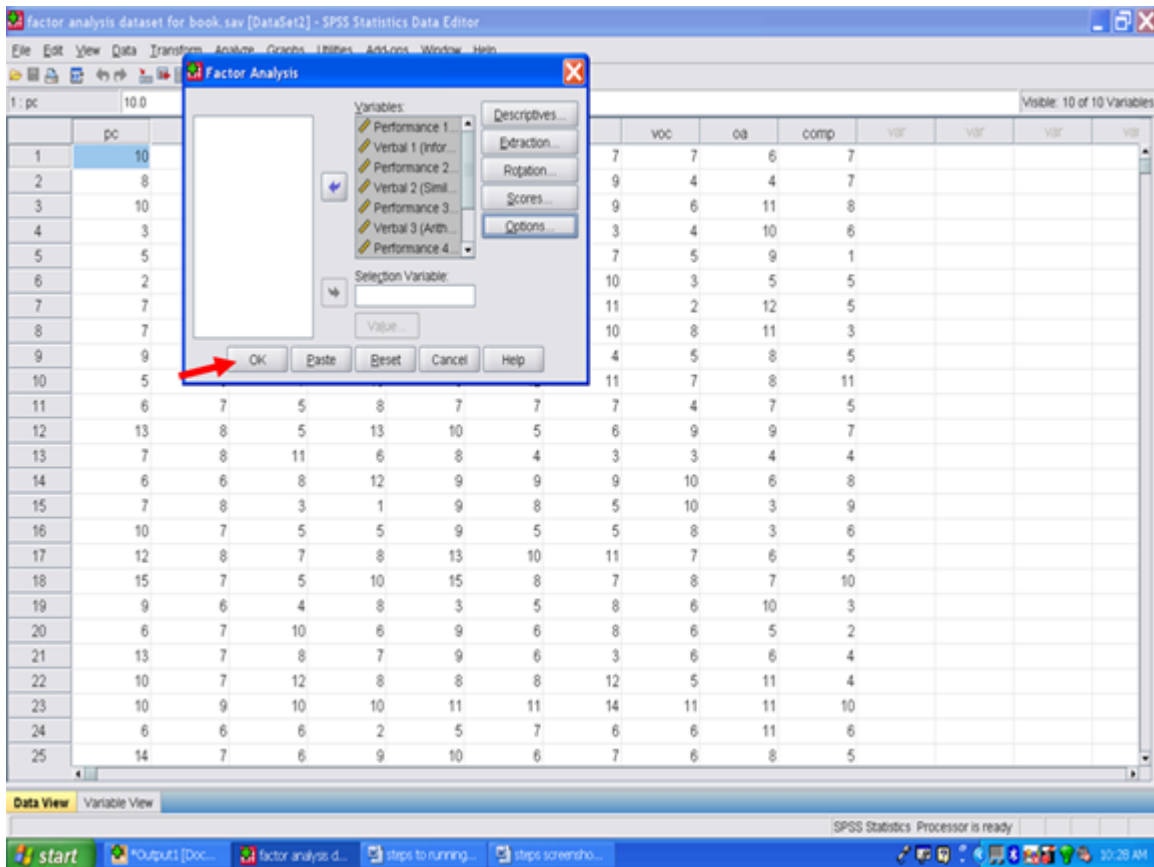
Click on Continue.

⁵<http://cnx.org/content/m40727/latest/9.2.png/image>



After clicking on Continue, the screen below should be present.
Now click on OK.

⁶<http://cnx.org/content/m40727/latest/9.3.png/image>



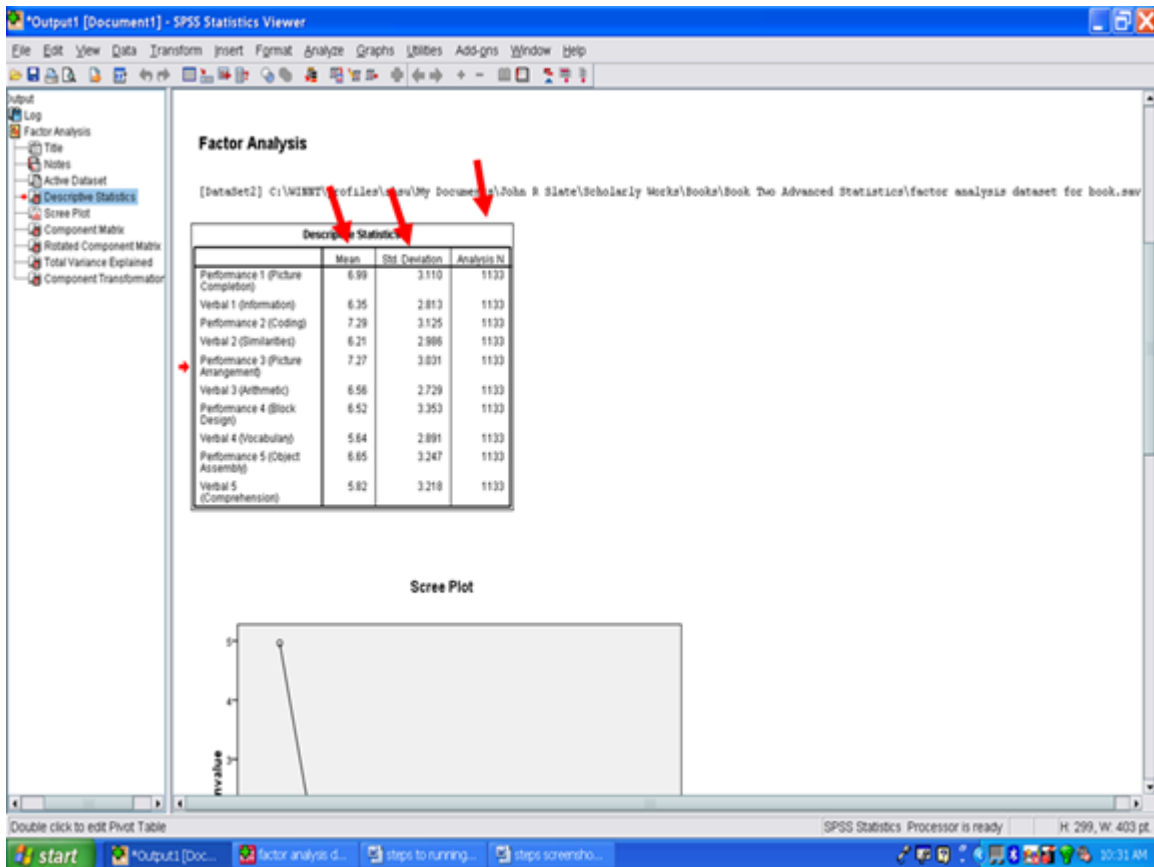
You should now have factor analysis results. If SPSS does not send you to your output screen, click on the Output icon at the bottom of your screen.

⁷ <http://cnx.org/content/m40727/latest/9.4.png/image>

	pc	inf	cod	sim	pa	ari	bd	voc	oa	comp	var	var	var	var
1	10	7	8	7	10	5	7	7	6	7				
2	8	8	7	6	7	9	9	4	4	7				
3	10	8	2	9	10	11	9	6	11	8				
4	3	4	11	4	7	6	3	4	10	6				
5	5	7	17	6	8	10	7	5	9	1				
6	2	5	8	8	9	7	10	3	5	5				
7	7	4	5	8	7	11	11	2	12	5				
8	7	7	7	7	14	8	10	8	11	3				
9	9	6	10	6	9	8	4	5	8	5				
10	5	9	7	10	9	12	11	7	8	11				
11	6	7	5	8	7	7	7	4	7	5				
12	13	8	5	13	10	5	6	9	9	7				
13	7	8	11	6	8	4	3	3	4	4				
14	6	6	8	12	9	9	9	10	6	8				
15	7	8	3	1	9	8	5	10	3	9				
16	10	7	5	5	9	5	5	8	3	6				
17	12	8	7	8	13	10	11	7	6	5				
18	15	7	5	10	15	8	7	8	7	10				
19	9	6	4	8	3	5	8	6	10	3				
20	6	7	10	6	9	6	8	6	5	2				
21	13	7	8	7	9	6	3	6	6	4				
22	10	7	12	8	8	8	12	5	11	4				
23	10	9	10	10	11	11	14	11	11	10				
24	6	6	6	2	5	7	6	6	11	6				
25	14	7	6	9	10	6	7	6	8	5				

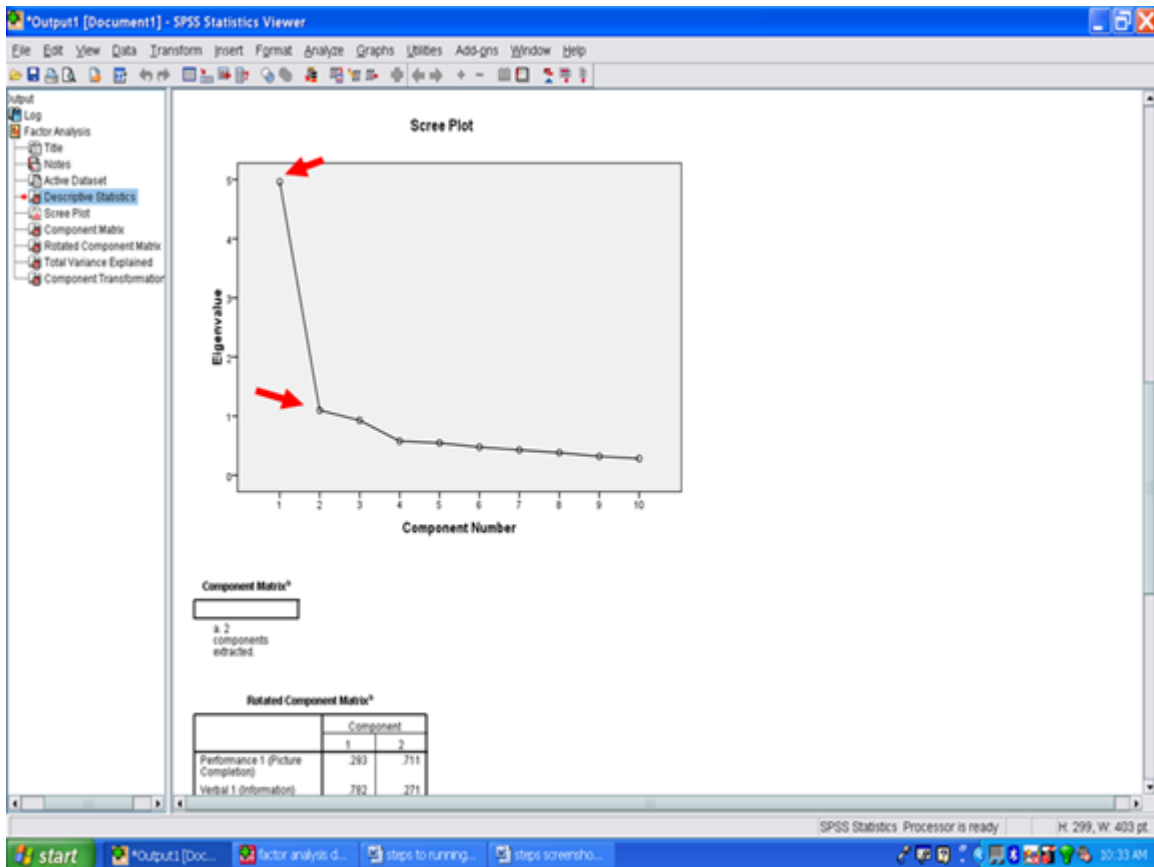
In your SPSS Output screen, the first table of importance is the Descriptive Statistics table. The sample size (n), M , and SD for each of the 10 subscales used in the factor analysis are present in this table.

⁸<http://cnx.org/content/m40727/latest/9.5.png/image>



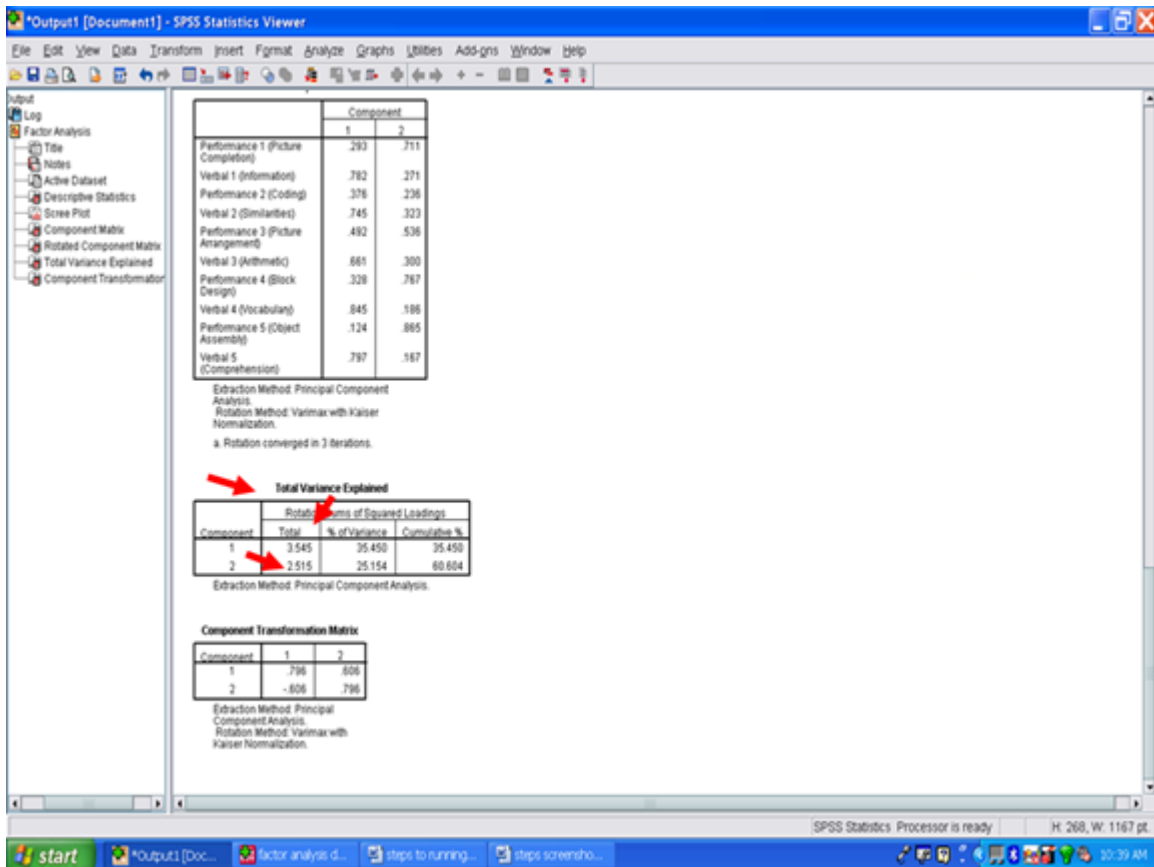
Underneath the Descriptive Statistics table is the Scree plot. What the scree plot does is to plot the eigenvalue against the factor number (Cattell, 1966; Zwick & Velicer, 1986). The eigenvalue and factor number are present in the table labeled Total Variance Explained. You will note in the Scree Plot below that the line is essentially flat after the second factor. This plot below depicts the presence of two possible factors, 1 and 2.

⁹<http://cnx.org/content/m40727/latest/9.6.png/image>



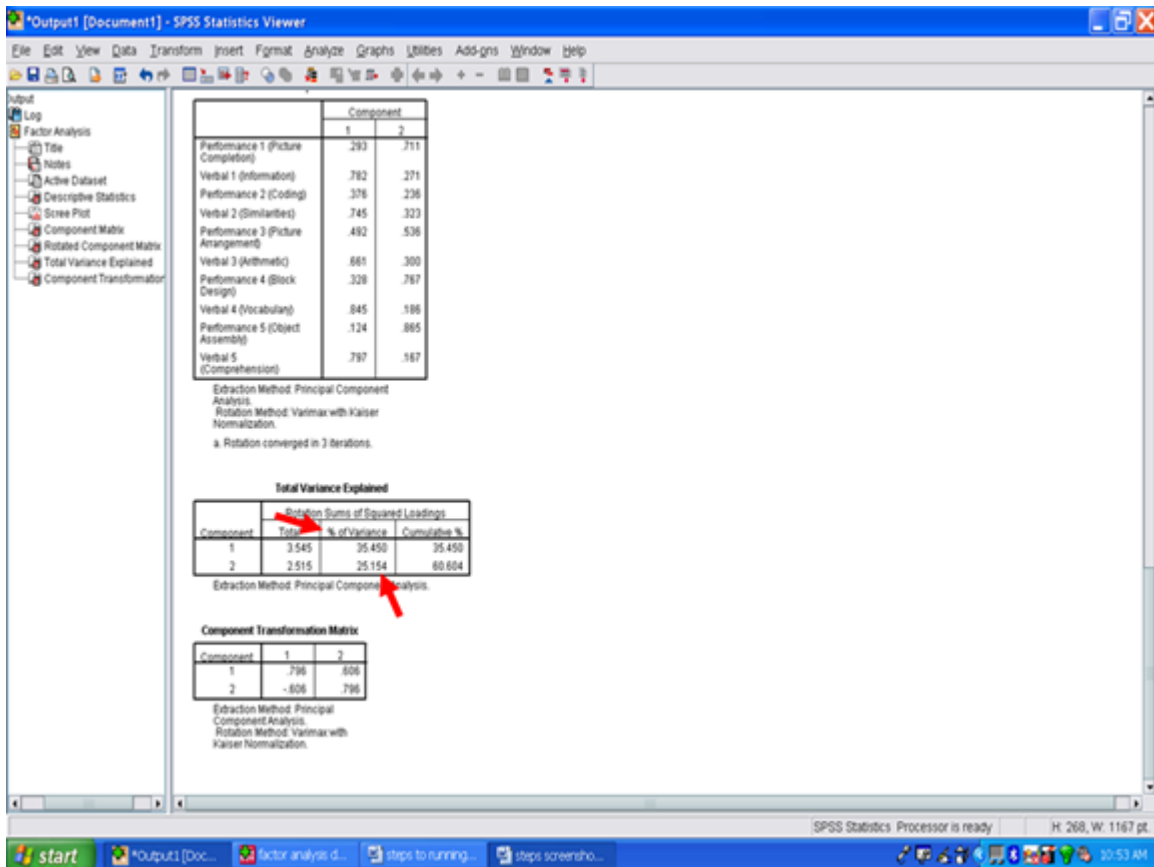
In the Total Variance Explained table, statistical information that is depicted in the Scree Plot is presented. The column labeled Total reflects the eigenvalues for the two factors. To determine the number of factors you may have: look at the eigenvalue column. All factors with values less than 1 are considered to be statistically insignificant and are disregarded (Kaiser, 1958). Thus, an eigenvalue of at least 1.0 must be present for a factor to be possible. In the Total Variance Explained table below, two factors are present with eigenvalues greater than 1.00: Factor 1 has an eigenvalue of 3.545 and Factor 3 has an eigenvalue of 2.515.

¹⁰<http://cnx.org/content/m40727/latest/9.7.png/image>



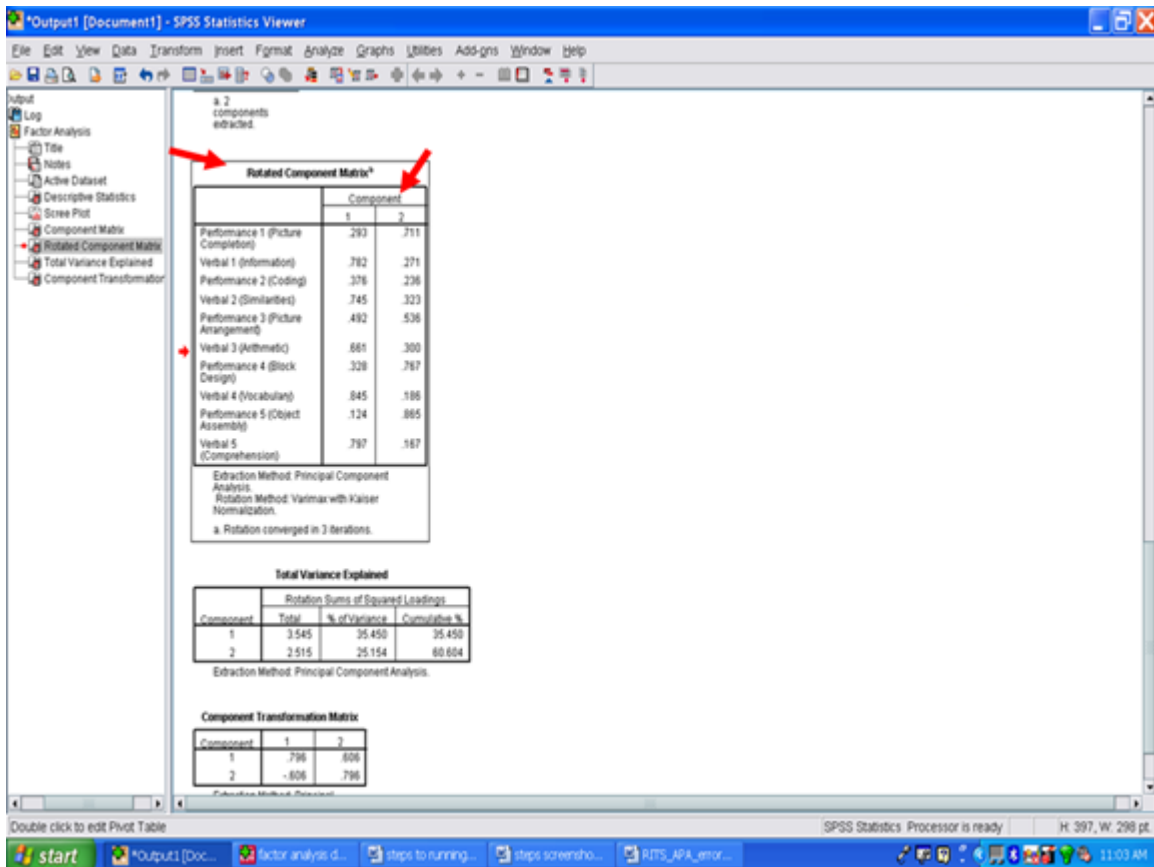
After checking the eigenvalues to ensure that they are greater than one (Kaiser, 1958), the percent of variance must be examined. To constitute a viable factor, the factor should account for at least 5 percent of the variance to be used. In the example below, Factor 1 accounts for 35.45% of the variance and Factor 2 explains 25.15% of the variance. In both cases, the factors account for much more than the minimum 5%. Together these two factors account for 60.60% of the variance.

¹¹ <http://cnx.org/content/m40727/latest/9.8.png/image>



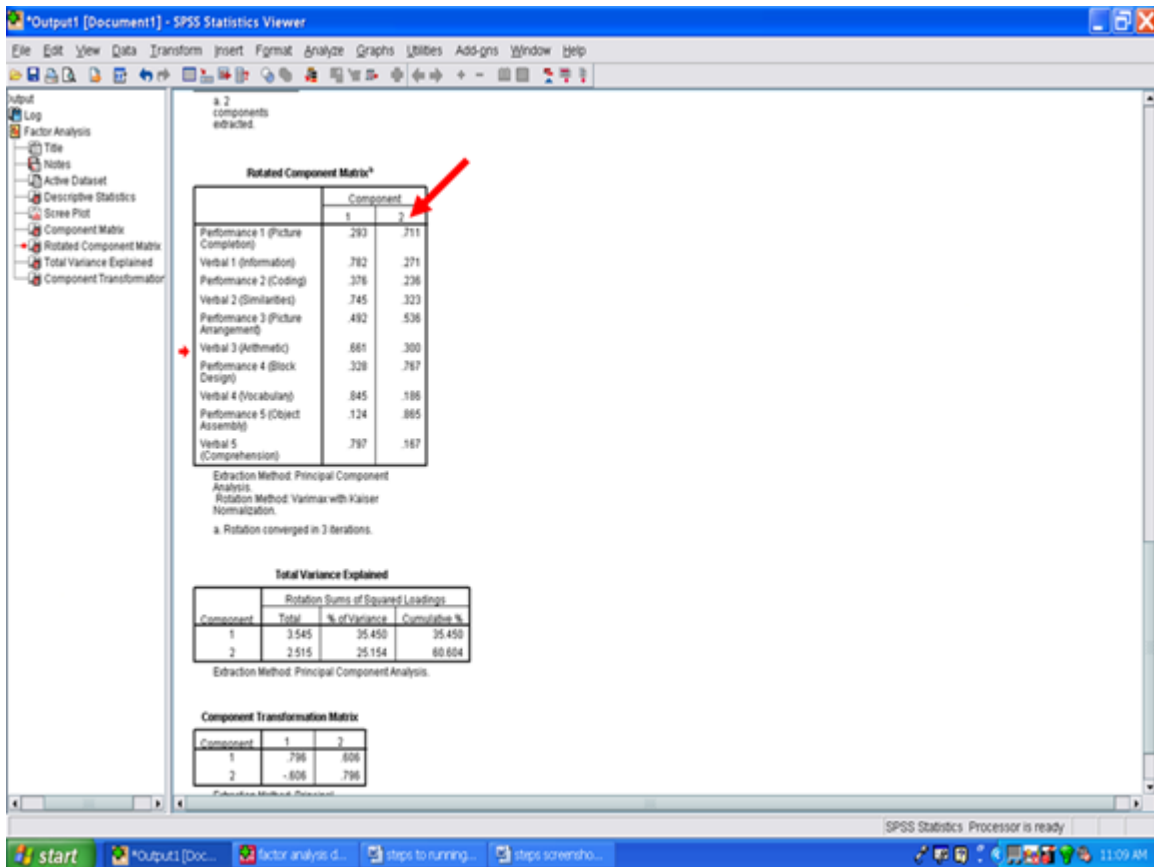
Next we will examine the Rotated Component Matrix. This table contains the factor loadings or pattern/structure coefficients for the 10 subscales analyzed in this Varimax procedure. To determine whether a subscale is a component of a specific factor, a cutoff value of .3 (Lambert & Durand, 1975) is recommended as an acceptable minimum value for pattern/structure coefficients. In this example, Verbal 1 (Information) has a pattern/structure coefficient of .782; Verbal 1 (Similarities) of .745; Verbal 3 (Arithmetic) of .661; Verbal 4 (Vocabulary) of .845; and Verbal 5 (Comprehension) of .797. Though above the cutoff of .3, Performance 2 (Coding), Performance 3 (Picture Arrangement), and Performance 4 (Block Design) are well below the coefficients for the Verbal subscales.

¹²<http://cnx.org/content/m40727/latest/9.9.png/image>



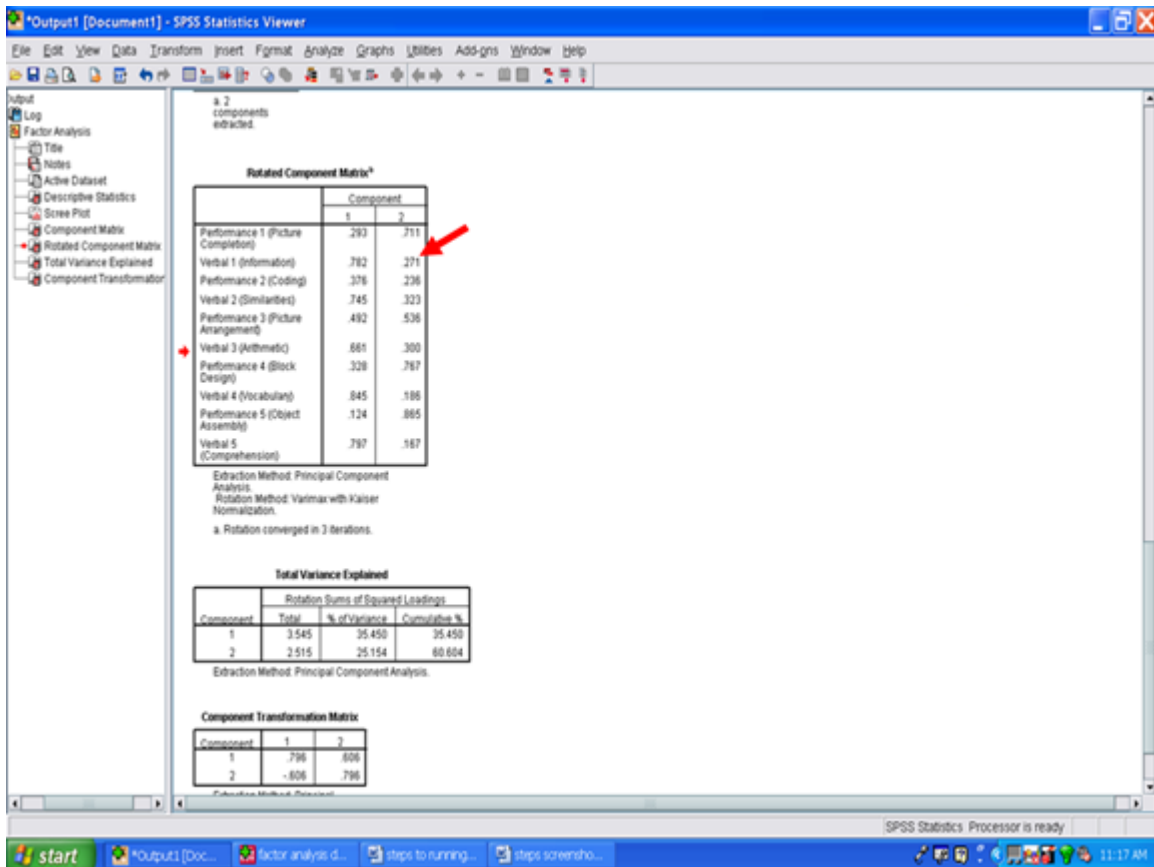
For Factor 2, the pattern/structure coefficients were .711 for Performance 1 (Picture Completion); .536 for Performance 3 (Picture Arrangement); .767 for Performance 4 (Block Design); and .865 for Performance 5 (Object Assembly). Two of the Verbal subscales were at or above the cutoff value of .3, however, their pattern/structure coefficients were markedly lower than the ones for the four Performance subscales mentioned above.

¹³<http://cnx.org/content/m40727/latest/9.10.png/image>



In this example, it appears that the Performance 2 (Coding) subscale is not part of Factor 2. It may be a component of Factor 1.

¹⁴<http://cnx.org/content/m40727/latest/9.11.png/image>



You have now conducted a Varimax factor rotation of the 10 subscales of the Wechsler Intelligence Scale for Children-Third Edition. To determine whether your factors are internally consistent, you would need to perform an internal consistency analysis for the subscales that constitute Factor 1 and a separate internal consistency analysis for the subscales that constitute Factor 2. See our chapter on steps and screenshots for conducting internal consistency analyses.

10.1 References

- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245–276.
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*, 187–200.
- Lambert, Z. V., & Durand, R. M. (1975). Some precautions in using canonical analysis. *Journal of Market Research, 12*, 468–475.
- Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432–442.

¹⁵<http://cnx.org/content/m40727/latest/9.12.png/image>

Chapter 11

11. Multivariate Analysis of Variance: Part I¹



NOTE: This chapter is published by NCPEA Press² and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of education administration.

About the Authors

John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website (Writing and Statistical Help³) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

Ana Rojas-LeBouef is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. Dr. LeBoeuf recently completed her doctoral degree in Reading, where she conducted a 16-year analysis of Texas statewide data regarding the achievement gap. Her research interests lie in examining the inequities in achievement among ethnic groups. Dr. Rojas-LeBouef also assists students and faculty in their writing and statistical needs on the Writing and Statistical Help website.

In this set of steps, readers will calculate a multivariate analysis of variance procedure, following the determination of the extent to which data for the dependent variables reflect normal distributions. Although a parametric statistical procedure requires that its data be reflective of a normal curve, the multivariate analysis of variance procedure is

¹This content is available online at <<http://cnx.org/content/m40728/1.2/>>.

²<http://www.ncpeapublications.org/books.html>

³<http://cnx.org/content/m40728/latest/www.writingandstatisticalhelp>

regarded as being sufficiently robust that it can withstand most violations. For detailed information regarding the assumptions underlying the multivariate analysis of variance (MANOVA) procedure, readers are referred to the Hyperstats Online Statistics Textbook at <http://davidmlane.com/hyperstat/>⁴; to the *Electronic Statistics Textbook* (2011) at <http://www.statsoft.com/textbook/>⁵; or to Andy Field's (2009) *Discovering Statistics Using SPSS* at http://www.amazon.com/Discovering-Statistics-Introducing-Statistical-Method/dp/1847879071/ref=sr_1_1?s=books&ie=UTF8&qid=1304967862&sr=1-1⁶

Research questions for which a MANOVA procedure is appropriate involve asking for differences in multiple dependent variables by group membership (i.e., more than two groups may be present). In addition to multiple dependent variables being present, multiple independent variables can be present as well. That is, differences in several achievement variables could be analyzed by student gender, student ethnicity, student socioeconomic status, and the like. A specific research question that could be addressed is, "What is the difference in academic achievement among elementary school students as a function of ethnic membership, gender, and grade level?" Academic achievement in this example could be reading, math, science, and social studies scores. The independent variables are ethnicity, gender, and grade level.

For this particular chapter, the research question on which we will conduct a MANOVA will be: What is the difference in verbal aptitude among elementary reading groups? Verbal aptitude consists of Verbal 1 (inf), Verbal 2 (sim), Verbal 3 (ari), and Verbal 4 (voc).

	group	pc	inf	cod	sim	pa	ari	bd	voc	sa	Gender
1	3	8	12	6	12	9	9	8	8	8	
2	3	8	12	6	12	9	9	8	8	8	
3	3	9	5	6	5	11	6	10	4	7	
4	3	3	4	11	4	7	6	3	4	10	1.00
5	3	5	7	17	6	8	10	7	5	9	1.00
6	3	2	5	8	8	9	7	10	3	5	1.00
7	3	9	6	10	6	9	8	4	5	8	1.00
8	3	5	9	7	10	9	12	11	7	8	1.00
9	3	6	7	5	8	7	7	7	4	7	1.00
10	3	7	8	11	6	8	4	3	3	4	1.00
11	3	6	6	8	12	9	9	9	10	6	1.00
12	3	7	8	3	1	9	8	5	10	3	1.00
13	3	12	8	7	8	13	10	11	7	6	1.00
14	3	10	9	10	10	11	11	14	11	11	1.00
15	3	3	4	4	8	7	7	5	6	7	1.00
16	3	3	4	4	8	7	7	5	6	7	1.00
17	3	8	4	9	7	9	8	10	7	7	1.00
18	3	5	9	6	6	5	7	9	4	4	1.00
19	3	6	8	11	8	10	11	2	8	4	1.00
20	3	3	6	8	6	6	8	3	2	6	1.00
21	3	11	8	8	10	9	9	12	11	9	1.00
22	3	17	11	6	11	13	9	14	11	13	1.00
23	3	9	12	6	11	11	7	9	9	12	1.00
24	1	7	10	11	4	8	7	11	7	10	1.00
25	3	7	8	7	10	11	10	7	9	8	1.00
26	3	5	5	8	6	6	8	4	8	4	1.00
27	3	8	7	7	7	2	8	8	6	11	1.00
28	3	9	2	6	8	10	7	4	6	6	1.00
29	3	14	8	8	8	12	8	8	8	9	1.00
30	3	8	8	7	10	11	11	3	9	8	1.00
31	3	11	3	4	7	8	8	10	4	10	1.00
32	3	10	7	9	8	10	8	10	5	11	1.00
33	3	4	11	9	6	13	11	9	4	7	1.00
34	3	12	12	7	8	4	8	9	9	8	1.00
35	3	11	9	7	8	10	11	9	11	9	1.00
36	3	8	5	9	1	7	9	11	8	8	1.00
37	3	7	6	6	7	6	11	6	7	6	1.00

⁴<http://davidmlane.com/hyperstat/>

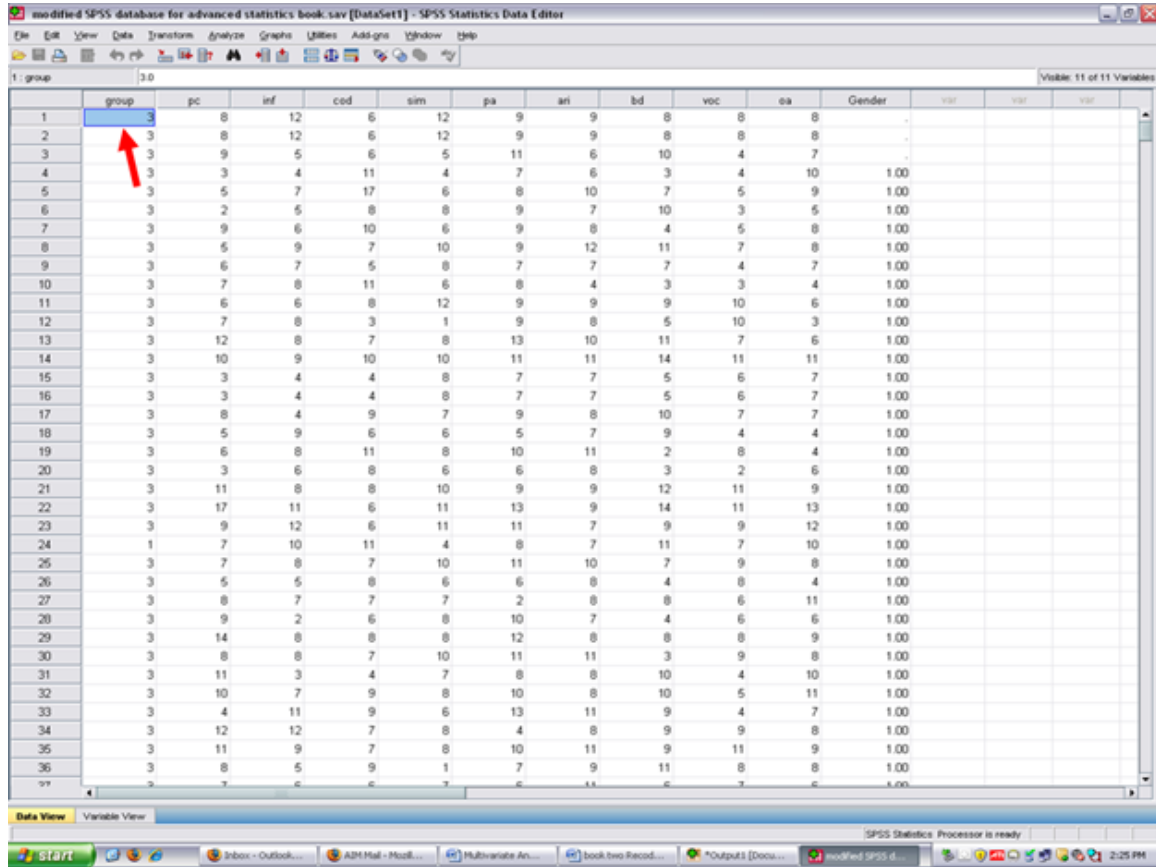
⁵<http://www.statsoft.com/textbook/>

⁶http://www.amazon.com/Discovering-Statistics-Introducing-Statistical-Method/dp/1847879071/ref=sr_1_1?s=books&ie=UTF8&qid=1304967862&sr=1-1

¹

⁷<http://cnx.org/content/m40728/latest/10.1.png/image>

In this example, our independent or grouping variable is elementary reading group.



By clicking on variable view, you can see how many groups of students are present in the reading group variable; the names of the groups; and the numbers which have been assigned to each group.

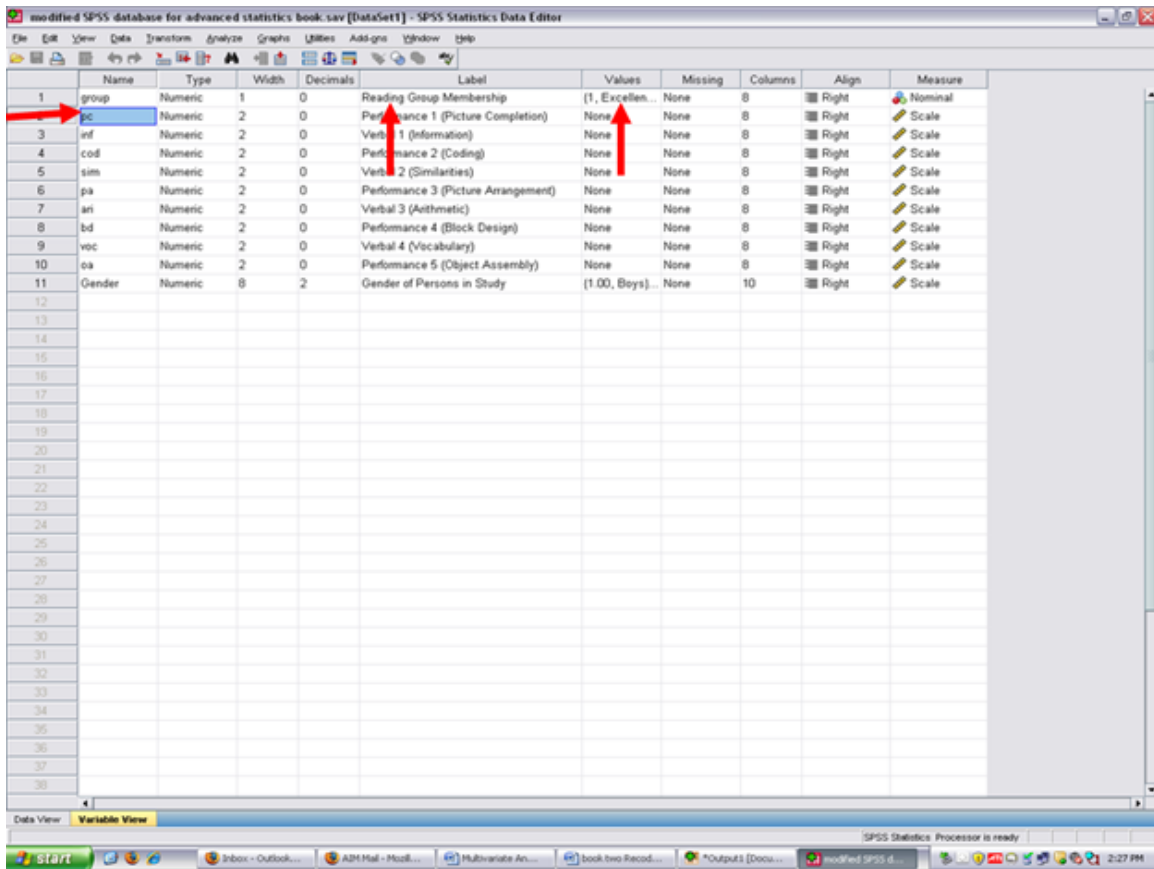
⁸<http://cnx.org/content/m40728/latest/10.2.png/image>

The screenshot displays the SPSS Statistics Data Editor window. The main area shows a data grid with 36 rows and 11 columns. The columns are labeled: group, pc, inf, cod, sim, pa, an, bd, voc, oa, and Gender. The 'group' column is highlighted in blue. A red arrow points to the 'group' column header. The bottom of the window shows the 'Data View' and 'Variable View' tabs, with 'Variable View' selected. The Windows taskbar at the bottom shows the Start button and several open applications.

	group	pc	inf	cod	sim	pa	an	bd	voc	oa	Gender	vis1	vis2	vis3
1	3	8	12	6	12	9	9	8	8	8	.			
2	3	8	12	6	12	9	9	8	8	8	.			
3	3	9	5	6	5	11	6	10	4	7	.			
4	3	3	4	11	4	7	6	3	4	10	1.00			
5	3	5	7	17	6	8	10	7	5	9	1.00			
6	3	2	5	8	8	9	7	10	3	5	1.00			
7	3	9	6	10	6	9	8	4	5	8	1.00			
8	3	5	9	7	10	9	12	11	7	8	1.00			
9	3	6	7	5	8	7	7	7	4	7	1.00			
10	3	7	8	11	6	8	4	3	3	4	1.00			
11	3	6	6	8	12	9	9	9	10	6	1.00			
12	3	7	8	3	1	9	8	5	10	3	1.00			
13	3	12	8	7	8	13	10	11	7	6	1.00			
14	3	10	9	10	10	11	11	14	11	11	1.00			
15	3	3	4	4	8	7	7	5	6	7	1.00			
16	3	3	4	4	8	7	7	5	6	7	1.00			
17	3	8	4	9	7	9	8	10	7	7	1.00			
18	3	5	9	6	6	5	7	9	4	4	1.00			
19	3	6	8	11	8	10	11	2	8	4	1.00			
20	3	3	6	8	6	6	8	3	2	6	1.00			
21	3	11	8	8	10	9	9	12	11	9	1.00			
22	3	17	11	6	11	13	9	14	11	13	1.00			
23	3	9	12	6	11	11	7	9	9	12	1.00			
24	1	7	10	11	4	8	7	11	7	10	1.00			
25	3	7	8	7	10	11	10	7	9	8	1.00			
26	3	5	5	8	6	6	8	4	8	4	1.00			
27	3	8	7	7	7	2	8	8	6	11	1.00			
28	3	9	2	6	8	10	7	4	6	6	1.00			
29	3	14	8	8	8	12	8	8	8	9	1.00			
30	3	8	8	7	10	11	11	3	9	8	1.00			
31	3	11	3	4	7	8	8	10	4	10	1.00			
32	3	10	7	9	8	10	8	10	5	11	1.00			
33	3	4	11	9	6	13	11	9	4	7	1.00			
34	3	12	12	7	8	4	8	9	9	8	1.00			
35	3	11	9	7	8	10	11	9	11	9	1.00			
36	3	8	5	9	1	7	9	11	8	8	1.00			

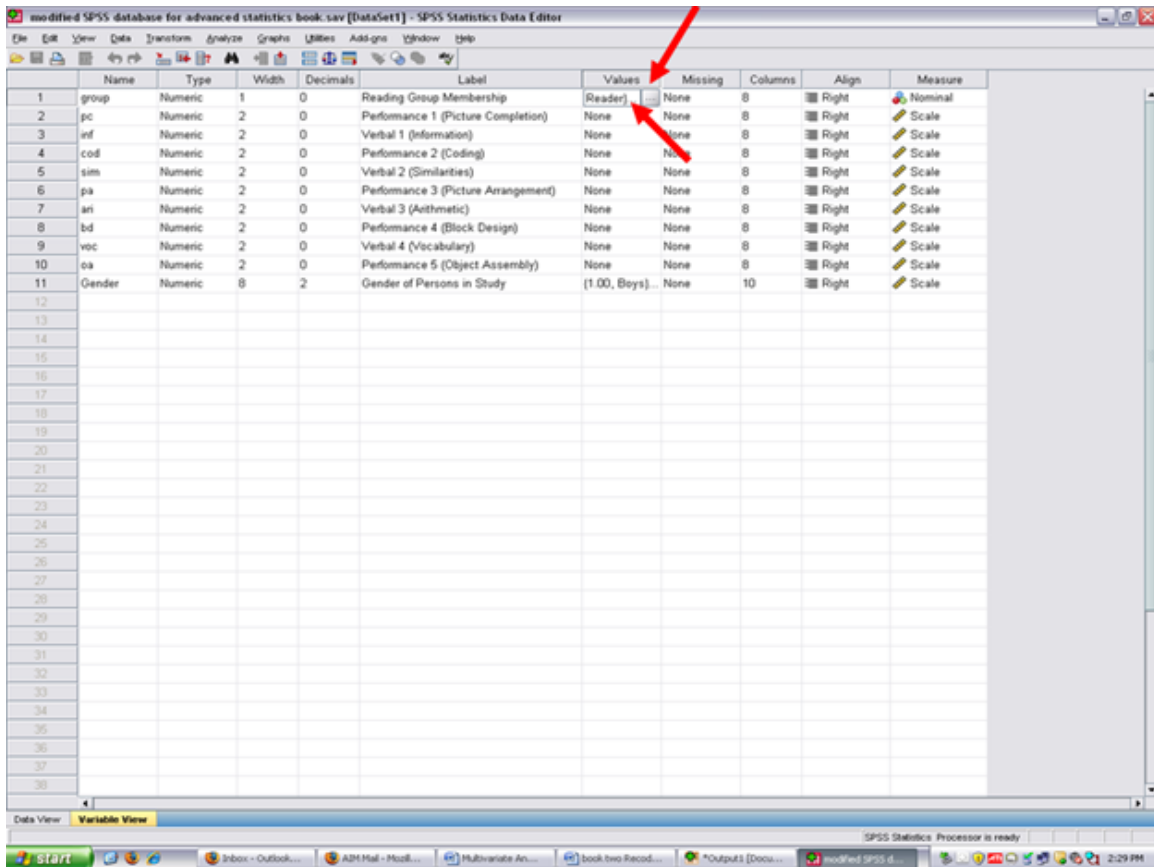
After clicking on Variable View, the following screen will appear: Arrows have been placed toward the name of the independent or grouping variable; toward the label assigned to this variable; and then toward the values of each group.

⁹<http://cnx.org/content/m40728/latest/10.3.png/image>



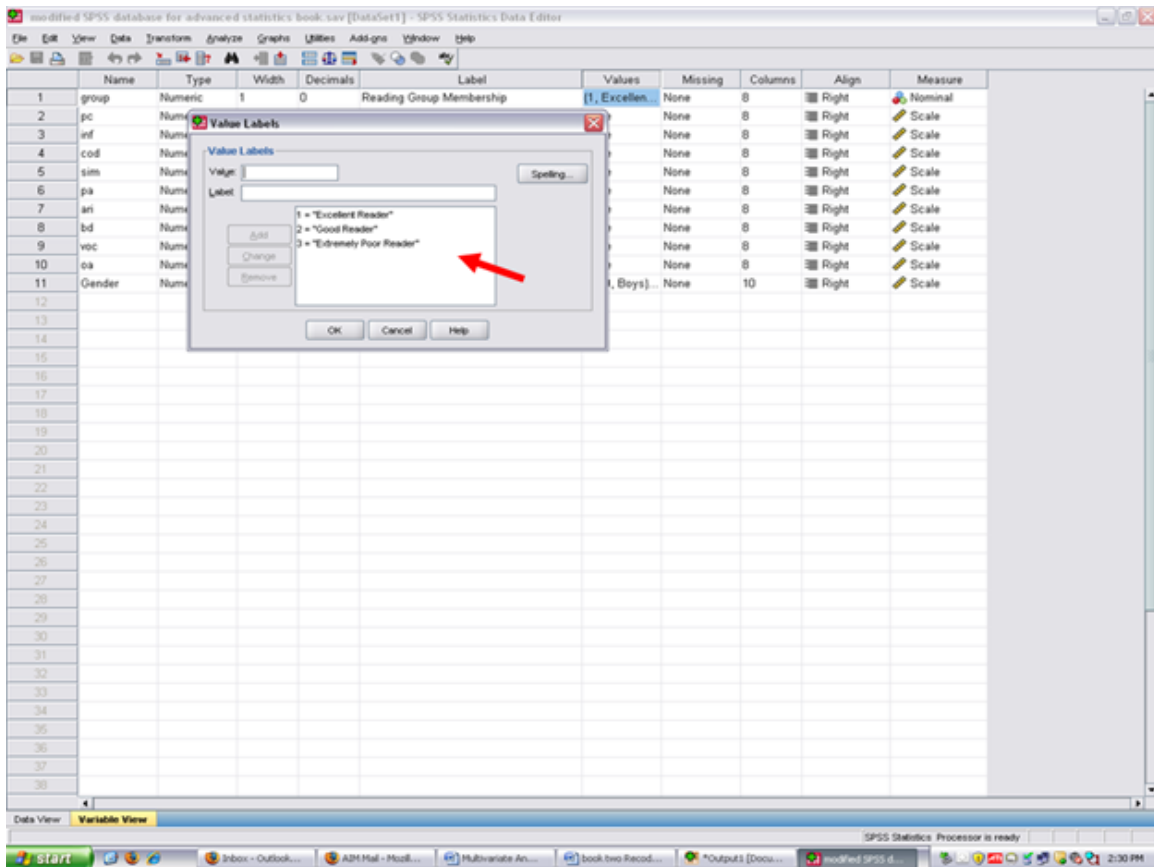
Now click on the cell for Reading Group Membership. Three dots will appear, indicating that another screen is beneath this one.

¹⁰<http://cnx.org/content/m40728/latest/10.4.png/image>



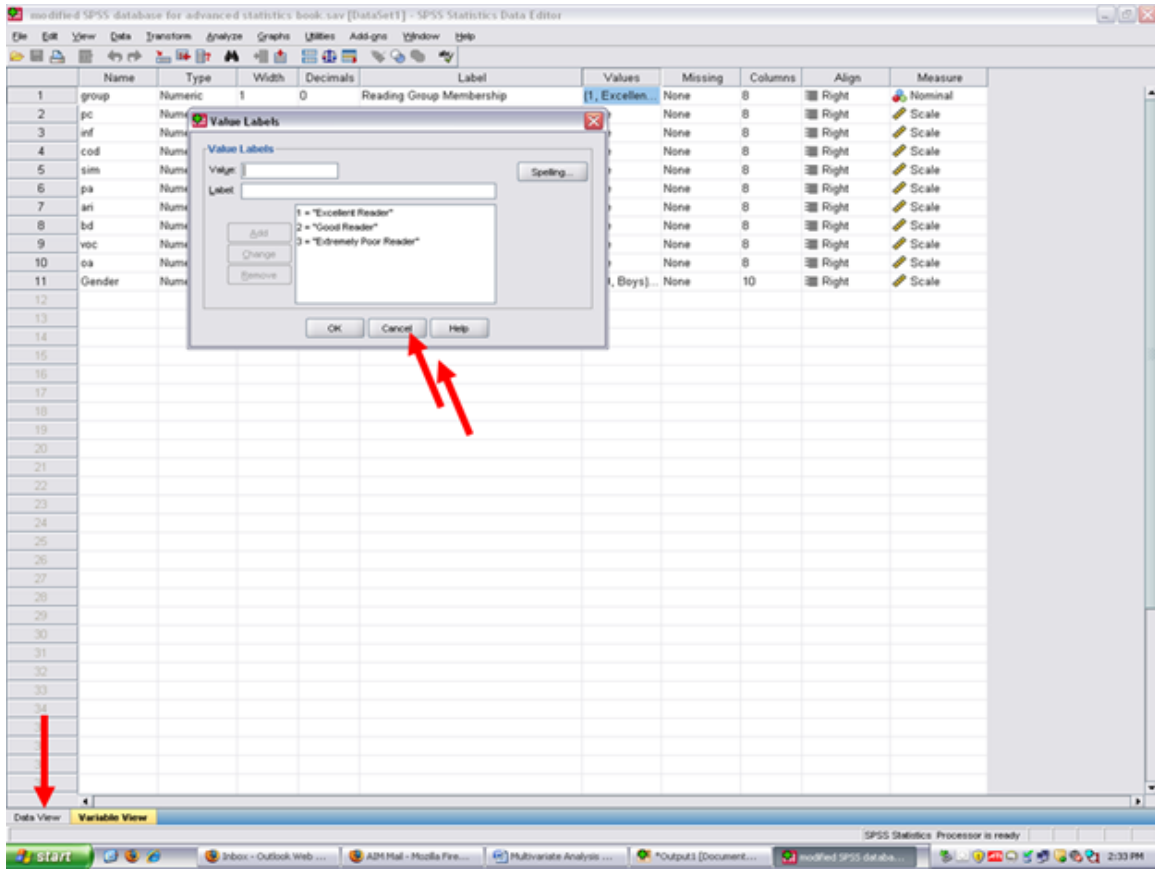
Click on the three dots shown in the screen above and the value and name of each of the reading groups will be shown. Our three groups are Excellent Reader (1), Good Reader (2), and Extremely Poor Reader (3).

¹¹<http://cnx.org/content/m40728/latest/10.5.png/image>



Now that you know what your independent (grouping) variable is and your dependent variables are, it is time to determine the extent to which the dependent variable data are normally distributed. Cancel out of the screen above. Then click on data view.

¹²<http://cnx.org/content/m40728/latest/10.6.png/image>



Your screen should now look like the one below.

¹³<http://cnx.org/content/m40728/latest/10.7.png/image>

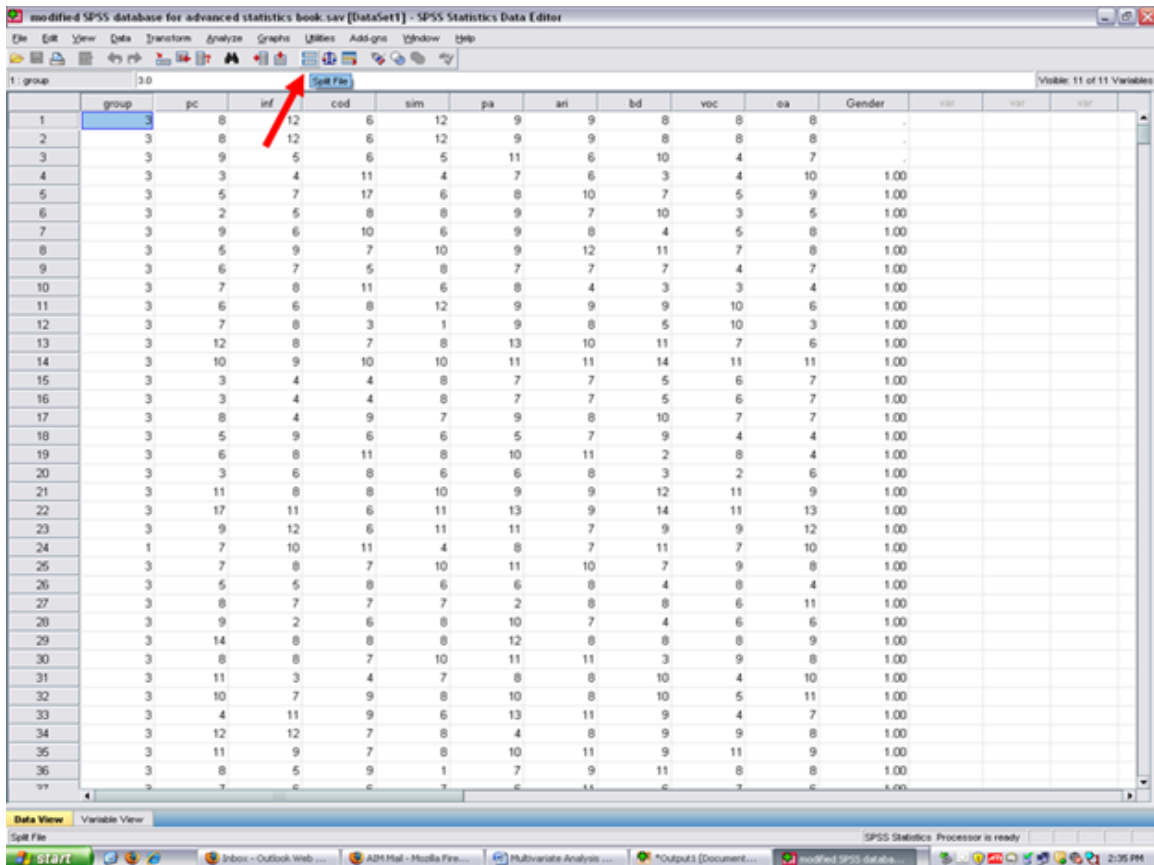
	group	pc	inf	cod	sim	pa	an	bd	voc	oa	Gender	vis1	vis2	vis3
1	3	8	12	6	12	9	9	8	8	8	.			
2	3	8	12	6	12	9	9	8	8	8	.			
3	3	9	5	6	5	11	6	10	4	7	.			
4	3	3	4	11	4	7	6	3	4	10	1.00			
5	3	5	7	17	6	8	10	7	5	9	1.00			
6	3	2	5	8	8	9	7	10	3	5	1.00			
7	3	9	6	10	6	9	8	4	5	8	1.00			
8	3	5	9	7	10	9	12	11	7	8	1.00			
9	3	6	7	5	8	7	7	7	4	7	1.00			
10	3	7	8	11	6	8	4	3	3	4	1.00			
11	3	6	6	8	12	9	9	9	10	6	1.00			
12	3	7	8	3	1	9	8	5	10	3	1.00			
13	3	12	8	7	8	13	10	11	7	6	1.00			
14	3	10	9	10	10	11	11	14	11	11	1.00			
15	3	3	4	4	8	7	7	5	6	7	1.00			
16	3	3	4	4	8	7	7	5	6	7	1.00			
17	3	8	4	9	7	9	8	10	7	7	1.00			
18	3	5	9	6	6	5	7	9	4	4	1.00			
19	3	6	8	11	8	10	11	2	8	4	1.00			
20	3	3	6	8	6	6	8	3	2	6	1.00			
21	3	11	8	8	10	9	9	12	11	9	1.00			
22	3	17	11	6	11	13	9	14	11	13	1.00			
23	3	9	12	6	11	11	7	9	9	12	1.00			
24	1	7	10	11	4	8	7	11	7	10	1.00			
25	3	7	8	7	10	11	10	7	9	8	1.00			
26	3	5	5	8	6	6	8	4	8	4	1.00			
27	3	8	7	7	7	2	8	8	6	11	1.00			
28	3	9	2	6	8	10	7	4	6	6	1.00			
29	3	14	8	8	8	12	8	8	8	9	1.00			
30	3	8	8	7	10	11	11	3	9	8	1.00			
31	3	11	3	4	7	8	8	10	4	10	1.00			
32	3	10	7	9	8	10	8	10	5	11	1.00			
33	3	4	11	9	6	13	11	9	4	7	1.00			
34	3	12	12	7	8	4	8	9	9	8	1.00			
35	3	11	9	7	8	10	11	9	11	9	1.00			
36	3	8	5	9	1	7	9	11	8	8	1.00			

14

Check for Skewness and Kurtosis values falling within/without the parameters of normality (-3 to +3). To do so, you need to obtain these values for each of your three groups for each of your four dependent variables.

✓ Split your file on the basis on your independent variable/fixed factor/grouping variable. To do so, click on the icon next to the scales. Holding your cursor on it will reveal Split File, as shown below.

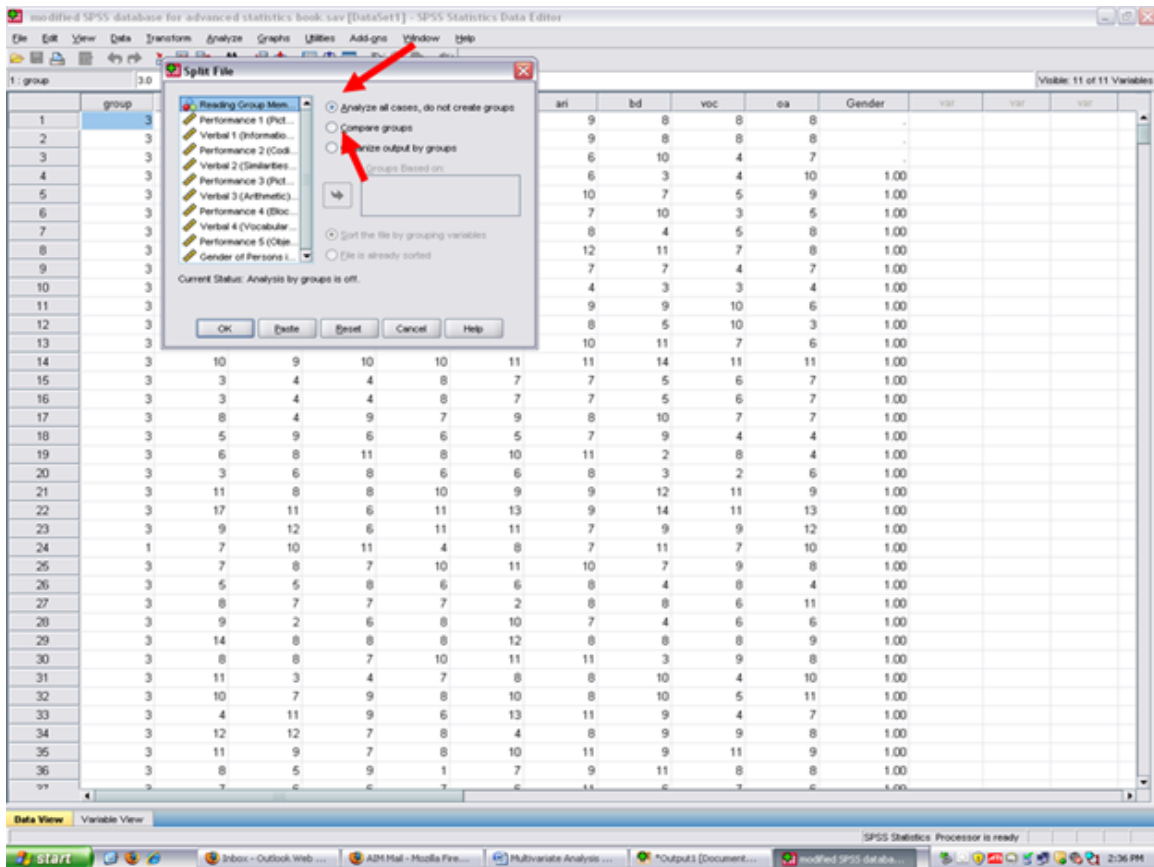
¹⁴<http://cnx.org/content/m40728/latest/10.8.png/image>



15

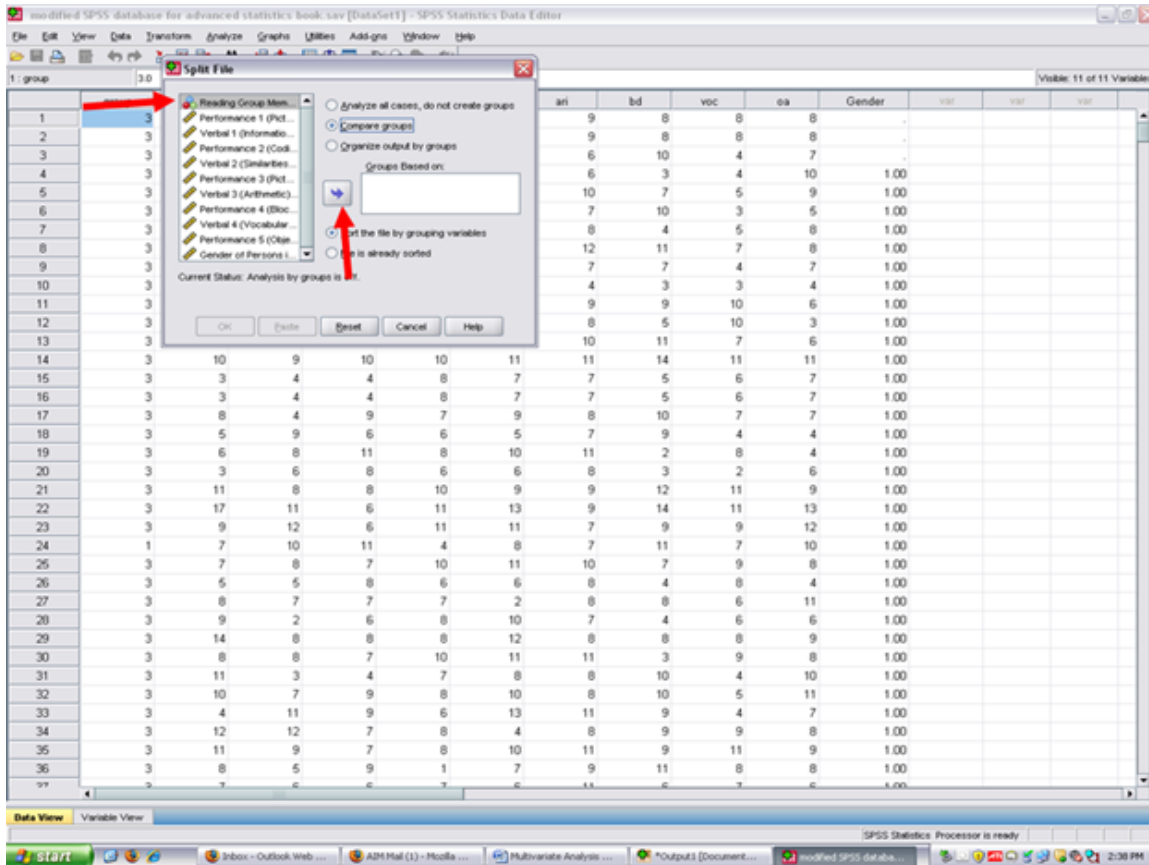
After clicking on the split file icon, the following screen will appear: You will note that the Analyze all cases, do not create groups is checked. This value is the default for SPSS as all cases are analyzed, unless otherwise specified. To obtain measures of normality for each of the three reading groups, the Compare groups button will need to be clicked.

¹⁵<http://cnx.org/content/m40728/latest/10.9.png/image>



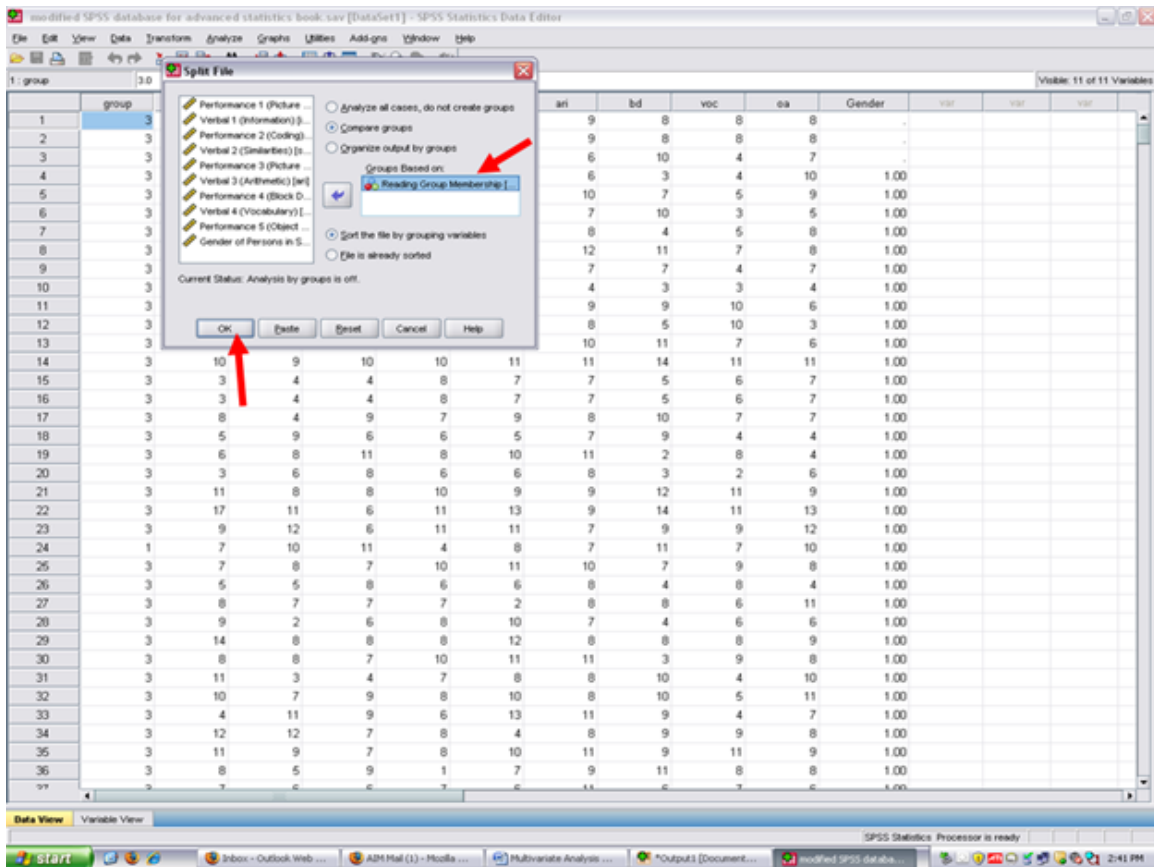
The screen below will appear after the Compare groups button has been clicked. Note that the Groups Based on rectangle has now become active. The independent (grouping) variable should be highlighted, as it already is, and then moved to the Groups Based on cell. After highlighting the Reading Group Membership variable, then click on the arrow below.

¹⁶<http://cnx.org/content/m40728/latest/10.10.png/image>



After clicking on the arrow, your independent variable of Reading Group Membership is now in the Groups Based on cell. Now click on OK.

¹⁷<http://cnx.org/content/m40728/latest/10.11.png/image>



18

Now all SPSS analyses will be calculated separately for each of the three reading groups.

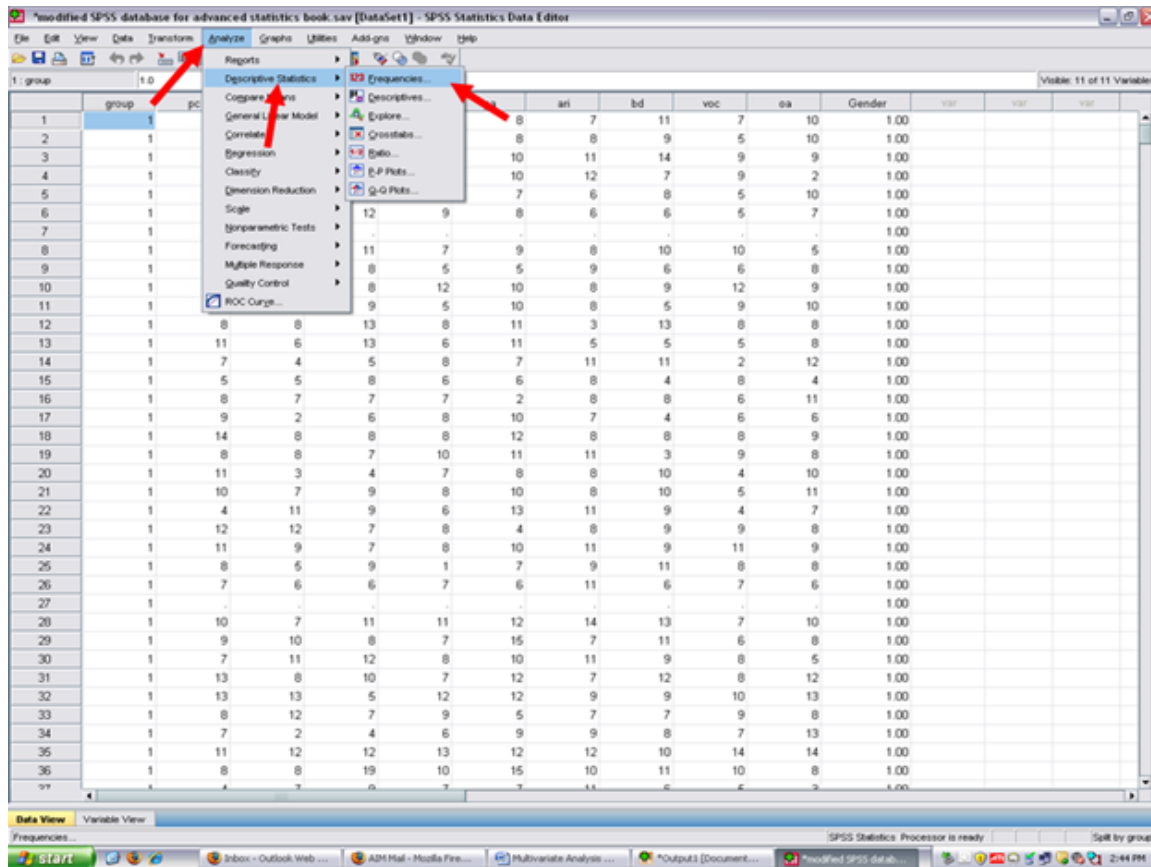
¹⁸<http://cnx.org/content/m40728/latest/10.12.png/image>

The screenshot shows the SPSS Statistics Data Editor window. The title bar reads "modified SPSS database for advanced statistics book.sav [DataSet1] - SPSS Statistics Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, and Help. The toolbar contains various icons for file operations and analysis. The main window displays a data grid with 36 rows and 11 columns. The columns are labeled: group, pc, inf, cod, sim, pa, an, bd, voc, oa, and Gender. The Gender column contains the value 1.00 for all rows. The bottom status bar indicates "SPSS Statistics Processor is ready" and "Split by group".

group	pc	inf	cod	sim	pa	an	bd	voc	oa	Gender
1	7	10	11	4	8	7	11	7	10	1.00
2	9	6	10	7	8	8	9	5	10	1.00
3	8	6	8	8	10	11	14	9	9	1.00
4	2	7	14	8	10	12	7	9	2	1.00
5	6	4	6	4	7	6	8	5	10	1.00
6	6	9	12	9	8	6	6	5	7	1.00
7	1.00
8	10	11	11	7	9	8	10	10	5	1.00
9	9	4	8	5	5	9	6	6	8	1.00
10	9	13	8	12	10	8	9	12	9	1.00
11	7	8	9	5	10	8	5	9	10	1.00
12	8	8	13	8	11	3	13	8	8	1.00
13	11	6	13	6	11	5	5	5	8	1.00
14	7	4	5	8	7	11	11	2	12	1.00
15	5	5	8	6	6	8	4	8	4	1.00
16	8	7	7	7	2	8	8	6	11	1.00
17	9	2	6	8	10	7	4	6	6	1.00
18	14	8	8	8	12	8	8	8	9	1.00
19	8	8	7	10	11	11	3	9	8	1.00
20	11	3	4	7	8	8	10	4	10	1.00
21	10	7	9	8	10	8	10	5	11	1.00
22	4	11	9	6	13	11	9	4	7	1.00
23	12	12	7	8	4	8	9	9	8	1.00
24	11	9	7	8	10	11	9	11	9	1.00
25	8	5	9	1	7	9	11	8	8	1.00
26	7	6	6	7	6	11	6	7	6	1.00
27	1.00
28	10	7	11	11	12	14	13	7	10	1.00
29	9	10	8	7	15	7	11	6	8	1.00
30	7	11	12	8	10	11	9	8	5	1.00
31	13	8	10	7	12	7	12	8	12	1.00
32	13	13	5	12	12	9	9	10	13	1.00
33	8	12	7	9	5	7	7	9	8	1.00
34	7	2	4	6	9	9	8	7	13	1.00
35	11	12	12	13	12	12	10	14	14	1.00
36	8	8	19	10	15	10	11	10	8	1.00
37	1.00

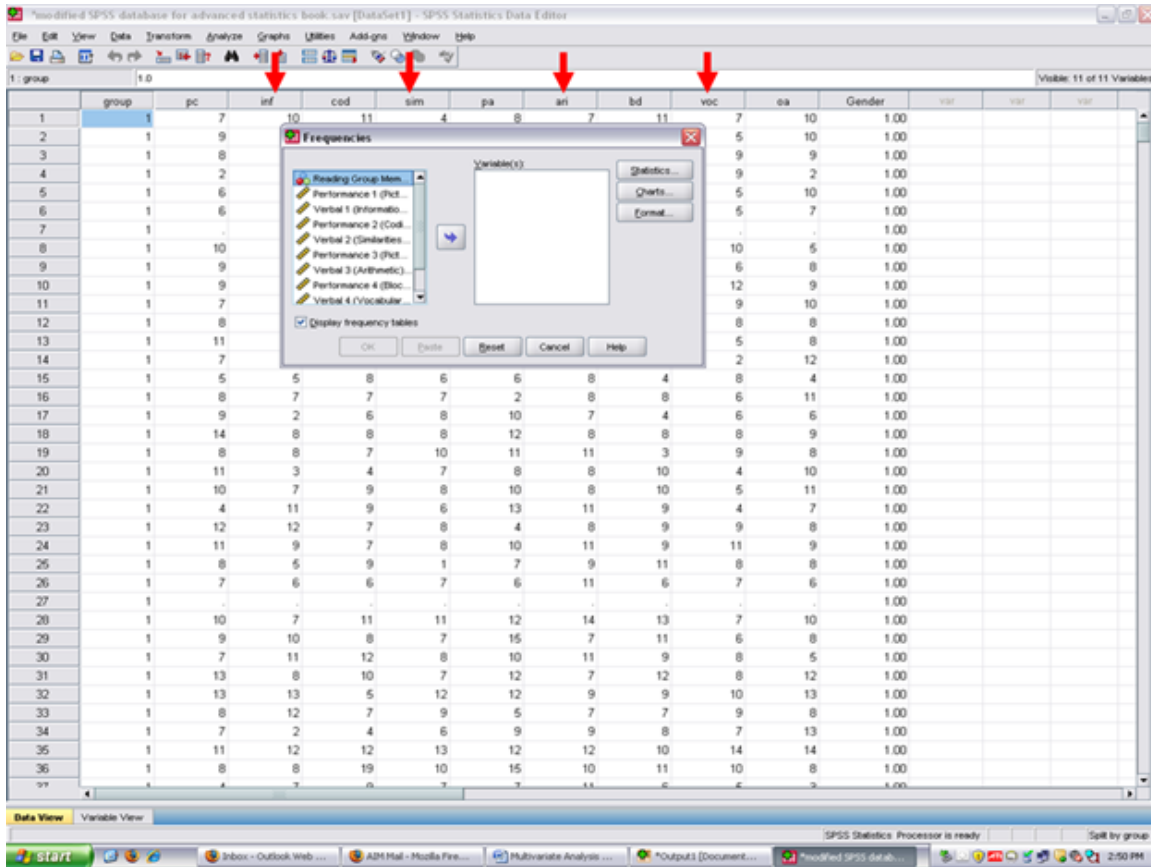
- ✓ Analyze
- * Descriptive Statistics
- * Frequencies

¹⁹<http://cnx.org/content/m40728/latest/10.13.png/image>



After clicking on frequencies, the following screen should now be present. Remember that the dependent variables in this example are Verbal 1 through Verbal 4 (i.e., inf, ari, voc)

²⁰<http://cnx.org/content/m40728/latest/10.14.png/image>



21

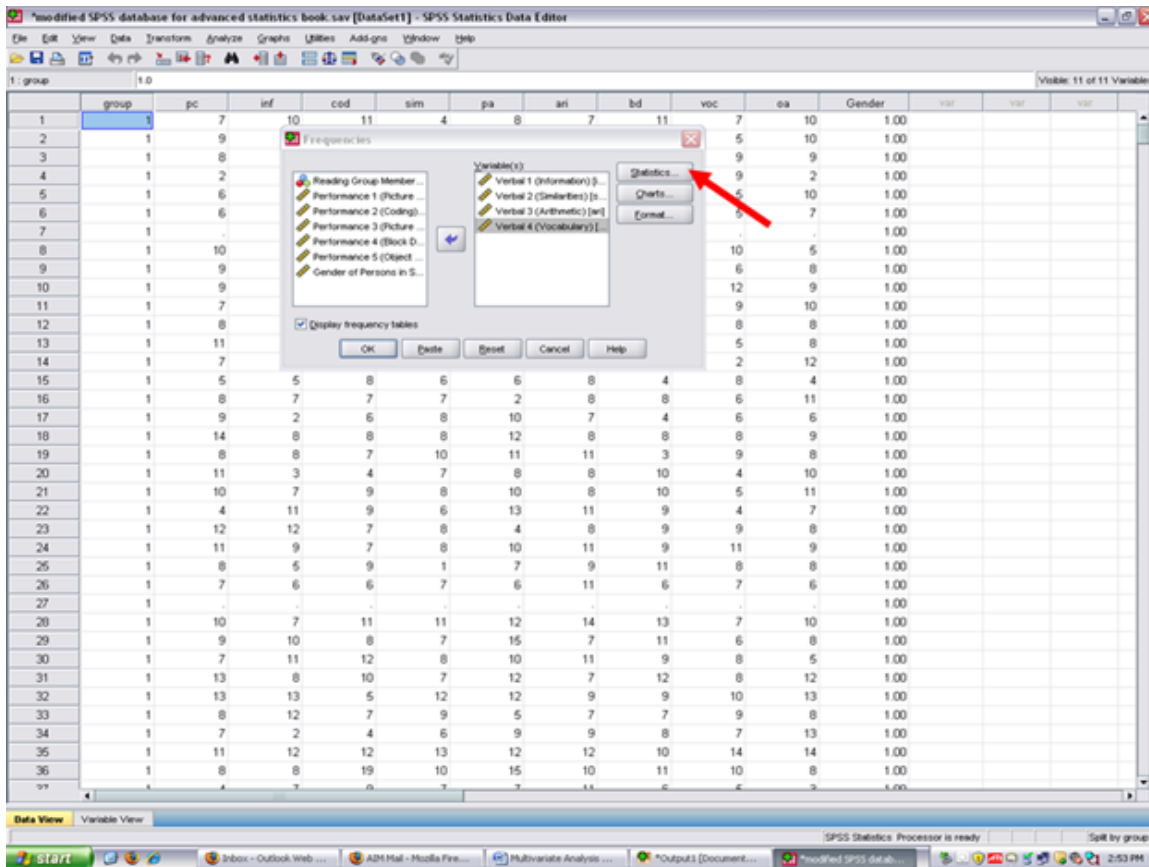
✓ Move over the dependent variables. In this example, highlight each of the dependent variables one at a time and click on the right arrow.

²¹ <http://cnx.org/content/m40728/latest/10.15.png/image>

The screenshot shows the SPSS Statistics Data Editor window. A 'Frequencies' dialog box is open, displaying a list of variables on the left and a 'Variable(s)' box on the right. The variables listed are: Reading Group Mem..., Performance 1 (Flt..., Verbal 1 (Informatio..., Performance 2 (Cook..., Verbal 2 (Similes..., Performance 3 (Flt..., Verbal 3 (Arithmetic), Performance 4 (Bloc..., and Verbal 4 (Vocabular... The 'Verbal 1 (Informatio...', 'Performance 2 (Cook...', 'Verbal 3 (Arithmetic)', and 'Performance 4 (Bloc...' variables are selected. A red arrow points from the 'Statistics...' button in the dialog box to the 'Variable(s)' box. Another red arrow points from the 'Display frequency tables' checkbox to the 'Display frequency tables' checkbox. The background shows a data table with columns for 'group', 'pc', 'inf', 'cod', 'sim', 'pa', 'an', 'bd', 'voc', 'oa', 'Gender', and three 'vis' columns.

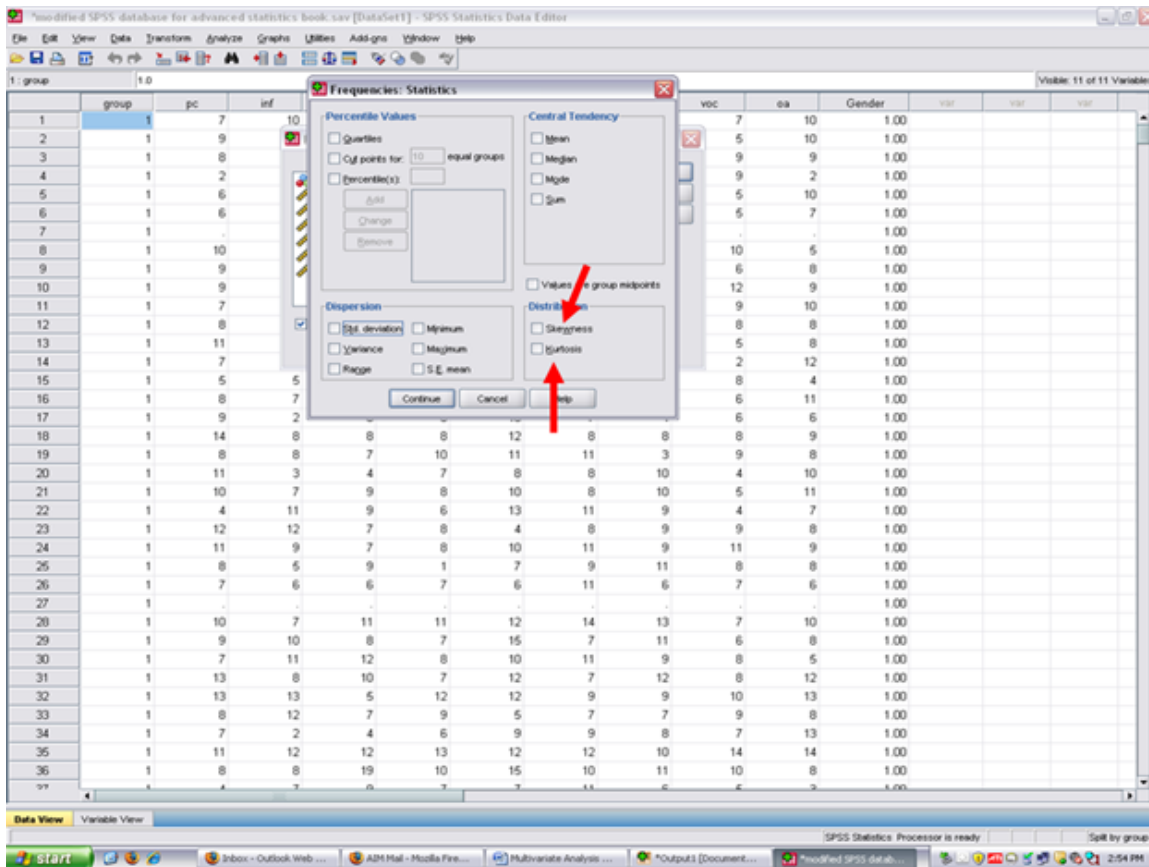
When all four dependent variables are in the Variable cell, then click on Statistics.

²²<http://cnx.org/content/m40728/latest/10.16.png/image>



After clicking on statistics, the screen below will appear. Note that no statistics items are checked. For purposes of this example, only the skewness and kurtosis items will be checked. Although readers may obtain descriptive statistics at this screen, the MANOVA procedure itself can be used to obtain that information.

²³<http://cnx.org/content/m40728/latest/10.17.png/image>



24

Though previously discussed in the steps and screenshots chapters for the basic statistical procedures, readers may find the following information helpful in understanding skewness and kurtosis and their importance for conducting statistical procedures.

* **Skewness** [Note. Skewness refers to the extent to which the data are normally distributed around the mean. Skewed data involve having either mostly high scores with a few low ones or having mostly low scores with a few high ones.] Readers are referred to the following sources for a more detailed definition of skewness: http://www.statistics.com/index.php?page=glossary&term_id=356²⁵ and <http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb>²⁶

To standardize the skewness value so that its value can be constant across datasets and across studies, the following calculation must be made: Take the skewness value from the SPSS output and divide it by the Std. error of skewness. If the resulting calculation is within -3 to +3, then the skewness of the dataset is within the range of normality (Onwuegbuzie & Daniel, 2002). If the resulting calculation is outside of this +/-3 range, the dataset is not normally distributed.

* **Kurtosis** [Note. Kurtosis also refers to the extent to which the data are normally distributed around the mean. This time, the data are piled up higher than normal around the mean or piled up higher than normal at the ends of the distribution.] Readers are referred to the following sources for a more detailed definition of kurtosis: http://www.statistics.com/index.php?page=glossary&term_id=326²⁷ and

²⁴ <http://cnx.org/content/m40728/latest/10.18.png/image>

²⁵ http://www.statistics.com/index.php?page=glossary&term_id=356

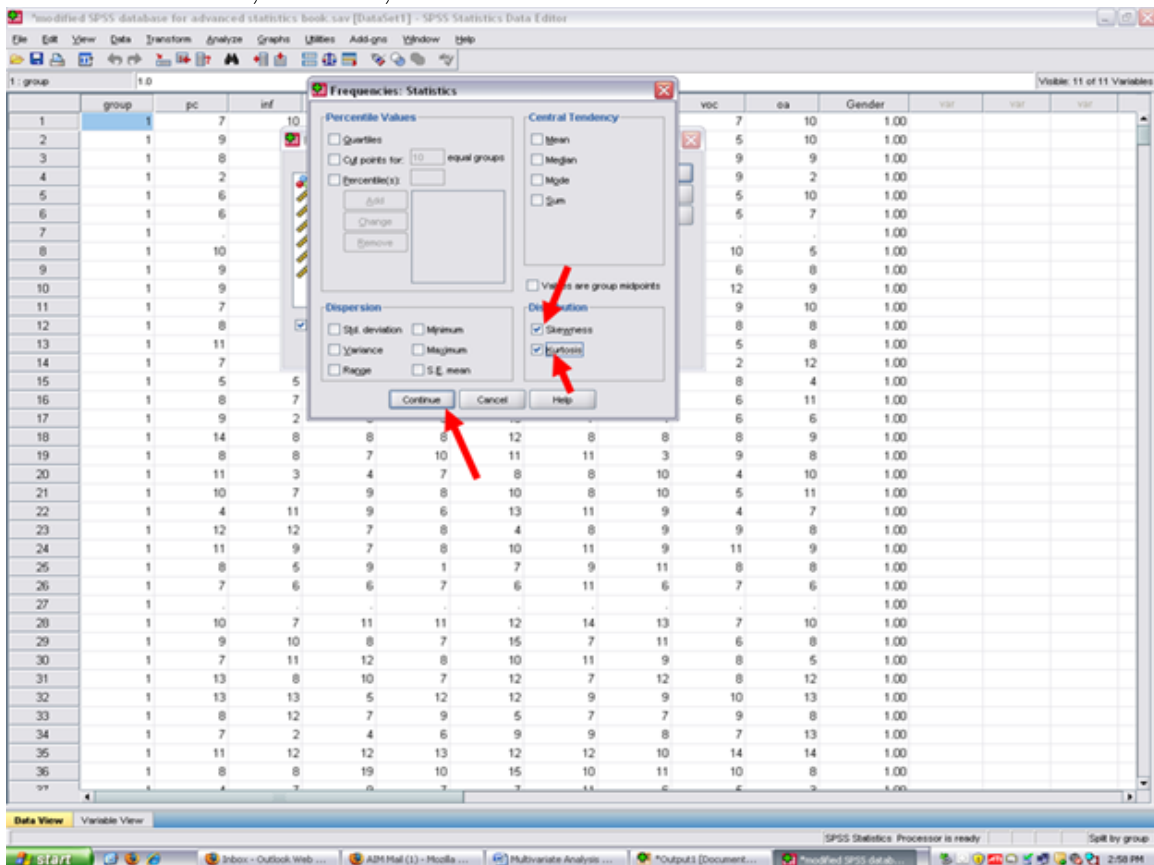
²⁶ <http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb>

²⁷ http://www.statistics.com/index.php?page=glossary&term_id=326

<http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb>²⁸

To standardize the kurtosis value so that its value can be constant across datasets and across studies, the following calculation must be made: Take the kurtosis value from the SPSS output and divide it by the Std. error of kurtosis. If the resulting calculation is within -3 to $+3$, then the kurtosis of the dataset is within the range of normality (Onwuegbuzie & Daniel, 2002). If the resulting calculation is outside of this ± 3 range, the dataset is not normally distributed. Then the kurtosis of the dataset is within the range of normality (Onwuegbuzie & Daniel, 2002). If the resulting calculation is outside of this ± 3 range, the dataset is not normally distributed.

Click on Skewness, Kurtosis, and Continue



29

Now click on OK.

²⁸<http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb>

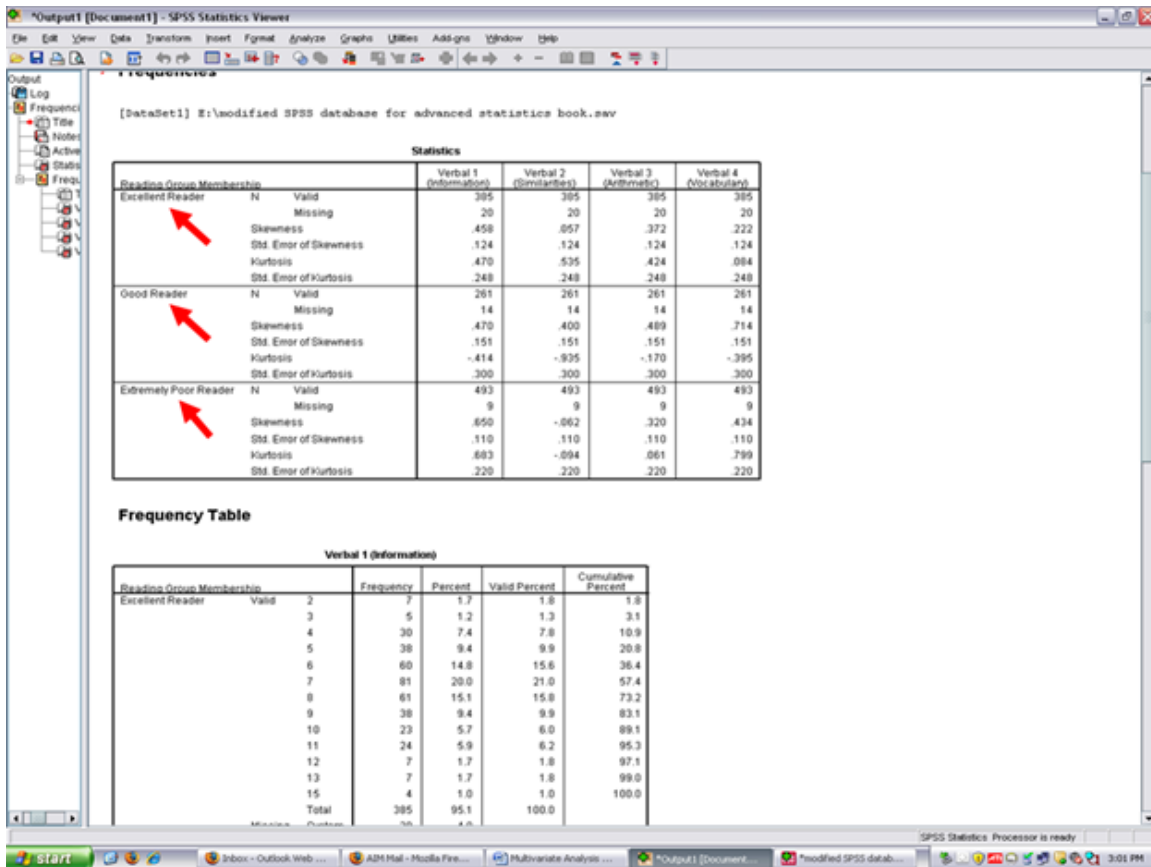
²⁹<http://cnx.org/content/m40728/latest/10.19.png/image>

The screenshot shows the SPSS Statistics Data Editor window with a data table and a 'Frequencies' dialog box. The data table has columns: group, pc, inf, cod, sim, pa, an, bd, voc, oa, Gender, and three unlabeled columns. The 'Frequencies' dialog box is open, showing a list of variables on the left and a 'Variable(s)' list on the right. The 'Display frequency tables' checkbox is checked. A red arrow points to this checkbox. The taskbar at the bottom shows the Start button and several open applications, including 'Inbox - Outlook Web...', 'AIM Mail - Mozilla Fire...', 'Multivariate Analysis...', and 'Output1 [Document...]'.

30

If SPSS does not send you to the Output screen, click on the Output button at the bottom of your screen so that you can view the results of the statistics you just had SPSS calculate for you.

³⁰<http://cnx.org/content/m40728/latest/10.20.png/image>



In the columns to the right of the three reading groups are the skewness and kurtosis values for the four dependent variables.

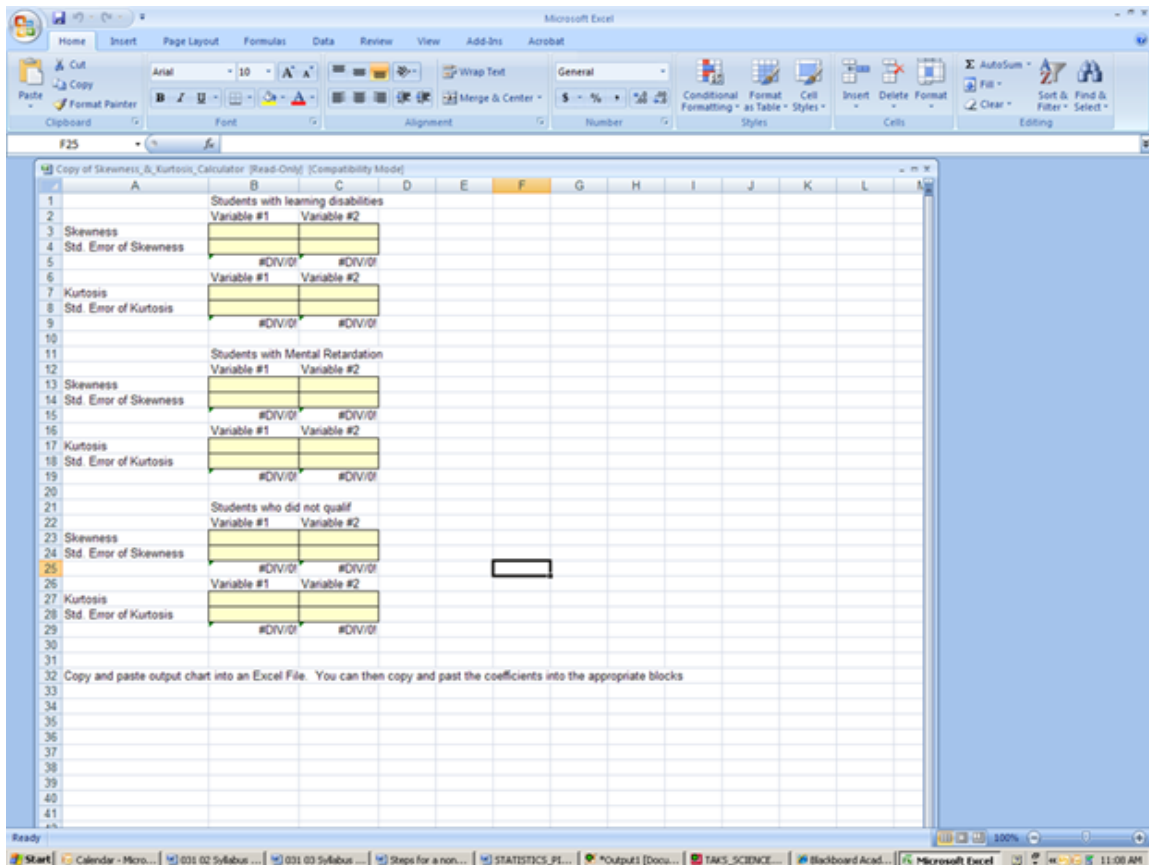
³²<http://cnx.org/content/m40728/latest/10.22.png/image>

The screenshot shows the SPSS Statistics Viewer window. The main content is a 'Frequencies' table. The table has a header row for 'Statistics' and four columns for 'Verbal 1 (Information)', 'Verbal 2 (Similarity)', 'Verbal 3 (Arithmetic)', and 'Verbal 4 (Vocabulary)'. The rows are grouped by 'Reading Group Membership' (Excellent Reader, Good Reader, Extremely Poor Reader) and then by 'N Valid', 'N Missing', 'Skewness', 'Std. Error of Skewness', 'Kurtosis', and 'Std. Error of Kurtosis'. Red arrows point to the 'Statistics' header and the first three 'Verbal' columns.

Reading Group Membership		Verbal 1 (Information)	Verbal 2 (Similarity)	Verbal 3 (Arithmetic)	Verbal 4 (Vocabulary)
Excellent Reader	N Valid	395	395	395	395
	N Missing	20	20	20	20
	Skewness	.458	.657	.372	.222
	Std. Error of Skewness	.124	.124	.124	.124
	Kurtosis	.470	.535	.424	.084
	Std. Error of Kurtosis	.248	.248	.248	.248
Good Reader	N Valid	261	261	261	261
	N Missing	14	14	14	14
	Skewness	.470	.400	.489	.714
	Std. Error of Skewness	.151	.151	.151	.151
	Kurtosis	-.414	-.935	-.170	-.395
	Std. Error of Kurtosis	.300	.300	.300	.300
Extremely Poor Reader	N Valid	493	493	493	493
	N Missing	9	9	9	9
	Skewness	.850	-.062	.320	.434
	Std. Error of Skewness	.110	.110	.110	.110
	Kurtosis	.893	-.094	.061	.799
	Std. Error of Kurtosis	.220	.220	.220	.220

Using the skewness and kurtosis value for each dependent variable above for each of the three groups above, type them in one at a time into the standardized coefficients calculator. To be regarded as being normally distributed, the coefficient should be within -3 to $+3$ (Onwuegbuzie & Daniel, 2002).

³³<http://cnx.org/content/m40728/latest/10.23.png/image>



34

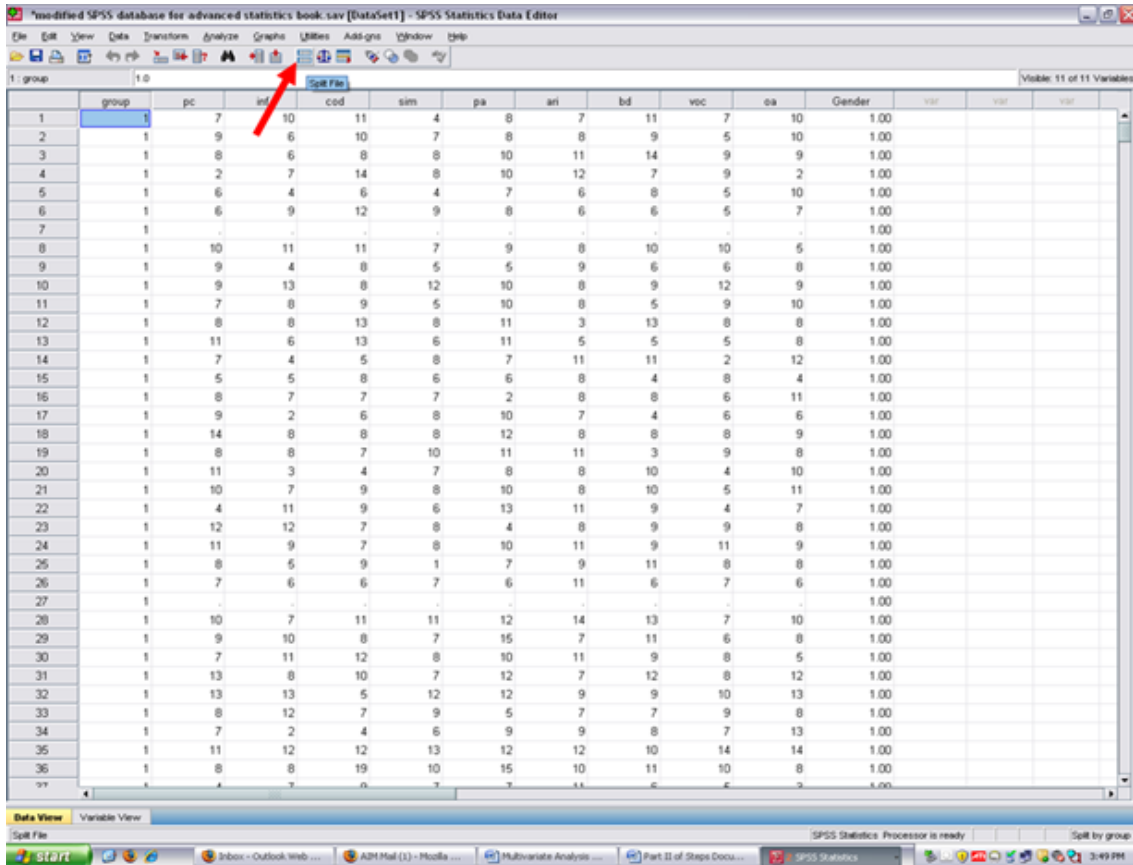
After calculating the standardized coefficients, record the number that are within ± 3 and the number that are outside of these boundaries. This information needs to be discussed in your results section where you document your checks of the assumptions underlying this particular statistical technique.

Now, before conducting the MANOVA to answer our research question, this dataset must be put back together. Remember that the dataset is currently split. As you were in the SPSS Output screen, make sure that you go back to the Data screen before continuing.

To unsplit the file,

✓ Split Files (the icon next to the scales)

³⁴<http://cnx.org/content/m40728/latest/10.24.png/image>



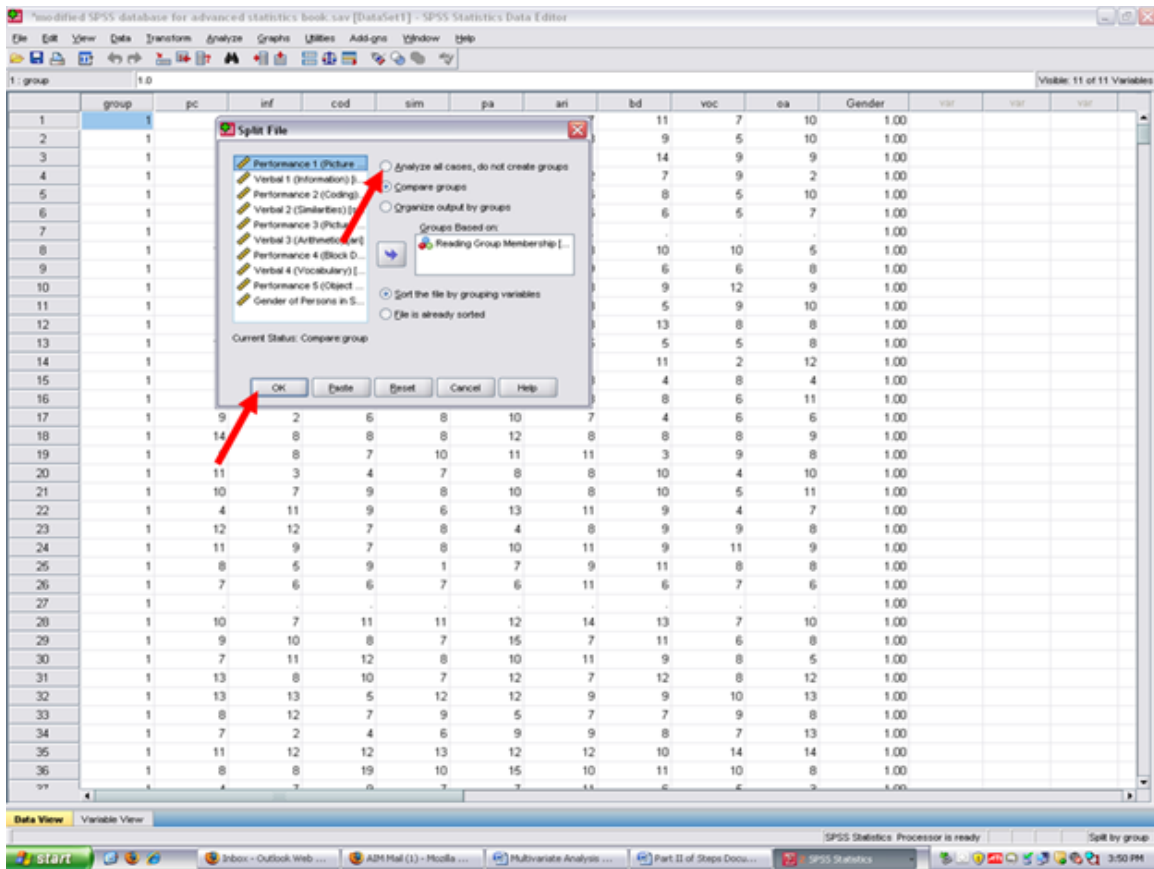
35

Clicking on split files will reveal this screen. Although two ways exist in which to have SPSS analyze all cases, the easiest is simply to click on:

✓ Analyze all cases, do not create groups

Then, ✓ OK

³⁵<http://cnx.org/content/m40728/latest/10.25.png/image>



After clicking on OK, your screen should now look like the one below.

³⁶<http://cnx.org/content/m40728/latest/10.26.png/image>

group	pc	inf	cod	sim	pa	an	bd	voc	oa	Gender			
1	7	10	11	4	8	7	11	7	10	1.00			
2	9	6	10	7	8	8	9	5	10	1.00			
3	8	6	8	8	10	11	14	9	9	1.00			
4	2	7	14	8	10	12	7	9	2	1.00			
5	6	4	6	4	7	6	8	5	10	1.00			
6	6	9	12	9	8	6	6	5	7	1.00			
7	1.00			
8	10	11	11	7	9	8	10	10	5	1.00			
9	9	4	8	5	5	9	6	6	8	1.00			
10	9	13	8	12	10	8	9	12	9	1.00			
11	7	8	9	5	10	8	5	9	10	1.00			
12	8	8	13	8	11	3	13	8	8	1.00			
13	11	6	13	6	11	5	5	5	8	1.00			
14	7	4	5	8	7	11	11	2	12	1.00			
15	5	5	8	6	6	8	4	8	4	1.00			
16	8	7	7	7	2	8	8	6	11	1.00			
17	9	2	6	8	10	7	4	6	6	1.00			
18	14	8	8	8	12	8	8	8	9	1.00			
19	8	8	7	10	11	11	3	9	8	1.00			
20	11	3	4	7	8	8	10	4	10	1.00			
21	10	7	9	8	10	8	10	5	11	1.00			
22	4	11	9	6	13	11	9	4	7	1.00			
23	12	12	7	8	4	8	9	9	8	1.00			
24	11	9	7	8	10	11	9	11	9	1.00			
25	8	5	9	1	7	9	11	8	8	1.00			
26	7	6	6	7	6	11	6	7	6	1.00			
27	1.00			
28	10	7	11	11	12	14	13	7	10	1.00			
29	9	10	8	7	15	7	11	6	8	1.00			
30	7	11	12	8	10	11	9	8	5	1.00			
31	13	8	10	7	12	7	12	8	12	1.00			
32	13	13	5	12	12	9	9	10	13	1.00			
33	8	12	7	9	5	7	7	9	8	1.00			
34	7	2	4	6	9	9	8	7	13	1.00			
35	11	12	12	13	12	12	10	14	14	1.00			
36	8	8	19	10	15	10	11	10	8	1.00			
37	4	7	6	7	7	14	6	6	3	1.00			

³⁷ <http://cnx.org/content/m40728/latest/10.27.png/image>

Chapter 12

12. Multivariate Analysis of Variance: Part II¹



NOTE: This chapter is published by NCPEA Press² and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of education administration.

About the Authors

John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website (Writing and Statistical Help³) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

Ana Rojas-LeBouef is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. Dr. LeBoeuf recently completed her doctoral degree in Reading, where she conducted a 16-year analysis of Texas statewide data regarding the achievement gap. Her research interests lie in examining the inequities in achievement among ethnic groups. Dr. Rojas-LeBouef also assists students and faculty in their writing and statistical needs on the Writing and Statistical Help website.

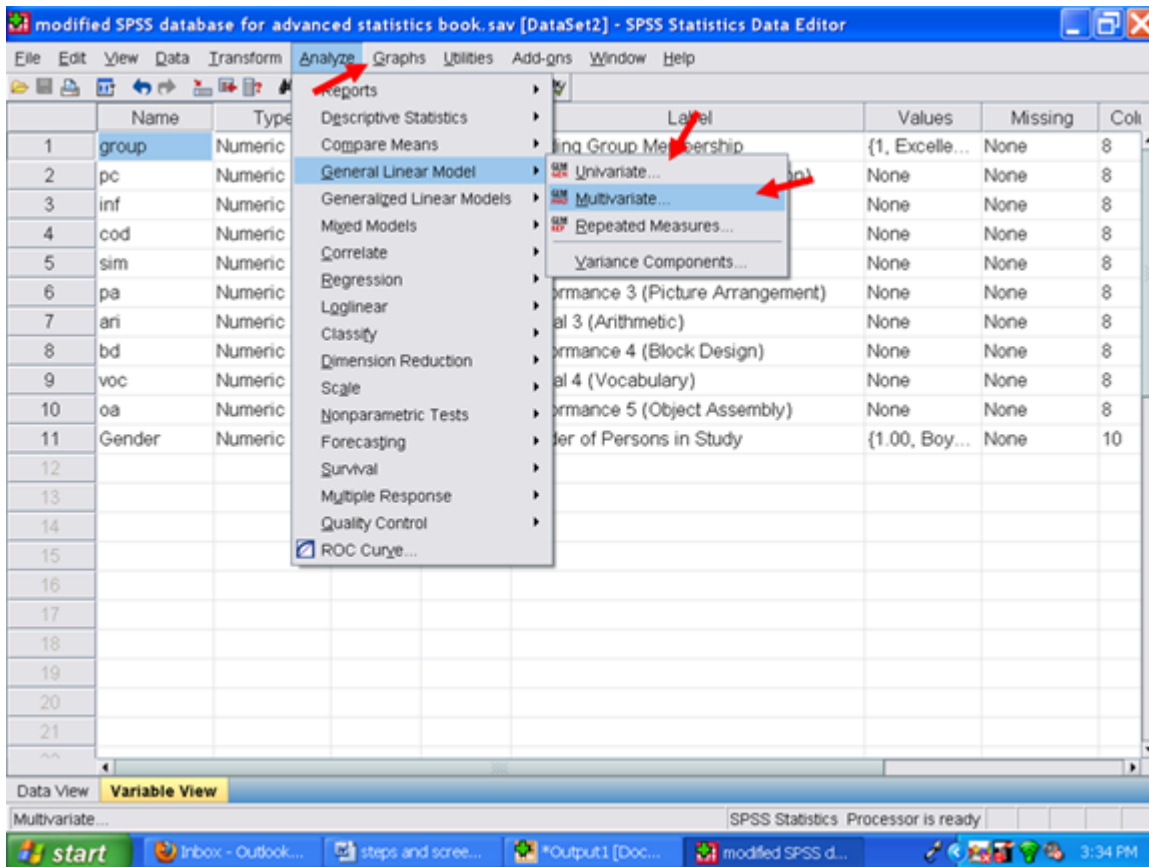
We are now ready to conduct the MANOVA procedure.

- ✓ **Analyze**
- ✓ **General Linear Model**
- ✓ **Multivariate**

¹This content is available online at <<http://cnx.org/content/m40729/1.2/>>.

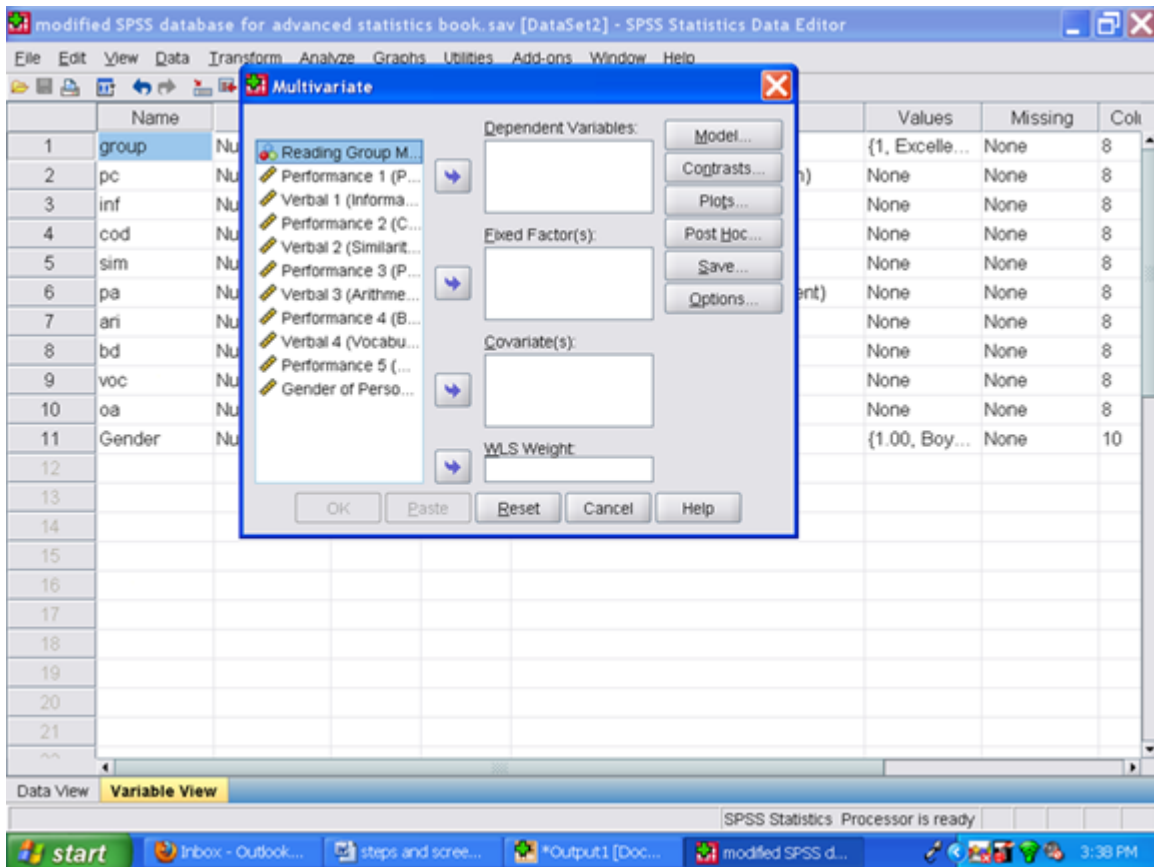
²<http://www.ncpeapublications.org/books.html>

³<http://cnx.org/content/m40729/latest/www.writingandstatisticalhelp>



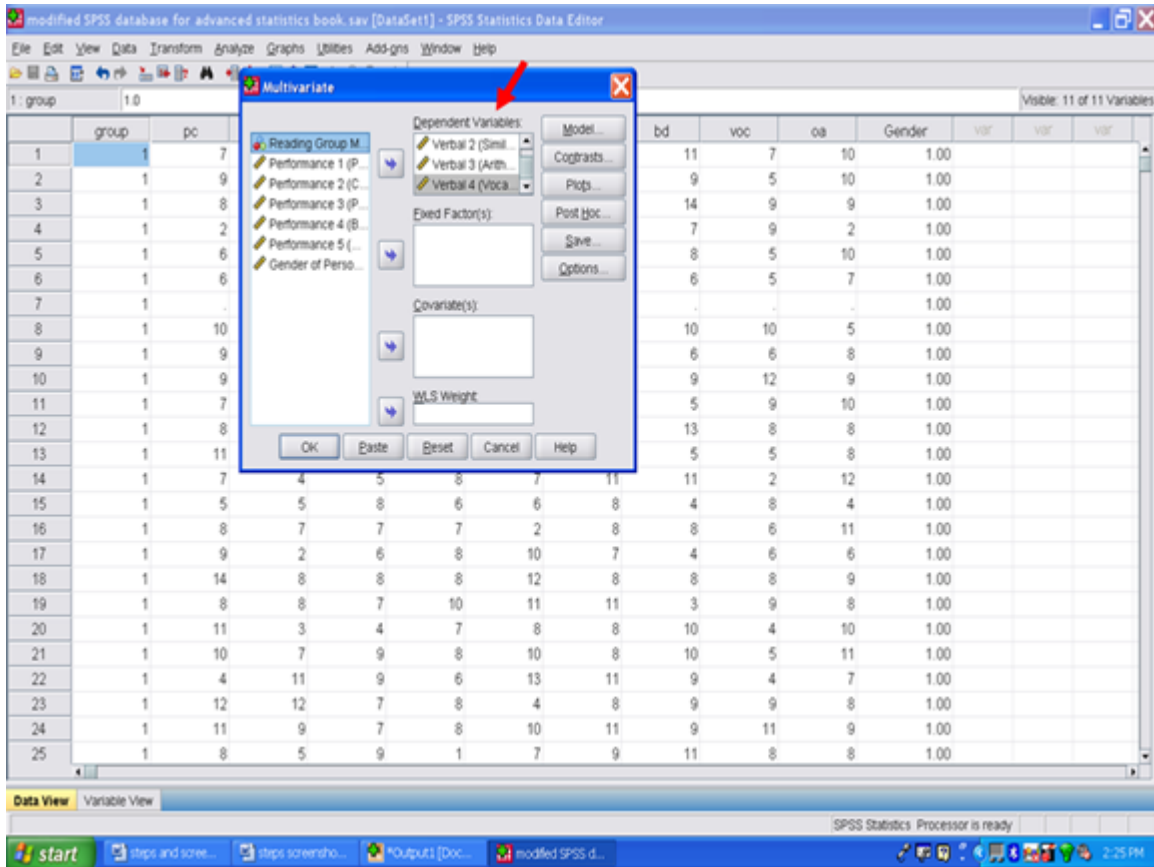
After clicking on Multivariate, the following screen will appear. You will send your dependent variables to the Dependent Variables box and your independent variable to the Fixed Factor box. Remember that in this example, Verbal 1 through Verbal 4 (i.e., Information, Similarities, Arithmetic, and Vocabulary) constitute the dependent variables and Reading Group Membership is the independent variable.

⁴<http://cnx.org/content/m40729/latest/11.1.png/image>



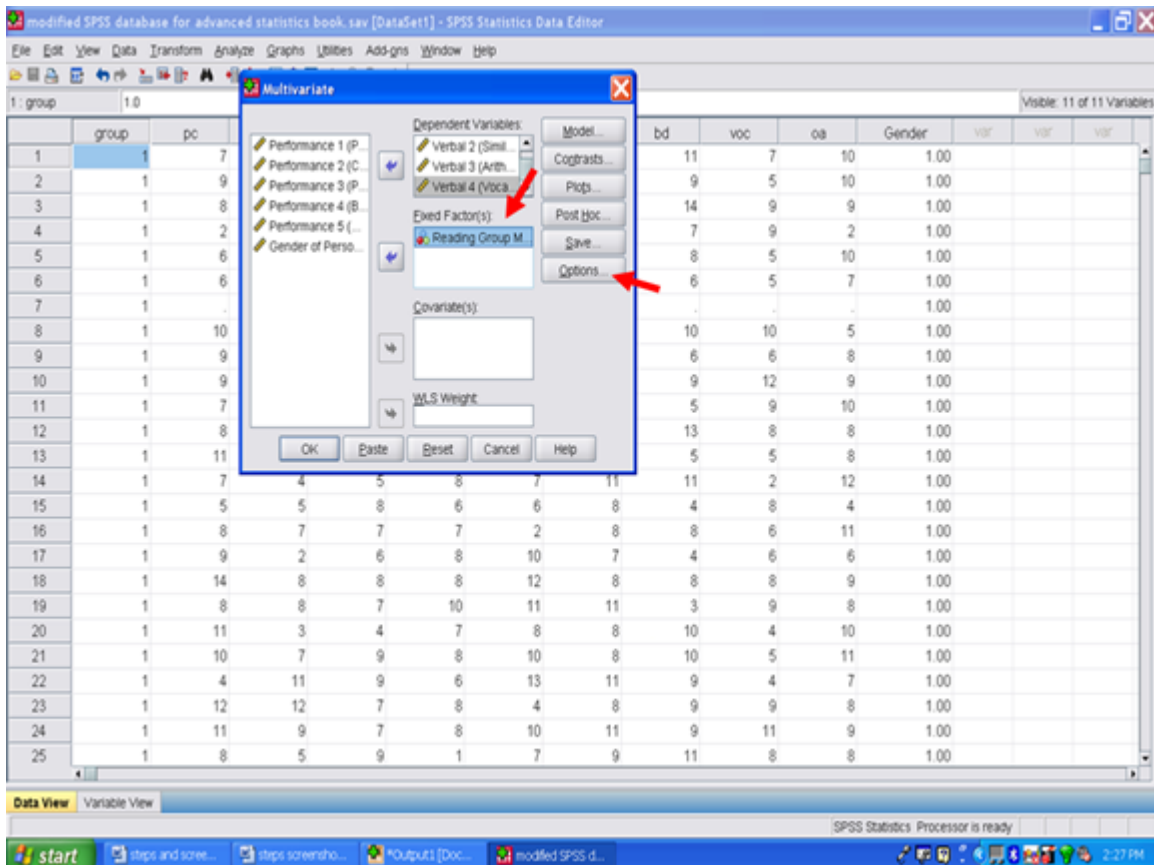
Now send the four dependent variables (i.e., Verbal 1 through Verbal 4) over one at a time to the Dependent Variables screen.

⁵<http://cnx.org/content/m40729/latest/11.2.png/image>



Then send over the independent variable, Reading Group Membership, to the Fixed Factor box. Then click on Options.

⁶<http://cnx.org/content/m40729/latest/113.png/image>



Clicking on options will then give you this screen.

We will use this screen to obtain descriptive statistics of our four dependent variables for each of our three reading groups; to obtain effect size estimates; and to determine the extent to which the assumptions underlying use of the MANOVA procedure are met.

Click on (Overall) and on group and send them to the Display Means for: box.

⁷ <http://cnx.org/content/m40729/latest/11.4.png/image>

The screenshot shows the SPSS 'Multivariate: Options' dialog box. In the 'Estimated Marginal Means' section, 'Factor(s) and Factor Interactions' is set to '(OVERALL)'. The 'Display Means for:' field is empty. Under the 'Display' section, the following options are checked: 'Descriptive statistics', 'Estimates of effect size', 'Homogeneity tests', 'Observed power', 'Parameter estimates', 'SSCP matrices', 'Residual SSCP matrix', 'Transformation matrix', 'Homogeneity tests', 'Spread vs. level plot', 'Residual plot', 'Lack of fit', and 'General estimable function'. The 'Significance level' is set to .05 and 'Confidence intervals are 95.0%'. The background shows a data table with columns 'group', 'pc', 'bd', 'voc', 'oa', 'Gender', and three 'var' columns.

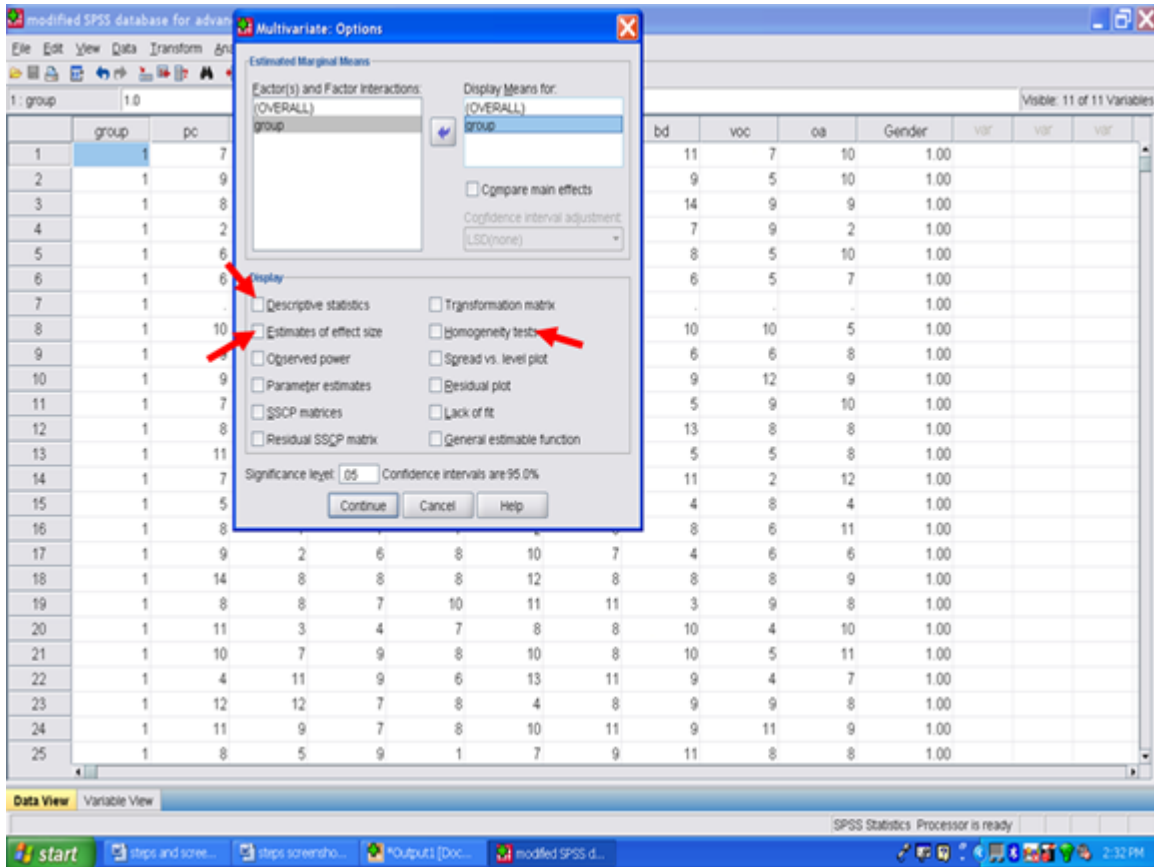
To obtain the information just mentioned, you will need to click on:

Descriptive Statistics

Estimates of Effect Size

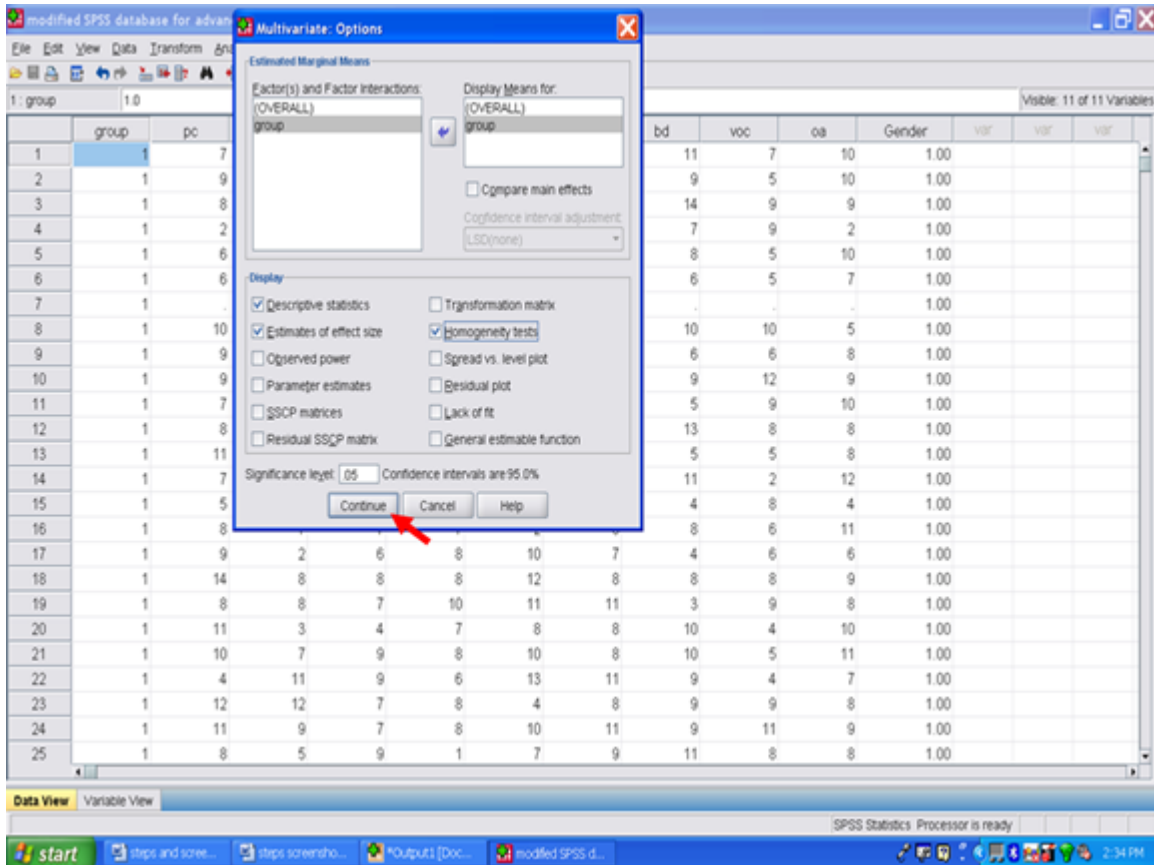
Homogeneity tests

⁸ <http://cnx.org/content/m40729/latest/11.5.png/image>



After checking the three boxes mentioned, then click on Continue

⁹ <http://cnx.org/content/m40729/latest/11.6.png/image>



After clicking on Continue, the following screen will appear. Now click on Post Hoc so that pairwise analyses can be conducted.

¹⁰<http://cnx.org/content/m40729/latest/11.7.png/image>

The screenshot shows the SPSS Multivariate dialog box with the following settings:

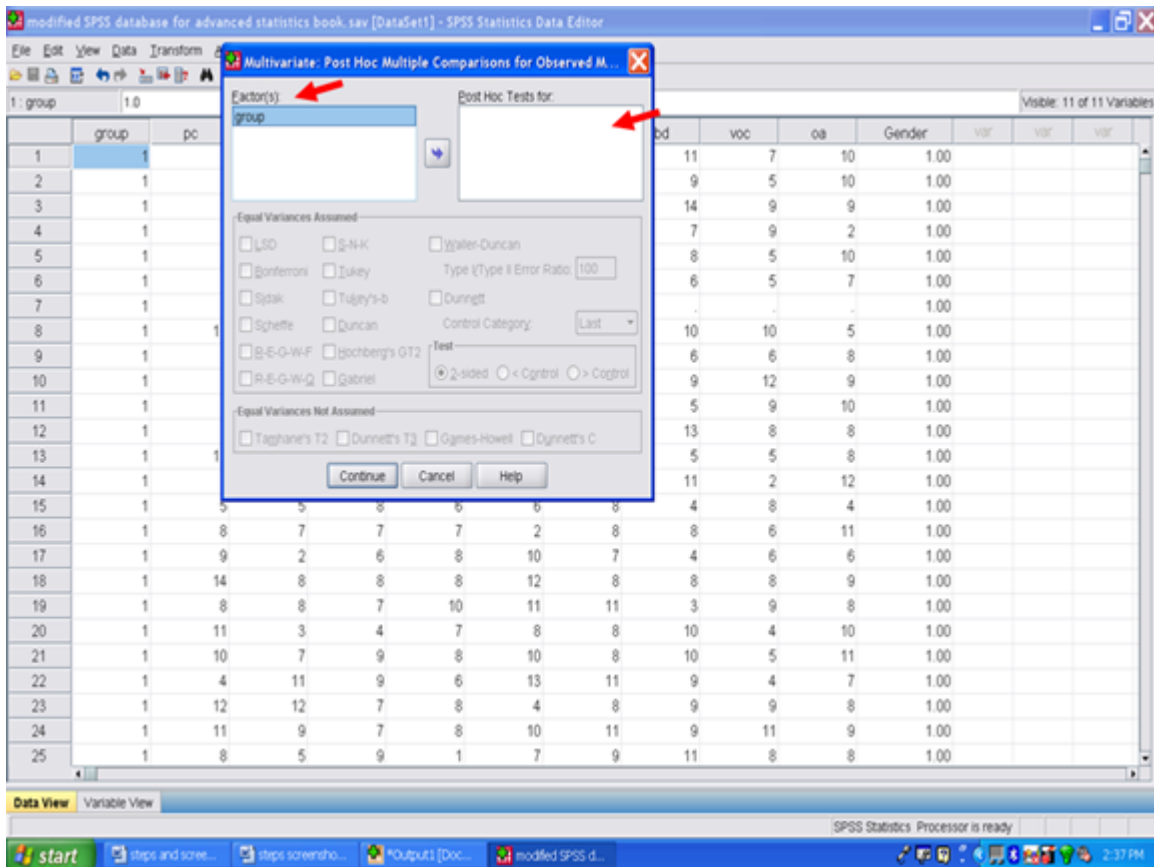
- Dependent Variables:** Verbal 2 (Smi...), Verbal 3 (Arith...), Verbal 4 (Voca...)
- Fixed Factor(s):** Reading Group M...
- Model:** (empty)
- Post Hoc:** (highlighted with a red arrow)
- Save:** (empty)
- Options:** (empty)

The data table in the background is as follows:

group	pc	bd	voc	oa	Gender	var	var	var
1	1	7	11	7	10	1.00		
2	1	9	9	5	10	1.00		
3	1	8	14	9	9	1.00		
4	1	2	7	9	2	1.00		
5	1	6	8	5	10	1.00		
6	1	6	6	5	7	1.00		
7	1	1.00		
8	1	10	10	10	5	1.00		
9	1	9	6	6	8	1.00		
10	1	9	9	12	9	1.00		
11	1	7	5	9	10	1.00		
12	1	8	13	8	8	1.00		
13	1	11	5	5	8	1.00		
14	1	7	4	5	8	7	11	11
15	1	5	5	8	6	6	8	4
16	1	8	7	7	7	2	8	8
17	1	9	2	6	8	10	7	4
18	1	14	8	8	8	12	8	8
19	1	8	8	7	10	11	11	3
20	1	11	3	4	7	8	8	10
21	1	10	7	9	8	10	8	10
22	1	4	11	9	6	13	11	9
23	1	12	12	7	8	4	8	9
24	1	11	9	7	8	10	11	9
25	1	8	5	9	1	7	9	11

Clicking on Post Hoc will then give you the screen below. Click on group and send it to the box labeled Post Hoc Tests for:

¹¹<http://cnx.org/content/m40729/latest/11.8.png/image>



Once you send the independent variable over to the box labeled Post Hoc Tests for:, the different types of post hoc procedures become activated. For our purposes, we will click on the Scheffe' post hoc procedure. If you are analyzing data for your dissertation or theses, you might check with your chair to see if s/he has a preference for a different type. Scheffe' is a conservative post hoc procedure, thus the reason that we recommend its use. Note that the way that Scheffe is spelled in SPSS is not written correctly in compliance with APA. The way that we have typed it is correct.

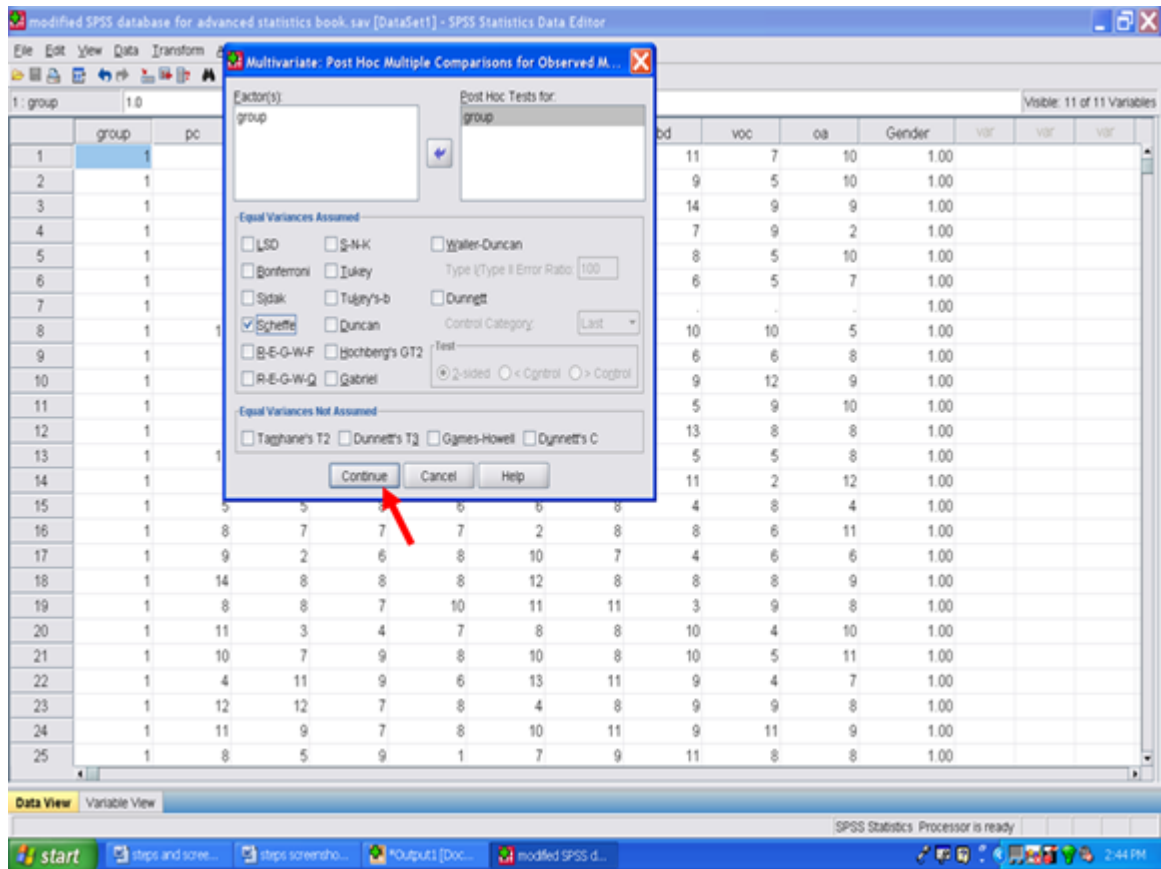
¹²<http://cnx.org/content/m40729/latest/11.9.png/image>

The screenshot shows the SPSS 'Multivariate: Post Hoc Multiple Comparisons for Observed M...' dialog box. The 'Factor(s)' field contains 'group', and the 'Post Hoc Tests for:' field also contains 'group'. Under the 'Equal Variances Assumed' section, the 'Scheffe'' checkbox is selected. A red arrow points to the 'Scheffe'' checkbox, and another red arrow points to the 'Post Hoc Tests for:' field. The background shows a data table with columns 'group', 'pc', 'bd', 'voc', 'oa', 'Gender', and three 'var' columns.

group	pc	bd	voc	oa	Gender	var	var	var
1	1	11	7	10	1.00			
2	1	9	5	10	1.00			
3	1	14	9	9	1.00			
4	1	7	9	2	1.00			
5	1	8	5	10	1.00			
6	1	6	5	7	1.00			
7	1	.	.	.	1.00			
8	1	10	10	5	1.00			
9	1	6	6	8	1.00			
10	1	9	12	9	1.00			
11	1	5	9	10	1.00			
12	1	13	8	8	1.00			
13	1	5	5	8	1.00			
14	1	11	2	12	1.00			
15	1	5	5	8	1.00			
16	1	8	7	7	1.00			
17	1	9	2	6	1.00			
18	1	14	8	8	1.00			
19	1	8	8	7	1.00			
20	1	11	3	4	1.00			
21	1	10	7	9	1.00			
22	1	4	11	9	1.00			
23	1	12	12	7	1.00			
24	1	11	9	7	1.00			
25	1	8	5	9	1.00			

After clicking on the Scheffe' post hoc procedure, then click on Continue.

¹³<http://cnx.org/content/m40729/latest/11.10.png/image>



Then click on OK.

¹⁴<http://cnx.org/content/m40729/latest/11.11.png/image>

The screenshot shows the SPSS Multivariate dialog box. The 'Dependent Variables' list contains 'Performance 1 (P...)', 'Performance 2 (C...)', 'Performance 3 (P...)', 'Performance 4 (B...)', 'Performance 5 (...)', and 'Gender of Perso...'. The 'Fixed Factor(s)' list contains 'Reading Group M...'. The 'Covariate(s)' list is empty. The 'WLS Weight' field is empty. The 'OK' button is highlighted with a red arrow.

group	pc	bd	voc	oa	Gender	var	var	var
1	1	7						
2	1	9						
3	1	8						
4	1	2						
5	1	6						
6	1	6						
7	1							
8	1	10						
9	1	9						
10	1	9						
11	1	7						
12	1	8						
13	1	11						
14	1	7	4	5	8	7	11	
15	1	5	5	8	6	6	8	
16	1	8	7	7	7	2	8	
17	1	9	2	6	8	10	7	
18	1	14	8	8	8	12	8	
19	1	8	8	7	10	11	11	
20	1	11	3	4	7	8	8	
21	1	10	7	9	8	10	8	
22	1	4	11	9	6	13	11	
23	1	12	12	7	8	4	8	
24	1	11	9	7	8	10	11	
25	1	8	5	9	1	7	9	

Your MANOVA has now been calculated. If you are not sent to the SPSS Output file, click on Output at the bottom of your screen so that you may view your MANOVA results.

¹⁵<http://cnx.org/content/m40729/latest/11.12.png/image>

Chapter 13

13. Multiple Analysis of Variance: Part III¹



NOTE: This chapter is published by NCPEA Press² and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of education administration.

About the Authors

- **John R. Slate** is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website (Writing and Statistical Help³) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.
- **Ana Rojas-LeBouef** is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. Dr. LeBoeuf recently completed her doctoral degree in Reading, where she conducted a 16-year analysis of Texas statewide data regarding the achievement gap. Her research interests lie in examining the inequities in achievement among ethnic groups. Dr. Rojas-LeBouef also assists students and faculty in their writing and statistical needs on the Writing and Statistical Help website.

Now that you are in the SPSS output screen, your screen should look like the following: Your first table is titled Between-Subjects Factors and it is entirely redundant with the table immediately below it. The column of n in the Between-Subjects Factors table is duplicated in the Descriptive Statistics table. Therefore, we will not use the Between-Subjects Factors table.

¹This content is available online at <<http://cnx.org/content/m40731/1.2/>>.

²<http://www.ncpeapublications.org/books.html>

³<http://cnx.org/content/m40731/latest/www.writingandstatisticalhelp>

General Linear Model

Between-Subjects Factors

Reading Group Membership	Value Label	N
1	Excellent Reader	395
2	Good Reader	261
3	Extremely Poor Reader	493

Descriptive Statistics

Verbal Test	Reading Group Membership	Mean	Std. Deviation	N
Verbal 1 (Information)	Excellent Reader	7.34	2.383	395
	Good Reader	3.46	1.878	261
	Extremely Poor Reader	7.12	2.491	493
	Total	6.36	2.910	1139
Verbal 2 (Similarities)	Excellent Reader	7.44	2.486	395
	Good Reader	3.18	1.894	261
	Extremely Poor Reader	6.83	2.729	493
	Total	6.20	2.984	1139
Verbal 3 (Arithmetic)	Excellent Reader	7.19	2.214	395
	Good Reader	3.90	2.236	261
	Extremely Poor Reader	7.46	2.427	493
	Total	6.55	2.730	1139
Verbal 4 (Vocabulary)	Excellent Reader	6.47	2.555	395
	Good Reader	2.87	1.593	261
	Extremely Poor Reader	6.54	2.611	493
	Total	5.63	2.889	1139

Bonferroni Test of Equality of Covariance Matrices^a

Bonferroni M	154.351
F	7.675
df1	20
df2	2953487.651
Sig.	.000

Tests the null hypothesis that the covariance matrices are equal.

Focusing on the Descriptive Statistics table will show you a column for the M , for the SD , and for the sample size, n . This information has been provided to you for each of the four dependent variables for each of the three groups. This information will be used in your Results section and should be kept.

⁴<http://cnx.org/content/m40731/latest/12.1.png/image>

The screenshot displays the SPSS Statistics Viewer interface. The main window shows the 'Descriptive Statistics' table, which is highlighted with a red arrow. Below this table is the 'Box's Test of Equality of Covariance Matrices' table, also highlighted with a red arrow. The 'Multivariate Tests' table is visible at the bottom of the main window. The taskbar at the bottom shows the Windows Start button and several open applications, including 'Output1 [Doc...]' and 'modified SPSS d...'. The system tray shows the time as 3:20 PM on 11/27/11.

Descriptive Statistics

Reading Group Membership	Mean	Std. Deviation	N
Verbal 1 (Informal)			
Excellent Reader	7.34	2.383	395
Good Reader	3.48	1.878	261
Extremely Poor Reader	7.12	2.481	493
Total	6.36	2.810	1139
Verbal 2 (Similar)			
Excellent Reader	7.44	2.486	395
Good Reader	3.18	1.894	261
Extremely Poor Reader	6.83	2.729	493
Total	6.20	2.984	1139
Verbal 3 (Arithmetic)			
Excellent Reader	7.19	2.214	395
Good Reader	3.90	2.238	261
Extremely Poor Reader	7.46	2.427	493
Total	6.55	2.730	1139
Verbal 4 (Vocabulary)			
Excellent Reader	6.47	2.555	395
Good Reader	2.87	1.593	261
Extremely Poor Reader	6.54	2.811	493
Total	5.83	2.889	1139

Box's Test of Equality of Covariance Matrices^a

Box's M	154.351
F	1.675
df1	20
df2	295307.651
Sig.	.234

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.
a. Design Intercept * group

Multivariate Tests^a

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	
Intercept	Pillai's Trace	.815	2047.526 ^a	4.000	1133.000	.000	.915

Underneath the Descriptive Statistics table is the Box's Test of Equality of Covariance Matrices. The information in this box involves checking one of the assumptions underlying use of a MANOVA procedure. The Sig. value in this box should be greater than .05 if the assumption is met. In the example below, the Sig. value is .234 which means that this particular assumption has been met. Readers should be informed whether the Box's *M* assumption was met or was violated. See Field (2010) for a detailed explanation of Box's *M* and the use of a MANOVA, even when this assumption is violated.

⁵ <http://cnx.org/content/m40731/latest/12.2.png/image>

The screenshot displays the SPSS Statistics Viewer interface. The main window shows the following data:

Descriptive Statistics				
Reading Group Membership	Mean	Std. Deviation	N	
Verbal 1 (Informal)	Excellent Reader	7.34	2.383	395
	Good Reader	3.48	1.878	261
	Extremely Poor Reader	7.12	2.481	493
	Total	6.36	2.810	1139
Verbal 2 (Similar)	Excellent Reader	7.44	2.486	395
	Good Reader	3.18	1.894	261
	Extremely Poor Reader	6.83	2.729	493
	Total	6.20	2.984	1139
Verbal 3 (Arithmetic)	Excellent Reader	7.19	2.214	395
	Good Reader	3.90	2.238	261
	Extremely Poor Reader	7.46	2.427	493
	Total	6.55	2.730	1139
Verbal 4 (Vocabulary)	Excellent Reader	6.47	2.555	395
	Good Reader	2.87	1.593	261
	Extremely Poor Reader	6.54	2.811	493
	Total	5.83	2.889	1139

Box's Test of Equality of Covariance Matrices ^a	
Box's M	154.351
F	1.675
df1	20
df2	295307.651
Sig.	.234

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.
a. Design: Intercept * group

Multivariate Tests ^a							
Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	
Intercept	Pillai's Trace	.815	2047.526 ^a	4.000	1133.000	.000	.915

The screenshot also shows the Windows taskbar at the bottom with the system clock at 3:20 PM and the SPSS Statistics Processor ready.

Underneath the Box's Test of Equality of Covariance Matrices is the Multivariate Tests table. This table is important because it indicates whether or not a statistically significant difference is present among the reading groups (i.e., our independent variable) in the aggregated dependent variable (i.e., aggregated Verbal 1 through Verbal 4). For our purposes, we will use Wilks' Lambda to determine whether or not a difference is present. In this example, a statistically significant difference is present.

⁶ <http://cnx.org/content/m40731/latest/12.3.png/image>

Multivariate Tests^a

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	
Intercept	Pillai's Trace	.915	3047.526 ^a	4.000	1133.000	.000	.915
	Wilks' Lambda	.005	3047.526 ^a	4.000	1133.000	.000	.915
	Hotelling's Trace	10.759	3047.526 ^a	4.000	1133.000	.000	.915
	Roy's Largest Root	10.759	3047.526 ^a	4.000	1133.000	.000	.915
group	Pillai's Trace	.469	86.881	8.000	2268.000	.000	.235
	Wilks' Lambda	.542	101.858 ^a	8.000	2268.000	.000	.264
	Hotelling's Trace	.827	117.002	8.000	2268.000	.000	.293
	Roy's Largest Root	.802	227.447 ^b	4.000	1134.000	.000	.445

a. Exact statistic.
b. The statistic is an upper bound on F that yields a lower bound on the significance level.
c. Design: Intercept + group

Levene's Test of Equality of Error Variances^a

Source	F	df1	df2	Sig.
Verbal 1 (Information)	.310	2	1136	.453
Verbal 2 (Similarities)	1.266	2	1136	.319
Verbal 3 (Arithmetic)	.981	2	1136	.651
Verbal 4 (Vocabulary)	1.894	2	1136	.249

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a. Design: Intercept + group

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Verbal 1 (Information)	2856.693 ^a	2	1428.346	264.859	.000	.318
	Verbal 2 (Similarities)	3166.903 ^a	2	1583.451	258.105	.000	.312
	Verbal 3 (Arithmetic)	2401.448 ^a	2	1200.725	224.261	.000	.283
	Verbal 4 (Vocabulary)	2975.028 ^a	2	1487.514	259.160	.000	.313
Intercept	Verbal 1 (Information)	37959.390	1	37959.390	7038.812	.000	.861
	Verbal 2 (Similarities)	36045.112	1	36045.112	5875.418	.000	.838
	Verbal 3 (Arithmetic)	40666.157	1	40666.157	7595.282	.000	.870
	Verbal 4 (Vocabulary)	29074.830	1	29074.830	5065.535	.000	.817
group	Verbal 1 (Information)	2856.693	2	1428.346	264.859	.000	.318
	Verbal 2 (Similarities)	3166.903	2	1583.451	258.105	.000	.312

Next, the Levene's Test of Equality of Error Variances table is examined. This table involves checking the assumption that the variability is consistent within each dependent variable for each cell in the research design. The Sig. value in each row must be greater than .05 for each assumption to be met. In the table below, the assumption of equality of error variances is met for all four dependent variables as the Sig. values range from .249 to .651. Readers should be informed that this assumption for each dependent variable was met or was not met. See Field (2010) for a detailed discussion regarding Levene's Test and the robustness of MANOVA, even when this assumption is violated.

⁷ <http://cnx.org/content/m40731/latest/12.4.png/image>

The screenshot shows the SPSS Statistics Viewer interface. The main window displays the following tables:

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.915	3047.526 ^a	4.000	1133.000	.000	.915
	Wilks' Lambda	.005	3047.526 ^a	4.000	1133.000	.000	.915
	Hotelling's Trace	10.759	3047.526 ^a	4.000	1133.000	.000	.915
	Roy's Largest Root	10.759	3047.526 ^a	4.000	1133.000	.000	.915
group	Pillai's Trace	.469	86.881	8.000	2268.000	.000	.235
	Wilks' Lambda	.542	101.656 ^a	8.000	2268.000	.000	.264
	Hotelling's Trace	.827	117.002	8.000	2268.000	.000	.293
	Roy's Largest Root	.802	227.447 ^a	4.000	1134.000	.000	.445

a. Exact statistic.
b. The statistic is an upper bound on F that yields a lower bound on the significance level.
c. Design: Intercept + group

Levene's Test of Equality of Error Variances ^a				
	F	df1	df2	Sig.
Verbal 1 (Information)	.918	2	1136	.453
Verbal 2 (Similarities)	1.286	2	1136	.319
Verbal 3 (Arithmetic)	.981	2	1136	.651
Verbal 4 (Vocabulary)	1.894	2	1136	.349

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a. Design: Intercept + group

Tests of Between-Subjects Effects							
Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Verbal 1 (Information)	2956.893 ^a	2	1428.346	264.858	.000	.318
	Verbal 2 (Similarities)	3166.903 ^a	2	1583.451	258.105	.000	.312
	Verbal 3 (Arithmetic)	2401.448 ^a	2	1200.725	224.261	.000	.283
	Verbal 4 (Vocabulary)	2975.028 ^a	2	1487.514	258.180	.000	.313
Intercept	Verbal 1 (Information)	37959.390	1	37959.390	7038.812	.000	.861
	Verbal 2 (Similarities)	36045.112	1	36045.112	5875.418	.000	.838
	Verbal 3 (Arithmetic)	40686.157	1	40686.157	7595.282	.000	.870
	Verbal 4 (Vocabulary)	29074.930	1	29074.930	5065.535	.000	.817
group	Verbal 1 (Information)	2956.893	2	1428.346	264.858	.000	.318
	Verbal 2 (Similarities)	3166.903	2	1583.451	258.105	.000	.312
	Verbal 3 (Arithmetic)	2401.448	2	1200.725	224.261	.000	.283

NUM LOCK: OFF

Underneath the Levene's Test of Equality table is the Tests of Between-Subjects Effects table. Contained in this table are the univariate ANOVAs regarding whether the reading groups differ on each of the four dependent variables. The Wilks' Lambda simply indicated whether an overall difference across the aggregated dependent variables was present. It did not provide information on whether the difference was for one, two, three, or all four of the dependent variables.

⁸ <http://cnx.org/content/m40731/latest/12.5.png/image>

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Verbal 1 (Information)	2956.893 ^a	2	1428.346	264.858	.000	.318
	Verbal 2 (Similarities)	3166.903 ^b	2	1583.451	258.105	.000	.312
	Verbal 3 (Arithmetic)	2401.449 ^c	2	1200.725	224.261	.000	.263
	Verbal 4 (Vocabulary)	2975.028 ^d	2	1487.514	258.160	.000	.313
Intercept	Verbal 1 (Information)	37959.390	1	37959.390	7038.812	.000	.861
	Verbal 2 (Similarities)	36045.112	1	36045.112	5875.418	.000	.838
	Verbal 3 (Arithmetic)	40666.157	1	40666.157	7595.282	.000	.870
	Verbal 4 (Vocabulary)	29074.930	1	29074.930	5065.535	.000	.817
group	Verbal 1 (Information)	2956.893	2	1428.346	264.858	.000	.318
	Verbal 2 (Similarities)	3166.903	2	1583.451	258.105	.000	.312
	Verbal 3 (Arithmetic)	2401.449	2	1200.725	224.261	.000	.263
	Verbal 4 (Vocabulary)	2975.028	2	1487.514	258.160	.000	.313
Error	Verbal 1 (Information)	6126.300	1136	5.393			
	Verbal 2 (Similarities)	6869.248	1136	6.135			
	Verbal 3 (Arithmetic)	6082.296	1136	5.354			
	Verbal 4 (Vocabulary)	8520.361	1136	7.490			
Total	Verbal 1 (Information)	54991.000	1139				
	Verbal 2 (Similarities)	53959.000	1139				
	Verbal 3 (Arithmetic)	57370.000	1139				
	Verbal 4 (Vocabulary)	45603.000	1139				
Corrected Total	Verbal 1 (Information)	8982.992	1138				
	Verbal 2 (Similarities)	10136.151	1138				
	Verbal 3 (Arithmetic)	8483.745	1138				
	Verbal 4 (Vocabulary)	9495.389	1138				

a. R Squared = .318 (Adjusted R Squared = .317)
b. R Squared = .312 (Adjusted R Squared = .311)
c. R Squared = .263 (Adjusted R Squared = .262)
d. R Squared = .313 (Adjusted R Squared = .312)

Estimated Marginal Means

1. Grand Mean

95% Confidence Interval

Find the row section that begins with the name of the independent variable. In our case, the independent variable is labeled group. Each row in this group section is a separate analysis of variance result. Examining the Sig. column for our four rows shows that each univariate ANOVA yielded a statistically significant result.

⁹<http://cnx.org/content/m40731/latest/12.6.png/image>

SPSS Statistics Viewer - *Output [Document1] - SPSS Statistics Viewer

General Linear Model

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Verbal 1 (Information)	2956.893 ^a	2	1428.346	264.858	.000	.318
	Verbal 2 (Similarities)	3166.903 ^b	2	1583.451	258.105	.000	.312
	Verbal 3 (Arithmetic)	2401.449 ^c	2	1200.725	224.261	.000	.283
	Verbal 4 (Vocabulary)	2875.028 ^d	2	1437.514	258.180	.000	.313
Intercept	Verbal 1 (Information)	37959.390	1	37959.390	7038.812	.000	.861
	Verbal 2 (Similarities)	36045.112	1	36045.112	5875.418	.000	.838
	Verbal 3 (Arithmetic)	40666.157	1	40666.157	7595.282	.000	.870
	Verbal 4 (Vocabulary)	29074.930	1	29074.930	5085.535	.000	.817
group	Verbal 1 (Information)	2956.893	2	1428.346	264.858	.000	.318
	Verbal 2 (Similarities)	3166.903	2	1583.451	258.105	.000	.312
	Verbal 3 (Arithmetic)	2401.449	2	1200.725	224.261	.000	.283
	Verbal 4 (Vocabulary)	2875.028	2	1437.514	258.180	.000	.313
Error	Verbal 1 (Information)	6126.300	1136	5.393			
	Verbal 2 (Similarities)	6869.248	1136	6.135			
	Verbal 3 (Arithmetic)	6082.296	1136	5.354			
	Verbal 4 (Vocabulary)	8520.361	1136	7.490			
Total	Verbal 1 (Information)	54991.000	1139				
	Verbal 2 (Similarities)	53959.000	1139				
	Verbal 3 (Arithmetic)	57370.000	1139				
	Verbal 4 (Vocabulary)	45603.000	1139				
Corrected Total	Verbal 1 (Information)	8982.992	1138				
	Verbal 2 (Similarities)	10136.151	1138				
	Verbal 3 (Arithmetic)	8483.745	1138				
	Verbal 4 (Vocabulary)	9495.389	1138				

a. R Squared = .318 (Adjusted R Squared = .317)
 b. R Squared = .312 (Adjusted R Squared = .311)
 c. R Squared = .283 (Adjusted R Squared = .282)
 d. R Squared = .313 (Adjusted R Squared = .312)

Estimated Marginal Means

1. Grand Mean

95% Confidence Interval

These results may be interpreted to mean that a statistically significant difference was present for Verbal 1, for Verbal 1, for Verbal 3, and for Verbal 4 as a function of the reading group. The effect size information for each statistically significant difference is present in the last column, the one labeled Partial Eta Squared. Partial eta squared, or η^2 , is the effect size metric provided to you in the SPSS output. To interpret whether it is small, moderate, or large, see Cohen (1988).

¹⁰<http://cnx.org/content/m40731/latest/12.7.png/image>

SPSS Statistics Viewer - *Output [Document] - SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help

General Linear Model

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Verbal 1 (Information)	2956.893 ^a	2	1428.346	264.858	.000	.318
	Verbal 2 (Similarities)	3166.903 ^b	2	1583.451	258.105	.000	.312
	Verbal 3 (Arithmetic)	2401.449 ^c	2	1200.725	224.261	.000	.263
	Verbal 4 (Vocabulary)	2975.028 ^d	2	1487.514	258.160	.000	.313
Intercept	Verbal 1 (Information)	37959.390	1	37959.390	7038.812	.000	.861
	Verbal 2 (Similarities)	36045.112	1	36045.112	5875.418	.000	.838
	Verbal 3 (Arithmetic)	40666.157	1	40666.157	7595.282	.000	.870
	Verbal 4 (Vocabulary)	29074.930	1	29074.930	5065.535	.000	.817
group	Verbal 1 (Information)	2956.893	2	1428.346	264.858	.000	.318
	Verbal 2 (Similarities)	3166.903	2	1583.451	258.105	.000	.312
	Verbal 3 (Arithmetic)	2401.449	2	1200.725	224.261	.000	.263
	Verbal 4 (Vocabulary)	2975.028	2	1487.514	258.160	.000	.313
Error	Verbal 1 (Information)	6126.300	1136	5.393			
	Verbal 2 (Similarities)	6869.248	1136	6.135			
	Verbal 3 (Arithmetic)	6082.296	1136	5.354			
	Verbal 4 (Vocabulary)	8520.361	1136	5.740			
Total	Verbal 1 (Information)	54991.000	1139				
	Verbal 2 (Similarities)	53959.000	1139				
	Verbal 3 (Arithmetic)	57370.000	1139				
	Verbal 4 (Vocabulary)	45603.000	1139				
Corrected Total	Verbal 1 (Information)	8982.992	1138				
	Verbal 2 (Similarities)	10136.151	1138				
	Verbal 3 (Arithmetic)	8483.745	1138				
	Verbal 4 (Vocabulary)	9495.389	1138				

a. R Squared = .318 (Adjusted R Squared = .317)
b. R Squared = .312 (Adjusted R Squared = .311)
c. R Squared = .263 (Adjusted R Squared = .262)
d. R Squared = .313 (Adjusted R Squared = .312)

Estimated Marginal Means

1. Grand Mean

95% Confidence Interval

SPSS Statistics Processor is ready H: 160, W: 408 pt.

Because a statistically significant difference is present for each dependent variable, post hoc procedures must be examined to determine if all groups differ from each other or if only some of the groups differ. Post hoc procedure information is provided in the Multiple Comparisons table. The arrow below is pointed toward the specific type of post hoc we requested.

¹¹ <http://cnx.org/content/m40731/latest/12.8.png/image>

The screenshot shows the SPSS Statistics Viewer interface. The main window displays a 'Multiple Comparisons' table for 'Verbal 1 (Informal)'. A red arrow points to the 'Sig.' column. The table lists pairwise comparisons between three reading groups: Excellent Reader, Good Reader, and Extremely Poor Reader. The 'Sig.' column shows values of .000 for comparisons between Excellent and Extremely Poor Readers, and .17 for comparisons between Excellent and Good Readers. Below the table, it states 'Based on observed means. The error term is Mean Square(Error) = 5.740. * The mean difference is significant at the .05 level.' Below that, it says 'Homogeneous Subsets' and 'Verbal 1 (Informal)'. The taskbar at the bottom shows the Windows taskbar with the Start button and several open applications.

Dependent Variable	i Reading Group Membership	j Reading Group Membership	Mean Difference (i - j)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Verbal 1 (Informal)	Excellent Reader	Good Reader	.21	.158	.400	-.17	.60
		Extremely Poor Reader	3.88*	.186	.000	3.43	4.34
	Good Reader	Excellent Reader	-.21	.158	.400	-.60	.17
		Extremely Poor Reader	3.67*	.178	.000	3.23	4.10
	Extremely Poor Reader	Excellent Reader	-3.88*	.186	.000	-4.34	-3.43
		Good Reader	-3.67*	.178	.000	-4.10	-3.23
Verbal 2 (Similarities)	Excellent Reader	Good Reader	.81*	.168	.001	.28	1.03
		Extremely Poor Reader	4.26*	.199	.000	3.77	4.75
	Good Reader	Excellent Reader	-.81*	.168	.001	-1.03	-.20
		Extremely Poor Reader	3.65*	.190	.000	3.18	4.11
	Extremely Poor Reader	Excellent Reader	-4.26*	.199	.000	-4.75	-3.77
		Good Reader	-3.65*	.190	.000	-4.11	-3.18
Verbal 3 (Arithmetic)	Excellent Reader	Good Reader	-.26	.157	.245	-.65	.12
		Extremely Poor Reader	3.30*	.186	.000	2.84	3.75
	Good Reader	Excellent Reader	.26	.157	.245	-.12	.65
		Extremely Poor Reader	3.56*	.177	.000	3.13	3.99
	Extremely Poor Reader	Excellent Reader	-3.30*	.186	.000	-3.75	-2.84
		Good Reader	-3.56*	.177	.000	-3.99	-3.13
Verbal 4 (Vocabulary)	Excellent Reader	Good Reader	-.67	.163	.915	-.47	.33
		Extremely Poor Reader	3.81*	.192	.000	3.34	4.28
	Good Reader	Excellent Reader	.67	.163	.915	-.33	.47
		Extremely Poor Reader	3.87*	.183	.000	3.43	4.32
	Extremely Poor Reader	Excellent Reader	-3.81*	.192	.000	-4.28	-3.34
		Good Reader	-3.87*	.183	.000	-4.32	-3.43

The most important column in this post hoc table is the Sig. column. For a pairwise comparison to be statistically significant, the Sig. value must be .05 or below. In the example below for the Verbal 1 set of comparisons, Excellent Readers did not differ from Good Readers, however, Excellent readers did differ from Extremely Poor Readers.

¹²<http://cnx.org/content/m40731/latest/12.9.png/image>

Multiple Comparisons

Dependent Variable	i Reading Group Membership	j Reading Group Membership	Mean Difference (i - j)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Verbal 1 (Information)	Excellent Reader	Good Reader	.21	.158	.400	-.17	.60
		Extremely Poor Reader	3.88*	.186	.000	3.43	4.34
	Good Reader	Excellent Reader	-.21	.158	.400	-.60	.17
		Extremely Poor Reader	3.67*	.178	.000	3.23	4.10
	Extremely Poor Reader	Excellent Reader	-3.88*	.186	.000	-4.34	-3.43
		Good Reader	-3.67*	.178	.000	-4.10	-3.23
Verbal 2 (Similarities)	Excellent Reader	Good Reader	.81*	.168	.001	.28	1.03
		Extremely Poor Reader	4.26*	.199	.000	3.77	4.75
	Good Reader	Excellent Reader	-.81*	.168	.001	-1.03	-.20
		Extremely Poor Reader	3.65*	.190	.000	3.18	4.11
	Extremely Poor Reader	Excellent Reader	-4.26*	.199	.000	-4.75	-3.77
		Good Reader	-3.65*	.190	.000	-4.11	-3.18
Verbal 3 (Arithmetic)	Excellent Reader	Good Reader	-.26	.157	.245	-.65	.12
		Extremely Poor Reader	3.30*	.186	.000	2.84	3.75
	Good Reader	Excellent Reader	.26	.157	.245	-.12	.65
		Extremely Poor Reader	3.56*	.177	.000	3.13	3.99
	Extremely Poor Reader	Excellent Reader	-3.30*	.186	.000	-3.75	-2.84
		Good Reader	-3.56*	.177	.000	-3.99	-3.13
Verbal 4 (Vocabulary)	Excellent Reader	Good Reader	-.67	.163	.915	-.47	.33
		Extremely Poor Reader	3.81*	.192	.000	3.34	4.28
	Good Reader	Excellent Reader	.67	.163	.915	-.33	.47
		Extremely Poor Reader	3.87*	.183	.000	3.43	4.32
	Extremely Poor Reader	Excellent Reader	-3.81*	.192	.000	-4.28	-3.34
		Good Reader	-3.87*	.183	.000	-4.32	-3.43

Based on observed means.
The error term is Mean Square(Error) = 5.740.
*. The mean difference is significant at the .05 level.

Homogeneous Subsets

Verbal 1 (Information)

Schafly#3

Readers should note that only three pairwise comparisons are unique in the table for each dependent variable. SPSS, however, provides information on six pairwise comparisons. When three groups are present, as they are in this example, row 1, row 2, and row 4 are unique. Row 1 is Excellent Reader compared to Good Reader. Row 2 is Excellent Reader compared to Extremely Poor Reader. Row 4 is Good Reader compared to Extremely Poor Reader.

¹³<http://cnx.org/content/m40731/latest/12.10.png/image>

Multiple Comparisons

Dependent Variable	i Reading Group Membership	j Reading Group Membership	Mean Difference (i - j)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Verbal 1 (Information)	Excellent Reader	Good Reader	.21	.158	.400	-1.17	.80
		Extremely Poor Reader	3.88*	.186	.000	3.43	4.34
	Good Reader	Excellent Reader	-.21	.158	.400	-.80	.17
		Extremely Poor Reader	3.67*	.170	.000	3.23	4.10
	Extremely Poor Reader	Excellent Reader	-3.88*	.186	.000	-4.34	-3.43
		Good Reader	-3.67*	.170	.000	-4.10	-3.23
Verbal 2 (Similarities)	Excellent Reader	Good Reader	.81*	.168	.001	.28	1.33
		Extremely Poor Reader	4.26*	.199	.000	3.77	4.75
	Good Reader	Excellent Reader	-.81*	.168	.001	-1.03	-.20
		Extremely Poor Reader	3.65*	.190	.000	3.18	4.11
	Extremely Poor Reader	Excellent Reader	-4.26*	.199	.000	-4.75	-3.77
		Good Reader	-3.65*	.190	.000	-4.11	-3.18
Verbal 3 (Arithmetic)	Excellent Reader	Good Reader	-.26	.157	.245	-.85	.12
		Extremely Poor Reader	3.30*	.186	.000	2.84	3.75
	Good Reader	Excellent Reader	.26	.157	.245	-.12	.85
		Extremely Poor Reader	3.56*	.177	.000	3.13	3.99
	Extremely Poor Reader	Excellent Reader	-3.30*	.186	.000	-3.75	-2.84
		Good Reader	-3.56*	.177	.000	-3.99	-3.13
Verbal 4 (Vocabulary)	Excellent Reader	Good Reader	-.67	.163	.915	-.47	.33
		Extremely Poor Reader	3.81*	.192	.000	3.34	4.28
	Good Reader	Excellent Reader	.67	.163	.915	-.33	.47
		Extremely Poor Reader	3.87*	.183	.000	3.43	4.32
	Extremely Poor Reader	Excellent Reader	-3.81*	.192	.000	-4.28	-3.34
		Good Reader	-3.87*	.183	.000	-4.32	-3.43

Based on observed means.
The error term is Mean Square(Error) = 5.740.
*. The mean difference is significant at the .05 level.

Homogeneous Subsets

Verbal 1 (Information)

Schiffé's^a

Each of the four dependent variables in this example have post hoc procedures calculated for them and are present in the table above.

All of the useful information present in the SPSS output for the MANOVA procedure has now been covered.

¹⁴<http://cnx.org/content/m40731/latest/12.11.png/image>

Chapter 14

14. Discriminant Analysis: Assumptions¹



NOTE: This chapter is published by NCPEA Press² and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of education administration.

About the Authors

John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website (Writing and Statistical Help³) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

Ana Rojas-LeBouef is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. Dr. LeBoeuf recently completed her doctoral degree in Reading, where she conducted a 16-year analysis of Texas statewide data regarding the achievement gap. Her research interests lie in examining the inequities in achievement among ethnic groups. Dr. Rojas-LeBouef also assists students and faculty in their writing and statistical needs on the Writing and Statistical Help website.

In this set of steps, readers will learn how to conduct a canonical discriminant analysis procedure. For detailed information regarding the assumptions underlying use of a discriminant analysis, readers are referred to the Hyperstats Online Statistics Textbook at <http://davidmlane.com/hyperstat/>⁴; to the *Electronic Statistics Textbook* (2011)

¹This content is available online at <<http://cnx.org/content/m40733/1.2/>>.

²<http://www.ncpeapublications.org/books.html>

³<http://cnx.org/content/m40733/latest/www.writingandstatisticalhelp>

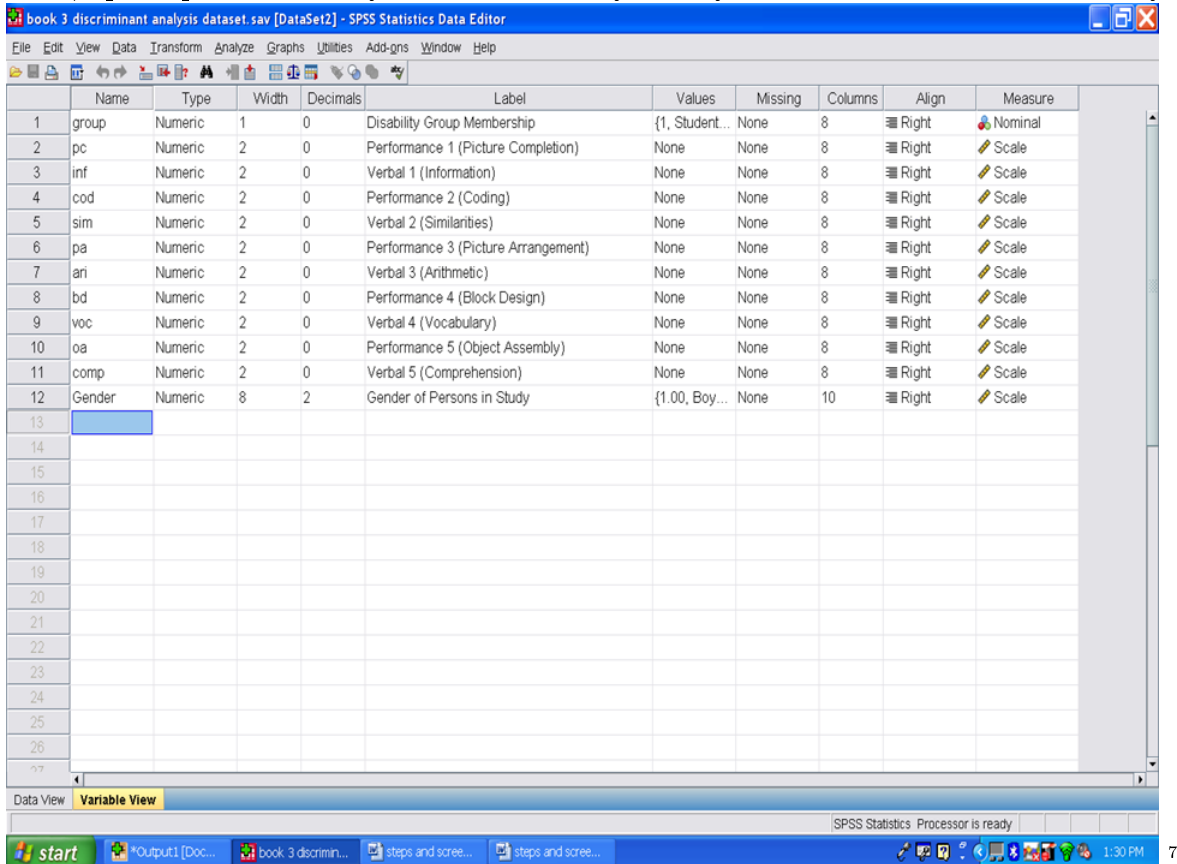
⁴<http://davidmlane.com/hyperstat/>

at <http://www.statsoft.com/textbook/>⁵ ; or to Andy Field's (2009) *Discovering Statistics Using SPSS* at http://www.amazon.com/Discovering-Statistics-Introducing-Statistical-Method/dp/1847879071/ref=sr_1_1?s=books&ie=UTF8&qid=1304967862&sr=1-1⁶

Research questions for which a discriminant analysis procedure is appropriate involve determining variables that predict group membership. For example, if two groups of persons are present such as completers and non-completers and archival data are available, then a discriminant analysis procedure could be utilized. Such a procedure could identify specific variables that differentiate group membership. As such, interventions could be developed and targeted toward the variables that predicted group membership. Other sample research questions for which a discriminant analysis might be appropriate: (a) What factors differentiates successful from unsuccessful students?; (b) What factors differentiate delinquents from nondelinquents?; (c) What set of test scores best differentiates students with LD, students who are failing, and students with MR?; and (d) What set of factors differentiates drop-outs from persisters?

For purposes of this chapter, our research question is: "What scholastic variables differentiate boys from girls?"

First, open up the dataset you intend to analyze for your canonical discriminant analysis.



Our independent variable is gender. Boys are labeled as group 1 and girls are labeled as group 2.

⁵<http://www.statsoft.com/textbook/>

⁶http://www.amazon.com/Discovering-Statistics-Introducing-Statistical-Method/dp/1847879071/ref=sr_1_1?s=books&ie=UTF8&qid=13049

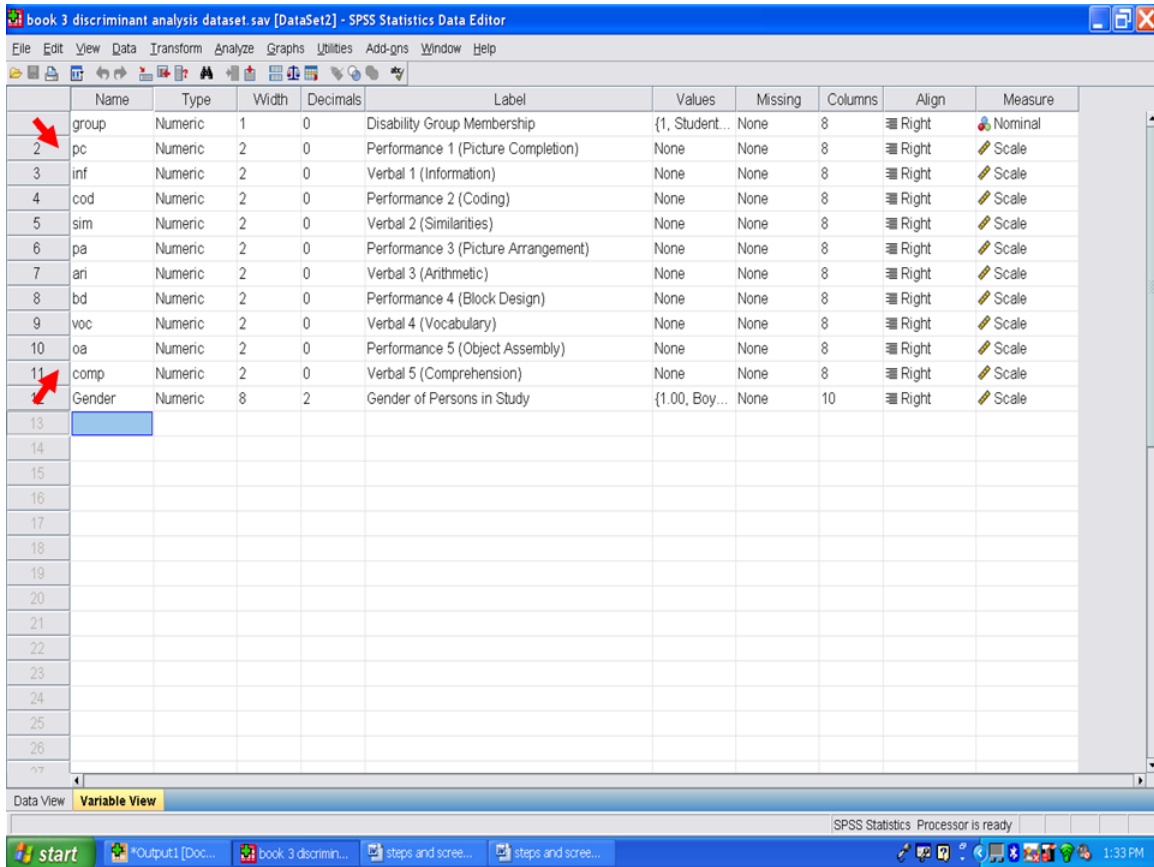
1

⁷<http://cnx.org/content/m40733/latest/13.1.png/image>

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	group	Numeric	1	0	Disability Group Membership	{1, Student...	None	8	Right	Nominal
2	pc	Numeric	2	0	Performance 1 (Picture Completion)	None	None	8	Right	Scale
3	inf	Numeric	2	0	Verbal 1 (Information)	None	None	8	Right	Scale
4	cod	Numeric	2	0	Performance 2 (Coding)	None	None	8	Right	Scale
5	sim	Numeric	2	0	Verbal 2 (Similarities)	None	None	8	Right	Scale
6	pa	Numeric	2	0	Performance 3 (Picture Arrangement)	None	None	8	Right	Scale
7	ari	Numeric	2	0	Verbal 3 (Arithmetic)	None	None	8	Right	Scale
8	bd	Numeric	2	0	Performance 4 (Block Design)	None	None	8	Right	Scale
9	voc	Numeric	2	0	Verbal 4 (Vocabulary)	None	None	8	Right	Scale
10	oa	Numeric	2	0	Performance 5 (Object Assembly)	None	None	8	Right	Scale
11	comp	Numeric	2	0	Verbal 5 (Comprehension)	None	None	8	Right	Scale
12	Gender	Numeric	8	2	Gender of Persons in Study	{1.00, Boy...	None	10	Right	Scale
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										

Our dependent variables, the ones we will use to differentiate boys from girls are 10 subscales from the Wechsler Intelligence Scale for Children-Third Edition: Picture Completion (pc), Information (inf), Coding (cod), Similarities (sim), Picture Arrangement (pa), Arithmetic (ari), Block Design (bd), Vocabulary (voc), Object Assembly (oa), and Comprehension (comp).

⁸<http://cnx.org/content/m40733/latest/13.2.png/image>



In the previous screenshots, we were in the variable view screen. Click on data view, shown below, so that your screen looks like the one below.

⁹<http://cnx.org/content/m40733/latest/13.3.png/image>

book 3 discriminant analysis dataset.sav [DataSet2] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1: pc 10.0 Visible: 12 of 12 Variables

	group	pc	inf	cod	sim	pa	ari	bd	voc	oa	comp	Gender	var	v
1	1	10	7	8	7	10	5	7	7	6	7	2.00		
2	1	8	8	7	6	7	9	9	4	4	7	2.00		
3	1	10	8	2	9	10	11	9	6	11	8	2.00		
4	1	3	4	11	4	7	6	3	4	10	6	1.00		
5	1	5	7	17	6	8	10	7	5	9	1	1.00		
6	1	2	5	8	8	9	7	10	3	5	5	1.00		
7	1	7	4	5	8	7	11	11	2	12	5	2.00		
8	1	7	7	7	7	14	8	10	8	11	3	2.00		
9	1	9	6	10	6	9	8	4	5	8	5	1.00		
10	1	5	9	7	10	9	12	11	7	8	11	1.00		
11	1	6	7	5	8	7	7	7	4	7	5	1.00		
12	1	13	8	5	13	10	5	6	9	9	7	2.00		
13	1	7	8	11	6	8	4	3	3	4	4	1.00		
14	1	6	6	8	12	9	9	9	10	6	8	1.00		
15	1	7	8	3	1	9	8	5	10	3	9	1.00		
16	1	10	7	5	5	9	5	5	8	3	6	2.00		
17	1	12	8	7	8	13	10	11	7	6	5	1.00		
18	1	15	7	5	10	15	8	7	8	7	10	2.00		
19	1	9	6	4	8	3	5	8	6	10	3	2.00		
20	1	6	7	10	6	9	6	8	6	5	2	2.00		
21	1	13	7	8	7	9	6	3	6	6	4	2.00		
22	1	10	7	12	8	8	8	12	5	11	4	2.00		
23	1	10	9	10	10	11	11	14	11	11	10	1.00		
24	1	6	6	6	2	5	7	6	6	11	6	2.00		
25	1	14	7	6	9	10	6	7	6	8	5	2.00		

Data View Variable View

SPSS Statistics Processor is ready

start *Output1 [Doc... book 3 discrim... steps and scree... steps and scree... 1:34 PM 10

Prior to conducting a canonical discriminant function, we need to check the assumptions that underlie its use.

14.1 Normal Distribution

It is assumed that the data (for the variables) represent a sample from a multivariate normal distribution. You can examine whether or not variables are normally distributed with histograms of frequency distributions. However, note that violations of the normality assumption are usually not "fatal," meaning, that the resultant significance tests etc. are still "trustworthy." You may use specific tests for normality in addition to graphs. <http://www.statsoft.com/textbook/discriminant-function-analysis/#assumptions>¹¹

We recommend that you calculate the standardized skewness coefficients and the standardized kurtosis coefficients, as discussed in other chapters.

* Skewness [Note. Skewness refers to the extent to which the data are normally distributed around the mean. Skewed data involve having either mostly high scores with a few low ones or having mostly low scores with a few high ones.] Readers are referred to the following sources for a more detailed definition of skewness: http://www.statistics.com/index.php?page=glossary&term_id=356¹² and <http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb>¹³

¹⁰<http://cnx.org/content/m40733/latest/13.4.png/image>

¹¹<http://www.statsoft.com/textbook/discriminant-function-analysis/#assumptions>

¹²http://www.statistics.com/index.php?page=glossary&term_id=356

¹³<http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb>

To standardize the skewness value so that its value can be constant across datasets and across studies, the following calculation must be made: Take the skewness value from the SPSS output and divide it by the Std. error of skewness. If the resulting calculation is within -3 to +3, then the skewness of the dataset is within the range of normality (Onwuegbuzie & Daniel, 2002). If the resulting calculation is outside of this +/-3 range, the dataset is not normally distributed.

* Kurtosis [Note. Kurtosis also refers to the extent to which the data are normally distributed around the mean. This time, the data are piled up higher than normal around the mean or piled up higher than normal at the ends of the distribution.] Readers are referred to the following sources for a more detailed definition of kurtosis: http://www.statistics.com/index.php?page=glossary&term_id=326¹⁴ and <http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb>¹⁵

To standardize the kurtosis value so that its value can be constant across datasets and across studies, the following calculation must be made: Take the kurtosis value from the SPSS output and divide it by the Std. error of kurtosis. If the resulting calculation is within -3 to +3, then the kurtosis of the dataset is within the range of normality (Onwuegbuzie & Daniel, 2002). If the resulting calculation is outside of this +/-3 range, the dataset is not normally distributed.

14.2 Homogeneity of Variances/Covariances

It is assumed that the variance/covariance matrices of variables are homogeneous across groups. Again, minor deviations are not that important. <http://www.statsoft.com/textbook/discriminant-function-analysis/#assumptions>¹⁶

14.3 Correlations between Means and Variances

The major "real" threat to the validity of significance tests occurs when the means for variables across groups are correlated with the variances (or standard deviations). Intuitively, if there is large variability in a group with particularly high means on some variables, then those high means are not reliable. However, the overall significance tests are based on pooled variances, that is, the average variance across all groups. Thus, the significance tests of the relatively larger means (with the large variances) would be based on the relatively smaller pooled variances, resulting erroneously in statistical significance. In practice, this pattern may occur if one group in the study contains a few extreme outliers, who have a large impact on the means, and also increase the variability. To guard against this problem, inspect the descriptive statistics, that is, the means and standard deviations or variances for such a correlation. <http://www.statsoft.com/textbook/discriminant-function-analysis/#assumptions>¹⁷

After calculating the means and standard deviations for your variables for each of your groups, check them to determine if large variability is present in the means for one of your groups compared to the means for the other group.

14.4 The Matrix Ill-Conditioning Problem

Another assumption of discriminant function analysis is that the variables that are used to discriminate between groups are not completely redundant. As part of the computations in-

¹⁴http://www.statistics.com/index.php?page=glossary&term_id=326

¹⁵<http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb>

¹⁶<http://www.statsoft.com/textbook/discriminant-function-analysis/#assumptions>

¹⁷<http://www.statsoft.com/textbook/discriminant-function-analysis/#assumptions>

volved in discriminant analysis, you will invert the variance/covariance matrix of the variables in the model. If any one of the variables is completely redundant with the other variables then the matrix is said to be *ill-conditioned*, and it cannot be inverted. For example, if a variable is the sum of three other variables that are also in the model, then the matrix is ill-conditioned. <http://www.statsoft.com/textbook/discriminant-function-analysis/#assumptions>¹⁸

What this assumption means is that each variable should be unique from any other variable in the analysis. Having one variable that includes another variable would be a violation of this assumption. An example of this would be using a total score that contains several subscale scores, all of which are used in the discriminant analysis.

14.5 Tolerance Values.

In order to guard against matrix ill-conditioning, constantly check the so-called tolerance value for each variable. This tolerance value is computed as *1 minus R-square* of the respective variable with all other variables included in the current model. Thus, it is the proportion of variance that is unique to the respective variable. In general, when a variable is almost completely redundant (and, therefore, the matrix ill-conditioning problem is likely to occur), the tolerance value for that variable will approach 0. <http://www.statsoft.com/textbook/discriminant-function-analysis/#assumptions>¹⁹

We will check this assumption, the tolerance values, when we examine the SPSS output.

¹⁸<http://www.statsoft.com/textbook/discriminant-function-analysis/#assumptions>

¹⁹<http://www.statsoft.com/textbook/discriminant-function-analysis/#assumptions>

Chapter 15

15. Discriminant Analysis: Part I¹



NOTE: This chapter is published by NCPEA Press² and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of education administration.

About the Authors

John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website (Writing and Statistical Help³) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

Ana Rojas-LeBouef is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. Dr. LeBoeuf recently completed her doctoral degree in Reading, where she conducted a 16-year analysis of Texas statewide data regarding the achievement gap. Her research interests lie in examining the inequities in achievement among ethnic groups. Dr. Rojas-LeBouef also assists students and faculty in their writing and statistical needs on the Writing and Statistical Help website.

Now that we have ascertained the assumptions underlying use of a discriminant analysis procedure, we will begin the steps.

Click on Analyze

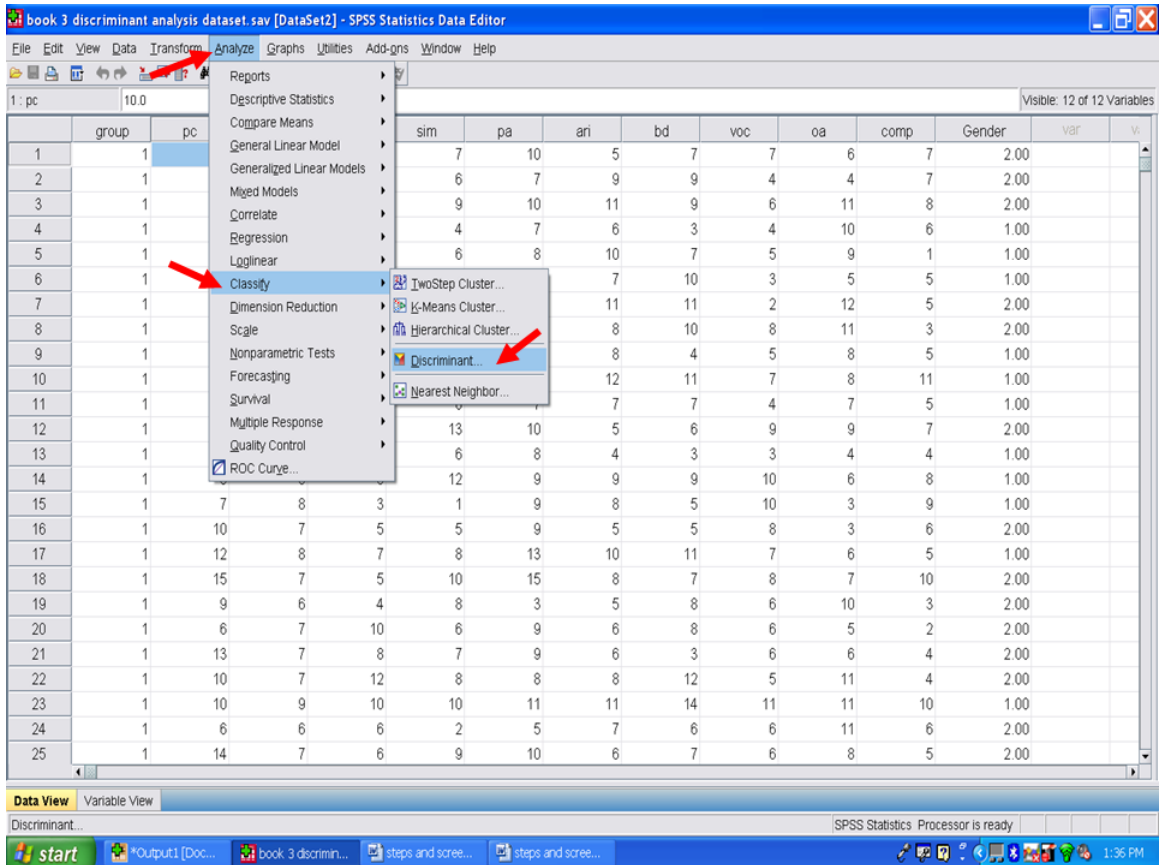
Click on Classify

Click on Discriminant

¹This content is available online at <<http://cnx.org/content/m40735/1.2/>>.

²<http://www.ncpeapublications.org/books.html>

³<http://cnx.org/content/m40735/latest/www.writingandstatisticalhelp>



The following screen will then appear. Grouping variable is your independent variable. Remember that in our example that gender is our independent or grouping variable.

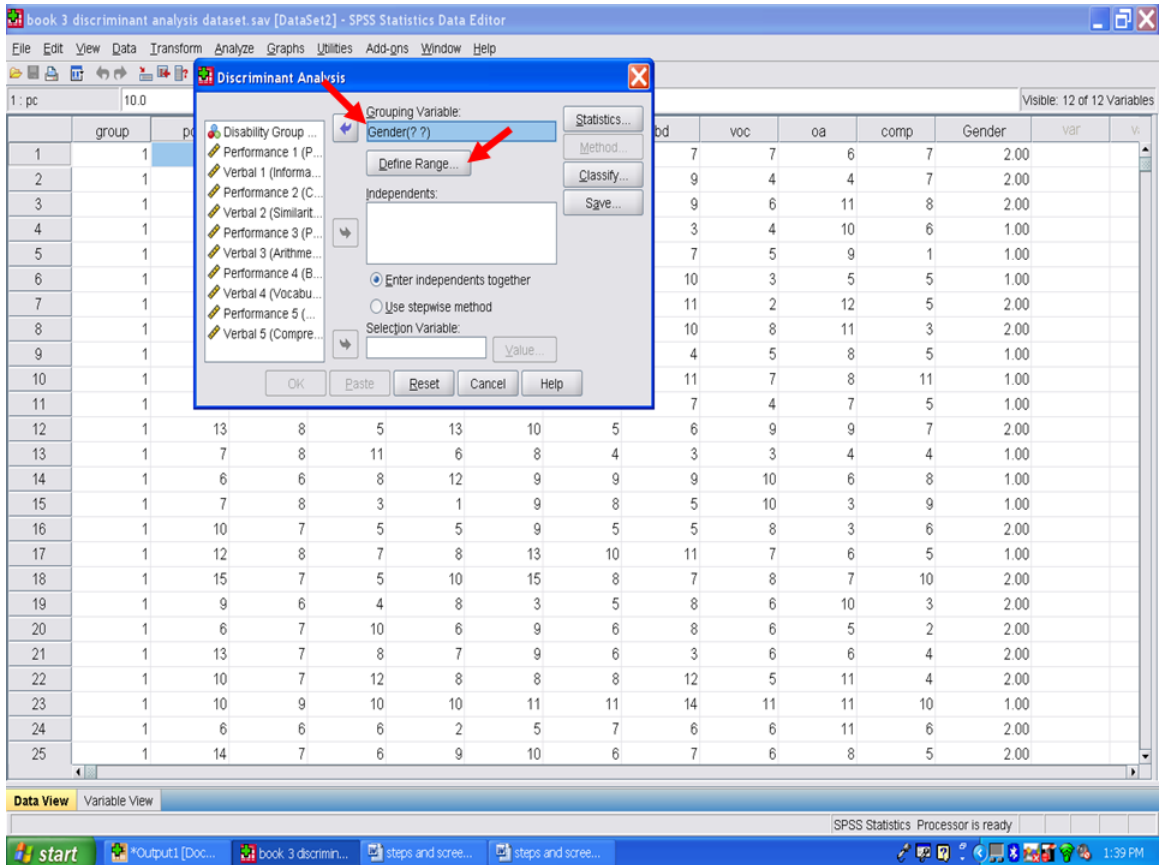
⁴<http://cnx.org/content/m40735/latest/14.1.png/image>

The screenshot shows the SPSS Discriminant Analysis dialog box. The 'Grouping Variable' field contains 'Gender'. The 'Enter independents together' radio button is selected. The 'Define Range...' button is highlighted with a red arrow. The background shows a data view with columns for 'group', 'pv', 'bd', 'voc', 'oa', 'comp', 'Gender', 'var', and 'v'.

group	pv	bd	voc	oa	comp	Gender	var	v						
1	1	7	7	6	7	2.00								
2	1	9	4	4	7	2.00								
3	1	9	6	11	8	2.00								
4	1	3	4	10	6	1.00								
5	1	7	5	9	1	1.00								
6	1	10	3	5	5	1.00								
7	1	11	2	12	5	2.00								
8	1	10	8	11	3	2.00								
9	1	4	5	8	5	1.00								
10	1	11	7	8	11	1.00								
11	1	7	4	7	5	1.00								
12	1	13	8	5	13	10	5	6	9	7	2.00			
13	1	7	8	11	6	8	4	3	3	4	4	1.00		
14	1	6	6	8	12	9	9	9	10	6	8	1.00		
15	1	7	8	3	1	9	8	5	10	3	9	1.00		
16	1	10	7	5	5	9	5	5	8	3	6	2.00		
17	1	12	8	7	8	13	10	11	7	6	5	1.00		
18	1	15	7	5	10	15	8	7	8	7	10	2.00		
19	1	9	6	4	8	3	5	8	6	10	3	2.00		
20	1	6	7	10	6	9	6	8	6	5	2	2.00		
21	1	13	7	8	7	9	6	3	6	6	4	2.00		
22	1	10	7	12	8	8	8	12	5	11	4	2.00		
23	1	10	9	10	10	11	11	14	11	11	10	1.00		
24	1	6	6	6	2	5	7	6	6	11	6	2.00		
25	1	14	7	6	9	10	6	7	6	8	5	2.00		

Highlight gender and send it into the Grouping variable cell. When you do so, you will note that after Gender is a set of () with question marks. Click on Define Range.

⁵<http://cnx.org/content/m40735/latest/14.2.png/image>



After clicking on Define Range, the following screen will appear: For the Minimum box, we will type in the number 1 because it represents boys. We will type in the number 2 in the Maximum box because it represents girls.

⁶ <http://cnx.org/content/m40735/latest/14.3.png/image>

The screenshot shows the SPSS Discriminant Analysis dialog box with the following settings:

- Grouping Variable: Gender(? ?)
- Enter in: Enter in
- Use stepwise selection: Use stepwise selection

The 'Discriminant Analysis: Define Range' sub-dialog box is open, showing:

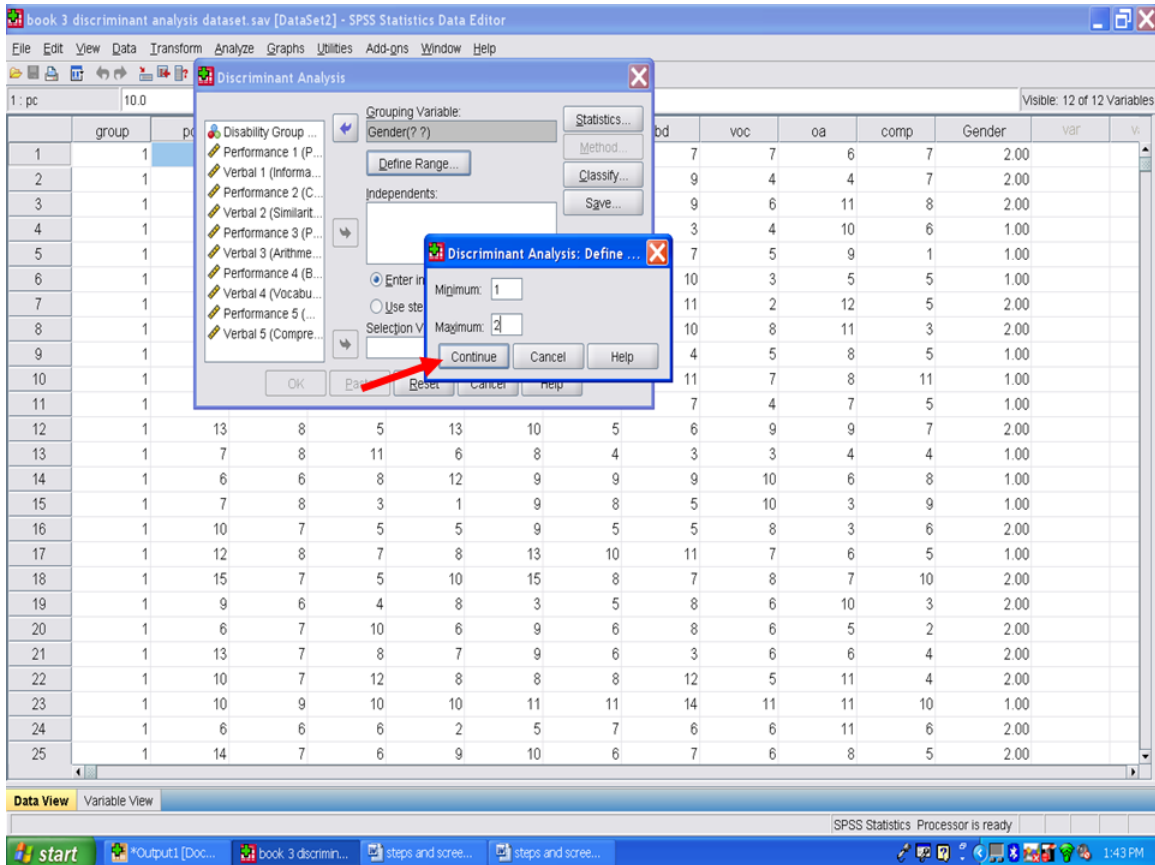
- Minimum:
- Maximum:
- Buttons: Continue, Cancel, Help

The background data table is as follows:

group	pc	bd	voc	oa	comp	Gender	var	v						
1	1	7	7	6	7	2.00								
2	1	9	4	4	7	2.00								
3	1	9	6	11	8	2.00								
4	1	3	4	10	6	1.00								
5	1	7	5	9	1	1.00								
6	1	10	3	5	5	1.00								
7	1	11	2	12	5	2.00								
8	1	10	8	11	3	2.00								
9	1	4	5	8	5	1.00								
10	1	11	7	8	11	1.00								
11	1	7	4	7	5	1.00								
12	1	13	8	5	13	10	5	6	9	9	7	2.00		
13	1	7	8	11	6	8	4	3	3	4	4	1.00		
14	1	6	6	8	12	9	9	9	10	6	8	1.00		
15	1	7	8	3	1	9	8	5	10	3	9	1.00		
16	1	10	7	5	5	9	5	5	8	3	6	2.00		
17	1	12	8	7	8	13	10	11	7	6	5	1.00		
18	1	15	7	5	10	15	8	7	8	7	10	2.00		
19	1	9	6	4	8	3	5	8	6	10	3	2.00		
20	1	6	7	10	6	9	6	8	6	5	2	2.00		
21	1	13	7	8	7	9	6	3	6	6	4	2.00		
22	1	10	7	12	8	8	8	12	5	11	4	2.00		
23	1	10	9	10	10	11	11	14	11	11	10	1.00		
24	1	6	6	6	2	5	7	6	6	11	6	2.00		
25	1	14	7	6	9	10	6	7	6	8	5	2.00		

After typing in 1 and 2, then click on Continue.

⁷ <http://cnx.org/content/m40735/latest/14.4.png/image>



Next we will send over our 10 dependent variables, the 10 subscales, to the Independents box. Highlight Performance 1 through Verbal 5; click on the middle arrow; send them to the Independents box.

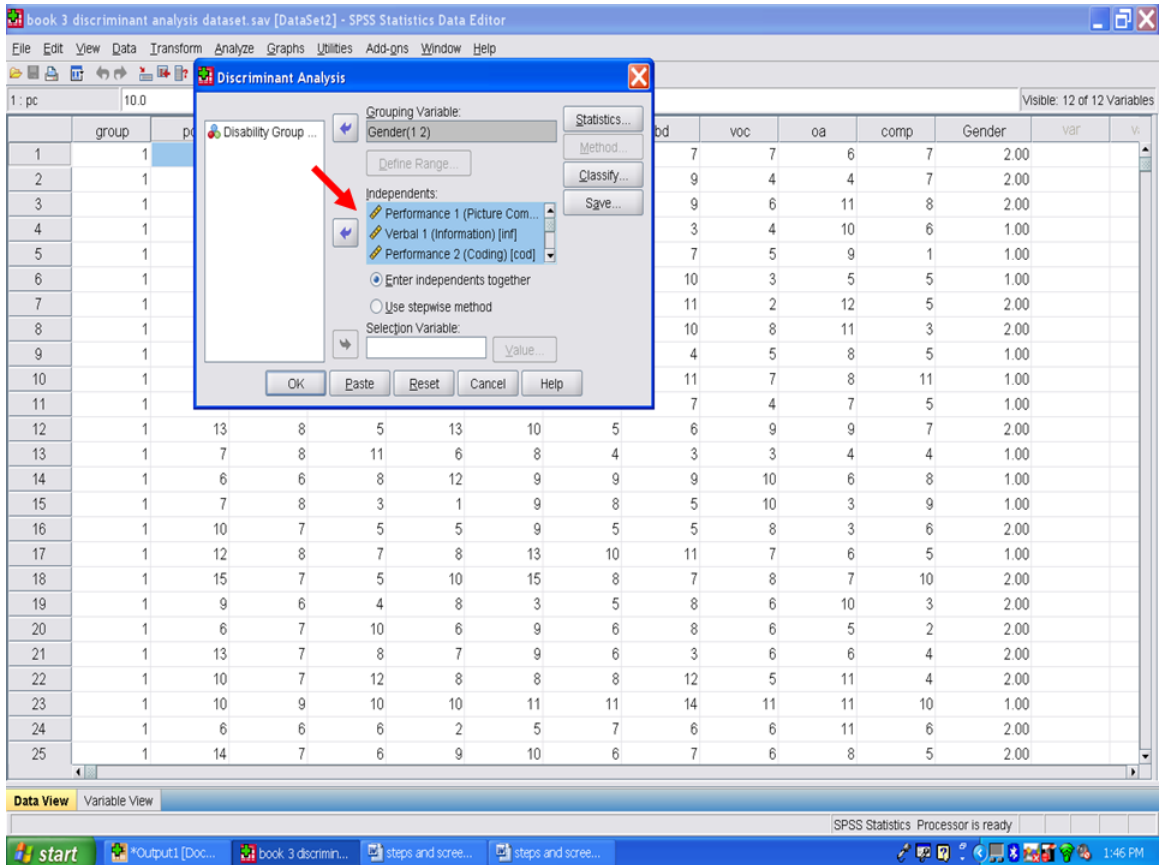
⁸ <http://cnx.org/content/m40735/latest/14.5.png/image>

The screenshot shows the SPSS Discriminant Analysis dialog box. The 'Grouping Variable' is 'Gender (1 2)'. The 'Independents' list includes 'Performance 1 (P...', 'Verbal 1 (Informa...', 'Performance 2 (C...', 'Verbal 2 (Similarit...', 'Performance 3 (P...', 'Verbal 3 (Arithme...', 'Performance 4 (B...', 'Verbal 4 (Vocabu...', 'Performance 5 (...', and 'Verbal 5 (Compre...'. The 'Enter independents together' option is selected. The 'Selection Variable' is empty. The background shows a data view with columns 'group', 'p', 'bd', 'voc', 'oa', 'comp', 'Gender', 'var', and 'v'. The 'Performance 5 (Compre...' variable is highlighted in the data view with a red arrow.

group	p	bd	voc	oa	comp	Gender	var	v					
1	1												
2	1												
3	1												
4	1												
5	1												
6	1												
7	1												
8	1												
9	1												
10	1												
11	1												
12	1	13	8	5	13	10	5	6	9	9	7	2.00	
13	1	7	8	11	6	8	4	3	3	4	4	1.00	
14	1	6	6	8	12	9	9	9	10	6	8	1.00	
15	1	7	8	3	1	9	8	5	10	3	9	1.00	
16	1	10	7	5	5	9	5	5	8	3	6	2.00	
17	1	12	8	7	8	13	10	11	7	6	5	1.00	
18	1	15	7	5	10	15	8	7	8	7	10	2.00	
19	1	9	6	4	8	3	5	8	6	10	3	2.00	
20	1	6	7	10	6	9	6	8	6	5	2	2.00	
21	1	13	7	8	7	9	6	3	6	6	4	2.00	
22	1	10	7	12	8	8	8	12	5	11	4	2.00	
23	1	10	9	10	10	11	11	14	11	11	10	1.00	
24	1	6	6	6	2	5	7	6	6	11	6	2.00	
25	1	14	7	6	9	10	6	7	6	8	5	2.00	

The screen should now look like the one below:

⁹ <http://cnx.org/content/m40735/latest/14.6.png/image>



Next we will change the specific type of discriminant analysis from the default of Enter Independents Together to Use Stepwise Method.

¹⁰<http://cnx.org/content/m40735/latest/14.7.png/image>

The screenshot displays the SPSS Discriminant Analysis dialog box. The 'Grouping Variable' is set to 'Gender (1 2)'. Under 'Independents', three variables are listed: 'Performance 1 (Picture Com...)', 'Verbal 1 (Information) [inf]', and 'Performance 2 (Coding) [cod]'. The 'Enter independents together' radio button is selected, with a red arrow pointing to it. The 'Use stepwise method' radio button is unselected. The 'Selection Variable' field is empty. The background shows a data view with columns for 'group', 'px', 'bd', 'voc', 'oa', 'comp', 'Gender', 'var', and 'v'.

group	px	bd	voc	oa	comp	Gender	var	v						
1	1	7	7	6	7	2.00								
2	1	9	4	4	7	2.00								
3	1	9	6	11	8	2.00								
4	1	3	4	10	6	1.00								
5	1	7	5	9	1	1.00								
6	1	10	3	5	5	1.00								
7	1	11	2	12	5	2.00								
8	1	10	8	11	3	2.00								
9	1	4	5	8	5	1.00								
10	1	11	7	8	11	1.00								
11	1	7	4	7	5	1.00								
12	1	13	8	5	13	10	5	6	9	9	7	2.00		
13	1	7	8	11	6	8	4	3	3	4	4	1.00		
14	1	6	6	8	12	9	9	9	10	6	8	1.00		
15	1	7	8	3	1	9	8	5	10	3	9	1.00		
16	1	10	7	5	5	9	5	5	8	3	6	2.00		
17	1	12	8	7	8	13	10	11	7	6	5	1.00		
18	1	15	7	5	10	15	8	7	8	7	10	2.00		
19	1	9	6	4	8	3	5	8	6	10	3	2.00		
20	1	6	7	10	6	9	6	8	6	5	2	2.00		
21	1	13	7	8	7	9	6	3	6	6	4	2.00		
22	1	10	7	12	8	8	8	12	5	11	4	2.00		
23	1	10	9	10	10	11	11	14	11	11	10	1.00		
24	1	6	6	6	2	5	7	6	6	11	6	2.00		
25	1	14	7	6	9	10	6	7	6	8	5	2.00		

¹¹ <http://cnx.org/content/m40735/latest/14.8.png/image>

Chapter 16

16. Discriminant Analysis: Part II¹



NOTE: This chapter is published by NCPEA Press² and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of education administration.

About the Authors

John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website (Writing and Statistical Help³) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

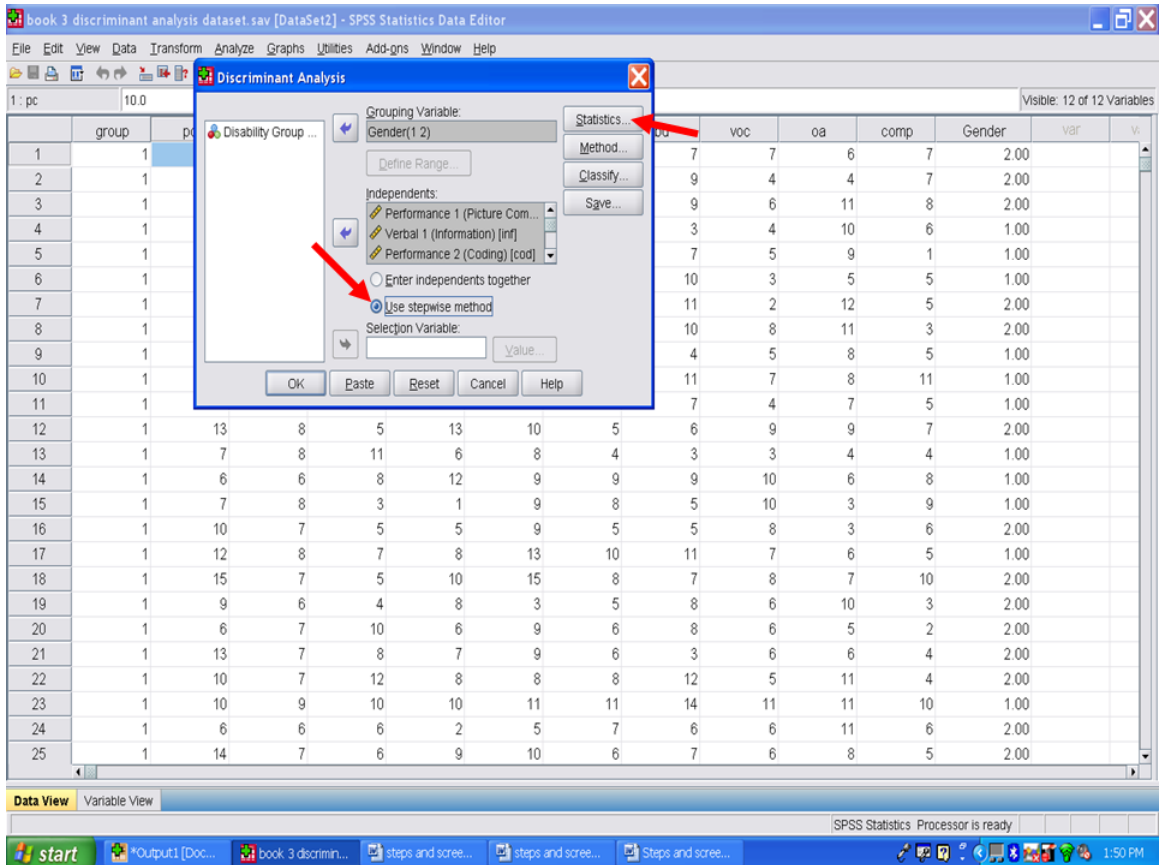
Ana Rojas-LeBouef is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. Dr. LeBoeuf recently completed her doctoral degree in Reading, where she conducted a 16-year analysis of Texas statewide data regarding the achievement gap. Her research interests lie in examining the inequities in achievement among ethnic groups. Dr. Rojas-LeBouef also assists students and faculty in their writing and statistical needs on the Writing and Statistical Help website.

Your screen should now look like the one below: Now click on Statistics.

¹This content is available online at <<http://cnx.org/content/m40736/1.2/>>.

²<http://www.ncpeapublications.org/books.html>

³<http://cnx.org/content/m40736/latest/www.writingandstatisticalhelp>



The following screen then appears.

Click on Means.

Click on Continue.

⁴<http://cnx.org/content/m40736/latest/15.1.png/image>

The screenshot shows the SPSS Discriminant Analysis dialog box with the Statistics sub-dialog open. The Statistics sub-dialog has the following options:

- Descriptives:**
 - Means
 - Univariate ANOVAs
 - Box's M
- Function Coefficients:**
 - Fisher's
 - Unstandardized
- Matrices:**
 - Within-groups correlation
 - Within-groups covariance
 - Separate-groups covariance
 - Total covariance

The 'Continue' button in the Statistics sub-dialog is highlighted with a red arrow. The background shows a data table with 25 rows and 12 columns, including variables like 'group', 'pc', 'bd', 'voc', 'oa', 'comp', 'Gender', 'var', and 'v'.

When the screen below appears, click on Method

⁵ <http://cnx.org/content/m40736/latest/15.2.png/image>

The screenshot shows the SPSS Discriminant Analysis dialog box. The 'Grouping Variable' is 'Gender (1 2)'. The 'Independents' list includes 'Performance 1 (Picture Com...)', 'Verbal 1 (Information) [inf]', and 'Performance 2 (Coding) [cod]'. The 'Method' button is highlighted with a red arrow. The background shows a data view with columns for 'group', 'px', 'voc', 'oa', 'comp', 'Gender', and 'var'.

group	px	voc	oa	comp	Gender	var						
1	1	7	7	6	7	2.00						
2	1	9	4	4	7	2.00						
3	1	9	6	11	8	2.00						
4	1	3	4	10	6	1.00						
5	1	7	5	9	1	1.00						
6	1	10	3	5	5	1.00						
7	1	11	2	12	5	2.00						
8	1	10	8	11	3	2.00						
9	1	4	5	8	5	1.00						
10	1	11	7	8	11	1.00						
11	1	7	4	7	5	1.00						
12	1	13	8	5	13	10	5	6	9	9	7	2.00
13	1	7	8	11	6	8	4	3	3	4	4	1.00
14	1	6	6	8	12	9	9	9	10	6	8	1.00
15	1	7	8	3	1	9	8	5	10	3	9	1.00
16	1	10	7	5	5	9	5	5	8	3	6	2.00
17	1	12	8	7	8	13	10	11	7	6	5	1.00
18	1	15	7	5	10	15	8	7	8	7	10	2.00
19	1	9	6	4	8	3	5	8	6	10	3	2.00
20	1	6	7	10	6	9	6	8	6	5	2	2.00
21	1	13	7	8	7	9	6	3	6	6	4	2.00
22	1	10	7	12	8	8	8	12	5	11	4	2.00
23	1	10	9	10	10	11	11	14	11	11	10	1.00
24	1	6	6	6	2	5	7	6	6	11	6	2.00
25	1	14	7	6	9	10	6	7	6	8	5	2.00

In the Method screen, we will use the default of Wilks' lambda.
Click on Continue.

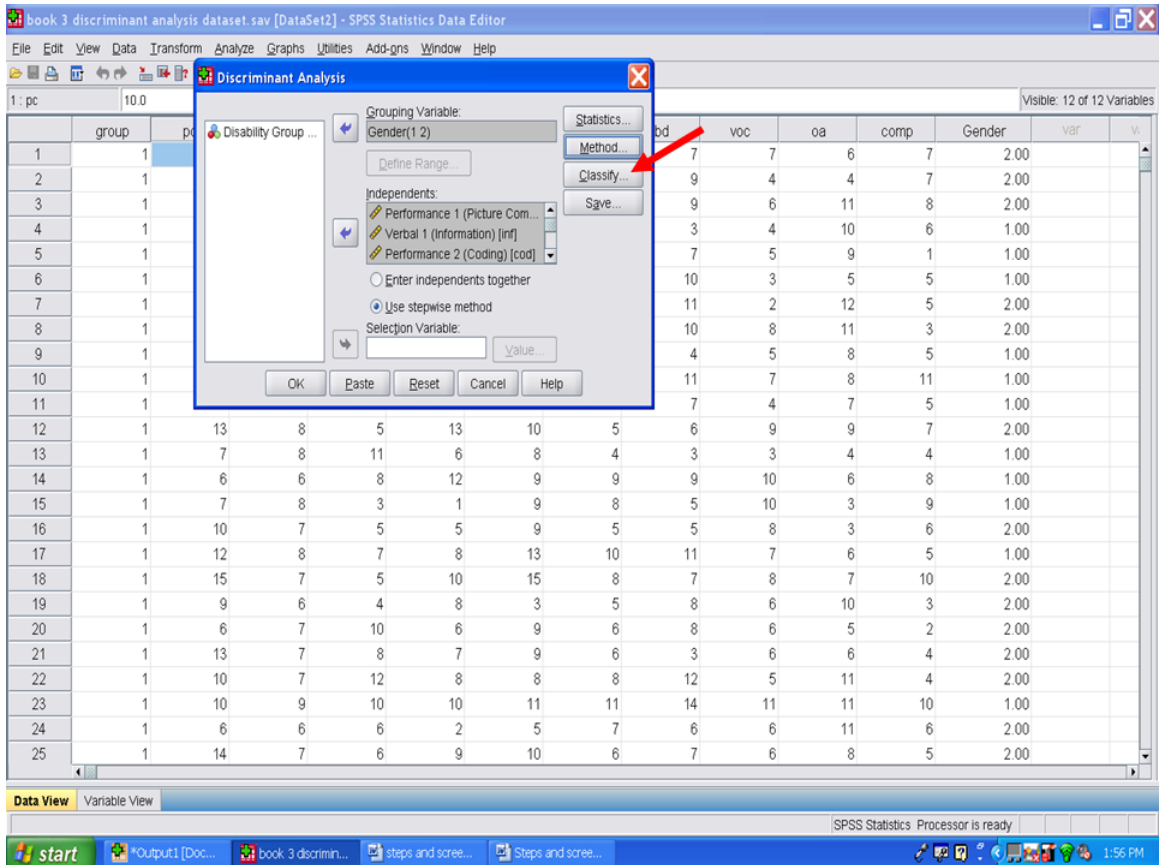
⁶ <http://cnx.org/content/m40736/latest/15.3.png/image>

The screenshot shows the SPSS 'Discriminant Analysis: Stepwise Method' dialog box. The 'Method' section includes radio buttons for 'Wilks' lambda', 'Unexplained variance', 'Mahalanobis distance', 'Smallest F ratio', and 'Bao's V'. The 'Criteria' section includes radio buttons for 'Use E value' (with 'Entry' at 3.84 and 'Removal' at 2.71) and 'Use probability of F' (with 'Entry' at .05 and 'Removal' at .10). The 'Display' section includes checkboxes for 'Summary of steps' (checked) and 'F for pairwise distances'. A red arrow points to the 'Continue' button.

group	pc	bd	voc	oa	comp	Gender	var	v						
1	1	7	7	6	7	2.00								
2	1	9	4	4	7	2.00								
3	1	9	6	11	8	2.00								
4	1	3	4	10	6	1.00								
5	1	7	5	9	1	1.00								
6	1	10	3	5	5	1.00								
7	1	11	2	12	5	2.00								
8	1	10	8	11	3	2.00								
9	1	4	5	8	5	1.00								
10	1	11	7	8	11	1.00								
11	1	7	4	7	5	1.00								
12	1	13	5	13	10	5	6	9	9	7	2.00			
13	1	7	8	11	6	8	4	3	3	4	4	1.00		
14	1	6	6	8	12	9	9	9	10	6	8	1.00		
15	1	7	8	3	1	9	8	5	10	3	9	1.00		
16	1	10	7	5	5	9	5	5	8	3	6	2.00		
17	1	12	8	7	8	13	10	11	7	6	5	1.00		
18	1	15	7	5	10	15	8	7	8	7	10	2.00		
19	1	9	6	4	8	3	5	8	6	10	3	2.00		
20	1	6	7	10	6	9	6	8	6	5	2	2.00		
21	1	13	7	8	7	9	6	3	6	6	4	2.00		
22	1	10	7	12	8	8	8	12	5	11	4	2.00		
23	1	10	9	10	10	11	11	14	11	11	10	1.00		
24	1	6	6	6	2	5	7	6	6	11	6	2.00		
25	1	14	7	6	9	10	6	7	6	8	5	2.00		

Then click on **Classify**.

⁷ <http://cnx.org/content/m40736/latest/15.4.png/image>



In the screen that appears, the All groups equal button is already clicked.
 Click on Summary table under the Display.
 Then click on Continue.

⁸ <http://cnx.org/content/m40736/latest/15.5.png/image>

book 3 discriminant analysis dataset.sav [DataSet2] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

Discriminant Analysis: Classification

Prior Probabilities

- All groups equal
- Compute from group sizes

Use Covariance Matrix

- Within-groups
- Separate-groups

Display

- Casewise results
- Limit cases to first:
- Summary table
- Leave-one-out classification
- Replace missing values with mean

Plots

- Combined-groups
- Separate-groups
- Territorial map

Continue Cancel Help

group	pc	bd	voc	oa	comp	Gender	var	v
1	1	7	7	6	7	2.00		
2	1	9	4	4	7	2.00		
3	1	9	6	11	8	2.00		
4	1	3	4	10	6	1.00		
5	1	7	5	9	1	1.00		
6	1	10	3	5	5	1.00		
7	1	11	2	12	5	2.00		
8	1	10	8	11	3	2.00		
9	1	4	5	8	5	1.00		
10	1	11	7	8	11	1.00		
11	1	7	4	7	5	1.00		
12	1	13	8	5	13	10	5	6
13	1	7	8	11	6	8	4	3
14	1	6	6	8	12	9	9	9
15	1	7	8	3	1	9	8	5
16	1	10	7	5	5	9	5	5
17	1	12	8	7	8	13	10	11
18	1	15	7	5	10	15	8	7
19	1	9	6	4	8	3	5	8
20	1	6	7	10	6	9	6	8
21	1	13	7	8	7	9	6	3
22	1	10	7	12	8	8	8	12
23	1	10	9	10	10	11	11	14
24	1	6	6	6	2	5	7	6
25	1	14	7	6	9	10	6	7

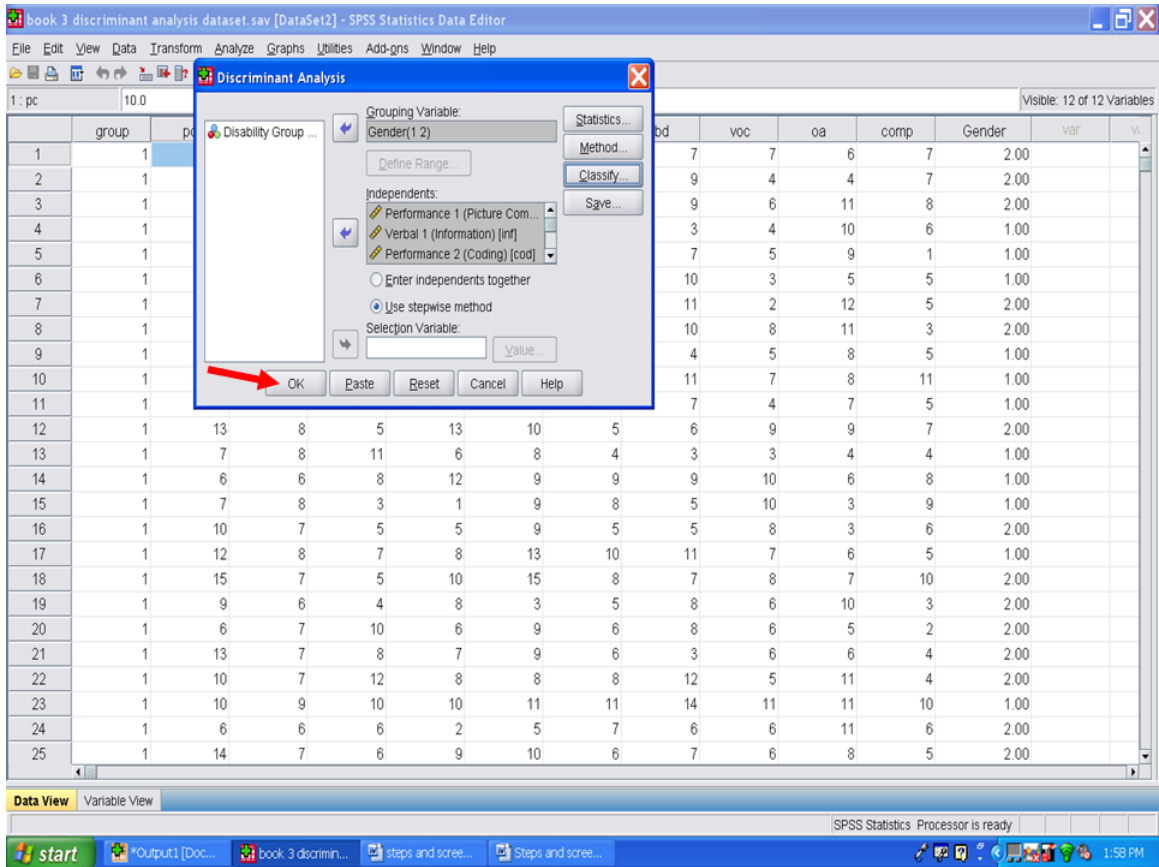
Data View Variable View

SPSS Statistics Processor is ready

start *Output1 [Doc... book 3 discrim... steps and scree... Steps and scree... 1:56 PM

We are now ready to have SPSS calculate this procedure.
Click on OK.

⁹<http://cnx.org/content/m40736/latest/15.6.png/image>



After clicking on OK, SPSS should send you to your output file. If not, click on the Output icon at the bottom of your screen.

¹⁰<http://cnx.org/content/m40736/latest/15.7.png/image>

The screenshot shows the SPSS Statistics Data Editor window. The title bar reads "book 3 discriminant analysis dataset.sav [DataSet2] - SPSS Statistics Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, and Help. The toolbar contains various icons for file operations and analysis. The main data grid has 25 rows and 14 columns. The columns are labeled: group, pc, inf, cod, sim, pa, ari, bd, voc, oa, comp, Gender, var, and v. The first row has values: 1, 10, 7, 8, 7, 10, 5, 7, 7, 6, 7, 2.00, and empty cells for the last two columns. A red arrow points to the value 10 in the 'pc' column of the first row. The status bar at the bottom indicates "SPSS Statistics Processor is ready" and the system clock shows "2:02 PM".

	group	pc	inf	cod	sim	pa	ari	bd	voc	oa	comp	Gender	var	v
1	1	10	7	8	7	10	5	7	7	6	7	2.00		
2	1	8	8	7	6	7	9	9	4	4	7	2.00		
3	1	10	8	2	9	10	11	9	6	11	8	2.00		
4	1	3	4	11	4	7	6	3	4	10	6	1.00		
5	1	5	7	17	6	8	10	7	5	9	1	1.00		
6	1	2	5	8	8	9	7	10	3	5	5	1.00		
7	1	7	4	5	8	7	11	11	2	12	5	2.00		
8	1	7	7	7	7	14	8	10	8	11	3	2.00		
9	1	9	6	10	6	9	8	4	5	8	5	1.00		
10	1	5	9	7	10	9	12	11	7	8	11	1.00		
11	1	6	7	5	8	7	7	7	4	7	5	1.00		
12	1	13	8	5	13	10	5	6	9	9	7	2.00		
13	1	7	8	11	6	8	4	3	3	4	4	1.00		
14	1	6	6	8	12	9	9	9	10	6	8	1.00		
15	1	7	8	3	1	9	8	5	10	3	9	1.00		
16	1	10	7	5	5	9	5	5	8	3	6	2.00		
17	1	12	8	7	8	13	10	11	7	6	5	1.00		
18	1	15	7	5	10	15	8	7	8	7	10	2.00		
19	1	9	6	4	8	3	5	8	6	10	3	2.00		
20	1	6	7	10	6	9	6	8	6	5	2	2.00		
21	1	13	7	8	7	9	6	3	6	6	4	2.00		
22	1	10	7	12	8	8	8	12	5	11	4	2.00		
23	1	10	9	10	10	11	11	14	11	11	10	1.00		
24	1	6	6	6	2	5	7	6	6	11	6	2.00		
25	1	14	7	6	9	10	6	7	6	8	5	2.00		

In the output screen, you should have the following output.
 The first table shows you how many of your participants' data were used in the analysis. In this example, we had a total of 52 participants whose data were not analyzed.

¹¹ <http://cnx.org/content/m40736/latest/15.8.png/image>

The screenshot displays the SPSS Statistics Viewer interface. The main window shows the output for a Discriminant analysis. The 'Analysis Case Processing Summary' table is highlighted with a red arrow, indicating the number of cases included and excluded from the analysis. The 'Group Statistics' table below it provides the number of cases for each group (Boys and Girls) across various performance and verbal tasks.

Analysis Case Processing Summary

Unweighted Cases	N	Percent
Valid	1130	95.6
Excluded		
Missing or out-of-range group codes	3	.3
At least one missing discriminating variable	49	4.1
Both missing or out-of-range group codes and at least one missing discriminating variable	0	.0
Total	52	4.4
Total	1182	100.0

Group Statistics

Gender of Persons in Study	Valid N (listwise)	
	Unweighted	Weighted
Boys		
Performance 1 (Picture Completion)	571	571.000
Verbal 1 (Information)	571	571.000
Performance 2 (Coding)	571	571.000
Verbal 2 (Similarities)	571	571.000
Performance 3 (Picture Arrangement)	571	571.000
Verbal 3 (Arithmetic)	571	571.000
Performance 4 (Block Design)	571	571.000
Verbal 4 (Vocabulary)	571	571.000
Performance 5 (Object Assembly)	571	571.000
Verbal 5 (Comprehension)	571	571.000
Girls		
Performance 1 (Picture Completion)	559	559.000
Verbal 1 (Information)	559	559.000
Performance 2 (Coding)	559	559.000

The second table in the output tells us we have 571 boys in group one and 559 girls in group two.

¹²<http://cnx.org/content/m40736/latest/15.9.png/image>

***Output1 [Document1] - SPSS Statistics Viewer**

File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help

Output

- Discriminant
 - Title
 - Notes
 - Active Dataset
 - Analysis Case Process
 - Group Statistics
 - Analysis 1
 - Title
 - Stepwise Statistics
 - Title
 - Variables Ent
 - Variables in th
 - Variables Not
 - Wilks' Lambd
 - Summary of Cano
 - Title
 - Eigenvalues
 - Wilks' Lambd
 - Standardized Canv
 - Structure Matrix
 - Functions at Group
 - Classification Stat
 - Title
 - Classification
 - Prior Probabil
 - Classification

Discriminant

[DataSet2] C:\WINNT\profiles\shzu\My Documents\John R Slate\Scholarly Works\Books\Book Two Advanced Statistics\book 3 discriminant analysis dataset

Analysis Case Processing Summary

Unweighted Cases	N	Percent
Valid	1130	95.6
Excluded		
Missing or out-of-range group codes	3	.3
At least one missing discriminating variable	49	4.1
Both missing or out-of-range group codes and at least one missing discriminating variable	0	.0
Total	52	4.4
Total	1182	100.0

Group Statistics

Gender of Persons in Study	Valid N (listwise)	
	Unweighted	Weighted
Boys		
Performance 1 (Picture Completion)	571	571.000
Verbal 1 (Information)	571	571.000
Performance 2 (Coding)	571	571.000
Verbal 2 (Similarities)	571	571.000
Performance 3 (Picture Arrangement)	571	571.000
Verbal 3 (Arithmetic)	571	571.000
Performance 4 (Block Design)	571	571.000
Verbal 4 (Vocabulary)	571	571.000
Performance 5 (Object Assembly)	571	571.000
Verbal 5 (Comprehension)	571	571.000
Girls		
Performance 1 (Picture Completion)	559	559.000
Verbal 1 (Information)	559	559.000
Performance 2 (Coding)	559	559.000

Information area

SPSS Statistics Processor is ready | H. 508, W. 1167 pt.

start *Output1 [Doc... book 3 discrimin... steps and scree... Steps and scree... 2:18 PM

¹³<http://cnx.org/content/m40736/latest/15.10.png/image>

Chapter 17

17. Discriminant Analysis: Part III¹



NOTE: This chapter is published by NCPEA Press² and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of education administration.

About the Authors

John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website (Writing and Statistical Help³) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

Ana Rojas-LeBoeuf is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. Dr. LeBoeuf recently completed her doctoral degree in Reading, where she conducted a 16-year analysis of Texas statewide data regarding the achievement gap. Her research interests lie in examining the inequities in achievement among ethnic groups. Dr. Rojas-LeBoeuf also assists students and faculty in their writing and statistical needs on the Writing and Statistical Help website.

The third table in the SPSS output for this canonical discriminant analysis is the Stepwise Statistics table. In this example, five of the 10 variables differentiated boys from girls.

¹This content is available online at <<http://cnx.org/content/m40737/1.2/>>.

²<http://www.ncpeapublications.org/books.html>

³<http://cnx.org/content/m40737/latest/www.writingandstatisticalhelp>

Analysis 1

Stepwise Statistics ←

Variables Entered/Removed^{a,b,c,d}

Step	Entered	Wilks' Lambda					Exact F		
		Statistic	df1	df2	df3	Statistic	df1	df2	Sig.
1	Verbal 3 (Arithmetic)	.961	1	1	1128.000	46.270	1	1128.000	.000
2	Performance 1 (Picture Completion)	.940	2	1	1128.000	35.762	2	1127.000	.000
3	Performance 2 (Coding)	.919	3	1	1128.000	33.010	3	1126.000	.000
4	Verbal 4 (Vocabulary)	.904	4	1	1128.000	29.892	4	1125.000	.000
5	Verbal 1 (Information)	.894	5	1	1128.000	26.521	5	1124.000	.000

At each step, the variable that minimizes the overall Wilks' Lambda is entered.

- Maximum number of steps is 20.
- Minimum partial F to enter is 3.84.
- Maximum partial F to remove is 2.71.
- F level, tolerance, or VIN insufficient for further computation.

Variables in the Analysis

Step		Tolerance	F to Remove	Wilks' Lambda
1	Verbal 3 (Arithmetic)	1.000	46.270	.961
2	Verbal 3 (Arithmetic)	.866	66.451	.996
2	Performance 1 (Picture Completion)	.866	24.298	.961
3	Verbal 3 (Arithmetic)	.812	41.561	.953
3	Performance 1 (Picture Completion)	.852	30.085	.944
3	Performance 2 (Coding)	.891	25.924	.940
4	Verbal 3 (Arithmetic)	.710	19.124	.919
4	Performance 1 (Picture Completion)	.777	42.617	.938
4	Performance 2 (Coding)	.887	22.401	.922

The next two tables indicate which variables remained in the discriminant analysis and which variables were not utilized. The variables that were not utilized did not contribute statistically significantly to differentiating boys from girls. The process of identifying variables in the analysis and variables not in the analysis is depicted for each of the five steps.

⁴<http://cnx.org/content/m40737/latest/16.1.png/image>

The screenshot displays the SPSS Statistics Viewer interface. The main window shows two tables: 'Variables in the Analysis' and 'Variables Not in the Analysis'. The 'Variables in the Analysis' table has columns for Step, Tolerance, F to Remove, and Wilks' Lambda. The 'Variables Not in the Analysis' table has columns for Step, Tolerance, Min. Tolerance, F to Enter, and Wilks' Lambda. Red arrows point to the 'F to Remove' and 'F to Enter' columns in their respective tables.

Step		Tolerance	F to Remove	Wilks' Lambda
1	Verbal 3 (Arithmetic)	1.000	46.270	
2	Verbal 3 (Arithmetic)	.866	66.451	.996
	Performance 1 (Picture Completion)	.866	24.298	.961
3	Verbal 3 (Arithmetic)	.812	41.561	.953
	Performance 1 (Picture Completion)	.852	30.085	.944
	Performance 2 (Coding)	.891	25.924	.940
4	Verbal 3 (Arithmetic)	.710	19.124	.919
	Performance 1 (Picture Completion)	.777	42.617	.938
	Performance 2 (Coding)	.887	22.401	.922
	Verbal 4 (Vocabulary)	.699	18.959	.919
5	Verbal 3 (Arithmetic)	.657	26.455	.916
	Performance 1 (Picture Completion)	.751	32.684	.920
	Performance 2 (Coding)	.880	24.945	.914
	Verbal 4 (Vocabulary)	.525	30.298	.919
	Verbal 1 (Information)	.478	11.881	.904

Step		Tolerance	Min. Tolerance	F to Enter	Wilks' Lambda
0	Performance 1 (Picture Completion)	1.000	1.000	4.795	.996
	Verbal 1 (Information)	1.000	1.000	4.018	.996
	Performance 2 (Coding)	1.000	1.000	41.626	.964
	Verbal 2 (Similarities)	1.000	1.000	11.807	.990
	Performance 3 (Picture Arrangement)	1.000	1.000	5.937	.995
	Verbal 3 (Arithmetic)	1.000	1.000	46.270	.961
	Performance 4 (Block Design)	1.000	1.000	5.498	.995
	Verbal 4 (Vocabulary)	1.000	1.000	35.587	.969
	Performance 5 (Object Assembly)	1.000	1.000	.228	1.000
	Verbal 5 (Comprehension)	1.000	1.000	18.444	.984

If you recall from the beginning of this chapter, one of the assumptions underlying use of a discriminant analysis procedure is tolerance values. The closer to 0 the tolerance values are, the more likely it is that the matrix is ill-conditioned and that overlap exists among the variables in the equation. In Step 5 below, you will note that all of the tolerance values are above 0, ranging from a low of .478 for Verbal 1 (Information) to a high of .880 for Performance 2 (Coding). Therefore, this assumption has not been violated.

⁵<http://cnx.org/content/m40737/latest/16.2.png/image>

The screenshot shows the SPSS Statistics Viewer window with the following data tables:

Variables in the Analysis

Step		Tolerance	F to Remove	Wilks' Lambda
1	Verbal 3 (Arithmetic)	1.000	46.270	
2	Verbal 3 (Arithmetic)	.866	66.451	.996
	Performance 1 (Picture Completion)	.866	24.298	.961
3	Verbal 3 (Arithmetic)	.812	41.561	.953
	Performance 1 (Picture Completion)	.852	30.085	.944
	Performance 2 (Coding)	.891	25.924	.940
4	Verbal 3 (Arithmetic)	.710	19.124	.919
	Performance 1 (Picture Completion)	.777	42.617	.938
	Performance 2 (Coding)	.887	22.401	.922
	Verbal 4 (Vocabulary)	.699	11.959	.919
5	Verbal 3 (Arithmetic)	.657	26.455	.916
	Performance 1 (Picture Completion)	.751	32.684	.920
	Performance 2 (Coding)	.880	24.945	.914
	Verbal 4 (Vocabulary)	.525	30.298	.919
	Verbal 1 (Information)	.478	11.881	.904

Variables Not in the Analysis

Step		Tolerance	Min. Tolerance	F to Enter	Wilks' Lambda
0	Performance 1 (Picture Completion)	1.000	1.000	4.795	.996
	Verbal 1 (Information)	1.000	1.000	4.018	.996
	Performance 2 (Coding)	1.000	1.000	41.626	.964
	Verbal 2 (Similarities)	1.000	1.000	11.807	.990
	Performance 3 (Picture Arrangement)	1.000	1.000	5.937	.995
	Verbal 3 (Arithmetic)	1.000	1.000	46.270	.961
	Performance 4 (Block Design)	1.000	1.000	5.498	.995
	Verbal 4 (Vocabulary)	1.000	1.000	35.587	.969
	Performance 5 (Object Assembly)	1.000	1.000	.228	1.000
	Verbal 5 (Comprehension)	1.000	1.000	18.444	.984

Under the Wilks' lambda table is a Summary of Canonical Discriminant Functions table labeled Eigenvalues. In this table is the canonical correlation, or R_c , which in this example is .325. This R_c value reflects the correlation between the groups and the discriminant function formed by the five variables that were statistically significant. The R_c may be interpreted as a Pearson correlation of the discriminant scores with the grouping variable. This information will be reported in your Results section.

⁶ <http://cnx.org/content/m40737/latest/16.3.png/image>

SPSS Statistics Processor is ready | H: 26, W: 1222 pt | 2:26 PM

SPSS Statistics Viewer - *Output1 [Document1]

File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help

4 | discriminant

Notes

Analysis 1

Title

Stepwise Statistics

Title

Variables Entered/Removed

Variables in the Analysis

Variables Not in the Analysis

Wilks' Lambda

Summary of Canonical Discriminant Functions

Title

Eigenvalues

Wilks' Lambda

Standardized Canonical Discriminant Function Coefficients

Structure Matrix

Functions at Group Centroids

Classification Statistics

Title

Classification Process

Prior Probabilities for Groups

Classification Results

Wilks' Lambda

Step	Number of Variables	Lambda	df1	df2	df3	Exact F			
						Statistic	df1	df2	Sig.
1	1	.961	1	1	1128	46.270	1	1128.000	.000
2	2	.940	2	1	1128	35.762	2	1127.000	.000
3	3	.919	3	1	1128	33.010	3	1126.000	.000
4	4	.904	4	1	1128	28.892	4	1125.000	.000
5	5	.894	5	1	1128	26.521	5	1124.000	.000

Summary of Canonical Discriminant Functions

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.118 ^a	100.0	100.0	.325

a. First 1 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.894	125.516	5	.000

Standardized Canonical Discriminant Function Coefficients

	Function
	1
Performance 1 (Picture Completion)	-.597
Verbal 1 (Information)	-.456
Performance 2 (Coding)	.483
Verbal 3 (Arithmetic)	.576
Verbal 4 (Vocabulary)	.688

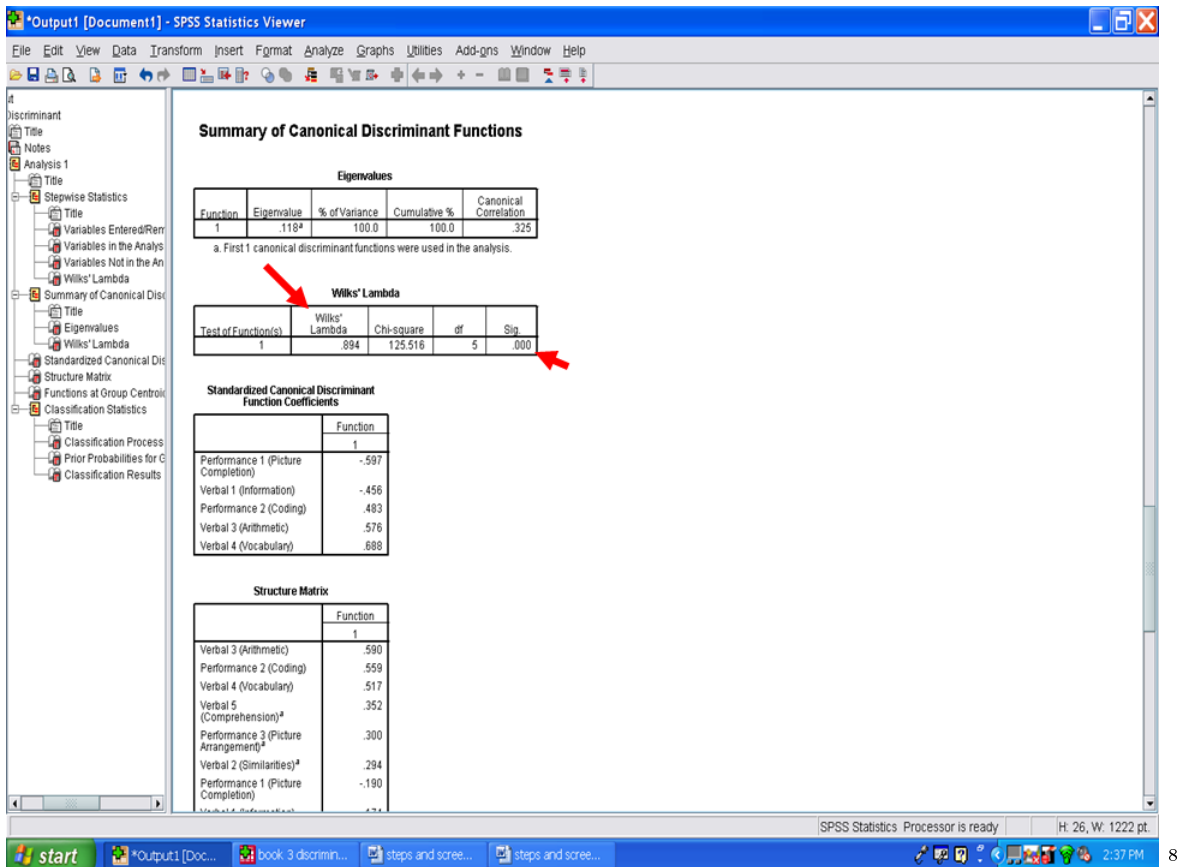
Structure Matrix

	Function
	1

start | *Output1 [Doc... | book 3 discrimin... | steps and scree... | steps and scree... | 7

In the table below the Eigenvalues one is a table labeled Wilks' Lambda. This measure is what is used to determine whether the discriminant function is statistically significant in differentiating group membership. In our example, the Wilks' Lambda of .894 is statistically significant at the .001 level. This result means that the discriminant function, consisting of five variables, statistically significantly differentiated the group of boys from the group of girls. This information will be reported in the Results section.

⁷ <http://cnx.org/content/m40737/latest/16.4.png/image>



The table immediately below the Wilks' Lambda table is labeled Standardized Canonical Discriminant Function Coefficients. These values are also referred to as Standardized Discriminant Coefficients. Contained in this table are the values that depict the relative importance of each variable in differentiating group membership. These coefficients are comparable to Beta Weights in multiple regression. In this example, Verbal 4 (Vocabulary) is more important, with a coefficient of .688, than is Performance 2 (Coding), with a coefficient of .483, in differentiating boys from girls. These coefficients should be provided in your Results section.

⁸ <http://cnx.org/content/m40737/latest/16.5.png/image>

The screenshot displays the SPSS Statistics Viewer interface. The main window shows the following tables:

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.894	125.516	5	.000

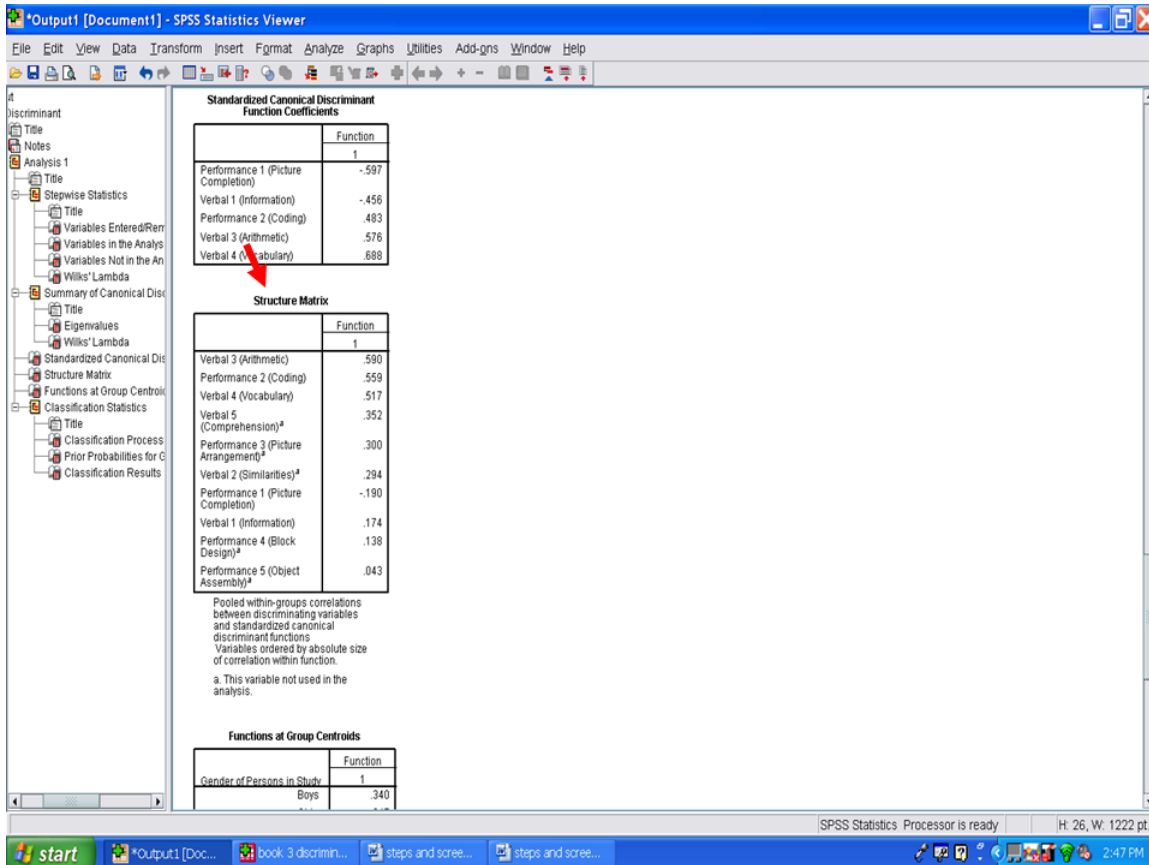
	Function
Performance 1 (Picture Completion)	-.597
Verbal 1 (Information)	-.456
Performance 2 (Coding)	.483
Verbal 3 (Arithmetic)	.576
Verbal 4 (Vocabulary)	.688

	Function
Verbal 3 (Arithmetic)	.590
Performance 2 (Coding)	.559
Verbal 4 (Vocabulary)	.517
Verbal 5 (Comprehension) ^a	.352
Performance 3 (Picture Arrangement) ^a	.300
Verbal 2 (Similarities) ^a	.294
Performance 1 (Picture Completion)	-.190
Verbal 1 (Information)	.174
Performance 4 (Block Design) ^a	.138
Performance 5 (Object Assembly) ^a	.043

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions. Variables ordered by absolute size of correlation within function.
a. This variable not used in the analysis.

Next, we examine the table underneath the Standardized Canonical Discriminant Function Coefficients, the one labeled as Structure Matrix. Each of these values depicts the relationship of that variable with the discriminant function scores. Each value is analogous to a correlation coefficient. In the example below, Verbal 3 (Arithmetic) has a correlation of .59 with the discriminant function scores. Similarly, Performance 2 (Coding) and Verbal 4 (Vocabulary) have correlations of .559 and .517, respectively, with the discriminant function scores. These values may be interpreted as factor loadings in assigning a name to the discriminant function. As such, the first three variables just mentioned contribute more to the discriminant function than does the variable of Performance 5 (Object Assembly).

⁹<http://cnx.org/content/m40737/latest/16.6.png/image>



Under this table is a table labeled Functions at Group Centroids. These values are z-scores, with a M of 0 and a SD of 1. The group centroid indicates the most typical location of any person from a particular group. For this example, the group centroid for boys is .34 and the group centroid for girls is -.347. Using these values we can compare how far apart boys and girls are differentiated by the statistically significant discriminant function in our example. The further apart these two means or centroids are, the better the separation or differentiation between the two groups. The centroids should be reported in the Results section.

¹⁰<http://cnx.org/content/m40737/latest/16.7.png/image>

Functions at Group Centroids

Gender of Persons in Study	Function
Boys	.340
Girls	-.347

Unstandardized canonical discriminant functions evaluated at group means

Classification Statistics

Classification Processing Summary

Processed	1182
Excluded	0
Missing or out-of-range group codes	
At least one missing discriminating variable	46
Used in Output	1136

Prior Probabilities for Groups

Gender of Persons in Study	Prior	Cases Used in Analysis	
		Unweighted	Weighted
Boys	.500	571	571.000
Girls	.500	559	559.000
Total	1.000	1130	1130.000

Classification Results^a

Gender of Persons in Study		Predicted Group Membership		Total
		Boys	Girls	
Original	Count	340	233	573
	Boys	193	367	560
	Girls	0	3	3
%	Boys	59.3	40.7	100.0
	Girls	34.5	65.5	100.0
	Ungrouped cases	.0	100.0	100.0

a. 62.4% of original grouped cases correctly classified.

You have now successfully conducted a canonical discriminant analysis.

¹¹ <http://cnx.org/content/m40737/latest/16.8.png/image>

Chapter 18

18. Multiple Regression: Assumptions¹



NOTE: This chapter is published by NCPEA Press² and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of education administration.

About the Authors

John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website (Writing and Statistical Help³) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

Ana Rojas-LeBouef is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. Dr. LeBoeuf recently completed her doctoral degree in Reading, where she conducted a 16-year analysis of Texas statewide data regarding the achievement gap. Her research interests lie in examining the inequities in achievement among ethnic groups. Dr. Rojas-LeBouef also assists students and faculty in their writing and statistical needs on the Writing and Statistical Help website.

In this set of steps, readers will learn how to conduct a multiple regression procedure. For detailed information regarding the assumptions underlying use of a multiple regression analysis, readers are referred to the Hyperstats Online Statistics Textbook at <http://davidmlane.com/hyperstat/>⁴; to the *Electronic Statistics Textbook* (2011)

¹This content is available online at <<http://cnx.org/content/m40738/1.2/>>.

²<http://www.ncpeapublications.org/books.html>

³<http://cnx.org/content/m40738/latest/www.writingandstatisticalhelp>

⁴<http://davidmlane.com/hyperstat/>

at <http://www.statsoft.com/textbook/>⁵ ; or to Andy Field's (2009) *Discovering Statistics Using SPSS* at http://www.amazon.com/Discovering-Statistics-Introducing-Statistical-Method/dp/1847879071/ref=sr_1_1?s=books&ie=UTF8&qid=1304967862&sr=1-1⁶

Research questions for which a multiple regression analysis is appropriate involve determining variables that predict a continuous variable. For example, if you want to predict the life expectancy of individuals and you have archival data available (e.g., health history, gender), then a multiple regression analysis procedure could be utilized. Such a procedure could identify specific variables that are predictive of a long or of a short life expectancy. As such, interventions could be developed and targeted toward the variables that were statistically significant predictors. Other sample research questions for which a multiple regression analysis might be appropriate: (a) What factors predict high scores on a scholastic aptitude measure?; and (b) What factors are predictive of high scores on a life satisfaction scale?

For the purposes of this chapter, our research question is: "What scholastic variables predict students' Full Scale IQ?"

Have your data set open in SPSS. In this dataset, we will determine what variables, if any, are predictive of students' Wechsler Full Scale IQ 3 (wfsiq). The 10 variables that we will use are: Picture Completion (pc), Information (inf), Coding (cod), Similarities (sim), Picture Arrangement (pa), Arithmetic (ari), Block Design (bd), Vocabulary (voc), Object Assembly (oa), and Comprehension (comp).

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	group	Numeric	1	0	Reading Group Membership	{1, Excele...	None	8	Right	Nominal
2	age	Numeric	2	0	Age of Students in Study	None	None	8	Right	Ordinal
3	recodage	Numeric	8	2	Recoded Age Into Categories	{1.00, Age...	None	8	Right	Nominal
4	wfsiq	Numeric	3	0	Wechsler Full Scale IQ 3	None	None	8	Right	Scale
5	wiviq	Numeric	3	0	Verbal IQ (Wechsler Verbal Intelligence 3)	None	None	8	Right	Scale
6	wipiq	Numeric	3	0	Performance IQ (Wechsler Performance Intell...	None	None	8	Right	Scale
7	pc	Numeric	2	0	Performance 1 (Picture Completion)	None	None	8	Right	Scale
8	inf	Numeric	2	0	Verbal 1 (Information)	None	None	8	Right	Scale
9	cod	Numeric	2	0	Performance 2 (Coding)	None	None	8	Right	Scale
10	sim	Numeric	2	0	Verbal 2 (Similarities)	None	None	8	Right	Scale
11	pa	Numeric	2	0	Performance 3 (Picture Arrangement)	None	None	8	Right	Scale
12	ari	Numeric	2	0	Verbal 3 (Arithmetic)	None	None	8	Right	Scale
13	bd	Numeric	2	0	Performance 4 (Block Design)	None	None	8	Right	Scale
14	voc	Numeric	2	0	Verbal 4 (Vocabulary)	None	None	8	Right	Scale
15	oa	Numeric	2	0	Performance 5 (Object Assembly)	None	None	8	Right	Scale
16	comp	Numeric	2	0	Verbal 5 (Comprehension)	None	None	8	Right	Scale
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										

As with every statistical procedure, we need to check the underlying assumptions. One

⁵<http://www.statsoft.com/textbook/>

⁶http://www.amazon.com/Discovering-Statistics-Introducing-Statistical-Method/dp/1847879071/ref=sr_1_1?s=books&ie=UTF8&qid=1304967862&sr=1-1

1

⁷<http://cnx.org/content/m40738/latest/17.1.png/image>

assumption involves the data being normally distributed. To check this assumption, we recommend that you calculate the standardized skewness coefficients and the standardized kurtosis coefficients, as discussed in other chapters.

* **Skewness** [Note. Skewness refers to the extent to which the data are normally distributed around the mean. Skewed data involve having either mostly high scores with a few low ones or having mostly low scores with a few high ones.] Readers are referred to the following sources for a more detailed definition of skewness: http://www.statistics.com/index.php?page=glossary&term_id=356⁸ and <http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb>⁹

To standardize the skewness value so that its value can be constant across datasets and across studies, the following calculation must be made: Take the skewness value from the SPSS output and divide it by the Std. error of skewness. If the resulting calculation is within -3 to +3, then the skewness of the dataset is within the range of normality (Onwuegbuzie & Daniel, 2002). If the resulting calculation is outside of this +/-3 range, the dataset is not normally distributed.

* **Kurtosis** [Note. Kurtosis also refers to the extent to which the data are normally distributed around the mean. This time, the data are piled up higher than normal around the mean or piled up higher than normal at the ends of the distribution.] Readers are referred to the following sources for a more detailed definition of kurtosis: http://www.statistics.com/index.php?page=glossary&term_id=326¹⁰ and <http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb>¹¹

To standardize the kurtosis value so that its value can be constant across datasets and across studies, the following calculation must be made: Take the kurtosis value from the SPSS output and divide it by the Std. error of kurtosis. If the resulting calculation is within -3 to +3, then the kurtosis of the dataset is within the range of normality (Onwuegbuzie & Daniel, 2002). If the resulting calculation is outside of this +/-3 range, the dataset is not normally distributed.

Now that you have verified that your data are normally distributed, the extent to which linearity is present between each of the 10 independent variables listed above and the dependent variable of Full Scale IQ must be determined. For linearity, we will have SPSS conduct scatterplots for each IV and DV pair.

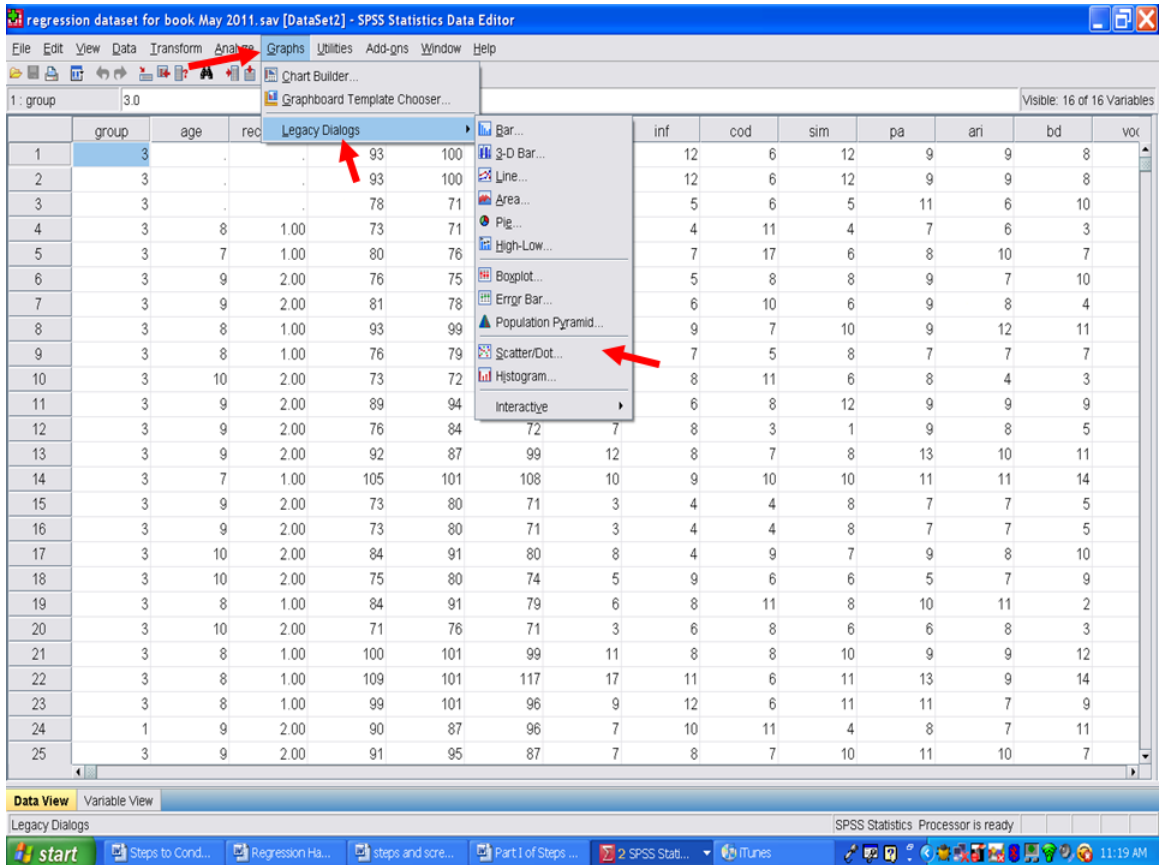
- ✓ **Graphs**
- ✓ **Legacy Dialogs**
- ✓ **Scatter/Dot**

⁸http://www.statistics.com/index.php?page=glossary&term_id=356

⁹<http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb>

¹⁰http://www.statistics.com/index.php?page=glossary&term_id=326

¹¹<http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb>



After clicking on Scatter/Dot, the following screen will appear. The Simple Scatter icon should be highlighted. If not, click on it.

✓ Click on Define

¹²<http://cnx.org/content/m40738/latest/17.2.png/image>

The screenshot shows the SPSS Statistics Data Editor window with the 'Scatter/Dot' dialog box open. The dialog box has several options: 'Simple Scatter', 'Matrix Scatter', 'Simple Dot', 'Overlay Scatter', and '3-D Scatter'. The 'Simple Scatter' option is selected. The 'Define' button is highlighted with a red arrow. The background shows a data table with columns for group, age, rec, pc, inf, cod, sim, pa, ari, bd, and vox.

group	age	rec	pc	inf	cod	sim	pa	ari	bd	vox
1	3		8	12	6	12	9	9	8	
2	3		8	12	6	12	9	9	8	
3	3		9	5	6	5	11	6	10	
4	3	8	3	4	11	4	7	6	3	
5	3	7	5	7	17	6	8	10	7	
6	3	9	2	5	8	8	9	7	10	
7	3	9	9	6	10	6	9	8	4	
8	3	8	5	9	7	10	9	12	11	
9	3	8	6	7	5	8	7	7	7	
10	3	10	7	8	11	6	8	4	3	
11	3	9	6	6	8	12	9	9	9	
12	3	9	7	8	3	1	9	8	5	
13	3	9	12	8	7	8	13	10	11	
14	3	7	10	9	10	10	11	11	14	
15	3	9	3	4	4	8	7	7	5	
16	3	9	3	4	4	8	7	7	5	
17	3	10	8	4	9	7	9	8	10	
18	3	10	5	9	6	6	5	7	9	
19	3	8	6	8	11	8	10	11	2	
20	3	10	3	6	8	6	6	8	3	
21	3	8	11	8	8	10	9	9	12	
22	3	8	17	11	6	11	13	9	14	
23	3	8	9	12	6	11	11	7	9	
24	1	9	7	10	11	4	8	7	11	
25	3	9	7	8	7	10	11	10	7	

The following screen should now be present.

¹³<http://cnx.org/content/m40738/latest/17.3.png/image>

The screenshot shows the SPSS Simple Scatterplot dialog box. The Y Axis is empty, and the X Axis is empty. The Set Markers by and Label Cases by fields are also empty. The Panel by section has Rows and Columns fields, both empty. The Template section has a checkbox for 'Use chart specifications from:' which is unchecked. The data view window in the background shows a table with columns: group, age, rec, od, sim, pa, ari, bd, vox. The data rows are numbered 1 through 25.

group	age	rec	od	sim	pa	ari	bd	vox
1	3							
2	3							
3	3							
4	3	8						
5	3	7						
6	3	9						
7	3	9						
8	3	8						
9	3	8						
10	3	10						
11	3	9						
12	3	9						
13	3	9						
14	3	7						
15	3	9						
16	3	9						
17	3	10						
18	3	10						
19	3	8						
20	3	10						
21	3	8	1.00	100	101	99	11	8
22	3	8	1.00	109	101	117	17	11
23	3	8	1.00	99	101	96	9	12
24	1	9	2.00	90	87	96	7	10
25	3	9	2.00	91	95	87	7	8

✓ Drag one of the two variables of interest to the first box (Y axis) on the right hand side and the other variable of interest to the second box (X axis) on the right hand side. It does not matter which variable goes in the X or Y axis because your scatterplot results will be the same. For our purposes, we will place the variable we are trying to predict, Wechsler Full Scale IQ 3, in the Y Axis box and one of the variables (i.e., Performance 1) we will use to try to predict it.

✓ Once you have a variable in each of the first two boxes, click on the OK tab on the bottom left hand corner of the screen.

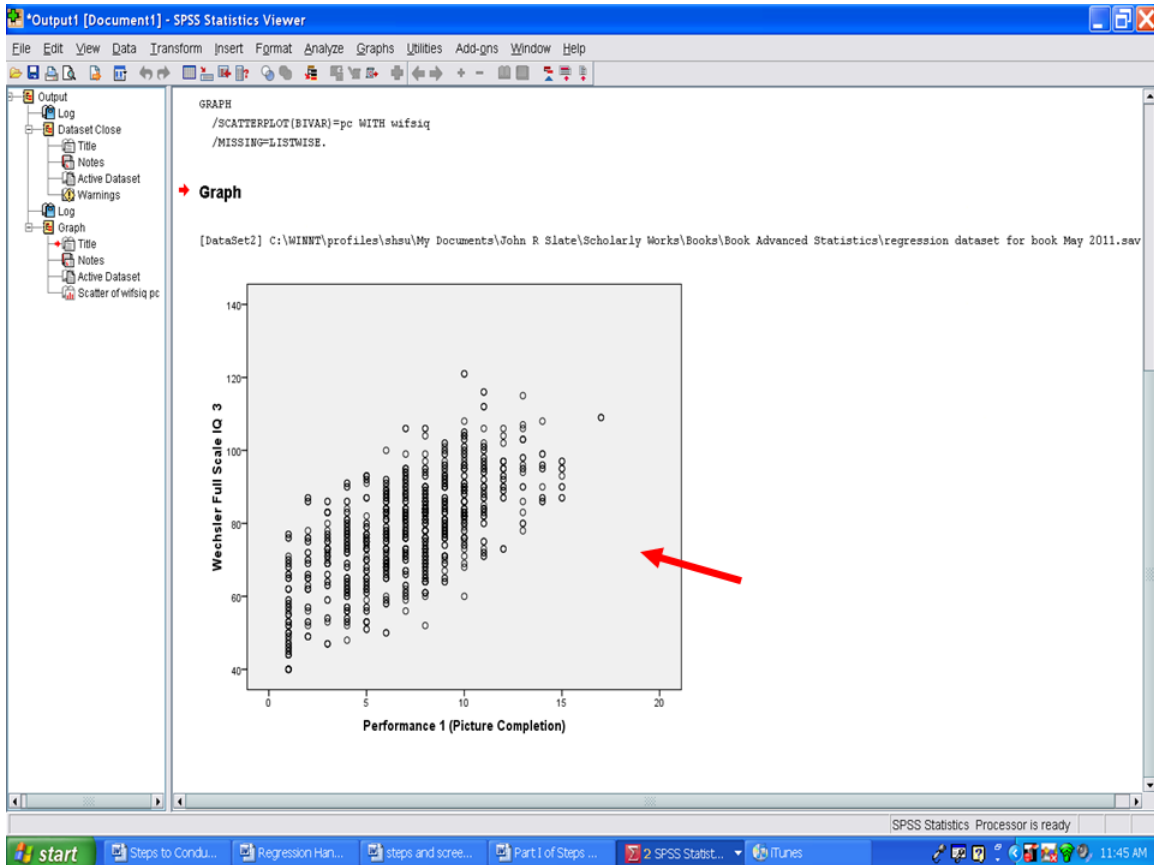
¹⁴<http://cnx.org/content/m40738/latest/17.4.png/image>

The screenshot shows the SPSS Simple Scatterplot dialog box. The Y-axis is set to 'Wechsler Full Scale IQ 3 [wfsiq]' and the X-axis is 'Performance 1 (Picture Comple...'. The dialog box is overlaid on a data view window showing a table with columns 'group', 'age', and 'rec'.

group	age	rec											
1	3												
2	3												
3	3												
4	3	8											
5	3	7											
6	3	9											
7	3	9											
8	3	8											
9	3	8											
10	3	10											
11	3	9											
12	3	9											
13	3	9											
14	3	7											
15	3	9											
16	3	9											
17	3	10											
18	3	10											
19	3	8											
20	3	10											
21	3	8	1.00	100	101	99	11	8	8	10	9	9	12
22	3	8	1.00	109	101	117	17	11	6	11	13	9	14
23	3	8	1.00	99	101	96	9	12	6	11	11	7	9
24	1	9	2.00	90	87	96	7	10	11	4	8	7	11
25	3	9	2.00	91	95	87	7	8	7	10	11	10	7

✓ Look at the scatterplot to determine whether a linear relationship is present. In the screenshot below, the relationship is very clearly linear.

¹⁵<http://cnx.org/content/m40738/latest/17.5.png/image>



You will need to repeat this process, for this example, nine more times. Leave the dependent variable of Wechsler Full Scale IQ 3 in the Y Axis box and replace the variable in the X Axis box with the next variable (i.e., Verbal 1). Then click on OK.

¹⁶<http://cnx.org/content/m40738/latest/17.6.png/image>

The screenshot shows the SPSS 'Simple Scatterplot' dialog box. The Y-axis is set to 'Wechsler Full Scale IQ 3 [wfsiq]' and the X-axis is set to 'Performance 1 (Picture Comple...'. A red arrow points to 'Verbal 1 (Informa...' in the list of variables. The background shows a data table with columns 'group', 'age', 'recod...', 'sim', 'pa', 'ari', 'bd', 'vo...'. The status bar at the bottom indicates 'SPSS Statistics Processor is ready' and the time is 11:47 AM.

After you have verified that linearity is present for each independent variable with the dependent variable, we will examine the extent to which multicollinearity is not present. Multicollinearity refers to having variables that are highly correlated with each other. When variables are highly correlated in a multiple regression analysis, the unique contribution of each variable in predicting the dependent variable is difficult to determine. The reason for this difficulty is that highly interrelated variables are being used to predict the same variance in the dependent variable. Researchers/statisticians disagree on the specific correlation value that must be present for multicollinearity to exist. Some persons contend that correlations above .70 are necessary whereas other persons contend that the correlations must be above .90 for multicollinearity to exist.

If multicollinearity is present, you can leave it as it is, and have SPSS calculate the multiple regression. Multicollinearity influences the results regarding each predictor's unique contribution. If your interest is in the overall or combined effect of the statistically significant predictors, then multicollinearity is not an issue. Other choices would be to remove one or more of the highly correlated variables from the regression analysis or to create an aggregate or composite of the highly correlated variables.

The choice that we recommend is to have SPSS calculate multicollinearity when the multiple regression analysis is calculated. More on this later.

¹⁷ <http://cnx.org/content/m40738/latest/17.7.png/image>

Chapter 19

19. Multiple Regression: Part I¹



NOTE: This chapter is published by NCPEA Press² and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of education administration.

About the Authors

John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website (Writing and Statistical Help³) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

Ana Rojas-LeBouef is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. Dr. LeBoeuf recently completed her doctoral degree in Reading, where she conducted a 16-year analysis of Texas statewide data regarding the achievement gap. Her research interests lie in examining the inequities in achievement among ethnic groups. Dr. Rojas-LeBouef also assists students and faculty in their writing and statistical needs on the Writing and Statistical Help website.

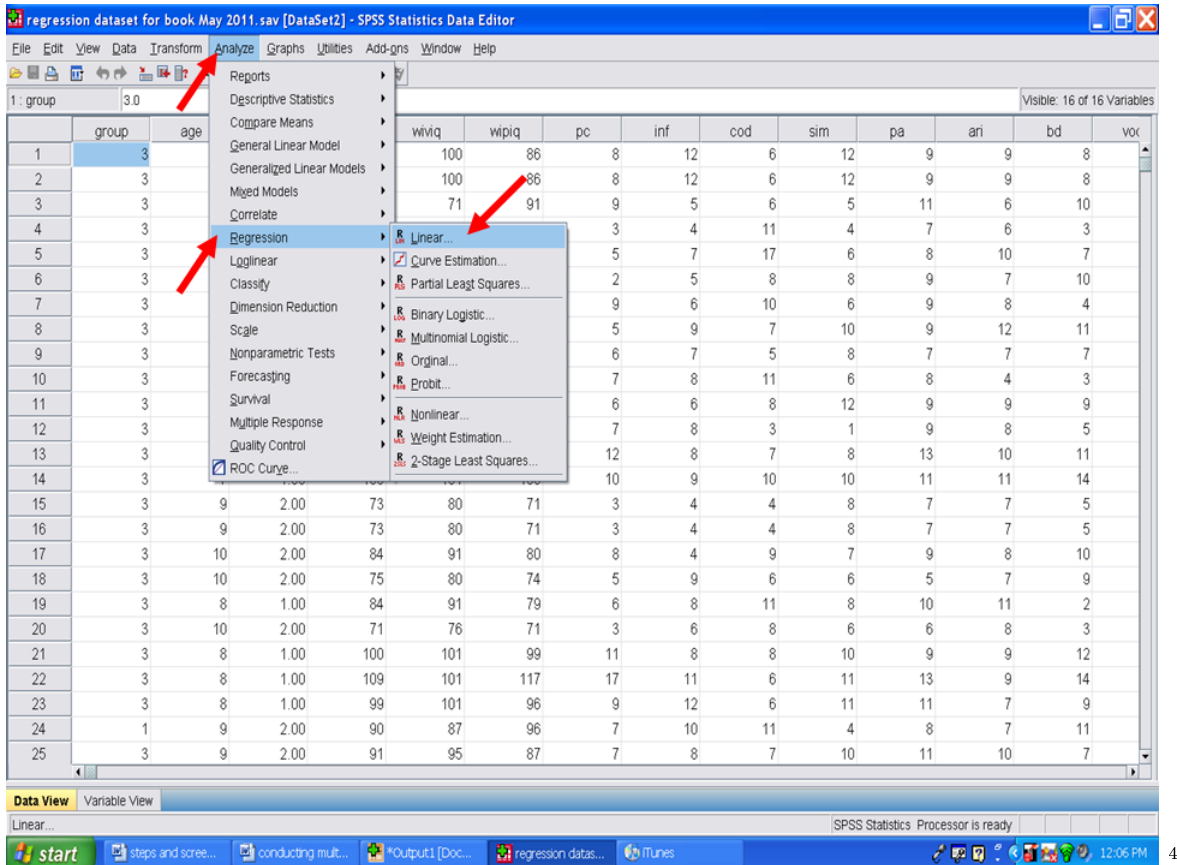
From the data screen, click on

- ✓ Analyze
- ✓ Regression
- ✓ Linear

¹This content is available online at <<http://cnx.org/content/m40739/1.2/>>.

²<http://www.ncpeapublications.org/books.html>

³<http://cnx.org/content/m40739/latest/www.writingandstatisticalhelp>



Your screen should now look like the one below. We will send the dependent variable, the one we are trying to predict, to the Dependent box. Click on Wechsler Full Scale IQ 3 and send it to the Dependent box.

⁴<http://cnx.org/content/m40739/latest/18.1.png/image>

The screenshot shows the SPSS Linear Regression dialog box. The dependent variable is 'Reading Group M...'. The independent variables are 'Performance 1 (P...', 'Verbal 1 (Informa...', 'Performance 2 (C...', 'Verbal 2 (Similarit...', 'Performance 3 (P...', 'Verbal 3 (Arithme...', 'Performance 4 (B...', 'Verbal 4 (Vocabu...', 'Performance 5 (...', and 'Verbal 5 (Compre...'. The method is set to 'Enter'. The background data view shows a table with columns 'inf', 'cod', 'sim', 'pa', 'ari', 'bd', and 'vox'.

inf	cod	sim	pa	ari	bd	vox
12	6	12	9	9	8	
12	6	12	9	9	8	
5	6	5	11	6	10	
4	11	4	7	6	3	
7	17	6	8	10	7	
5	8	8	9	7	10	
6	10	6	9	8	4	
9	7	10	9	12	11	
7	5	8	7	7	7	
8	11	6	8	4	3	
6	8	12	9	9	9	
8	3	1	9	8	5	
8	7	8	13	10	11	
9	10	10	11	11	14	
4	4	8	7	7	5	
4	4	8	7	7	5	
4	9	7	9	8	10	
6	6	6	5	7	9	
6	6	6	6	8	3	
8	8	10	9	9	12	
11	6	11	13	9	14	
12	6	11	11	7	9	
10	11	4	8	7	11	
8	7	10	11	10	7	

Your screen should now look like the following one. Next we will place all of the independent variables in the Independent(s) box. In this example, remember that we are using Performance 1 through Verbal 5.

⁵<http://cnx.org/content/m40739/latest/18.2.png/image>

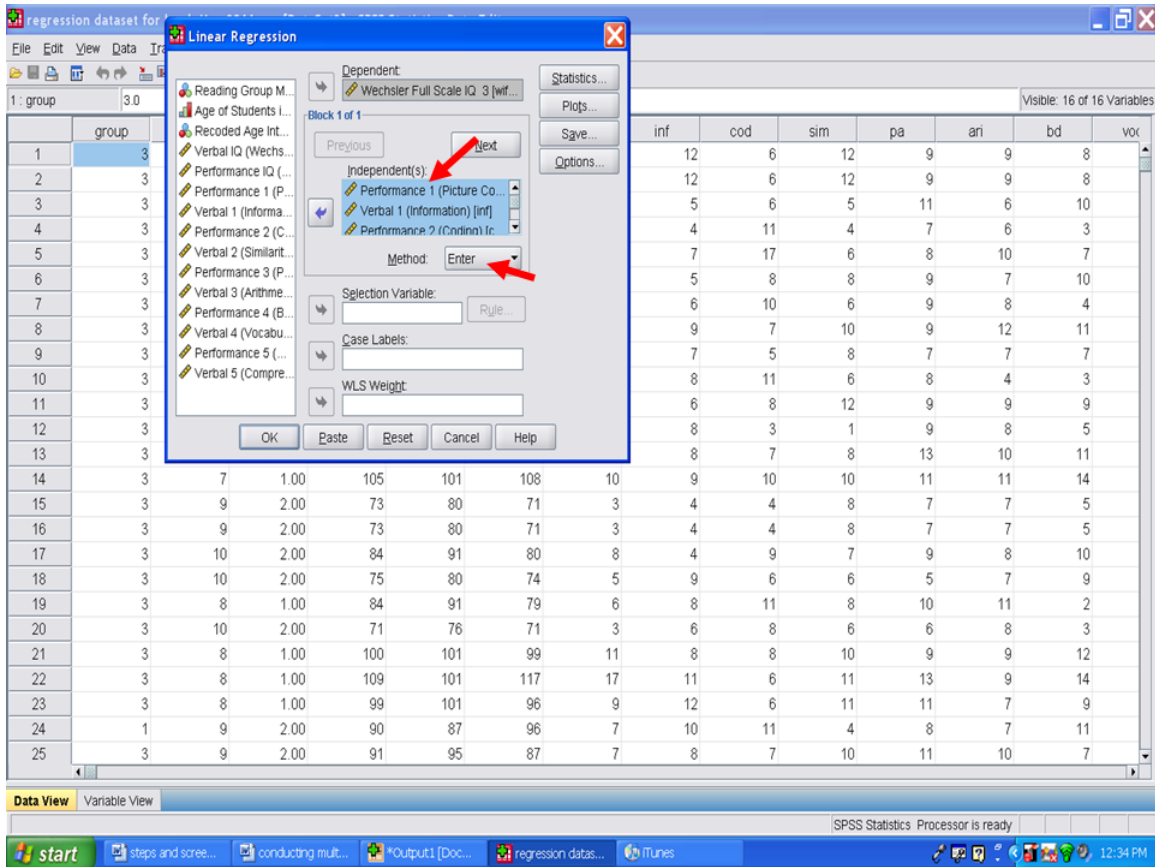
The screenshot shows the SPSS Linear Regression dialog box. The 'Dependent' box contains 'Wechsler Full Scale IQ 3 [wfr...]' and the 'Independent(s)' box contains several variables including 'Verbal IQ (Wechs...', 'Performance IQ (...', 'Performance 1 (P...', 'Verbal 1 (Informa...', 'Performance 2 (C...', 'Verbal 2 (Similarit...', 'Performance 3 (P...', 'Verbal 3 (Arithme...', 'Performance 4 (B...', 'Verbal 4 (Vocabu...', 'Performance 5 (...', and 'Verbal 5 (Compre...'. Red arrows point to the 'Dependent' box and the 'Independent(s)' box. The background shows a data table with columns 'inf', 'cod', 'sim', 'pa', 'ari', 'bd', and 'vox'.

	inf	cod	sim	pa	ari	bd	vox
12	6	12	9	9	8		
12	6	12	9	9	8		
5	6	5	11	6	10		
4	11	4	7	6	3		
7	17	6	8	10	7		
5	8	8	9	7	10		
6	10	6	9	8	4		
9	7	10	9	12	11		
7	5	8	7	7	7		
8	11	6	8	4	3		
6	8	12	9	9	9		
8	3	1	9	8	5		
8	7	8	13	10	11		
14	7	1.00	105	101	108	10	9
15	9	2.00	73	80	71	3	4
16	9	2.00	73	80	71	3	4
17	10	2.00	84	91	80	8	4
18	10	2.00	75	80	74	5	9
19	8	1.00	84	91	79	6	8
20	10	2.00	71	76	71	3	6
21	8	1.00	100	101	99	11	8
22	8	1.00	109	101	117	17	11
23	8	1.00	99	101	96	9	12
24	1	9	2.00	90	87	96	7
25	3	9	2.00	91	95	87	7

After you have placed all 10 of these variables in the Independent(s) box, your screen will look like the following one.

Next we will click on Method and change it from the default of Enter to Stepwise.

⁶ <http://cnx.org/content/m40739/latest/18.3.png/image>



Click on Stepwise.

⁷ <http://cnx.org/content/m40739/latest/18.4.png/image>

The screenshot shows the SPSS Linear Regression dialog box. The dependent variable is 'Wechsler Full Scale IQ 3 [wfs...]' and the independent variables are 'Performance 1 (Picture Co...', 'Verbal 1 (Information) [inf]', and 'Performance 2 (Coding) [c...'. The 'Method' dropdown menu is open, and 'Stepwise' is selected, indicated by a red arrow. Other options include 'Enter', 'Remove', 'Backward', and 'Forward'. The 'Selection Variable' is set to 'Enter'. The 'Case Labels' are set to 'Backward' and 'Forward'. The 'WLS Weight' is set to 'None'. The 'Statistics...' button is visible in the top right corner. The background shows a data view with 16 variables: inf, cod, sim, pa, ari, bd, and vox.

group	inf	cod	sim	pa	ari	bd	vox
1	12	6	12	9	9	8	
2	12	6	12	9	9	8	
3	5	6	5	11	6	10	
4	4	11	4	7	6	3	
5	7	17	6	8	10	7	
6	5	8	8	9	7	10	
7	6	10	6	9	8	4	
8	9	7	10	9	12	11	
9	7	5	8	7	7	7	
10	8	11	6	8	4	3	
11	6	8	12	9	9	9	
12	8	3	1	9	8	5	
13	8	7	8	13	10	11	
14	9	10	10	11	11	14	
15	4	4	8	7	7	5	
16	4	4	8	7	7	5	
17	4	9	7	9	8	10	
18	9	6	6	5	7	9	
19	8	11	8	10	11	2	
20	6	8	6	6	8	3	
21	8	8	10	9	9	12	
22	11	6	11	13	9	14	
23	12	6	11	11	7	9	
24	10	11	4	8	7	11	
25	8	7	10	11	10	7	

After clicking on Stepwise, your screen will look like the one below.
Click on Statistics

⁸<http://cnx.org/content/m40739/latest/18.5.png/image>

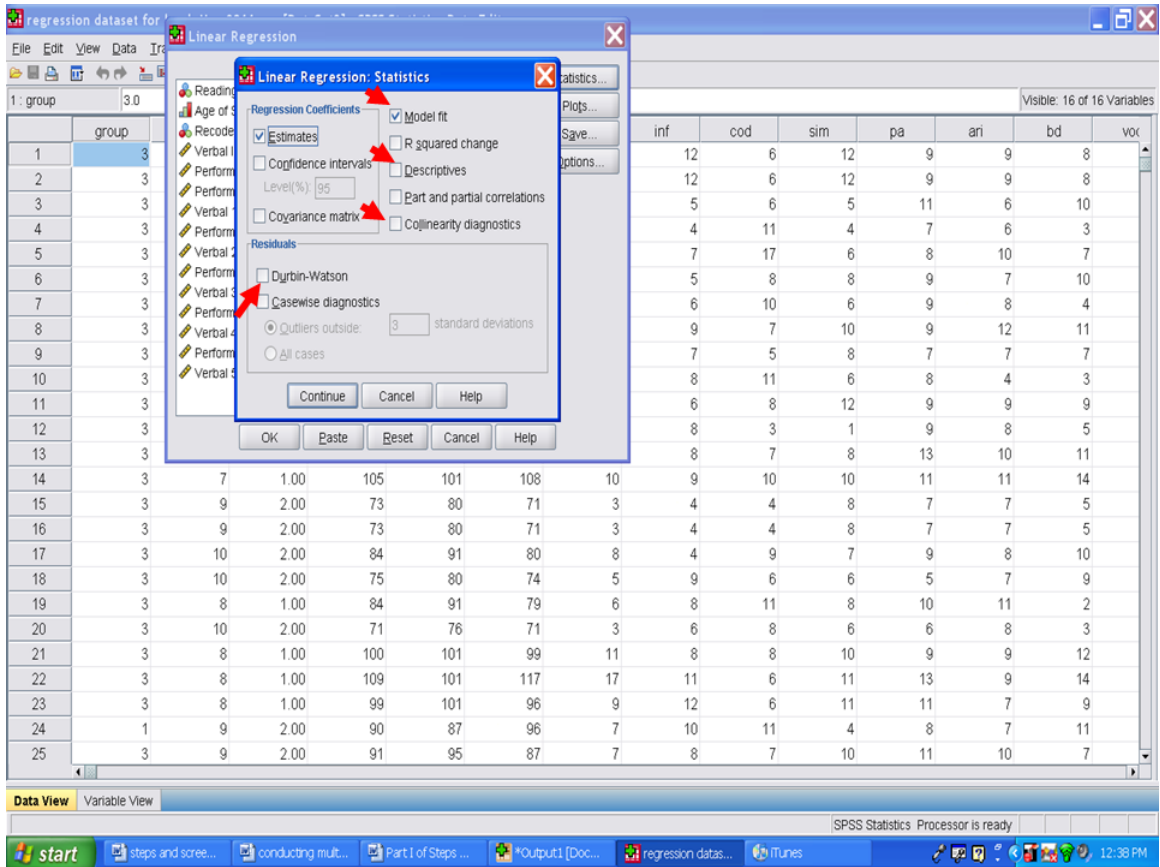
The screenshot shows the SPSS Linear Regression dialog box. The dependent variable is 'Wechsler Full Scale IQ 3 [wfs...]' and the independent variables are 'Performance 1 (Picture Co...', 'Verbal 1 (Information) [inf]', and 'Performance 2 (Coding) [c...]. The method is set to 'Stepwise'. A red arrow points to the 'Statistics...' button. The background data view shows the following table:

group	inf	cod	sim	pa	ari	bd	vox
1	12	6	12	9	9	8	
2	12	6	12	9	9	8	
3	5	6	5	11	6	10	
4	4	11	4	7	6	3	
5	7	17	6	8	10	7	
6	5	8	8	9	7	10	
7	6	10	6	9	8	4	
8	9	7	10	9	12	11	
9	7	5	8	7	7	7	
10	8	11	6	8	4	3	
11	6	8	12	9	9	9	
12	8	3	1	9	8	5	
13	8	7	8	13	10	11	
14	9	10	10	11	11	14	
15	4	4	8	7	7	5	
16	4	4	8	7	7	5	
17	4	9	7	9	8	10	
18	9	6	6	5	7	9	
19	6	8	11	8	10	2	
20	3	6	8	6	8	3	
21	11	8	8	10	9	12	
22	17	11	6	11	13	14	
23	9	12	6	11	11	9	
24	7	10	11	4	8	11	
25	7	8	7	10	10	7	

After clicking on Statistics, you will now

- ✓ Click on Estimates (default already checked)
- ✓ Model fit (default already checked)
- ✓ R squared change
- ✓ Descriptives
- ✓ Collinearity Diagnostics
- ✓ Durbin-Watson

⁹ <http://cnx.org/content/m40739/latest/18.6.png/image>



Your screen should like the one below. Now click on Continue.

¹⁰<http://cnx.org/content/m40739/latest/18.7.png/image>

The screenshot shows the SPSS Linear Regression dialog box with the Statistics sub-dialog open. The Statistics sub-dialog has the following options checked:

- Estimates
- Model fit
- R squared change
- Descriptives
- Durbin-Watson
- Collinearity diagnostics

The 'Continue' button is highlighted with a red arrow.

group	inf	cod	sim	pa	ari	bd	vox
1	3						
2	3						
3	3						
4	3						
5	3						
6	3						
7	3						
8	3						
9	3						
10	3						
11	3						
12	3						
13	3						
14	3	7	1.00	105	101	108	10
15	3	9	2.00	73	80	71	3
16	3	9	2.00	73	80	71	3
17	3	10	2.00	84	91	80	8
18	3	10	2.00	75	80	74	5
19	3	8	1.00	84	91	79	6
20	3	10	2.00	71	76	71	3
21	3	8	1.00	100	101	99	11
22	3	8	1.00	109	101	117	17
23	3	8	1.00	99	101	96	9
24	1	9	2.00	90	87	96	7
25	3	9	2.00	91	95	87	7

Next click on

✓ OK

Go to the Output file if SPSS does not automatically send you there.

¹¹<http://cnx.org/content/m40739/latest/18.8.png/image>

The screenshot shows the SPSS Linear Regression dialog box in the foreground, overlaid on a data view window. The dialog box has 'Wechsler Full Scale IQ 3 [wif...]' selected as the dependent variable and several independent variables: 'Performance 1 (Picture Co...', 'Verbal 1 (Information) [inf]', and 'Performance 2 (Coding) [c...'. The method is set to 'Stepwise'. A red arrow points to the 'OK' button. The background window shows a data table with 25 rows and 16 columns. The columns are labeled 'inf', 'cod', 'sim', 'pa', 'ari', 'bd', and 'vo'. The data view window title is 'regression dataset for...'. The taskbar at the bottom shows the Windows start button and several open applications including 'steps and screenshots canonical discriminant analysis Assumptions - Microsoft W...', 'SPSS Statistics Processor is ready', and 'regression datas...'. The system clock shows 12:53 PM.

group	inf	cod	sim	pa	ari	bd	vo
1	3						
2	3						
3	3						
4	3						
5	3						
6	3						
7	3						
8	3						
9	3						
10	3						
11	3						
12	3						
13	3						
14	3	1.00	105	101	108	10	9
15	3	9	2.00	73	80	71	3
16	3	9	2.00	73	80	71	3
17	3	10	2.00	84	91	80	8
18	3	10	2.00	75	80	74	5
19	3	8	1.00	84	91	79	6
20	3	10	2.00	71	76	71	3
21	3	8	1.00	100	101	99	11
22	3	8	1.00	109	101	117	17
23	3	8	1.00	99	101	96	9
24	1	9	2.00	90	87	96	7
25	3	9	2.00	91	95	87	7

¹²<http://cnx.org/content/m40739/latest/18.89.png/image>

Chapter 20

20. Multiple Regression: Part II¹



NOTE: This chapter is published by NCPEA Press² and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship and practice of education administration.

About the Authors

John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website (Writing and Statistical Help³) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

Ana Rojas-LeBouef is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. Dr. LeBoeuf recently completed her doctoral degree in Reading, where she conducted a 16-year analysis of Texas statewide data regarding the achievement gap. Her research interests lie in examining the inequities in achievement among ethnic groups. Dr. Rojas-LeBouef also assists students and faculty in their writing and statistical needs on the Writing and Statistical Help website.

The first table in your SPSS output file should be a Descriptive Statistics table. You will see a column for M , SD , and n . This information should be used in your Results section.

¹This content is available online at <<http://cnx.org/content/m40742/1.2/>>.

²<http://www.ncpeapublications.org/books.html>

³<http://cnx.org/content/m40742/latest/www.writingandstatisticalhelp>

The screenshot shows the SPSS Statistics Viewer interface. The 'Descriptive Statistics' table is highlighted with a red arrow. Below it is the 'Correlations' table. The taskbar at the bottom shows several open applications, including 'start', 'steps and sor...', 'conducting m...', 'Regression Ha...', 'steps screens...', '*Output1 [D...', 'regression dat...', and 'iTunes'. The system tray shows the time as 12:59 PM.

	Mean	Std. Deviation	N
Wechsler Full Scale IQ 3	77.86	13.508	1133
Performance 1 (Picture Completion)	6.99	3.110	1133
Verbal 1 (Information)	6.35	2.813	1133
Performance 2 (Coding)	7.29	3.125	1133
Verbal 2 (Similarities)	6.21	2.986	1133
Performance 3 (Picture Arrangement)	7.27	3.031	1133
Verbal 3 (Arithmetic)	6.56	2.729	1133
Performance 4 (Block Design)	6.52	3.353	1133
Verbal 4 (Vocabulary)	5.64	2.891	1133
Performance 5 (Object Assembly)	6.65	3.247	1133
Verbal 5 (Comprehension)	5.82	3.218	1133

	Wechsler Full Scale IQ 3	Performance 1 (Picture Completion)	Verbal 1 (Information)	Performance 2 (Coding)	Verbal 2 (Similarities)	Performance 3 (Picture Arrangement)	Verbal 3 (Arithmetic)	Performance 4 (Block Design)	Verbal 4 (Vocabulary)	Performance 5 (Object Assembly)	Verbal 5 (Comprehension)
Pearson Correlation	Wechsler Full Scale IQ 3	1.000	.674	.765	.487	.768	.719	.698	.736	.760	.760
	Performance 1 (Picture Completion)	.674	1.000	.441	.206	.434	.471	.345	.495	.406	.406
	Verbal 1 (Information)	.765	.441	1.000	.282	.624	.447	.533	.491	.662	.662
	Performance 2 (Coding)	.487	.206	.282	1.000	.229	.387	.332	.267	.256	.256
	Verbal 2 (Similarities)	.768	.434	.624	.229	1.000	.491	.502	.471	.663	.663
	Performance 3 (Picture Arrangement)	.719	.471	.447	.387	.491	1.000	.482	.467	.479	.479
	Verbal 3 (Arithmetic)	.698	.345	.533	.332	.502	.482	1.000	.440	.493	.493
	Performance 4 (Block Design)	.736	.495	.491	.267	.471	.467	.440	1.000	.417	.417
	Verbal 4 (Vocabulary)	.760	.406	.662	.256	.663	.479	.493	.417	1.000	.417
	Performance 5 (Object Assembly)	.760	.406	.662	.256	.663	.479	.493	.417	.417	1.000
	Verbal 5 (Comprehension)	.724	.367	.580	.261	.573	.403	.483	.414	.666	.666
Sig. (1-tailed)	Wechsler Full Scale IQ 3	.	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Performance 1 (Picture Completion)	.000	.	.000	.000	.000	.000	.000	.000	.000	.000

Underneath the Descriptive Statistics table is a table labeled Correlations. This table reflects the Pearson r s for each independent variable with the dependent variable, as well as the interrelationships among all of the variables.

Of the relationships of the independent variable with the Wechsler Full Scale IQ 3 (the dependent variable in this example), Verbal 2 (Similarities) has the highest correlation, .768. In a stepwise regression procedure, this variable should be the first statistically significant predictor.

⁴<http://cnx.org/content/m40742/latest/19.1.png/image>

*Output1 [Document1] - SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help

Verbal 5 (Comprehension) 5.82 3.218 1133

Correlations

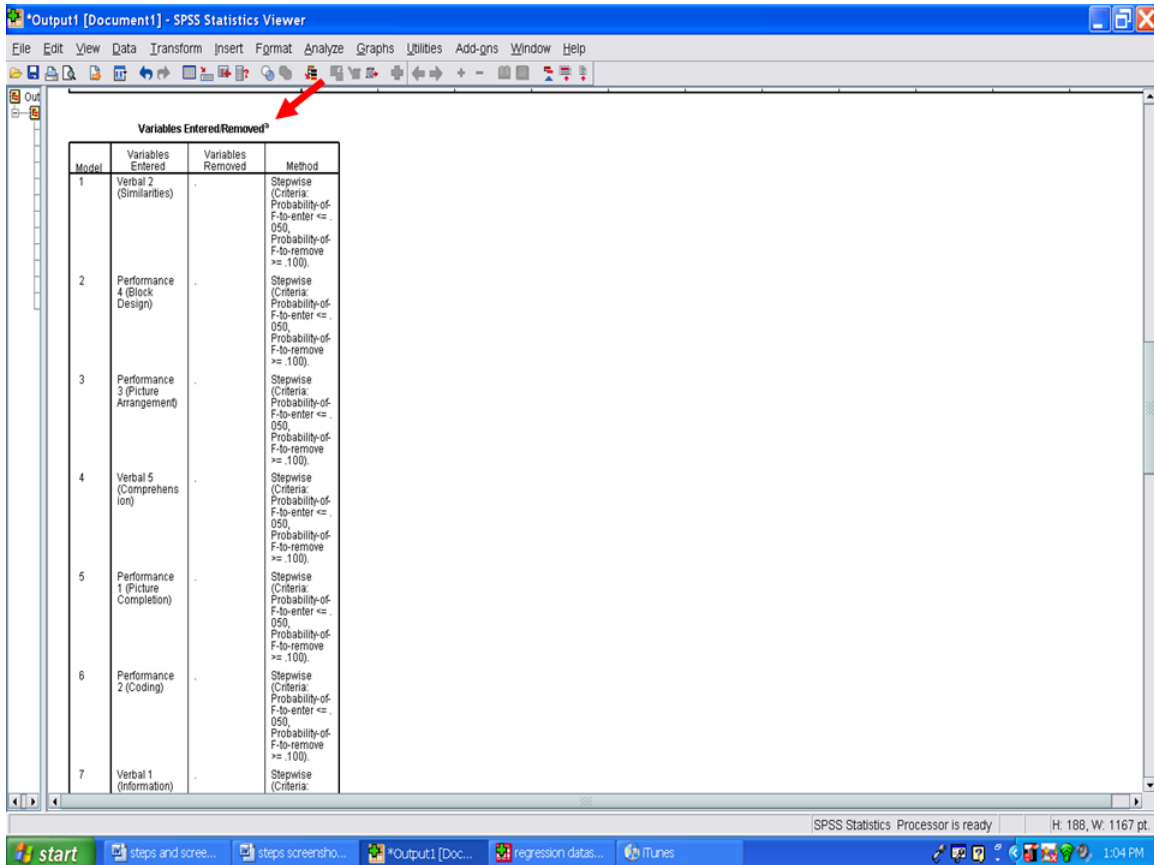
		Wechsler Full Scale IQ 3	Performance 1 (Picture Completion)	Verbal 1 (Information)	Performance 2 (Coding)	Verbal 2 (Similarities)	Performance 3 (Picture Arrangement)	Verbal 3 (Arithmetic)	Performance 4 (Block Design)	Verbal 4 (Vocabulary)	Performance 5 (Object Assembly)	Verbal 5 (Comprehension)
Pearson Correlation	Wechsler Full Scale IQ 3	1.000	.674	.765	.487	.768	.719	.698	.736	.760	.643	.724
	Performance 1 (Picture Completion)	.674	1.000	.441	.206	.434	.471	.345	.495	.406	.509	.367
	Verbal 1 (Information)	.765	.441	1.000	.262	.624	.447	.533	.491	.662	.334	.580
	Performance 2 (Coding)	.487	.206	.262	1.000	.229	.387	.332	.267	.256	.182	.261
	Verbal 2 (Similarities)	.768	.434	.624	.229	1.000	.491	.502	.471	.663	.403	.573
	Performance 3 (Picture Arrangement)	.719	.471	.447	.387	.491	1.000	.482	.467	.479	.419	.403
	Verbal 3 (Arithmetic)	.698	.345	.533	.332	.502	.482	1.000	.440	.493	.331	.483
	Performance 4 (Block Design)	.736	.495	.491	.267	.471	.467	.440	1.000	.417	.628	.414
	Verbal 4 (Vocabulary)	.760	.406	.662	.256	.663	.479	.493	.417	1.000	.317	.666
	Performance 5 (Object Assembly)	.643	.509	.334	.182	.403	.419	.331	.628	.317	1.000	.312
	Verbal 5 (Comprehension)	.724	.367	.580	.261	.573	.403	.483	.414	.666	.312	1.000
	Sig. (1-tailed)	Wechsler Full Scale IQ 3	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
Performance 1 (Picture Completion)		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
Verbal 1 (Information)		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
Performance 2 (Coding)		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
Verbal 2 (Similarities)		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
Performance 3 (Picture Arrangement)		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
Verbal 3 (Arithmetic)		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
Performance 4 (Block Design)		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
Verbal 4 (Vocabulary)		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
Performance 5 (Object Assembly)		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
Verbal 5 (Comprehension)		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
N		Wechsler Full Scale IQ 3	1133	1133	1133	1133	1133	1133	1133	1133	1133	1133
	Performance 1 (Picture Completion)	1133	1133	1133	1133	1133	1133	1133	1133	1133	1133	1133
	Verbal 1 (Information)	1133	1133	1133	1133	1133	1133	1133	1133	1133	1133	1133

SPSS Statistics Processor is ready H. 188, W. 1167 pt.

start steps and scree... steps screenho... *Output1 [Doc... regression datas... iTunes 1:01 PM

Next, you will see the Variables Entered/Removed table. We will not use the information in this table.

⁵ <http://cnx.org/content/m40742/latest/19.2.png/image>



The screenshot shows the SPSS Statistics Viewer window with the 'Variables Entered/Removed*' table. A red arrow points to the 'Variables Entered' column. The table contains the following data:

Model	Variables Entered	Variables Removed	Method
1	Verbal 2 (Similarities)		Stepwise (Criteria: Probability of F-to-enter <= .050, Probability of F-to-remove >= .100).
2	Performance 4 (Block Design)		Stepwise (Criteria: Probability of F-to-enter <= .050, Probability of F-to-remove >= .100).
3	Performance 3 (Picture Arrangement)		Stepwise (Criteria: Probability of F-to-enter <= .050, Probability of F-to-remove >= .100).
4	Verbal 5 (Comprehension)		Stepwise (Criteria: Probability of F-to-enter <= .050, Probability of F-to-remove >= .100).
5	Performance 1 (Picture Completion)		Stepwise (Criteria: Probability of F-to-enter <= .050, Probability of F-to-remove >= .100).
6	Performance 2 (Coding)		Stepwise (Criteria: Probability of F-to-enter <= .050, Probability of F-to-remove >= .100).
7	Verbal 1 (Information)		Stepwise (Criteria: Probability of F-to-enter <= .050, Probability of F-to-remove >= .100).

The table labeled Model Summary is an important table. This table is a summary of all of the steps entered/removed. Important columns in this table are:

Adjusted R Square
R Square Change
Durbin-Watson

⁶ <http://cnx.org/content/m40742/latest/19.3.png/image>

SPSS Statistics Processor is ready | H. 188, W. 1167 pt

7

start | steps and scree... | steps screenho... | *Output1 [Doc... | regression datas... | iTunes

SPSS Statistics Processor is ready | H. 188, W. 1167 pt

*Output1 [Document1] - SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help

a. Dependent Variable: Wechsler Full Scale IQ 3

Model Summary^a

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.768 ^a	.589	.589	8.660	.589	1623.156	1	1131	.000	
2	.877 ^b	.770	.769	6.491	.180	883.306	1	1130	.000	
3	.919 ^c	.845	.845	5.318	.076	554.226	1	1128	.000	
4	.949 ^d	.901	.901	4.250	.056	639.947	1	1128	.000	
5	.963 ^e	.927	.927	3.651	.026	401.185	1	1127	.000	
6	.975 ^f	.951	.950	3.005	.023	537.350	1	1126	.000	
7	.983 ^g	.967	.967	2.451	.017	567.513	1	1125	.000	
8	.989 ^h	.979	.979	1.964	.012	628.579	1	1124	.000	
9	.995 ⁱ	.990	.990	1.383	.011	1143.402	1	1123	.000	
10	.998 ^j	.997	.996	.803	.007	2213.928	1	1122	.000	1.912

a. Predictors: (Constant), Verbal 2 (Similarities)
b. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design)
c. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement)
d. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension)
e. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion)
f. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding)
g. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information)
h. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information), Performance 5 (Object Assembly)
i. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information), Performance 5 (Object Assembly), Verbal 3 (Arithmetic)
j. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information), Performance 5 (Object Assembly), Verbal 3 (Arithmetic), Verbal 4 (Vocabulary)
k. Dependent Variable: Wechsler Full Scale IQ 3

ANOVA^k

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	121732.471	1	121732.471	1623.156	.000 ^a
	Residual	84822.050	1131	74.997		
	Total	206554.521	1132			
2	Regression	158946.807	2	79473.404	1886.353	.000 ^b
	Residual	47607.713	1130	42.131		
	Total	206554.521	1132			

The Durbin-Watson statistic is a measure for significant residual autocorrelation. Ideally it should be close to 2. In our example, the Durbin-Watson statistic is 1.912. Therefore, this assumption has not been violated.

⁷ <http://cnx.org/content/m40742/latest/19.4.png/image>

SPSS Statistics Processor is ready | H. 188, W. 1167 pt.

start | steps and scree... | steps screenho... | *Output1 [Doc... | regression data... | iTunes | 1:12 PM

SPSS Statistics Processor is ready | H. 188, W. 1167 pt.

*Output1 [Document1] - SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help

a. Dependent Variable: Wechsler Full Scale IQ 3

Model Summary^a

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.768 ^a	.589	.589	8.660	.589	1623.156	1	1131	.000	
2	.877 ^b	.770	.769	6.491	.180	883.306	1	1130	.000	
3	.919 ^c	.845	.845	5.318	.076	554.226	1	1128	.000	
4	.949 ^d	.901	.901	4.250	.056	639.947	1	1128	.000	
5	.963 ^e	.927	.927	3.651	.026	401.185	1	1127	.000	
6	.975 ^f	.951	.950	3.005	.023	537.350	1	1126	.000	
7	.983 ^g	.967	.967	2.451	.017	567.513	1	1125	.000	
8	.989 ^h	.979	.979	1.964	.012	628.579	1	1124	.000	
9	.995 ⁱ	.990	.990	1.383	.011	1143.402	1	1123	.000	
10	.998 ^j	.997	.996	.803	.007	2213.928	1	1122	.000	1.912

a. Predictors: (Constant), Verbal 2 (Similarities)
 b. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design)
 c. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement)
 d. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension)
 e. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion)
 f. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding)
 g. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information)
 h. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information), Performance 5 (Object Assembly)
 i. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information), Performance 5 (Object Assembly), Verbal 3 (Arithmetic)
 j. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information), Performance 5 (Object Assembly), Verbal 3 (Arithmetic), Verbal 4 (Vocabulary)
 k. Dependent Variable: Wechsler Full Scale IQ 3

ANOVA^k

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	121732.471	1	121732.471	1623.156	.000 ^a
	Residual	84822.050	1131	74.997		
	Total	206554.521	1132			
2	Regression	158946.807	2	79473.404	1886.353	.000 ^b
	Residual	47607.713	1130	42.131		
	Total	206554.521	1132			

The Adjusted R Square column indicates the amount of variance that each model explains in the dependent variable (i.e., Wechsler Full Scale IQ 3). The first model has the letter a next to it. This model contains a single independent variable, Verbal 2 (Similarities). If you recall, this variable has the highest Pearson r with the Wechsler Full Scale IQ 3. In this example, Verbal 2 (Similarities) accounts for 58.9% of the variance in the Wechsler Full Scale IQ 3.

⁸ <http://cnx.org/content/m40742/latest/19.5.png/image>

SPSS Statistics Processor is ready | H. 188, W. 1167 pt.

9

SPSS Statistics Processor is ready | H. 188, W. 1167 pt.

start | steps and scree... | steps screenho... | *Output1 [Doc... | regression datas... | iTunes | 1:13 PM

a. Dependent Variable: Wechsler Full Scale IQ 3

Model Summary^a

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.768 ^a	.589	.589	8.660	.589	1623.156	1	1131	.000	
2	.877 ^b	.770	.769	6.491	.180	883.306	1	1130	.000	
3	.919 ^c	.845	.845	5.318	.076	554.226	1	1128	.000	
4	.949 ^d	.901	.901	4.250	.056	639.947	1	1128	.000	
5	.963 ^e	.927	.927	3.651	.026	401.185	1	1127	.000	
6	.975 ^f	.951	.950	3.005	.023	537.350	1	1126	.000	
7	.983 ^g	.967	.967	2.451	.017	567.513	1	1125	.000	
8	.989 ^h	.979	.979	1.964	.012	628.579	1	1124	.000	
9	.995 ⁱ	.990	.990	1.383	.011	1143.402	1	1123	.000	
10	.998 ^j	.997	.996	.803	.007	2213.928	1	1122	.000	1.912

a. Predictors: (Constant), Verbal 2 (Similarities)
b. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design)
c. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement)
d. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension)
e. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion)
f. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding)
g. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information)
h. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information), Performance 5 (Object Assembly)
i. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information), Performance 5 (Object Assembly), Verbal 3 (Arithmetic)
j. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information), Performance 5 (Object Assembly), Verbal 3 (Arithmetic), Verbal 4 (Vocabulary)
k. Dependent Variable: Wechsler Full Scale IQ 3

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	121732.471	1	121732.471	1623.156	.000 ^a
	Residual	84822.050	1131	74.997		
	Total	206554.521	1132			
2	Regression	158946.807	2	79473.404	1886.353	.000 ^b
	Residual	47607.713	1130	42.131		
	Total	206554.521	1132			

Each model, from 1 to 10, depicts the addition of another statistically significant variable in predicting the Wechsler Full Scale IQ 3. The final model, 10, indicates that all 10 independent variables accounted for 99.6% of the variance in the Wechsler Full Scale IQ 3.

⁹ <http://cnx.org/content/m40742/latest/19.6.png/image>

SPSS Statistics Processor is ready | H. 188, W. 1167 pt. | 10

10

Next, examine the R Square Change column. Each value reflects the unique variance in the Wechsler Full Scale IQ 3 explained by each statistically significant predictor variable. For Model 1, Verbal 2 (Similarities) explained the most variance, 58.9%. In Model 2, Performance 4 (Block Design) added 18.0% of unique variance that it explained.

¹⁰<http://cnx.org/content/m40742/latest/19.7.png/image>

SPSS Statistics Processor is ready | H. 188, W. 1167 pt | 1:19 PM

*Output1 [Document1] - SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help

a. Dependent Variable: Wechsler Full Scale IQ 3

Model Summary^k

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				Durbin-Watson
					R Square Change	F Change	df1	df2	
1	.768 ^a	.589	.589	8.660	.589	1623.156	1	1131	.000
2	.877 ^b	.770	.769	6.491	.180	883.306	1	1130	.000
3	.919 ^c	.845	.845	5.318	.076	554.226	1	1128	.000
4	.949 ^d	.901	.901	4.250	.056	639.947	1	1128	.000
5	.963 ^e	.927	.927	3.651	.026	401.185	1	1127	.000
6	.975 ^f	.951	.950	3.005	.023	537.350	1	1126	.000
7	.983 ^g	.967	.967	2.451	.017	567.513	1	1125	.000
8	.989 ^h	.979	.979	1.964	.012	628.579	1	1124	.000
9	.995 ⁱ	.990	.990	1.383	.011	1143.402	1	1123	.000
10	.998 ^j	.997	.996	.803	.007	2213.928	1	1122	.000

a. Predictors: (Constant), Verbal 2 (Similarities)
 b. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design)
 c. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement)
 d. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension)
 e. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion)
 f. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding)
 g. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information)
 h. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information), Performance 5 (Object Assembly)
 i. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information), Performance 5 (Object Assembly), Verbal 3 (Arithmetic)
 j. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information), Performance 5 (Object Assembly), Verbal 3 (Arithmetic), Verbal 4 (Vocabulary)
 k. Dependent Variable: Wechsler Full Scale IQ 3

ANOVA^k

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	121732.471	1	121732.471	1623.156	.000 ^a
	Residual	84822.050	1131	74.997		
	Total	206554.521	1132			
2	Regression	158946.807	2	79473.404	1886.353	.000 ^b
	Residual	47607.713	1130	42.131		
	Total	206554.521	1132			

The final model, 10, only added 0.7% of additional variance explained.

¹¹ <http://cnx.org/content/m40742/latest/19.8.png/image>

SPSS Statistics Processor is ready | H. 188, W. 1167 pt

12

a. Dependent Variable: Wechsler Full Scale IQ 3

Model Summary^k

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.768 ^a	.589	.589	8.660	.589	1623.156	1	1131	.000	
2	.877 ^b	.770	.769	6.491	.180	883.306	1	1130	.000	
3	.919 ^c	.845	.845	5.318	.076	554.226	1	1128	.000	
4	.949 ^d	.901	.901	4.250	.056	639.947	1	1128	.000	
5	.963 ^e	.927	.927	3.651	.026	401.185	1	1127	.000	
6	.975 ^f	.951	.950	3.005	.023	537.350	1	1126	.000	
7	.983 ^g	.967	.967	2.451	.017	567.513	1	1125	.000	
8	.989 ^h	.979	.979	1.964	.012	628.579	1	1124	.000	
9	.995 ⁱ	.990	.990	1.383	.011	1143.402	1	1123	.000	
10	.998 ^j	.997	.996	.803	.007	2213.928	1	1122	.000	1.912

a. Predictors: (Constant), Verbal 2 (Similarities)
b. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design)
c. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement)
d. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension)
e. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion)
f. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding)
g. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information)
h. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information), Performance 5 (Object Assembly)
i. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information), Performance 5 (Object Assembly), Verbal 3 (Arithmetic)
j. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information), Performance 5 (Object Assembly), Verbal 3 (Arithmetic), Verbal 4 (Vocabulary)
k. Dependent Variable: Wechsler Full Scale IQ 3

ANOVA^k

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	121732.471	1	121732.471	1623.156	.000 ^a
	Residual	84822.050	1131	74.997		
	Total	206554.521	1132			
2	Regression	158946.807	2	79473.404	1886.353	.000 ^b
	Residual	47607.713	1130	42.131		
	Total	206554.521	1132			

The next table of importance is the ANOVA table. Each model is examined to determine the extent to which it explains a statistically significant amount of the variance in the dependent variable. Of interest to us is the very last model, 10, which shows a statistically significant result, $F(10, 1122) = 31959.947, p < .001$.

¹²<http://cnx.org/content/m40742/latest/19.9.png/image>

The screenshot shows the SPSS Statistics Viewer window with the following ANOVA table:

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	121732.471	1	121732.471	1623.156	.000 ^a
	Residual	84822.050	1131	74.997		
	Total	206554.521	1132			
2	Regression	158946.807	2	79473.404	1886.353	.000 ^b
	Residual	47607.713	1130	42.131		
	Total	206554.521	1132			
3	Regression	174622.319	3	58207.440	2057.991	.000 ^c
	Residual	31932.202	1129	28.284		
	Total	206554.521	1132			
4	Regression	186180.877	4	46545.219	2577.006	.000 ^d
	Residual	20373.644	1128	18.062		
	Total	206554.521	1132			
5	Regression	191529.444	5	38305.889	2873.246	.000 ^e
	Residual	15025.077	1127	13.332		
	Total	206554.521	1132			
6	Regression	196383.338	6	32730.556	3623.434	.000 ^f
	Residual	10171.182	1126	9.033		
	Total	206554.521	1132			
7	Regression	199793.816	7	28541.974	4749.463	.000 ^g
	Residual	6760.705	1125	6.010		
	Total	206554.521	1132			
8	Regression	202218.608	8	25277.326	6552.648	.000 ^h
	Residual	4335.913	1124	3.858		
	Total	206554.521	1132			
9	Regression	204406.080	9	22711.787	11871.558	.000 ⁱ
	Residual	2148.440	1123	1.913		
	Total	206554.521	1132			
10	Regression	205831.918	10	20583.192	31959.947	.000 ^j
	Residual	722.603	1122	.644		
	Total	206554.521	1132			

Footnote:
a. Predictors: (Constant), Verbal 2 (Similarities)
b. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design)
c. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement)
d. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension)
e. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1

Underneath the ANOVA table is the Coefficients table. The important columns in this table are the Standardized Coefficients Beta and the Collinearity Statistics Tolerance and VIF ones. We will scroll down this table until we get to the final model, 10, information.

¹³<http://cnx.org/content/m40742/latest/19.10.png/image>

SPSS Statistics Processor is ready | H: 980, W: 537 pt

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	56.284	.594		94.727	.000		
	Verbal 2 (Similarities)	3.473	.086	.788	40.288	.000	1.000	1.000
2	(Constant)	50.021	.493		101.529	.000		
	Verbal 2 (Similarities)	2.448	.073	.541	33.426	.000	.778	1.285
	Performance 4 (Block Design)	1.939	.065	.481	29.720	.000	.778	1.285
3	(Constant)	45.153	.454		99.551	.000		
	Verbal 2 (Similarities)	1.923	.064	.425	30.048	.000	.684	1.462
	Performance 4 (Block Design)	1.534	.056	.381	27.317	.000	.705	1.418
	Performance 3 (Picture Arrangement)	1.481	.063	.332	23.542	.000	.687	1.455
4	(Constant)	43.891	.366		119.963	.000		
	Verbal 2 (Similarities)	1.325	.056	.293	23.519	.000	.564	1.774
	Performance 4 (Block Design)	1.352	.045	.336	29.756	.000	.688	1.454
	Performance 3 (Picture Arrangement)	1.332	.051	.299	26.324	.000	.678	1.475
	Verbal 5 (Comprehension)	1.245	.049	.297	25.297	.000	.636	1.571
5	(Constant)	42.034	.328		128.261	.000		
	Verbal 2 (Similarities)	1.199	.049	.265	24.553	.000	.554	1.805
	Performance 4 (Block Design)	1.122	.041	.278	27.562	.000	.633	1.591
	Performance 3 (Picture Arrangement)	1.122	.045	.252	25.073	.000	.641	1.561
	Verbal 5 (Comprehension)	1.185	.042	.282	27.972	.000	.633	1.579
6	(Constant)	39.128	.297		131.541	.000		
	Verbal 2 (Similarities)	1.223	.040	.270	30.421	.000	.554	1.805
	Performance 4 (Block Design)	1.059	.034	.263	31.521	.000	.629	1.591
	Performance 3 (Picture Arrangement)	.880	.038	.198	23.013	.000	.593	1.686

Scrolling down to the last model information shows us the following information. If you recall from the assumption checks, that multicollinearity was to be examined in the SPSS output. Multicollinearity is present when Tolerance values are below .1. As you can tell from the output below, Tolerance values range from a low of .386 to a high of .811. A second check for multicollinearity is the VIF column. Multicollinearity is present in the VIF column when the VIF values are greater than 10. In the example below, the VIF values range from a low of 1.233 to a high of 2.592. Therefore, multicollinearity is not present in this example.

¹⁴<http://cnx.org/content/m40742/latest/19.11.png/image>

*Output1 [Document1] - SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help

		Beta	Std. Error	t	Sig.	Partial Correlation	Partial	Standardized	Change in R Square
1	Performance 1 (Picture Completion)	.661	.017	38.965	.000	.608	1.645		
	Performance 2 (Coding)	.628	.015	42.978	.000	.811	1.233		
	Verbal 1 (Information)	.814	.021	38.209	.000	.471	2.123		
	Performance 5 (Object Assembly)	.606	.017	35.134	.000	.539	1.854		
	Verbal 3 (Arithmetic)	.665	.020	33.814	.000	.581	1.704		
	(Constant)	36.456	.083	437.681	.000				
	Verbal 2 (Similarities)	.647	.012	54.007	.000	.444	2.251		
	Performance 4 (Block Design)	.626	.010	61.016	.000	.482	2.076		
	Performance 3 (Picture Arrangement)	.619	.011	58.888	.000	.560	1.785		
	Verbal 5 (Comprehension)	.640	.011	60.551	.000	.492	2.033		
2	Performance 1 (Picture Completion)	.645	.010	65.543	.000	.607	1.647		
	Performance 2 (Coding)	.631	.008	74.406	.000	.811	1.233		
	Verbal 1 (Information)	.645	.013	50.123	.000	.434	2.302		
	Performance 5 (Object Assembly)	.620	.010	62.011	.000	.539	1.856		
	Verbal 3 (Arithmetic)	.645	.011	56.513	.000	.586	1.706		
	Verbal 4 (Vocabulary)	.625	.013	47.052	.000	.386	2.592		
	Performance 3 (Picture Arrangement)	.619	.011	58.888	.000	.560	1.785		
	Performance 4 (Block Design)	.626	.010	61.016	.000	.482	2.076		
	Verbal 2 (Similarities)	.647	.012	54.007	.000	.444	2.251		
	Performance 2 (Coding)	.628	.015	42.978	.000	.811	1.233		

a. Dependent Variable: Wechsler Full Scale IQ_3

Excluded Variables¹

Model	Beta In	t	Sig.	Partial Correlation	Collinearity Statistics			
					Tolerance	VIF	Minimum Tolerance	
1	Performance 1 (Picture Completion)	.419 ^a	24.509	.000	.589	.811	1.233	.811
	Verbal 1 (Information)	.468 ^a	23.391	.000	.571	.611	1.636	.611
	Performance 2 (Coding)	.328 ^a	19.338	.000	.499	.948	1.055	.948
	Performance 3 (Picture Arrangement)	.450 ^a	26.023	.000	.612	.759	1.318	.759
	Verbal 3 (Arithmetic)	.417 ^a	22.916	.000	.563	.748	1.337	.748
	Performance 4 (Block Design)	.481 ^a	29.720	.000	.662	.778	1.285	.778
	Verbal 4 (Vocabulary)	.447 ^a	20.607	.000	.523	.561	1.783	.561
	Performance 5 (Object Assembly)	.398 ^a	23.207	.000	.568	.837	1.194	.837
	Verbal 5 (Comprehension)	.423 ^a	21.610	.000	.541	.672	1.488	.672

SPSS Statistics Processor is ready | H: 980, W: 537 pt | 15

The Standardized Coefficients Beta column is to be examined next. Each of these values reflects the relative importance of each of these statistically significant predictors. In this column, you will see that the Betas range from a low of .130 to a high of .155, indicating that each has about the same degree of relative importance.

¹⁵<http://cnx.org/content/m40742/latest/19.12.png/image>

The screenshot shows the SPSS Statistics Viewer window with the following data:

Dependent Variable: Wechsler Full Scale IQ_3

Model	Variable	Beta In	t	Sig.	Partial Correlation	Tolerance	VIF	Minimum Tolerance
1	Performance 1 (Picture Completion)	.661	.017	.152	38.965	.000	.608	1.645
	Performance 2 (Coding)	.628	.015	.145	42.978	.000	.811	1.233
	Verbal 1 (Information)	.814	.021	.169	38.209	.000	.471	2.123
	Performance 5 (Object Assembly)	.606	.017	.146	35.134	.000	.539	1.854
	Verbal 3 (Arithmetic)	.665	.020	.134	33.814	.000	.587	1.704
	(Constant)	36.456	.083	.437	681	.000		
	Verbal 2 (Similarities)	.647	.012	.143	54.007	.000	.444	2.251
	Performance 4 (Block Design)	.626	.010	.155	61.016	.000	.482	2.076
	Performance 3 (Picture Arrangement)	.619	.011	.139	58.888	.000	.560	1.785
	Verbal 5 (Comprehension)	.640	.011	.152	60.551	.000	.492	2.033
2	Performance 1 (Picture Completion)	.645	.010	.149	65.543	.000	.607	1.647
	Performance 2 (Coding)	.631	.008	.146	74.406	.000	.811	1.233
	Verbal 1 (Information)	.645	.013	.134	50.123	.000	.434	2.302
	Performance 5 (Object Assembly)	.620	.010	.149	62.011	.000	.539	1.856
	Verbal 3 (Arithmetic)	.645	.011	.130	56.513	.000	.586	1.706
	Verbal 4 (Vocabulary)	.625	.013	.134	47.052	.000	.386	2.592
	Performance 3 (Picture Arrangement)	.619	.011	.139	58.888	.000	.560	1.785
	Performance 4 (Block Design)	.626	.010	.155	61.016	.000	.482	2.076
	Verbal 2 (Similarities)	.647	.012	.143	54.007	.000	.444	2.251
	Performance 2 (Coding)	.628	.015	.145	42.978	.000	.811	1.233

Excluded Variables¹

Model	Variable	Beta In	t	Sig.	Partial Correlation	Tolerance	VIF	Minimum Tolerance
1	Performance 1 (Picture Completion)	.419 ^a	24.509	.000	.589	.811	1.233	.811
	Verbal 1 (Information)	.468 ^a	23.391	.000	.571	.611	1.636	.611
	Performance 2 (Coding)	.328 ^a	19.338	.000	.499	.948	1.055	.948
	Performance 3 (Picture Arrangement)	.450 ^a	26.023	.000	.612	.759	1.318	.759
	Verbal 3 (Arithmetic)	.417 ^a	22.916	.000	.563	.748	1.337	.748
	Performance 4 (Block Design)	.481 ^a	29.720	.000	.662	.778	1.285	.778
	Verbal 4 (Vocabulary)	.447 ^a	20.607	.000	.523	.561	1.783	.561
	Performance 5 (Object Assembly)	.398 ^a	23.207	.000	.568	.837	1.194	.837
	Verbal 5 (Comprehension)	.423 ^a	21.610	.000	.541	.672	1.488	.672

In your Results section, you should discuss the assumptions that you checked; the extent to which each assumption was met or not met; the descriptive statistics; and the information in the columns that was discussed at each step of the regression process.

You have now successfully gone through the calculation of a multiple regression analysis.

¹⁶<http://cnx.org/content/m40742/latest/19.13.png/image>

Attributions

Collection: *Calculating Advanced Statistics*
Edited by: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/coll1346/1.1/>
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Introductory Chapter: Calculating Advanced Statistical Procedures in SPSS: A Self-Help and Practical Guide to Preparing Theses, Dissertations, and Manuscripts"

Used here as: "1. Introductory Chapter"

By: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/m40743/1.1/>
Pages: 1-2

Copyright: John R. Slate, Ana Rojas-LeBouef
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Selecting a Single Group for Statistical Analysis"

Used here as: "2. Selecting a Single Group for Statistical Analysis"

By: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/m40632/1.2/>
Pages: 3-24

Copyright: John R. Slate, Ana Rojas-LeBouef
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Recoding Data in SPSS"

Used here as: "3. Steps Involved in Recoding Data in SPSS"

By: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/m40708/1.2/>
Pages: 25-55

Copyright: John R. Slate, Ana Rojas-LeBouef
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Converting Data from An Excel File into SPSS"

Used here as: "4. Steps Involved in Converting Data from An Excel File into SPSS"

By: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/m40709/1.2/>
Pages: 57-74

Copyright: John R. Slate, Ana Rojas-LeBouef
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Typing Survey Data Into SPSS"

Used here as: "5. Typing Survey Data Into SPSS"

By: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/m40713/1.2/>
Pages: 75-100

Copyright: John R. Slate, Ana Rojas-LeBouef
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Internal Consistency Analysis: Part I"
Used here as: "6. Internal Consistency Analysis: Part I"
By: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/m40717/1.2/>
Pages: 101-112
Copyright: John R. Slate, Ana Rojas-LeBouef
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Internal Consistency Analysis: Part II"
Used here as: "7. Internal Consistency Analysis: Part II"
By: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/m40718/1.2/>
Pages: 113-122
Copyright: John R. Slate, Ana Rojas-LeBouef
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Internal Consistency Analysis: Part III"
Used here as: "8. Internal Consistency Analysis: Part III"
By: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/m40721/1.2/>
Pages: 123-133
Copyright: John R. Slate, Ana Rojas-LeBouef
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Factor Analysis: Part I"
Used here as: "9. Factor Analysis: Part I"
By: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/m40725/1.2/>
Pages: 135-145
Copyright: John R. Slate, Ana Rojas-LeBouef
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Factor Analysis: Part II"
Used here as: "10. Factor Analysis: Part II"
By: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/m40727/1.2/>
Pages: 147-159
Copyright: John R. Slate, Ana Rojas-LeBouef
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Multivariate Analysis of Variance: Part I"
Used here as: "11. Multivariate Analysis of Variance: Part I"
By: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/m40728/1.2/>
Pages: 161-188
Copyright: John R. Slate, Ana Rojas-LeBouef
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Multivariate Analysis of Variance: Part II"
Used here as: "12. Multivariate Analysis of Variance: Part II"
By: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/m40729/1.2/>
Pages: 189-201
Copyright: John R. Slate, Ana Rojas-LeBouef
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Multiple Analysis of Variance: Part III"
Used here as: "13. Multiple Analysis of Variance: Part III"
By: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/m40731/1.2/>
Pages: 203-214
Copyright: John R. Slate, Ana Rojas-LeBouef
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Discriminant Analysis: Assumptions"
Used here as: "14. Discriminant Analysis: Assumptions"
By: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/m40733/1.2/>
Pages: 215-221
Copyright: John R. Slate, Ana Rojas-LeBouef
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Discriminant Analysis: Part I"
Used here as: "15. Discriminant Analysis: Part I"
By: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/m40735/1.2/>
Pages: 223-231
Copyright: John R. Slate, Ana Rojas-LeBouef
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Discriminant Analysis: Part II"
Used here as: "16. Discriminant Analysis: Part II"
By: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/m40736/1.2/>
Pages: 233-243
Copyright: John R. Slate, Ana Rojas-LeBouef
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Discriminant Analysis: Part III"
Used here as: "17. Discriminant Analysis: Part III"
By: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/m40737/1.2/>
Pages: 245-253
Copyright: John R. Slate, Ana Rojas-LeBouef
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Multiple Regression: Assumptions"
Used here as: "18. Multiple Regression: Assumptions"
By: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/m40738/1.2/>
Pages: 255-263
Copyright: John R. Slate, Ana Rojas-LeBouef
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Multiple Regression: Part I"
Used here as: "19. Multiple Regression: Part I"
By: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/m40739/1.2/>
Pages: 265-274
Copyright: John R. Slate, Ana Rojas-LeBouef
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Multiple Regression: Part II"
Used here as: "20. Multiple Regression: Part II"
By: John R. Slate, Ana Rojas-LeBouef
URL: <http://cnx.org/content/m40742/1.2/>
Pages: 275-288
Copyright: John R. Slate, Ana Rojas-LeBouef
License: <http://creativecommons.org/licenses/by/3.0/>

About Connexions

Since 1999, Connexions has been pioneering a global system where anyone can create course materials and make them fully accessible and easily reusable free of charge. We are a Web-based authoring, teaching and learning environment open to anyone interested in education, including students, teachers, professors and lifelong learners. We connect ideas and facilitate educational communities.

Connexions's modular, interactive courses are in use worldwide by universities, community colleges, K-12 schools, distance learners, and lifelong learners. Connexions materials are in many languages, including English, Spanish, Chinese, Japanese, Italian, Vietnamese, French, Portuguese, and Thai. Connexions is part of an exciting new information distribution system that allows for **Print on Demand Books**. Connexions has partnered with innovative on-demand publisher QOOP to accelerate the delivery of printed course materials and textbooks into classrooms worldwide at lower prices than traditional academic publishers.