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

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

Chapter 1

1. Introductory Chapter[1]

NOTE: This chapter is published by NCPEA Press[2] 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 Help[3]) 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.

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

[1] This 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 Analysis[1]

NOTE: This chapter is published by NCPEA Press[2] 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 Help[3]) 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

[1] This 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 Publications[4] , 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.

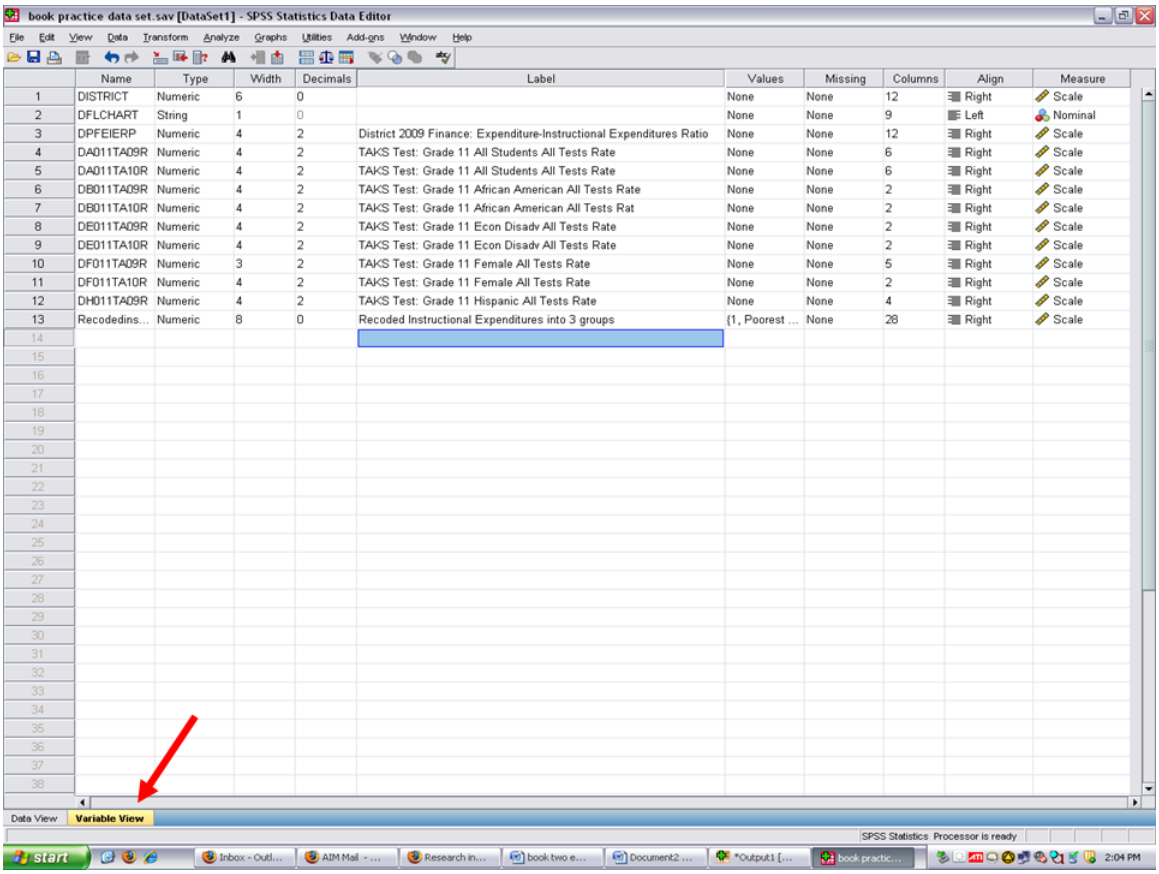
Janet Tareilo, is a Professor at Stephen F. Austin State University and serves as the Assistant Director of NCPEA Publications. Dr. Tareilo also serves as an Assistant Editor of the NCPEA Connexions Project and as a editor and reviewer for several national and international journals in educational leadership.

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

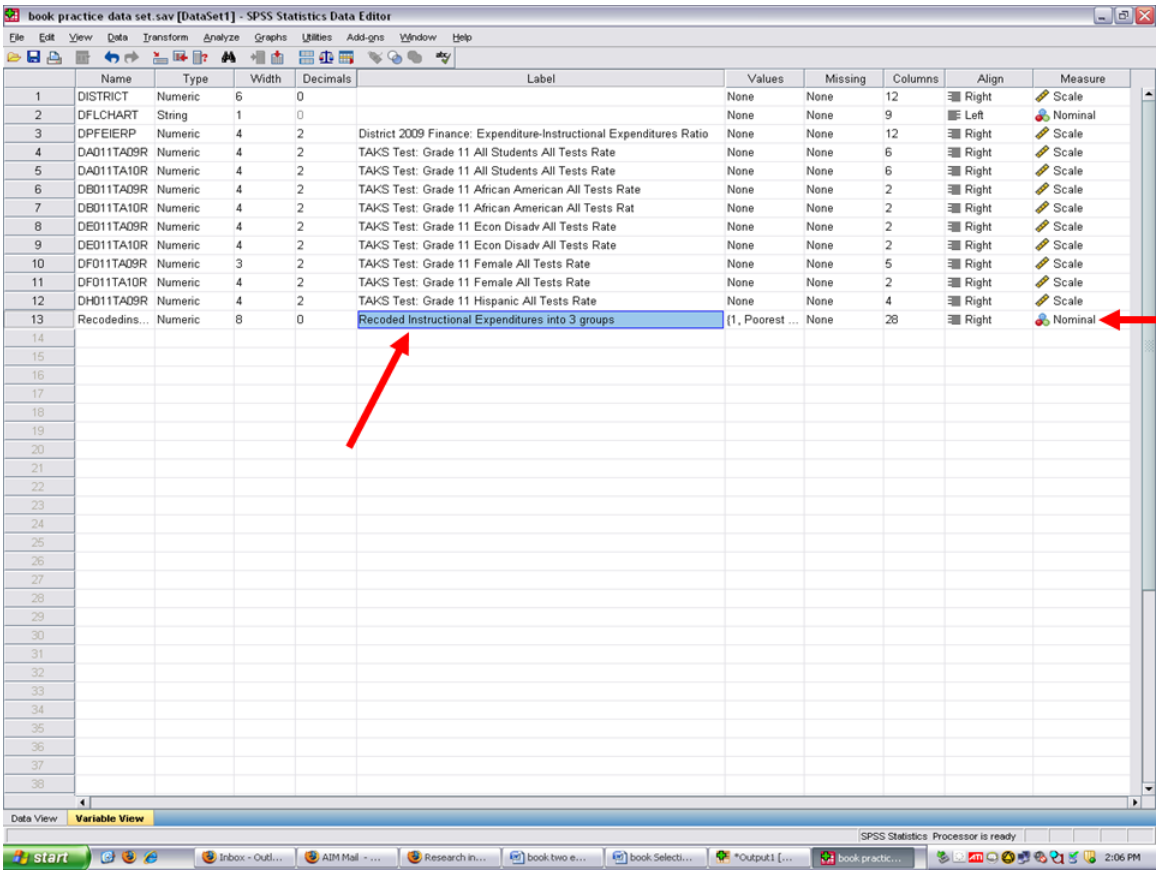


[6]

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.

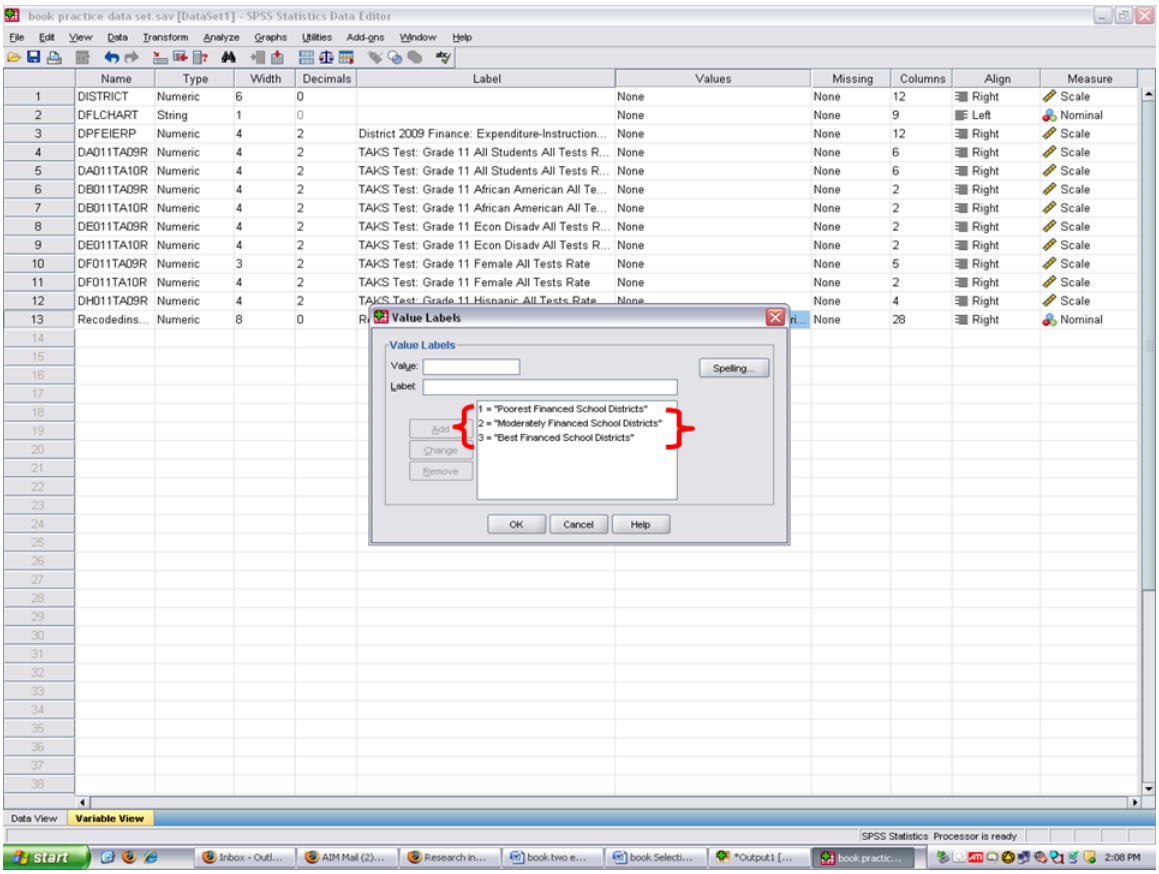
[6] http://cnx.org/content/m40632/latest/1.1.png/image



[7]

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.

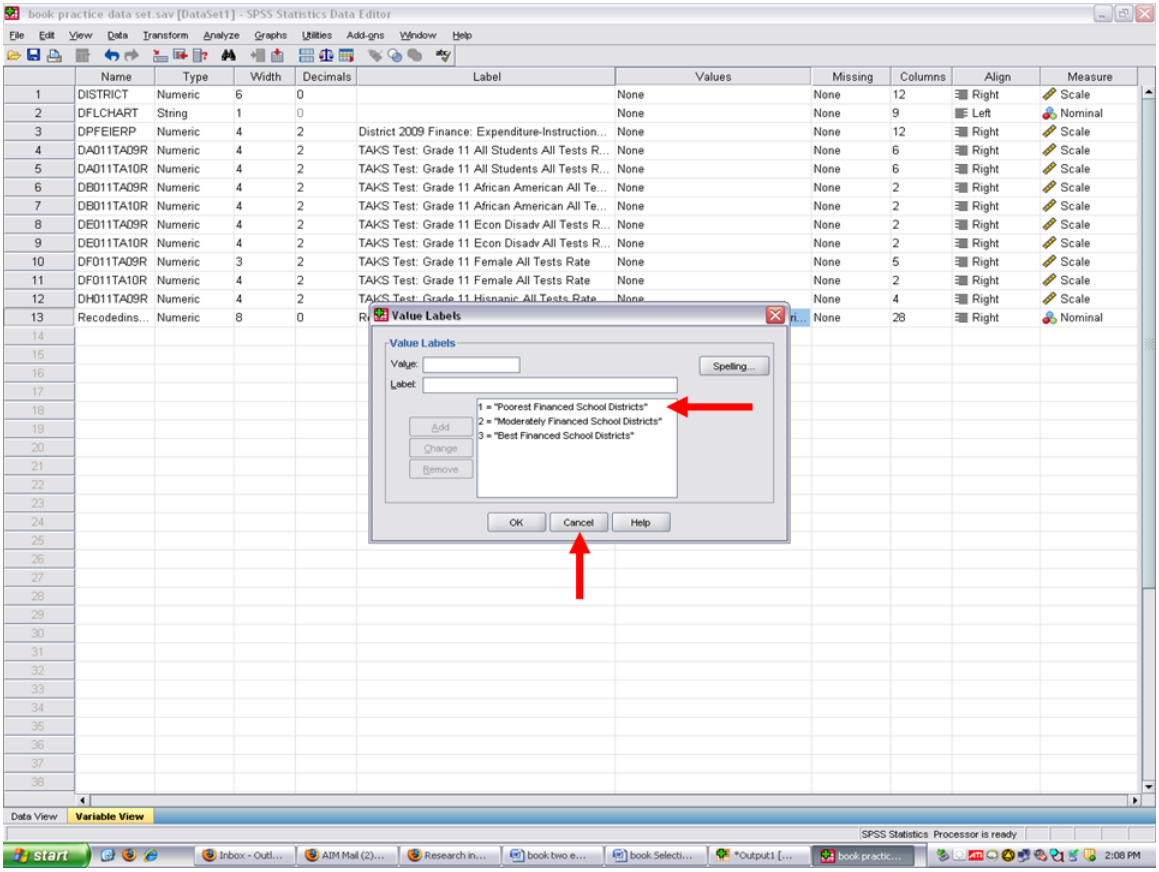
[7] http://cnx.org/content/m40632/latest/1.3.png/image



[8]

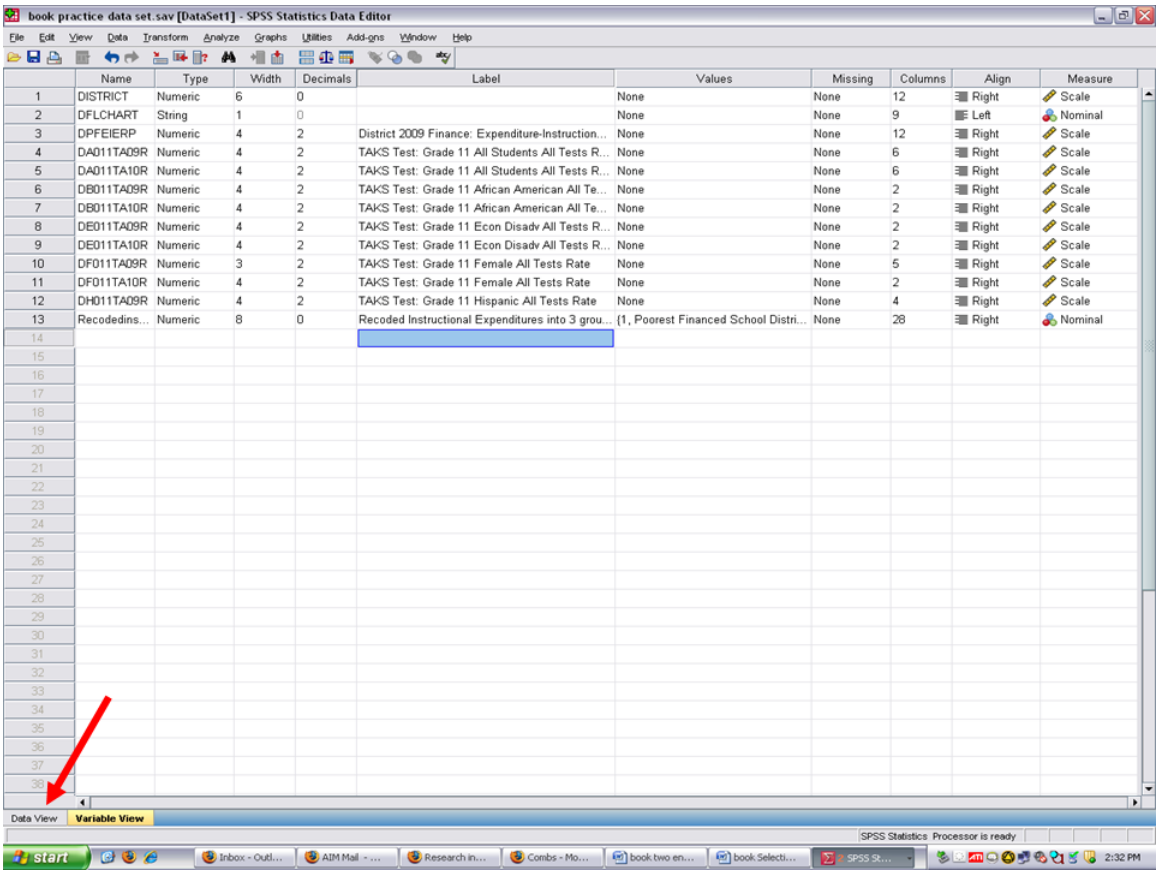
For purposes of this chapter, we will have statistics calculated for group 1 which is the Poorest Financed School Districts. Now that we know that this group is numbered as group 1, we will cancel out of this screen.

[8] http://cnx.org/content/m40632/latest/1.4.png/image



Click on the data view screen.

[9] http://cnx.org/content/m40632/latest/1.5.png/image

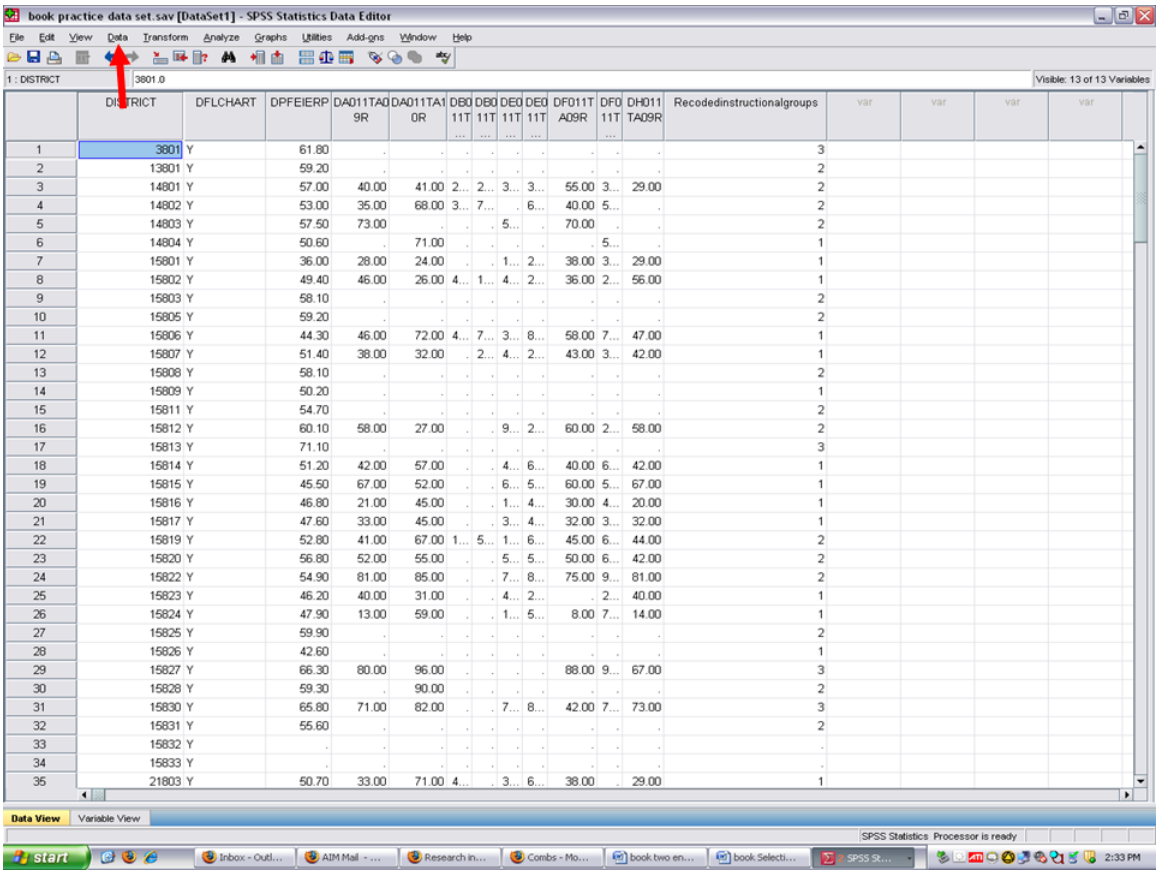


[10]

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

√ Data

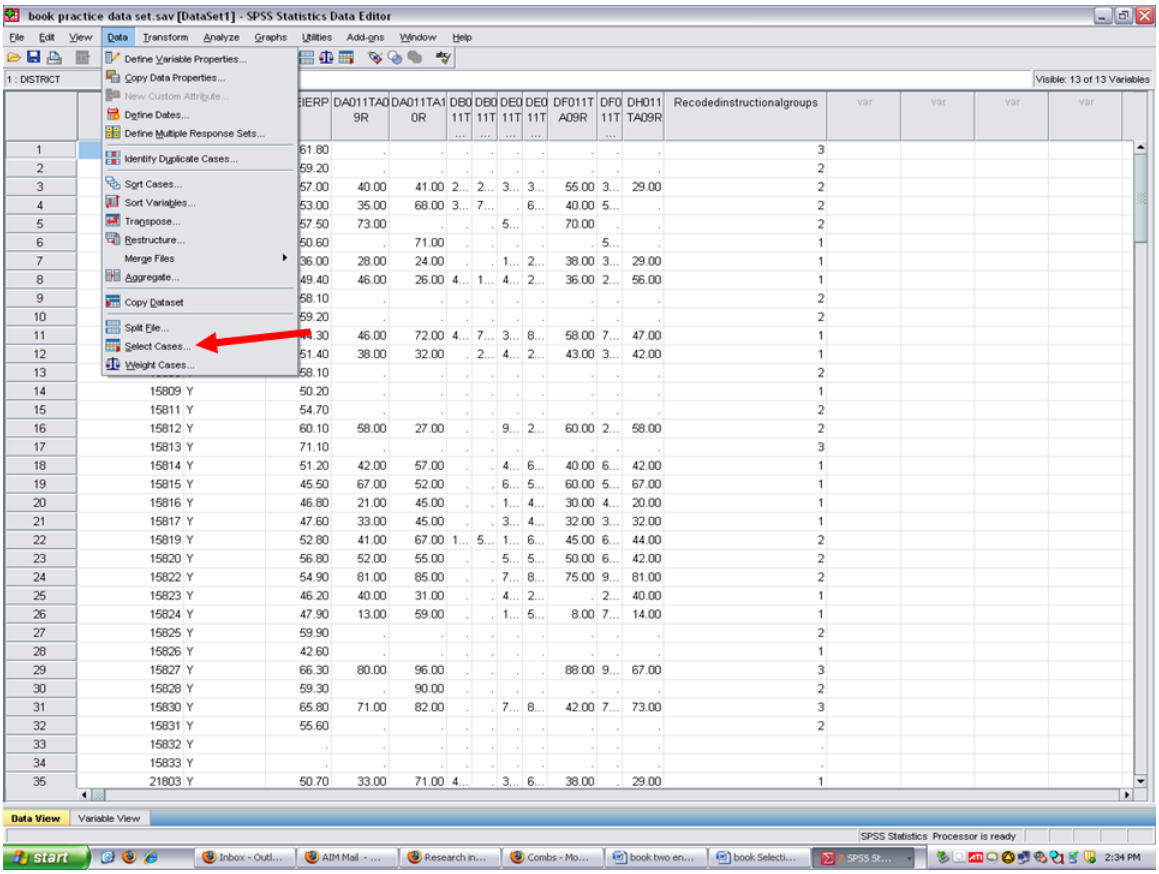
[10] http://cnx.org/content/m40632/latest/1.6.png/image



[11]

√ **Select Cases**

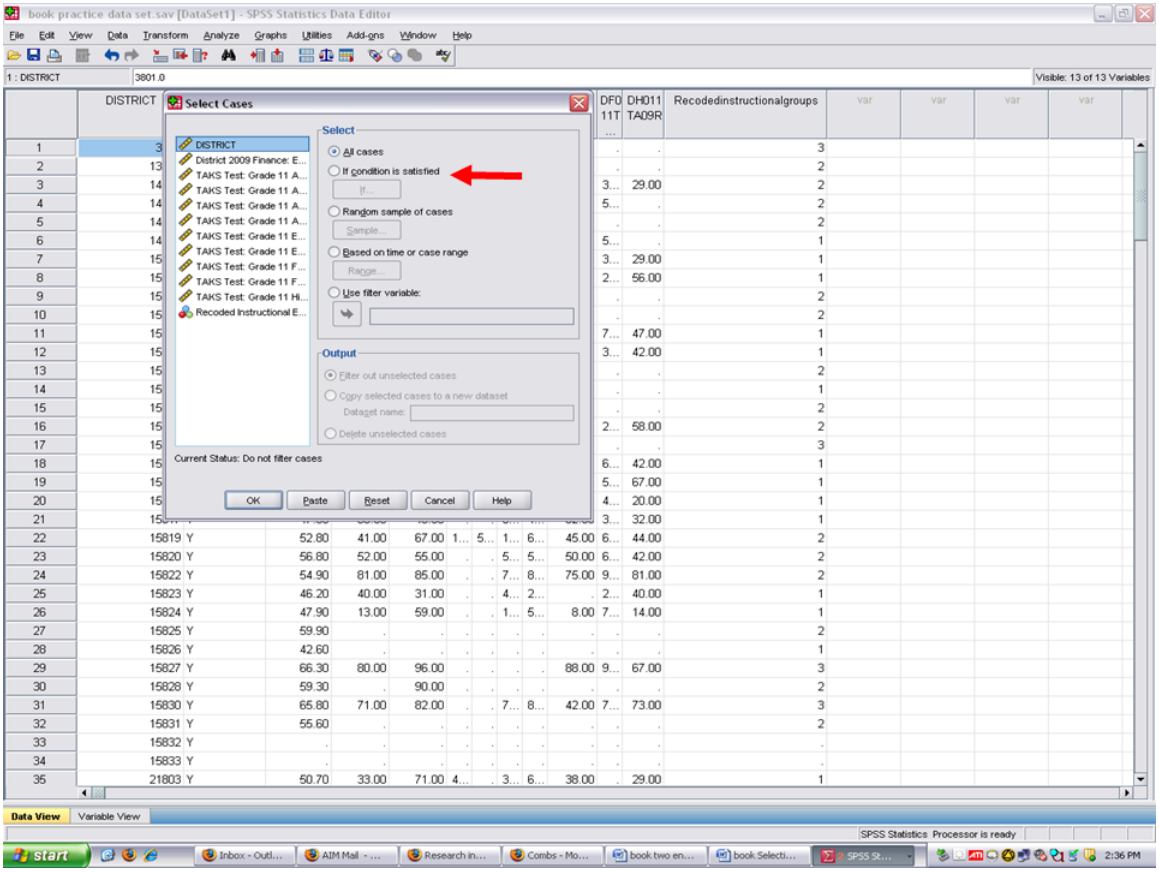
[11] http://cnx.org/content/m40632/latest/1.7.png/image



12

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

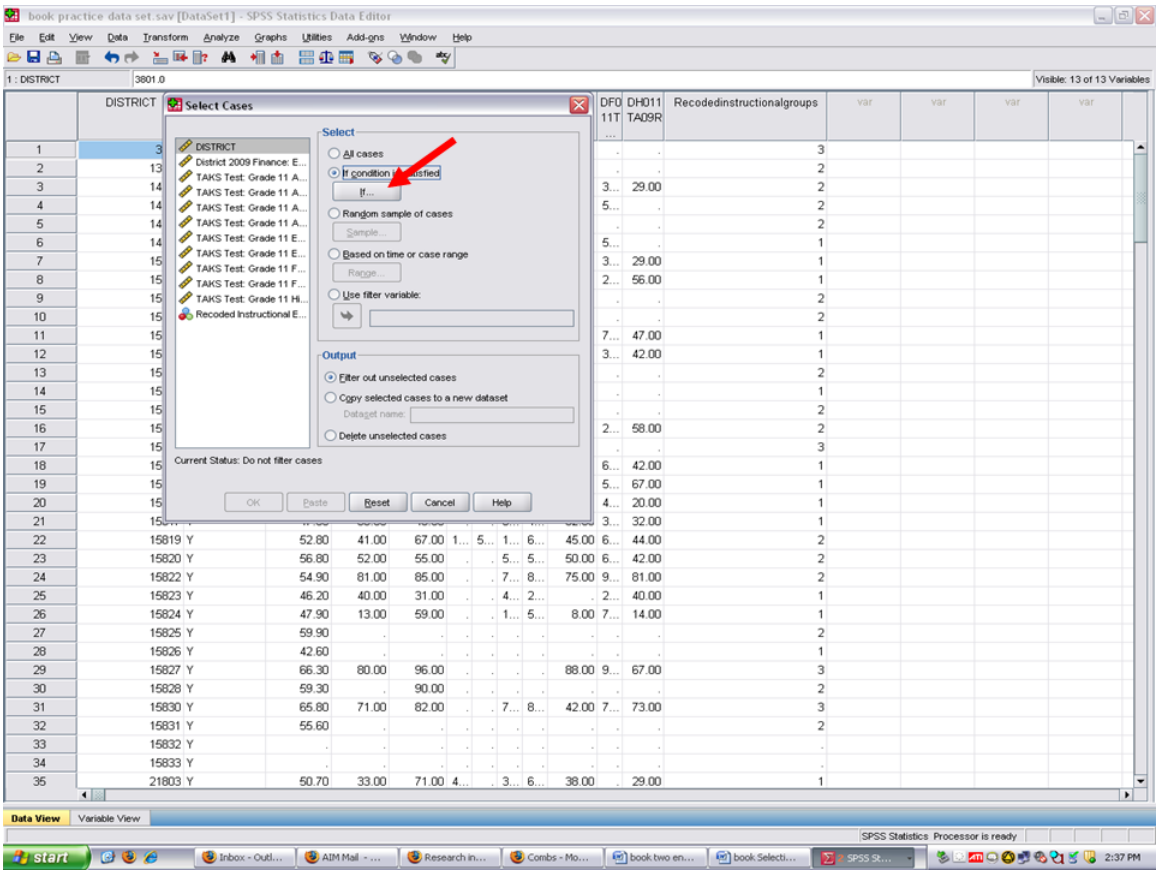
[12] http://cnx.org/content/m40632/latest/1.9.png/image



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.

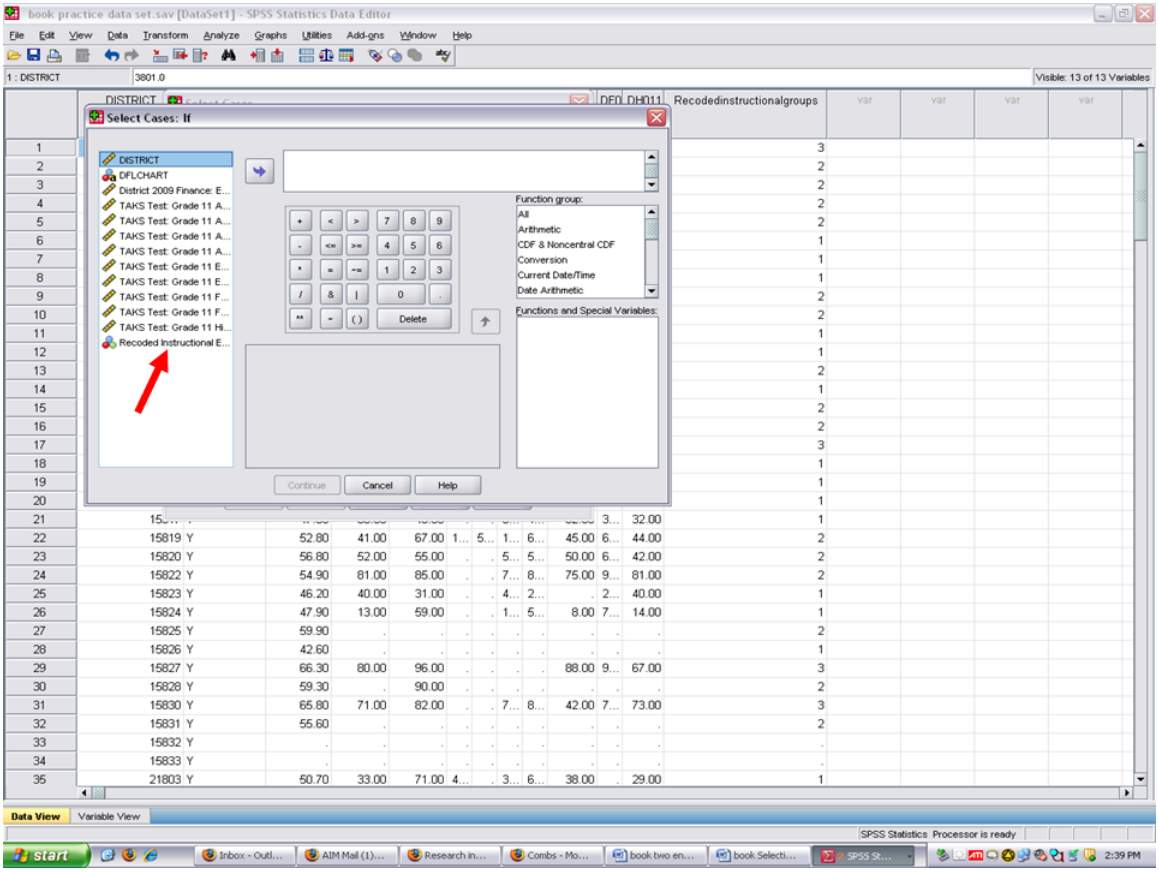
[13] http://cnx.org/content/m40632/latest/1.10.png/image



14

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.

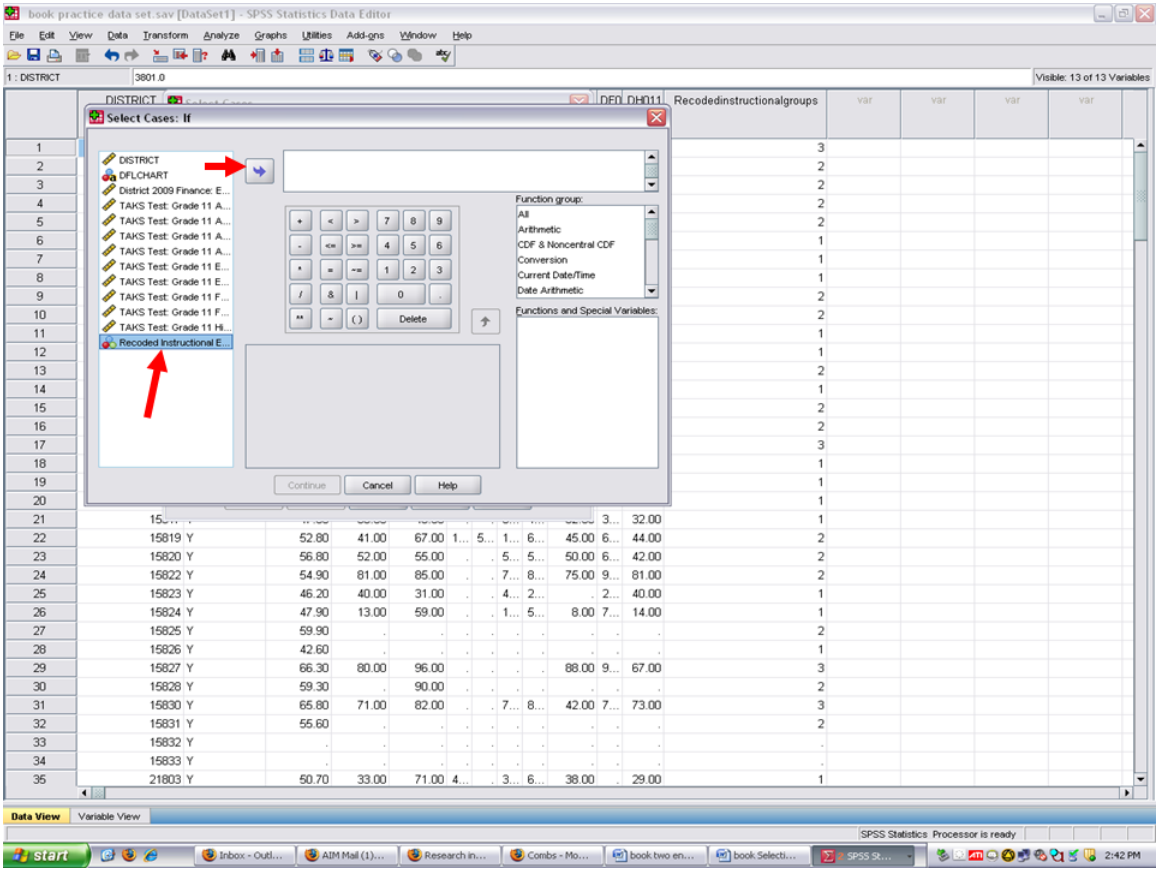
[14] http://cnx.org/content/m40632/latest/1.11.png/image



[15]

√ **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

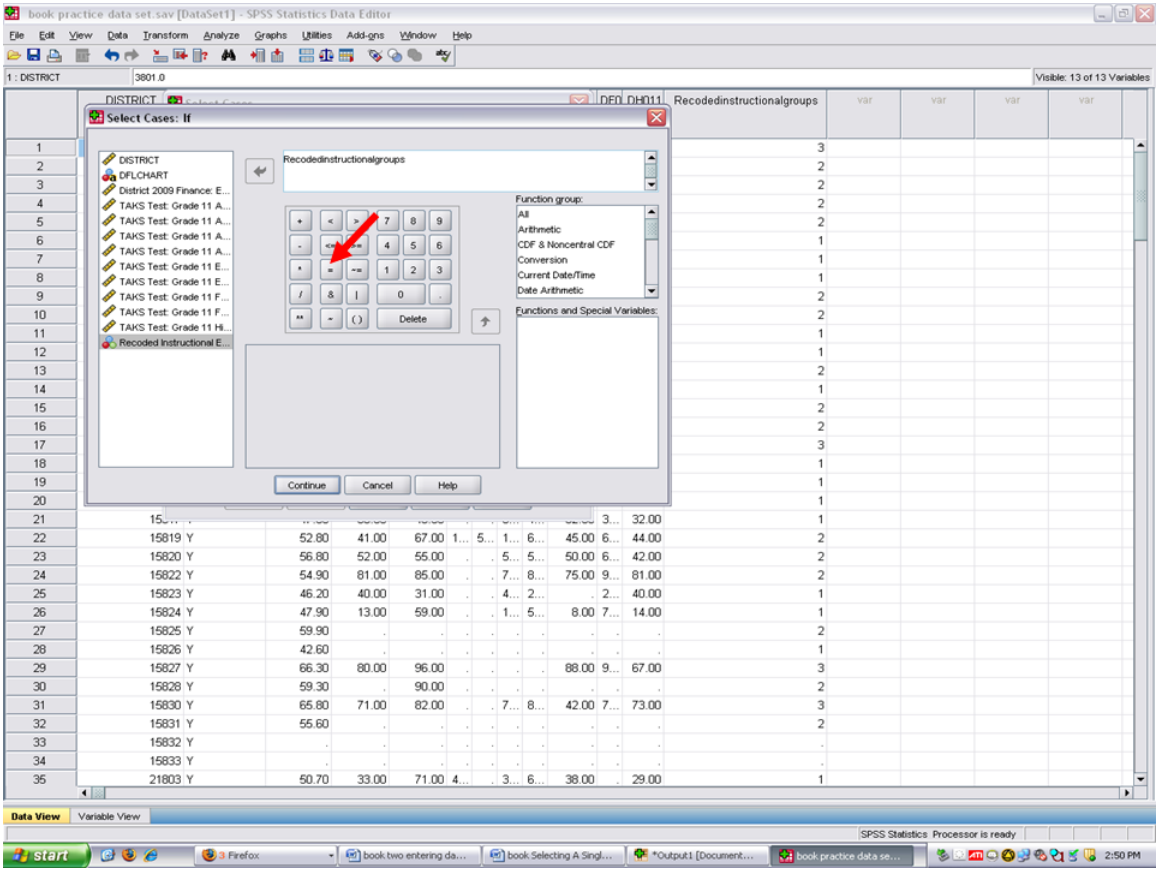


[16]

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



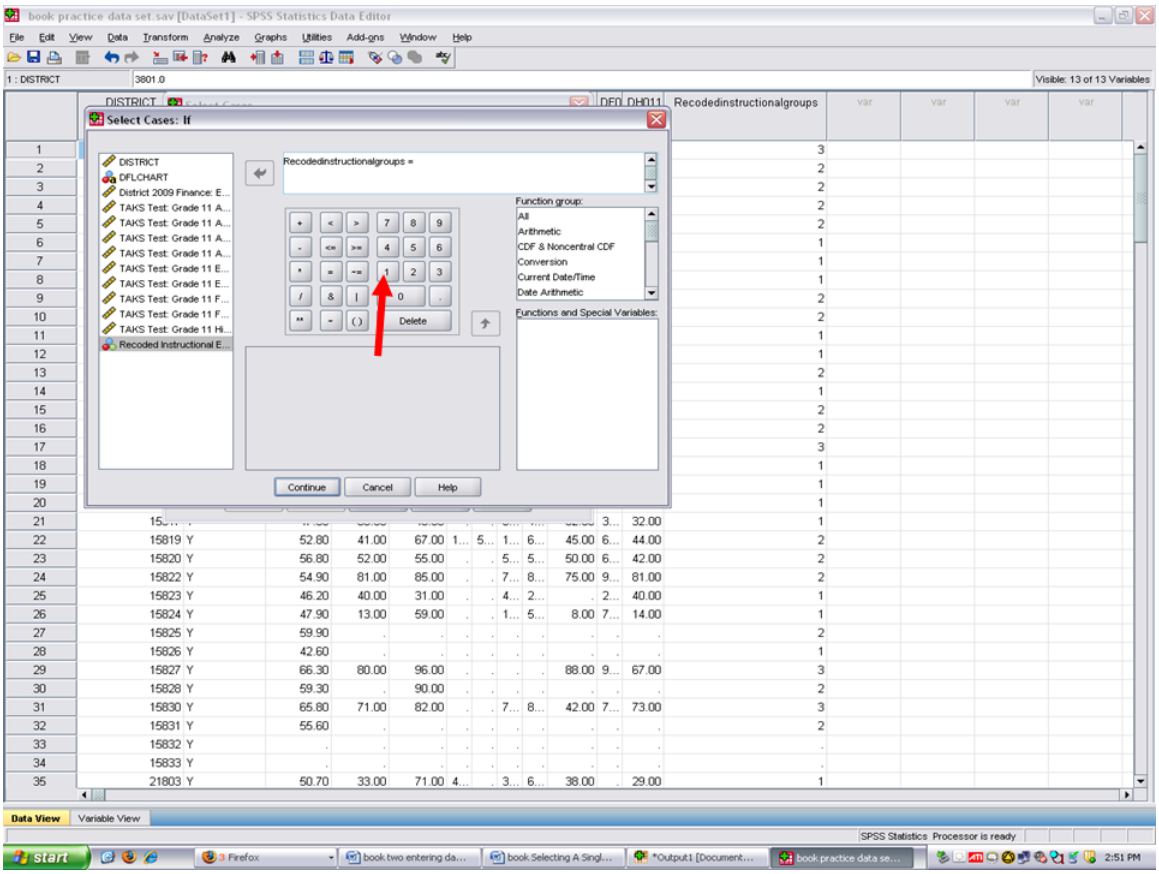
[17]

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.

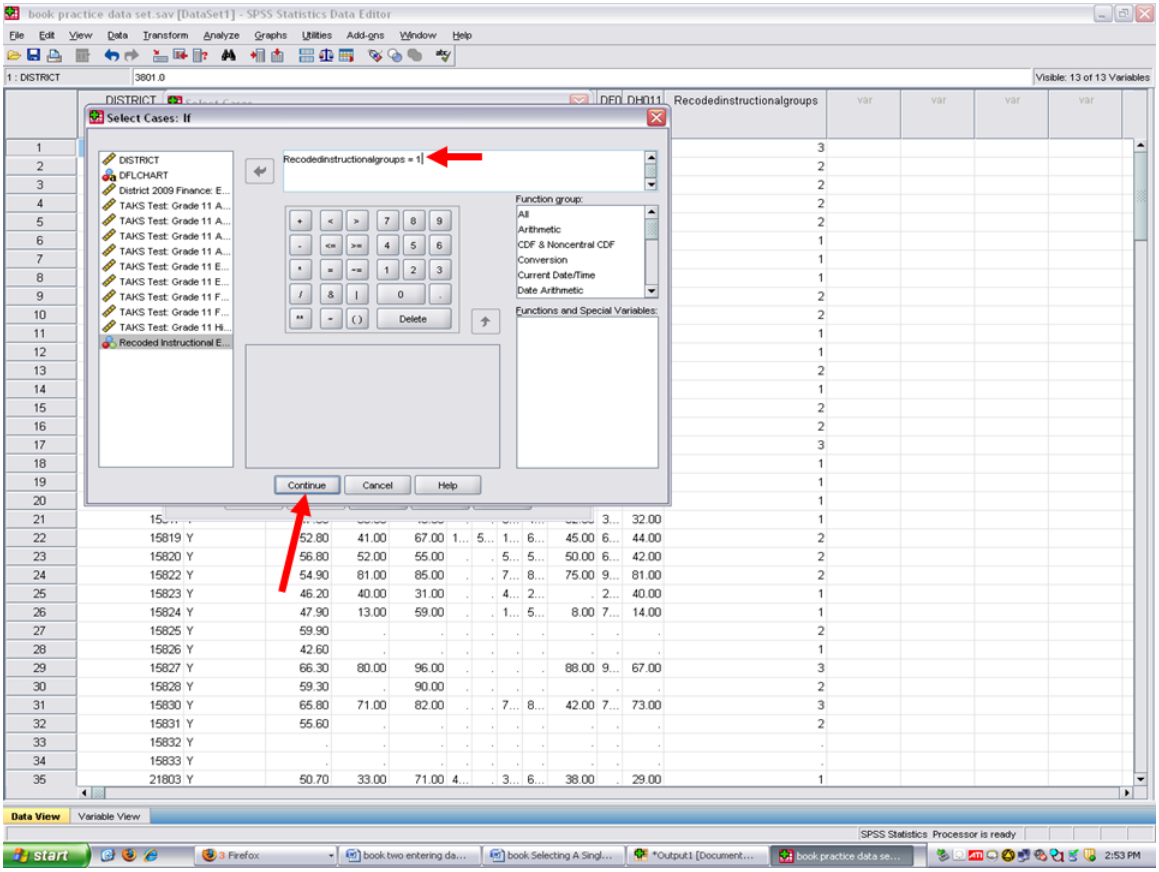
[17] http://cnx.org/content/m40632/latest/1.14.png/image



[18]

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

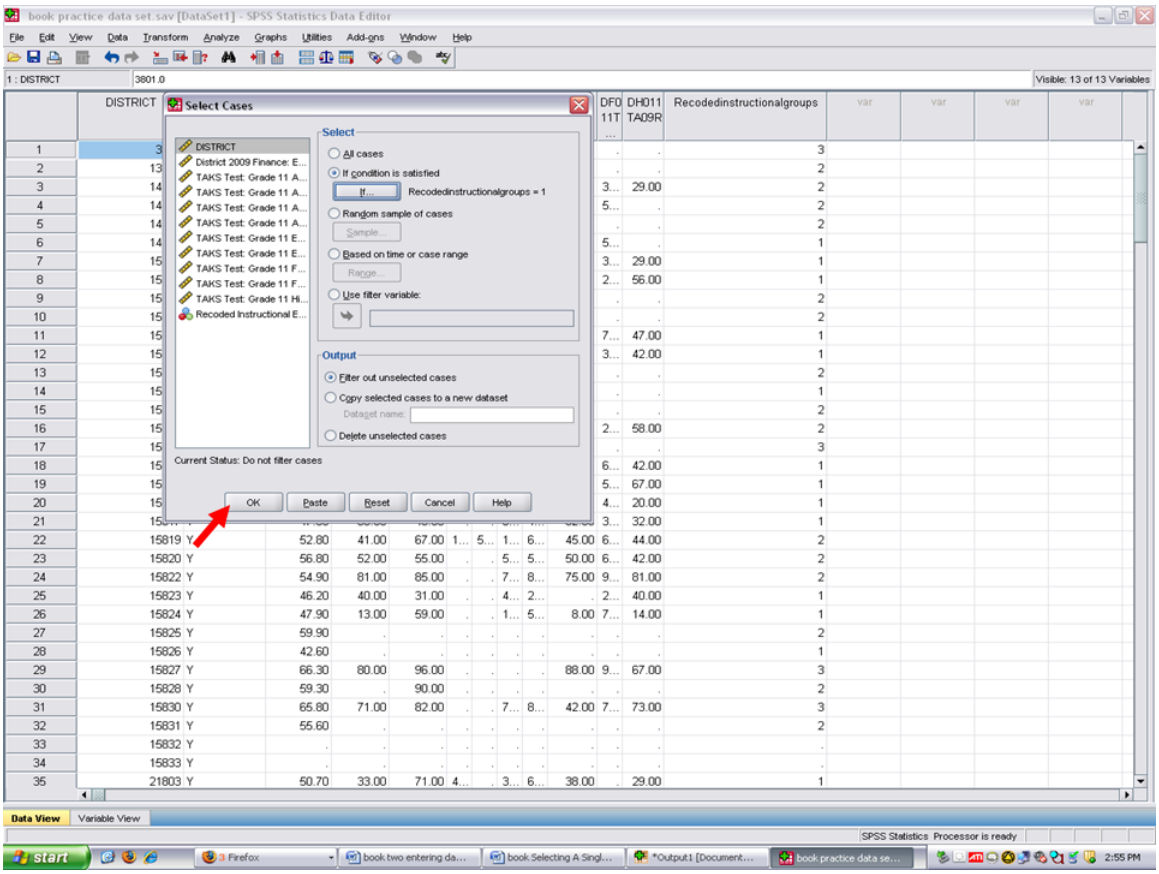
[18] http://cnx.org/content/m40632/latest/1.15.png/image



19

√ **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

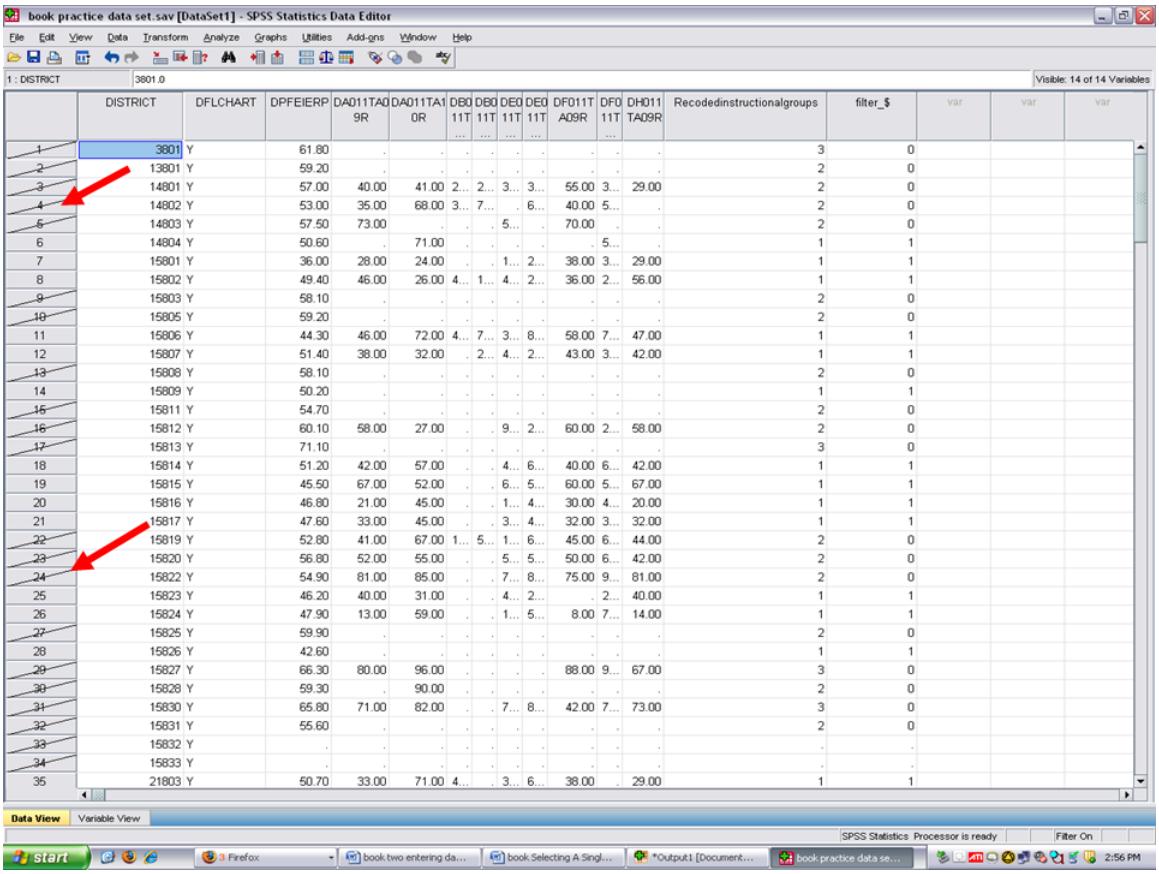


[20]

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



[21]

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

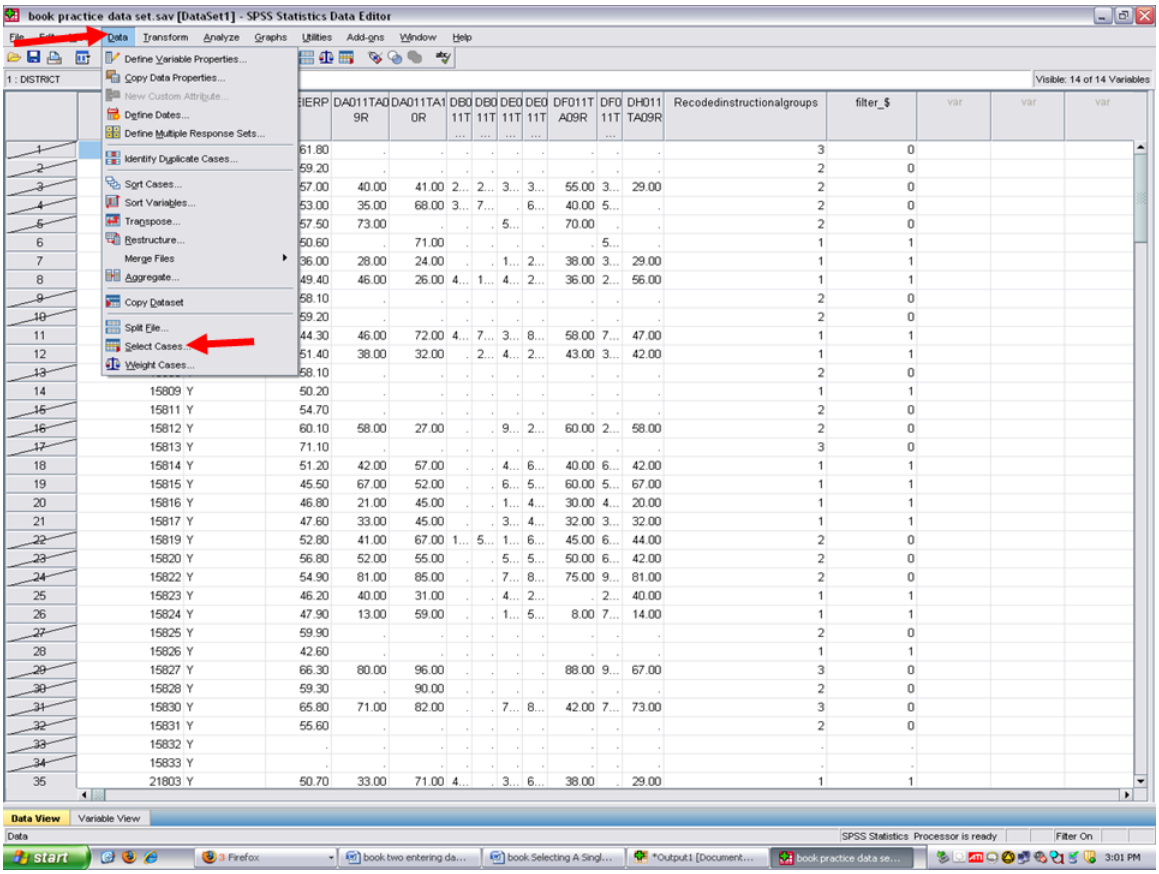
After obtaining the calculations of interest for this group, if you wish to select the Best Financed School Districts, you would follow the steps above in selecting group 1. Replace group = 1, with group = 2.

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

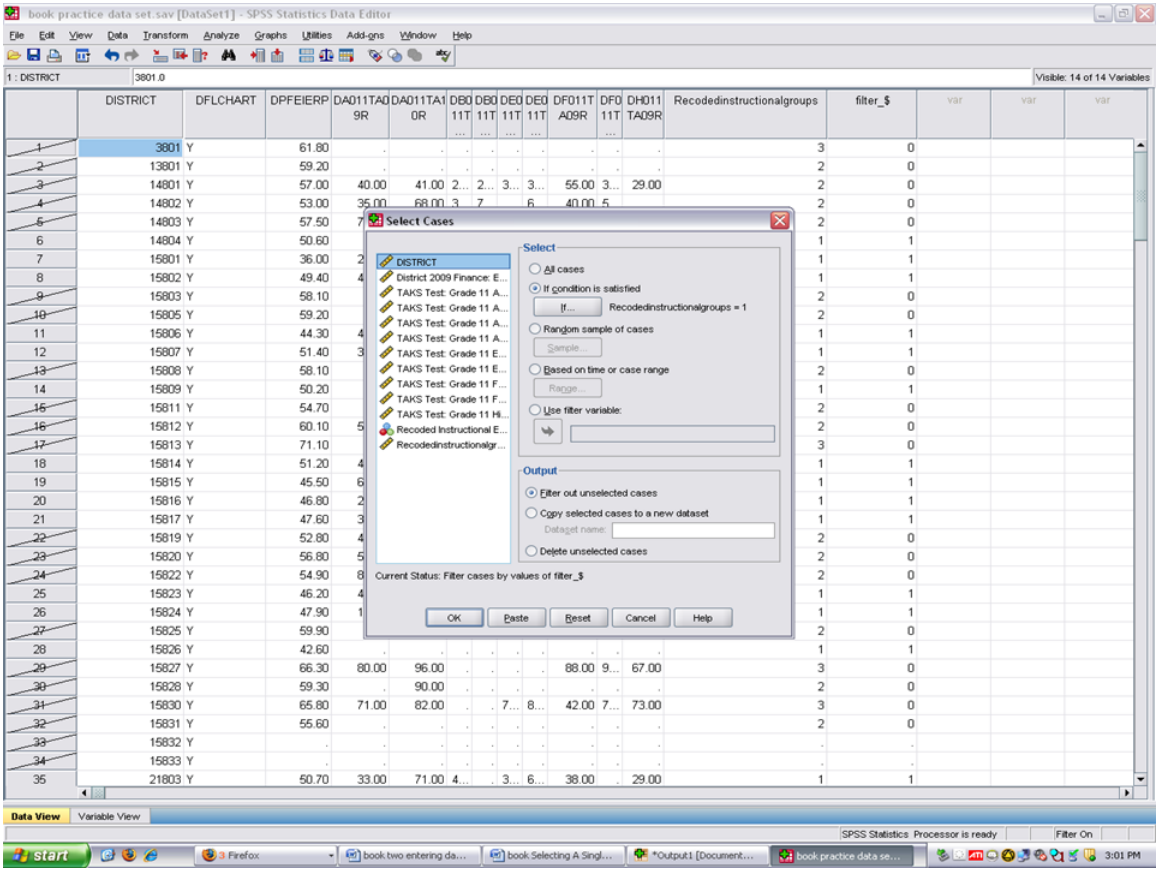
[21] http://cnx.org/content/m40632/latest/1.18.png/image



22

After clicking on Select Cases, this screen will appear.

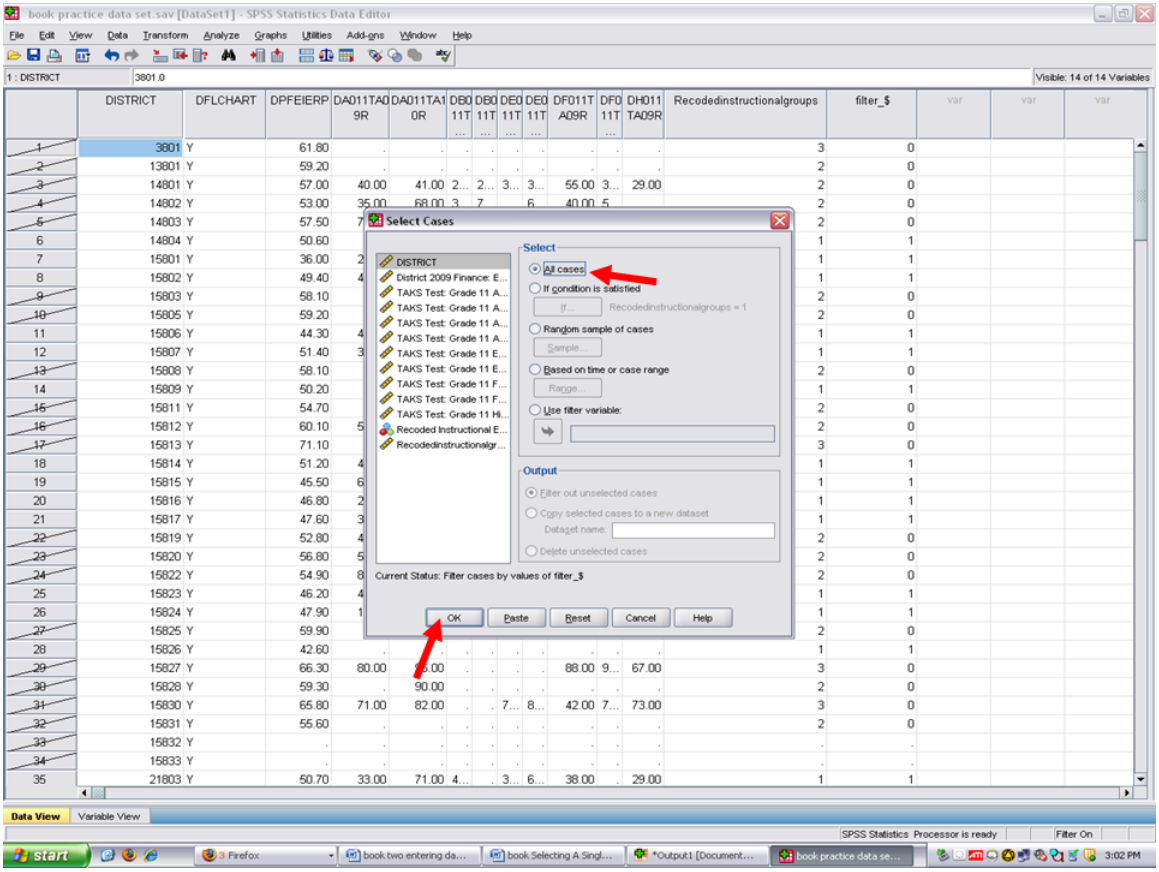
[22] http://cnx.org/content/m40632/latest/1.19.png/image



[23]

√ **Click on All Cases**

[23] http://cnx.org/content/m40632/latest/1.20.png/image

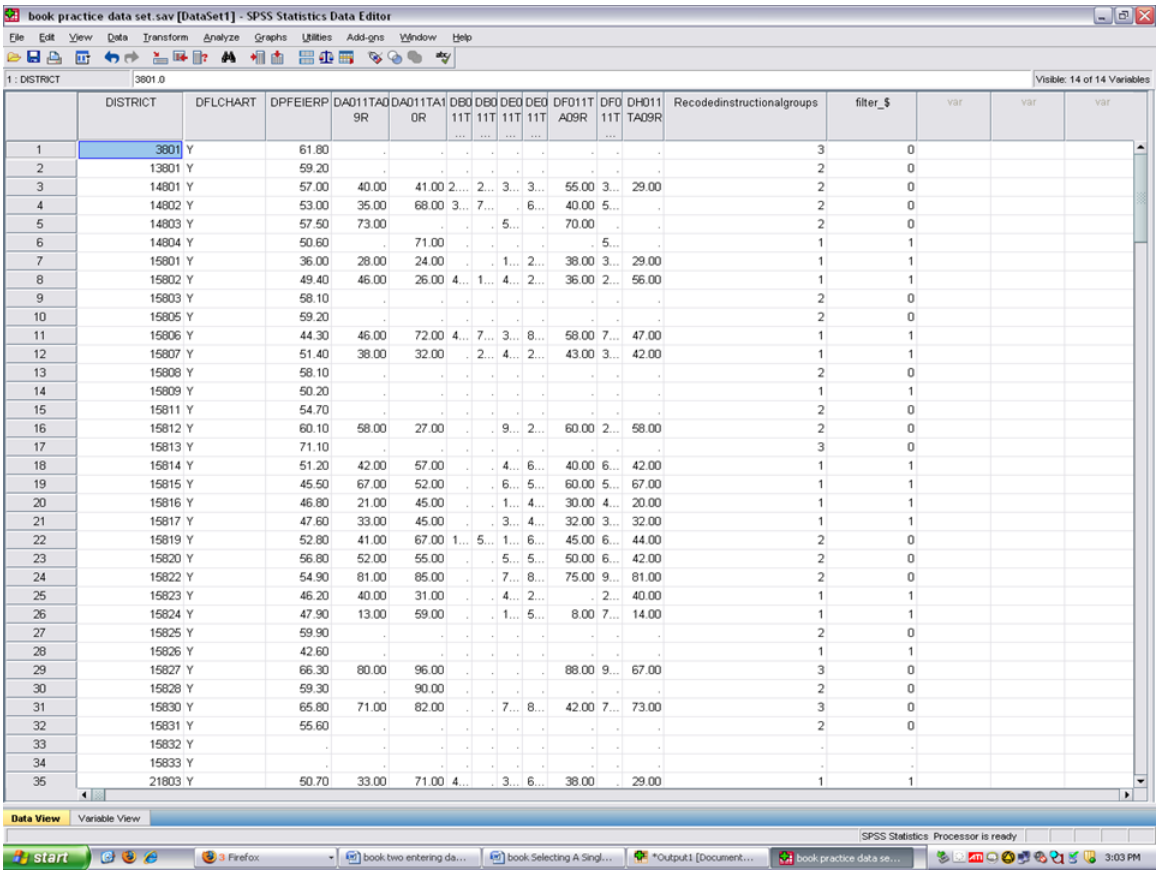


[24]

√ **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



[25]

[25] http://cnx.org/content/m40632/latest/1.22.png/image

Chapter 3

3. Steps Involved in Recoding Data in SPSS[1]

NOTE: This chapter is published by NCPEA Press[2] 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 Help[3]) 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

[1] This 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

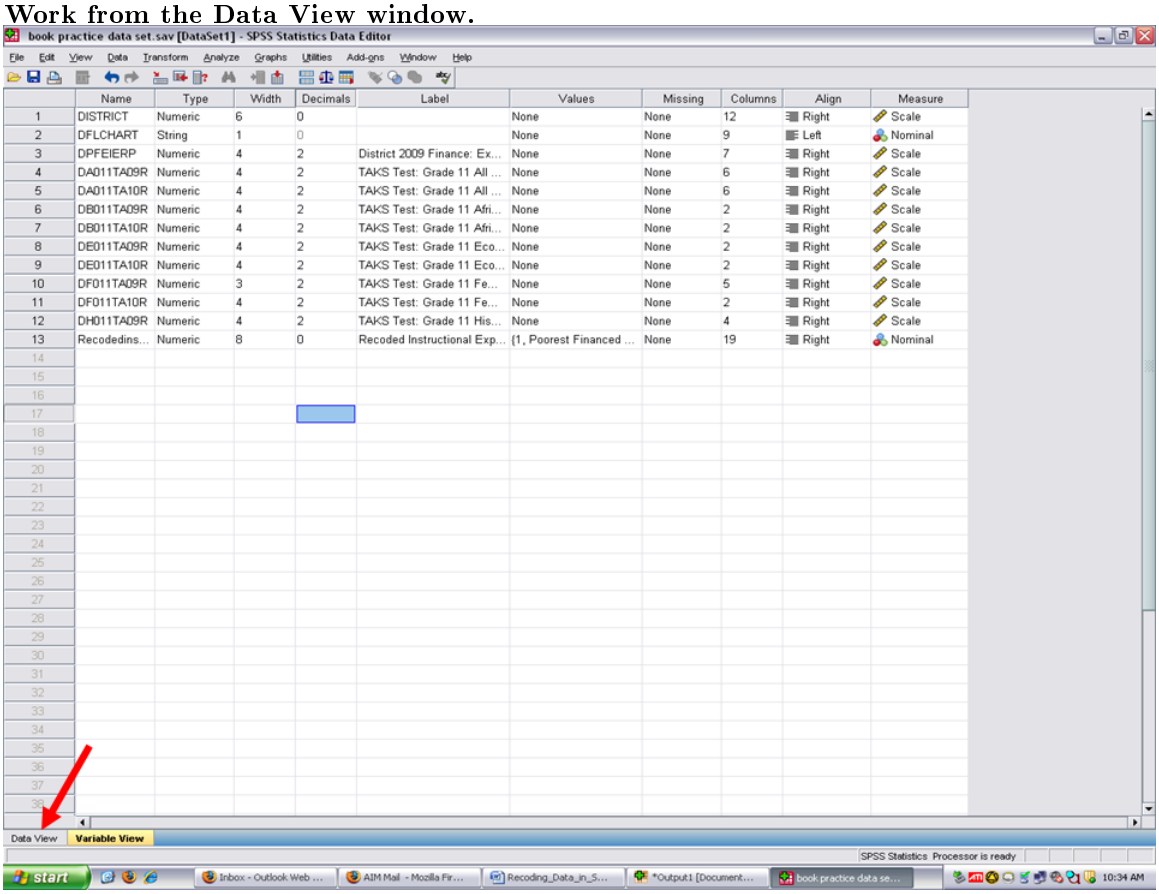
Theodore B. Creighton, is a Professor at Virginia Tech and the Publications Director for NCPEA Publications[4] , the Founding Editor of Education Leadership Review,[5] and the Senior Editor of the NCPEA Connexions Project.

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.

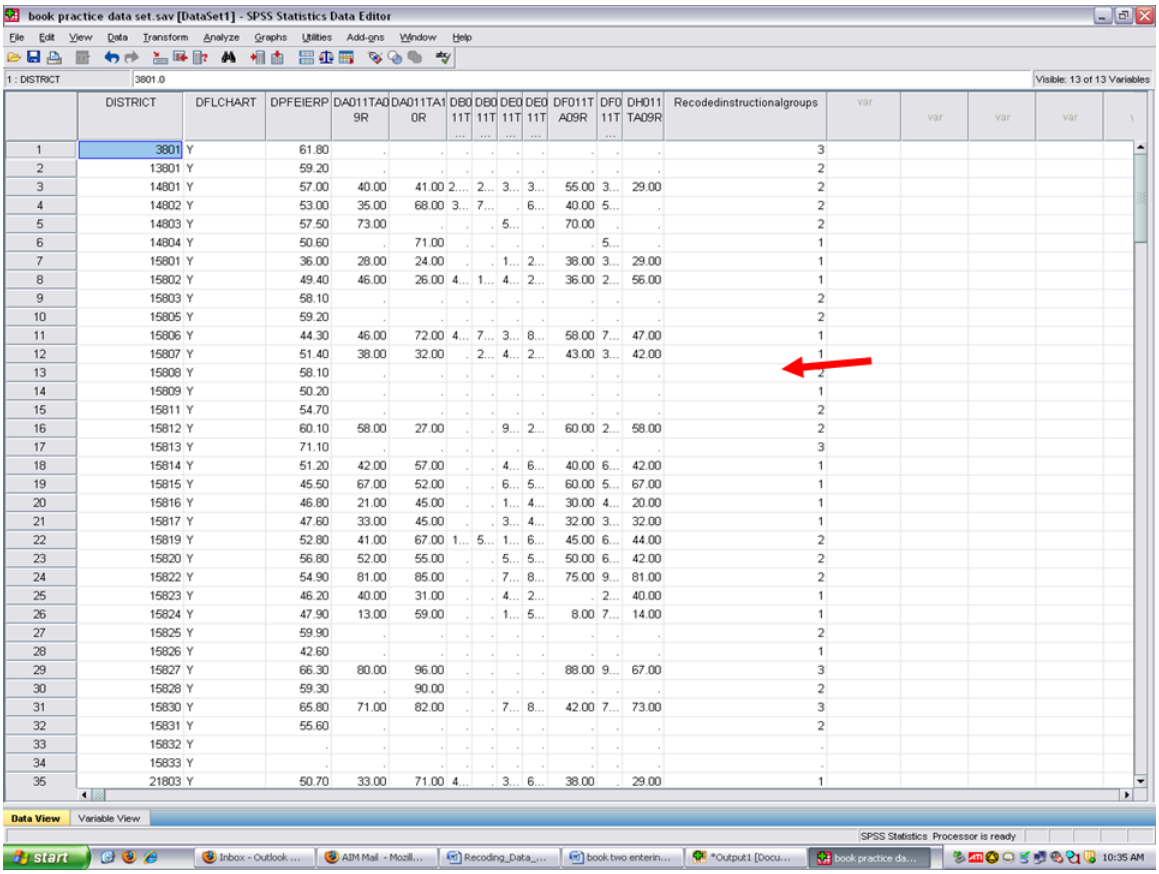
Work from the Data View window.



[6]

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

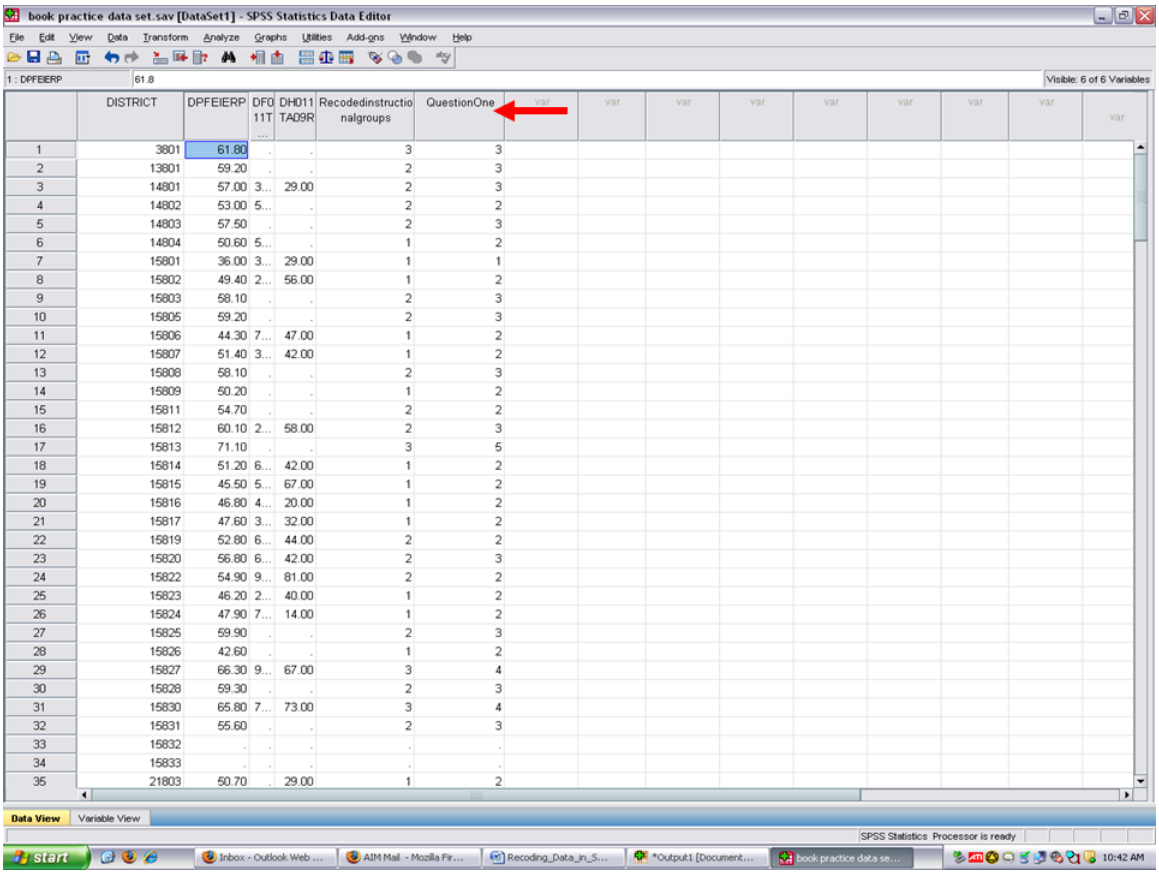
[4] http://www.ncpeapublications.org
[5] http://ncpeapublications.org/about-elr.html
[6] http://cnx.org/content/m40708/latest/2.1.png/image



[7]

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



[8]

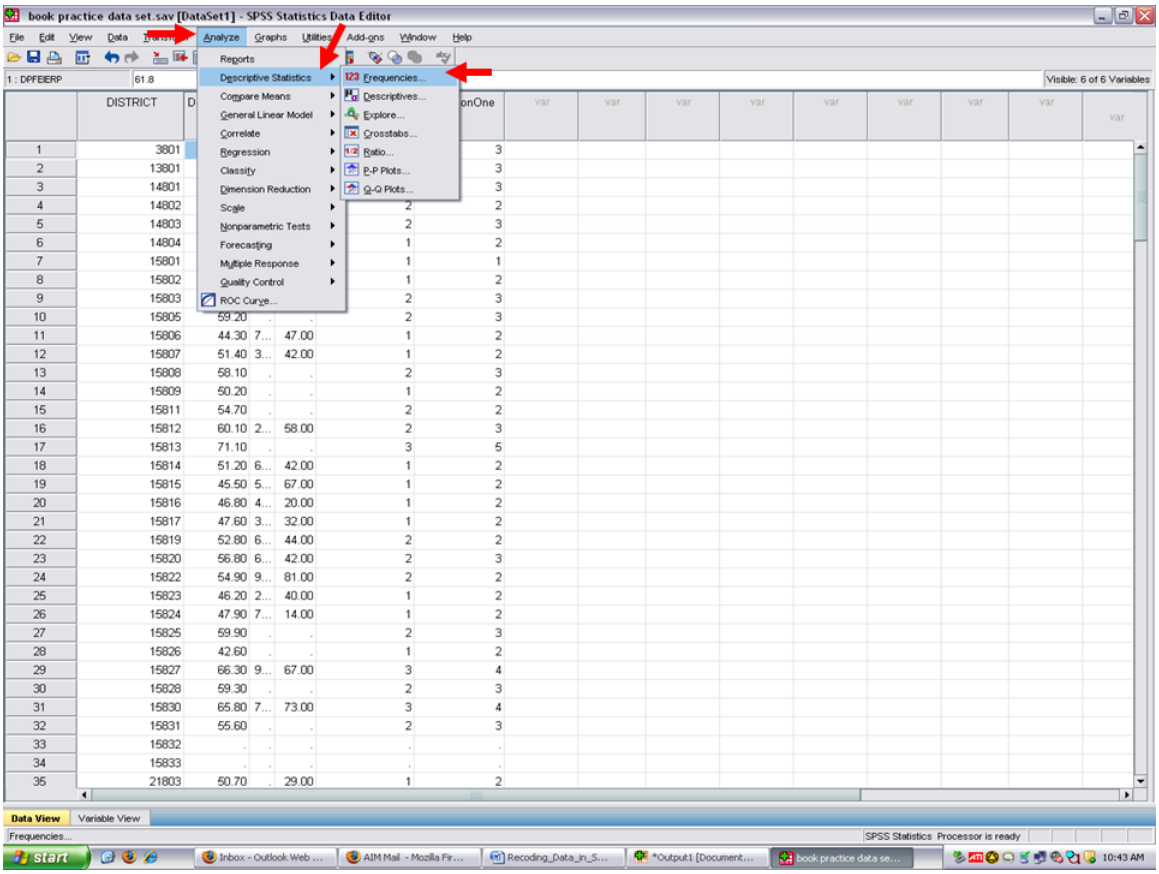
Identify the values within this variable that you wish to recode. You may need to run the frequencies procedure to list out all of the values present for the variable of interest.

√ Click on Analyze

√ Click on Descriptive Statistics

√ Click on Frequencies

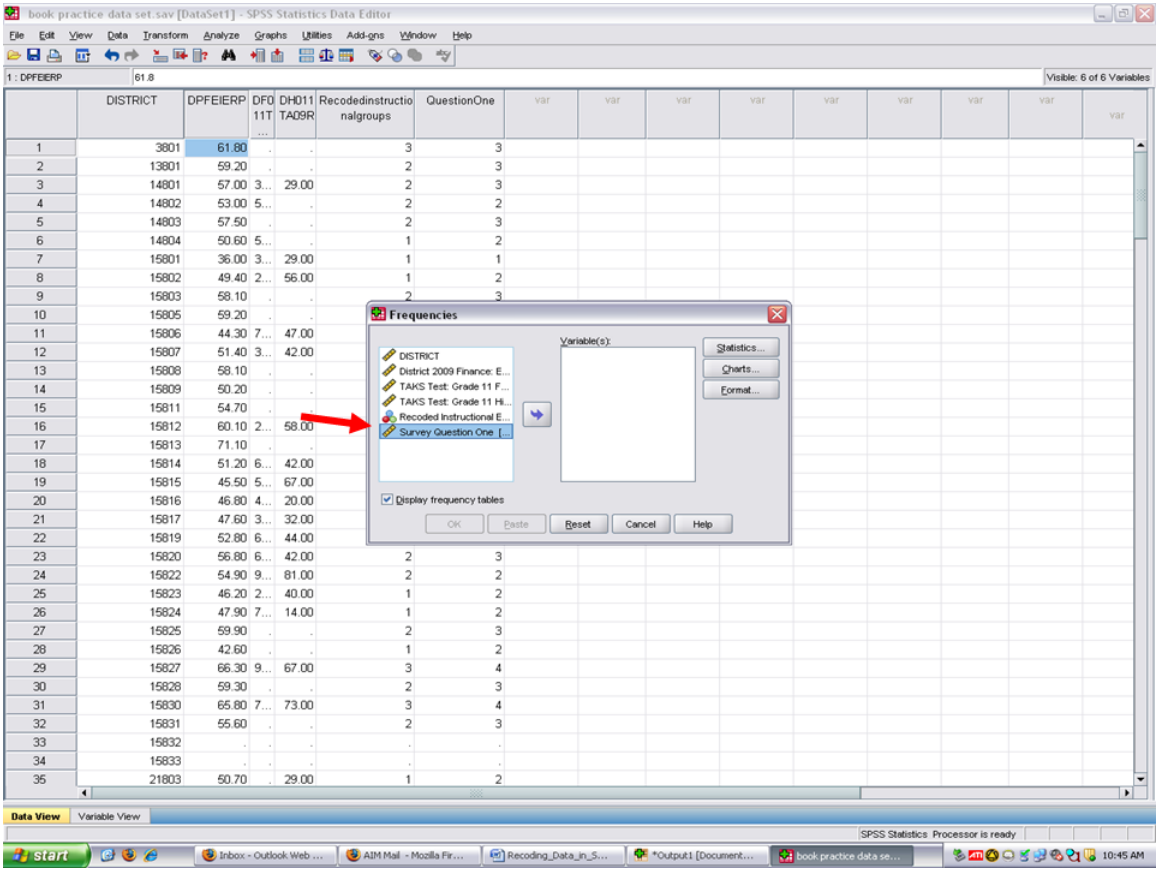
[8] http://cnx.org/content/m40708/latest/2.3.png/image



[9]

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.

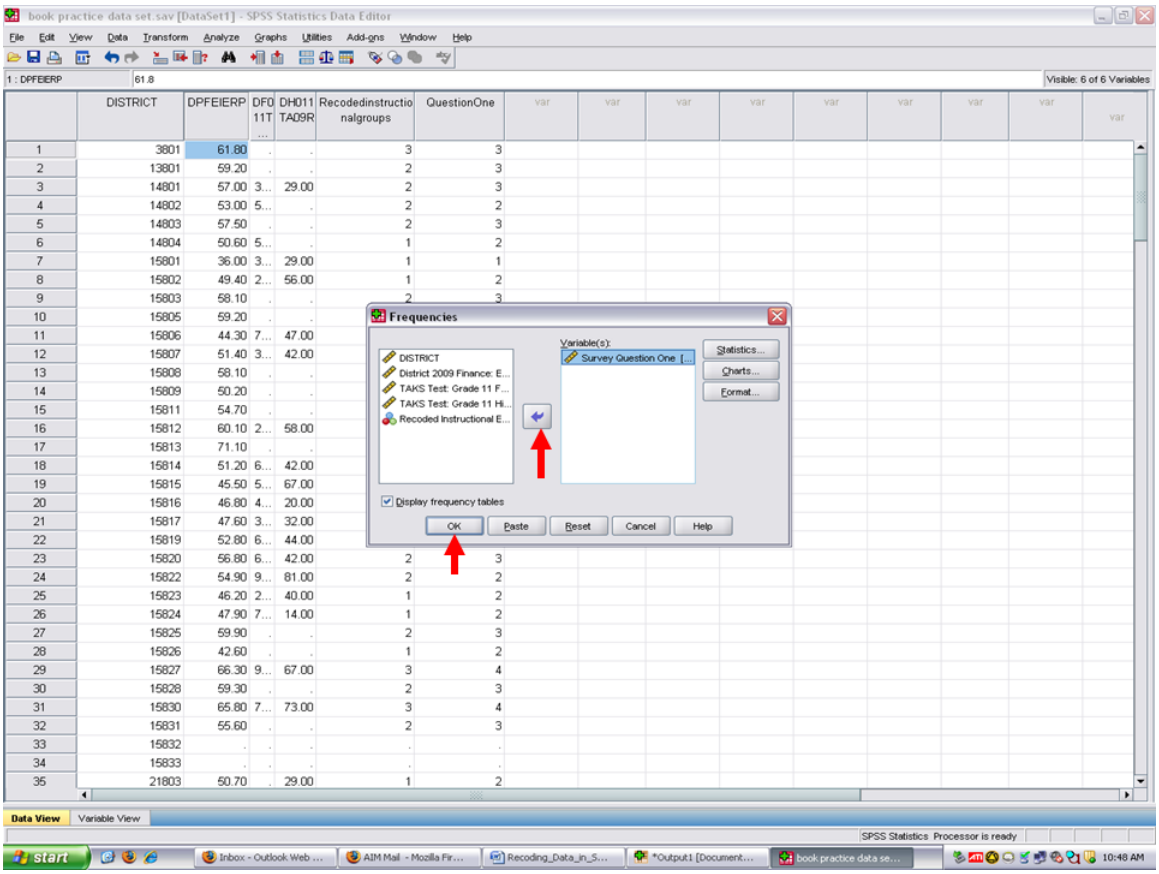
[9] http://cnx.org/content/m40708/latest/2.4.png/image



10

√ **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

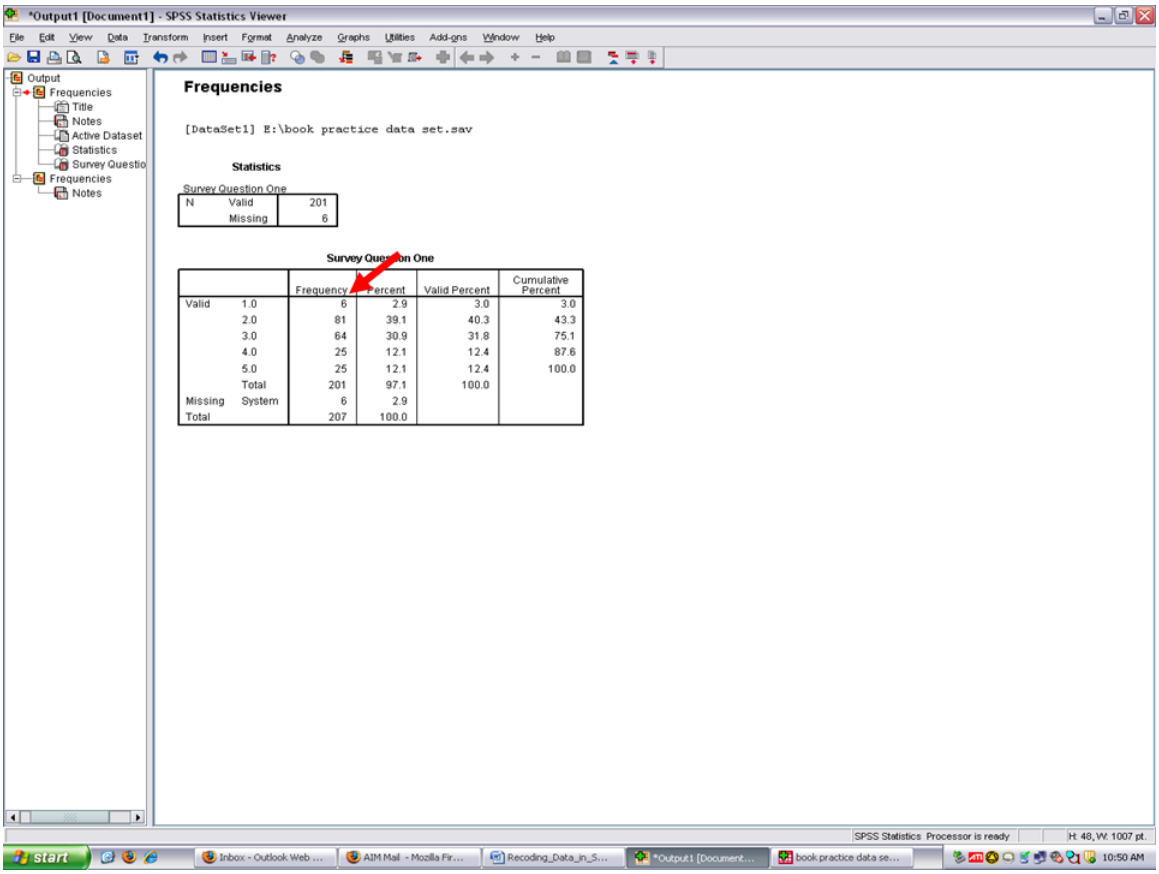


[11]

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

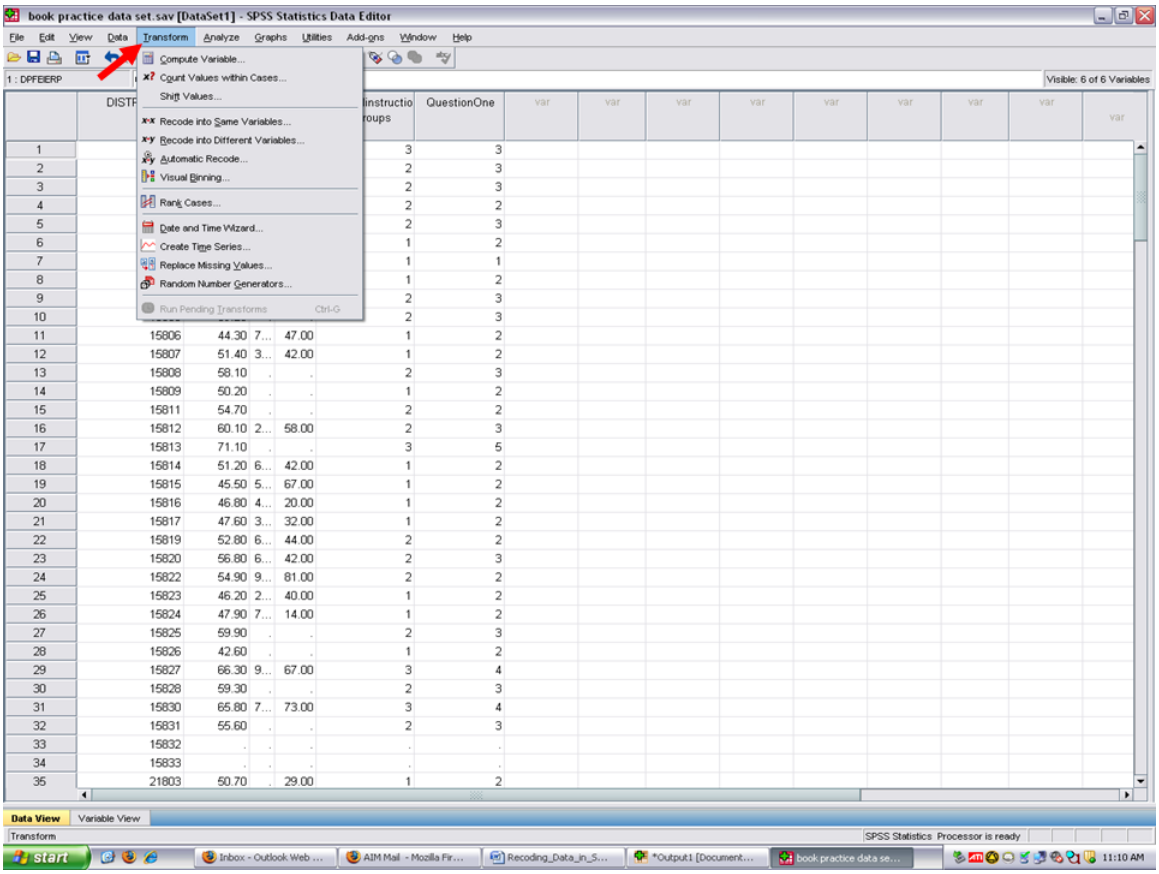


[12]

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.

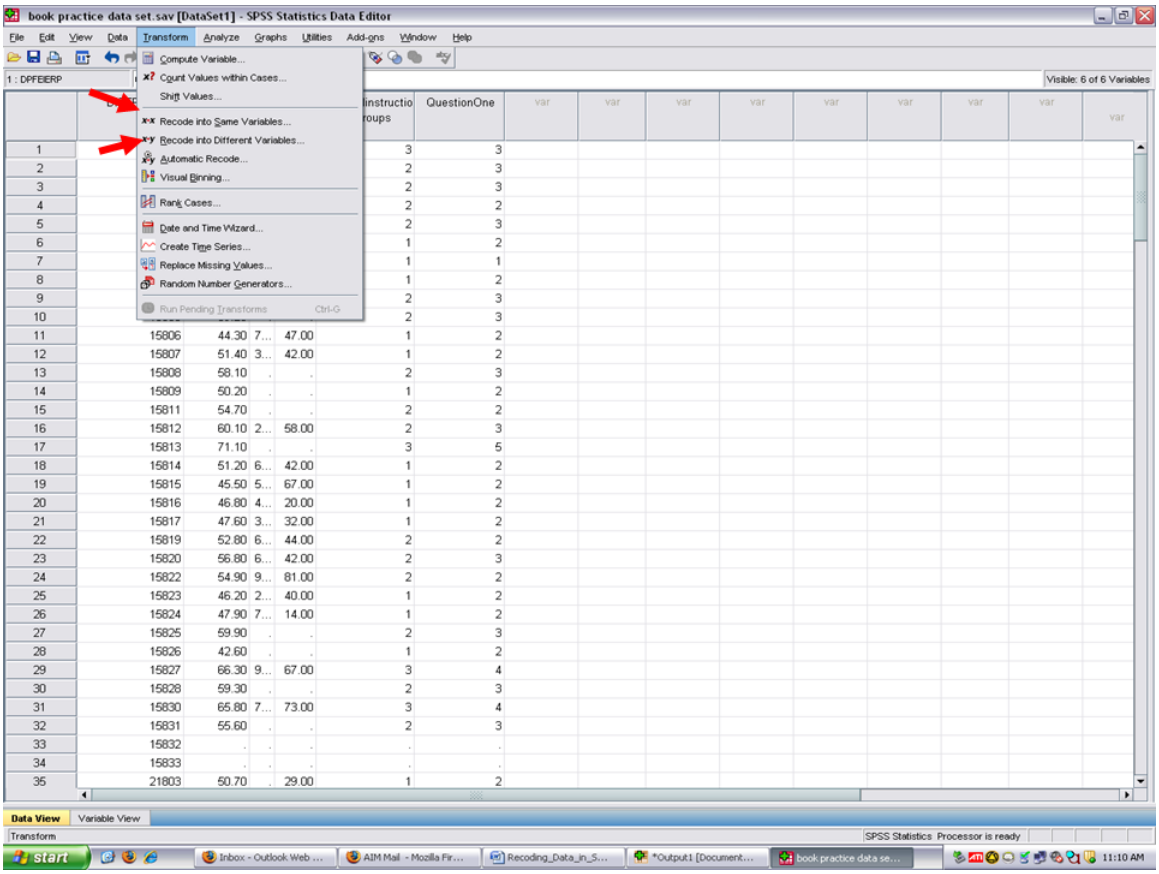
[12] http://cnx.org/content/m40708/latest/2.7.png/image



[13]

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.

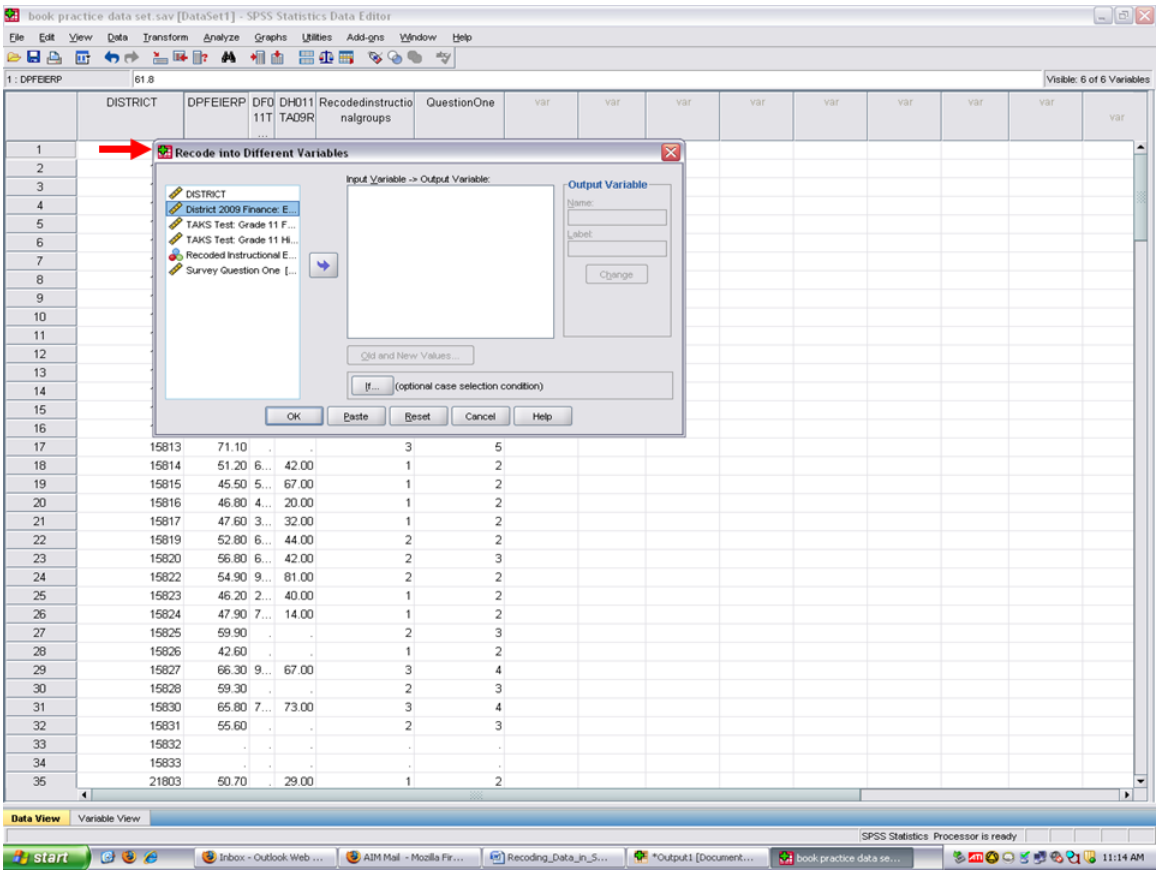
[13] http://cnx.org/content/m40708/latest/2.8.png/image



[14]

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

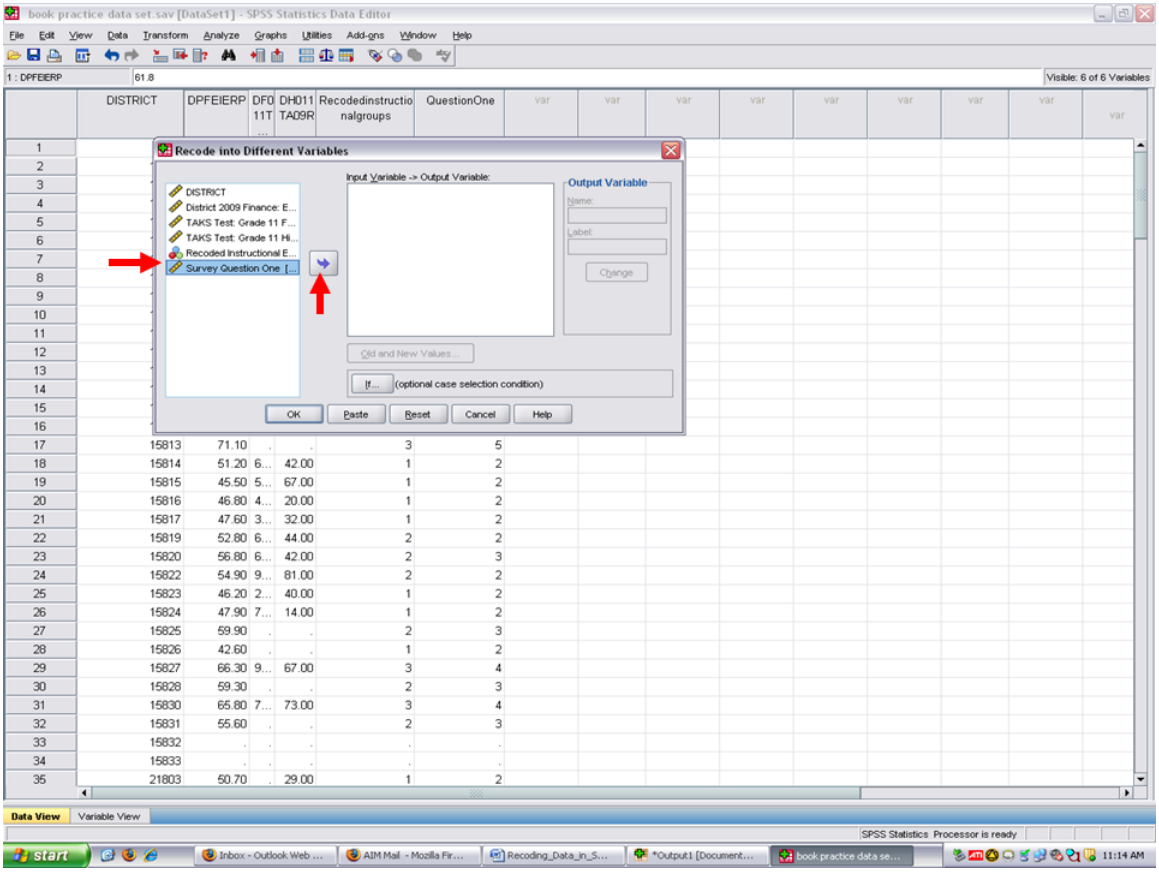
[14] http://cnx.org/content/m40708/latest/2.9.png/image



[15]

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.

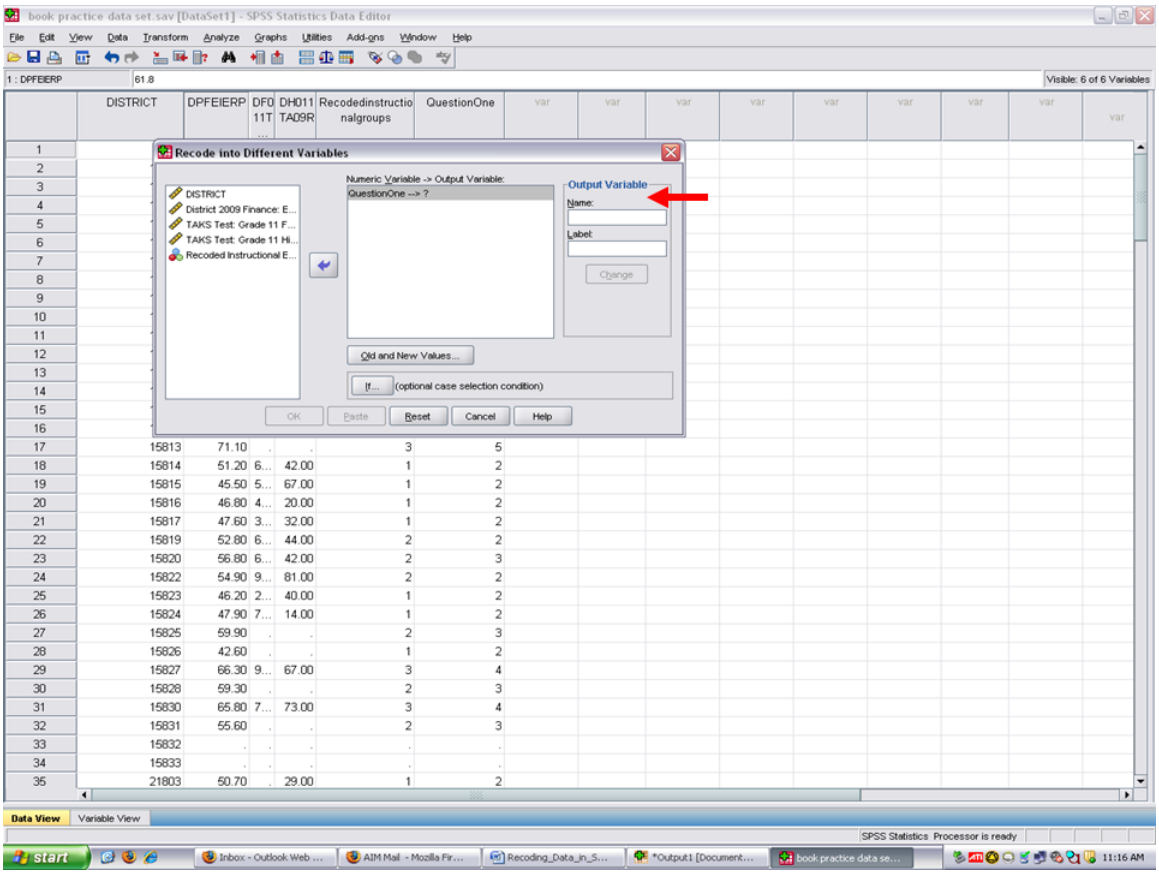
[15] http://cnx.org/content/m40708/latest/2.10.png/image



16

Click under the Output Variable where Name is present.

[16] http://cnx.org/content/m40708/latest/2.11.png/image

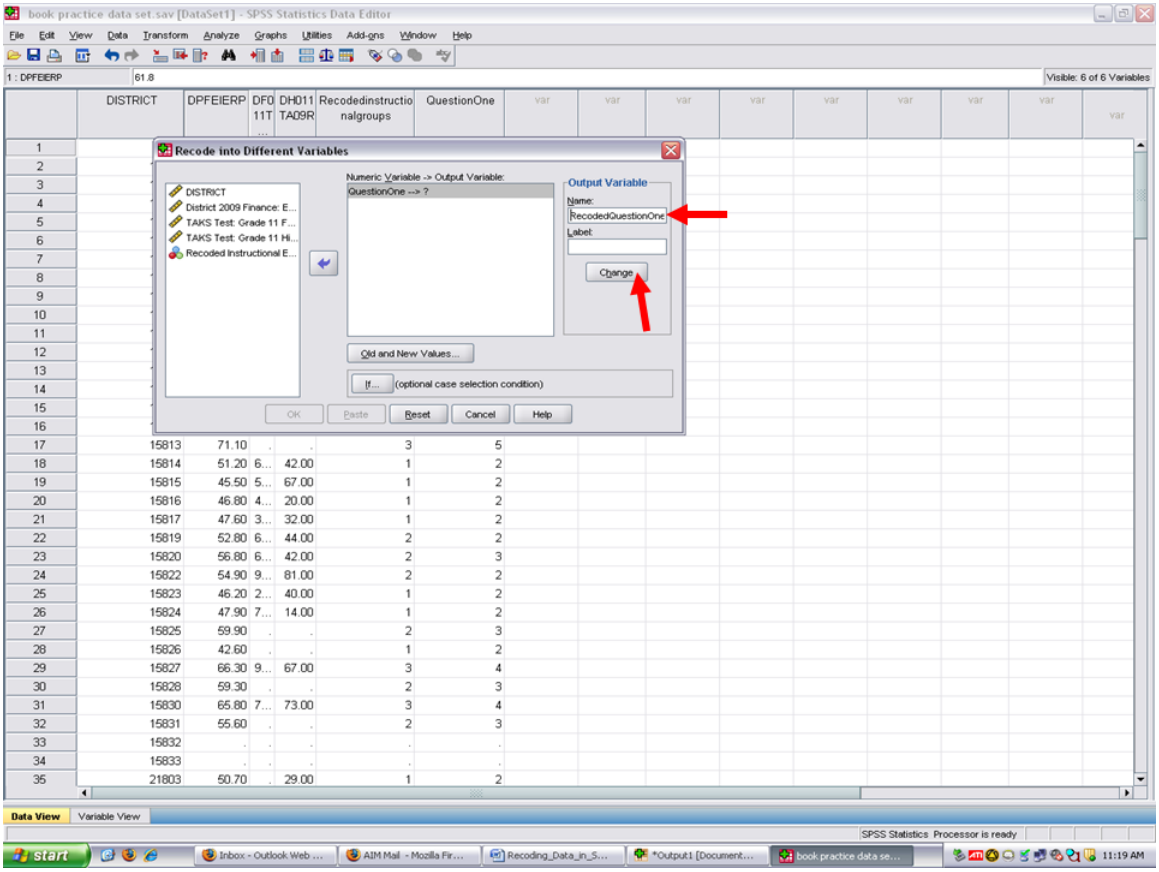


[17]

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.

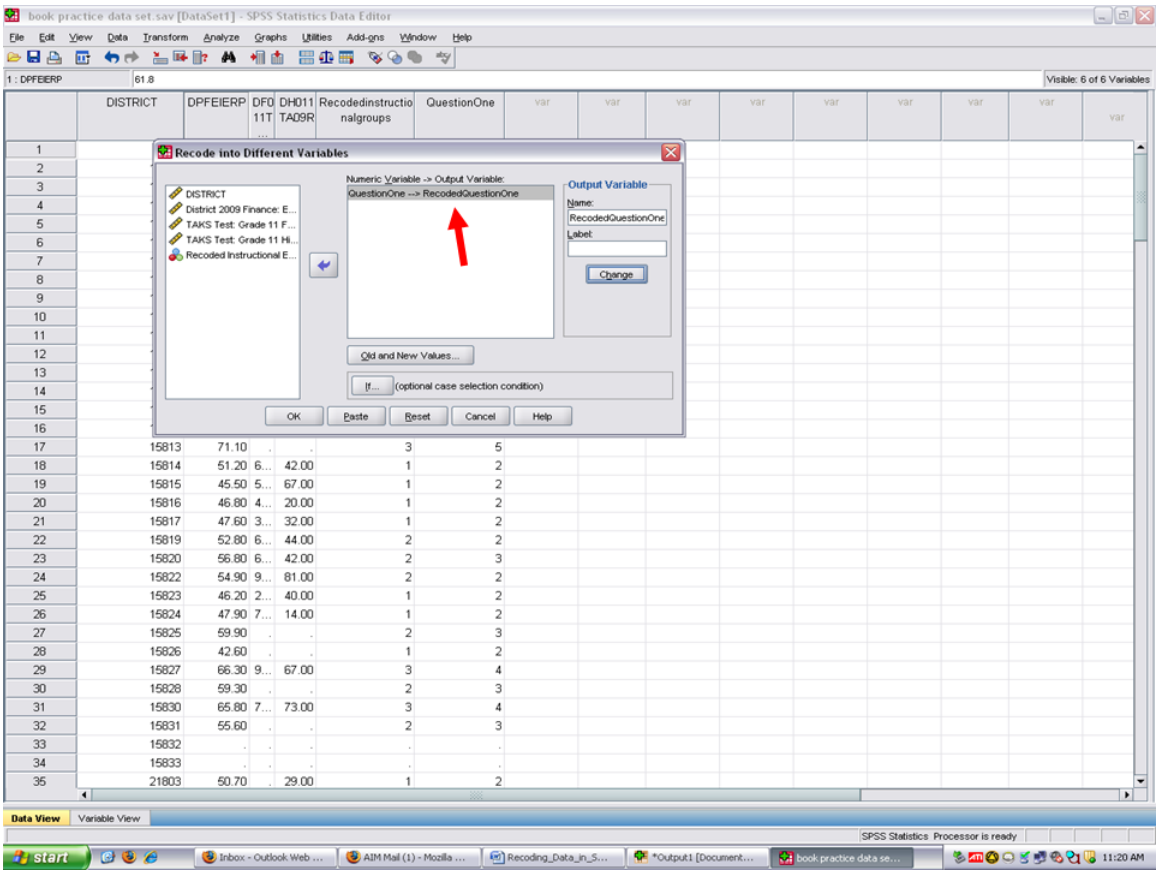
[17] http://cnx.org/content/m40708/latest/2.12.png/image



[18]

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

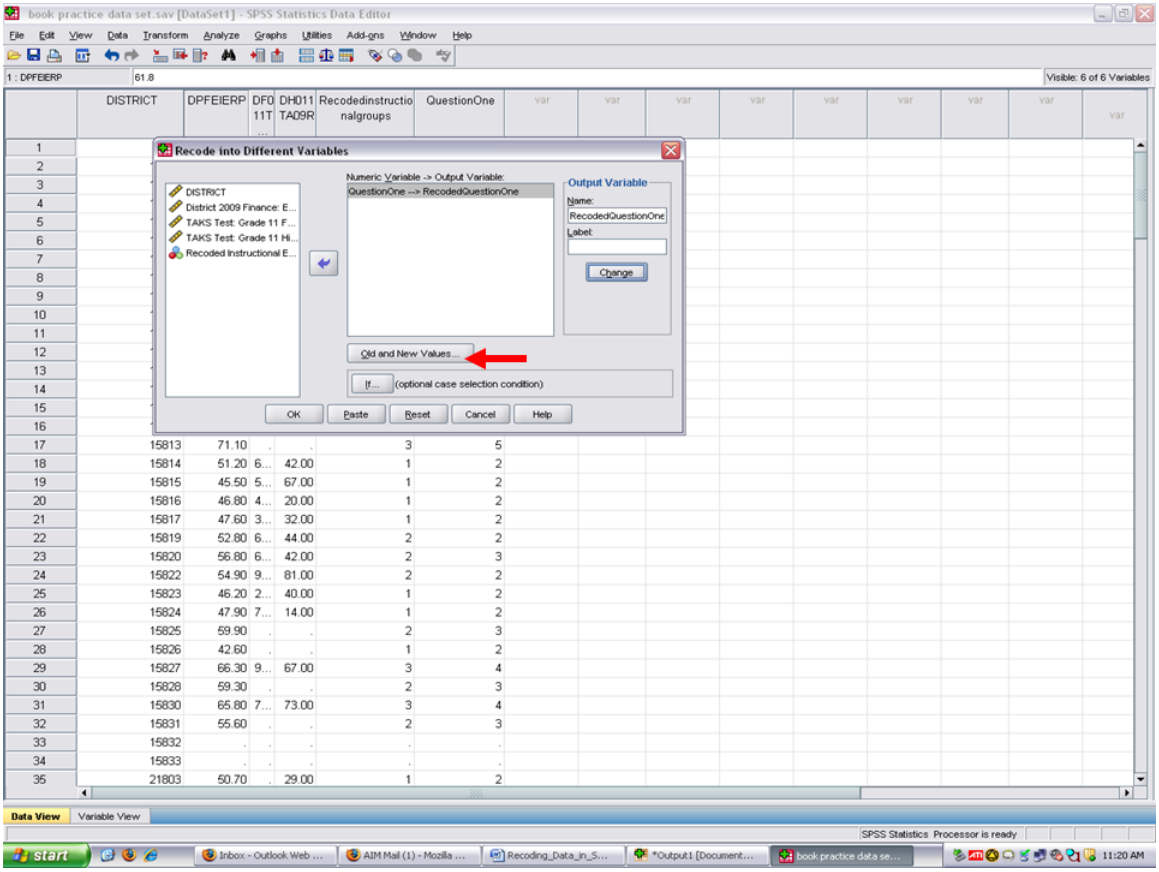
[18] http://cnx.org/content/m40708/latest/2.13.png/image



[19]

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

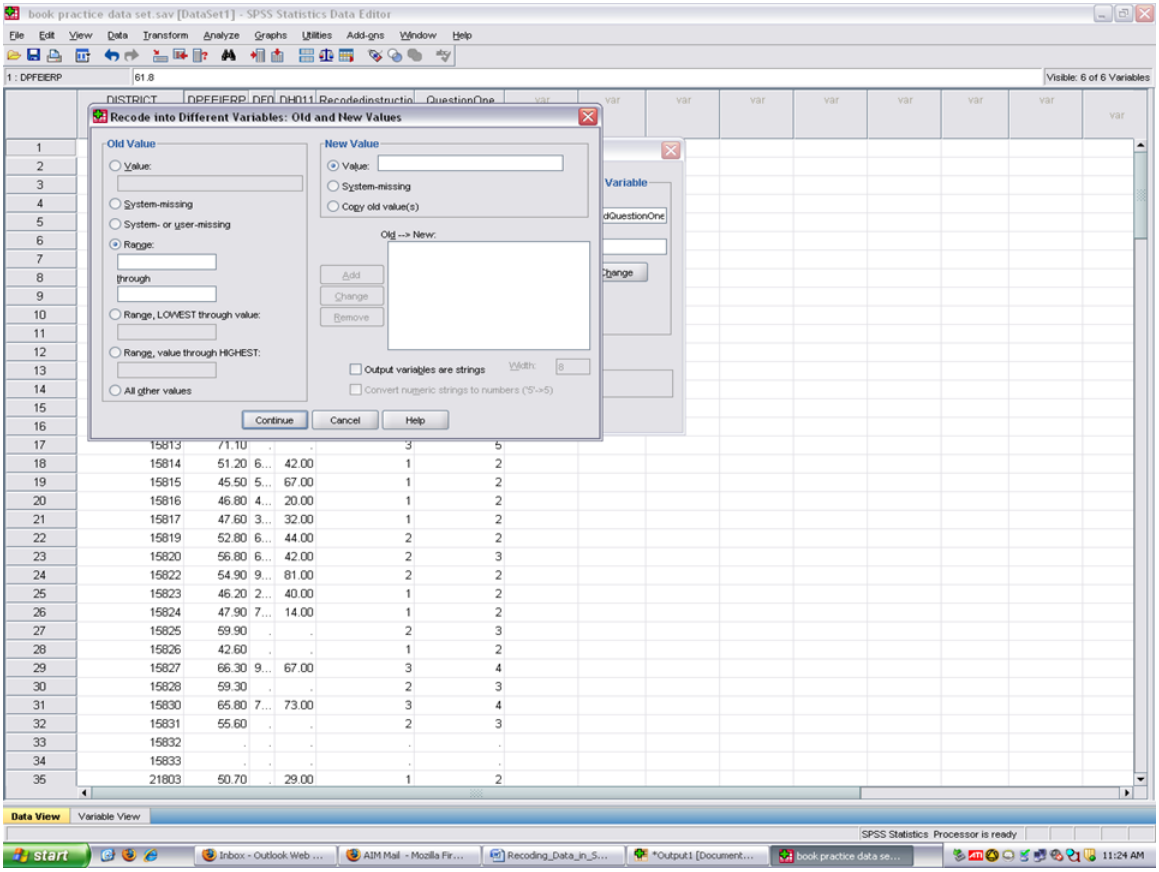
[19] http://cnx.org/content/m40708/latest/2.14.png/image



20

This screen will then appear.

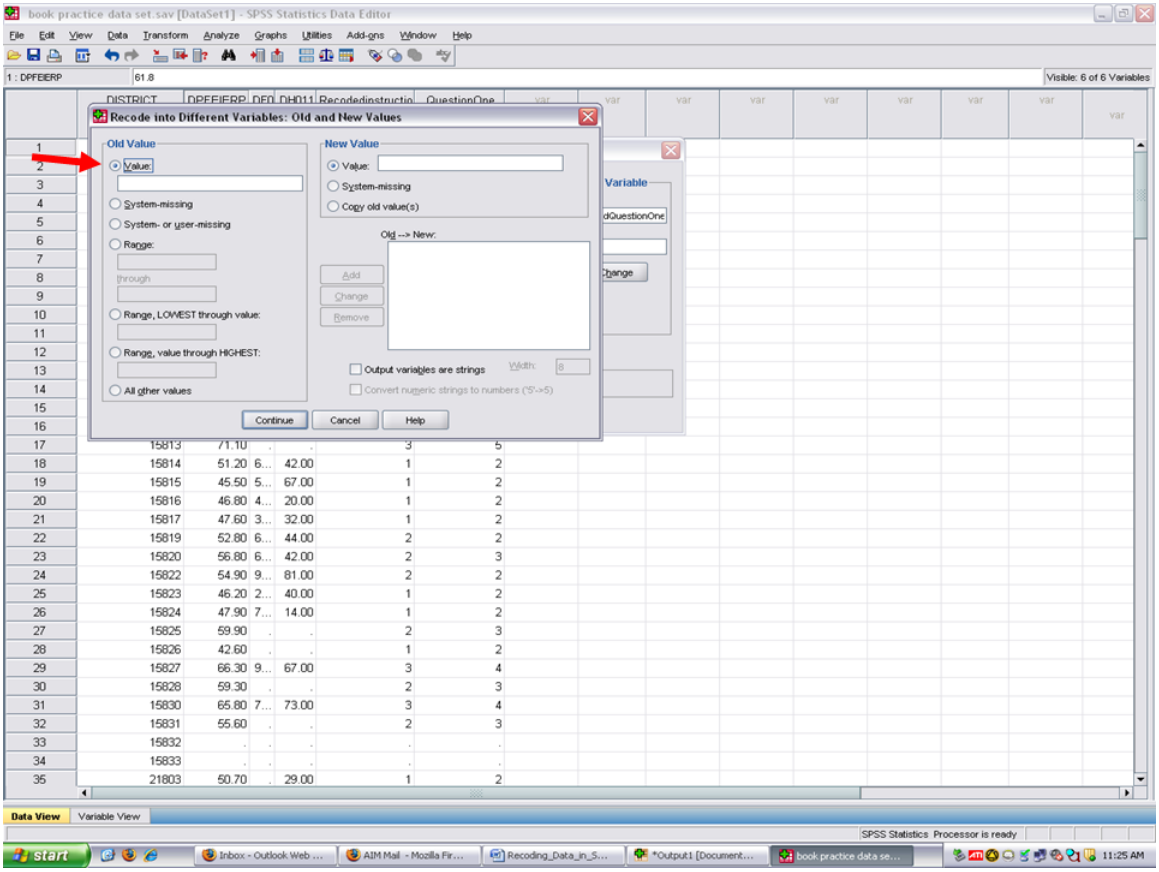
[20] http://cnx.org/content/m40708/latest/2.15.png/image



[21]

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

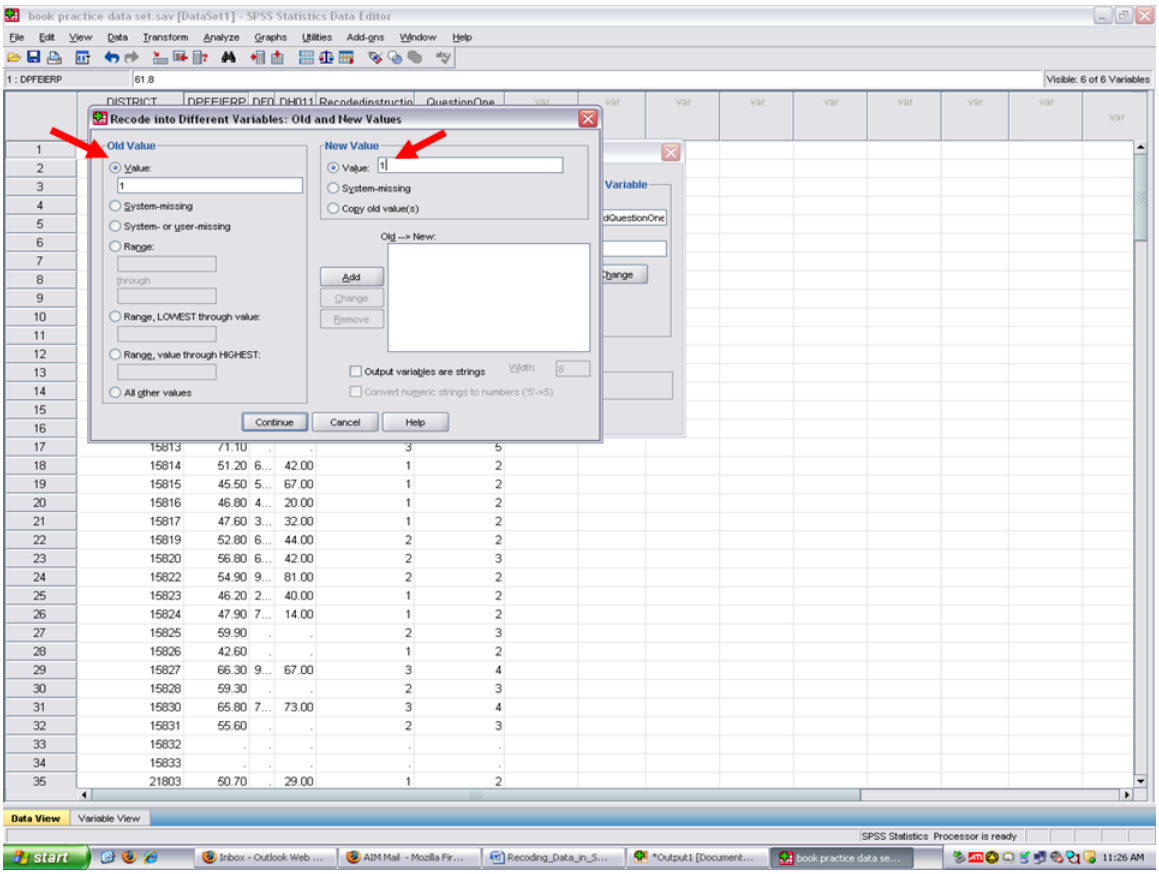
[21] http://cnx.org/content/m40708/latest/2.16.png/image



[22]

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.

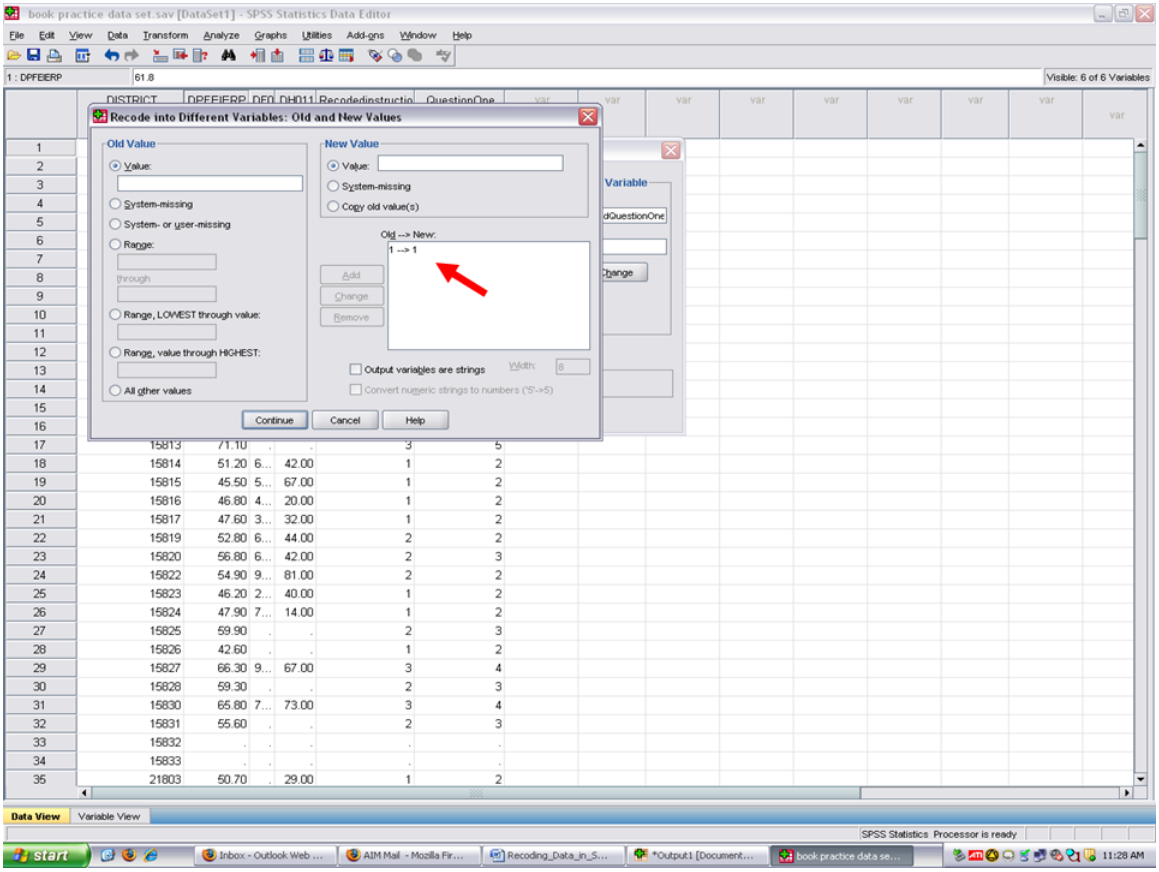
[22] http://cnx.org/content/m40708/latest/2.17.png/image



[23]

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

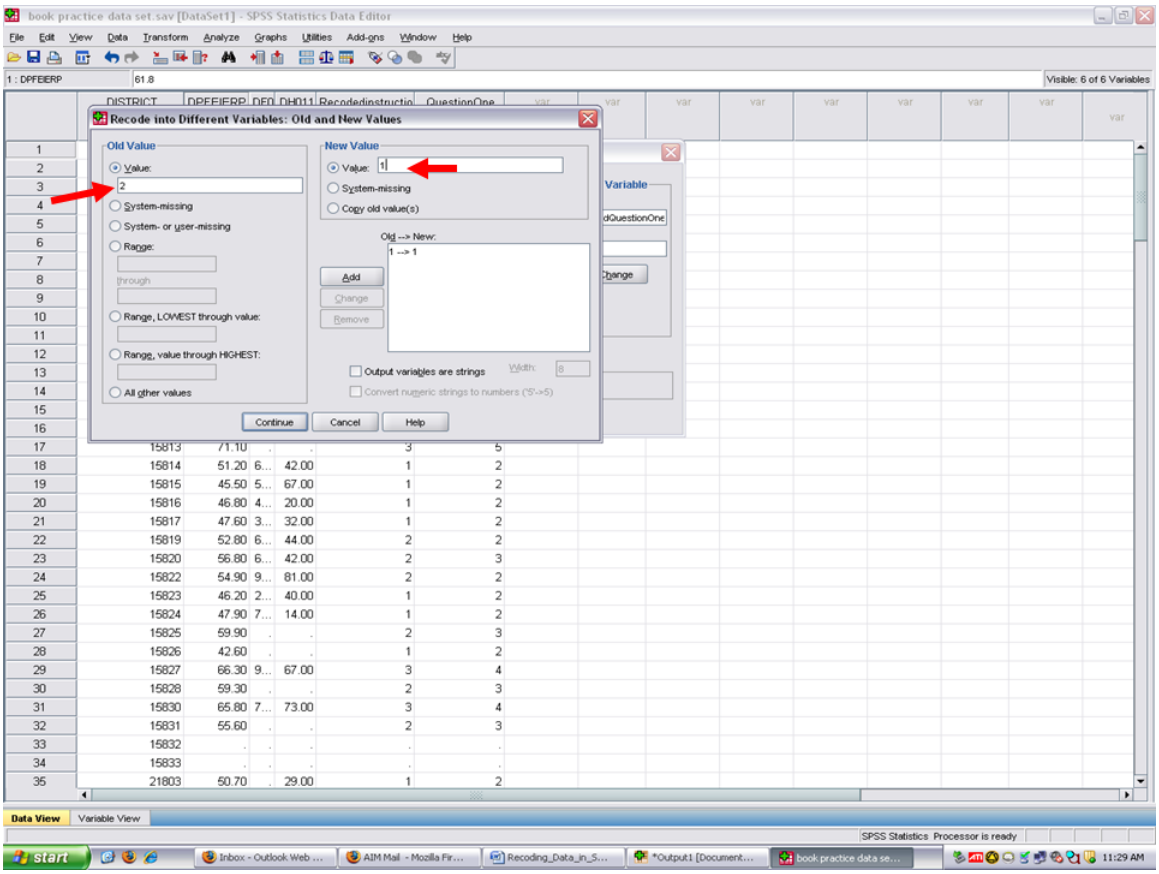
[23] http://cnx.org/content/m40708/latest/2.18.png/image



[24]

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.

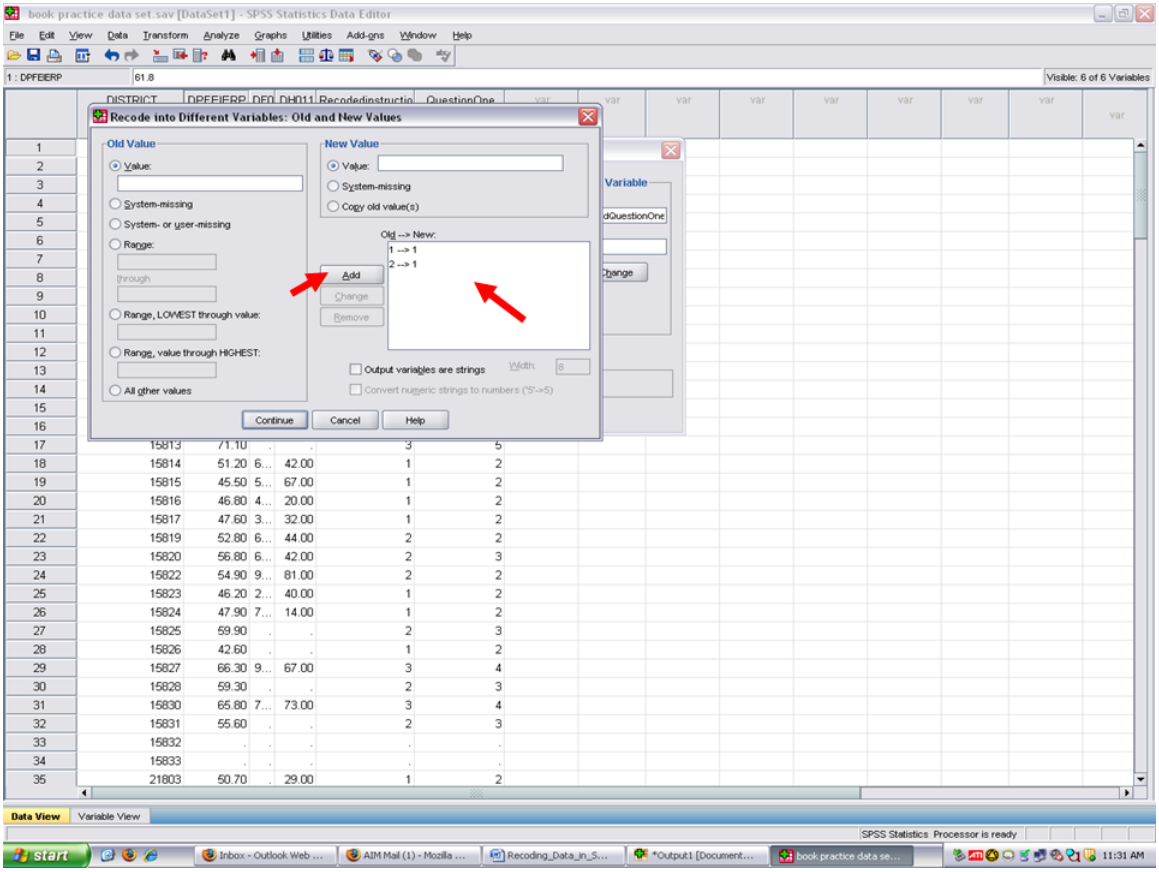
[24] http://cnx.org/content/m40708/latest/2.19.png/image



[25]

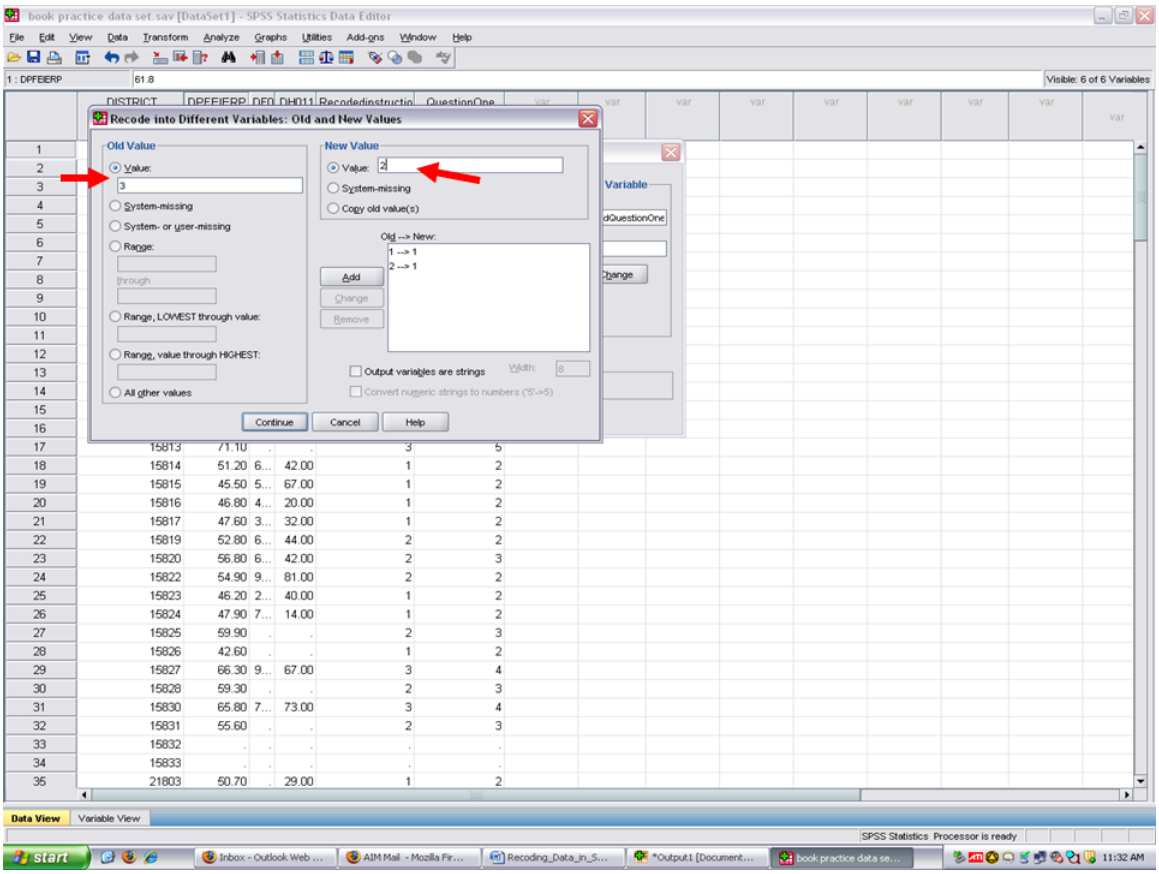
Click on Add. The Old Value of 2 and the New Value of 1 now appear in the larger box.

[25] http://cnx.org/content/m40708/latest/2.20.png/image



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

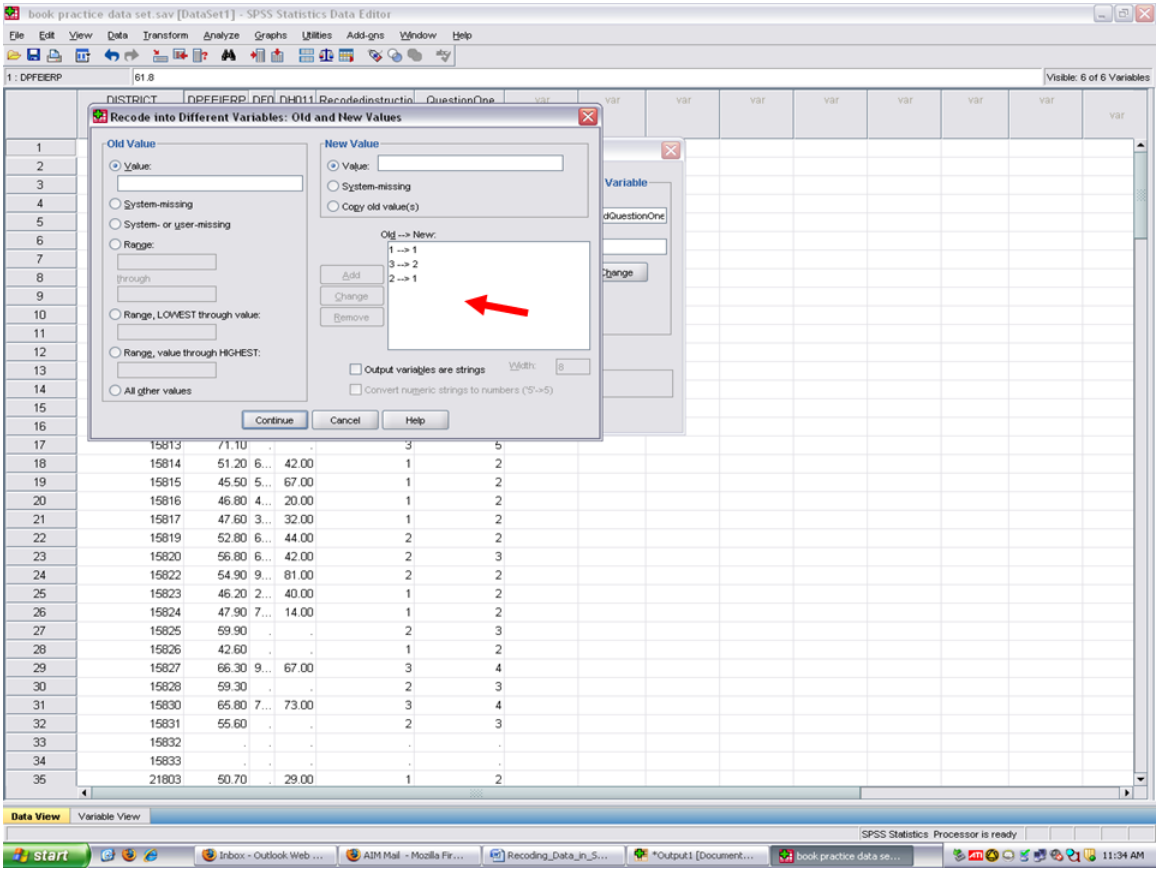
[26] http://cnx.org/content/m40708/latest/2.21.png/image



[27]

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

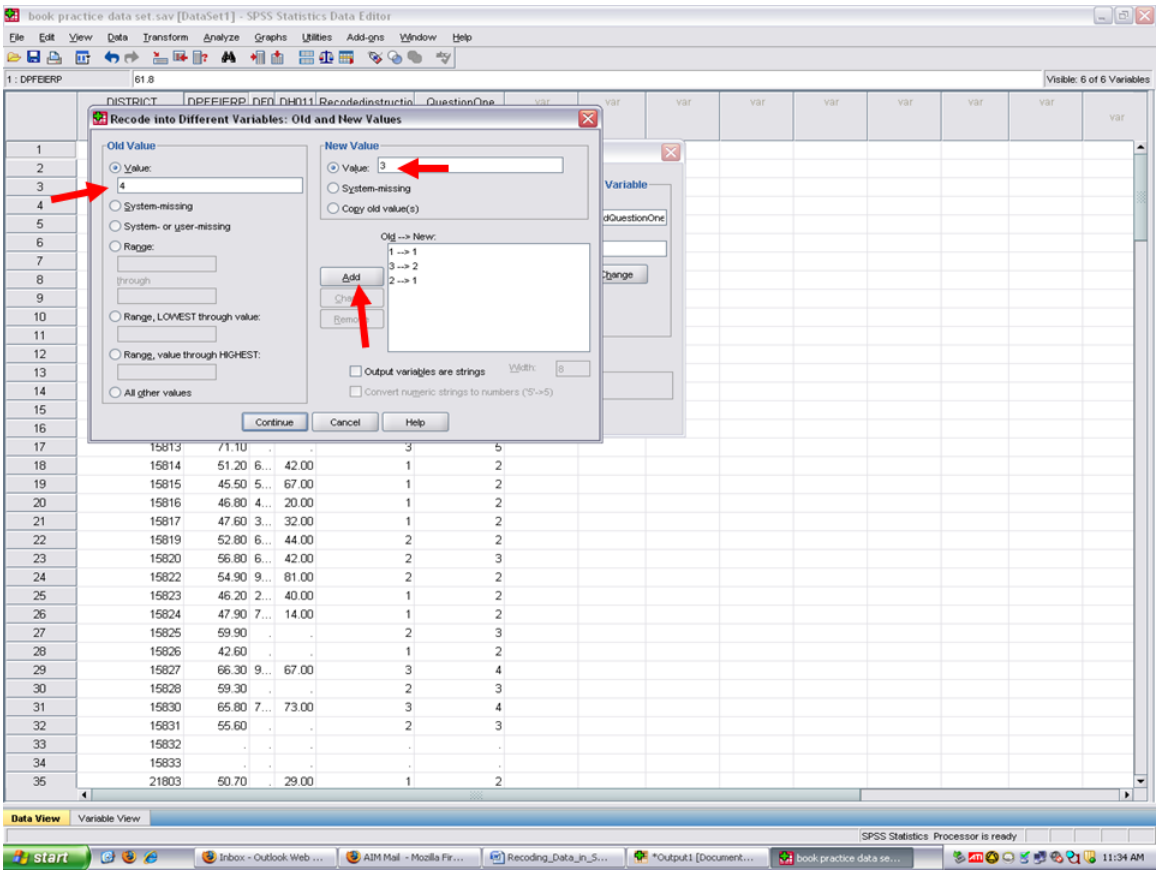
[27] http://cnx.org/content/m40708/latest/2.22.png/image



[28]

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

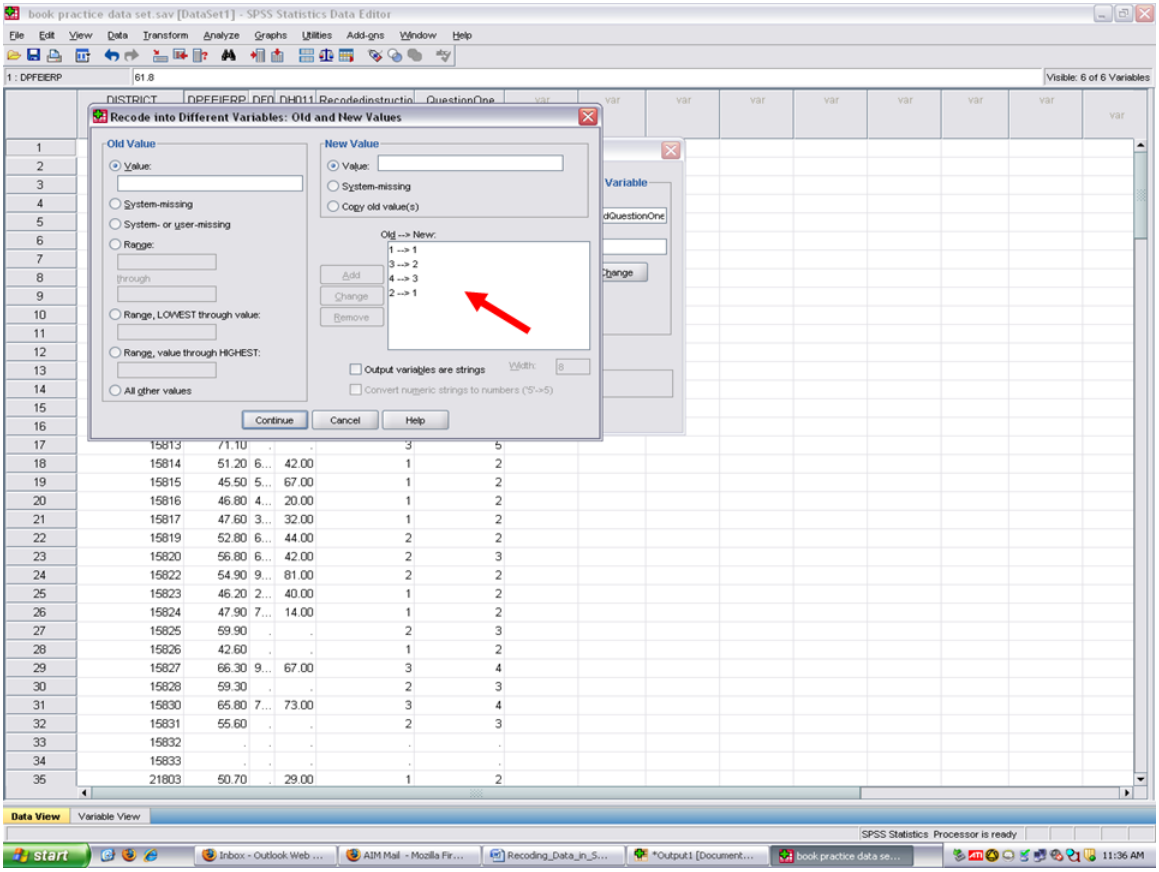
[28] http://cnx.org/content/m40708/latest/2.23.png/image



29

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

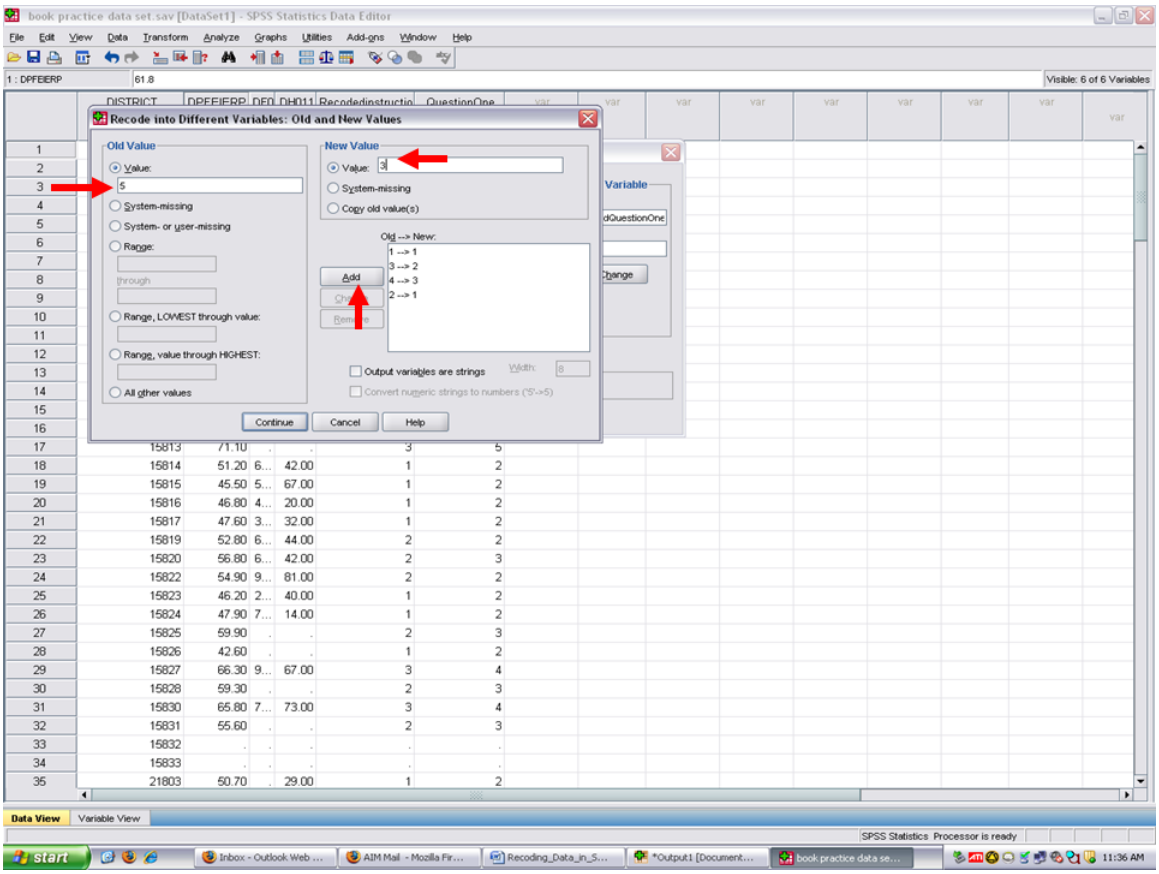
[29] http://cnx.org/content/m40708/latest/2.24.png/image



[30]

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

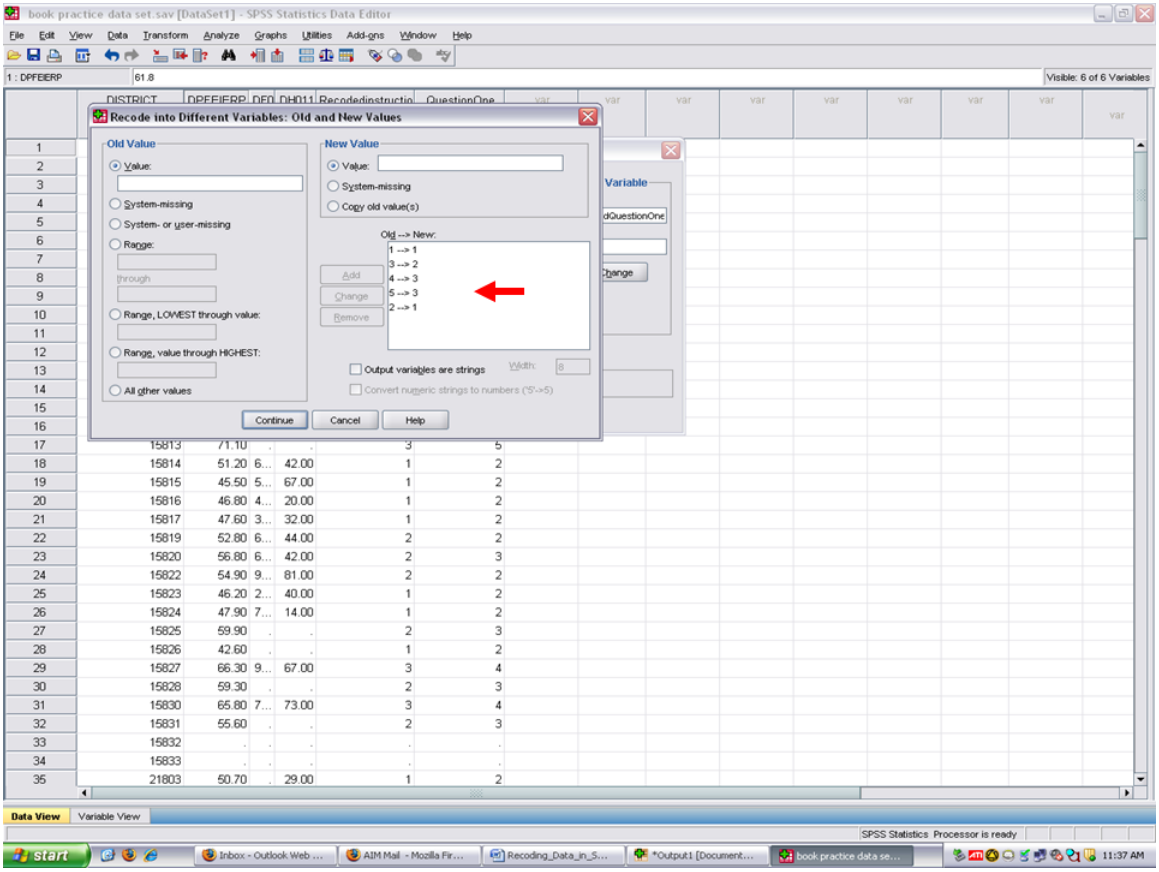
[30] http://cnx.org/content/m40708/latest/2.25.png/image



[31]

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.

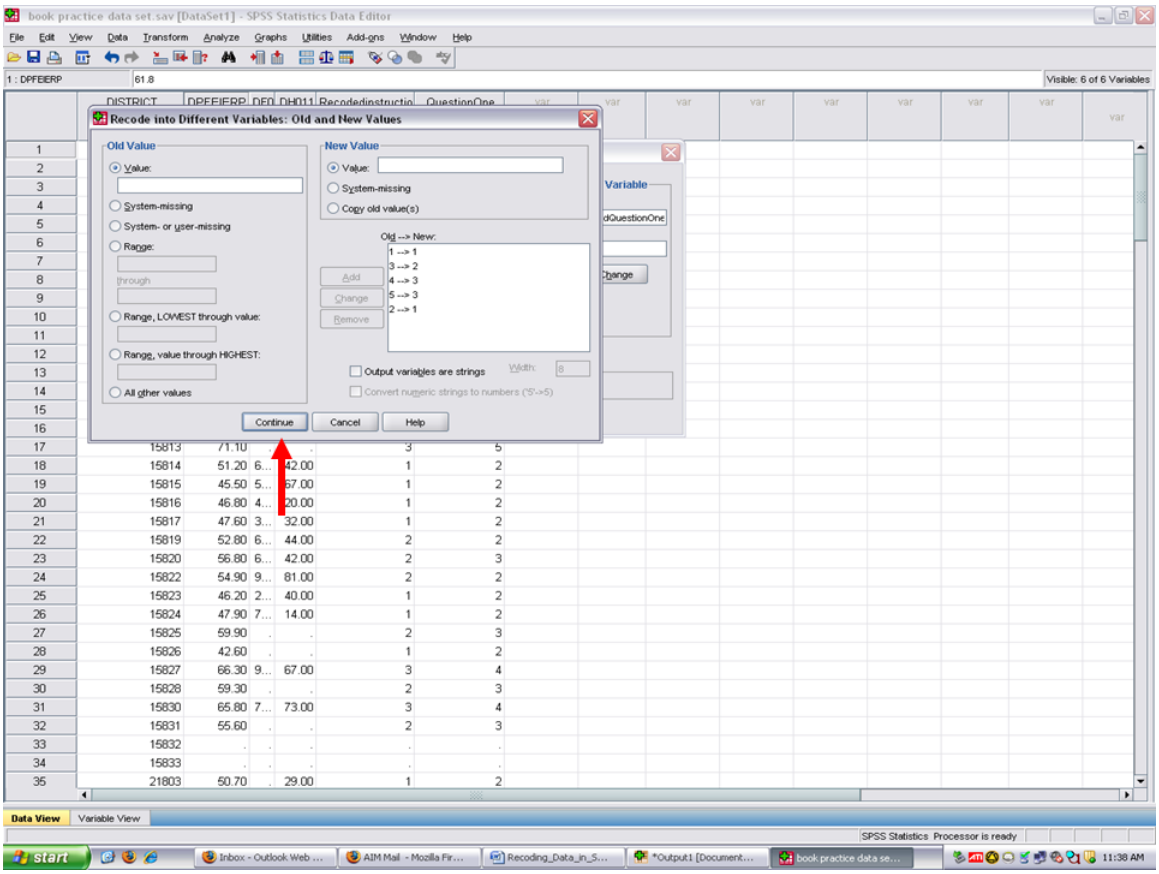
[31] http://cnx.org/content/m40708/latest/2.26.png/image



[32]

Click on Continue.

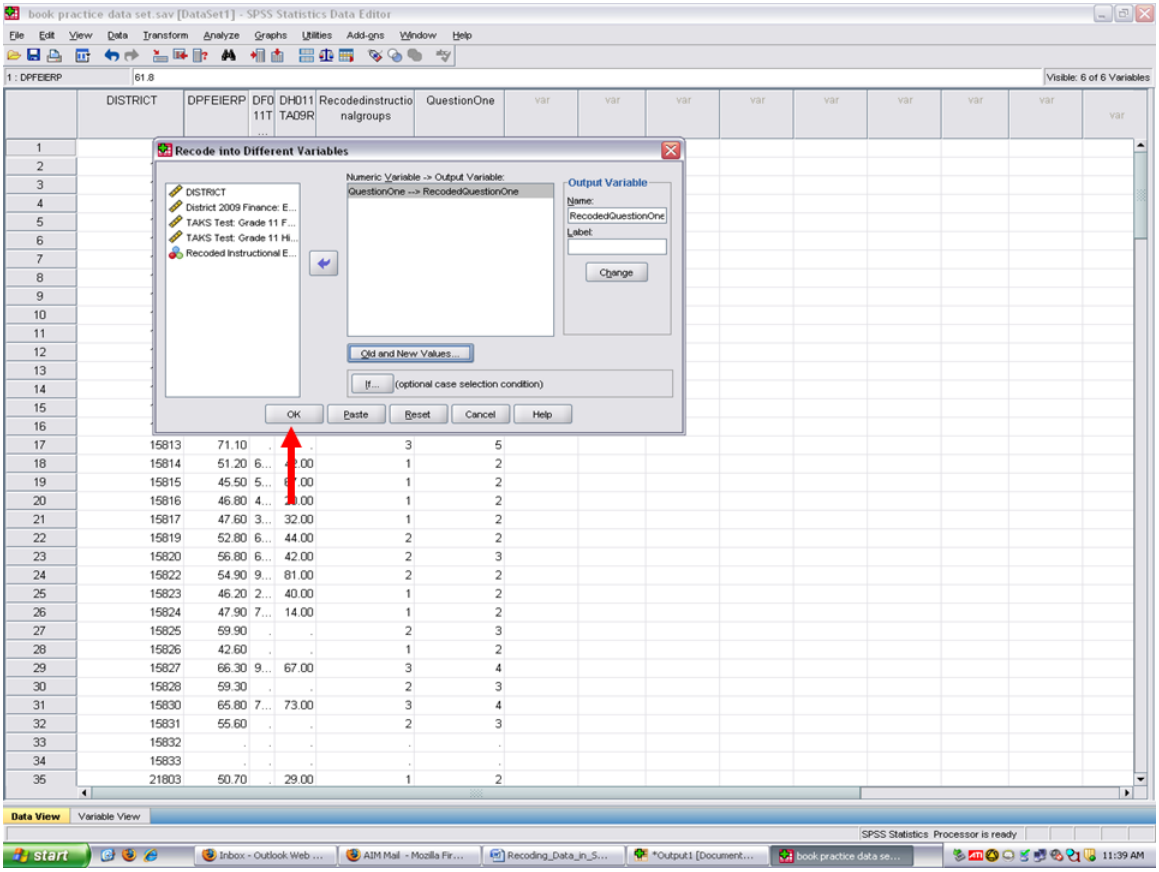
[32] http://cnx.org/content/m40708/latest/2.27.png/image



[33]

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

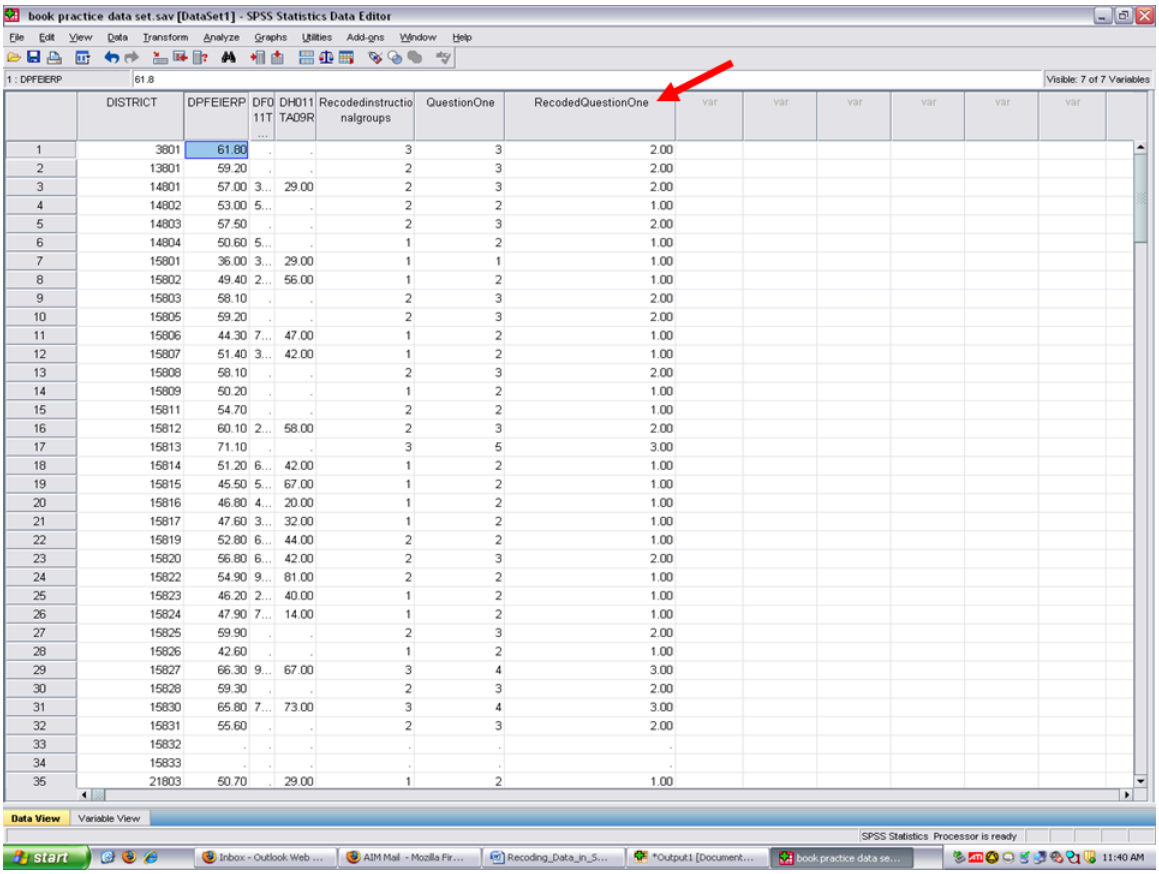
[33] http://cnx.org/content/m40708/latest/2.28.png/image



34

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



[35]

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 SPSS[1]

NOTE: This chapter is published by NCPEA Press[2] 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.

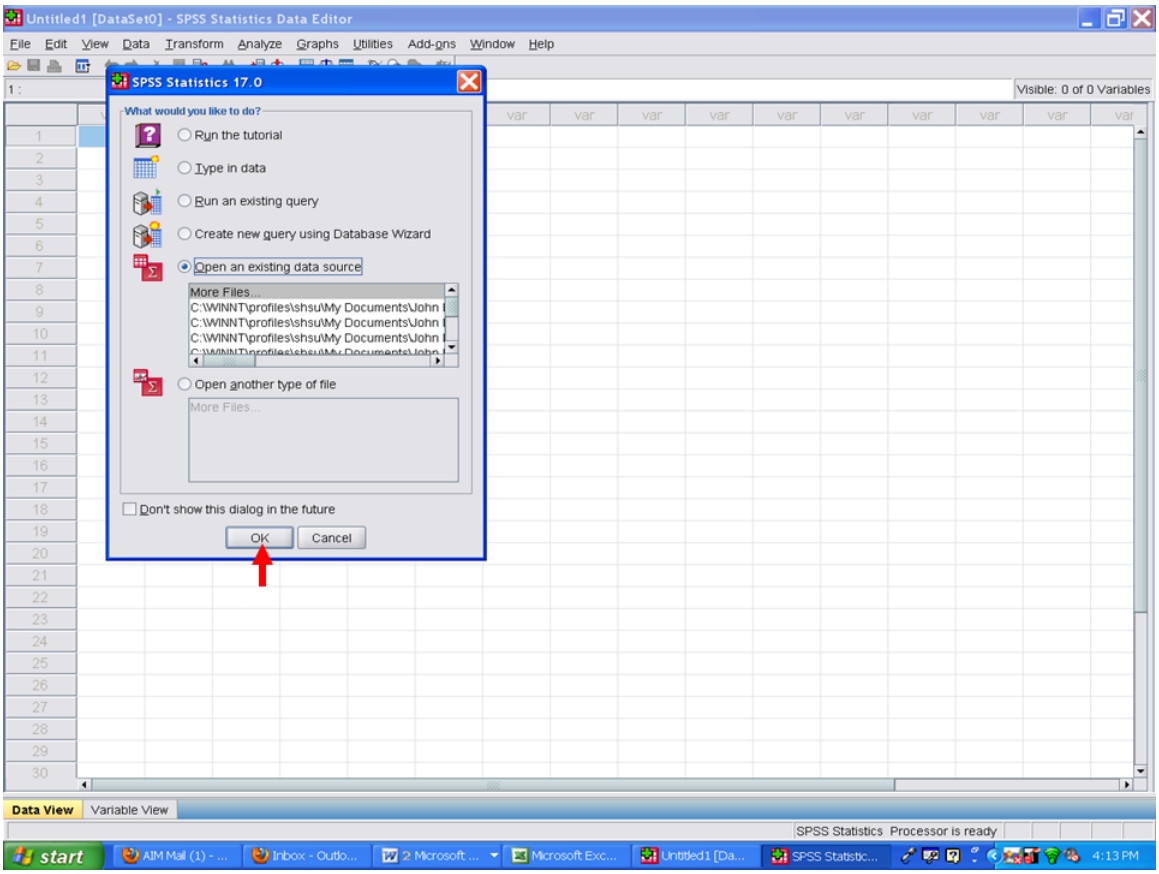
About the Authors

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 Help[3]) 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.

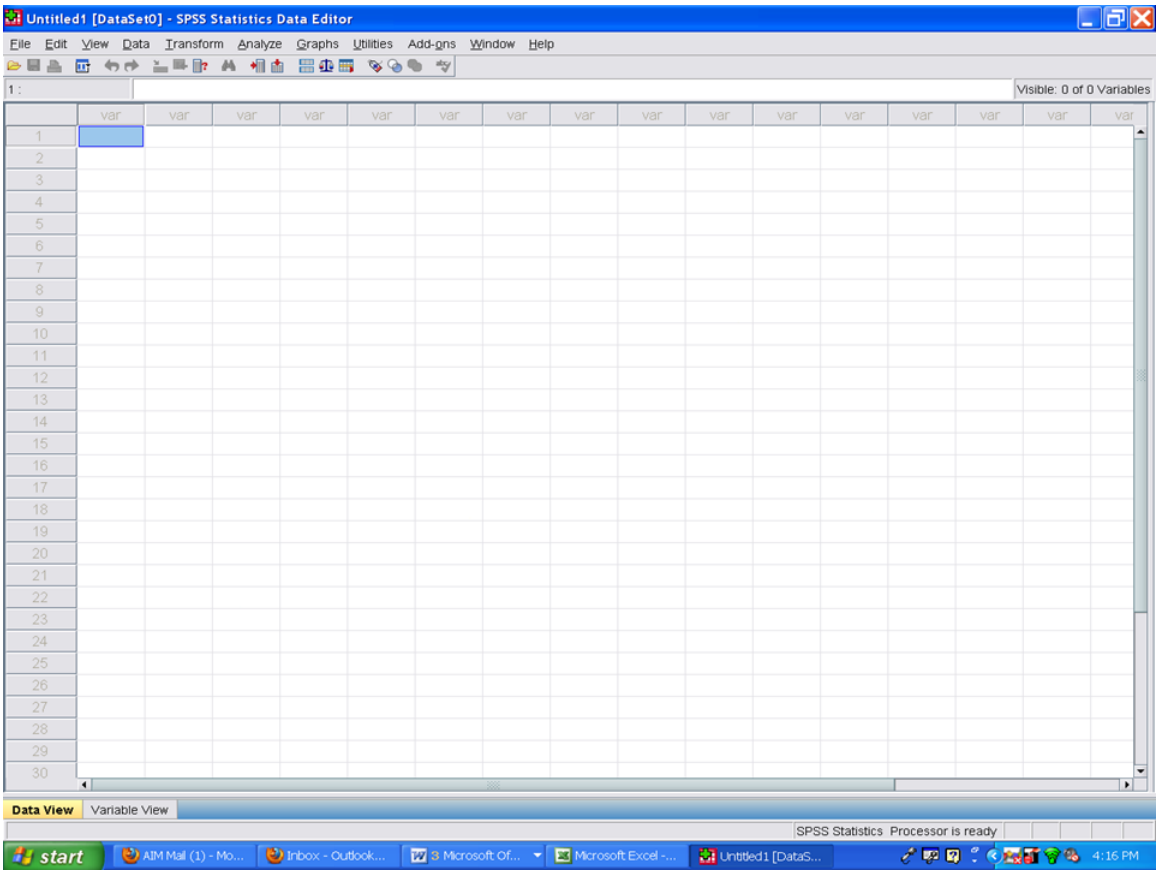
[1] This content is available online at <http://cnx.org/content/m40709/1.2/>.
[2] http://www.ncpeapublications.org/books.html
[3] http://cnx.org/content/m40709/latest/www.writingandstatisticalhelp



[4]

Your screen will now look like this.

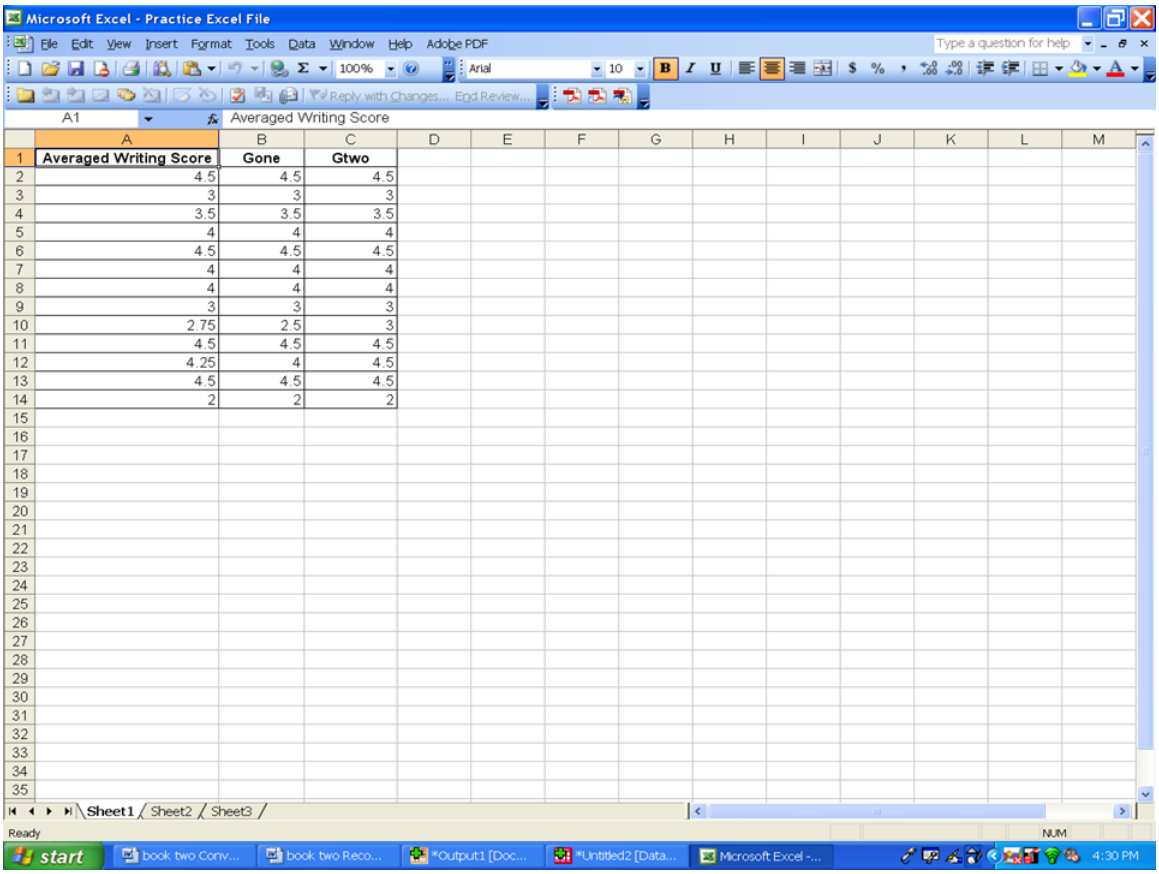
[4] http://cnx.org/content/m40709/latest/3.1.png/image



[5]

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
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

[6]

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.

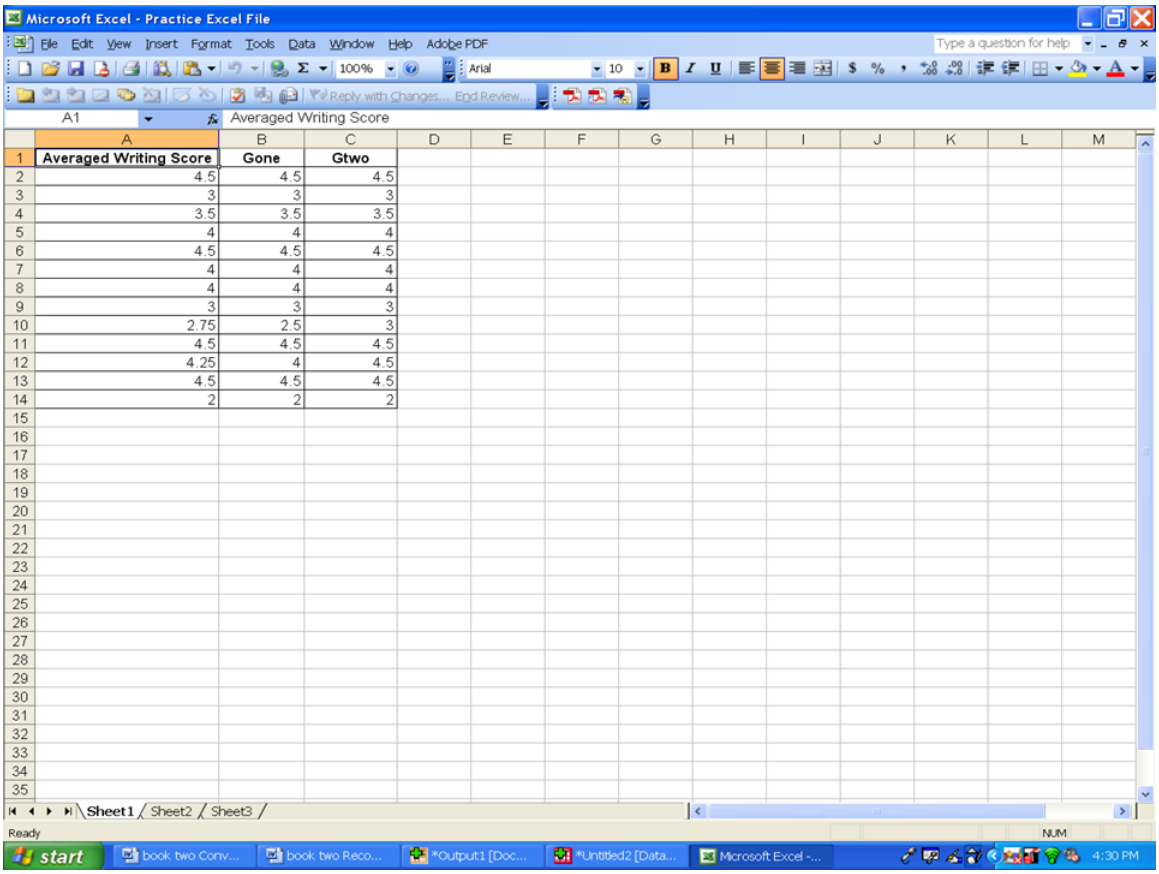
Once you are back in the SPSS data view window,

√ Click on File

√ Click on Open

√ Click on Data

[6] http://cnx.org/content/m40709/latest/3.3.png/image

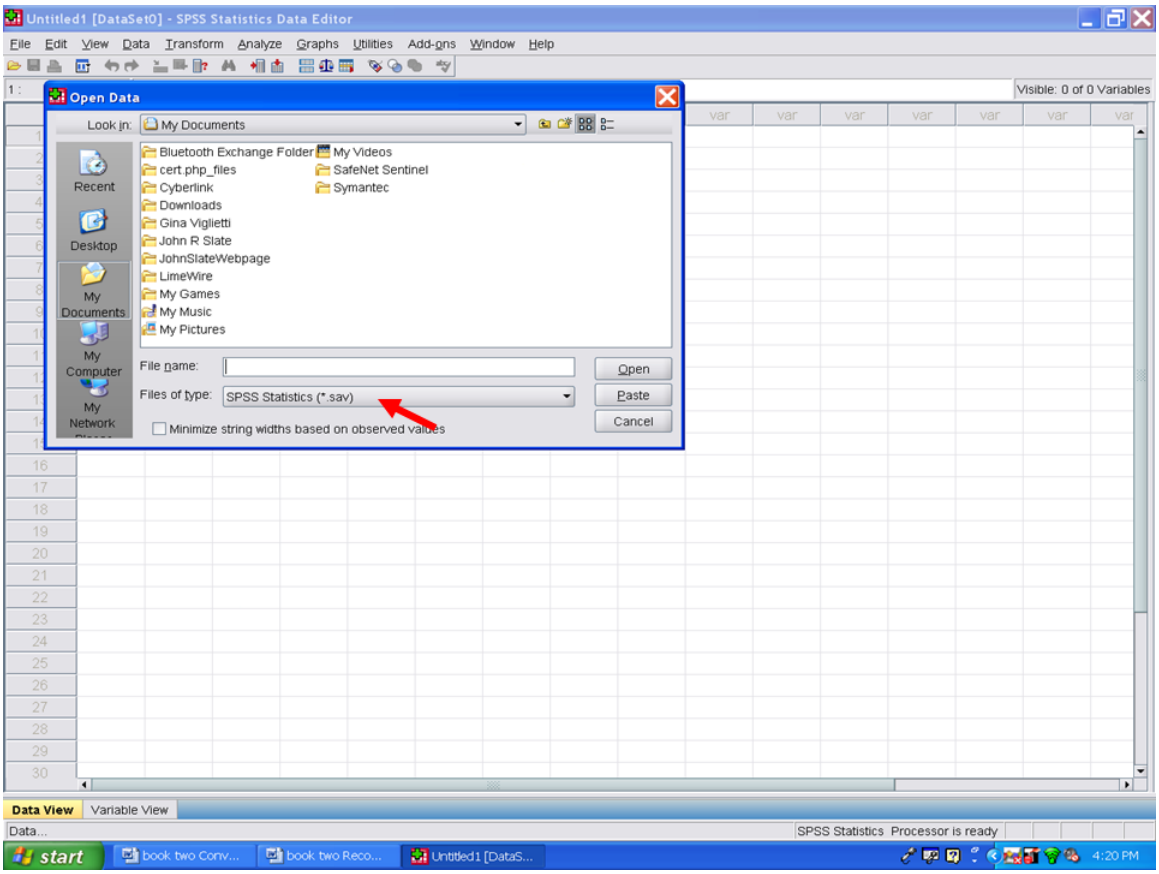


	A	B	C
1	Averaged Writing Score	Gone	Gtwo
2	4.5	4.5	4.5
3	3	3	3
4	3.5	3.5	3.5
5	4	4	4
6	4.5	4.5	4.5
7	4	4	4
8	4	4	4
9	3	3	3
10	2.75	2.5	3
11	4.5	4.5	4.5
12	4.25	4	4.5
13	4.5	4.5	4.5
14	2	2	2

[7]

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.

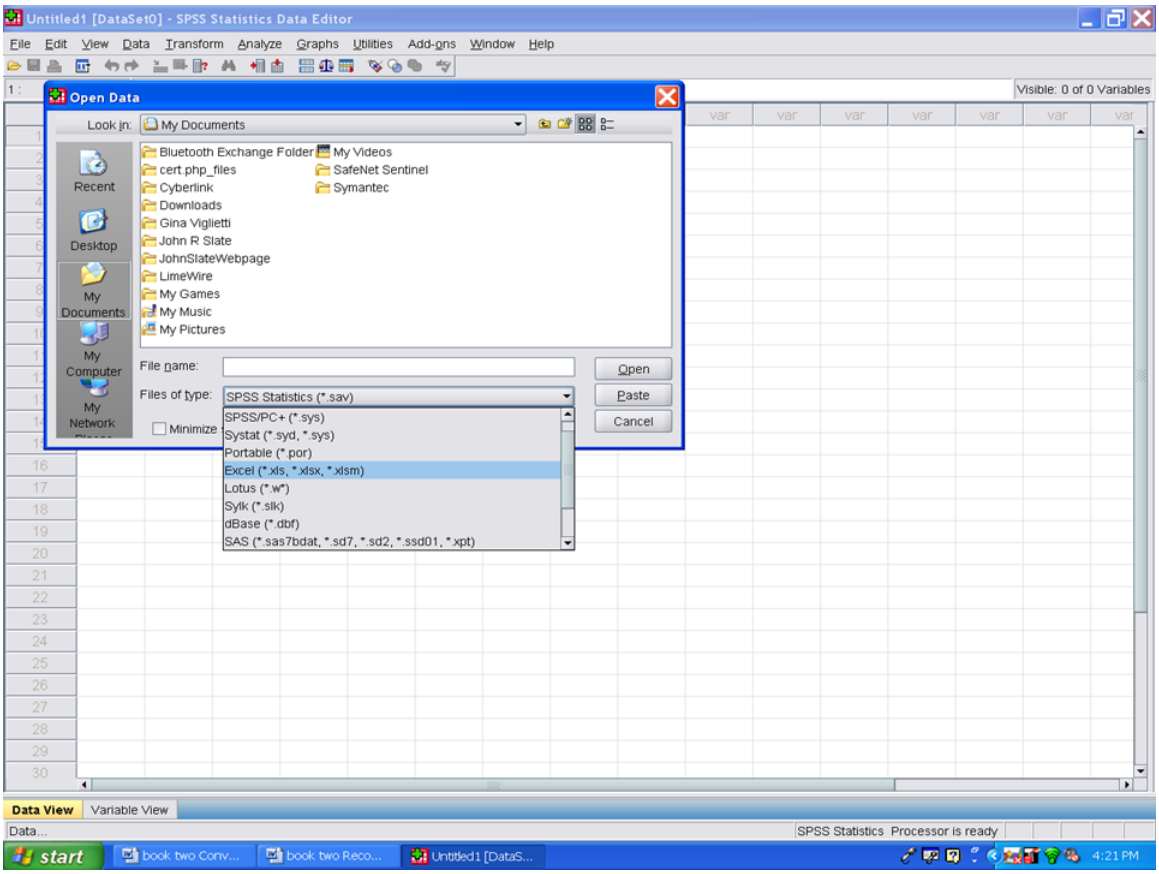
[7] http://cnx.org/content/m40709/latest/3.41.png/image



[8]

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

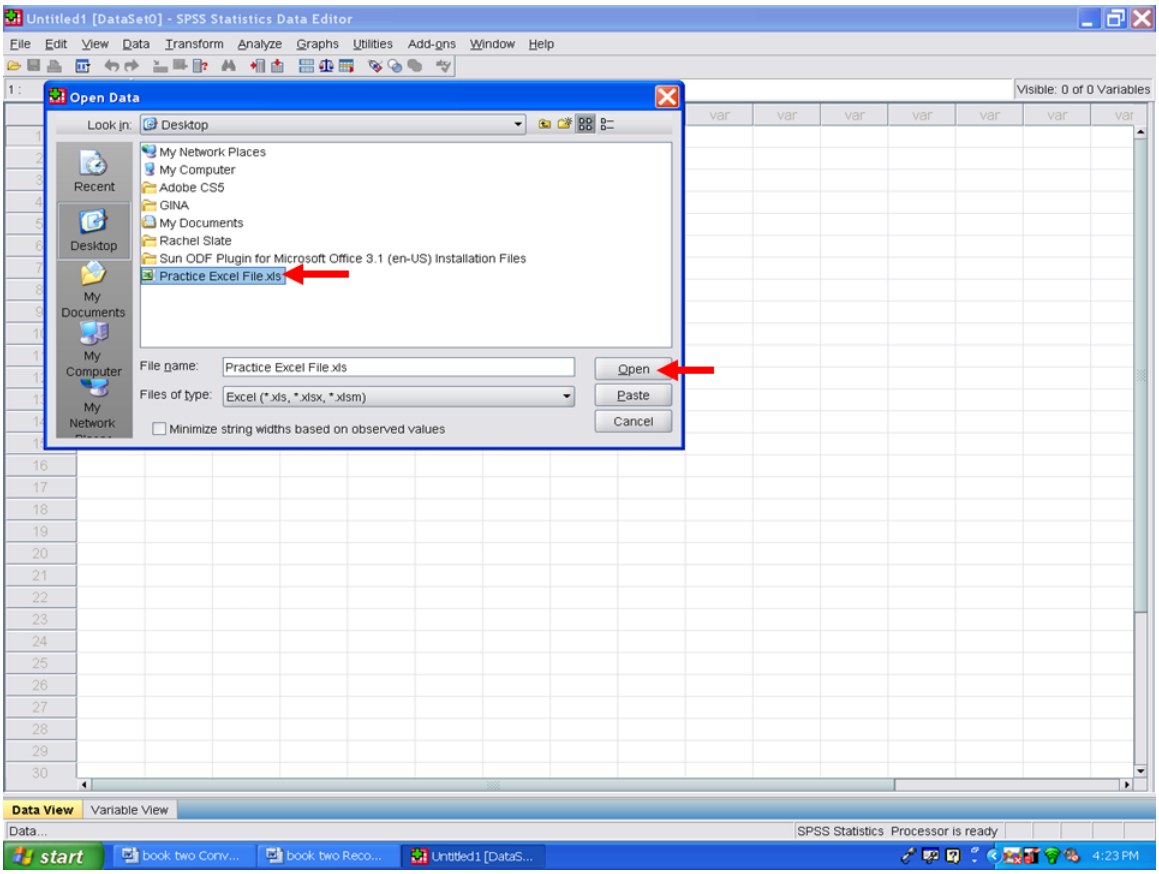
[8] http://cnx.org/content/m40709/latest/3.51.png/image



[9]

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

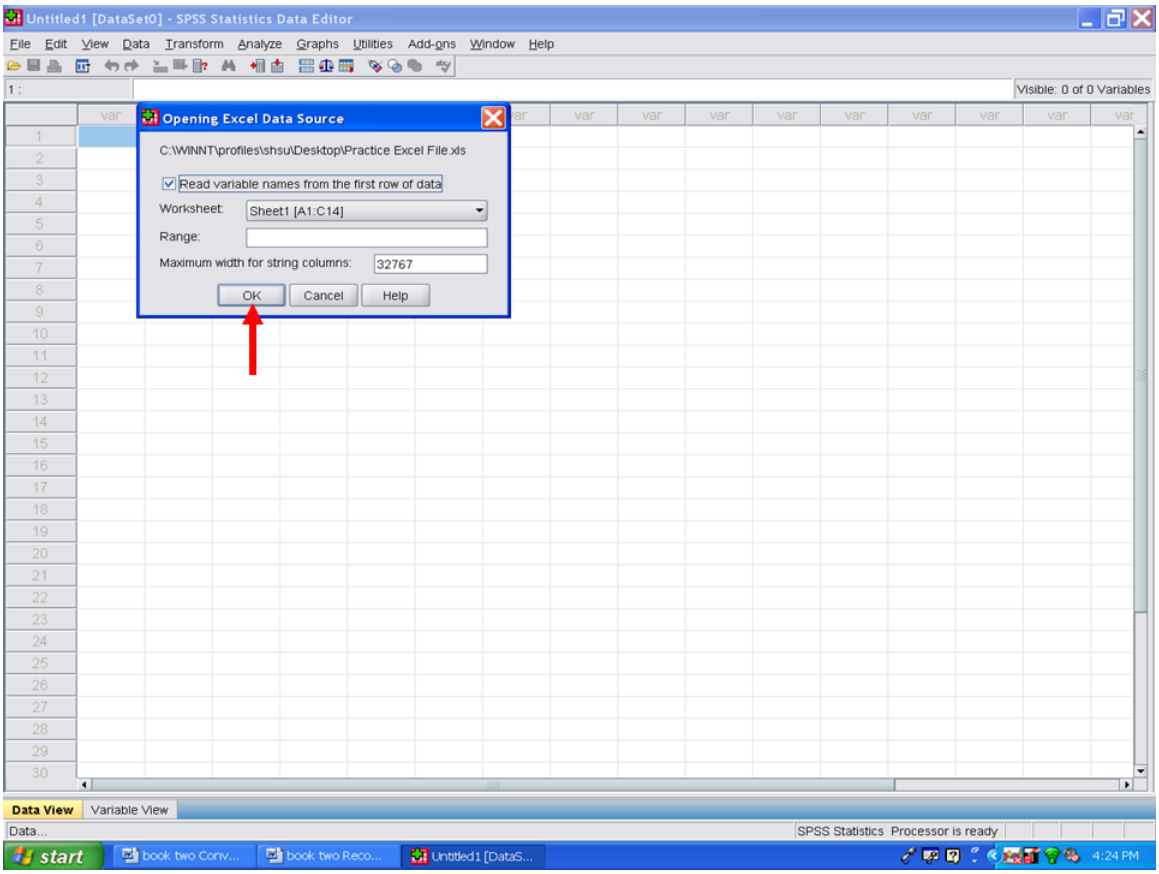


10

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.

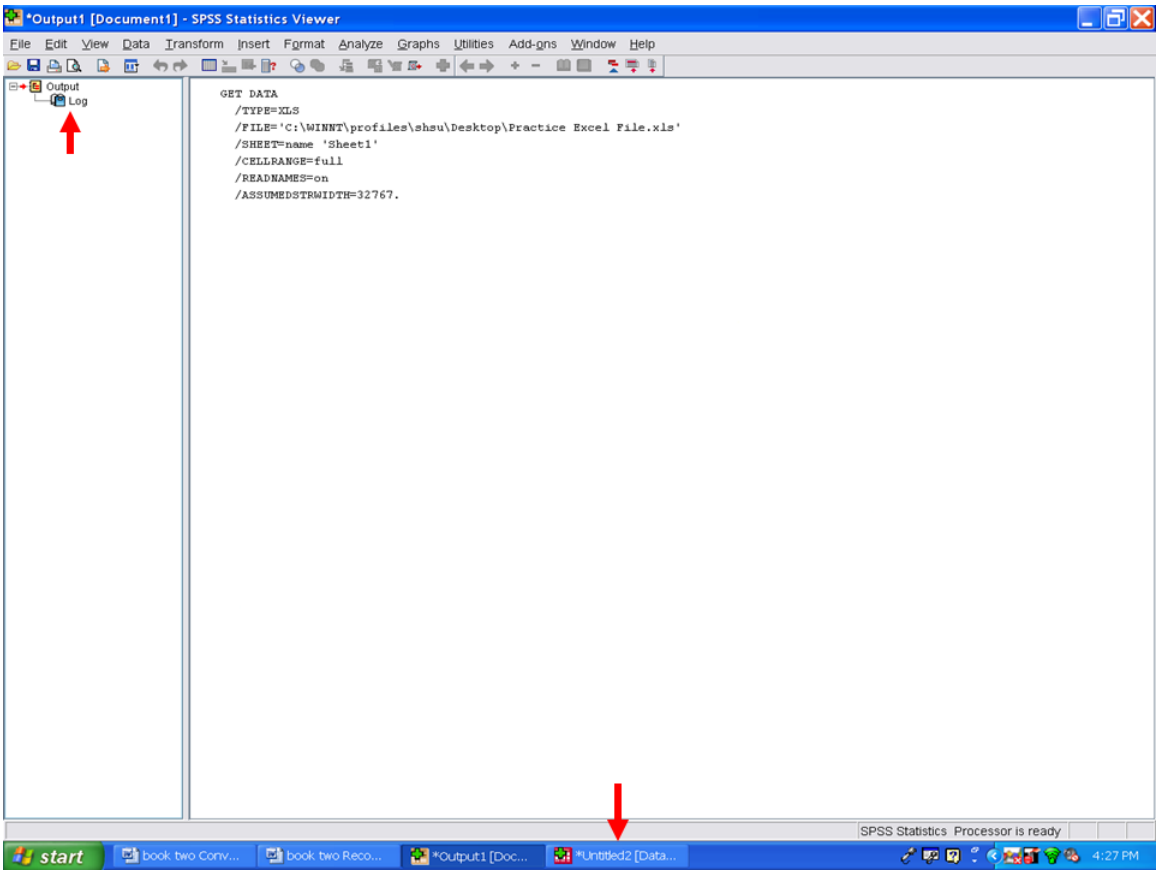
[10] http://cnx.org/content/m40709/latest/3.7.png/image



11

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

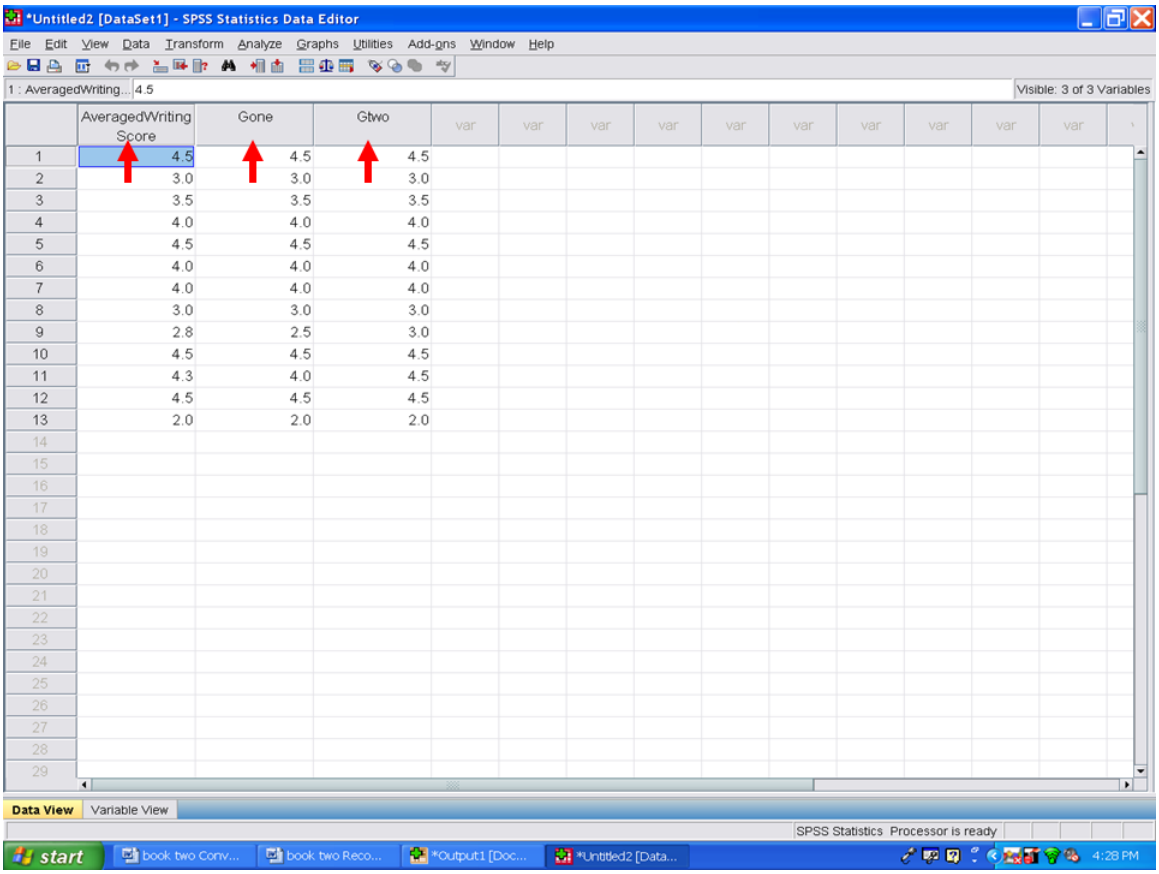
[11] http://cnx.org/content/m40709/latest/3.8.png/image



12

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.

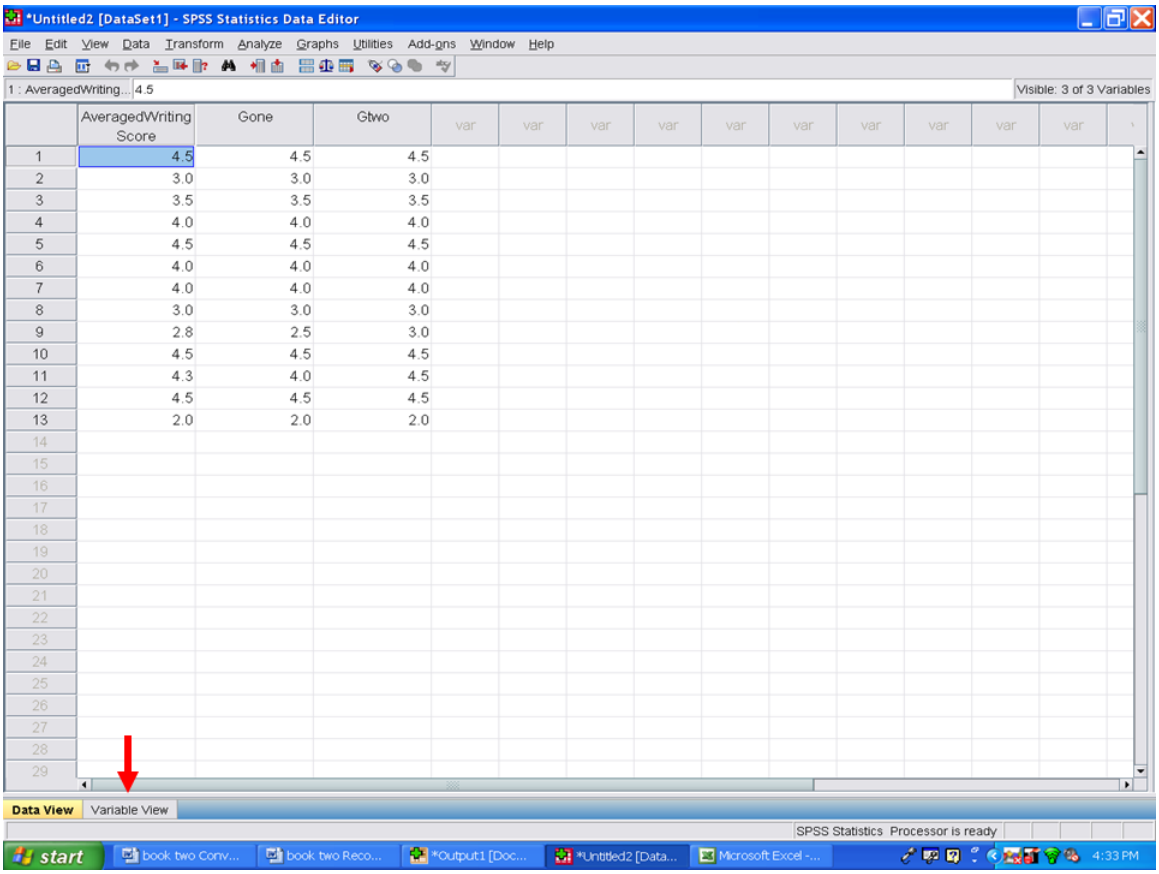
[12] http://cnx.org/content/m40709/latest/3.9.png/image



[13]

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

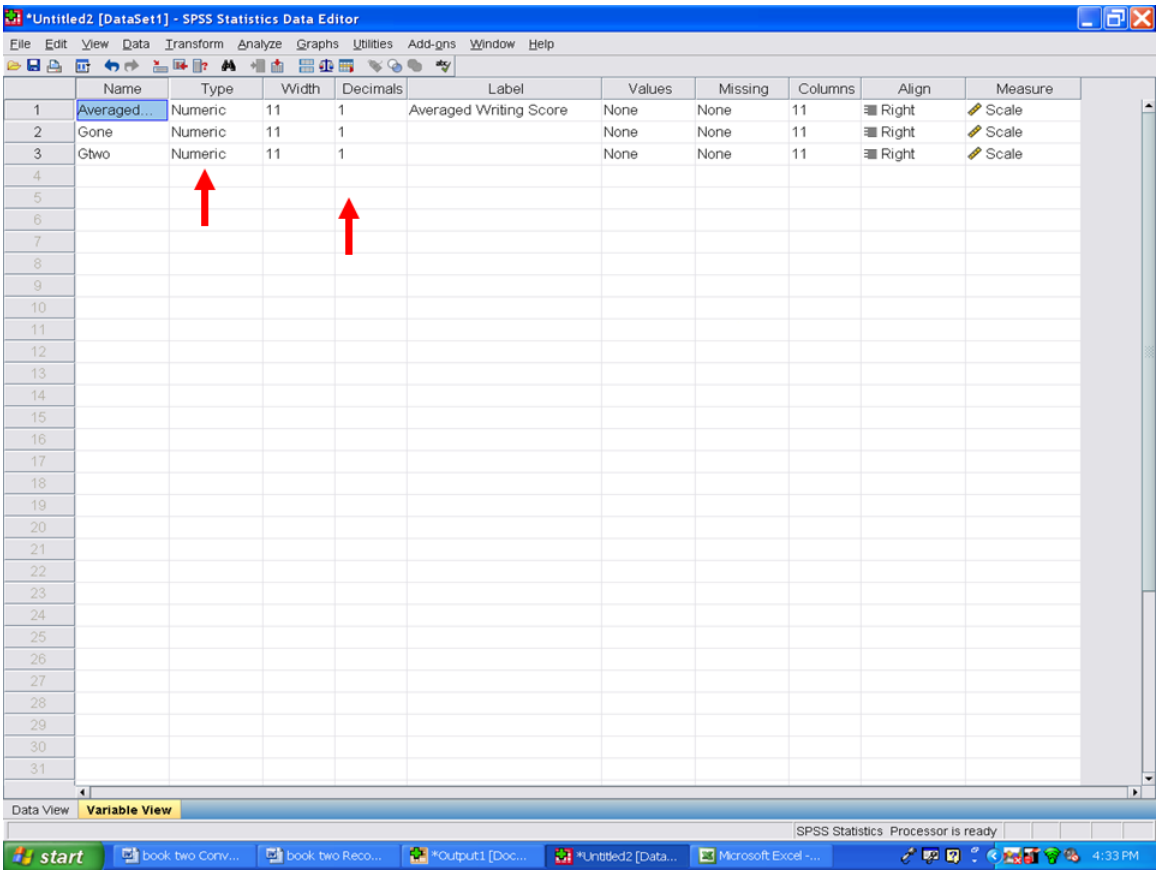
[13] http://cnx.org/content/m40709/latest/3.10.png/image



14

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

[14] http://cnx.org/content/m40709/latest/3.11.png/image

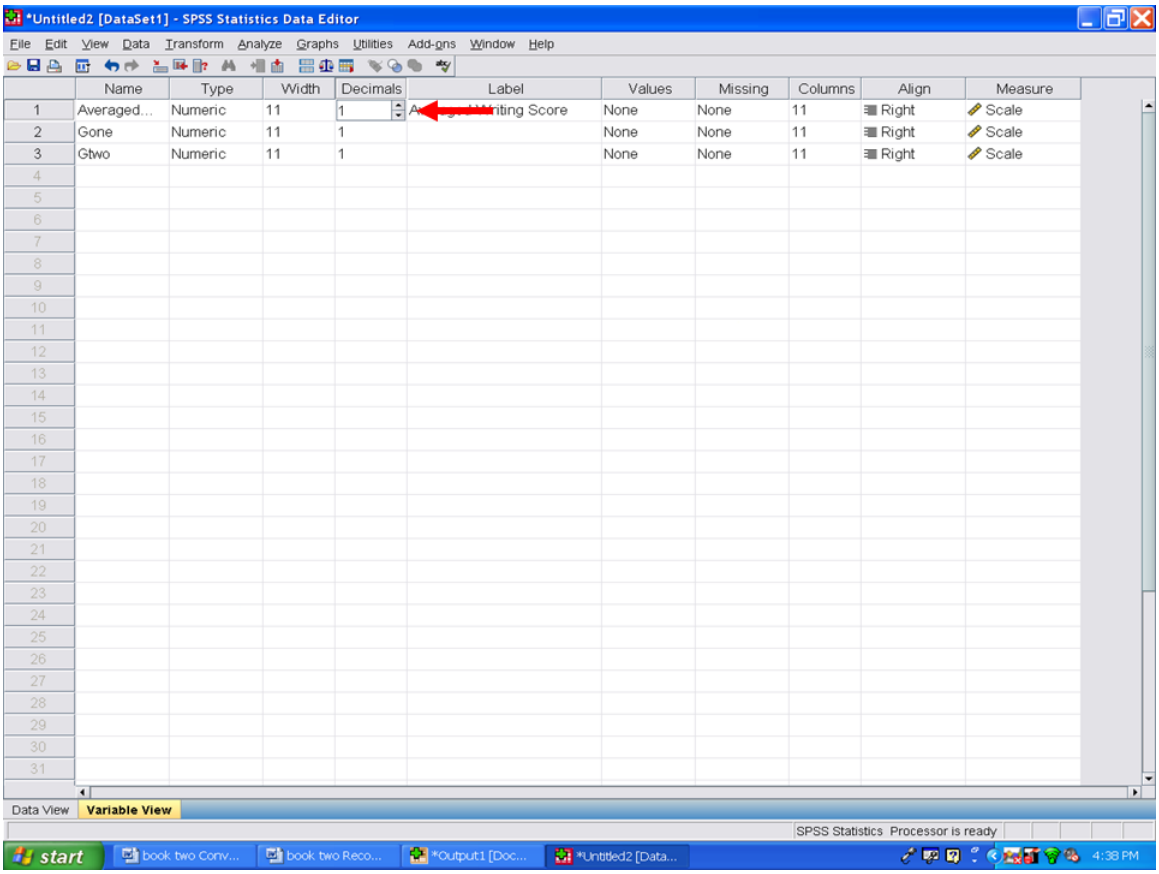


[15]

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.

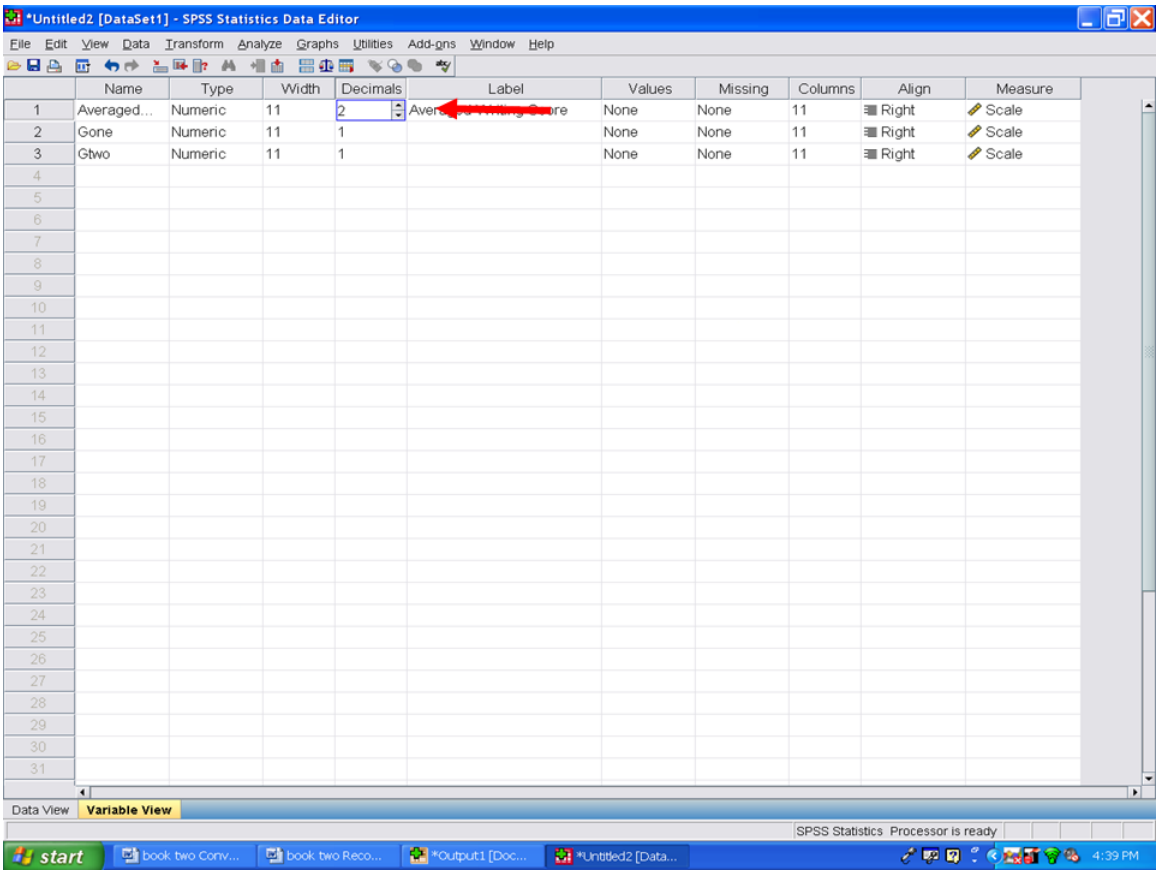
[15] http://cnx.org/content/m40709/latest/3.12.png/image



16

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.

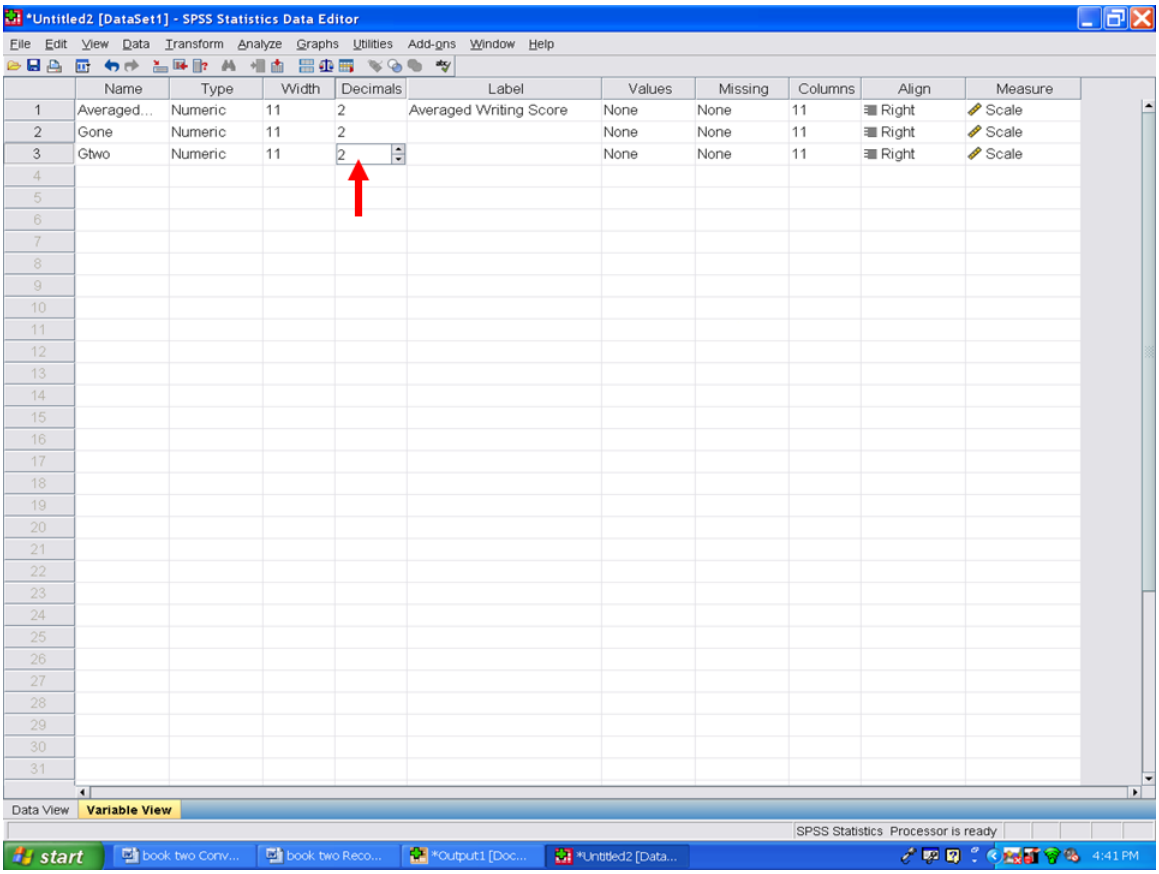
[16] http://cnx.org/content/m40709/latest/3.13.png/image



17

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

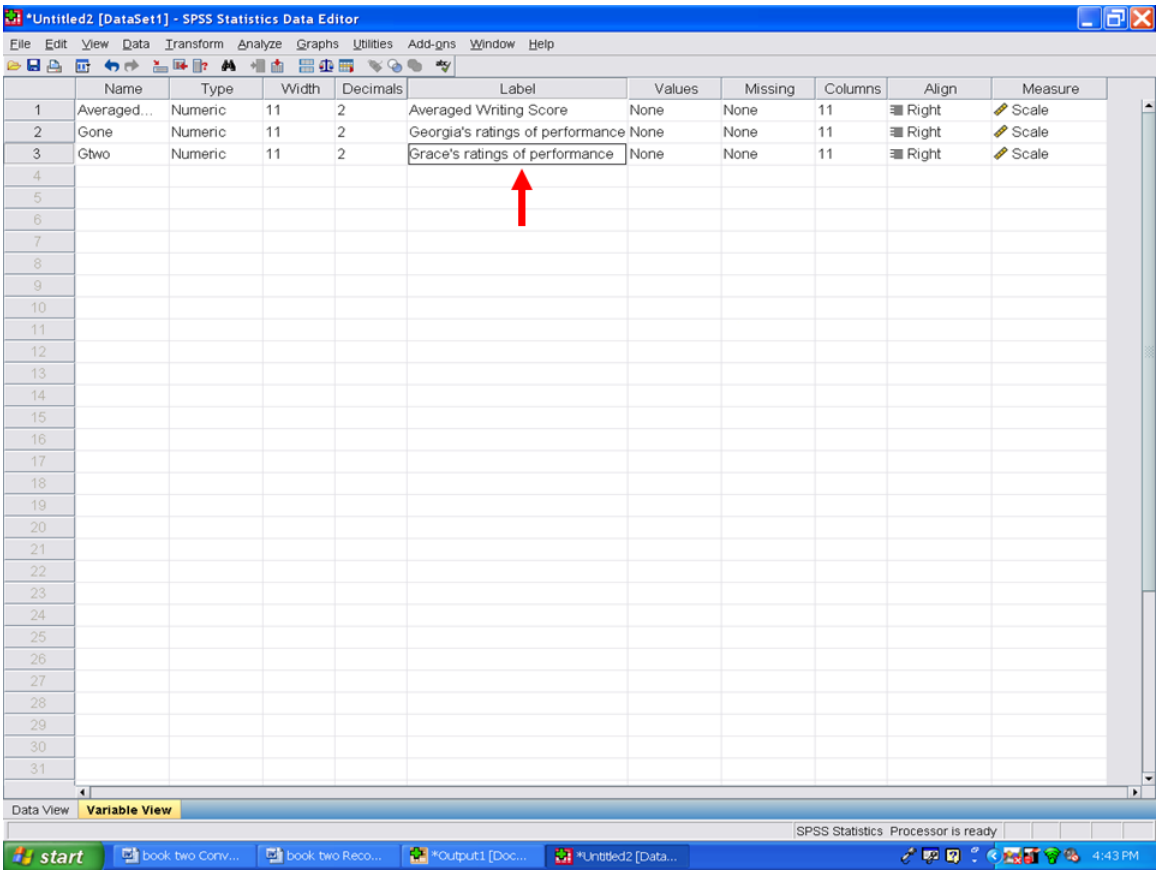
[17] http://cnx.org/content/m40709/latest/3.14.png/image



[18]

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



[19]

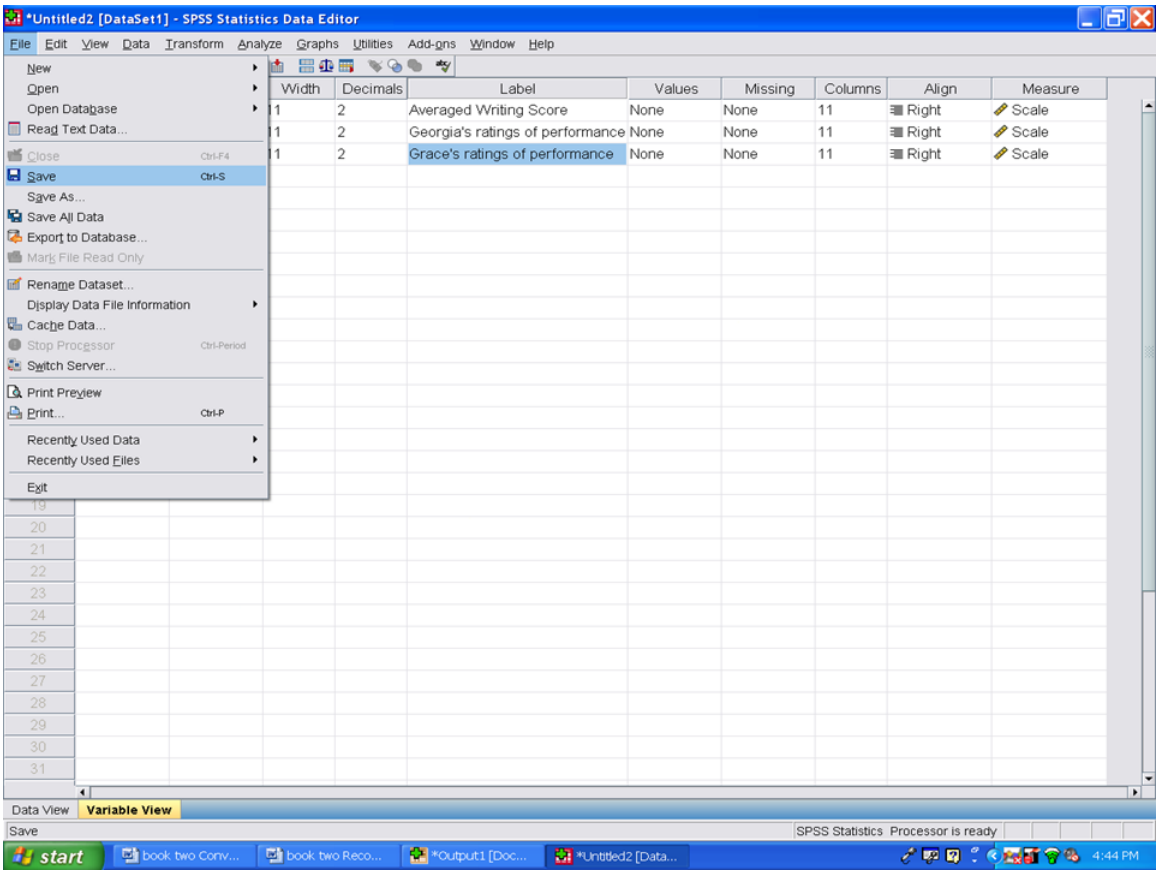
We have converted our Excel file into a SPSS dataset. Please save it prior to conducting any statistical analyses.

√ 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]

[20] http://cnx.org/content/m40709/latest/3.17.png/image

Chapter 5

5. Typing Survey Data Into SPSS[1]

NOTE: This chapter is published by NCPEA Press[2] 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.

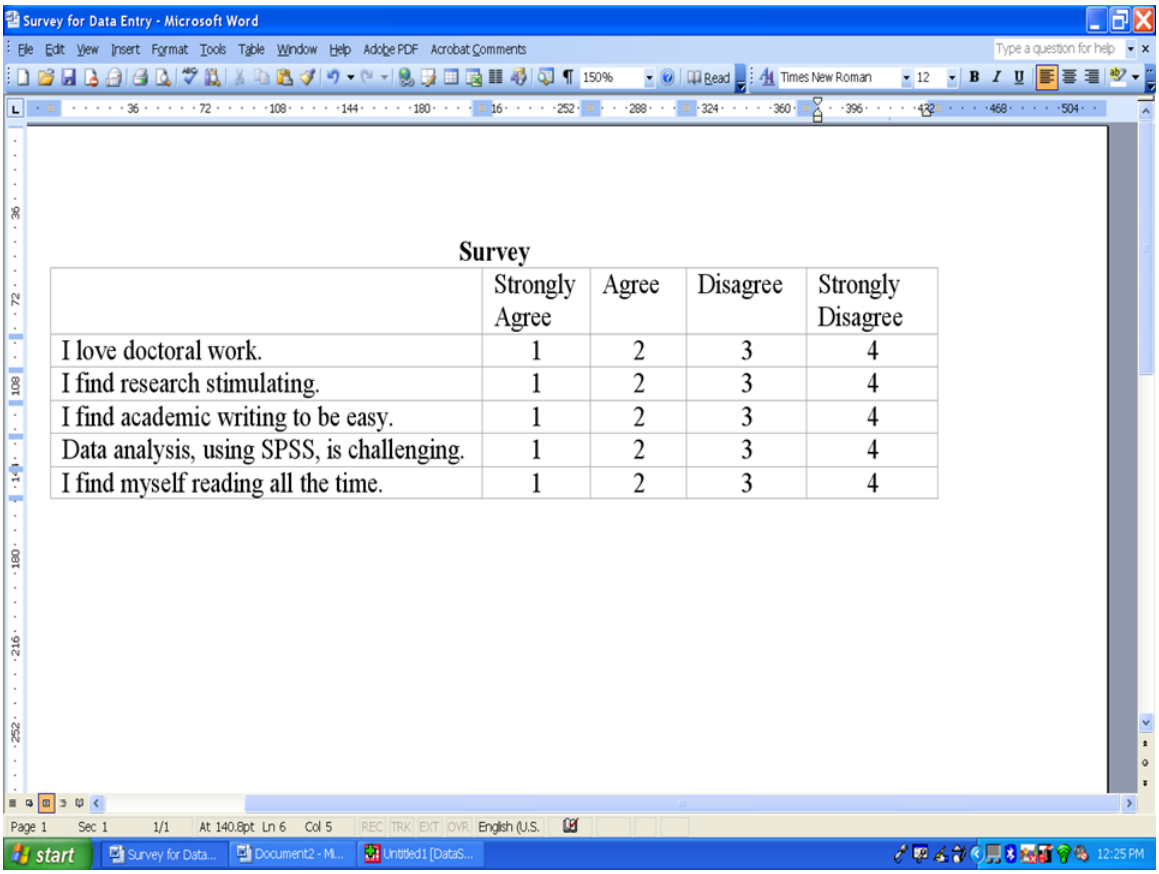
About the Authors

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 Help[3]) 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:

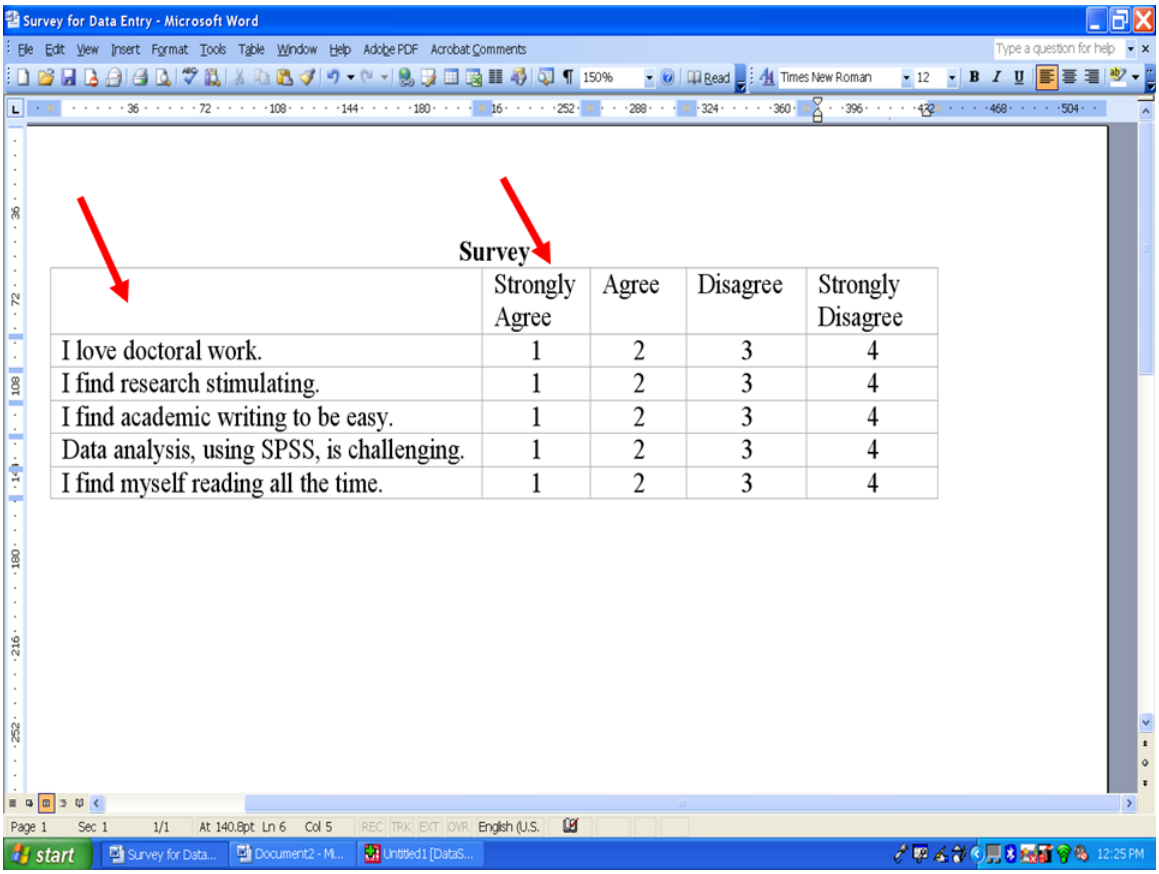
[1] This content is available online at <http://cnx.org/content/m40713/1.2/>.
[2] http://www.ncpeapublications.org/books.html
[3] http://cnx.org/content/m40713/latest/www.writingandstatisticalhelp



[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.

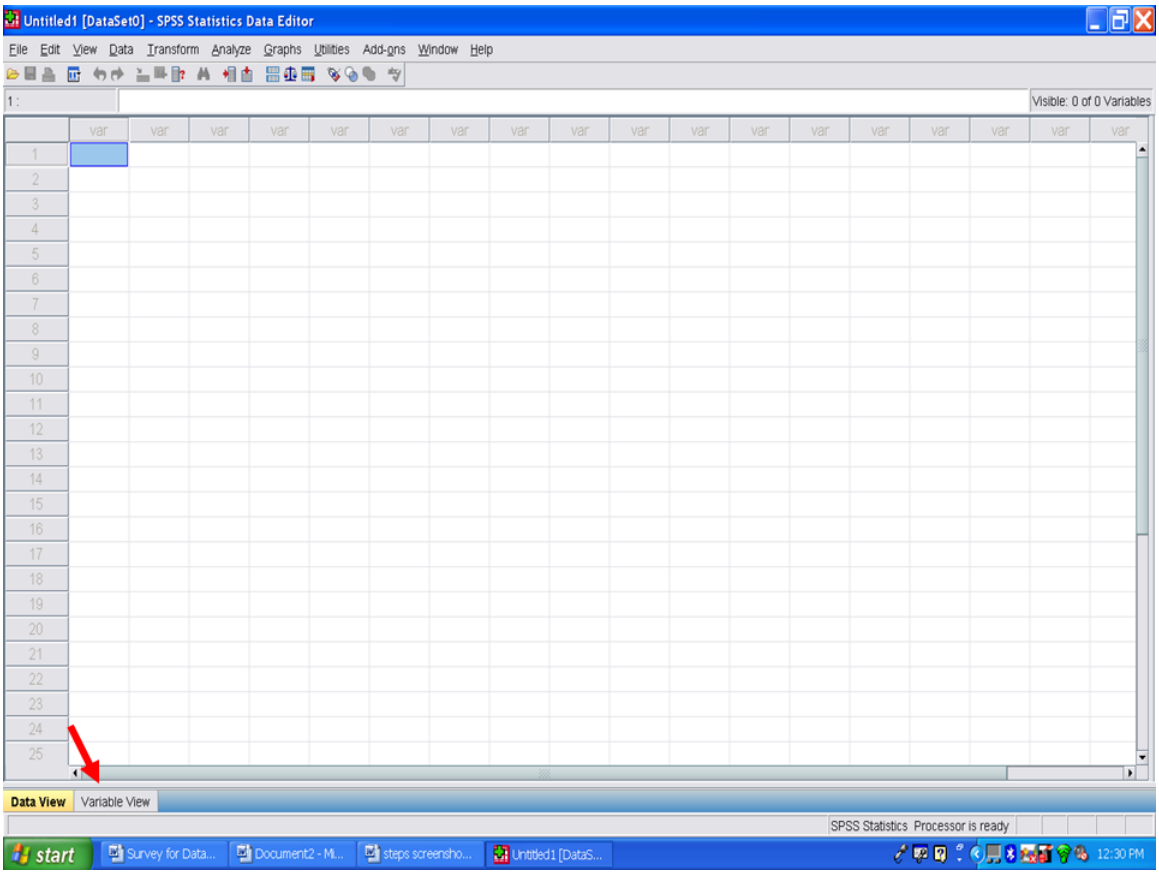
[4] http://cnx.org/content/m40713/latest/4.1.png/image



[5]

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.

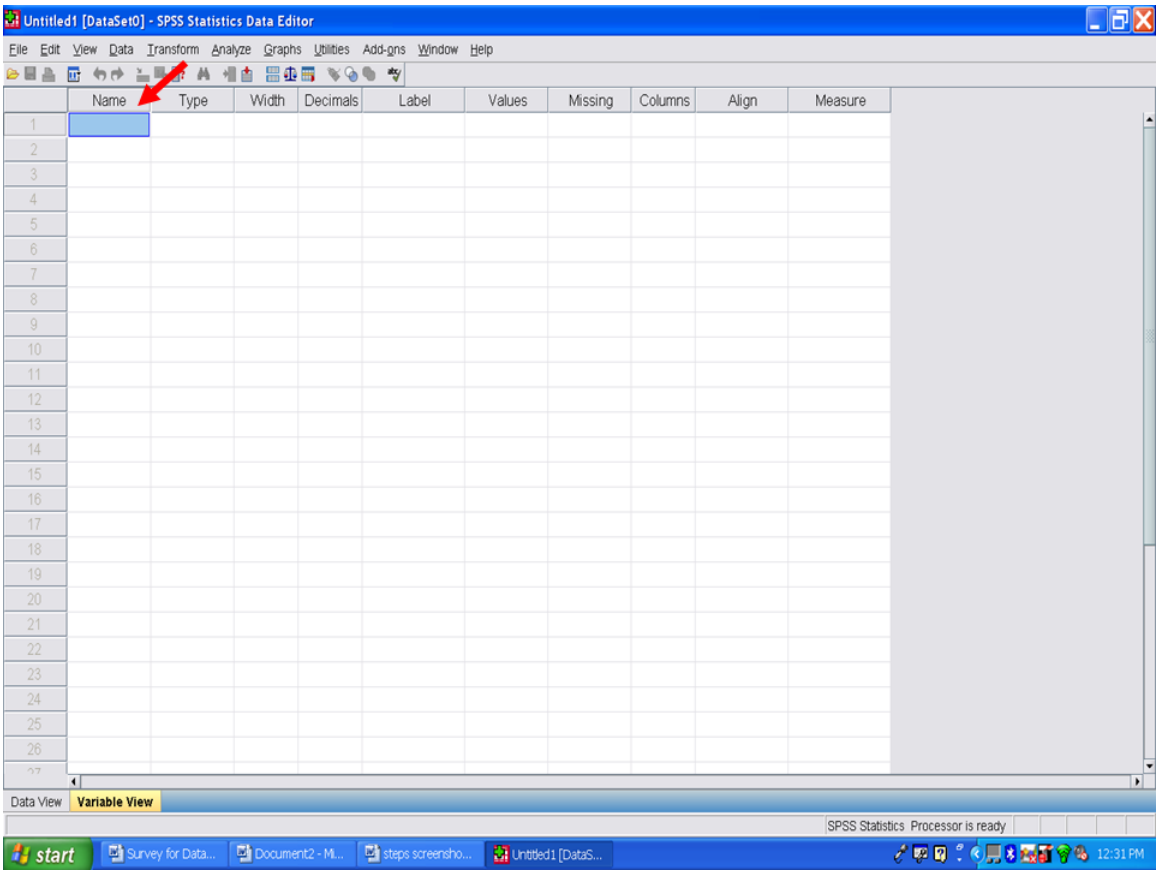
[5] http://cnx.org/content/m40713/latest/4.2.png/image



[6]

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.

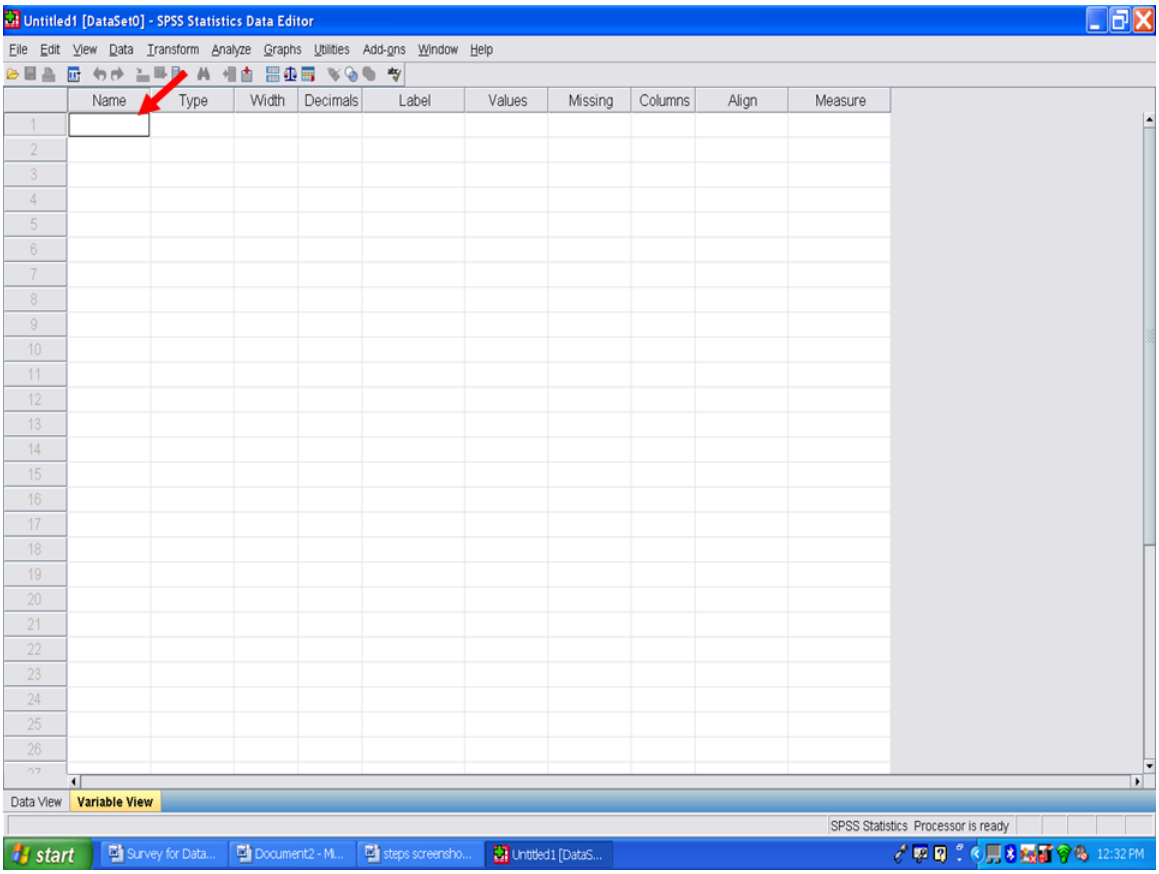
[6] http://cnx.org/content/m40713/latest/4.3.png/image



[7]

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

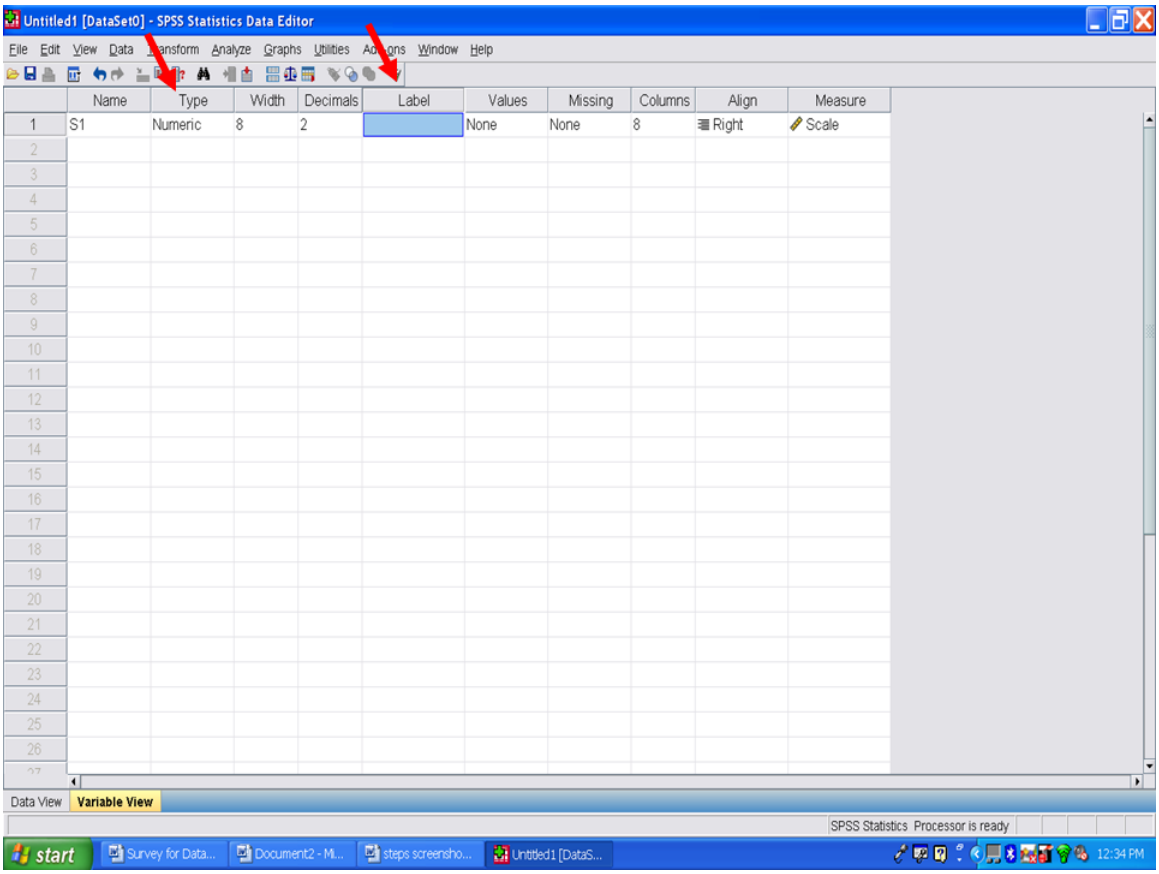
[7] http://cnx.org/content/m40713/latest/4.4.png/image



[8]

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.

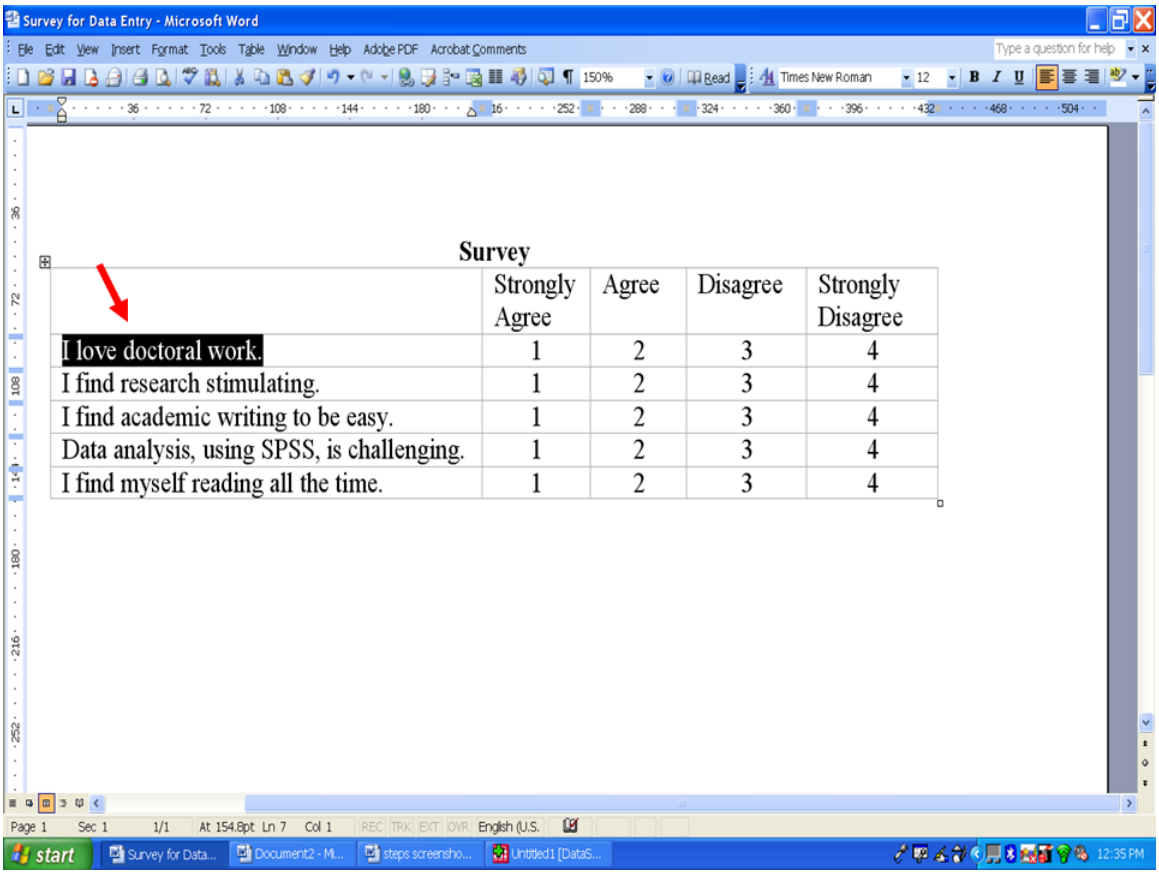
[8] http://cnx.org/content/m40713/latest/4.5.png/image



[9]

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

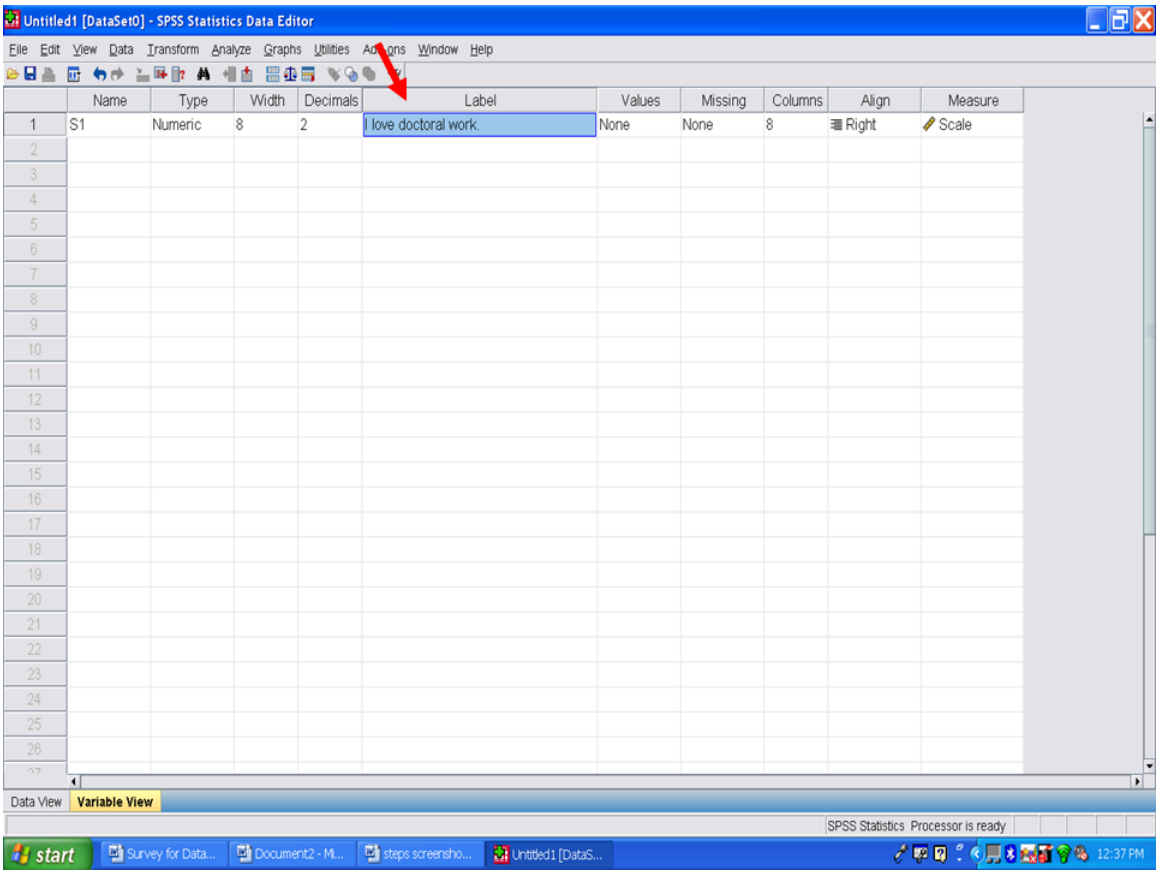
[9] http://cnx.org/content/m40713/latest/4.6.png/image



10

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

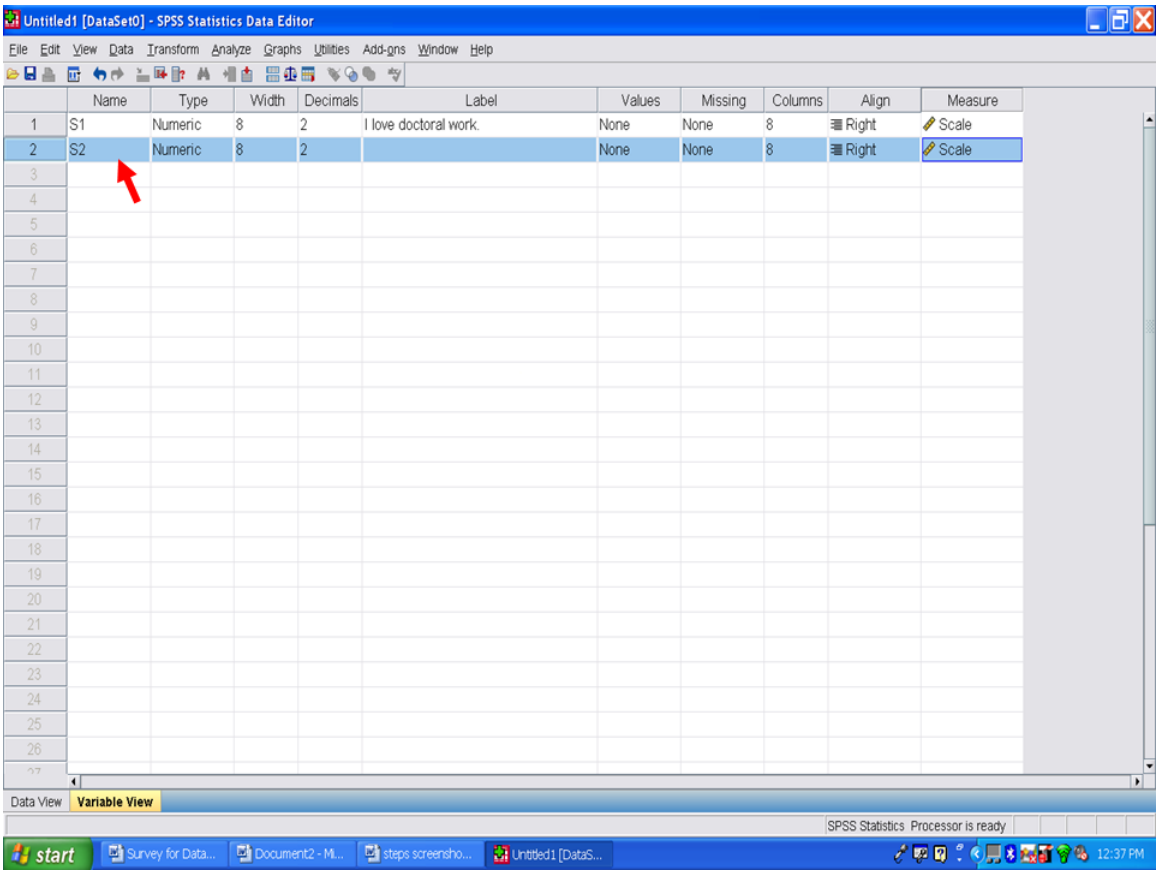
[10] http://cnx.org/content/m40713/latest/4.71.png/image



[11]

Repeat this process for S2.

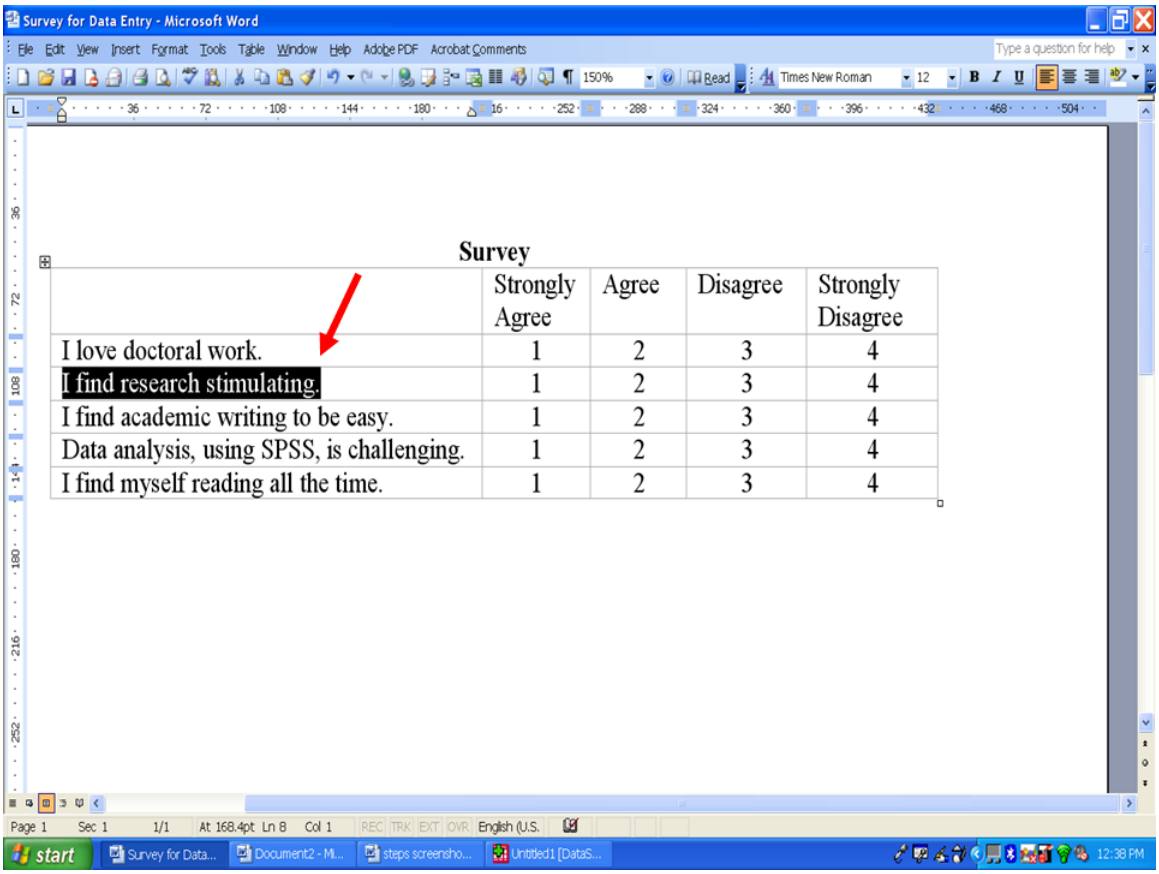
[11] http://cnx.org/content/m40713/latest/4.8.png/image



12

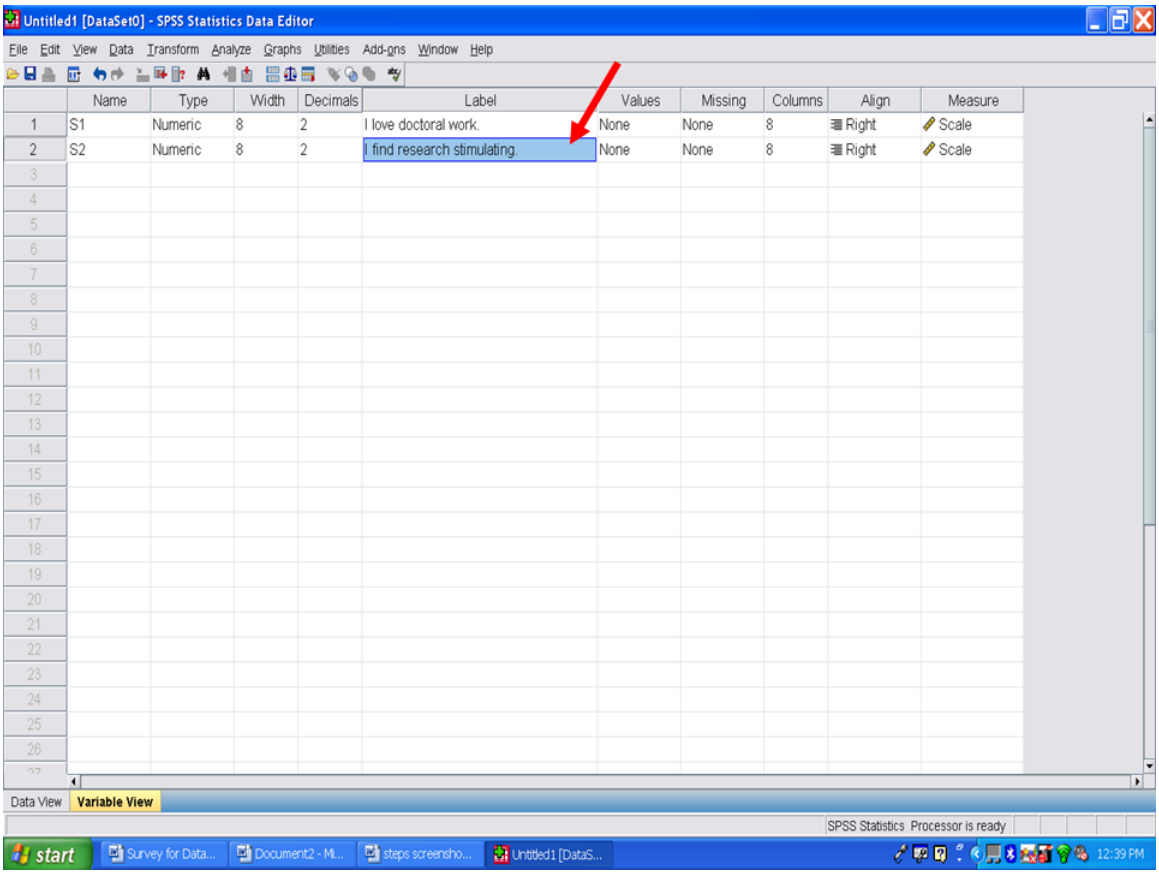
Go to your Word file and copy the second survey item.

[12] http://cnx.org/content/m40713/latest/4.9.png/image



Paste it in the Label column for the second row.

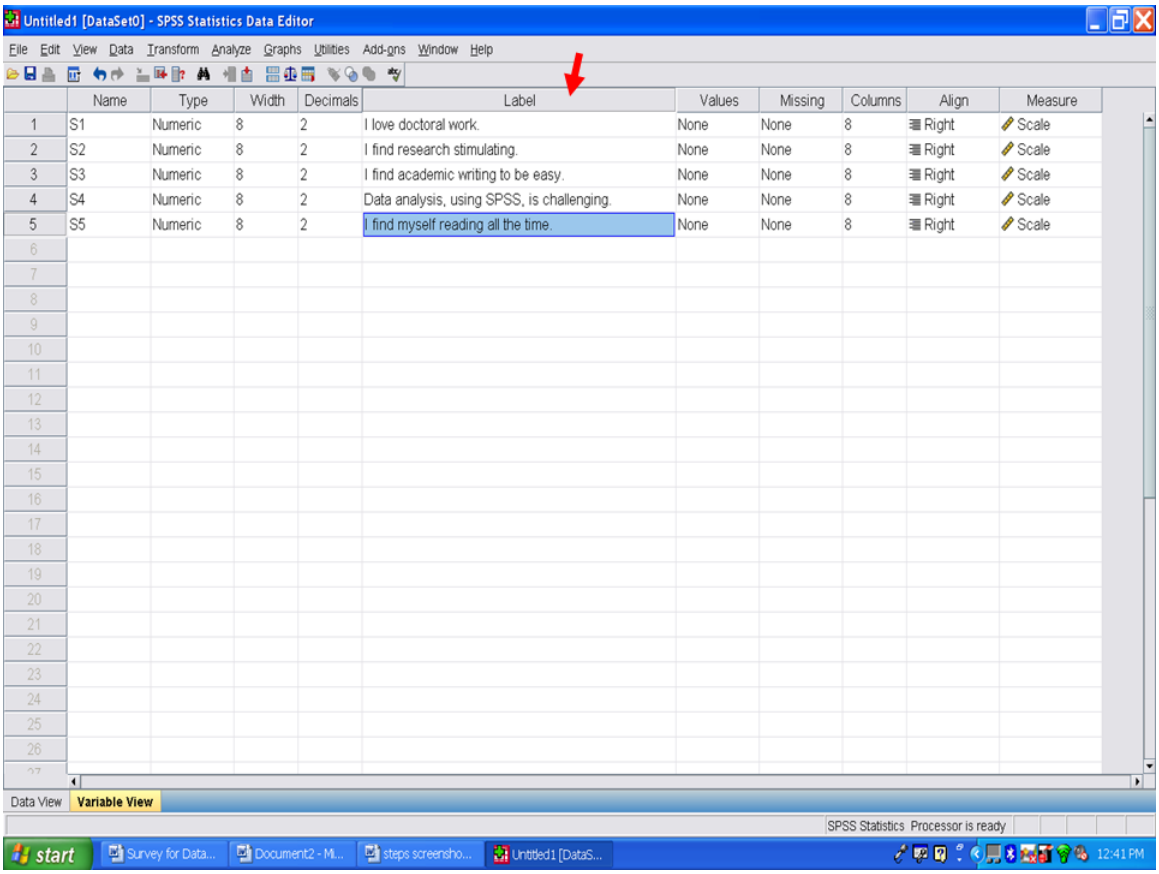
[13] http://cnx.org/content/m40713/latest/4.10.png/image



14

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

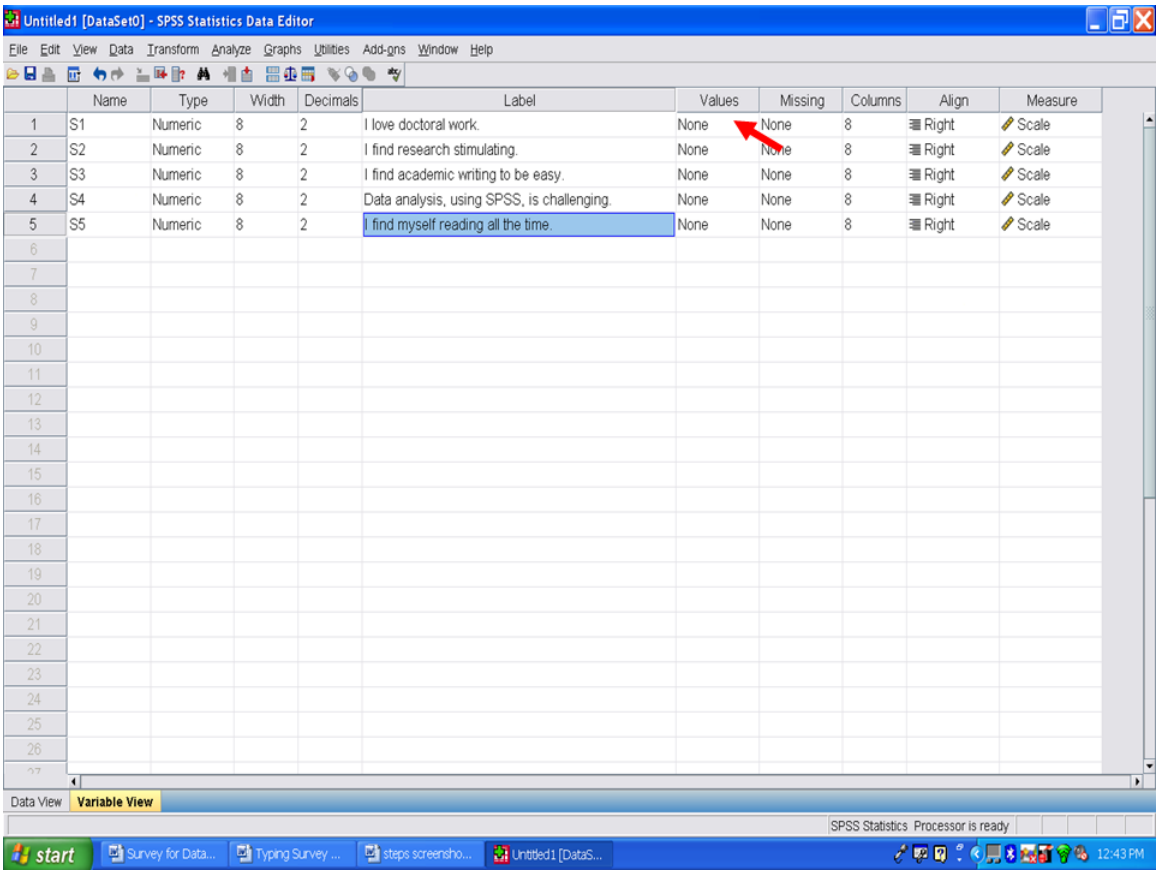
[14] http://cnx.org/content/m40713/latest/4.11.png/image



15

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.

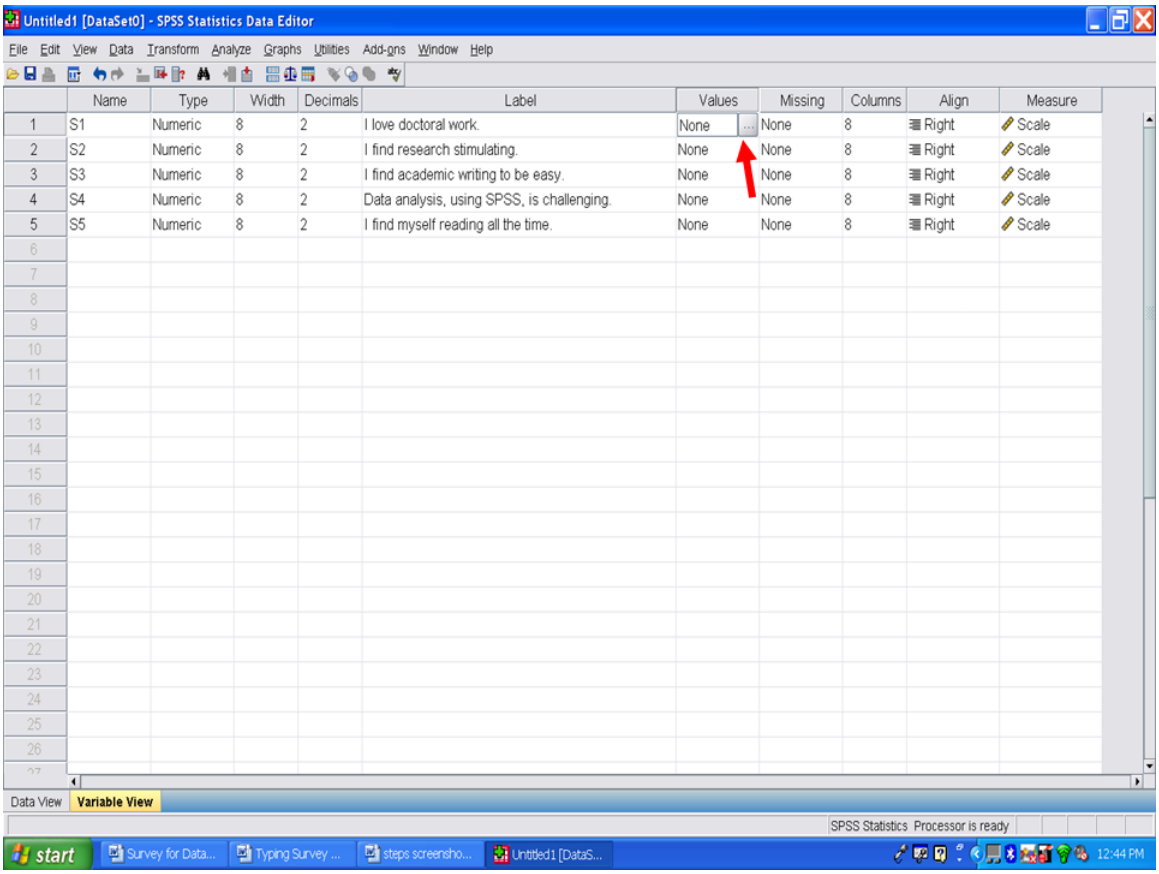
[15] http://cnx.org/content/m40713/latest/4.12.png/image



16

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.

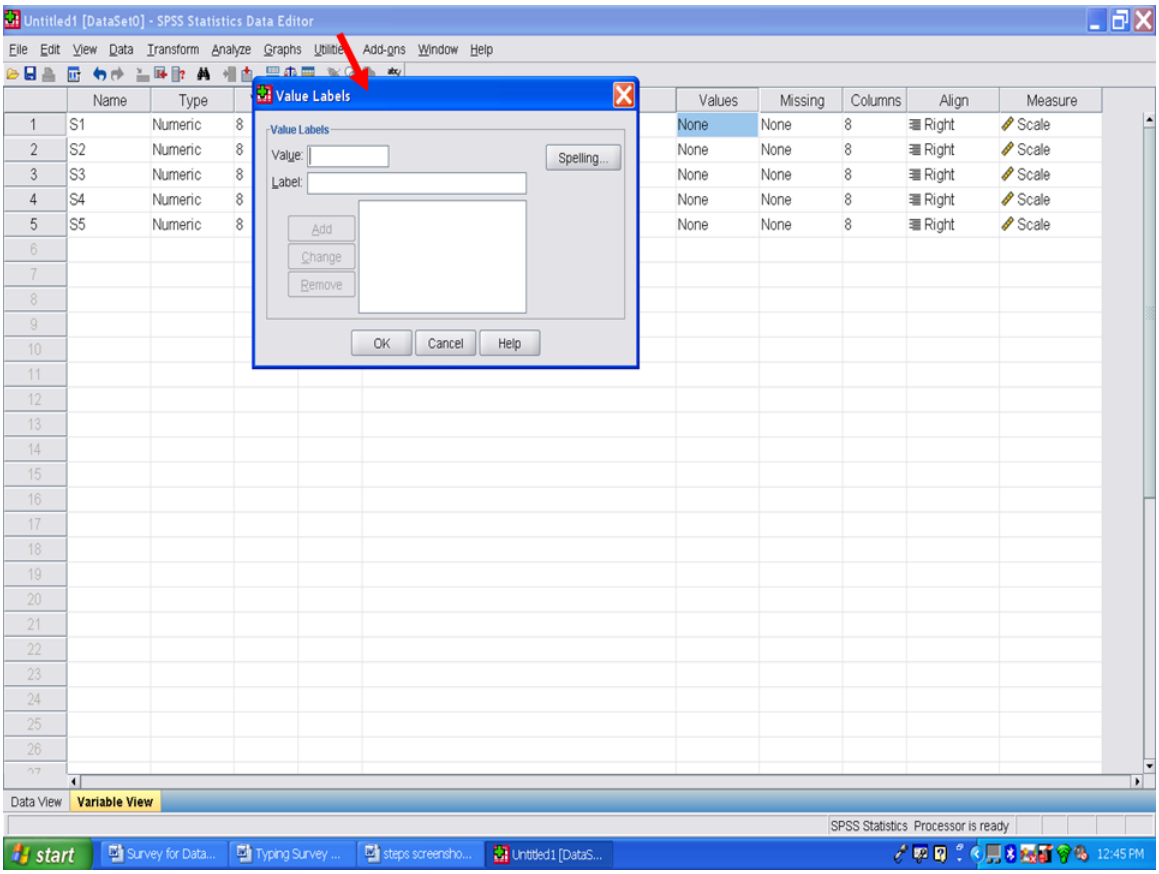
[16] http://cnx.org/content/m40713/latest/4.13.png/image



17

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

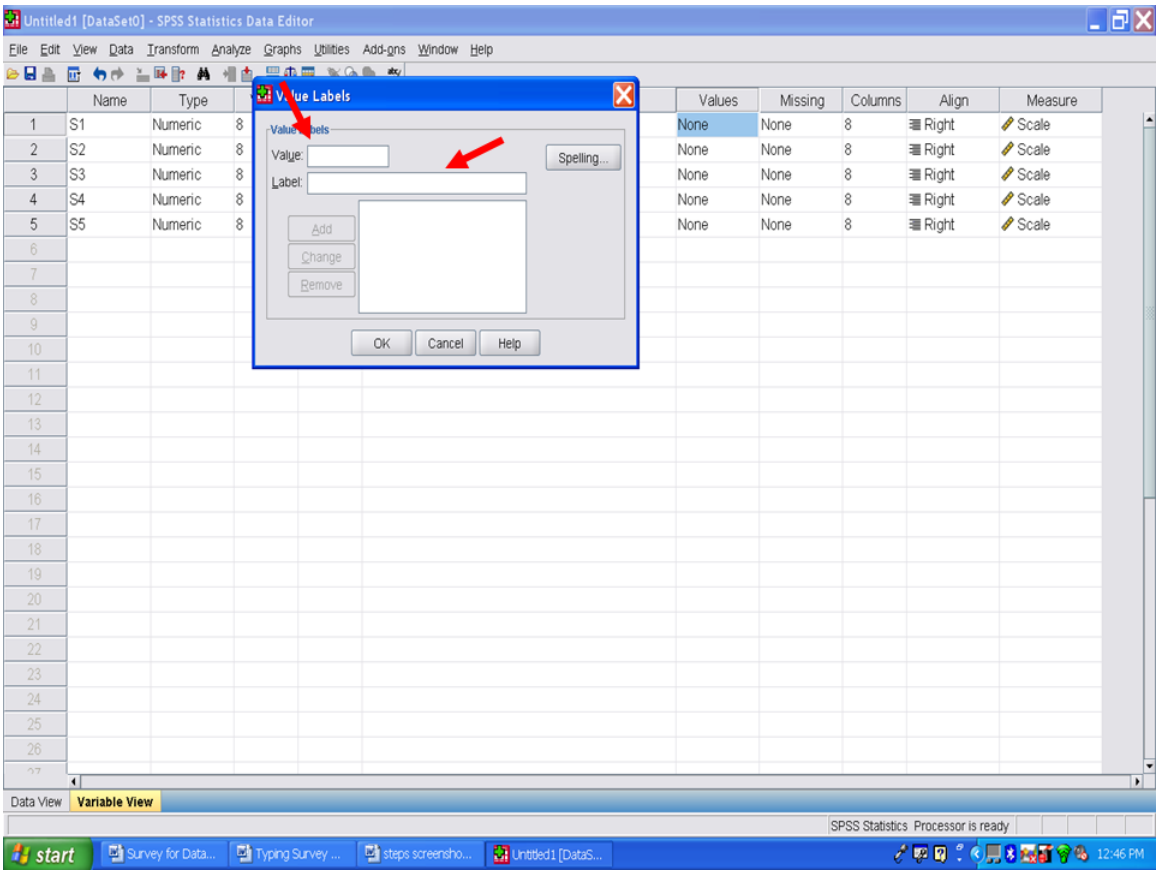
[17] http://cnx.org/content/m40713/latest/4.4.png/image



[18]

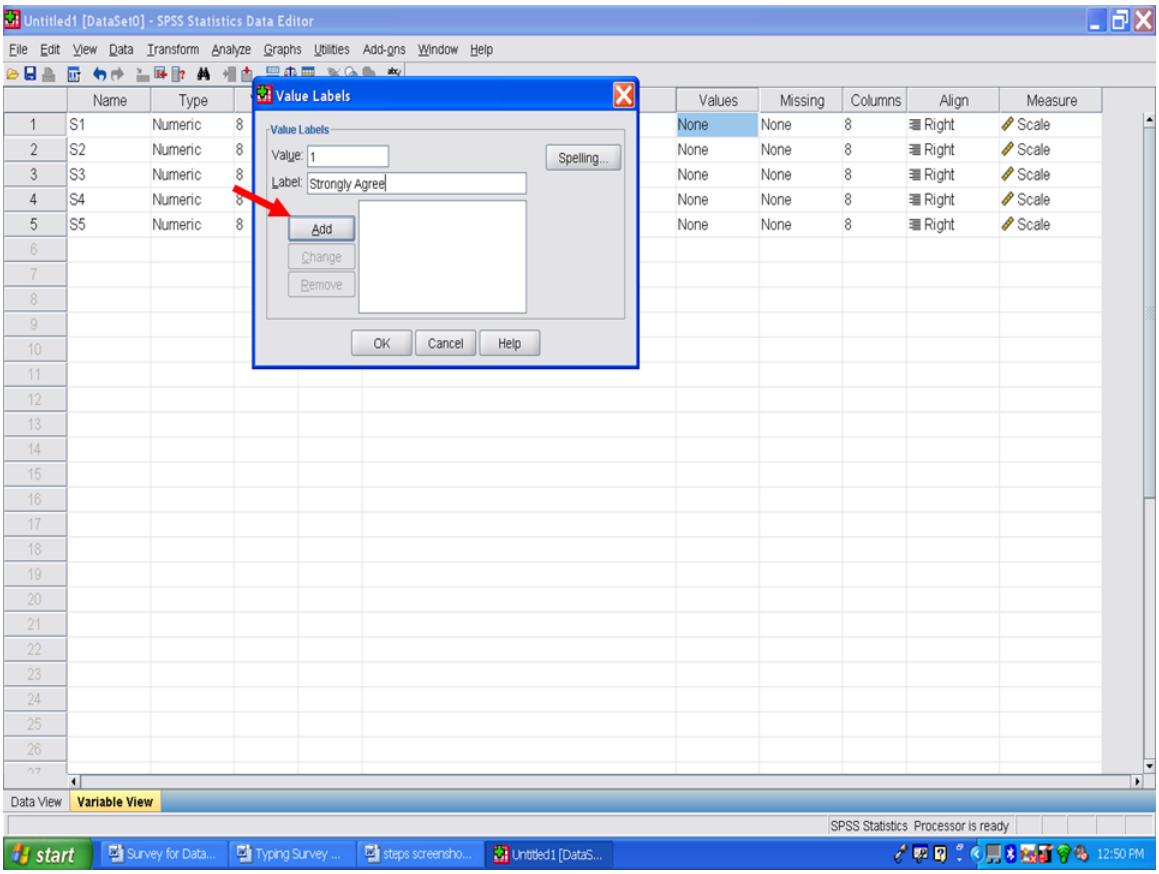
In this screen, we will type in a numeric value in the Value cell and the Likert-format response label in the Label cell. That is, our four values are 1, 2, 3, and 4, which correspond to a Strongly Agree, Agree, Disagree, and Strongly Disagree. Refer back to the survey which we are using for this example.

[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.

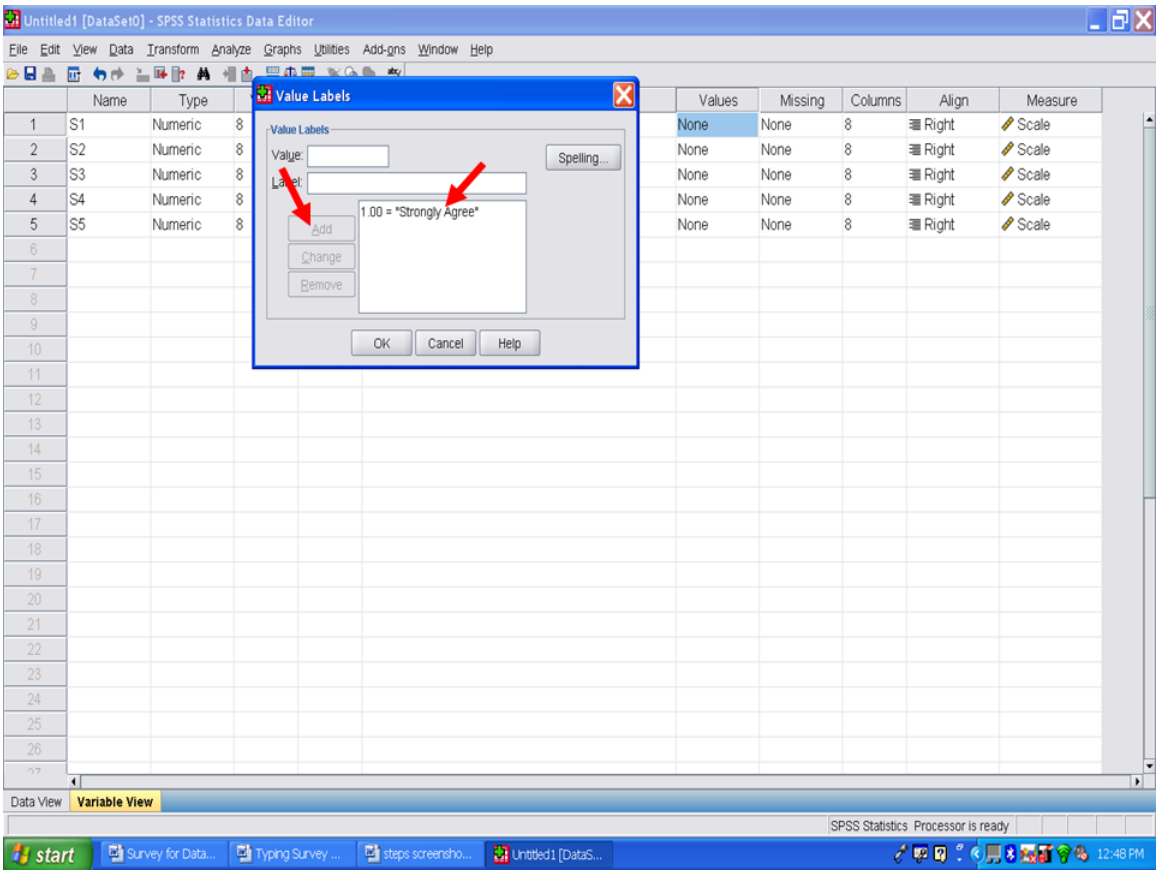
[19] http://cnx.org/content/m40713/latest/4.16png/image



[20]

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

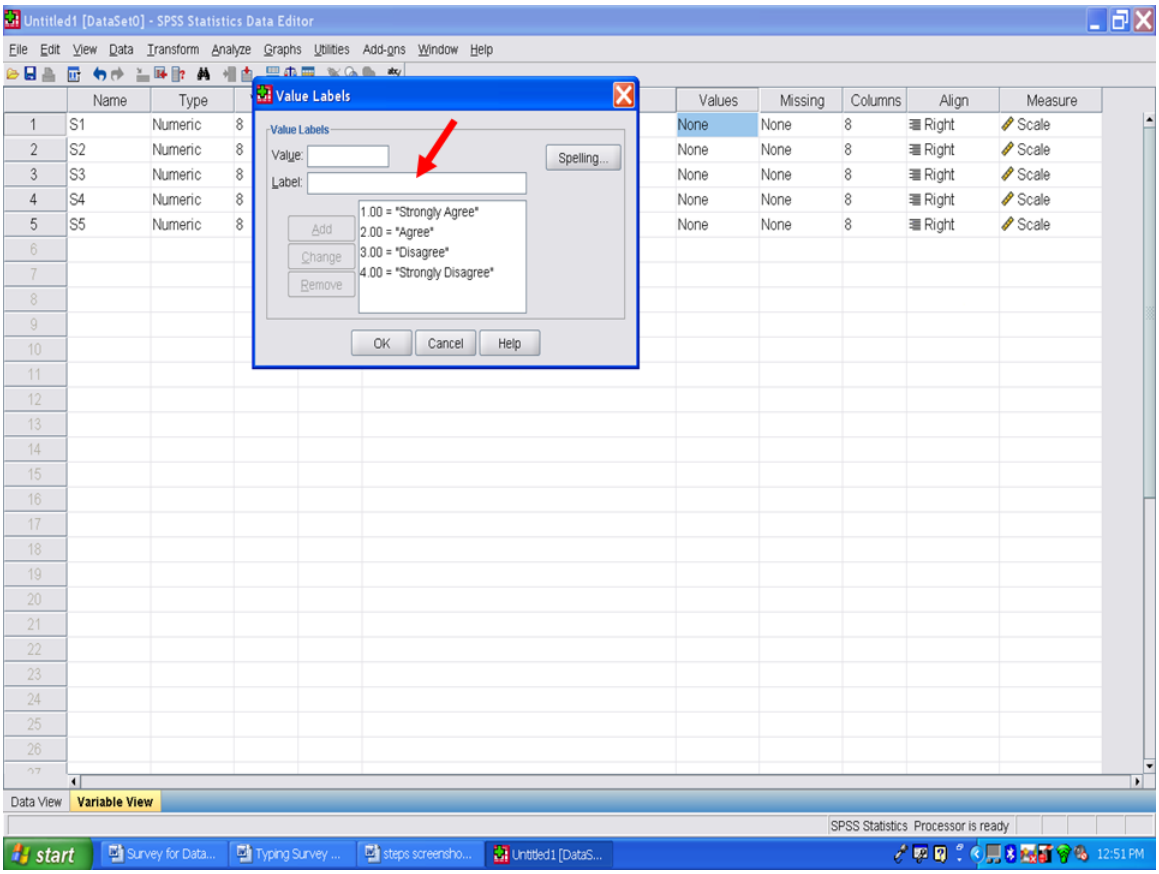
[20] http://cnx.org/content/m40713/latest/4.17.png/image



[21]

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

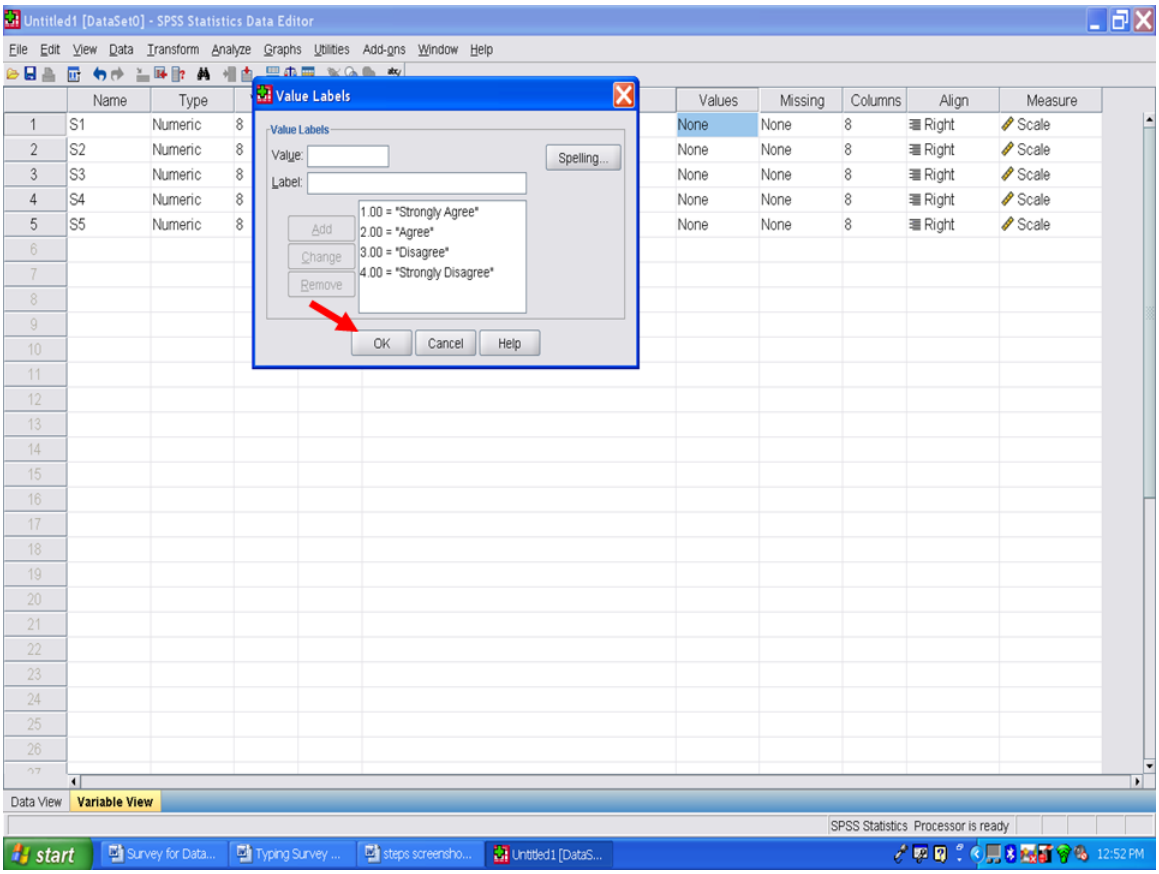
[21] http://cnx.org/content/m40713/latest/4.18.png/image



22

Now click on OK.

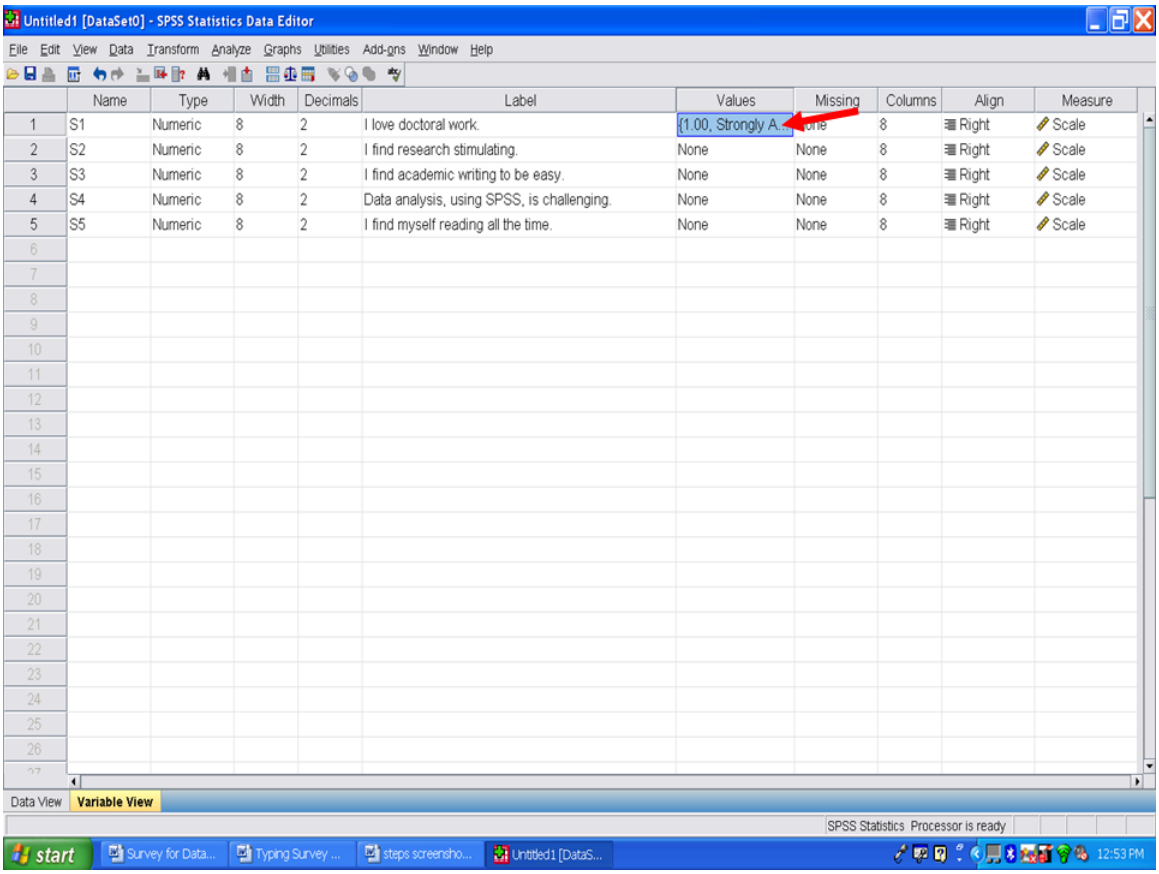
[22] http://cnx.org/content/m40713/latest/4.19.png/image



[23]

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.

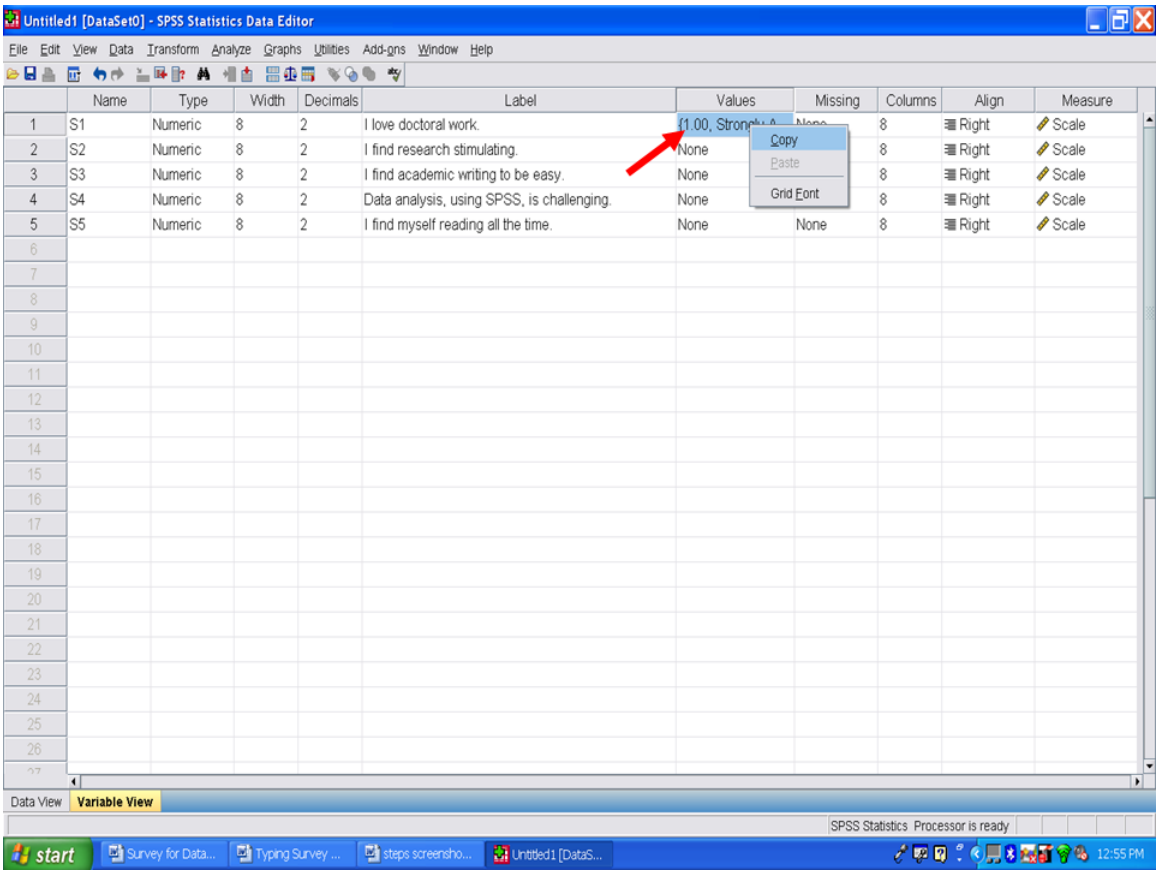
[23] http://cnx.org/content/m40713/latest/4.20.png/image



[24]

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

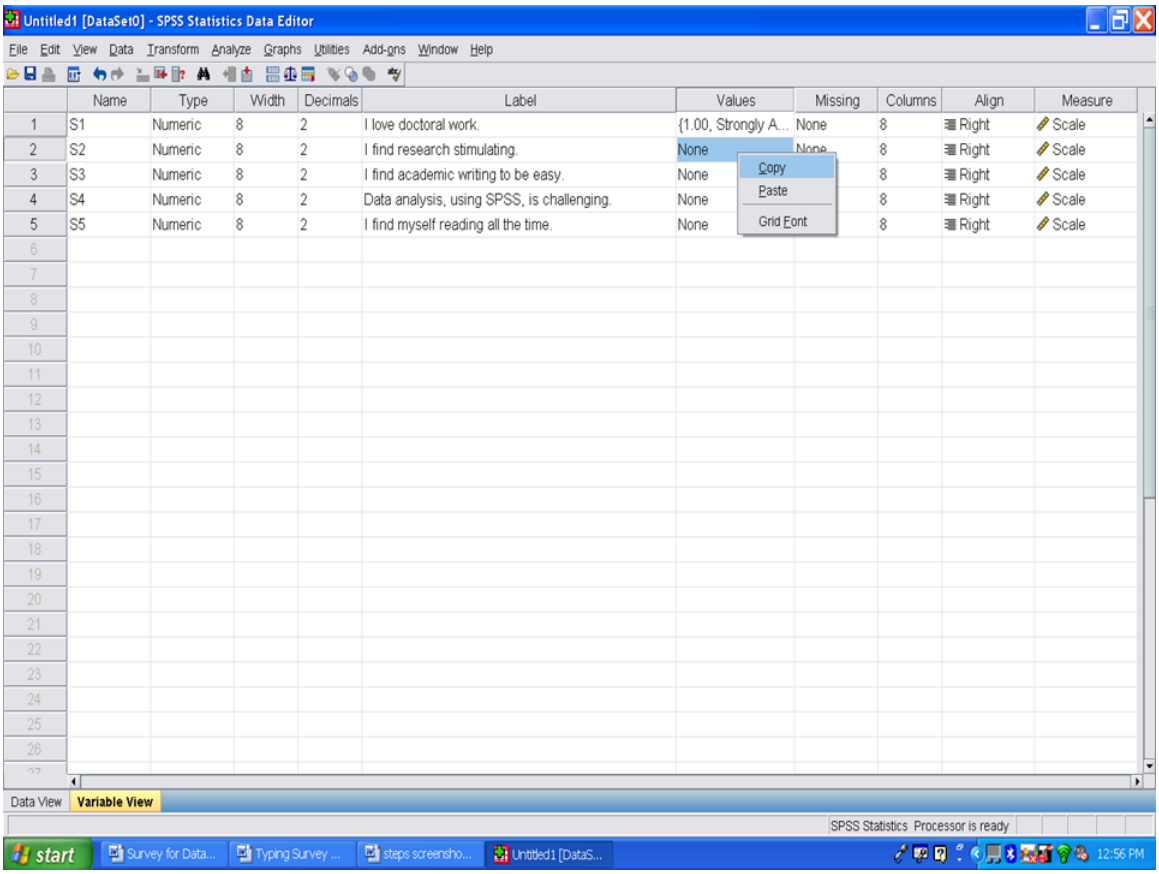
[24] http://cnx.org/content/m40713/latest/4.21.png/image



[25]

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

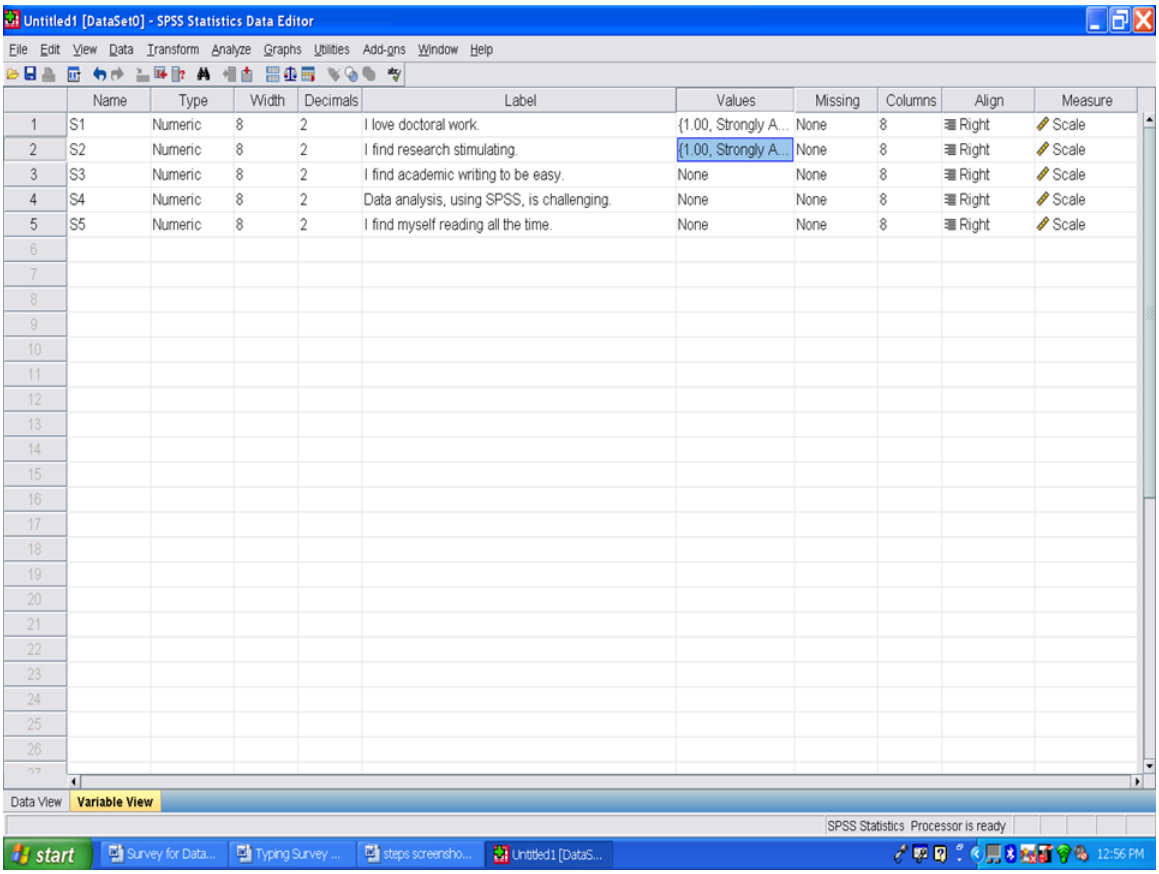
[25] http://cnx.org/content/m40713/latest/4.22.png/image



[26]

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.

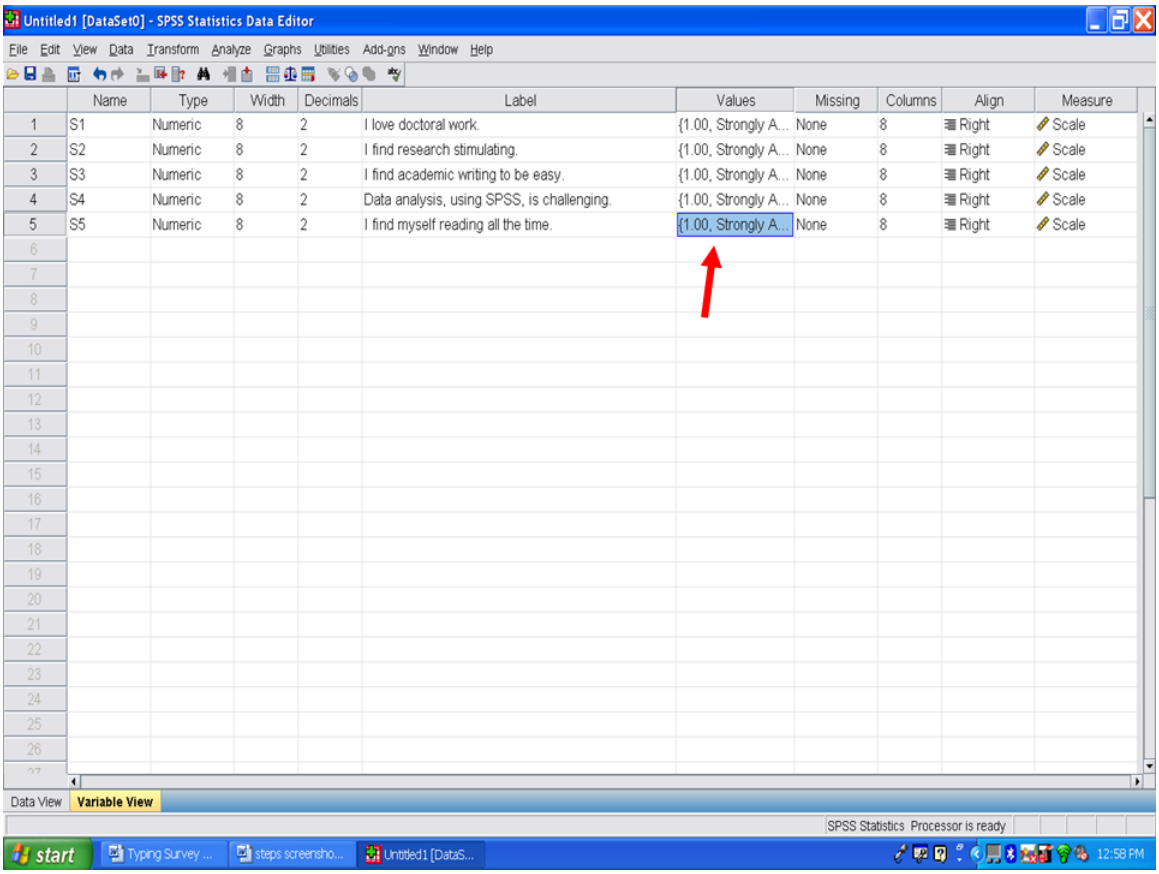
[26] http://cnx.org/content/m40713/latest/4.23.png/image



[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



[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 I[1]

NOTE: This chapter is published by NCPEA Press[2] 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 Help[3]) 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.

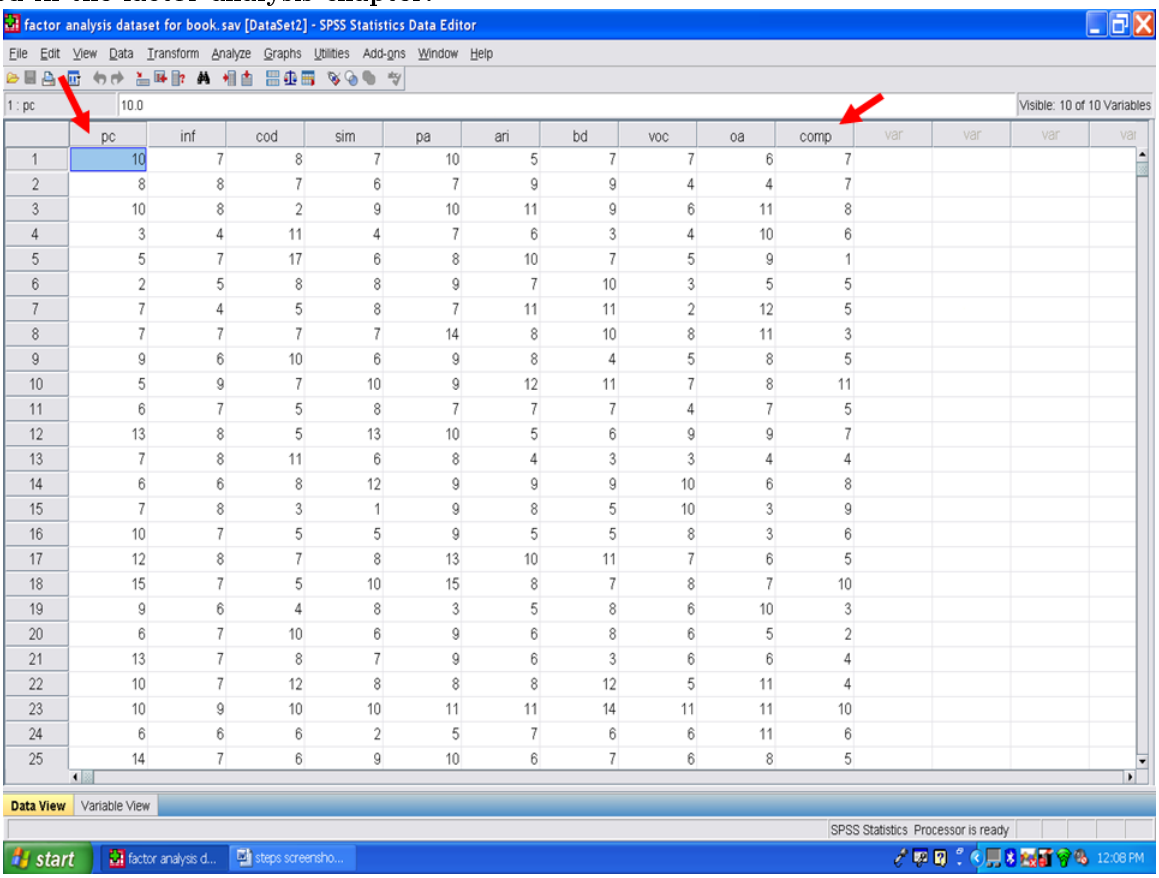
In this set of steps and screenshots, you will be presented with information regarding how to determine the internal consistency (i.e., reliability) of scores derived from a larger set of variables. In particular, this chapter follows up on the factor analysis chapter. The subscales that were determined 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

[1] This 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=1-1 [6]

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.

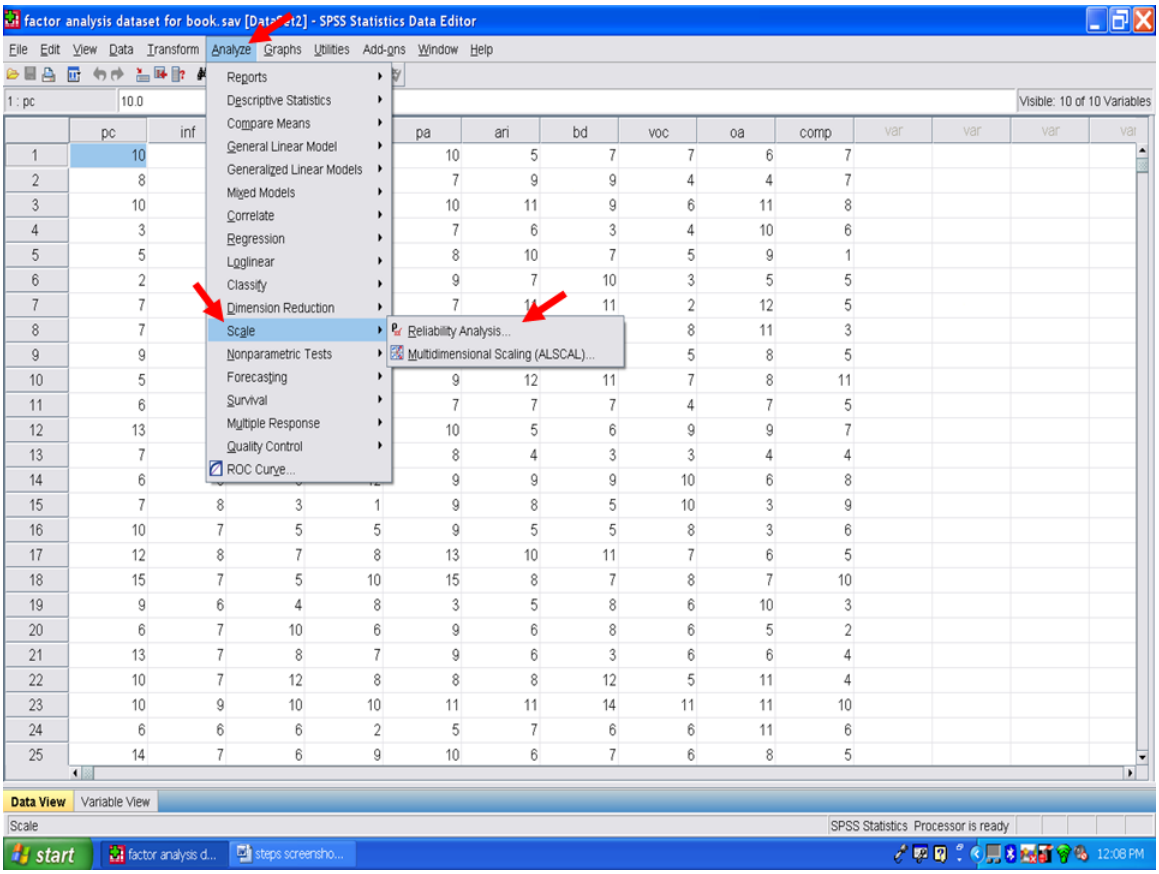
Have your data set pulled into SPSS-PC. In this data set, we have the same 10 subscales used in the factor analysis chapter.



[7]

Click on Analyze
Click on Scale
Click on Reliability Analysis

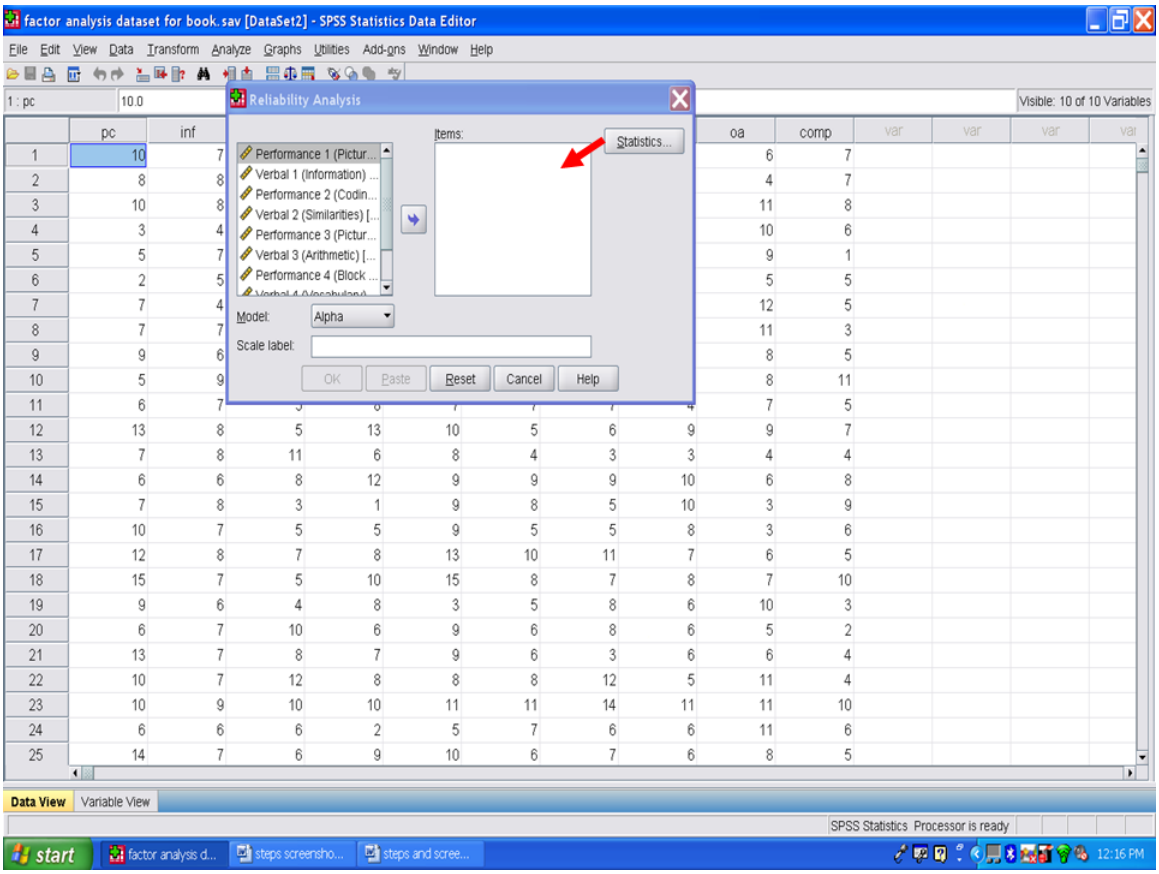
[4] http://davidmlane.com/hyperstat/
[5] http://www.statsoft.com/textbook/
[6] http://www.amazon.com/Discovering-Statistics-Introducing-Statistical-Method/dp/1847879071/ref=sr_1_1?s=books&ie=UTF8&qid=1304961
[7] http://cnx.org/content/m40717/latest/5.1.png/image



[8]

Using the information from the chapter on factor analysis, we want to determ ine 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.

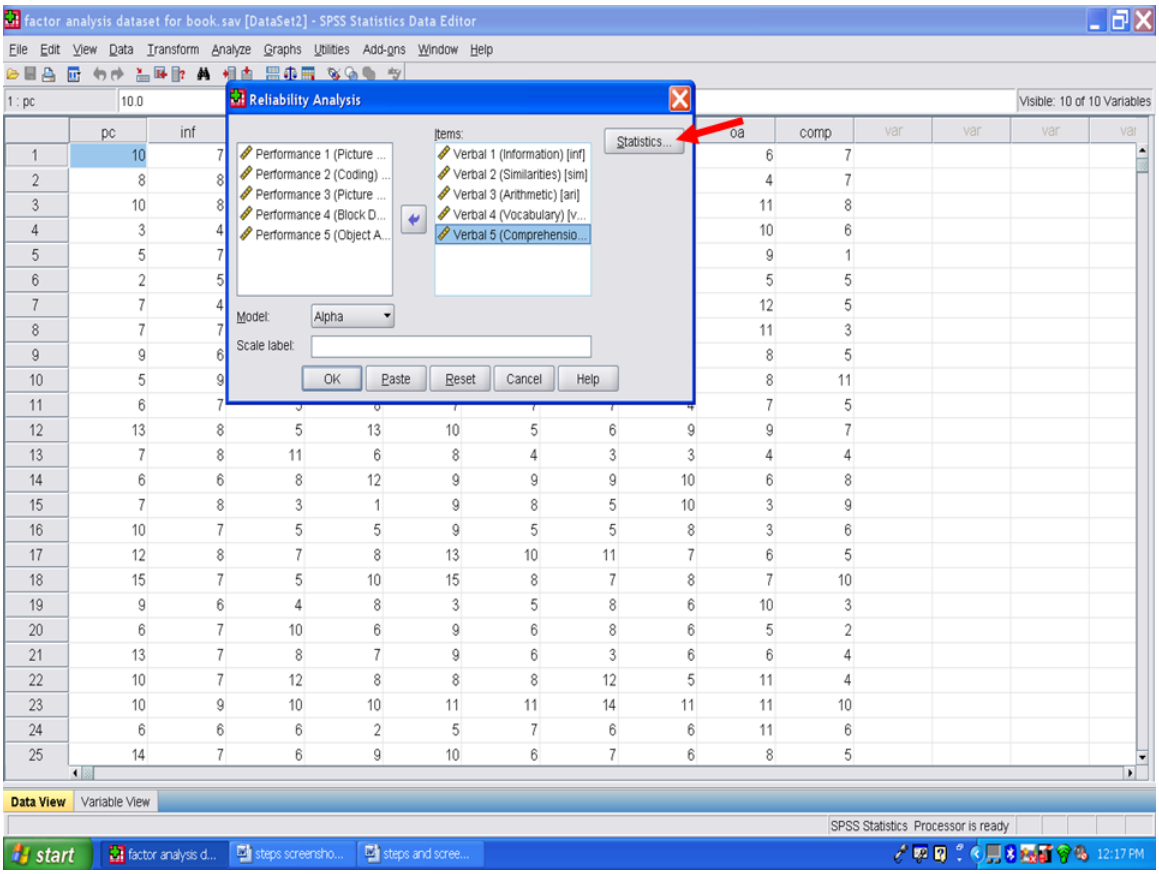
[8] http://cnx.org/content/m40717/latest/5.2.png/image



9

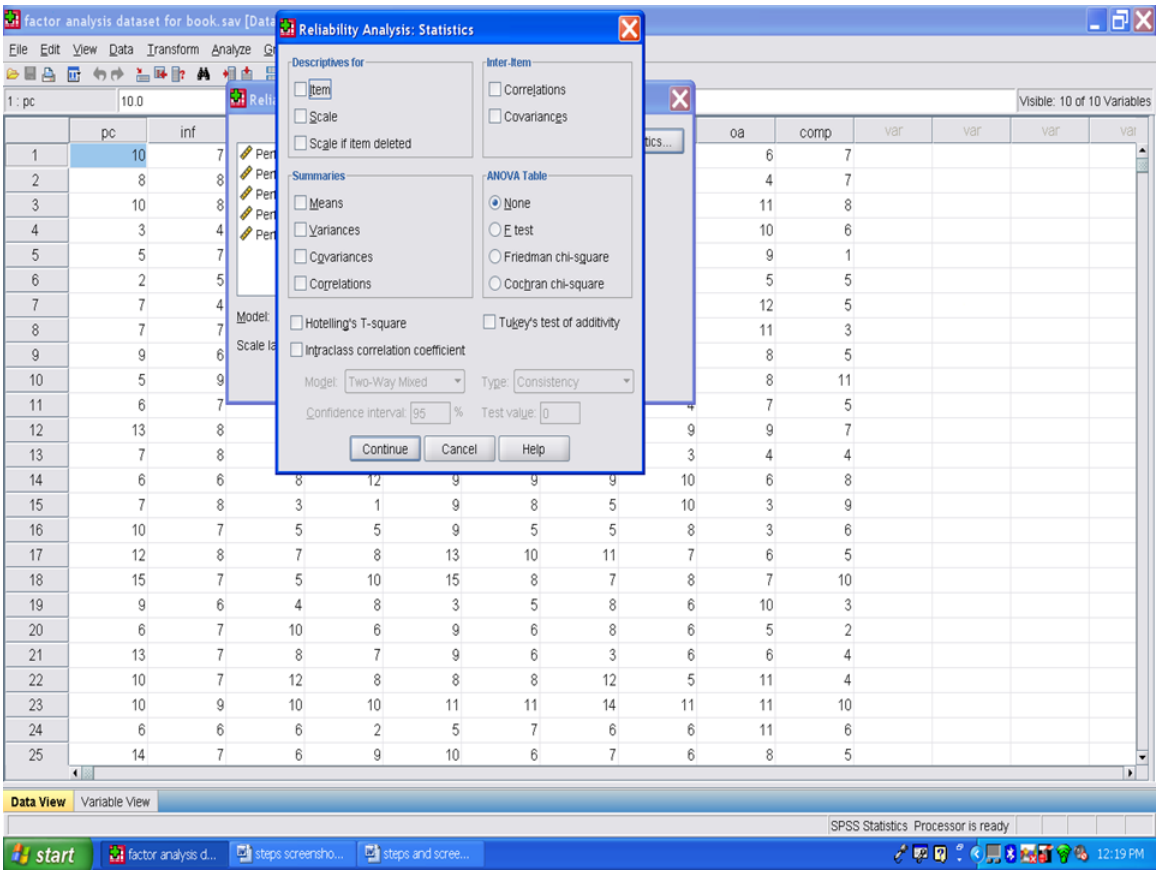
Now your screen should look like the one below.
Click on Statistics.

[9] http://cnx.org/content/m40717/latest/5.3.png/image



After clicking on Statistics, the following screen should appear.

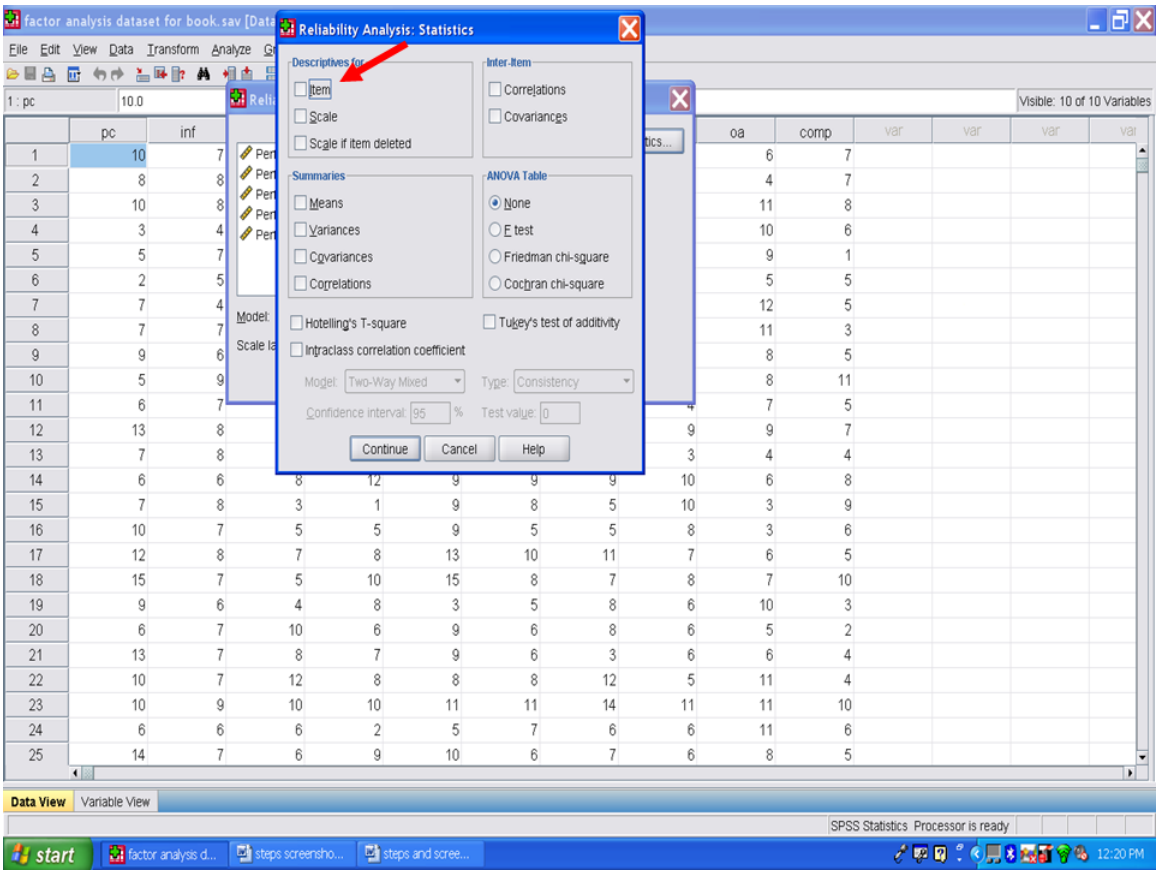
[10] http://cnx.org/content/m40717/latest/5.4.png/image



11

Click under Descriptives on
Item
Scale
Scale if item deleted

[11] http://cnx.org/content/m40717/latest/5.5.png/image

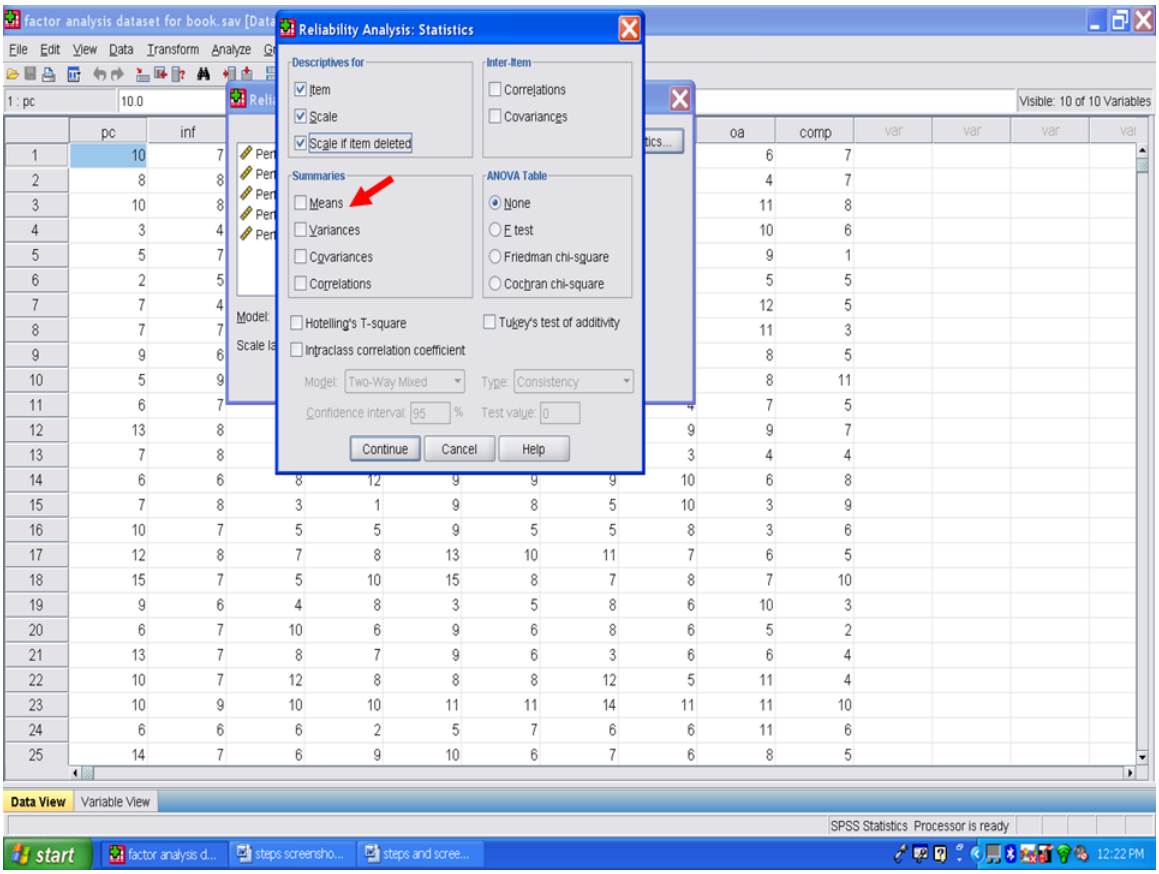


12

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

Means
Variances

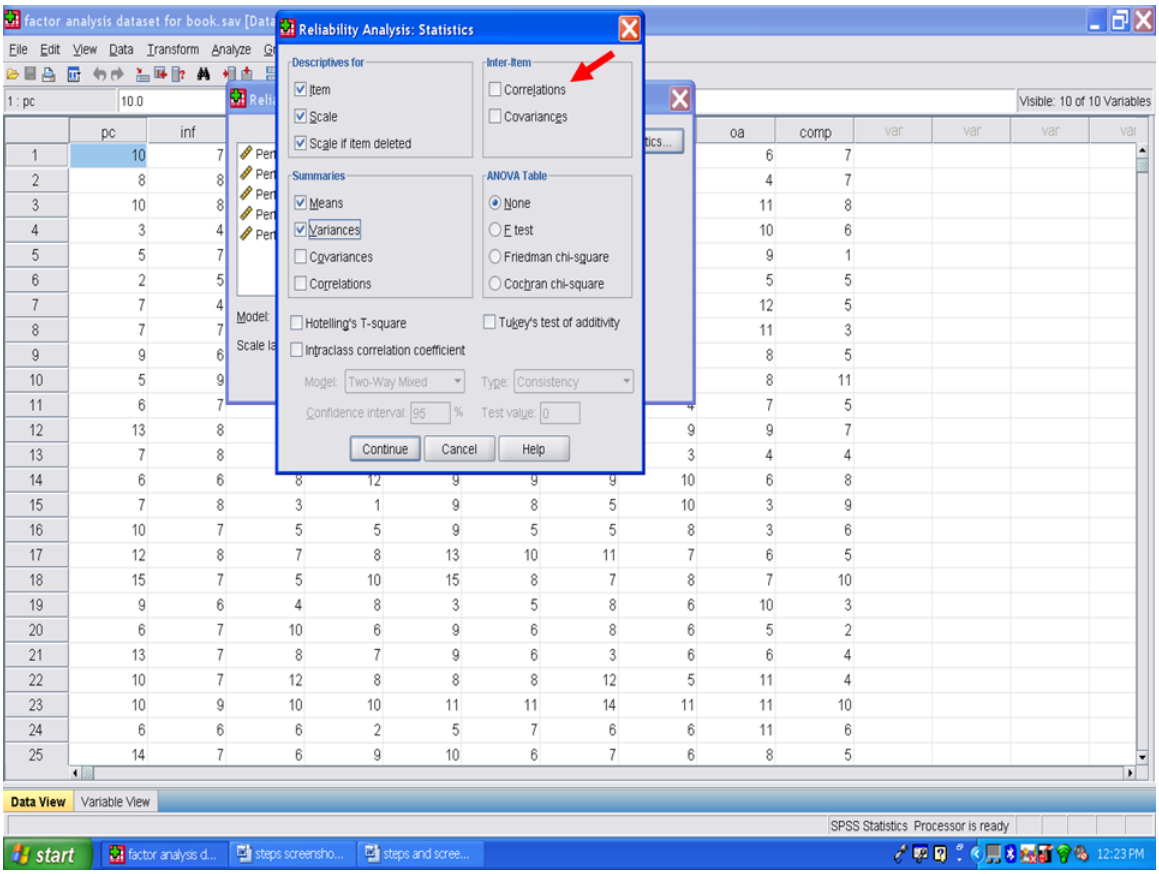
[12] http://cnx.org/content/m40717/latest/5.6.png/image



13

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

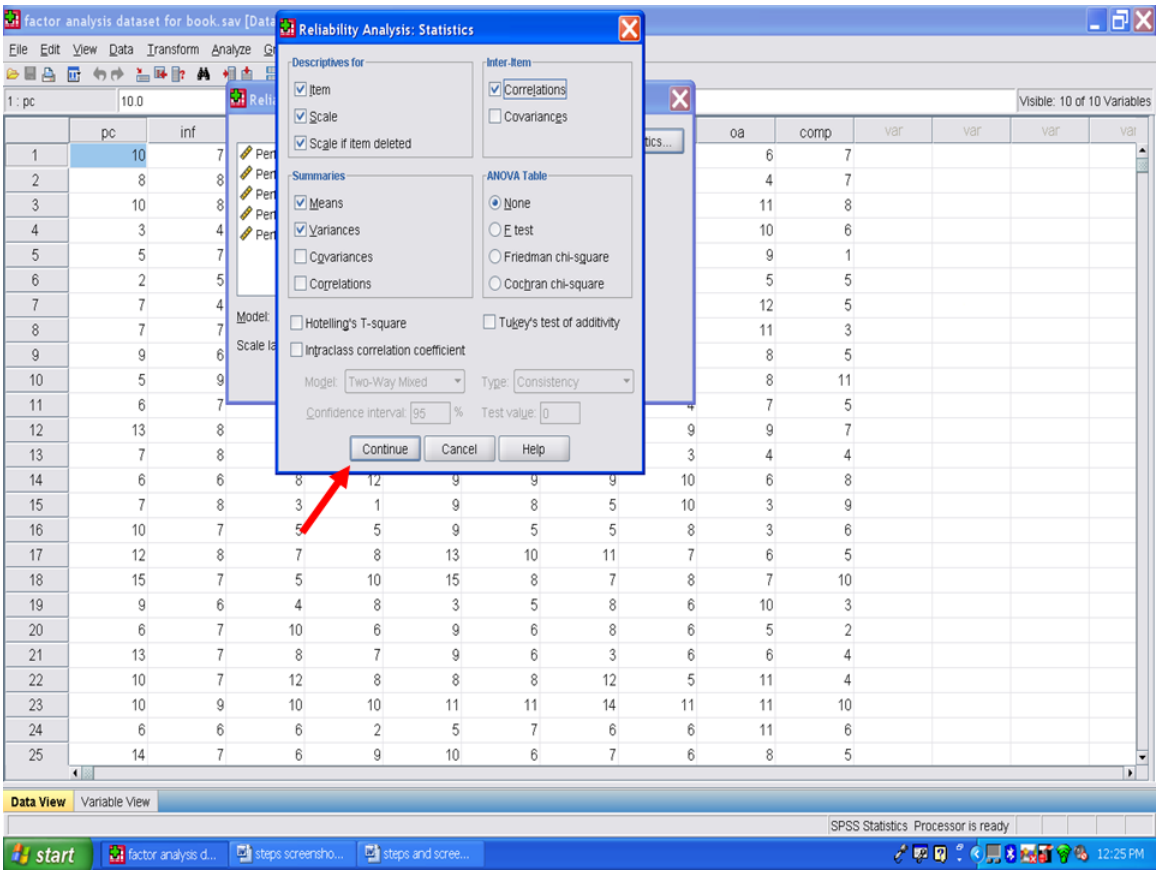
[13] http://cnx.org/content/m40717/latest/5.7.png/image



14

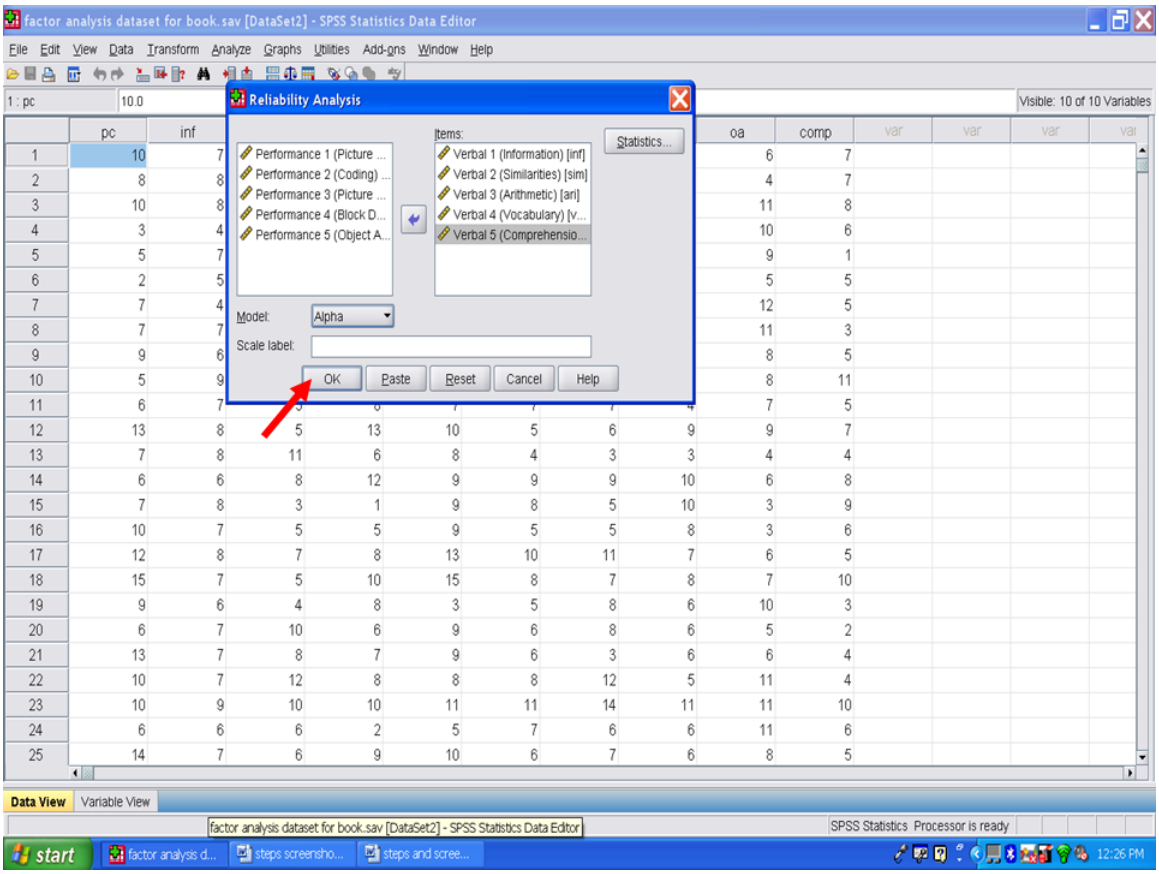
After you check to verify that your screen looks the one below, then you may click on Continue.

[14] http://cnx.org/content/m40717/latest/5.8.png/image



Next click on OK.

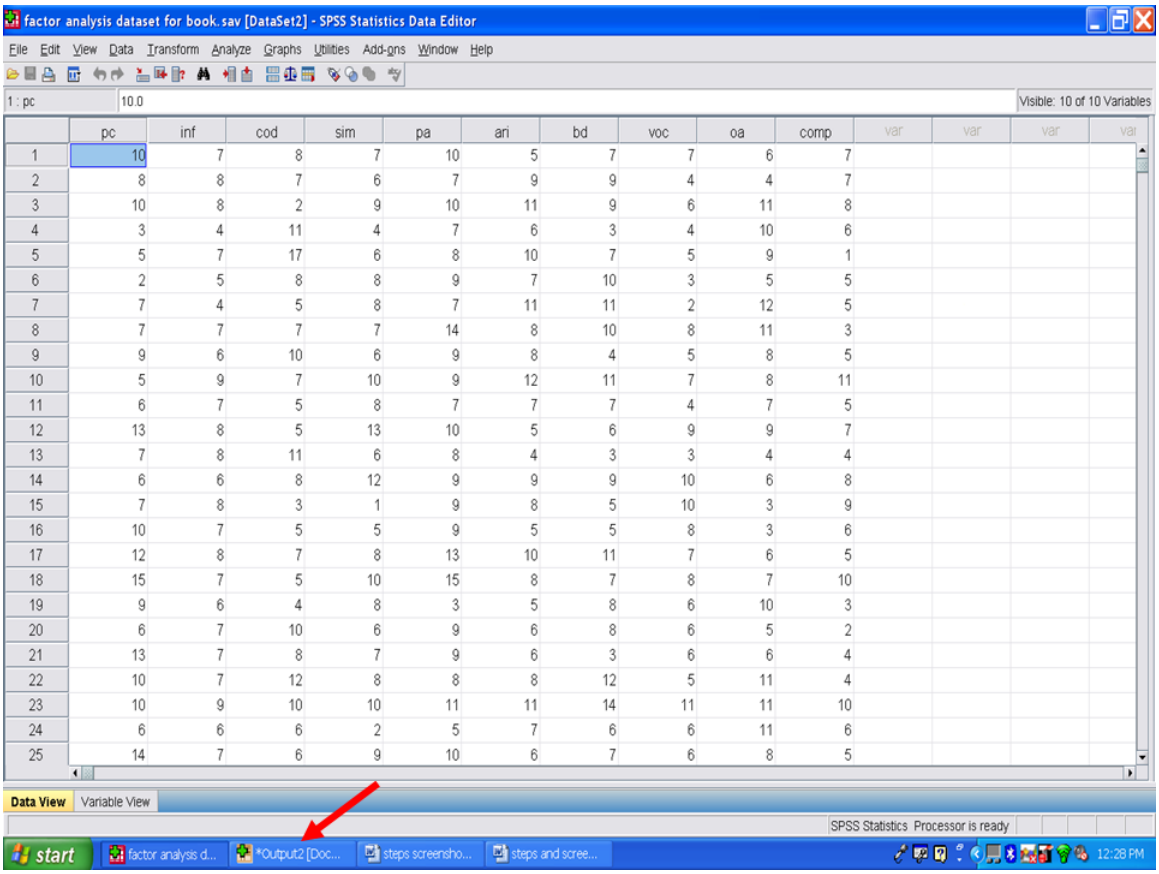
[15] http://cnx.org/content/m40717/latest/5.9.png/image



[16]

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



	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				

[17]

[17] http://cnx.org/content/m40717/latest/5.11.png/image

Chapter 7

7. Internal Consistency Analysis: Part II[1]

NOTE: This chapter is published by NCPEA Press[2] 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.

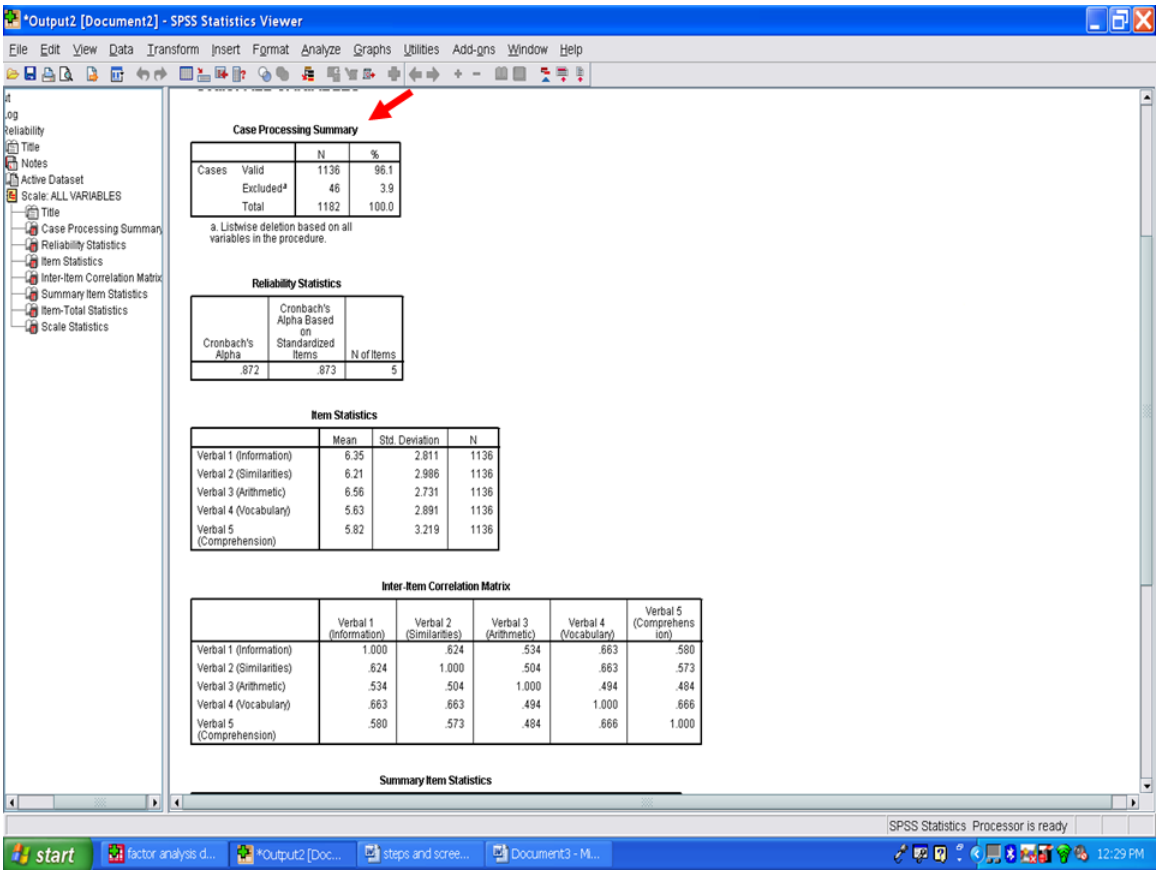
John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, 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 Help[3]) 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 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.

[1] This 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

		N	%
Cases	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

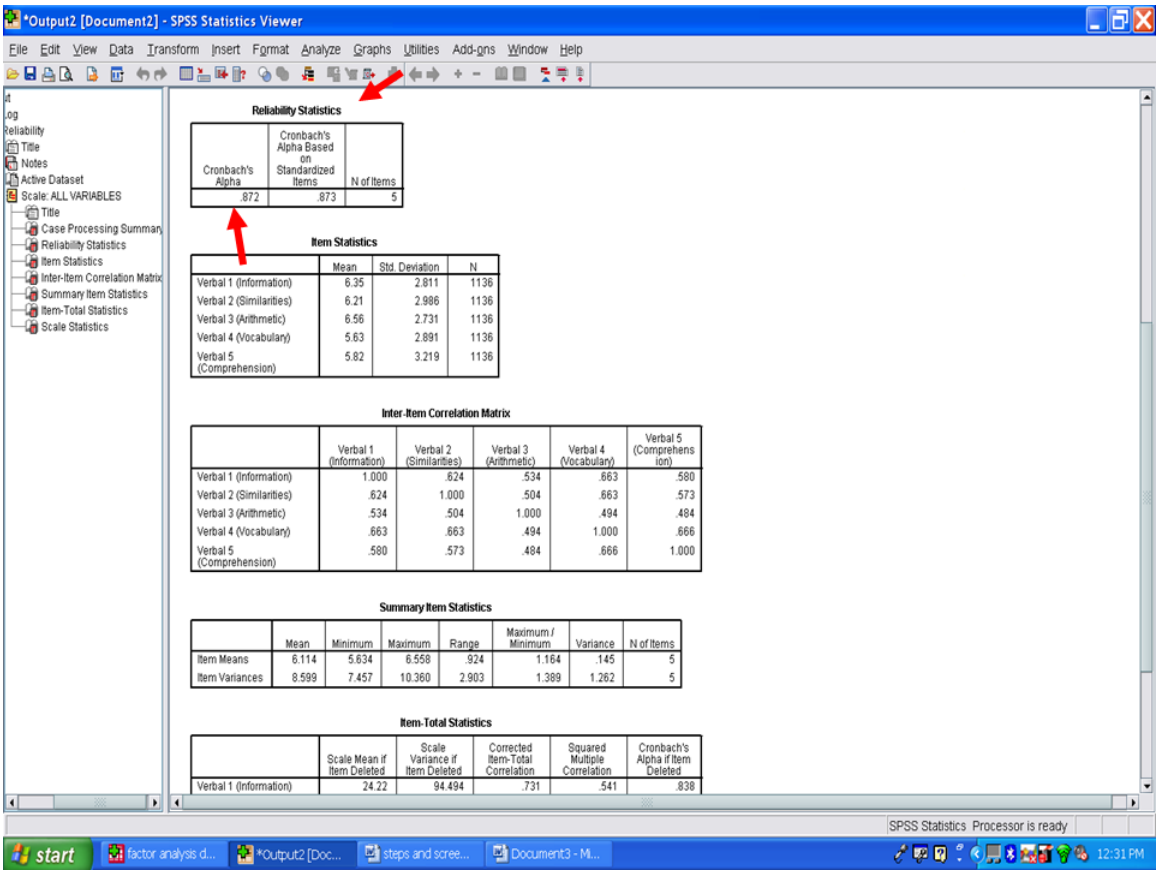
Inter-Item Correlation Matrix

	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 Alpa is .872 for these five Verbal subscales.

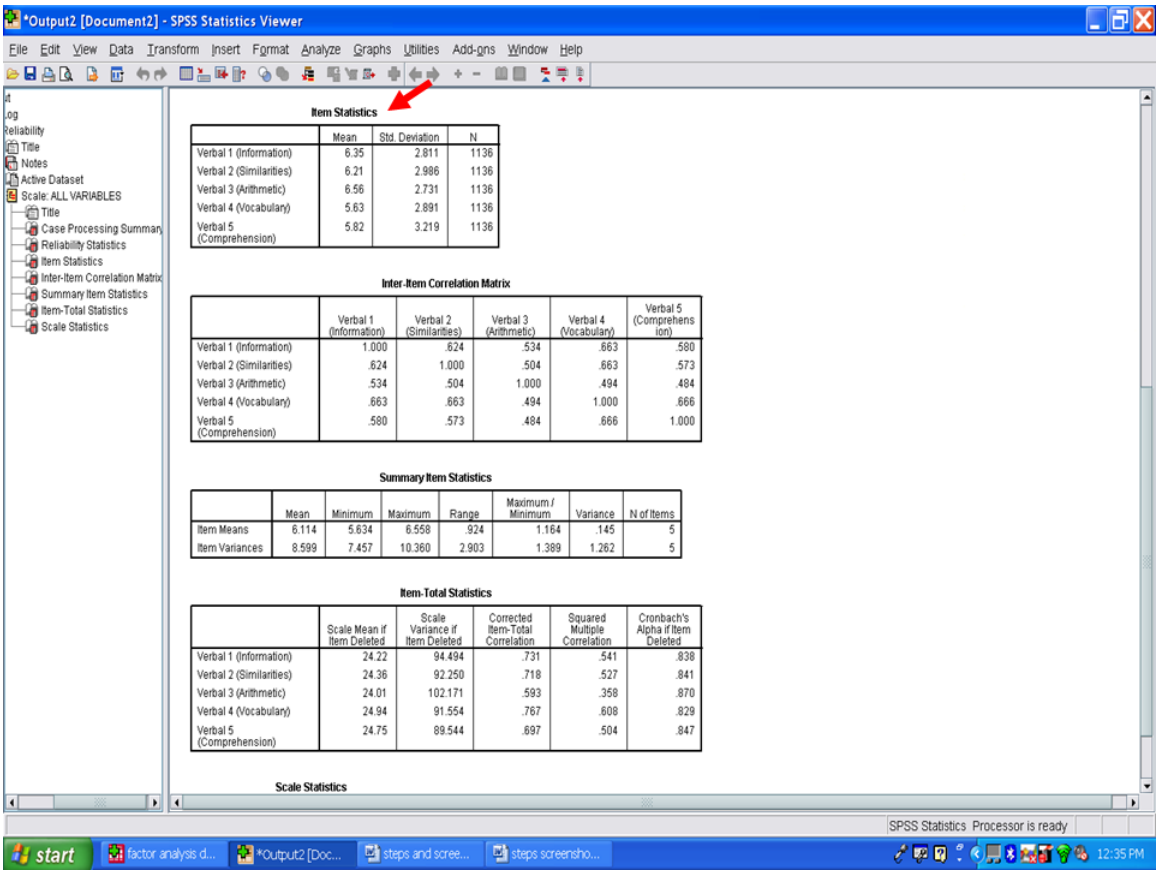
[4] http://cnx.org/content/m40718/latest/6.1.png/image



[5]

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.

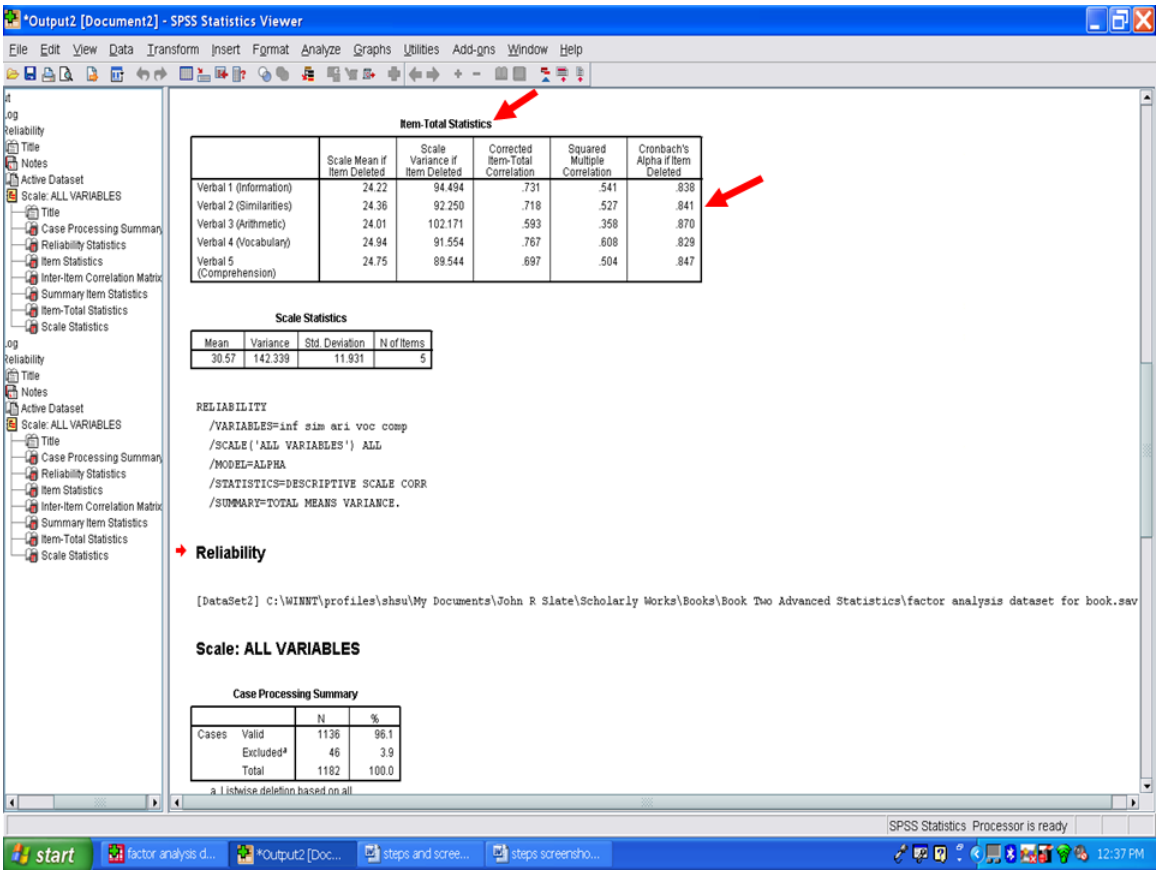
[5] http://cnx.org/content/m40718/latest/6.2.png/image



[6]

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.

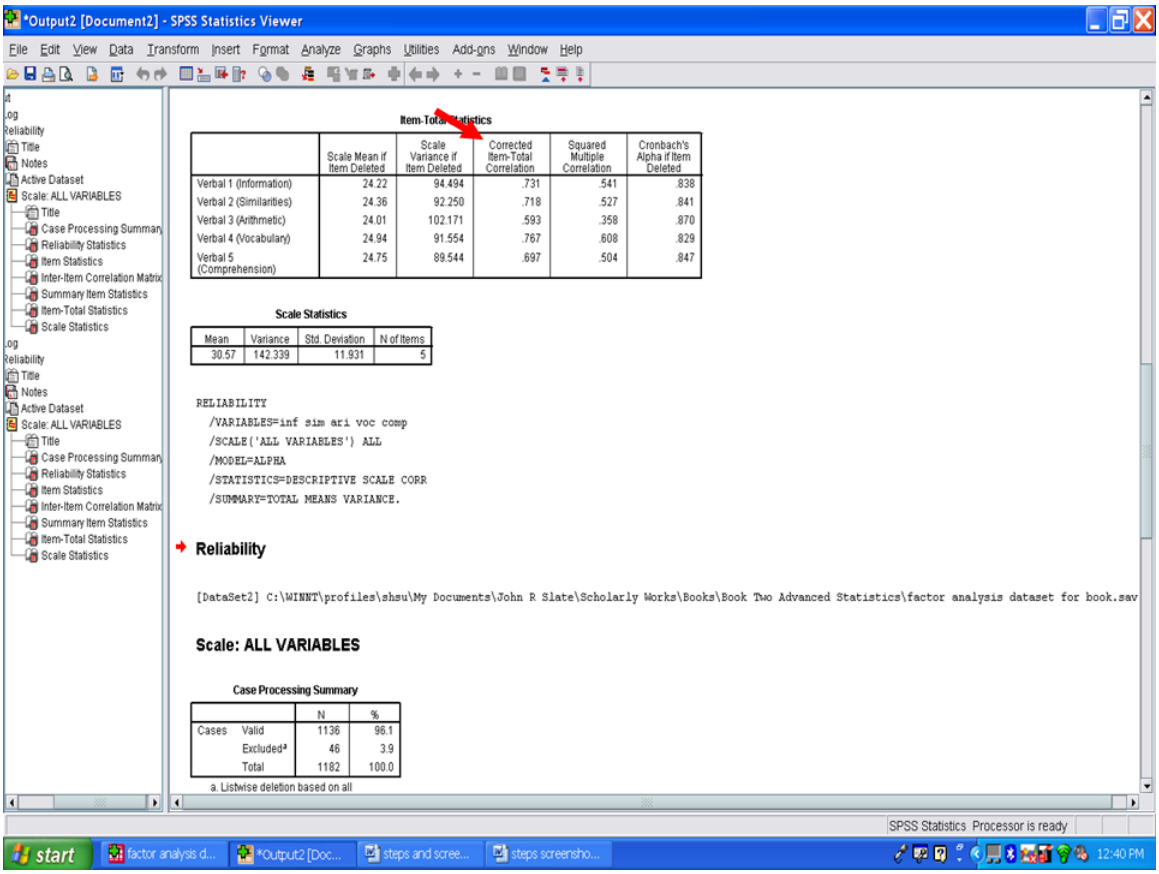
[6] http://cnx.org/content/m40718/latest/6.3.png/image



7

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



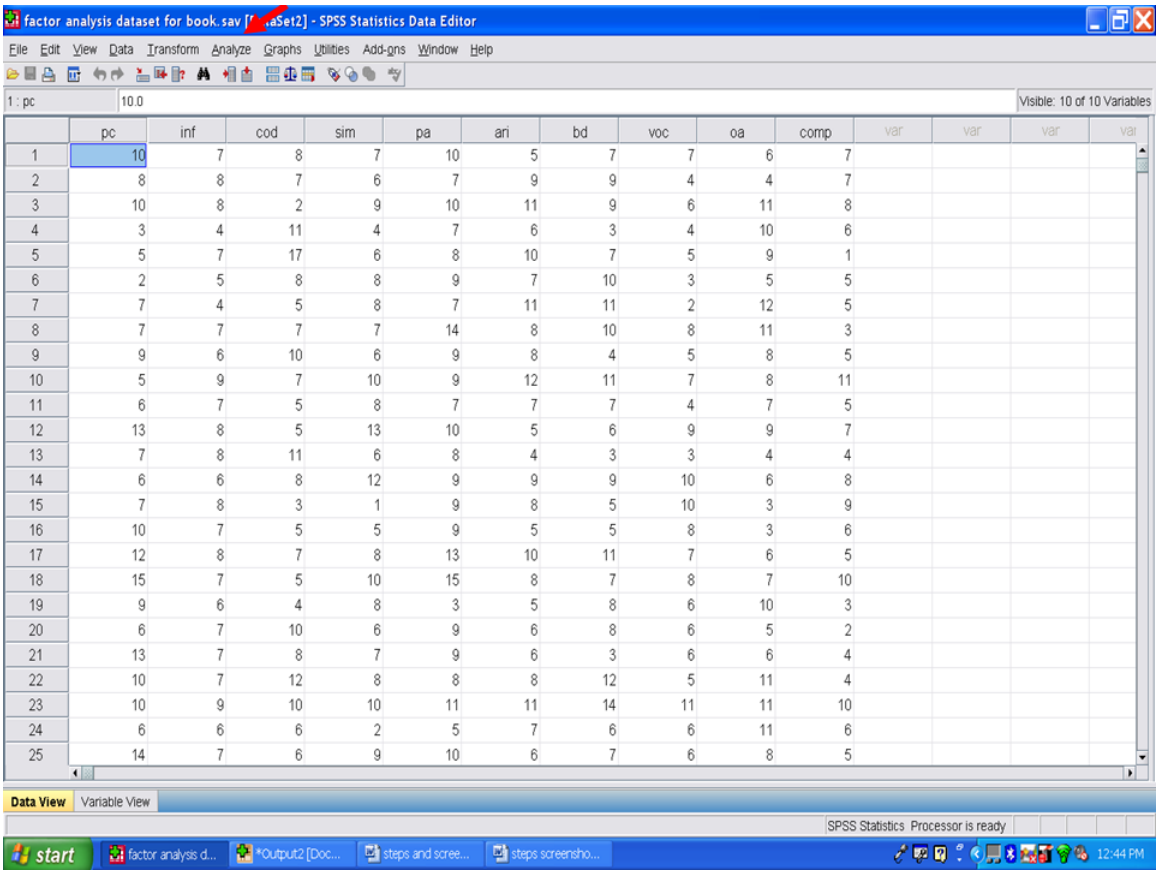
8

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

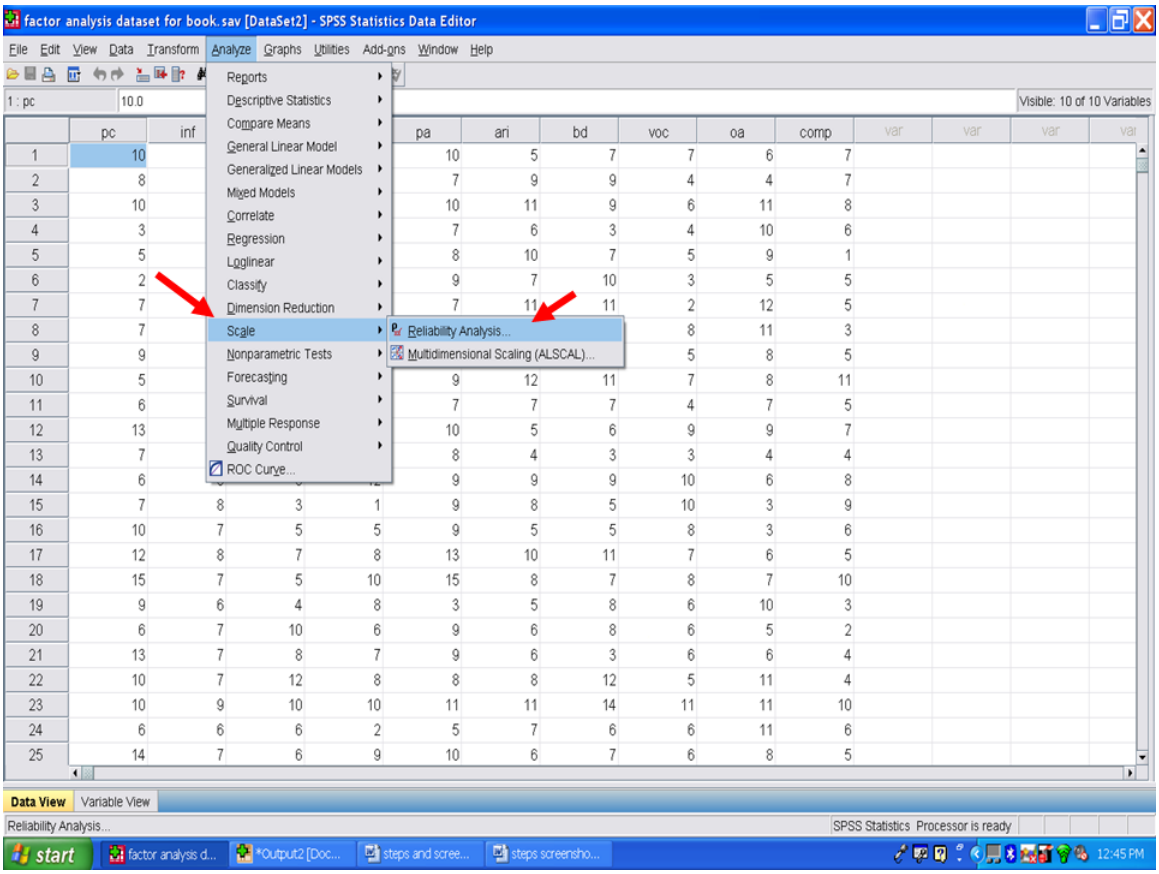
[8] http://cnx.org/content/m40718/latest/6.5.png/image



9

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

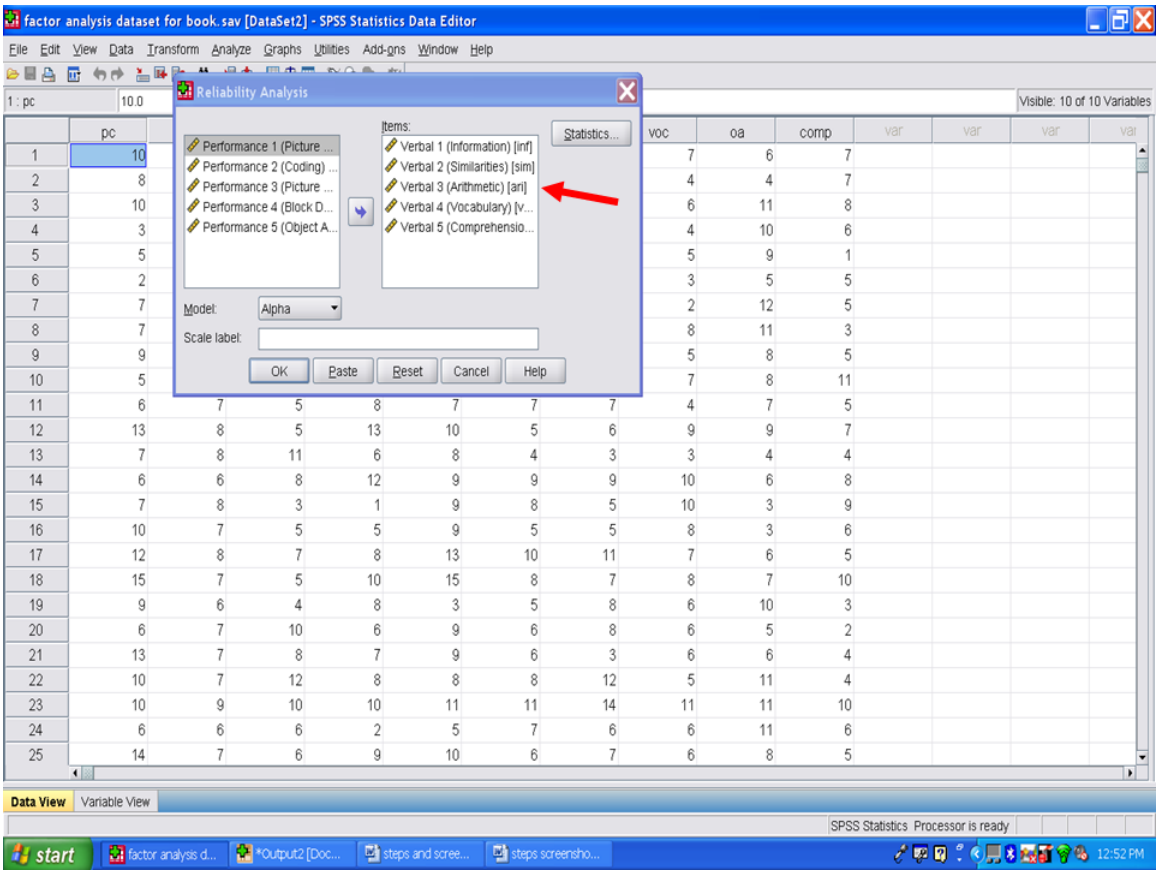
[9] http://cnx.org/content/m40718/latest/6.6.png/image



[10]

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.

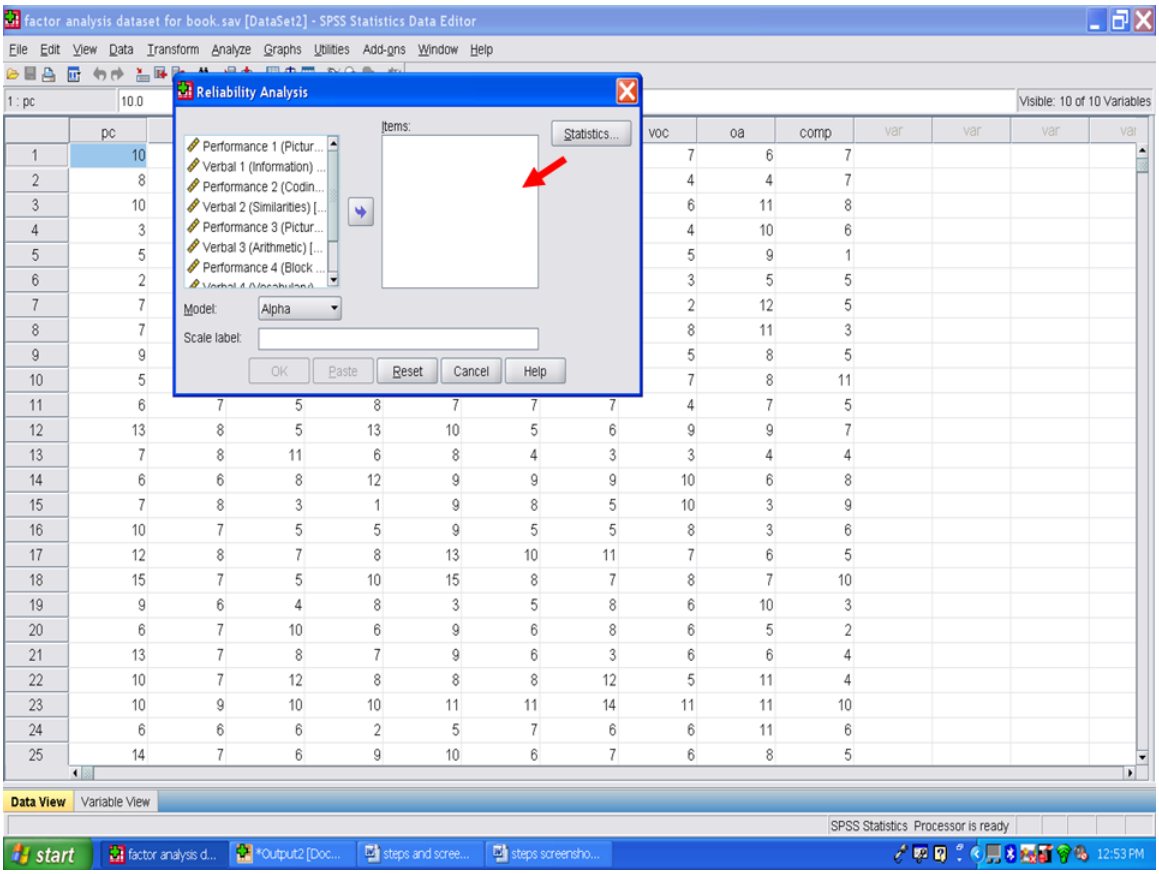
[10] http://cnx.org/content/m40718/latest/6.7.png/image



11

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]

[12] http://cnx.org/content/m40718/latest/6.9.png/image

Chapter 8

8. Internal Consistency Analysis: Part III[1]

NOTE: This chapter is published by NCPEA Press[2] 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.

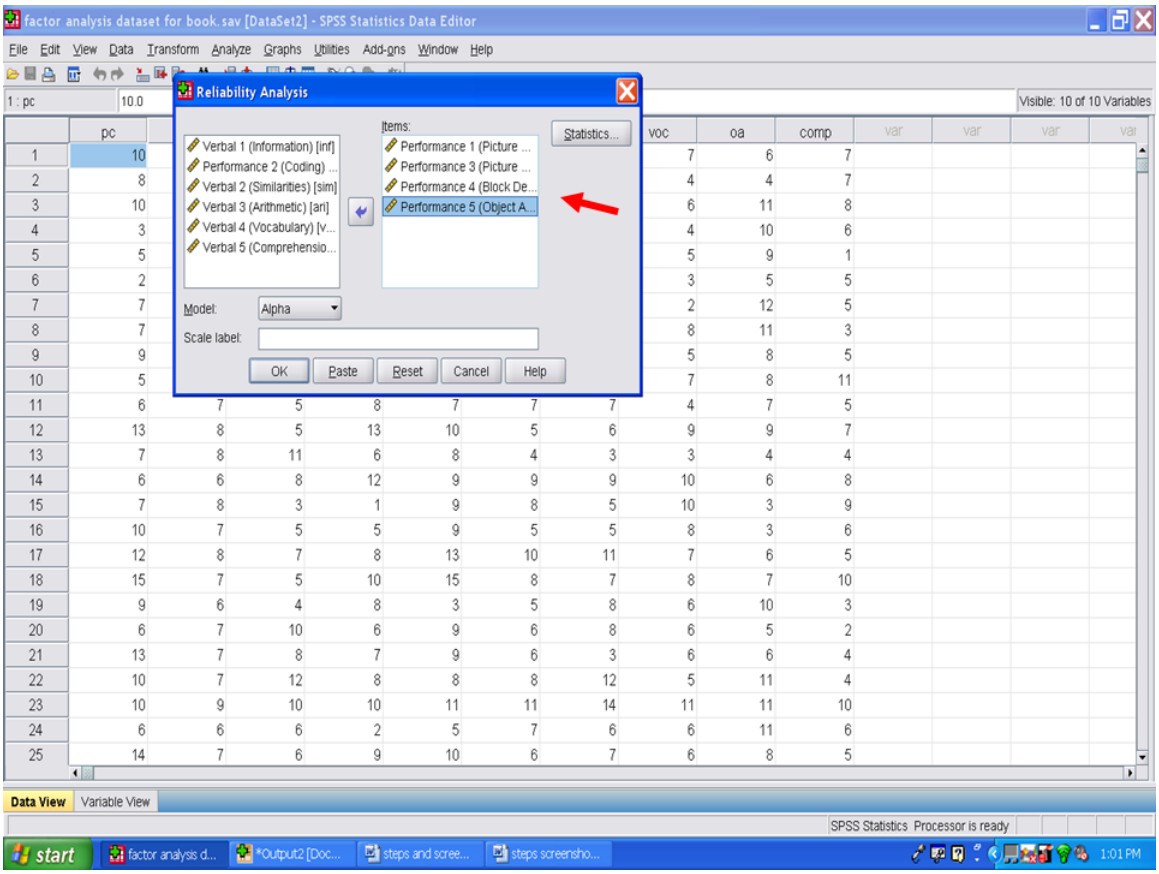
About the Authors

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 Help[3]) 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.

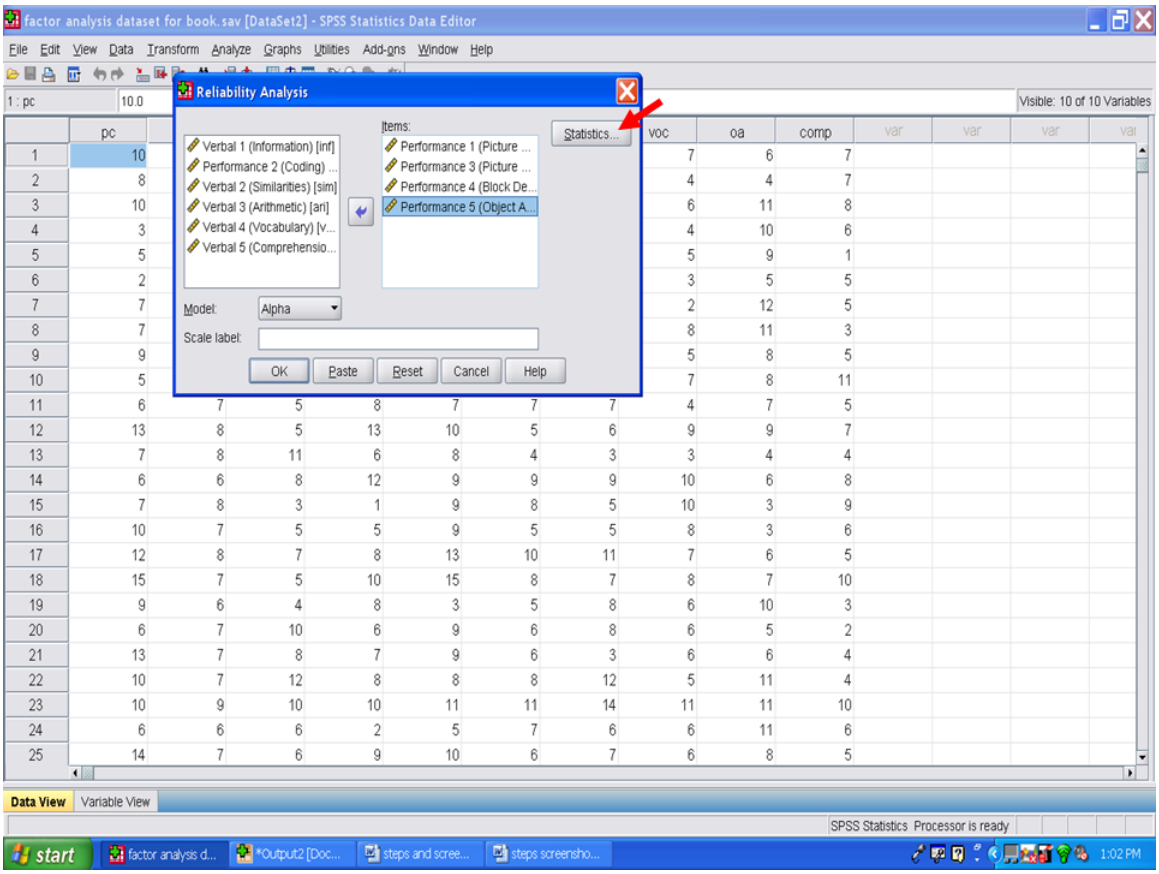
Once the four Performance subscales have been sent to the Items box, your screen would look like the following one:

[1] This 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

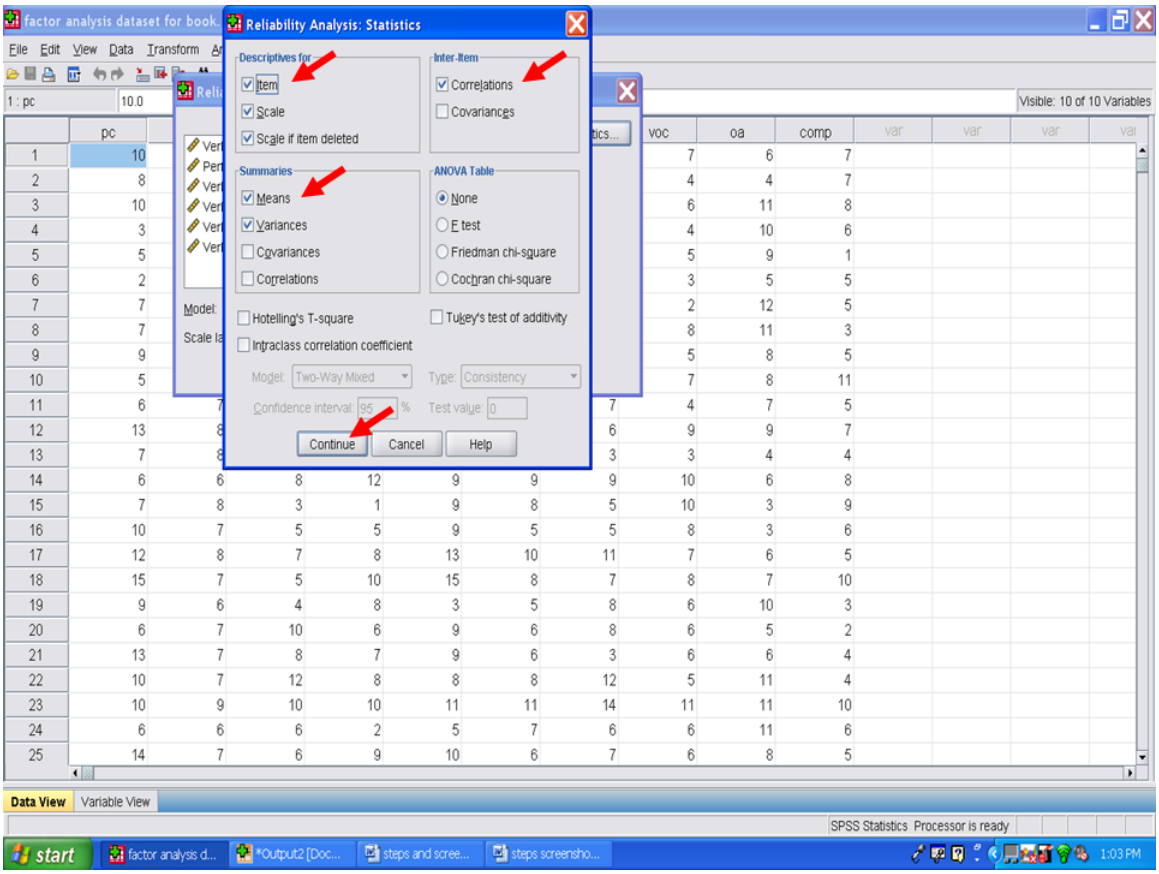


[5]

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

Click on Continue.

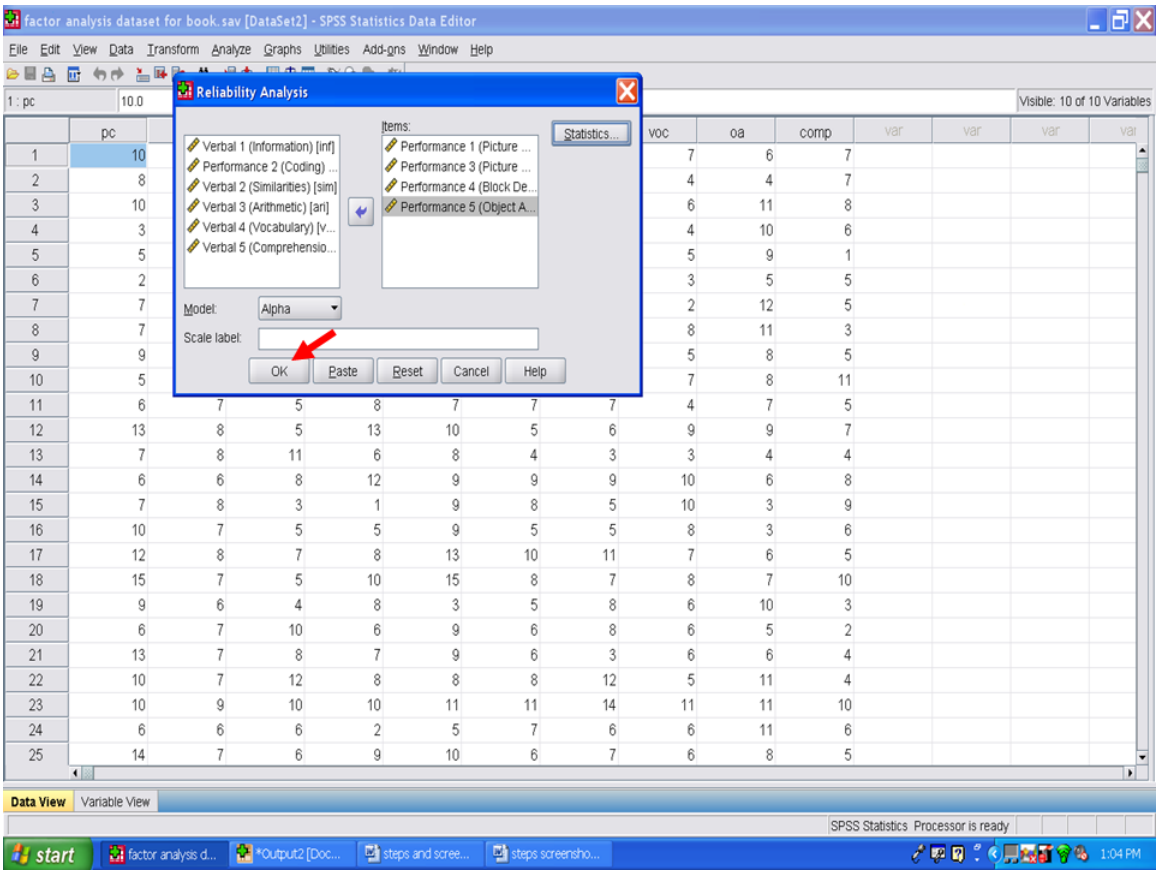
[5] http://cnx.org/content/m40721/latest/7.2.png/image



6

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

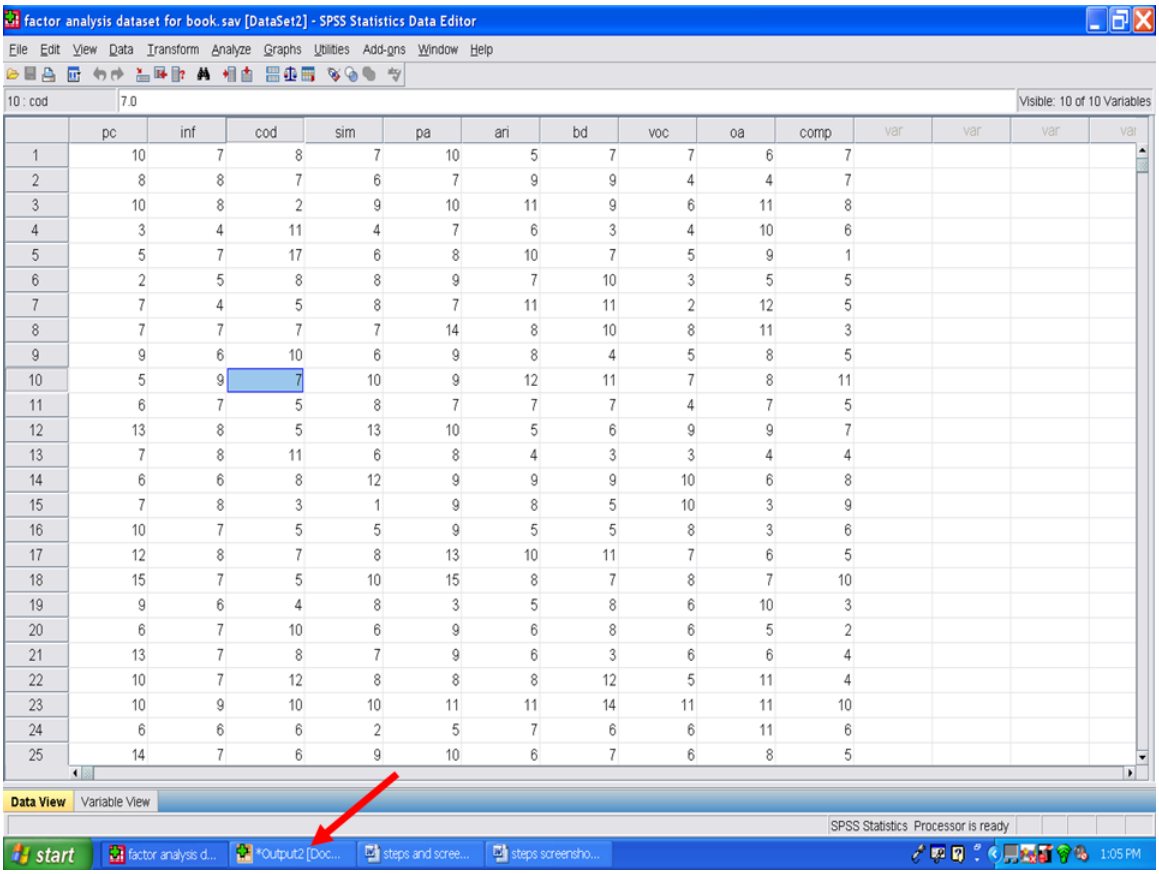
[6] http://cnx.org/content/m40721/latest/7.3.png/image



7

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

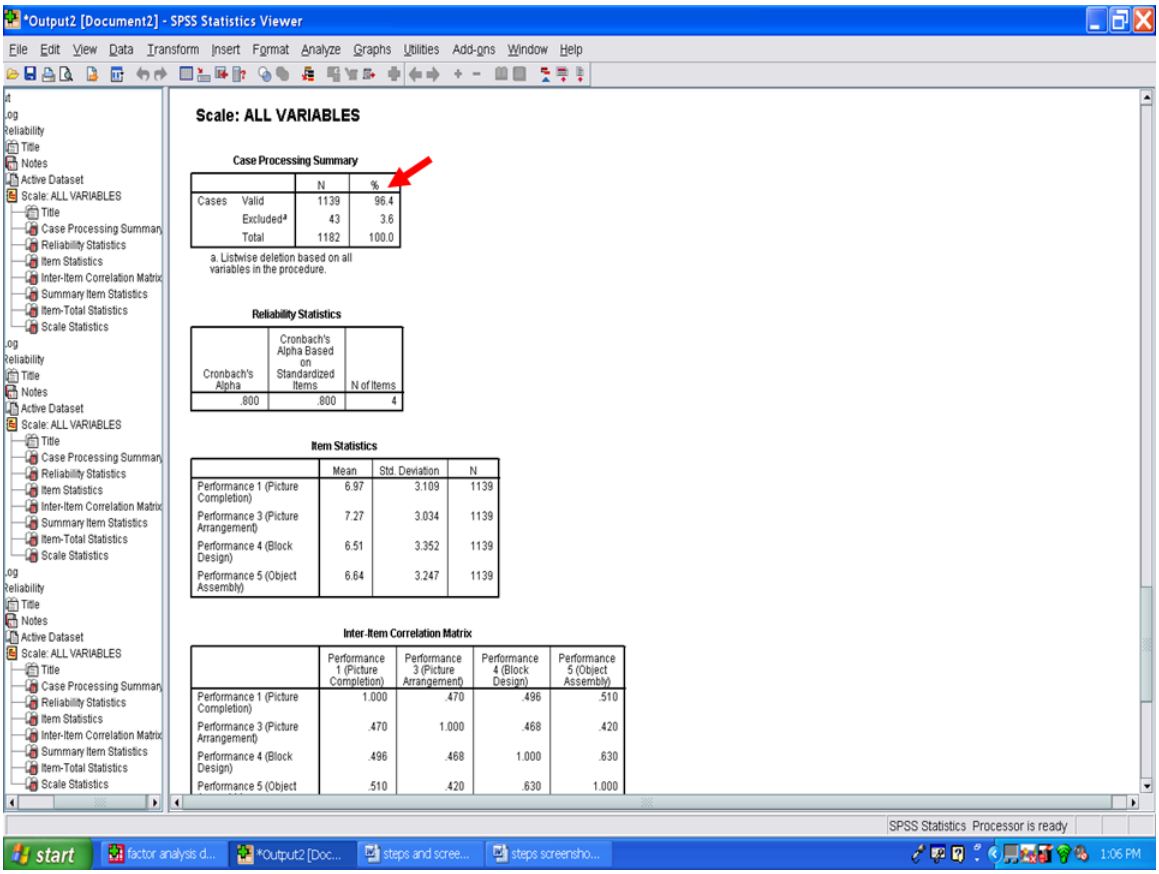
[7] http://cnx.org/content/m40721/latest/7.4.png/image



[8]

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.

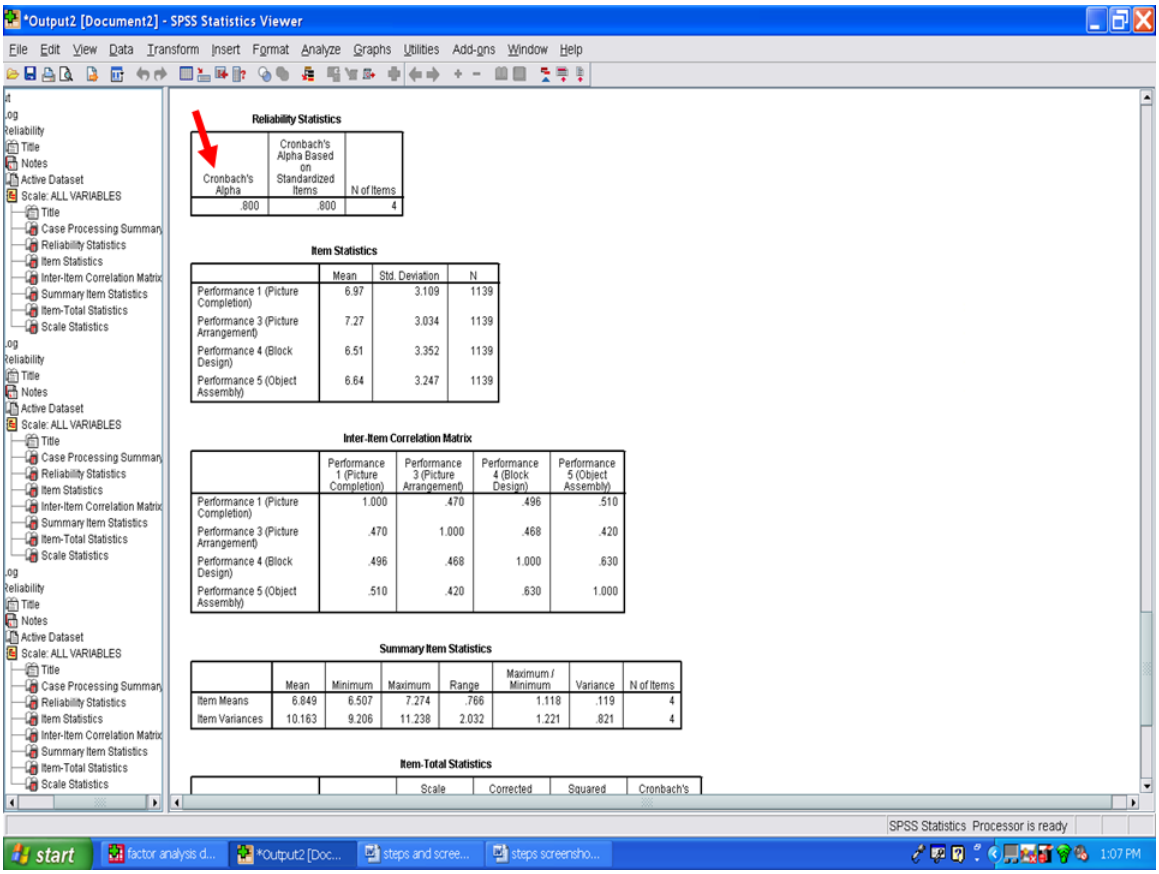
[8] http://cnx.org/content/m40721/latest/7.5.png/image



[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.

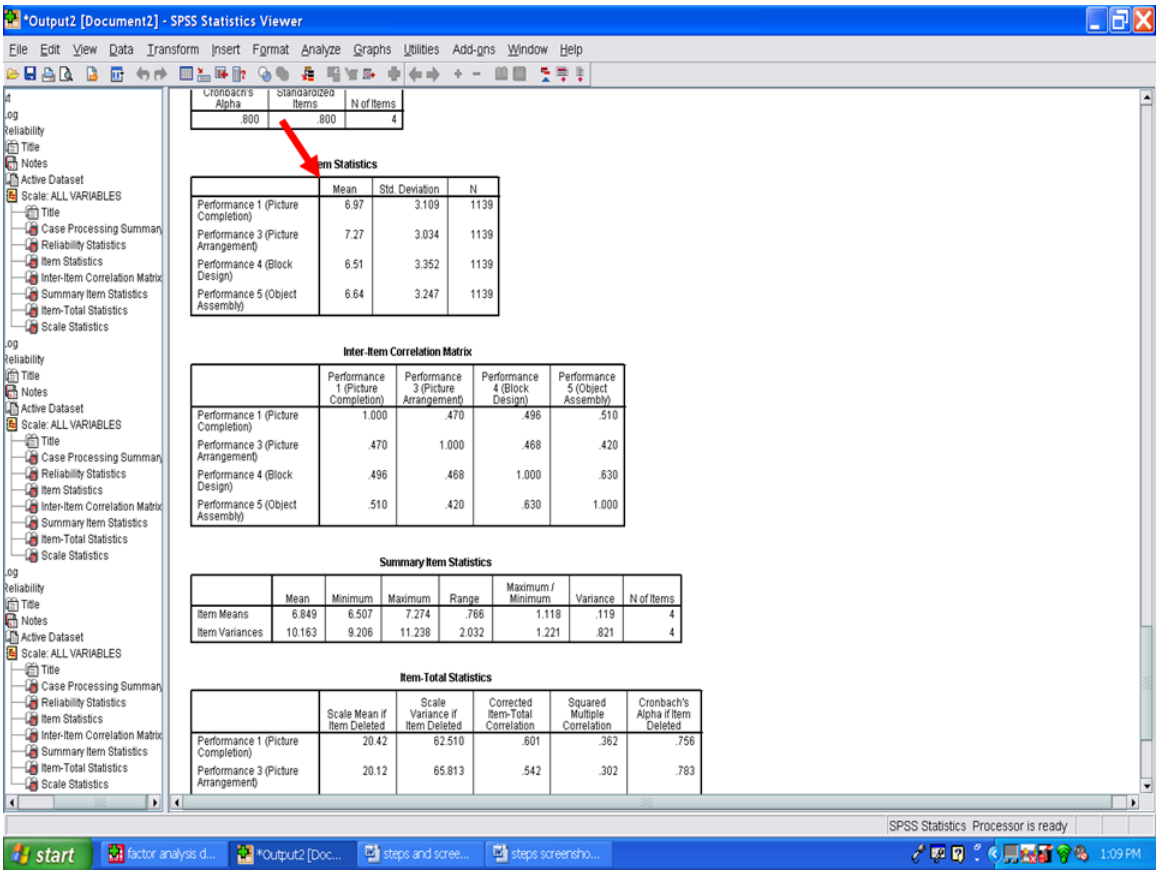
[9] http://cnx.org/content/m40721/latest/7.6.png/image



10

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

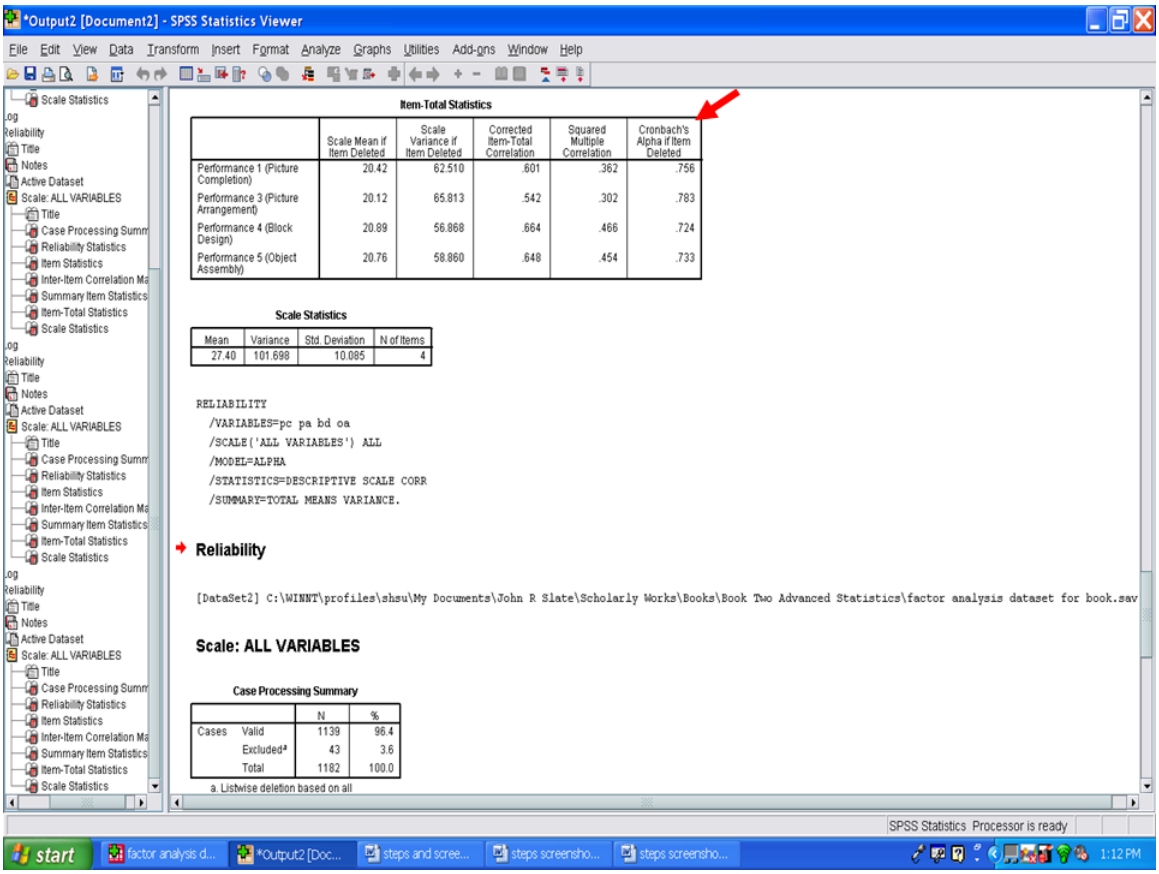
[10] http://cnx.org/content/m40721/latest/7.7.png/image



[11]

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.

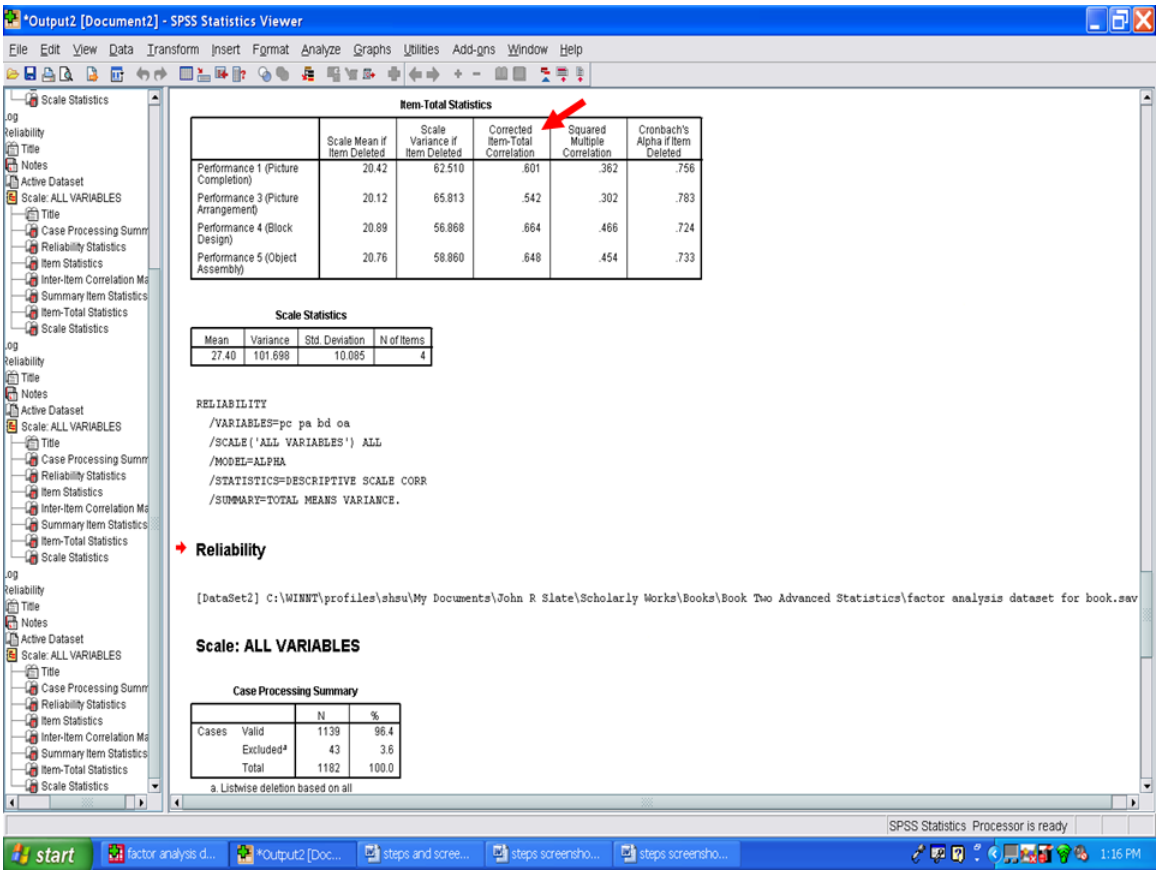
[11] http://cnx.org/content/m40721/latest/7.8.png/image



12

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



[13]

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 I[1]

NOTE: This chapter is published by NCPEA Press[2] 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 Help[3]) 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.

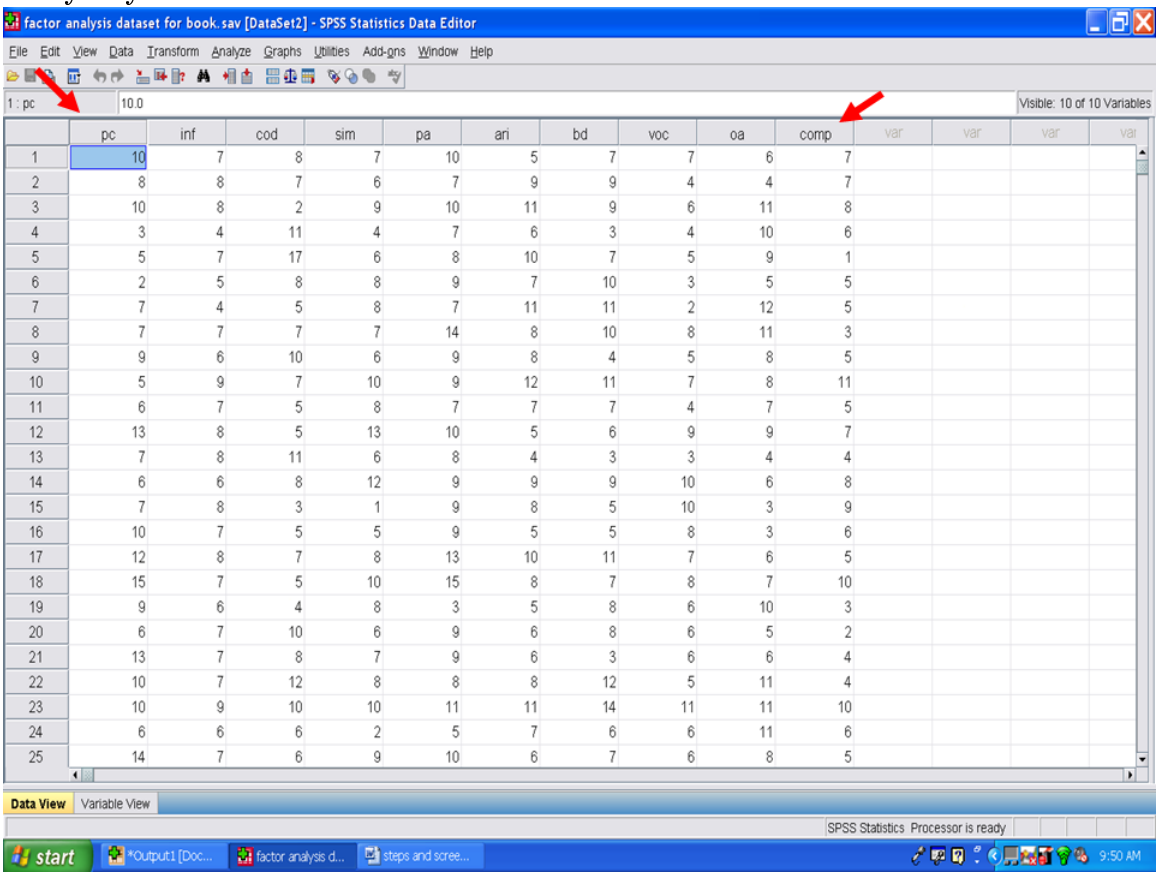
In this set of steps, readers will conduct a factor analysis procedure, in particular, a Varimax factor analysis. For detailed information regarding the assumptions underlying use of a factor analysis and reasons for using a Varimax factor analysis, readers are referred to the Hyperstats Online Statistics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

[1] This 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=1-1[6]

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.



[7]

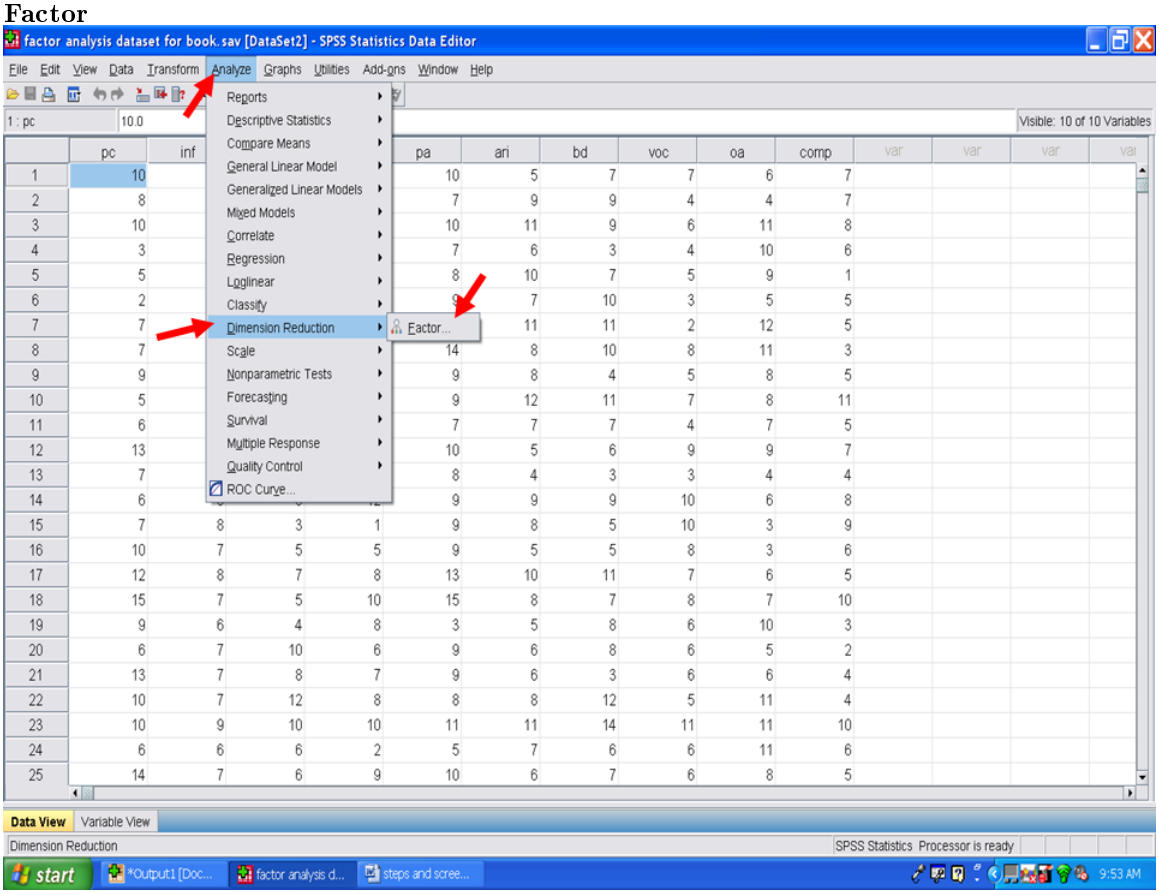
In this example, we will click on
Analyze
Dimension Reduction

[5] http://www.statsoft.com/textbook/

[6] http://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

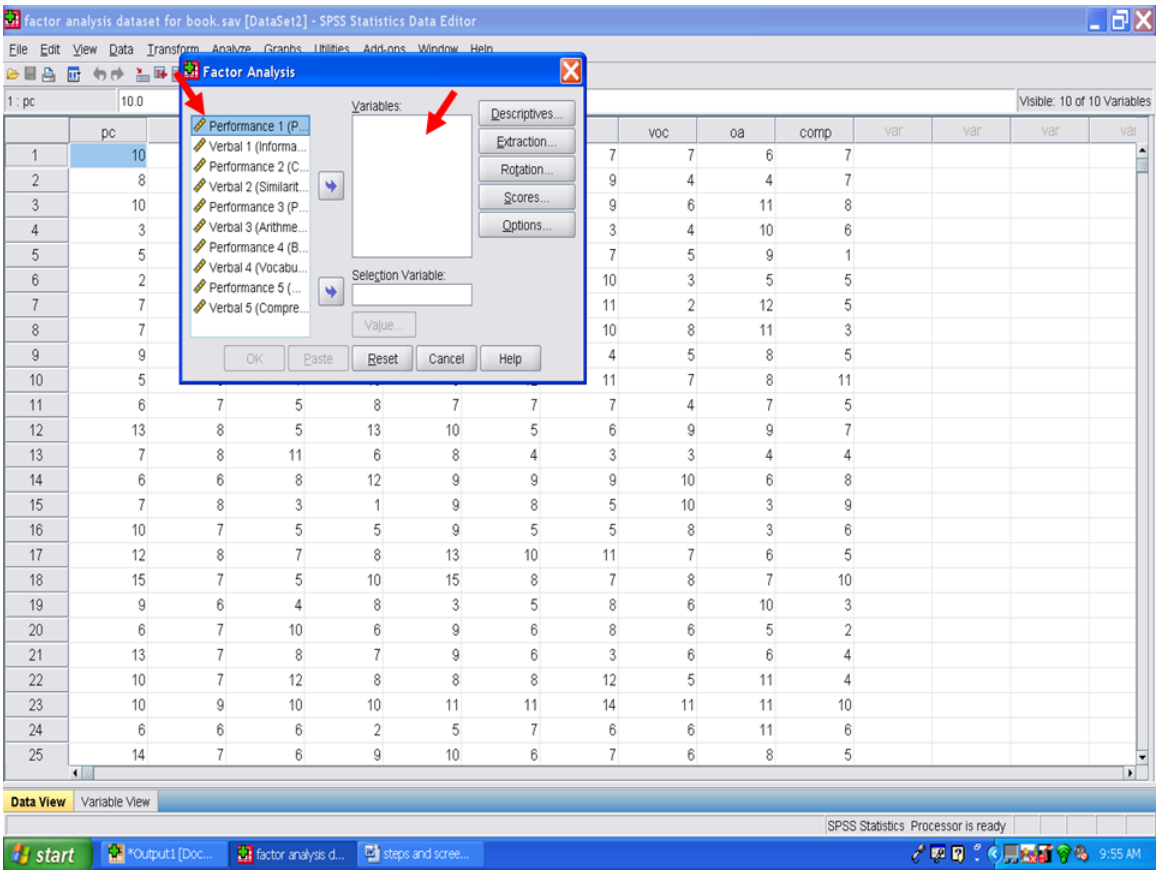
Factor



[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.

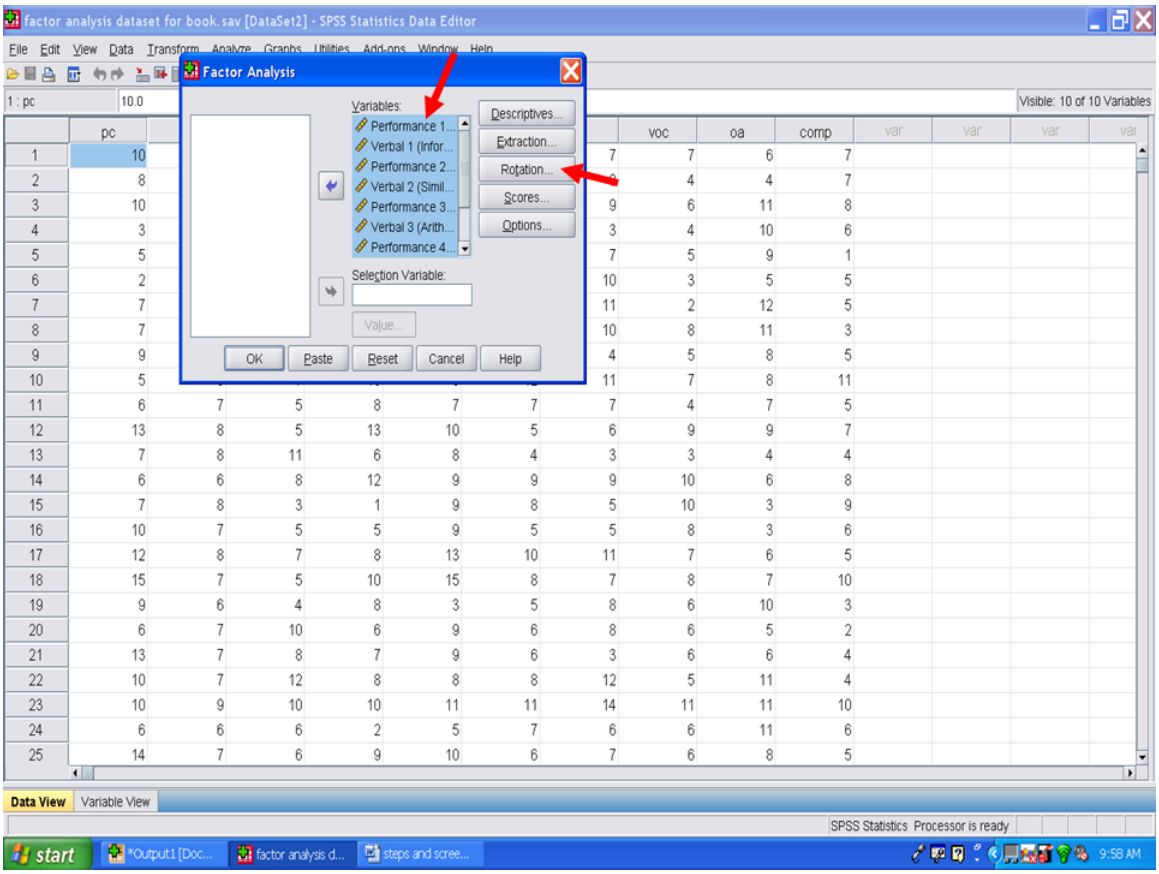
[8] http://cnx.org/content/m40725/latest/8.2.png/image



[9]

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

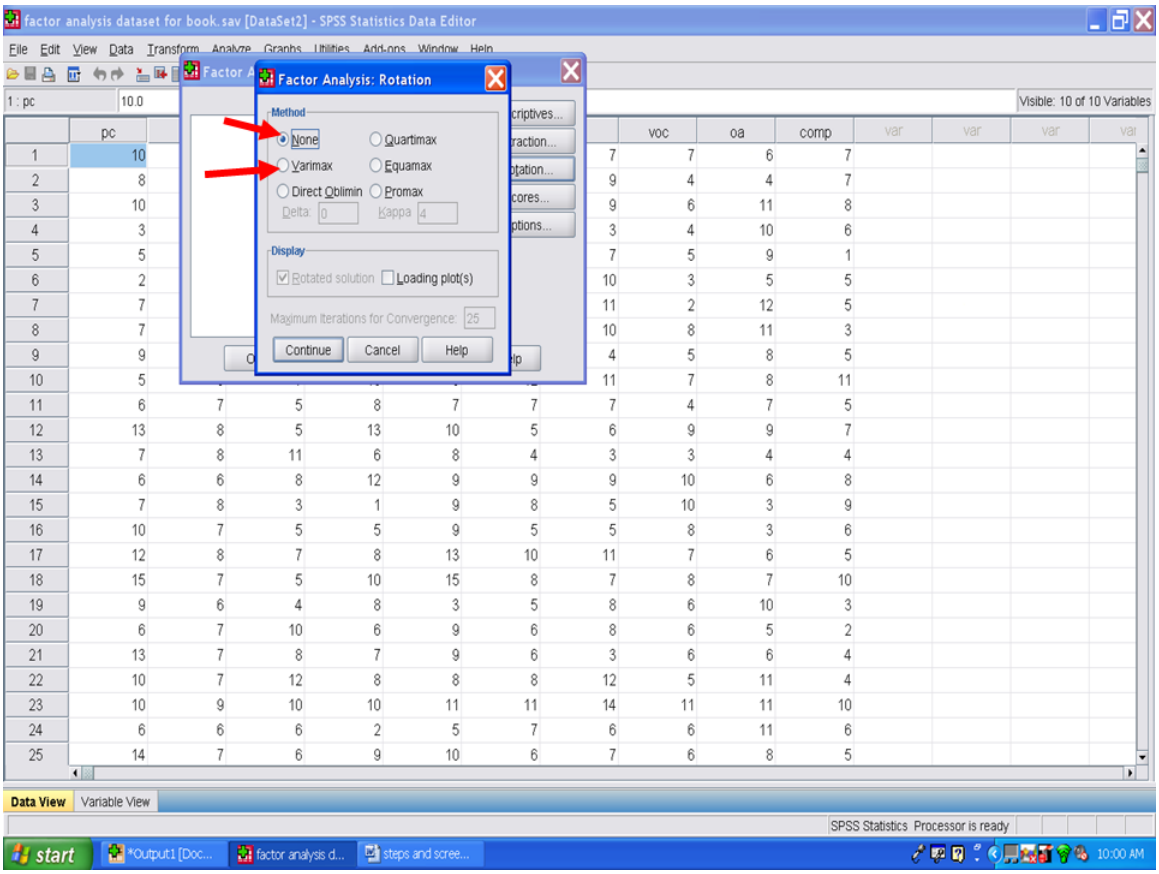
[9] http://cnx.org/content/m40725/latest/8.3.png/image



10

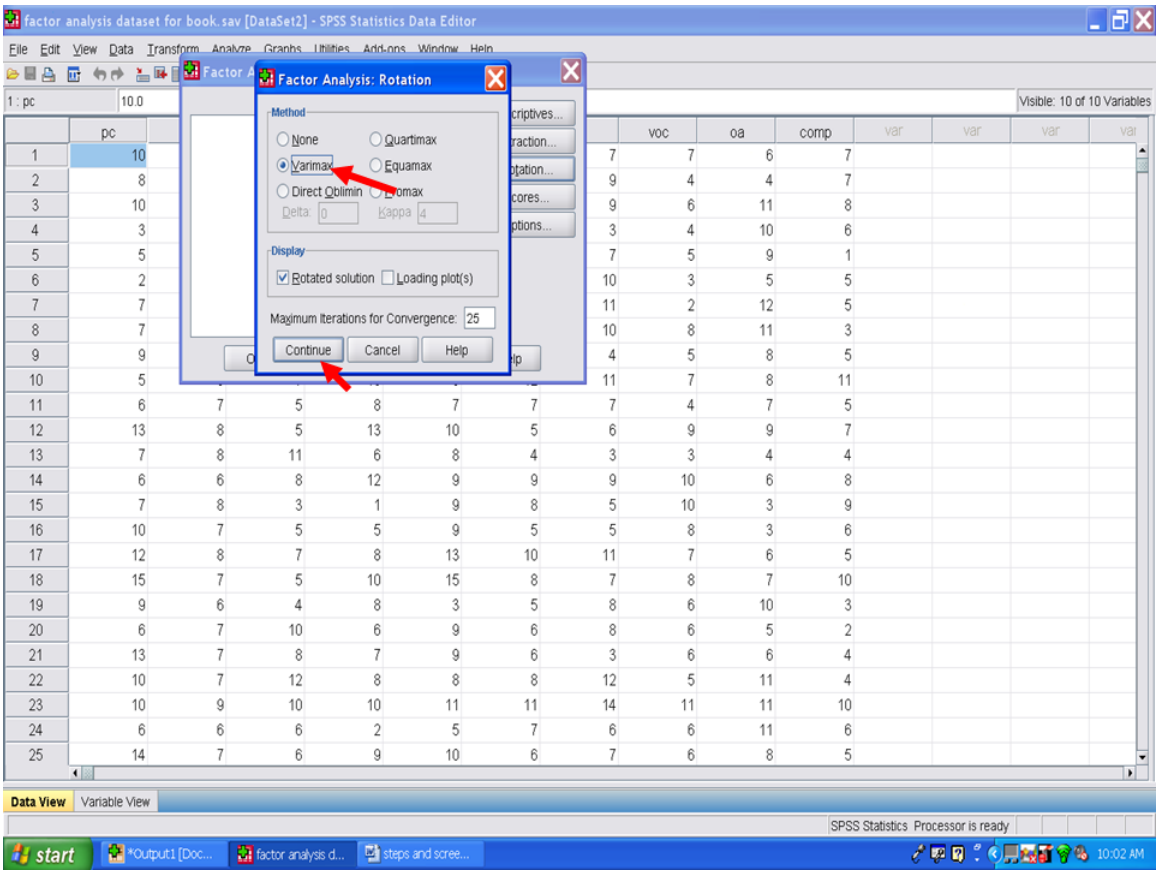
After clicking on Rotation, the following screen will appear. Under Method, the box for None is already checked because it is the SPSS default value. We will conduct a Varimax factor analysis. Click on Varimax.

[10] http://cnx.org/content/m40725/latest/8.4.png/image



Varimax has been clicked below. Click on Continue.

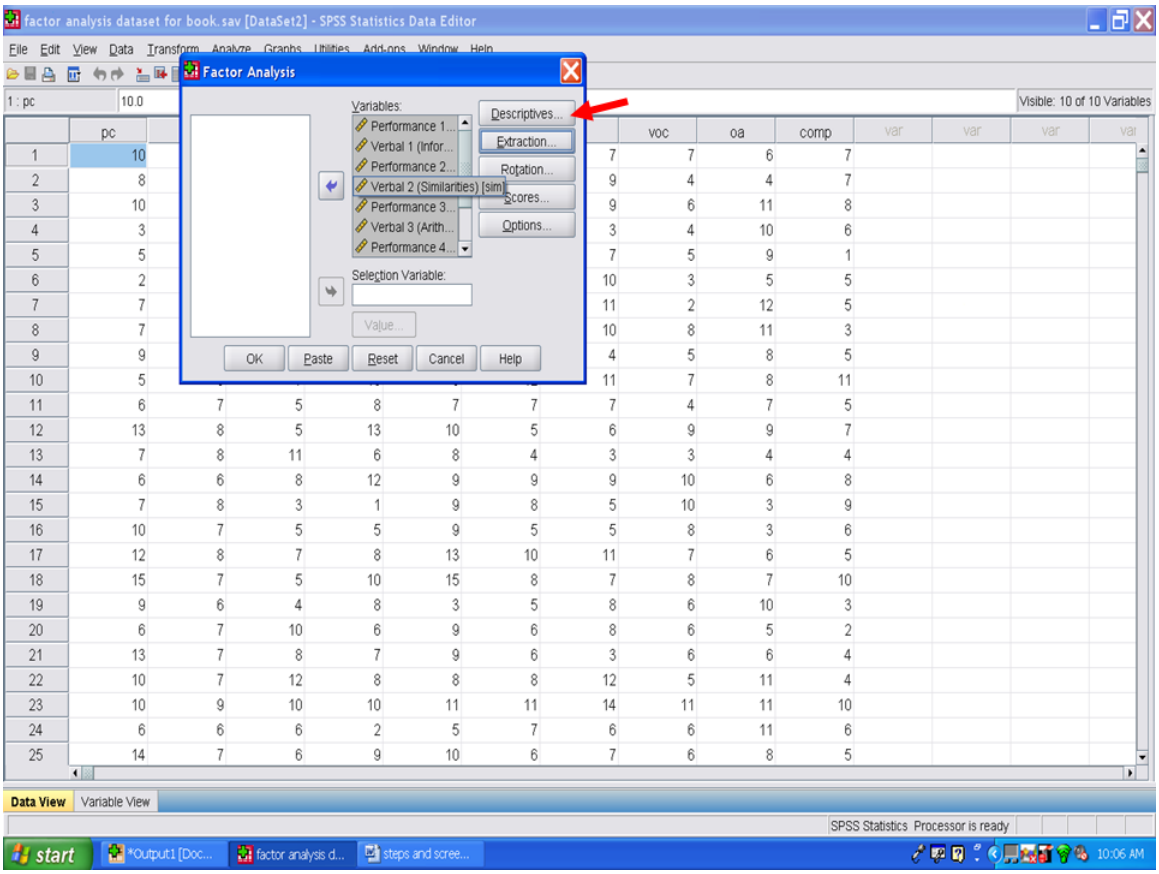
[11] http://cnx.org/content/m40725/latest/8.5.png/image



12

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

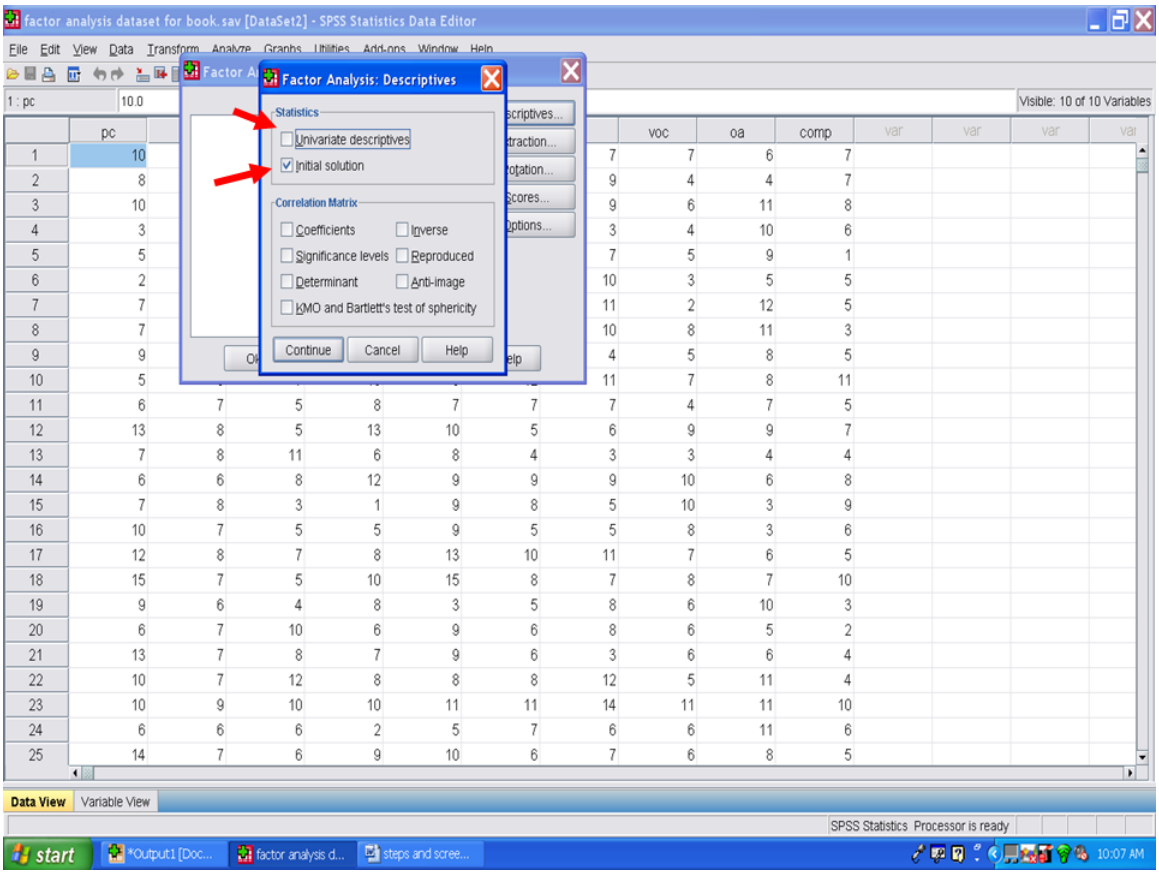
[12] http://cnx.org/content/m40725/latest/8.6.png/image



13

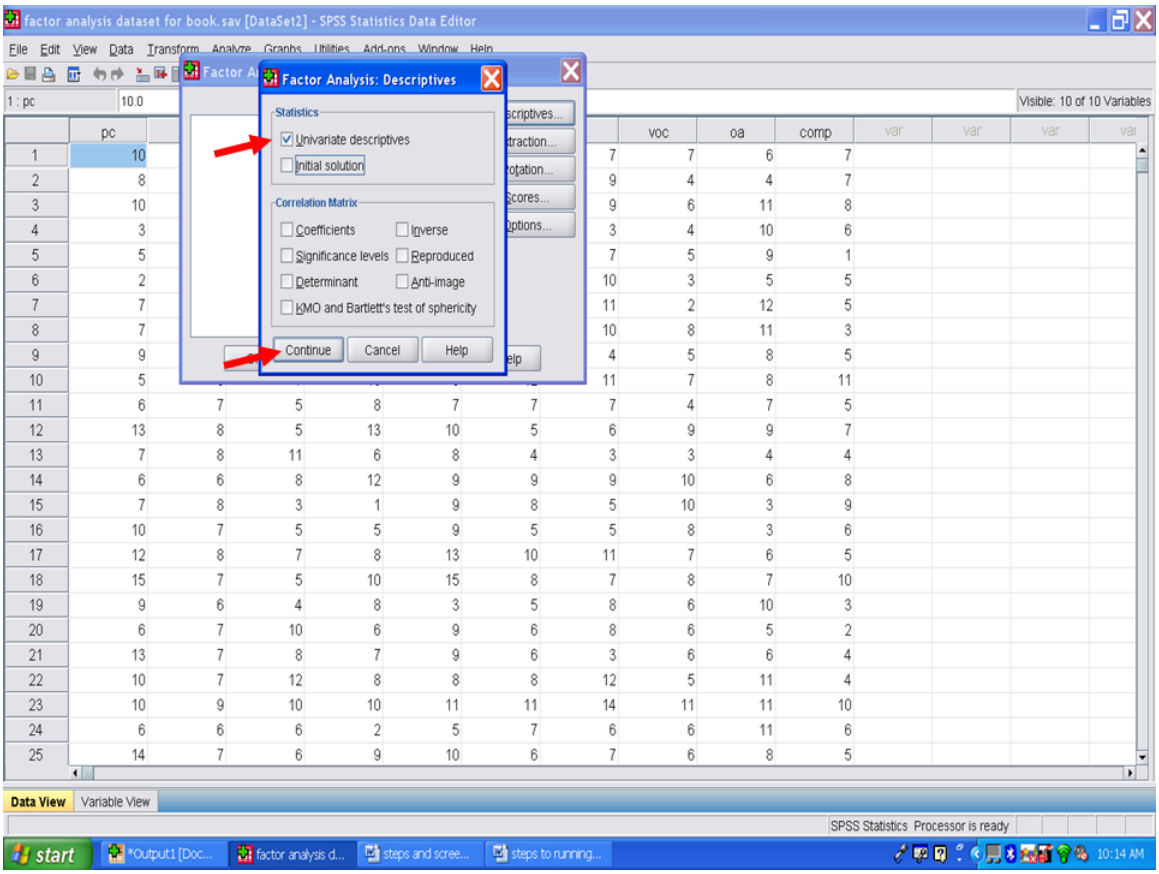
Your screen should now look like the one below. Click on Univariate Descriptives. Unclick Initial Solution.

[13] http://cnx.org/content/m40725/latest/8.7.png/image



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

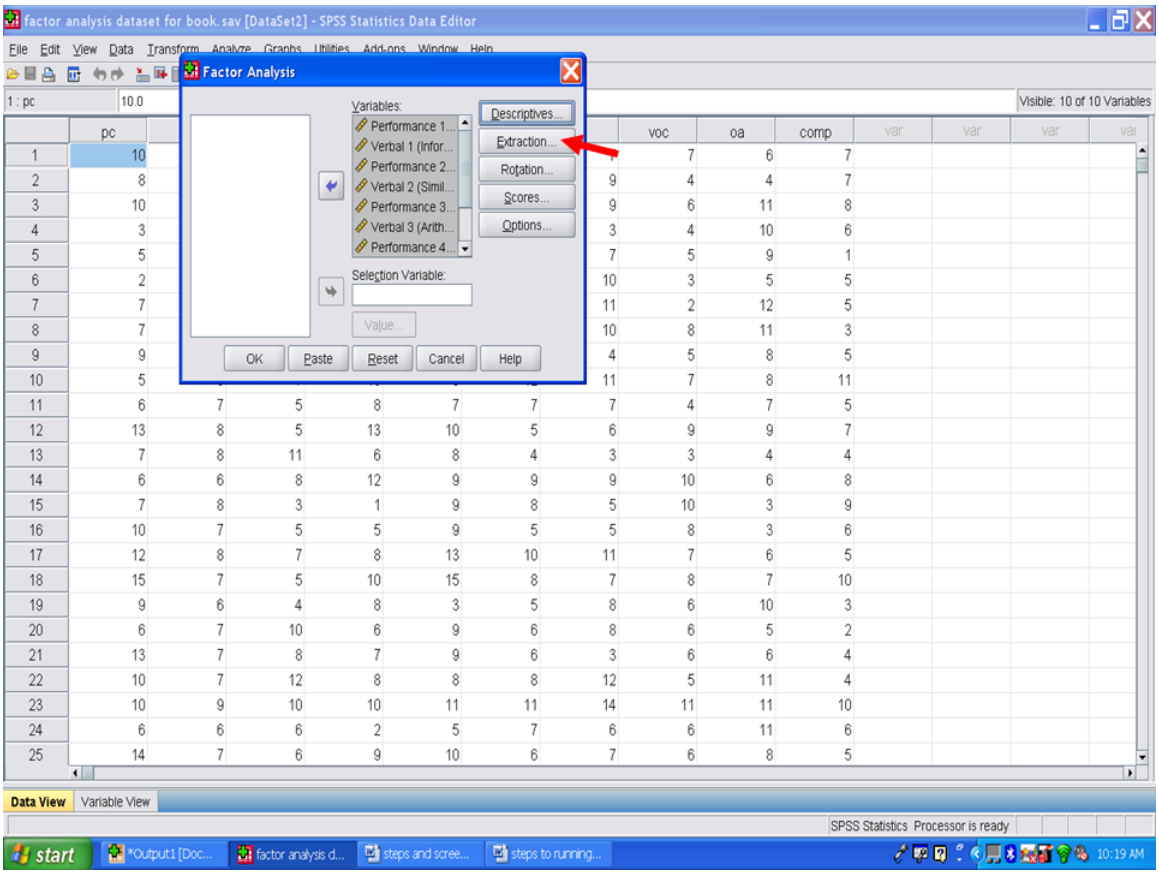
[14] http://cnx.org/content/m40725/latest/8.8.png/image



15

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

[15] http://cnx.org/content/m40725/latest/8.9.png/image



16

[16] http://cnx.org/content/m40725/latest/8.10.png/image

Chapter 10

10. Factor Analysis: Part II[1]

NOTE: This chapter is published by NCPEA Press[2] 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.

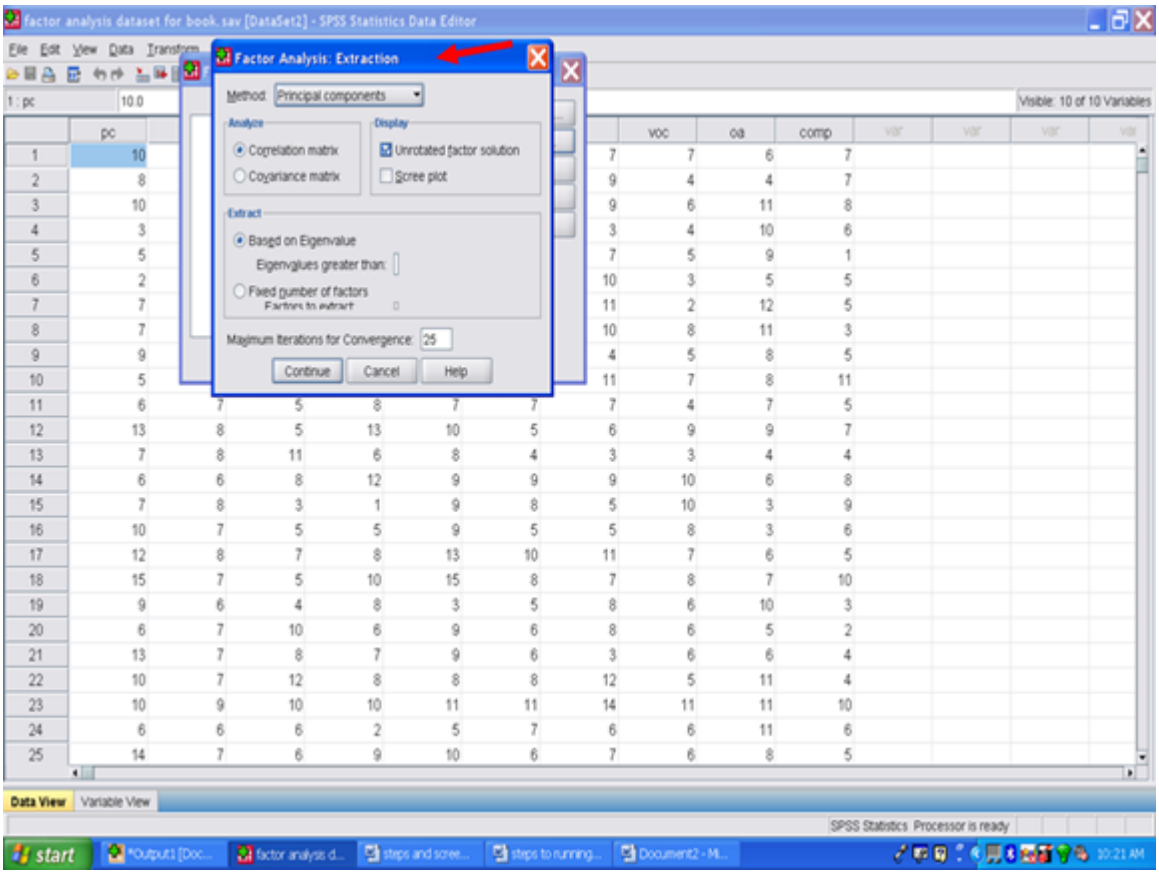
About the Authors

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 Help[3]) 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.

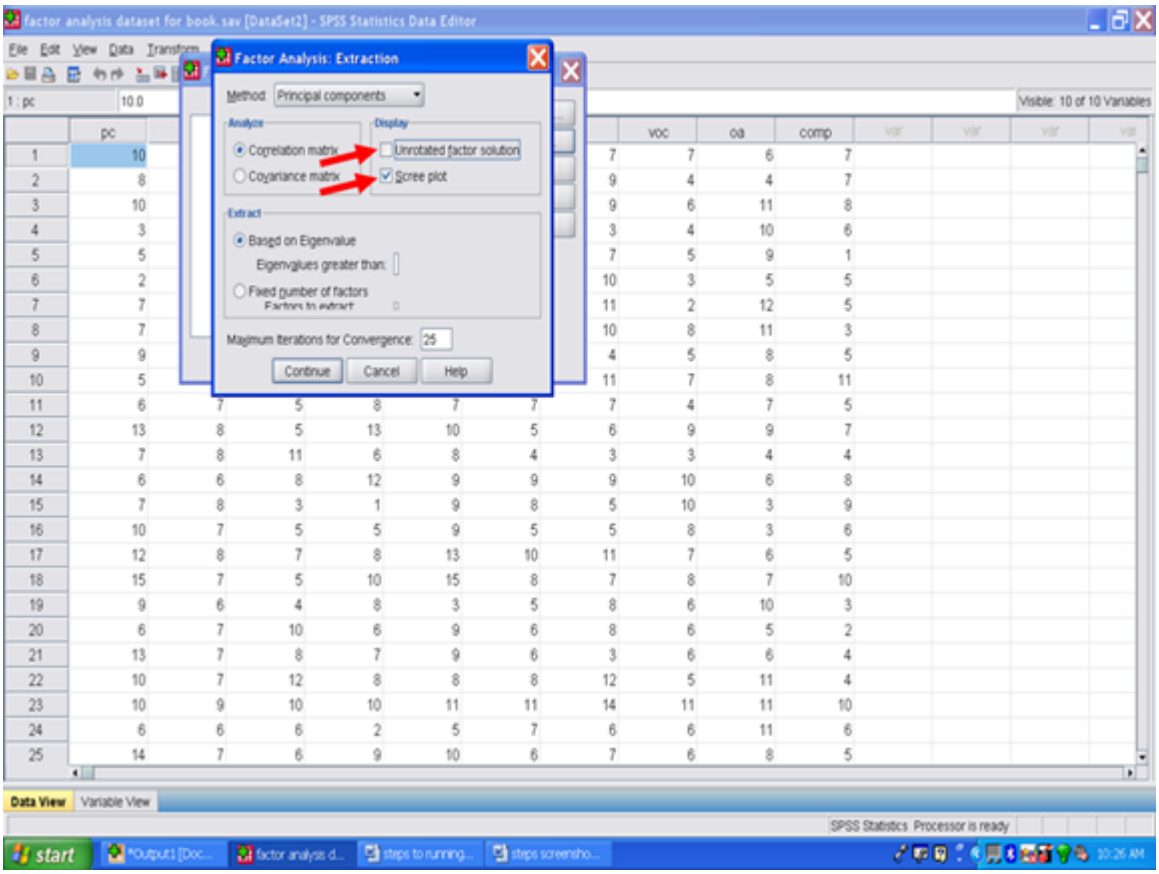
[1] This content is available online at <http://cnx.org/content/m40727/1.2/>.
[2] http://www.ncpeapublications.org/books.html
[3] http://cnx.org/content/m40727/latest/www.writingandstatisticalhelp



[4]

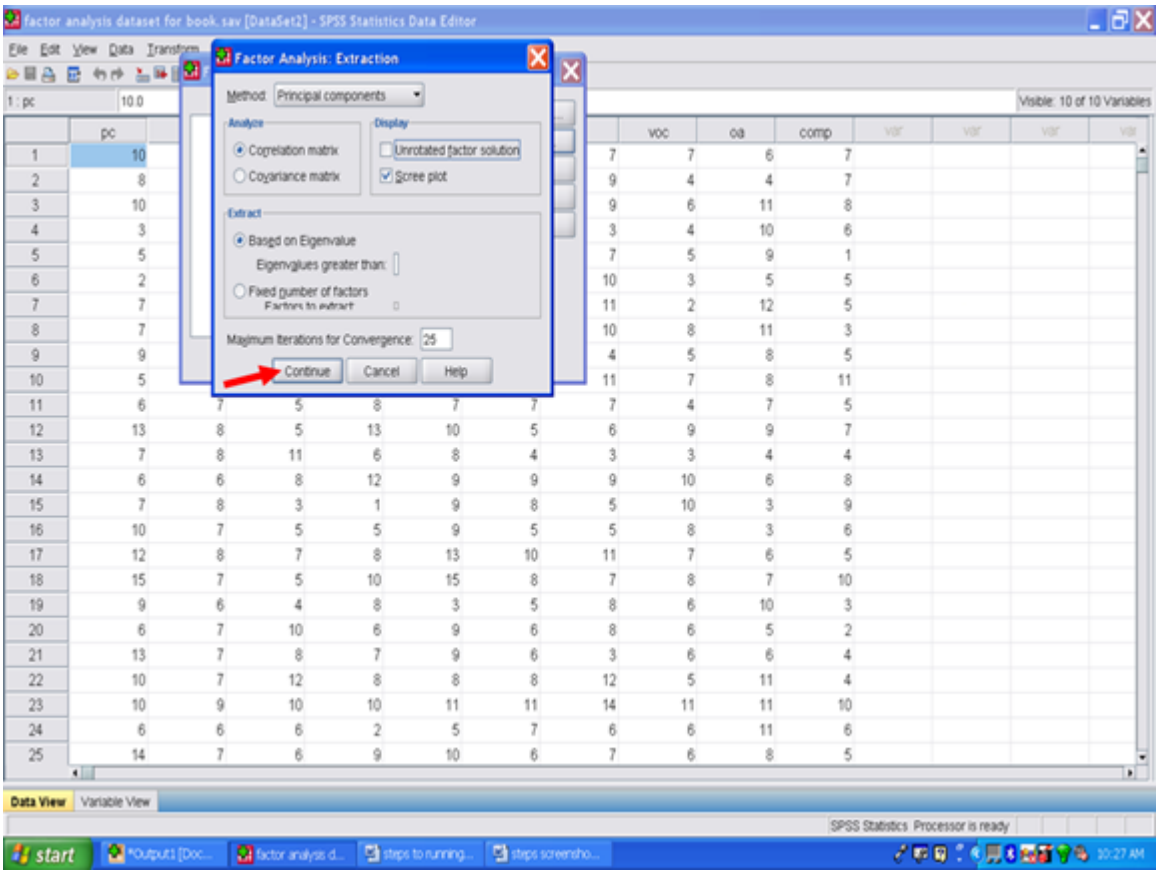
Under Display, click on Screen plot.
Unclick the Unrotated factor solution.

[4] http://cnx.org/content/m40727/latest/9.1.png/image



Click on Continue.

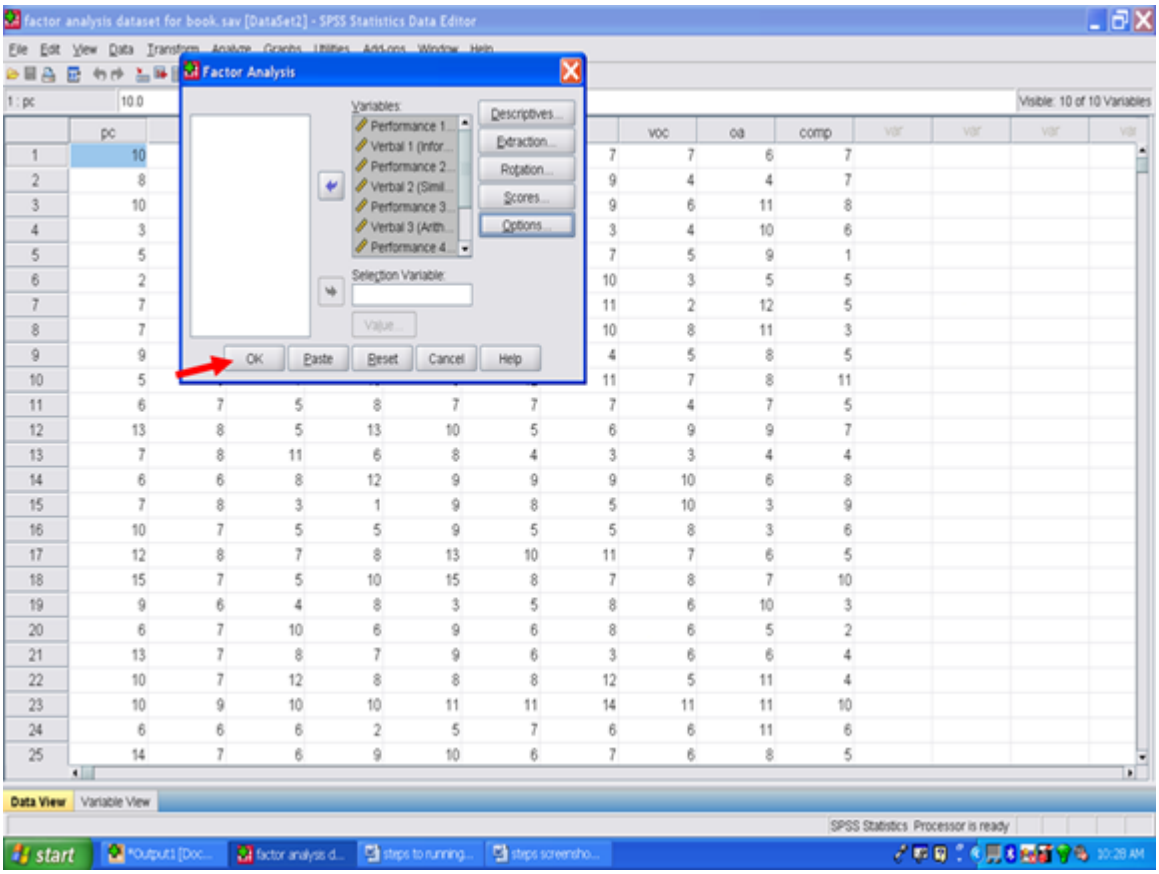
[5] http://cnx.org/content/m40727/latest/9.2.png/image



6

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

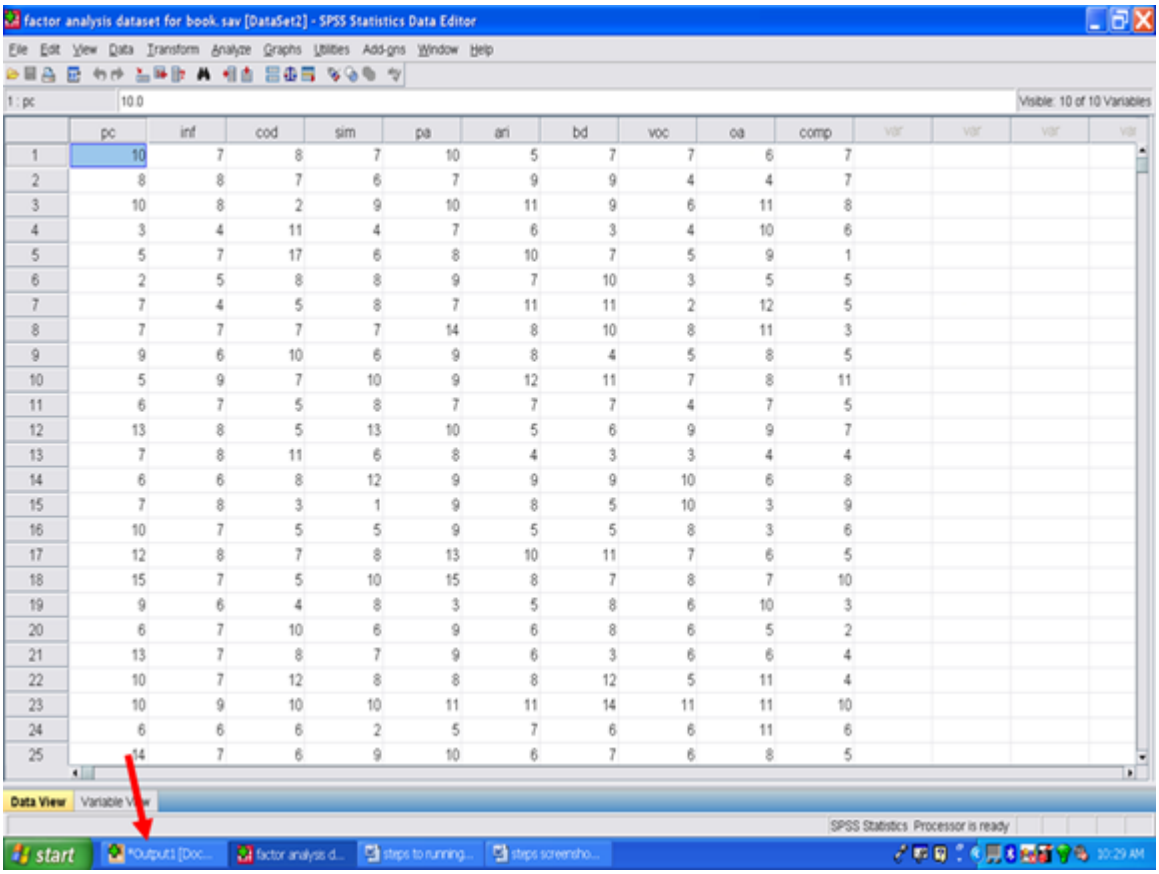
[6] http://cnx.org/content/m40727/latest/9.3.png/image



[7]

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.

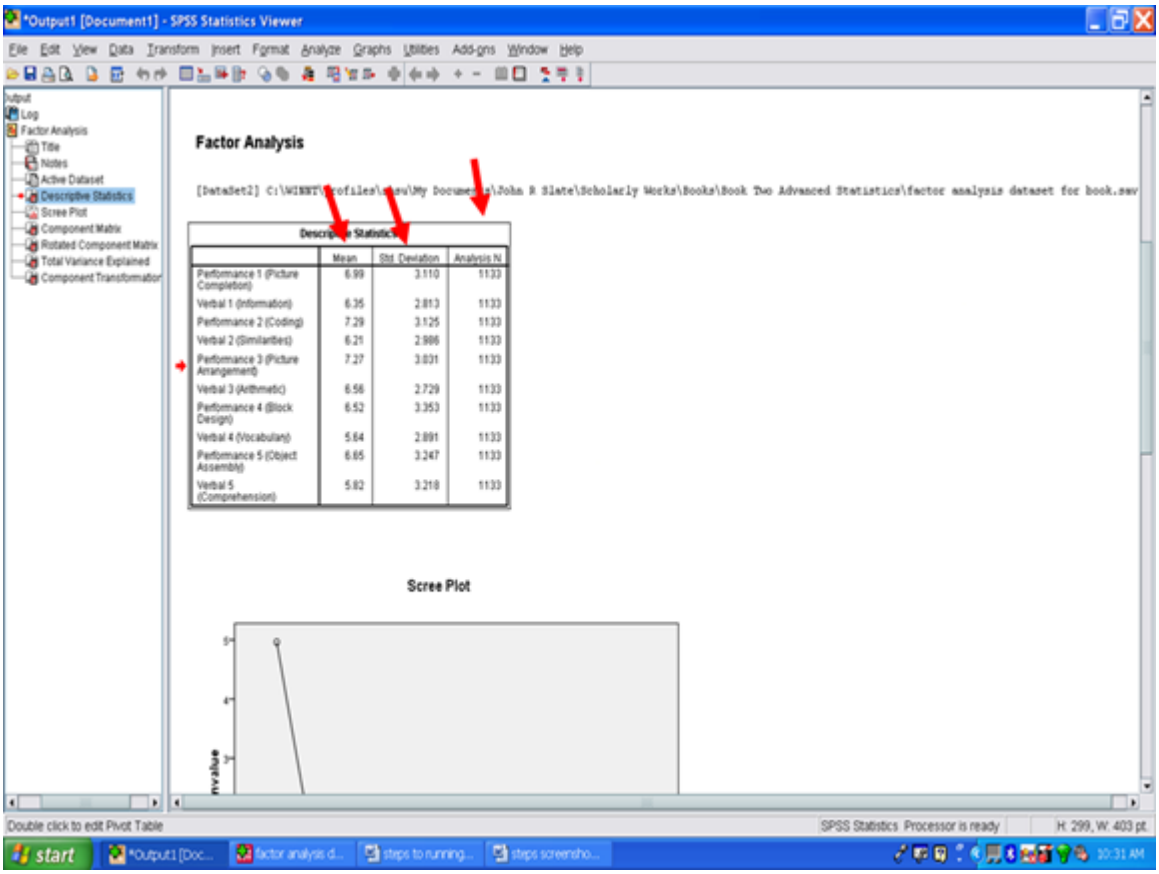
[7] http://cnx.org/content/m40727/latest/9.4.png/image



[8]

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.

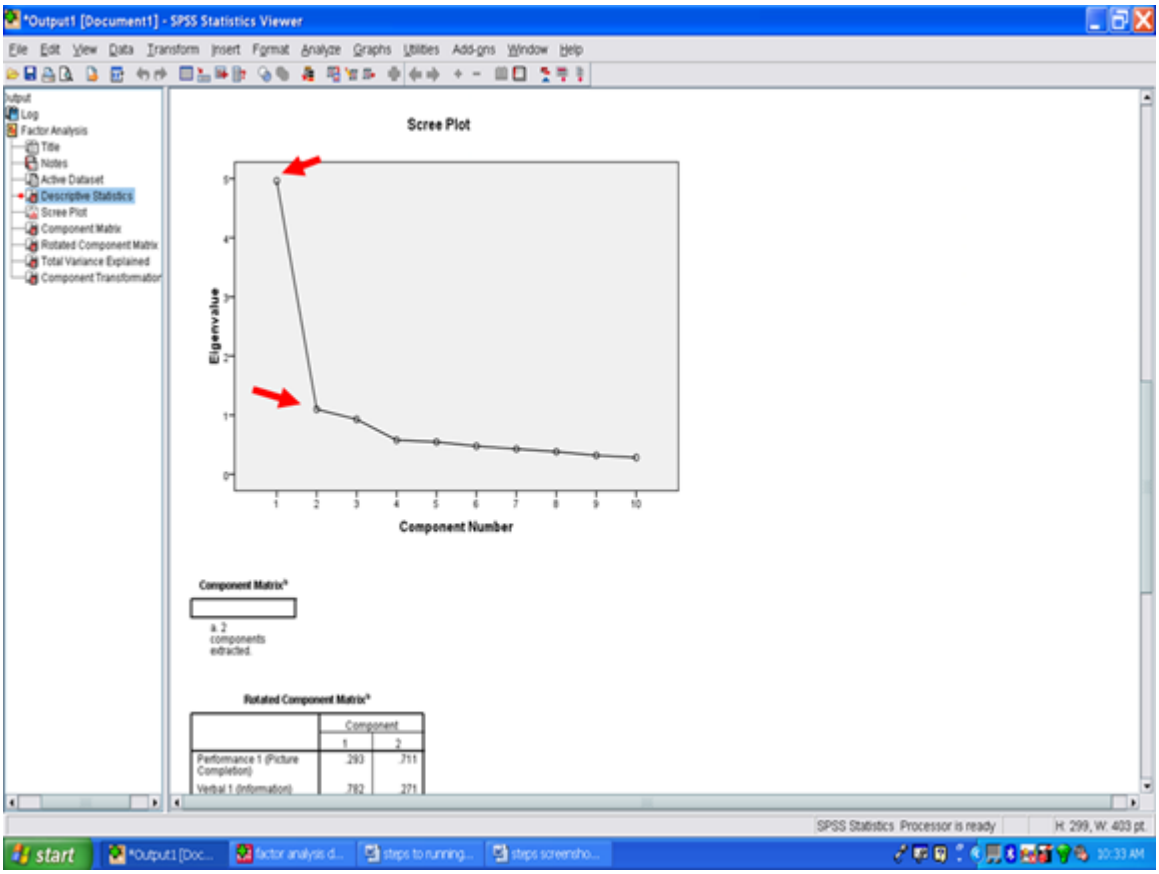
[8] http://cnx.org/content/m40727/latest/9.5.png/image



[9]

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.

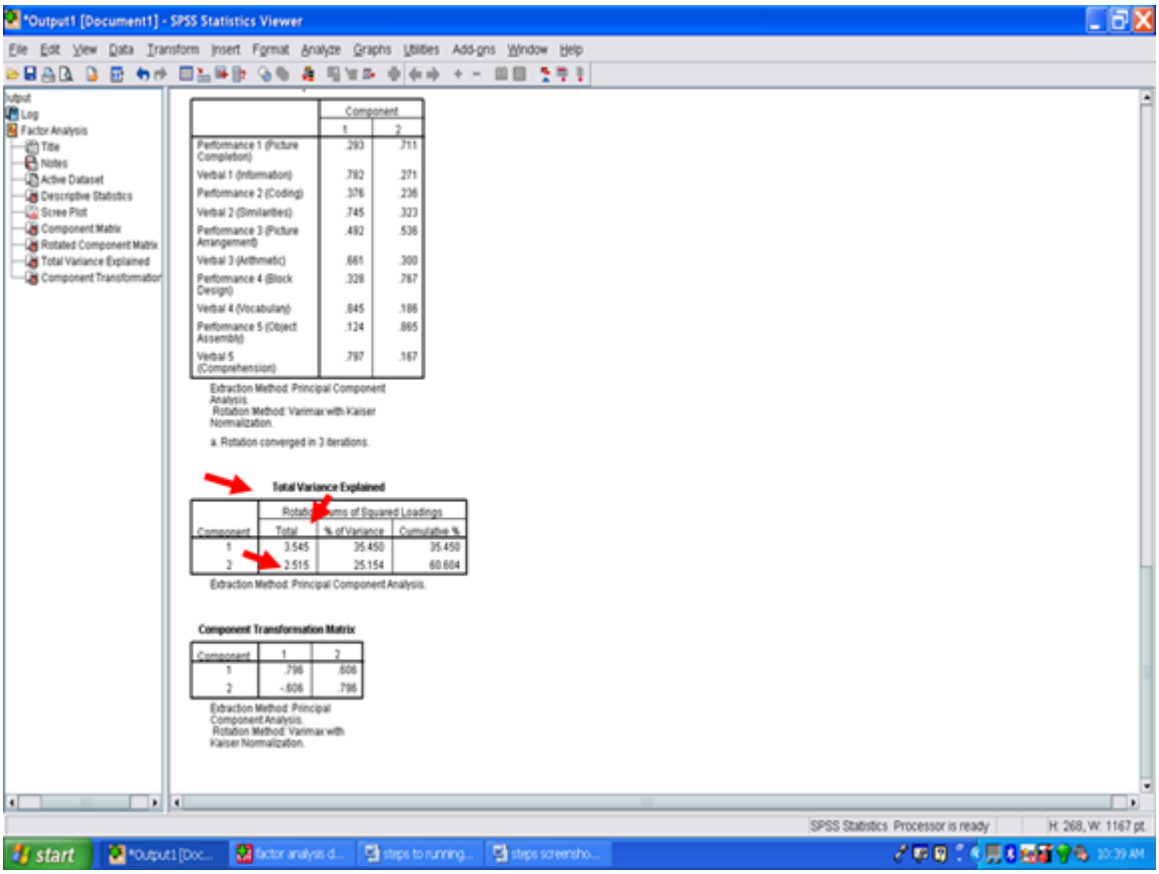
[9] http://cnx.org/content/m40727/latest/9.6.png/image



10

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.

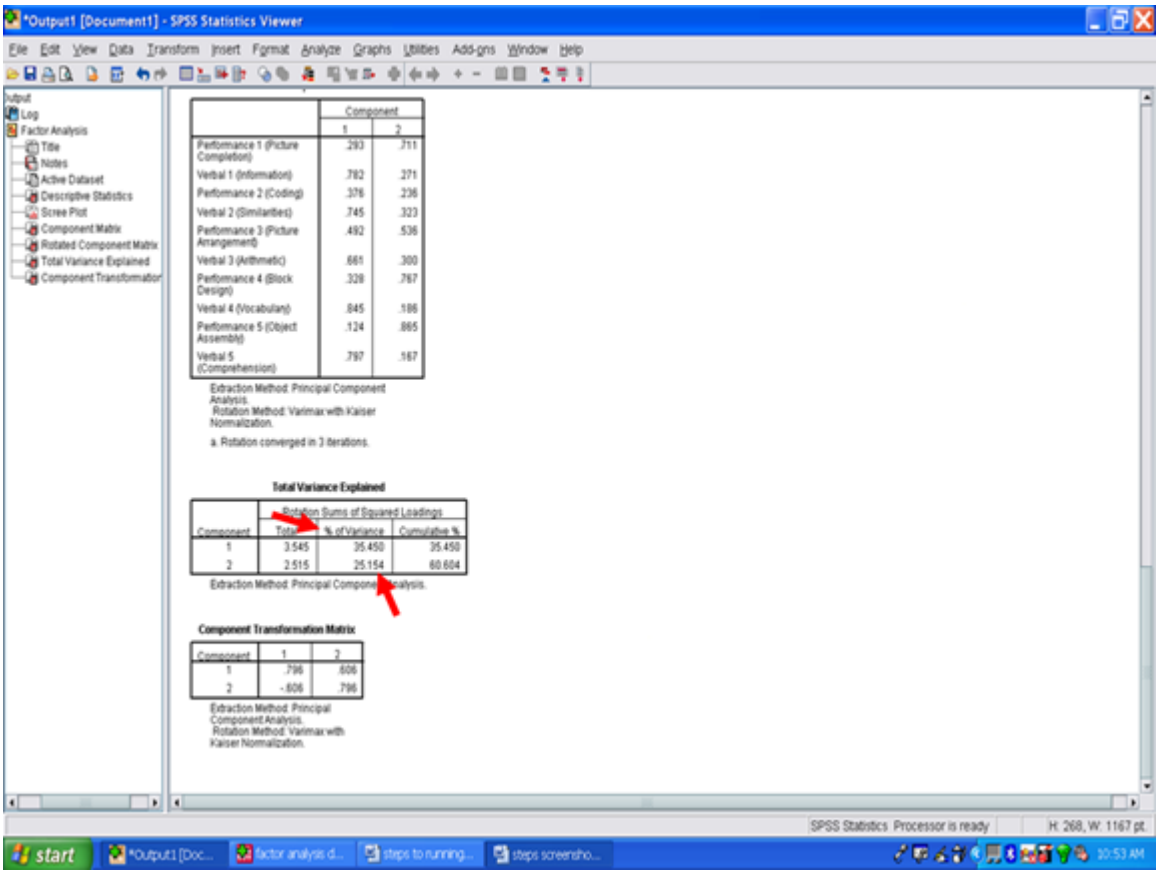
[10] http://cnx.org/content/m40727/latest/9.7.png/image



[11]

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.

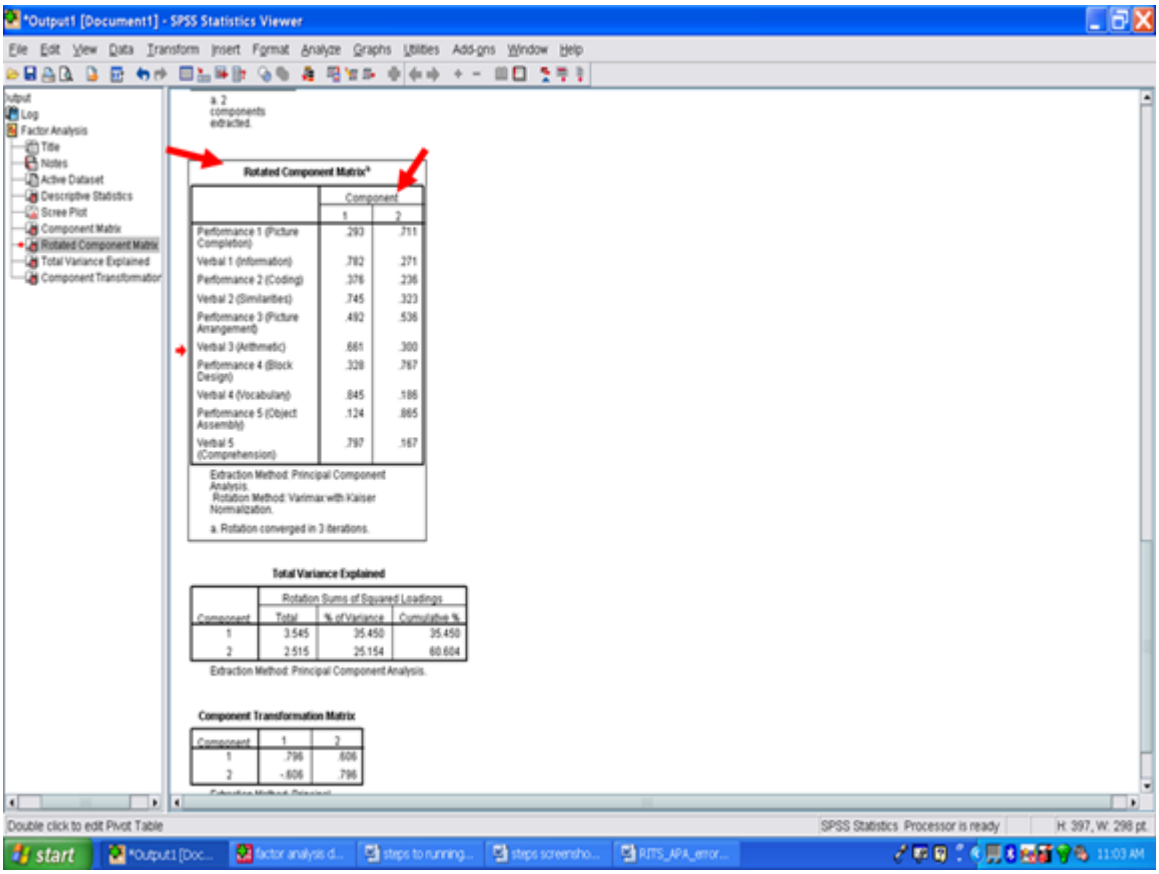
[11] http://cnx.org/content/m40727/latest/9.8.png/image



[12]

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.

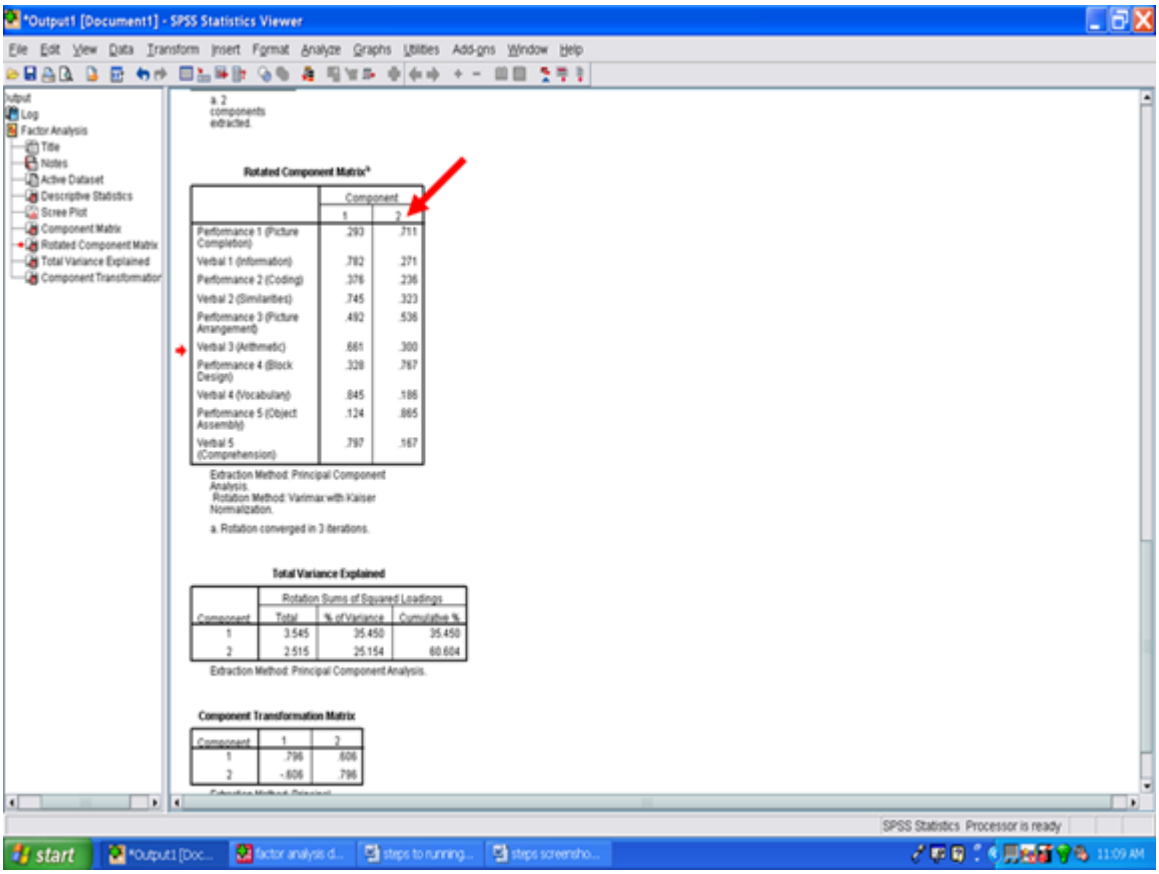
[12] http://cnx.org/content/m40727/latest/9.9.png/image



[13]

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.

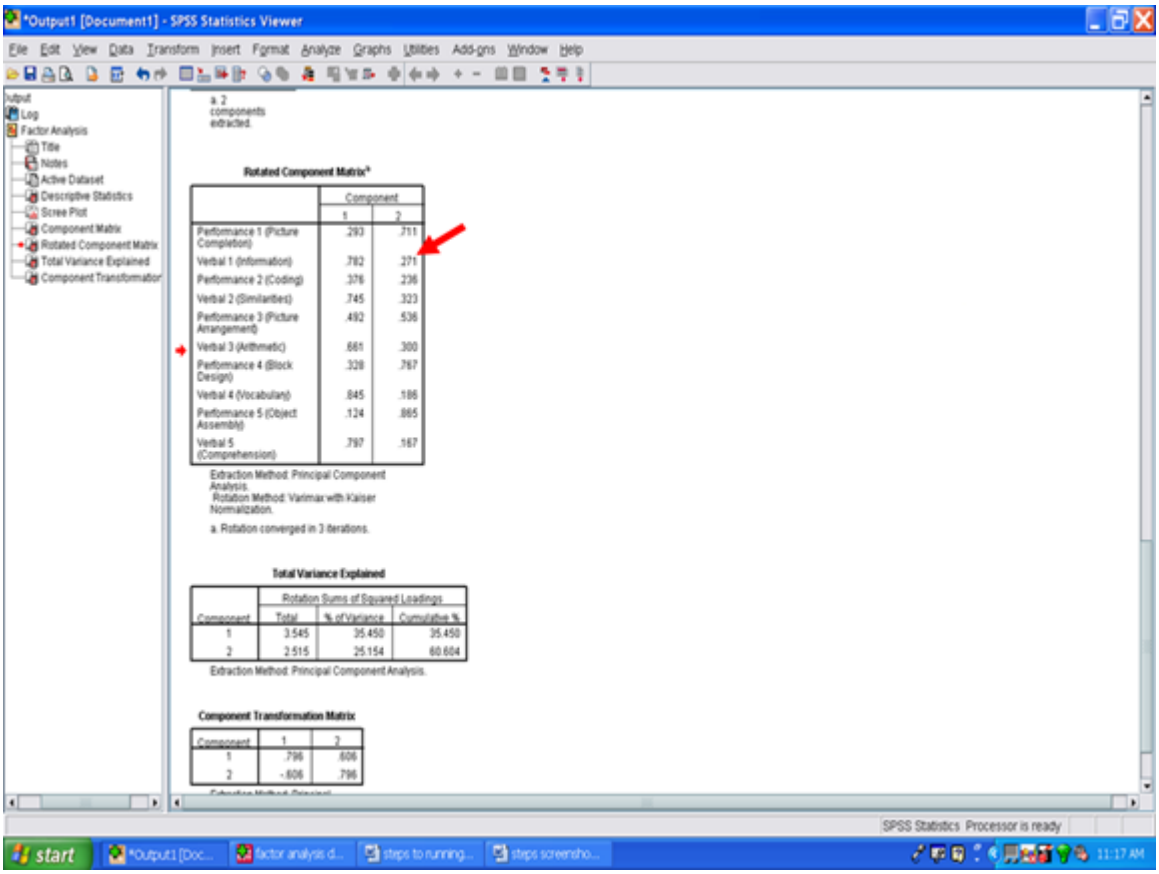
[13] http://cnx.org/content/m40727/latest/9.10.png/image



[14]

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



15

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 I[1]

NOTE: This chapter is published by NCPEA Press[2] 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 Help[3]) 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.

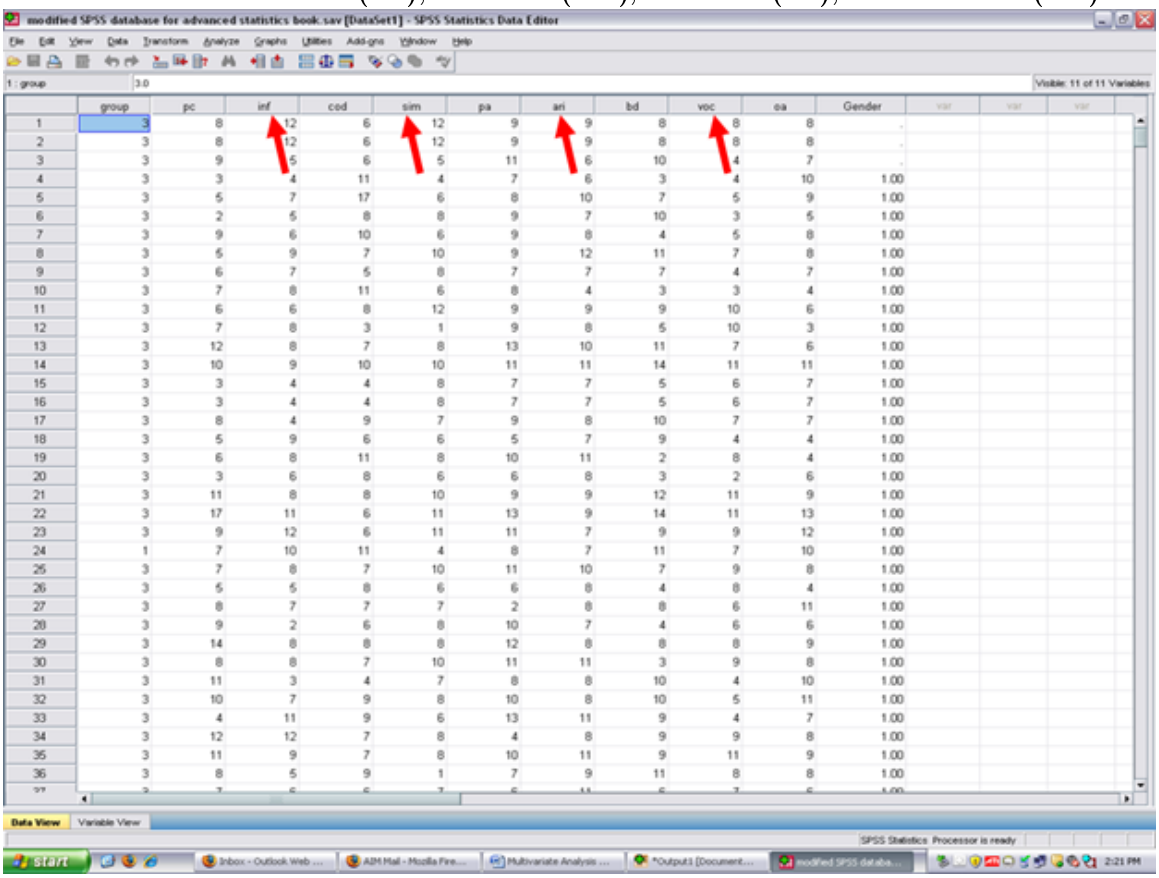
In this set of steps, readers will calculate a multivariate analysis of variance procedure, following the determination of the extent to which data for the dependent variables reflect normal distributions. Although a parametric statistical procedure requires that its data be reflective of a normal curve,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

[1] This 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=1-1[6]

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).



[7]

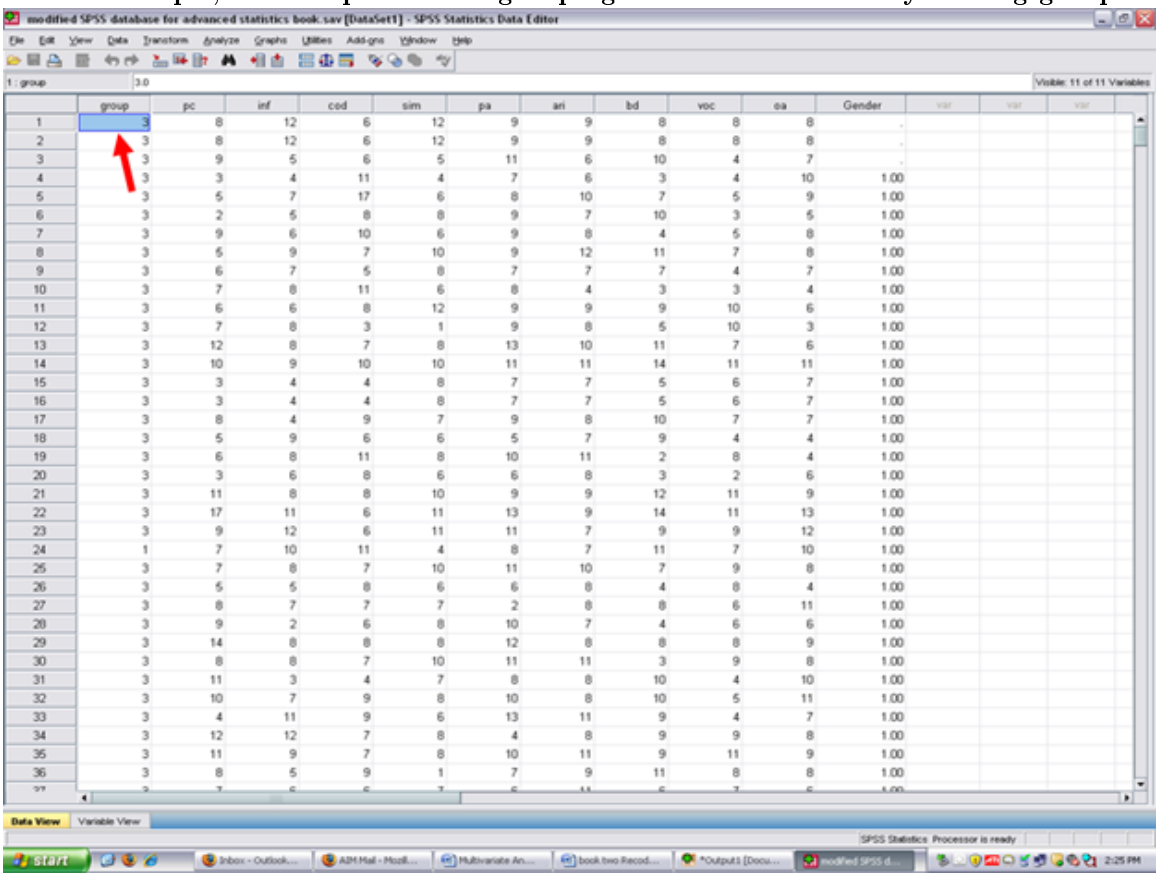
[4] http://davidmlane.com/hyperstat/

[5] http://www.statsoft.com/textbook/

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

[7] http://cnx.org/content/m40728/latest/10.1.png/image

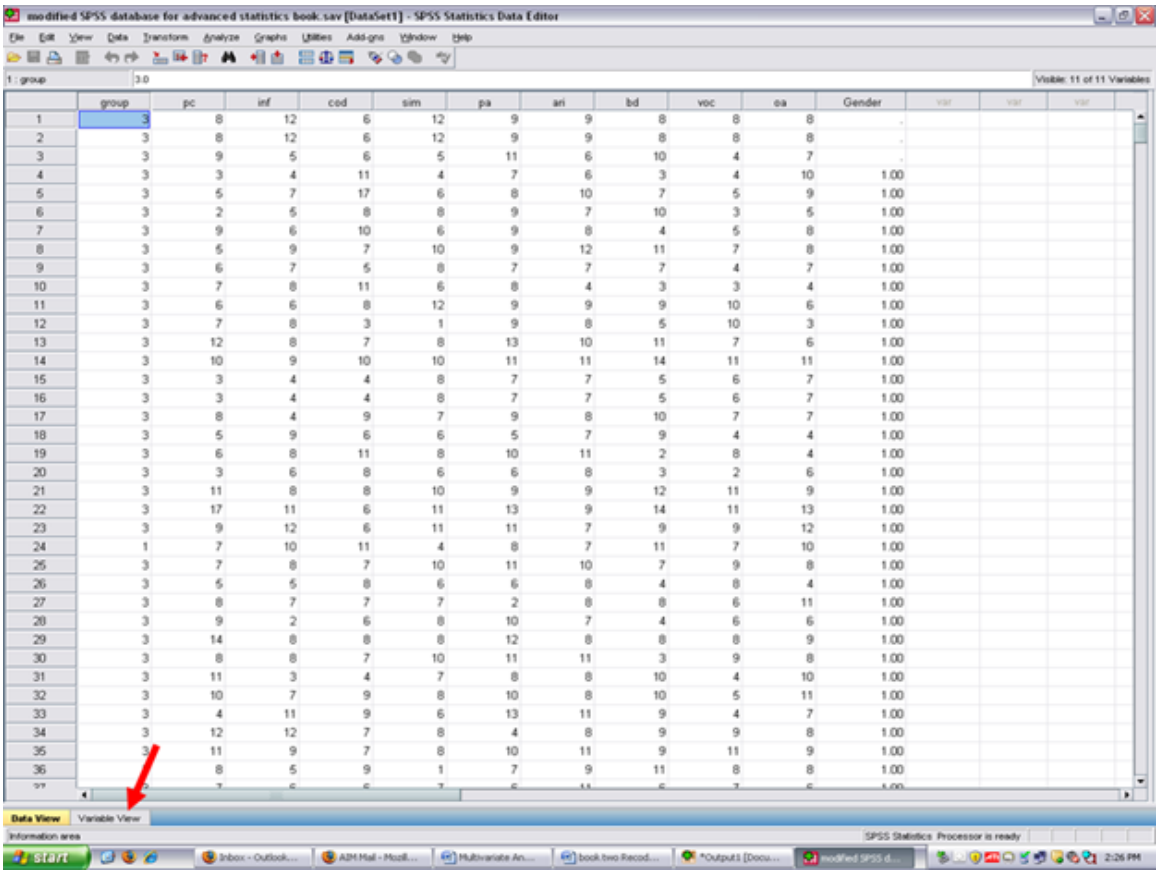
In this example, our independent or grouping variable is elementary reading group.



[8]

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.

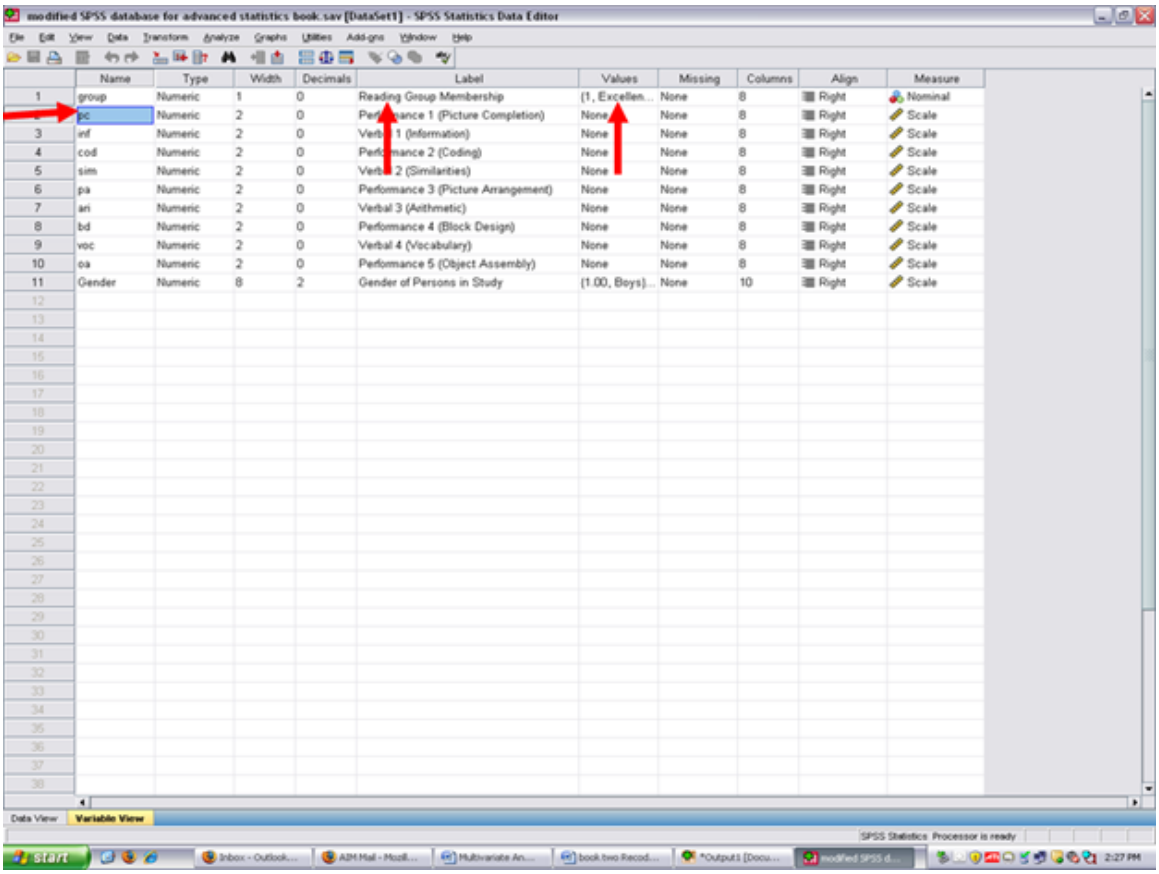
[8] http://cnx.org/content/m40728/latest/10.2.png/image



[9]

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.

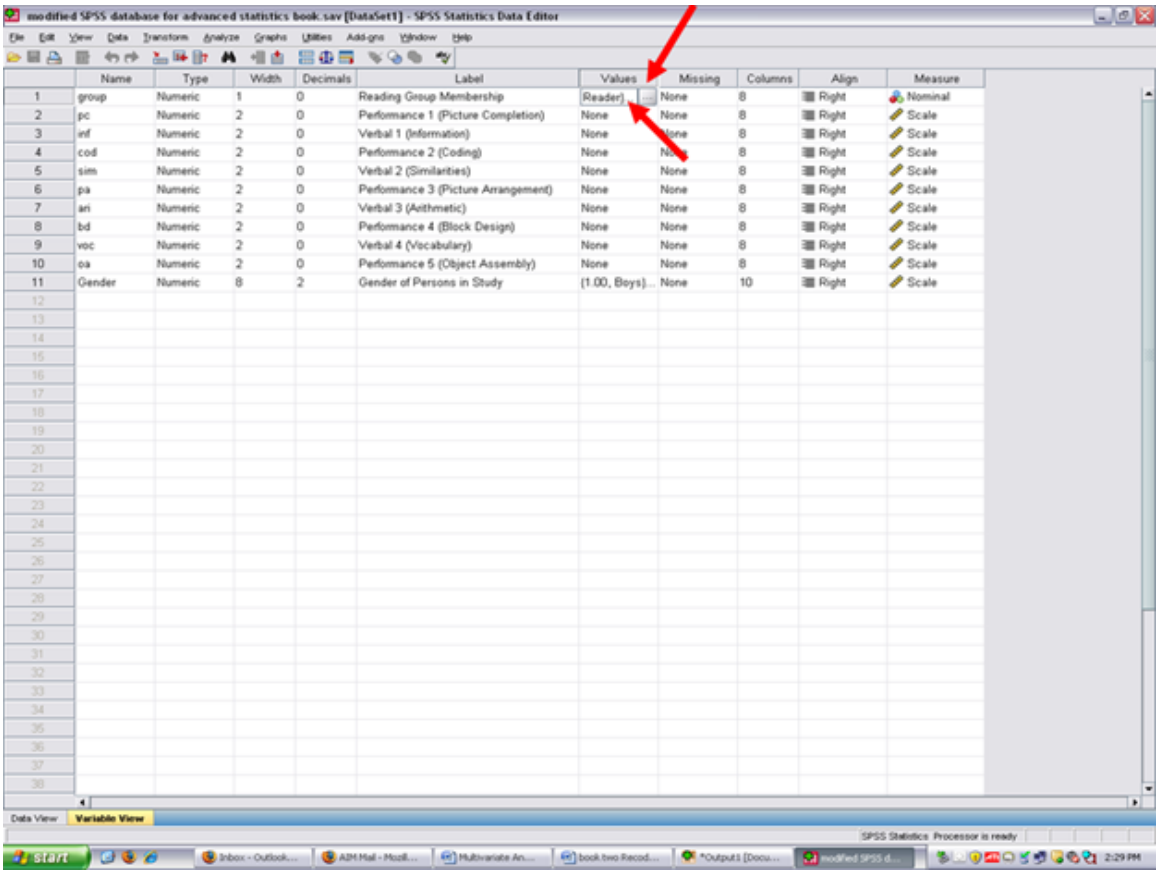
[9] http://cnx.org/content/m40728/latest/10.3.png/image



10

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

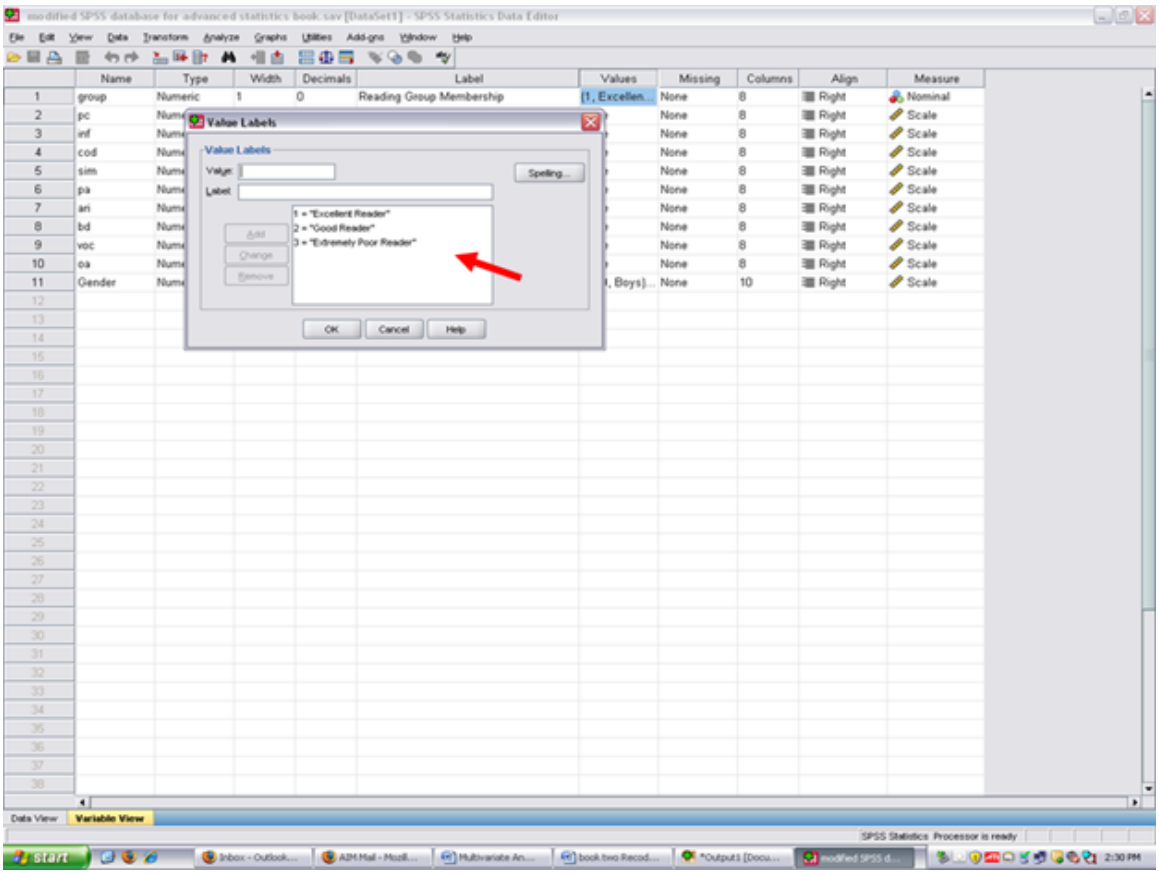
[10] http://cnx.org/content/m40728/latest/10.4.png/image



[11]

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).

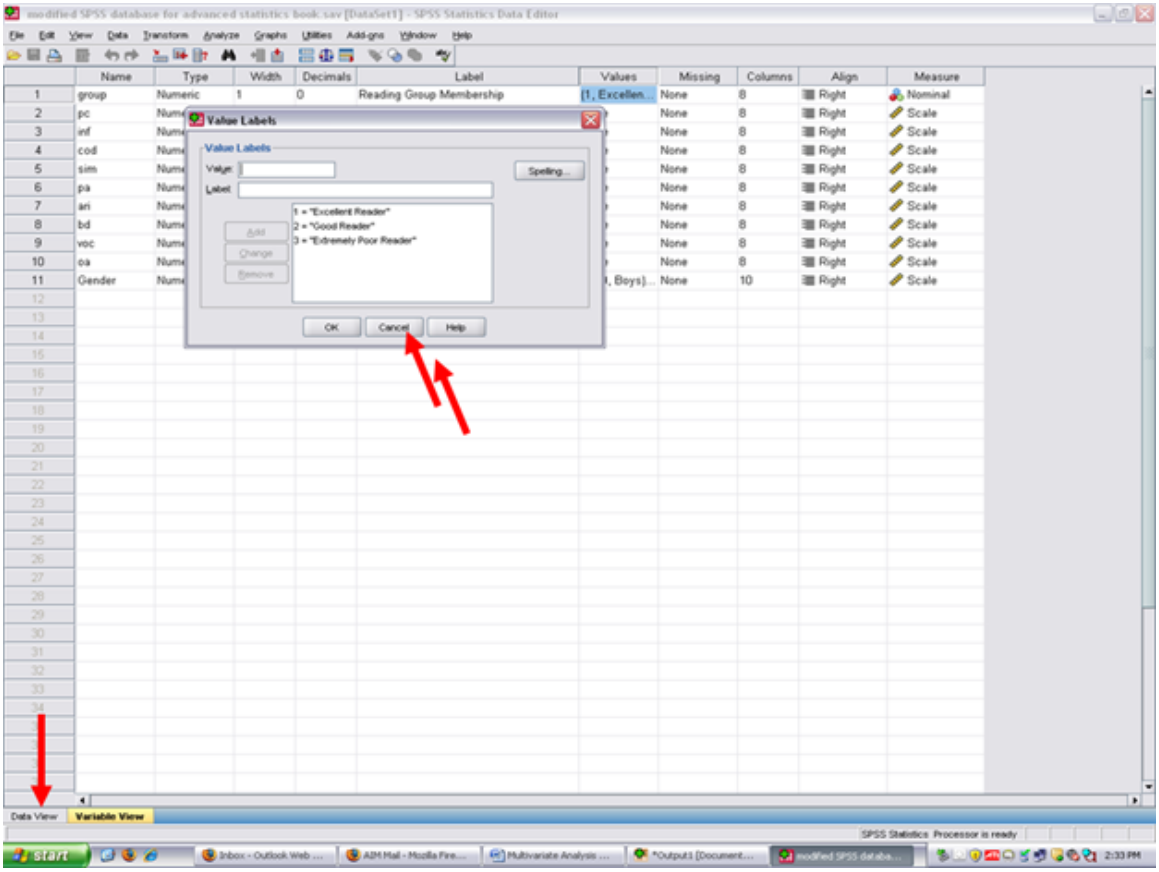
[11] http://cnx.org/content/m40728/latest/10.5.png/image



12

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.

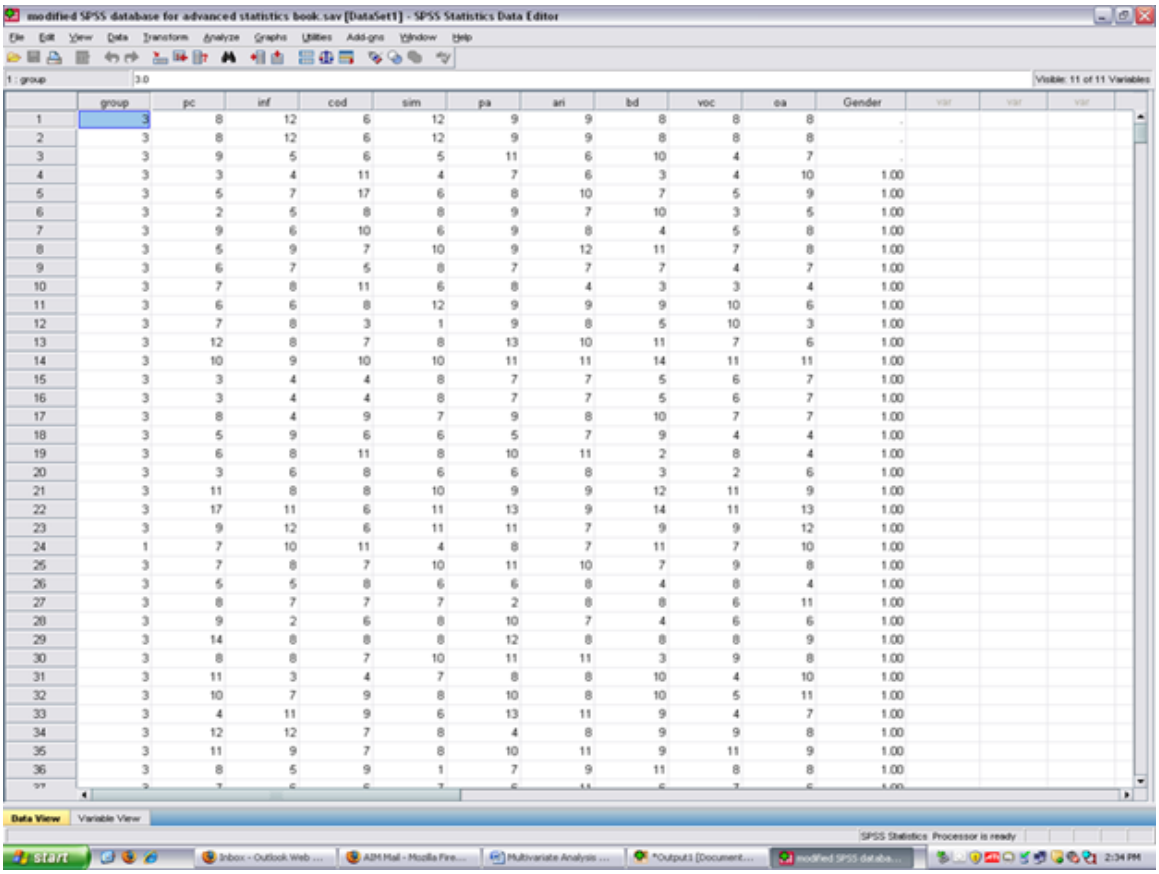
[12] http://cnx.org/content/m40728/latest/10.6.png/image



13

Your screen should now look like the one below.

[13] http://cnx.org/content/m40728/latest/10.7.png/image

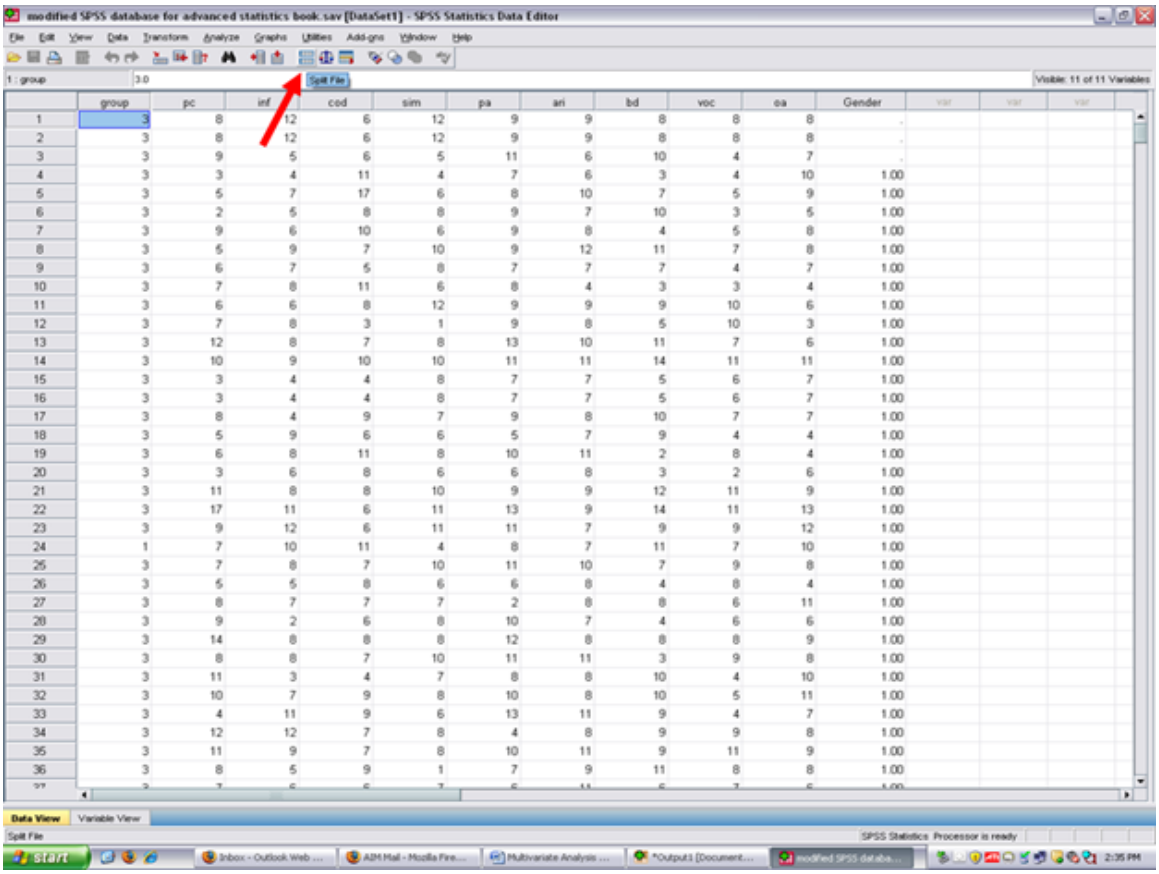


[14]

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.

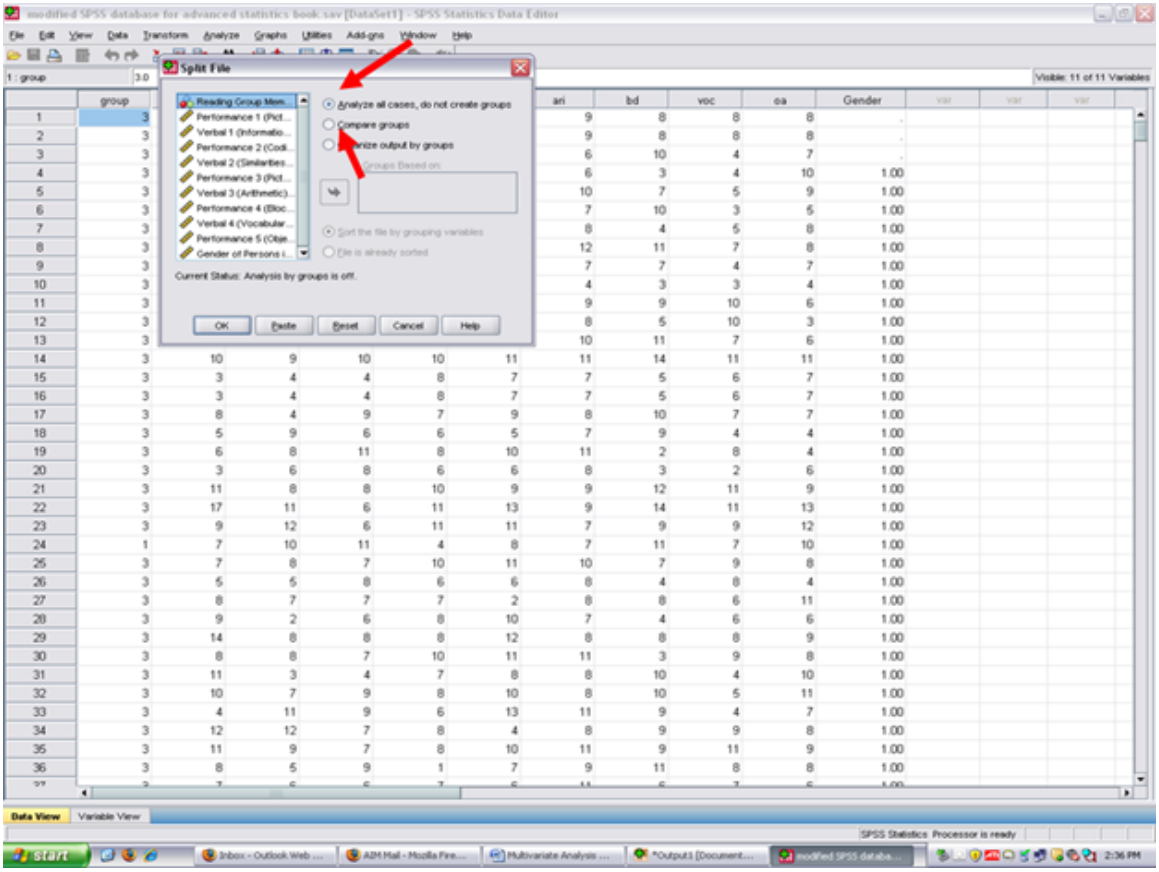
[14] http://cnx.org/content/m40728/latest/10.8.png/image



[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.

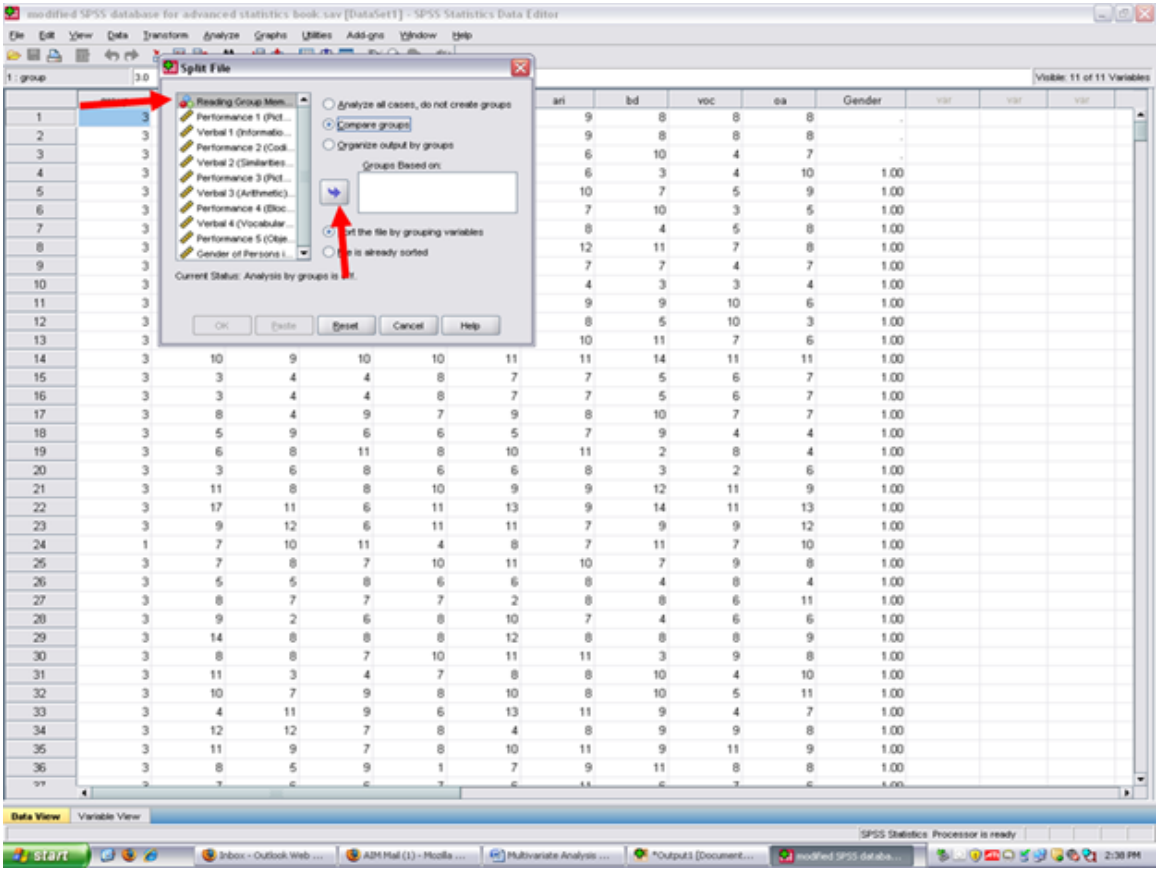
[15] http://cnx.org/content/m40728/latest/10.9.png/image



[16]

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.

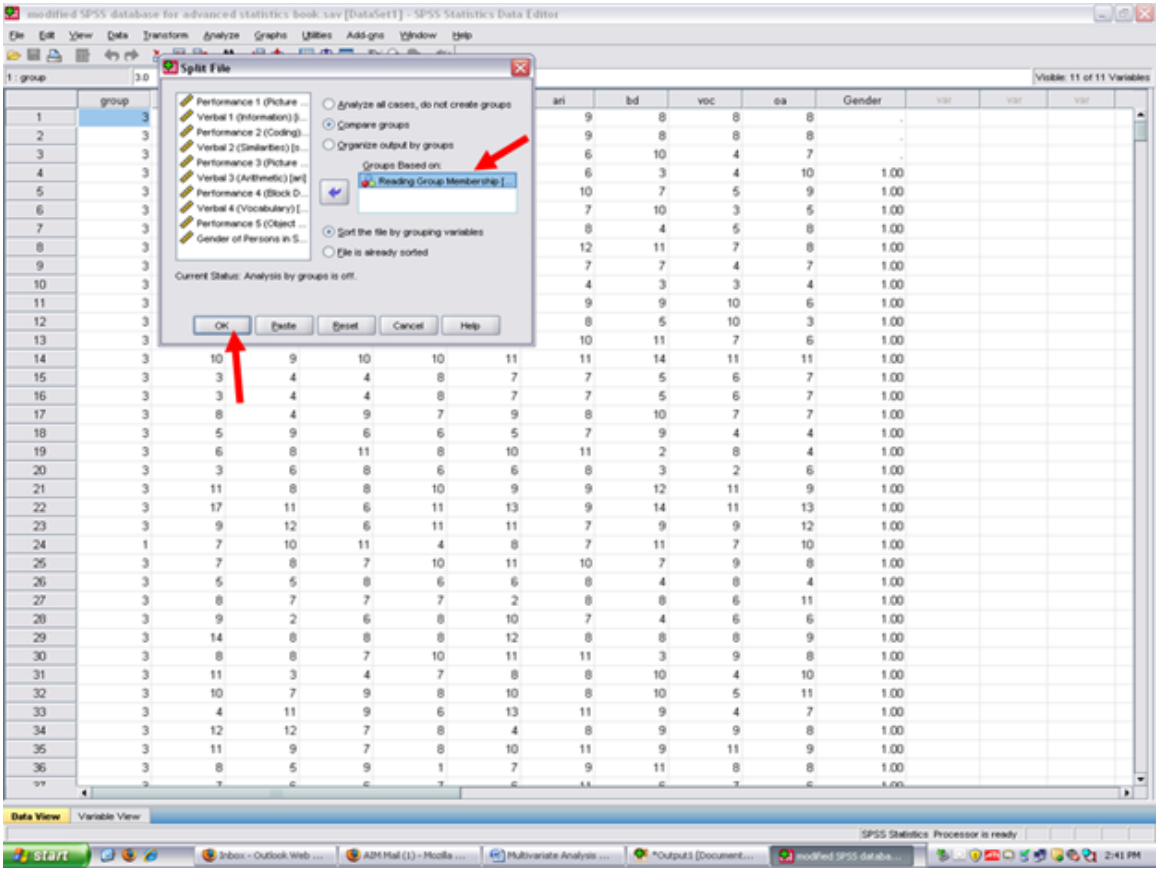
[16] http://cnx.org/content/m40728/latest/10.10.png/image



17

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

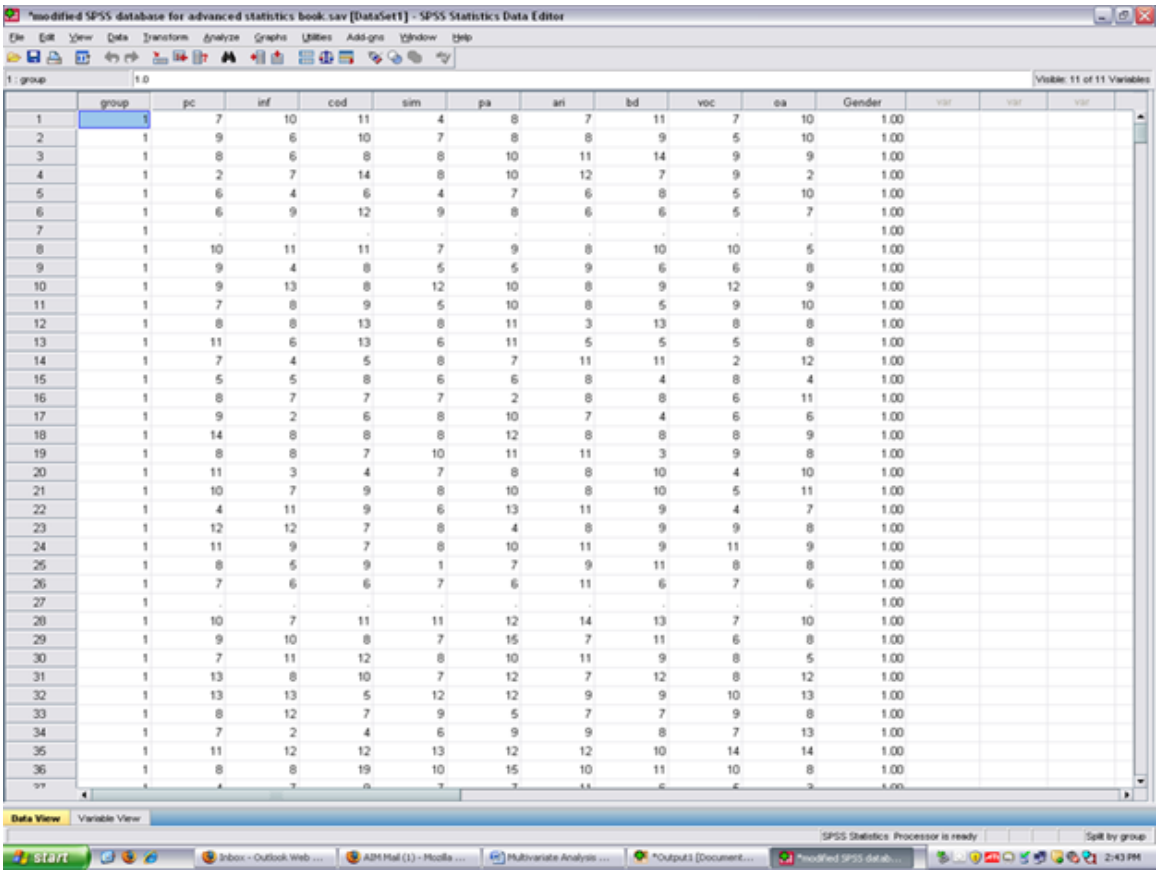
[17] http://cnx.org/content/m40728/latest/10.11.png/image



[18]

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

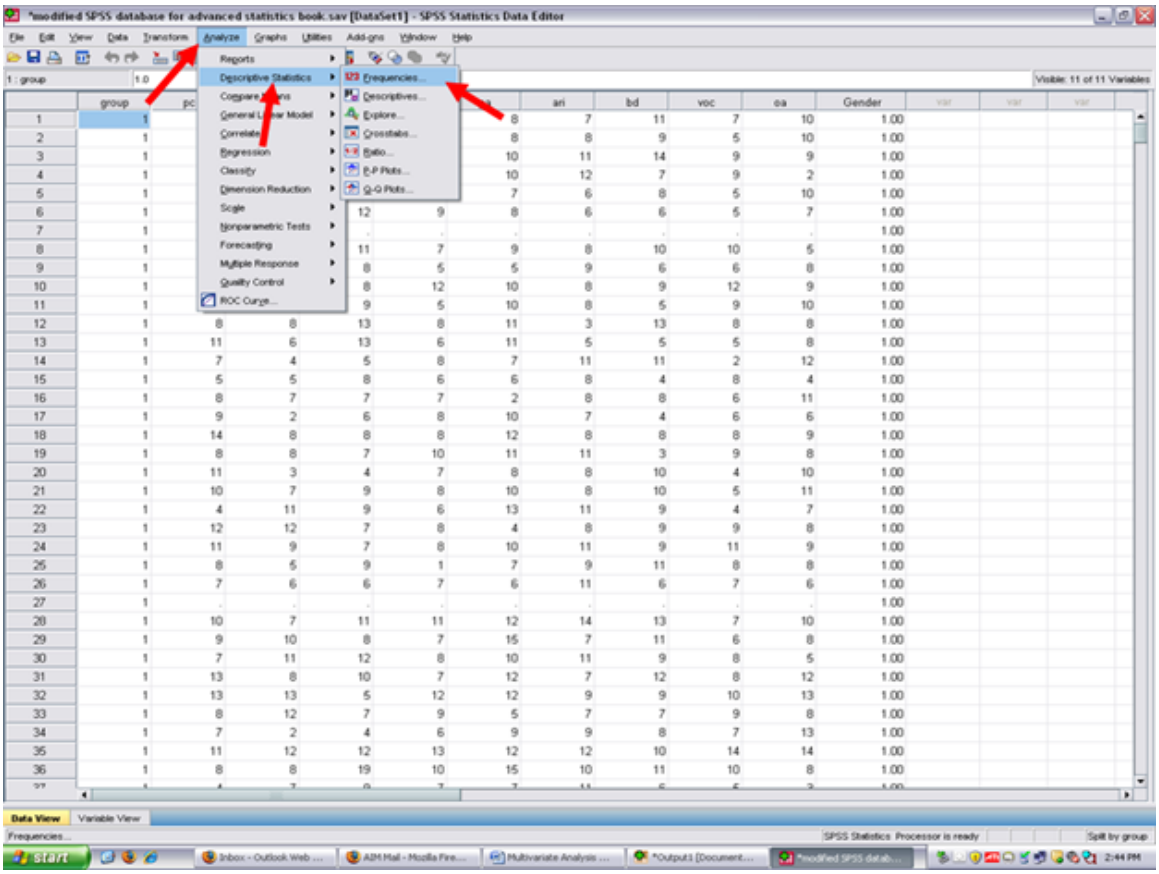
[18] http://cnx.org/content/m40728/latest/10.12.png/image



19

√ **Analyze**
*** Descriptive Statistics**
*** Frequencies**

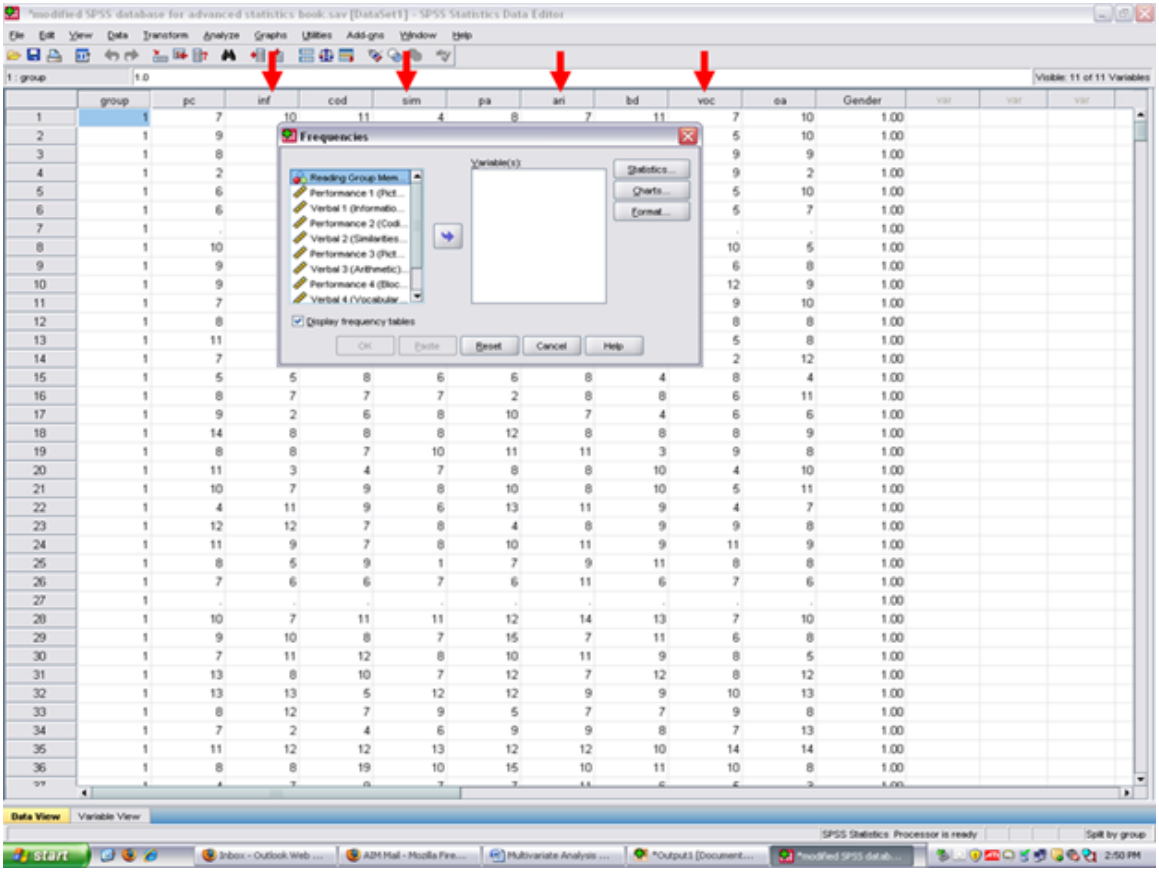
[19] http://cnx.org/content/m40728/latest/10.13.png/image



[20]

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)

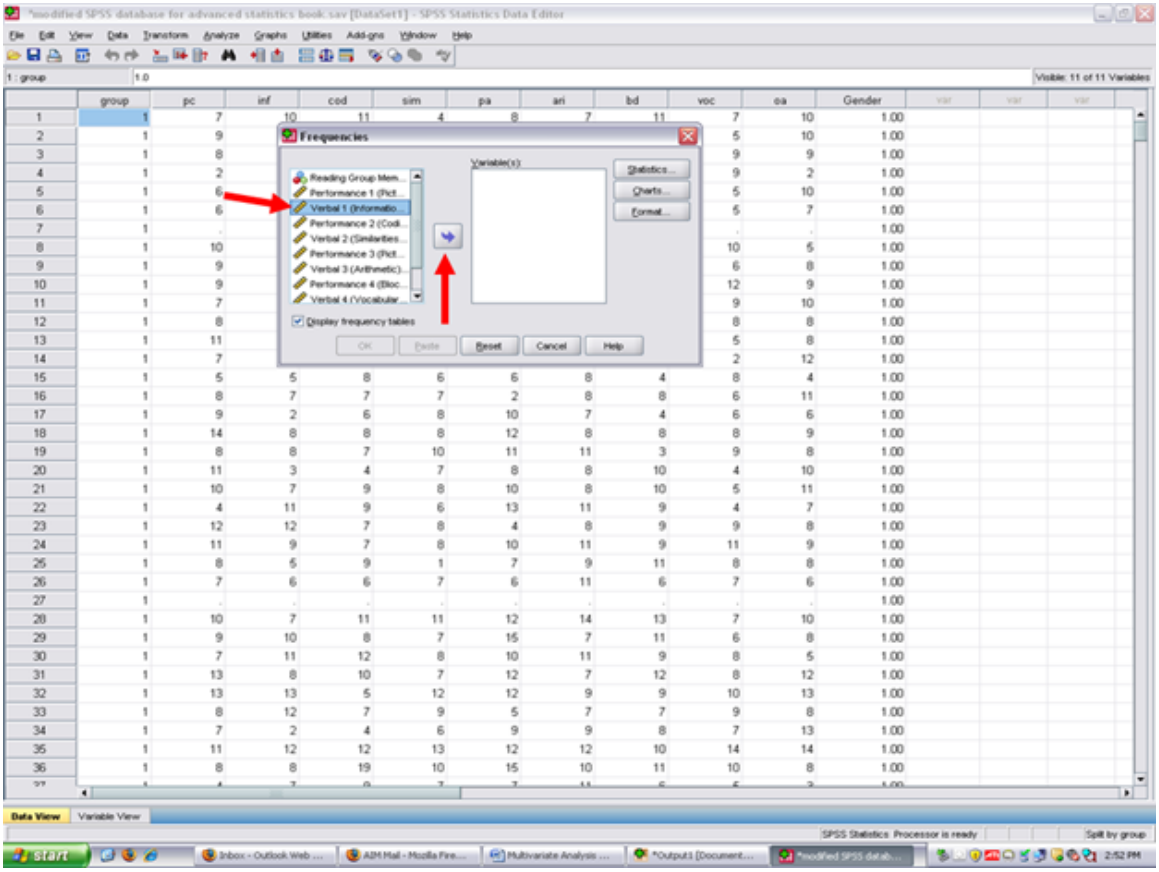
[20] http://cnx.org/content/m40728/latest/10.14.png/image



[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.**

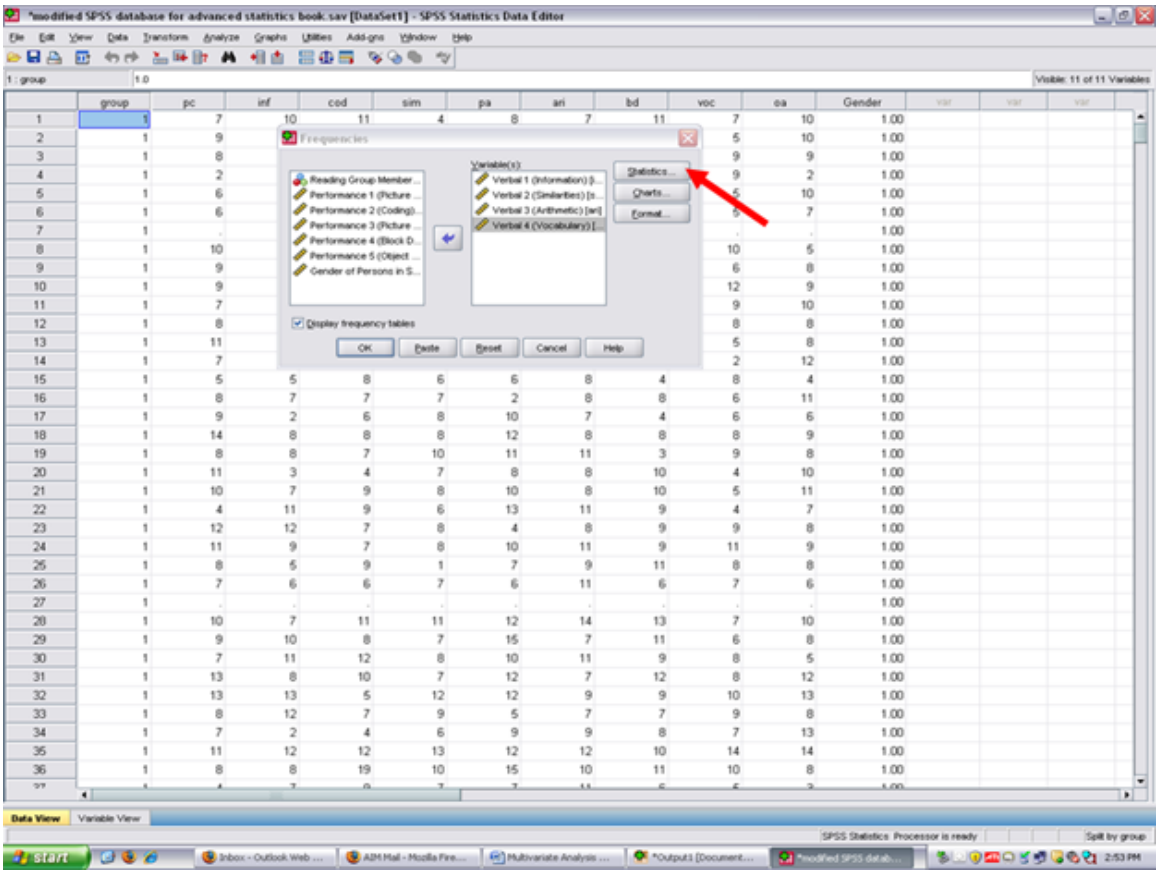
[21] http://cnx.org/content/m40728/latest/10.15.png/image



[22]

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

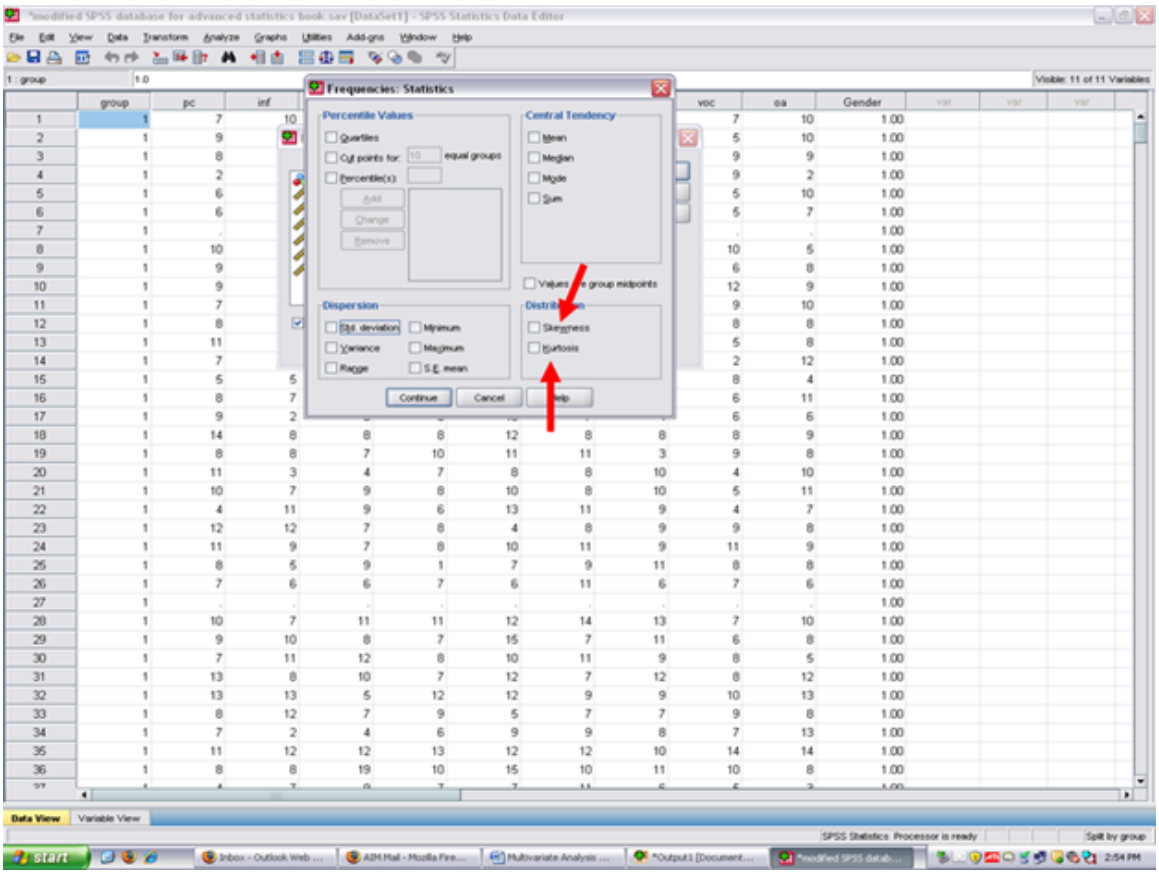
[22] http://cnx.org/content/m40728/latest/10.16.png/image



[23]

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=356[25] 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.

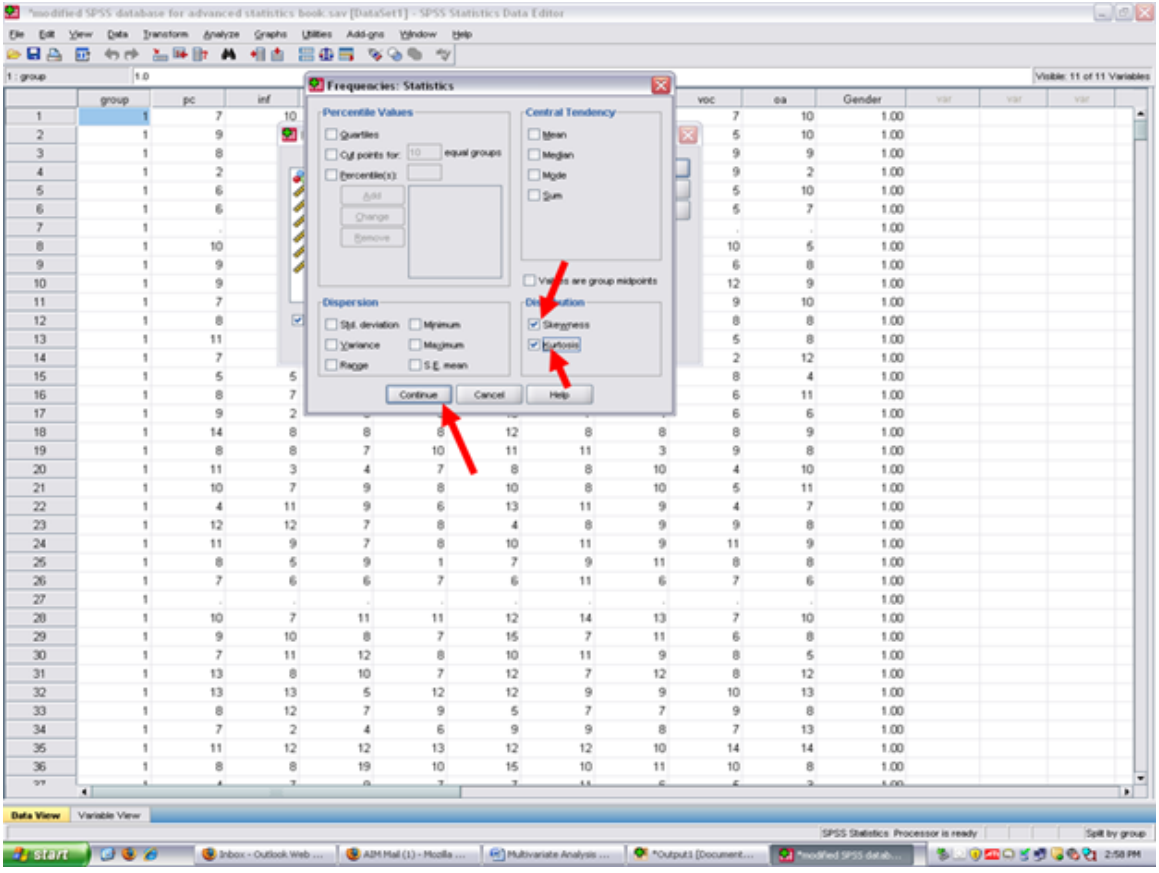
* Kurtosis [Note. Kurtosis also refers to the extent to which the data are normally distributed around the mean. This time, the data are piled up higher than normal around the mean or piled up higher than normal at the ends of the distribution.] Readers are referred to the following sources for a more detailed definition of kurtosis: http://www.statistics.com/index.php?page=glossary&term_id=326[27] 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.

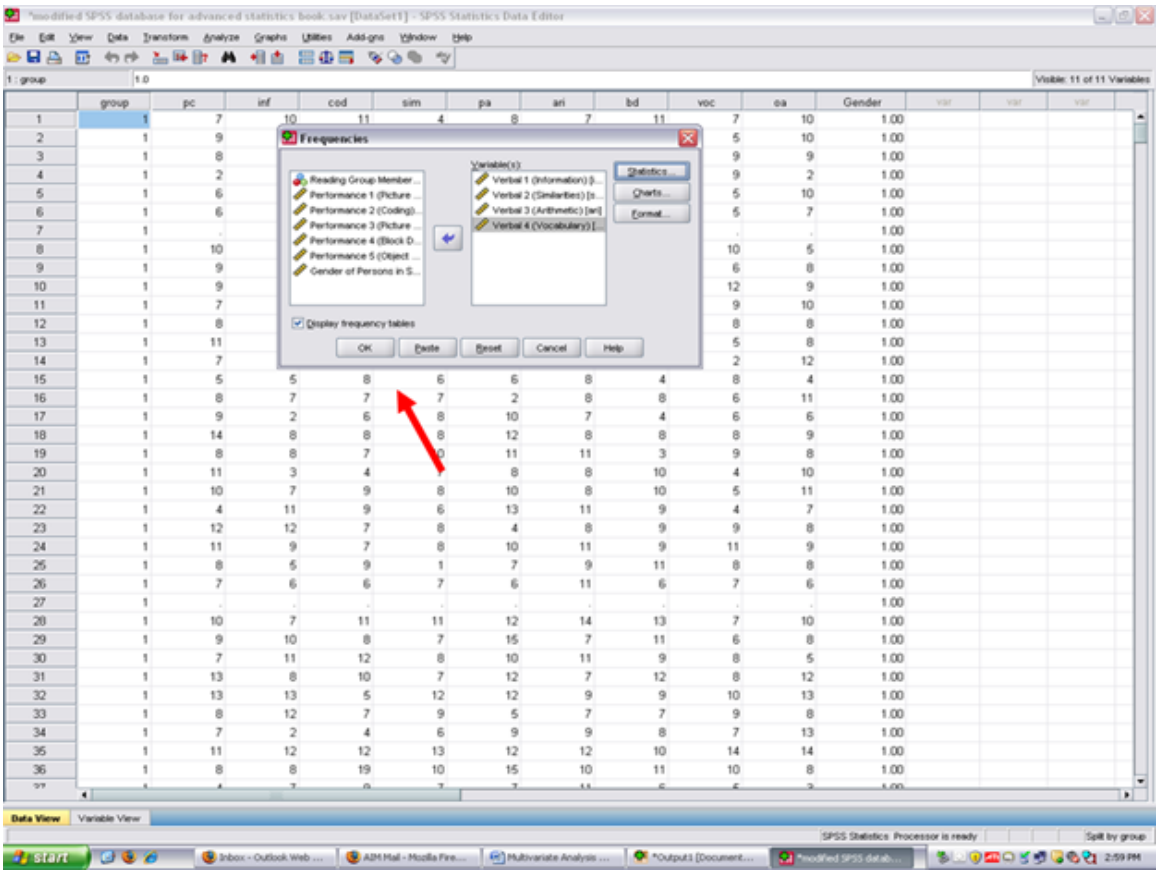
Click on Skewness, Kurtosis, and Continue



[29]

Now click on OK.

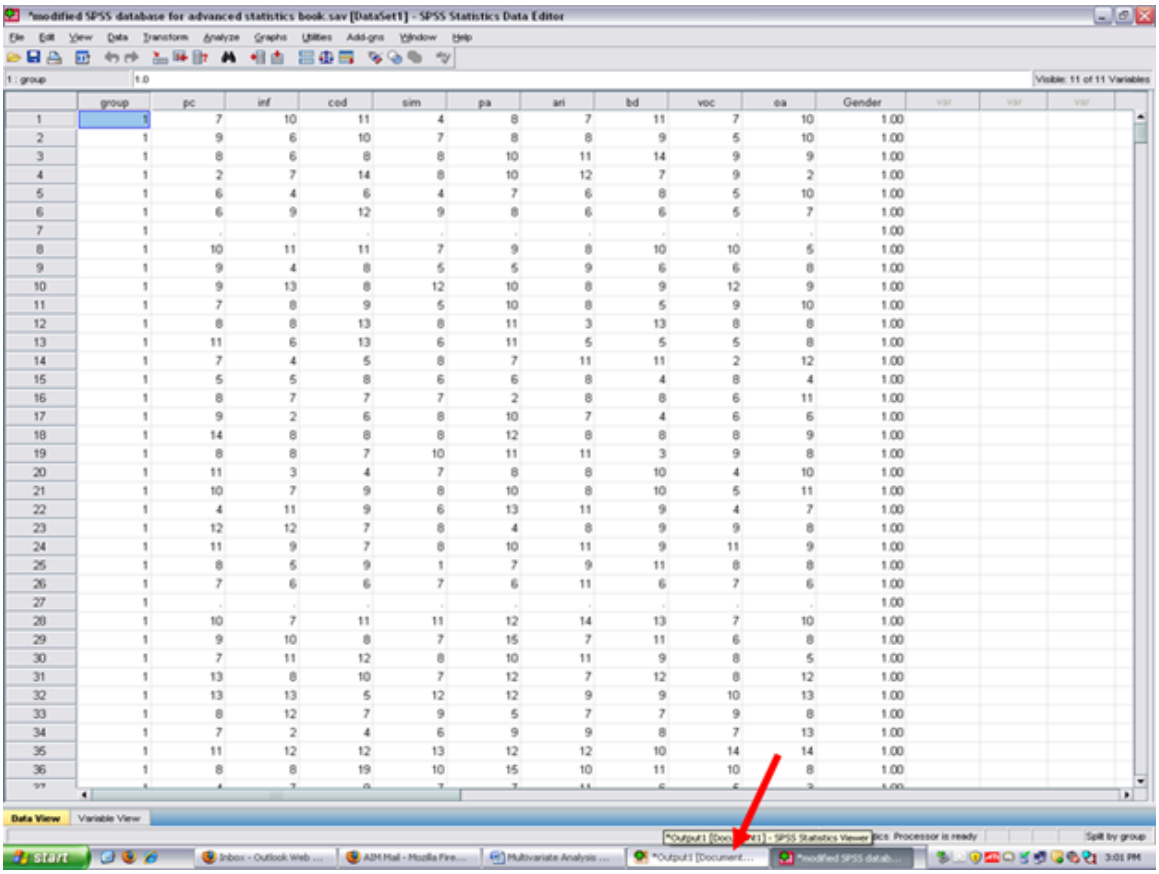
[28] http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb
[29] http://cnx.org/content/m40728/latest/10.19.png/image



[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.

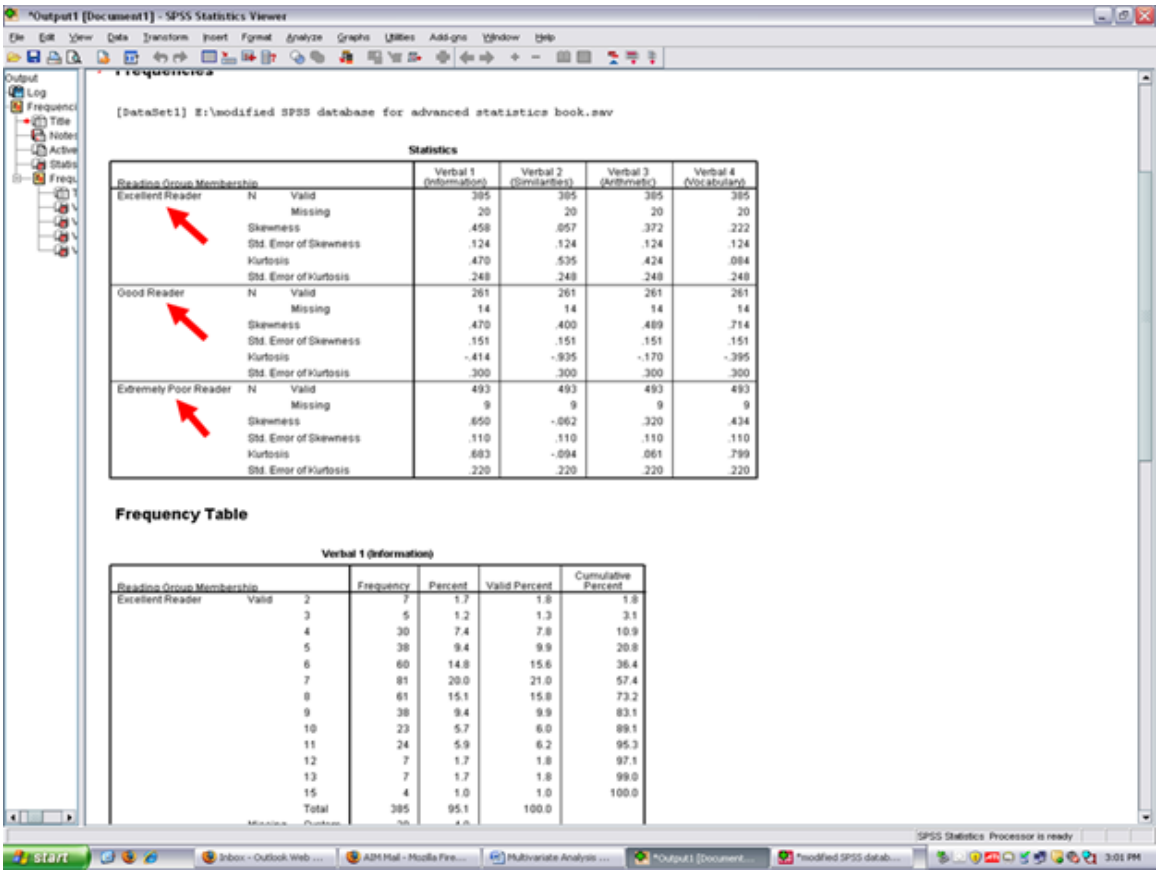
[30] http://cnx.org/content/m40728/latest/10.20.png/image



[31]

Clicking on the Output icon will result in the screen below being present. You will note that the Statistics table below contains statistics for the Excellent Reader group; for the Good Reader group; and for the Extremely Poor Reader group.

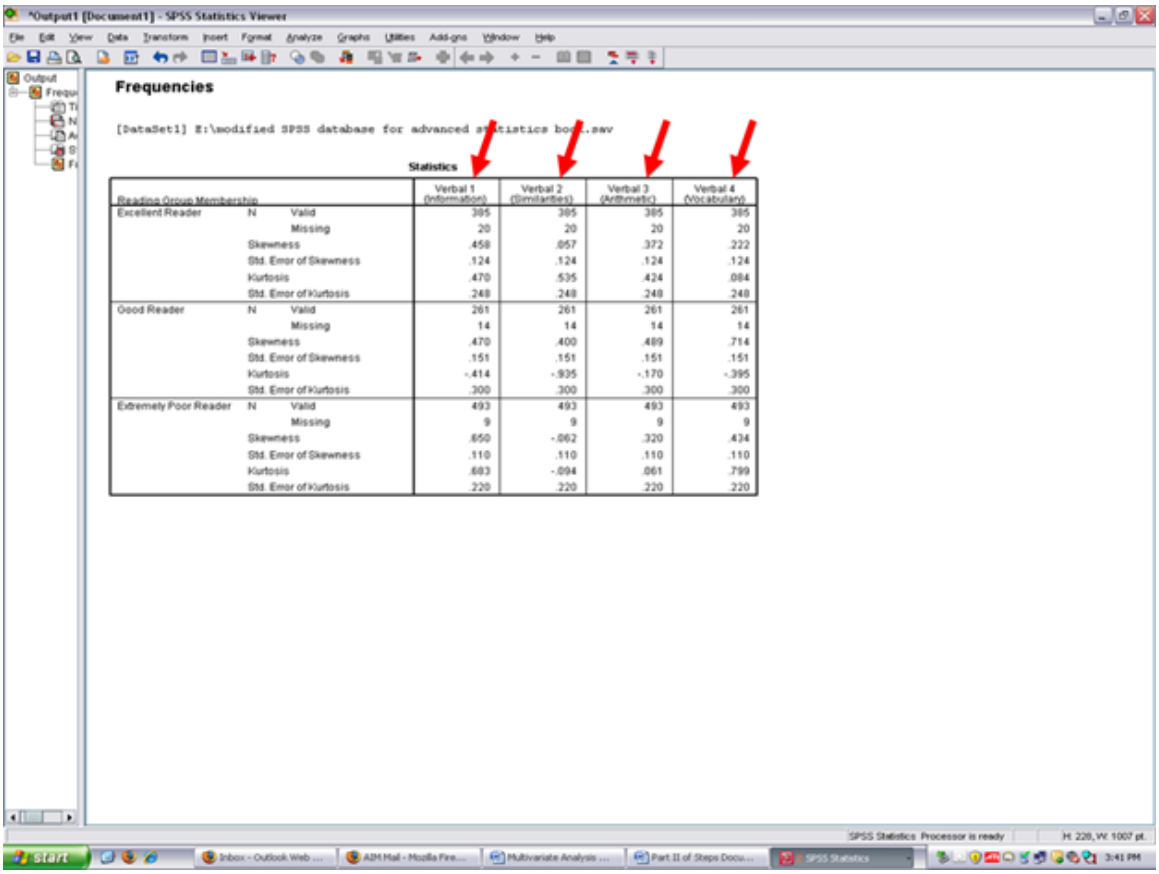
[31] http://cnx.org/content/m40728/latest/10.21.png/image



[32]

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

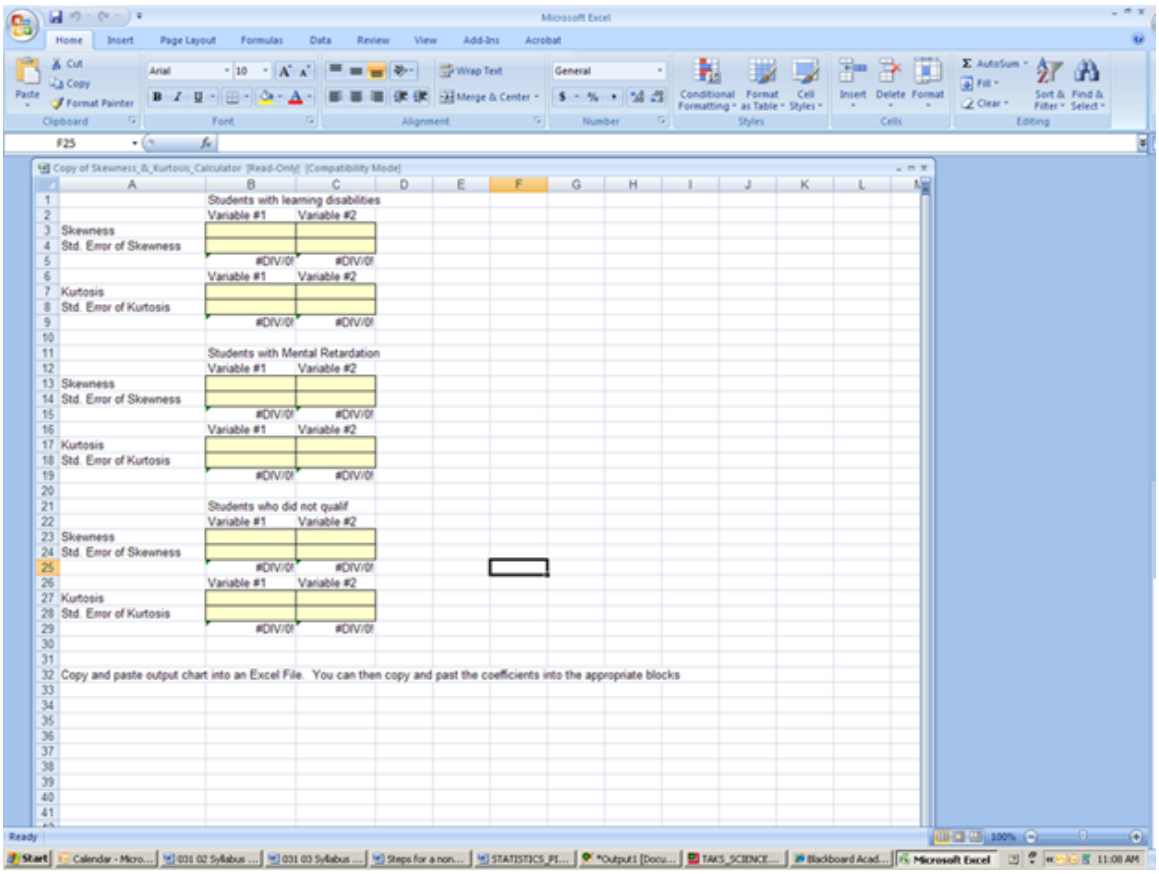
[32] http://cnx.org/content/m40728/latest/10.22.png/image



[33]

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]

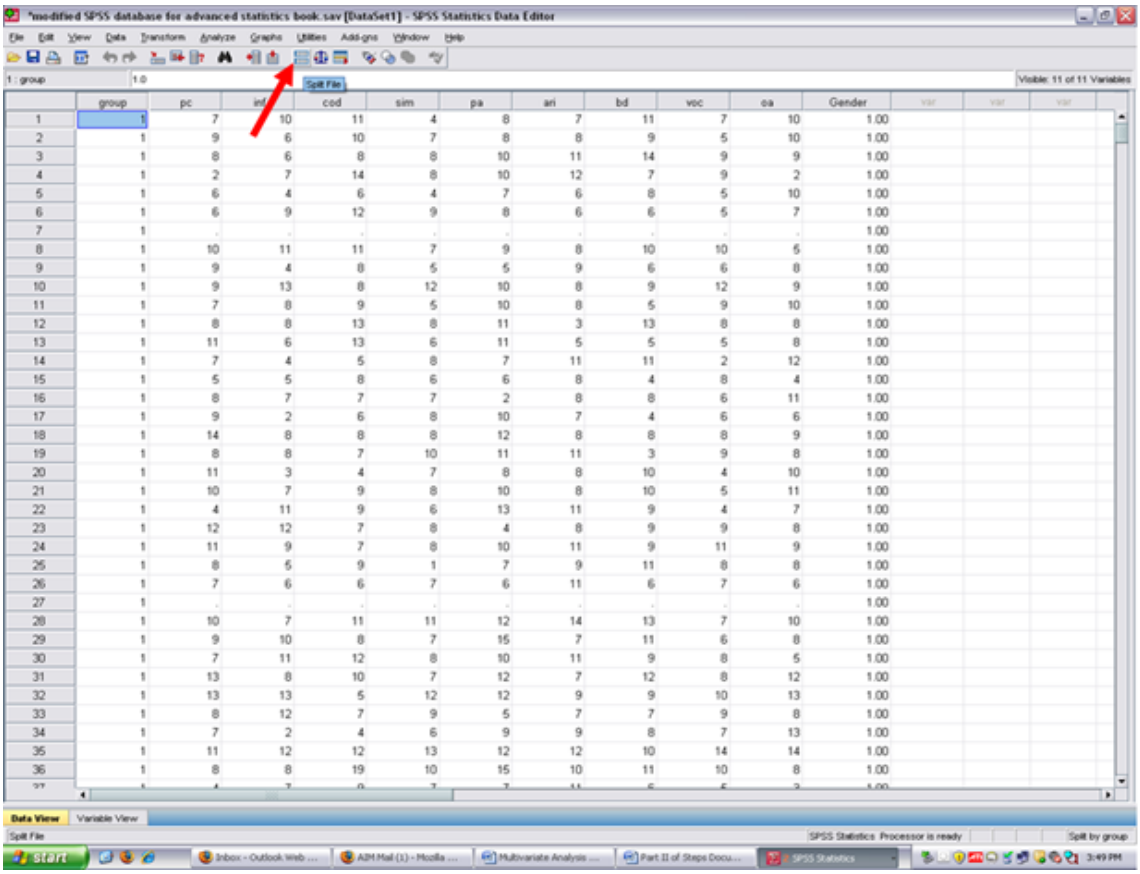
After calculating the standardized coefficients, record the number that are within +/-3 and the number that are outside of these boundaries. This information needs to be discussed in your results section where you document your checks of the assumptions underlying this particular statistical technique.

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



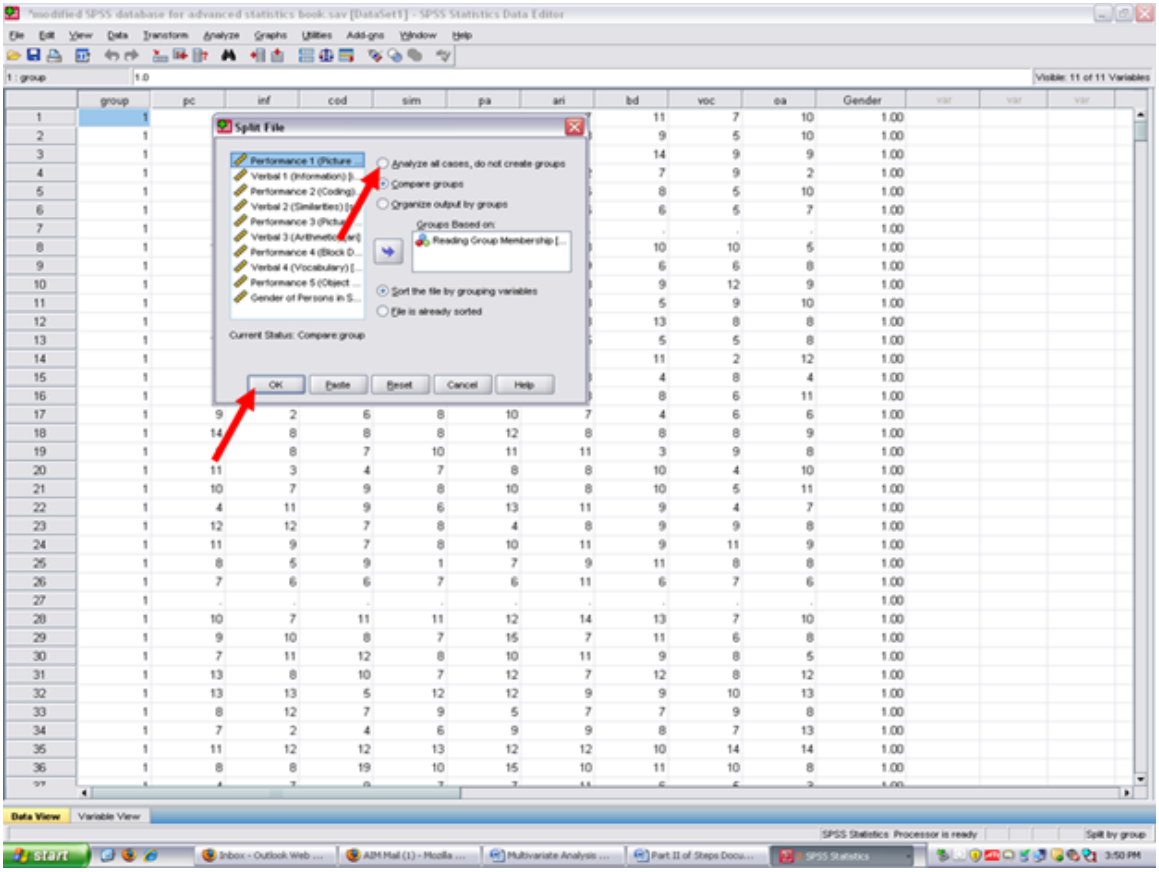
[35]

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

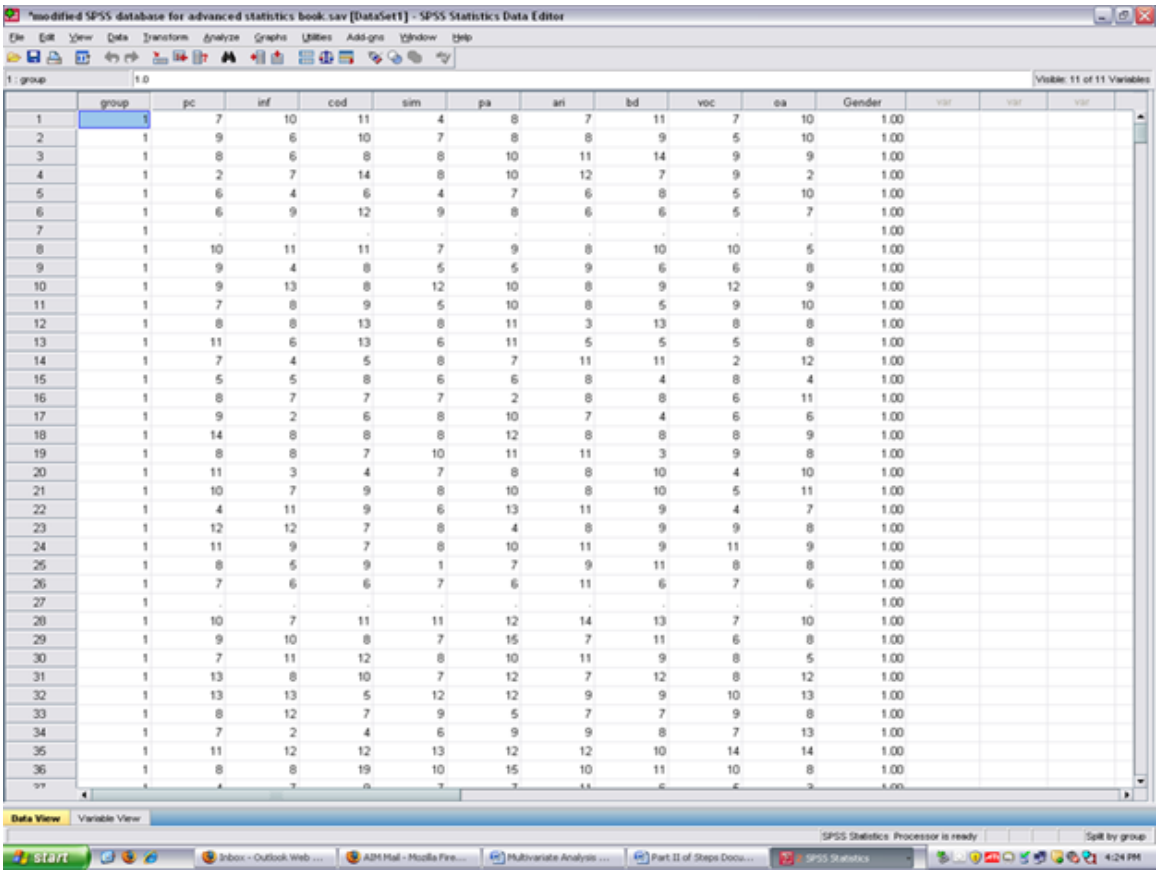
[35] http://cnx.org/content/m40728/latest/10.25.png/image



[36]

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

[36] http://cnx.org/content/m40728/latest/10.26.png/image



[37]

[37] http://cnx.org/content/m40728/latest/10.27.png/image

Chapter 12

12. Multivariate Analysis of Variance: Part II[1]

NOTE: This chapter is published by NCPEA Press[2] 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

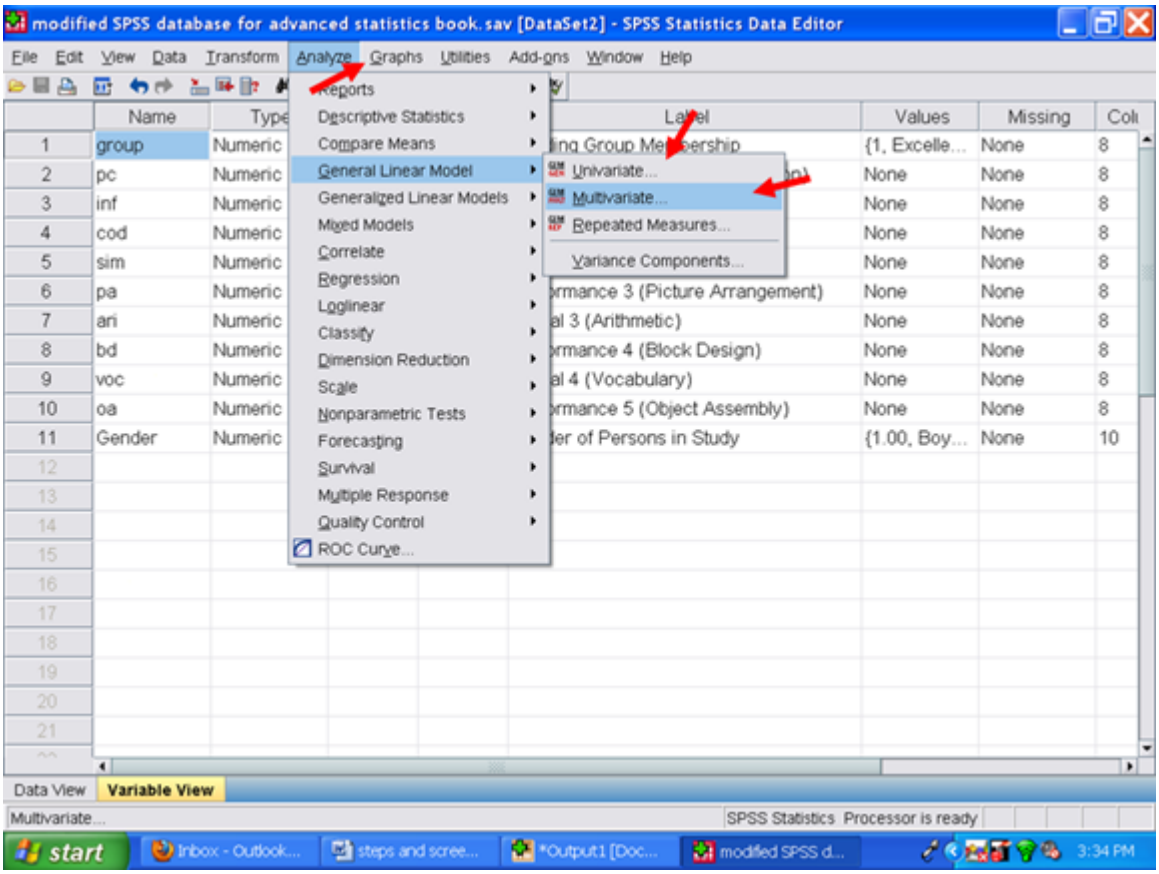
John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, 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 Help[3]) 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**

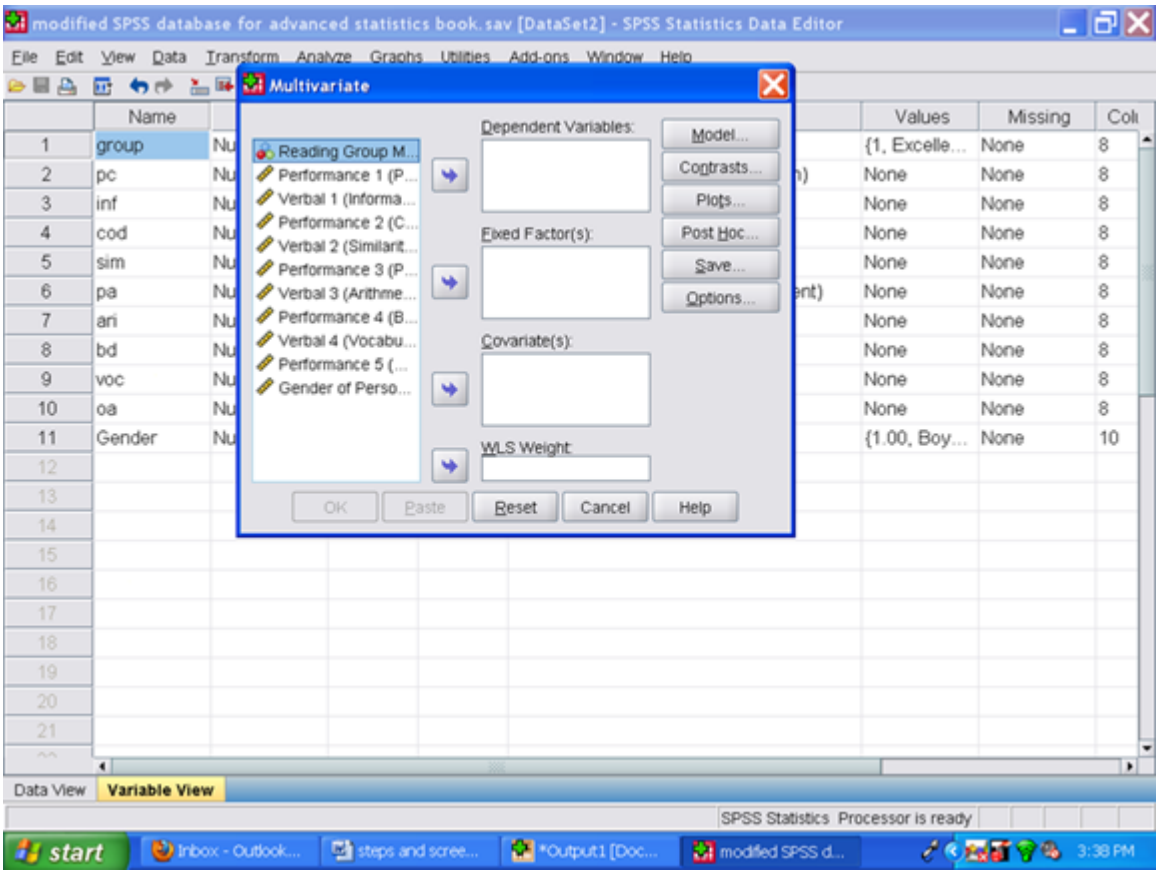
[1] This content is available online at <http://cnx.org/content/m40729/1.2/>.
[2] http://www.ncpeapublications.org/books.html
[3] http://cnx.org/content/m40729/latest/www.writingandstatisticalhelp



[4]

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.

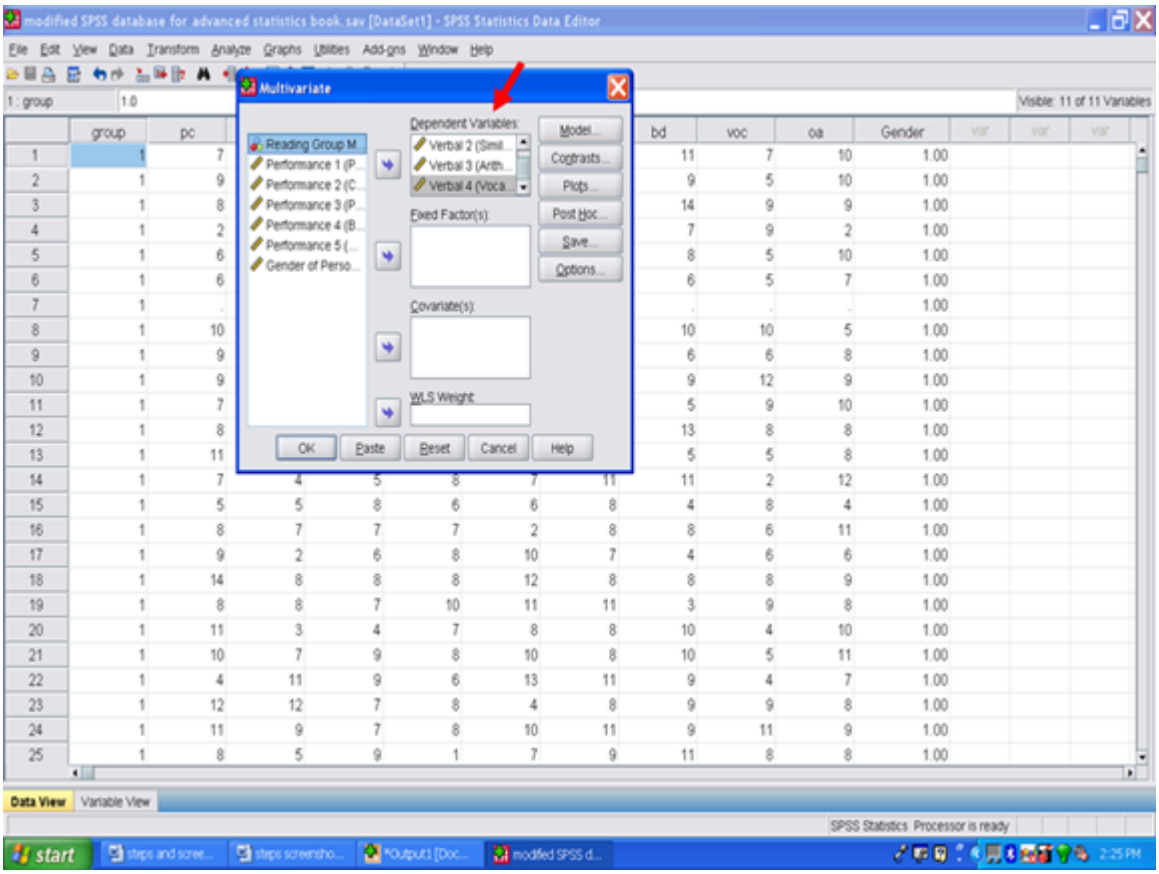
[4] http://cnx.org/content/m40729/latest/11.1.png/image



[5]

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

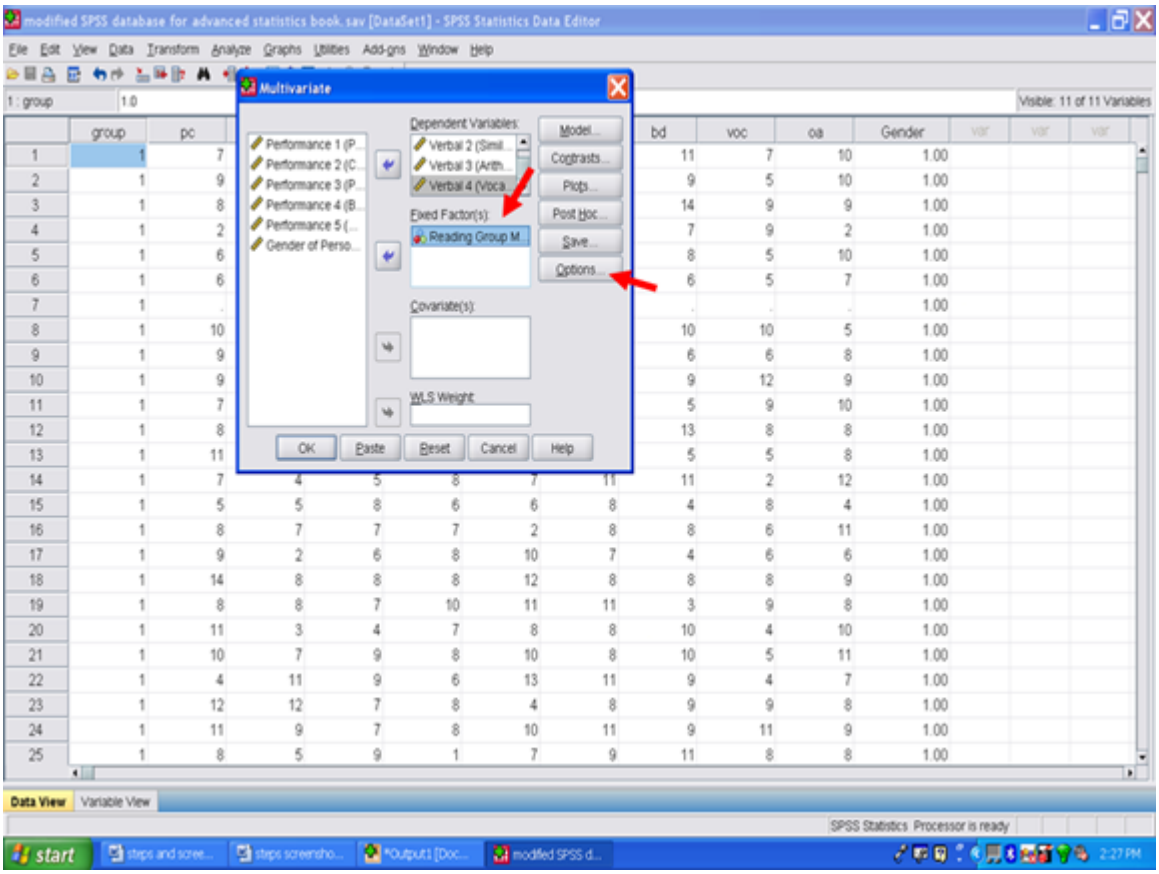
[5] http://cnx.org/content/m40729/latest/11.2.png/image



6

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



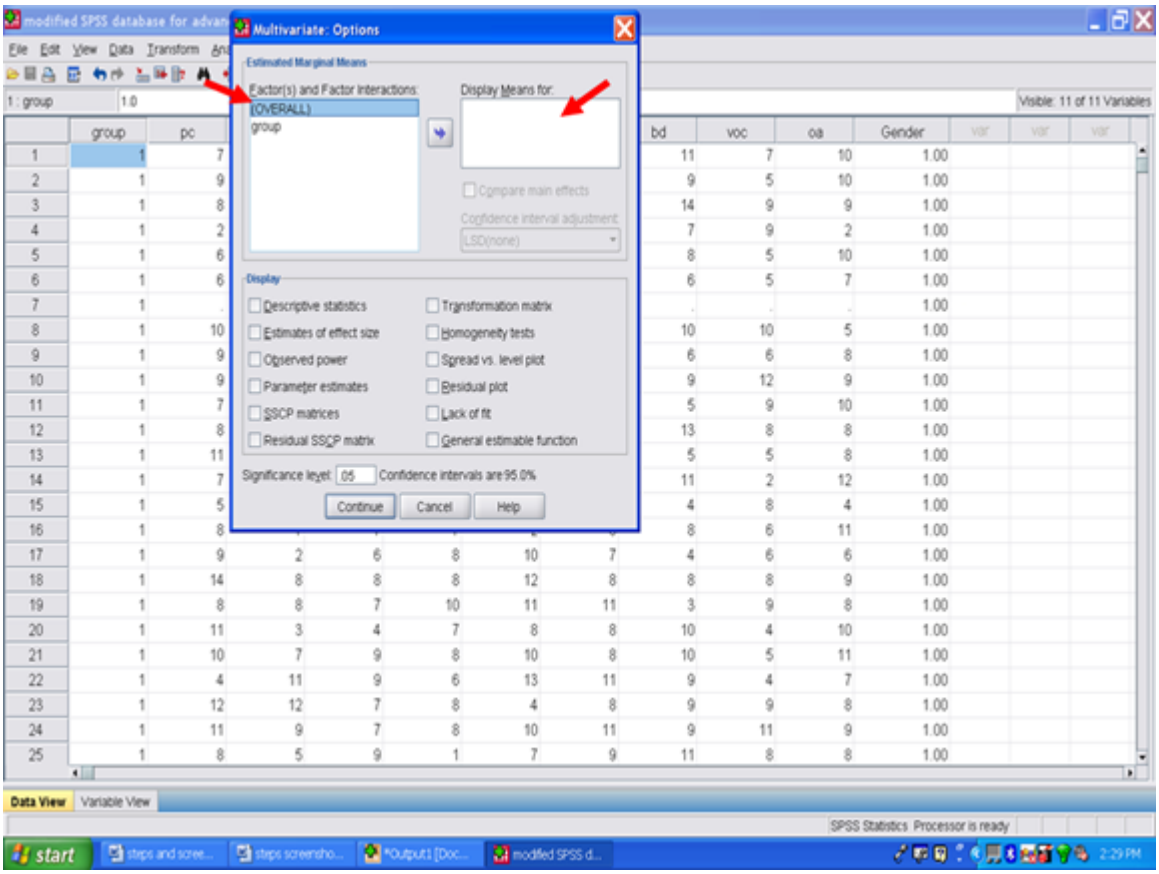
[7]

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.

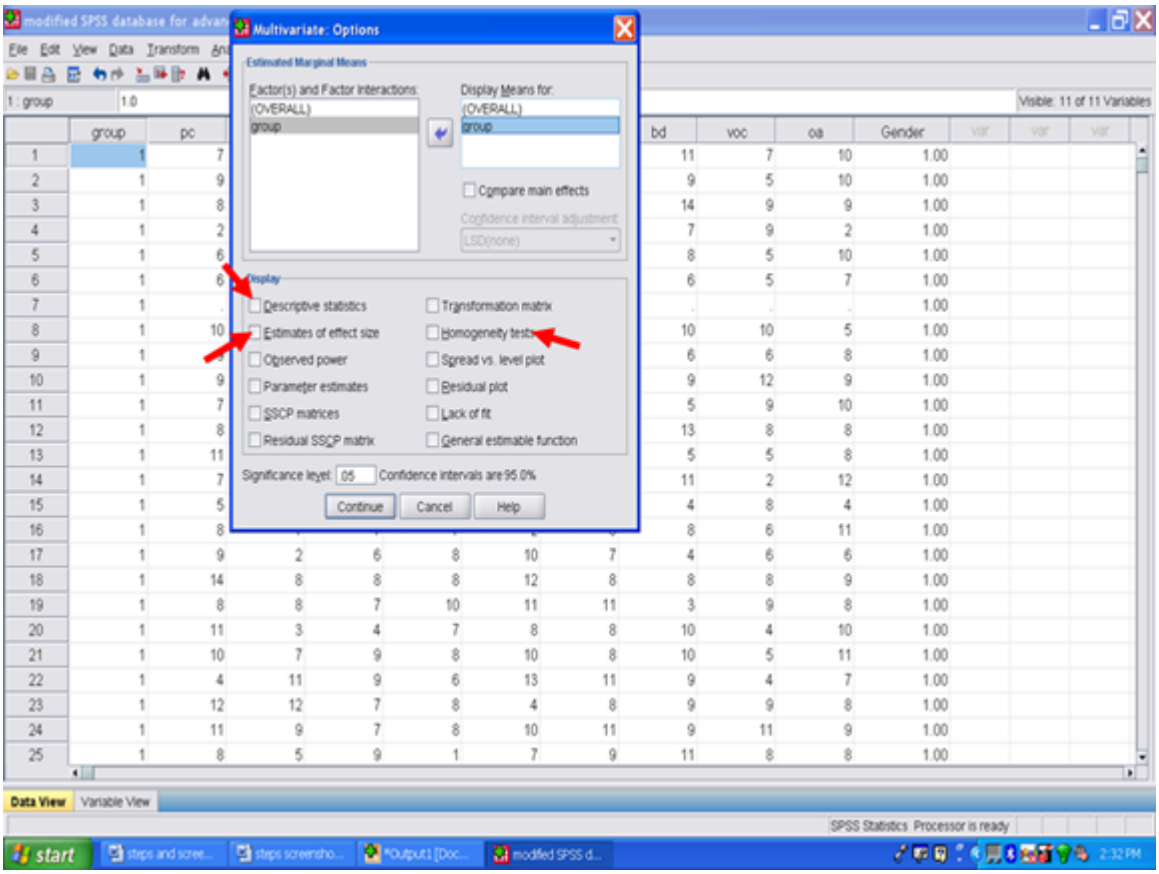
[7] http://cnx.org/content/m40729/latest/11.4.png/image



[8]

To obtain the information just mentioned, you will need to click on:
Descriptive Statistics
Estimates of Effect Size
Homogeneity tests

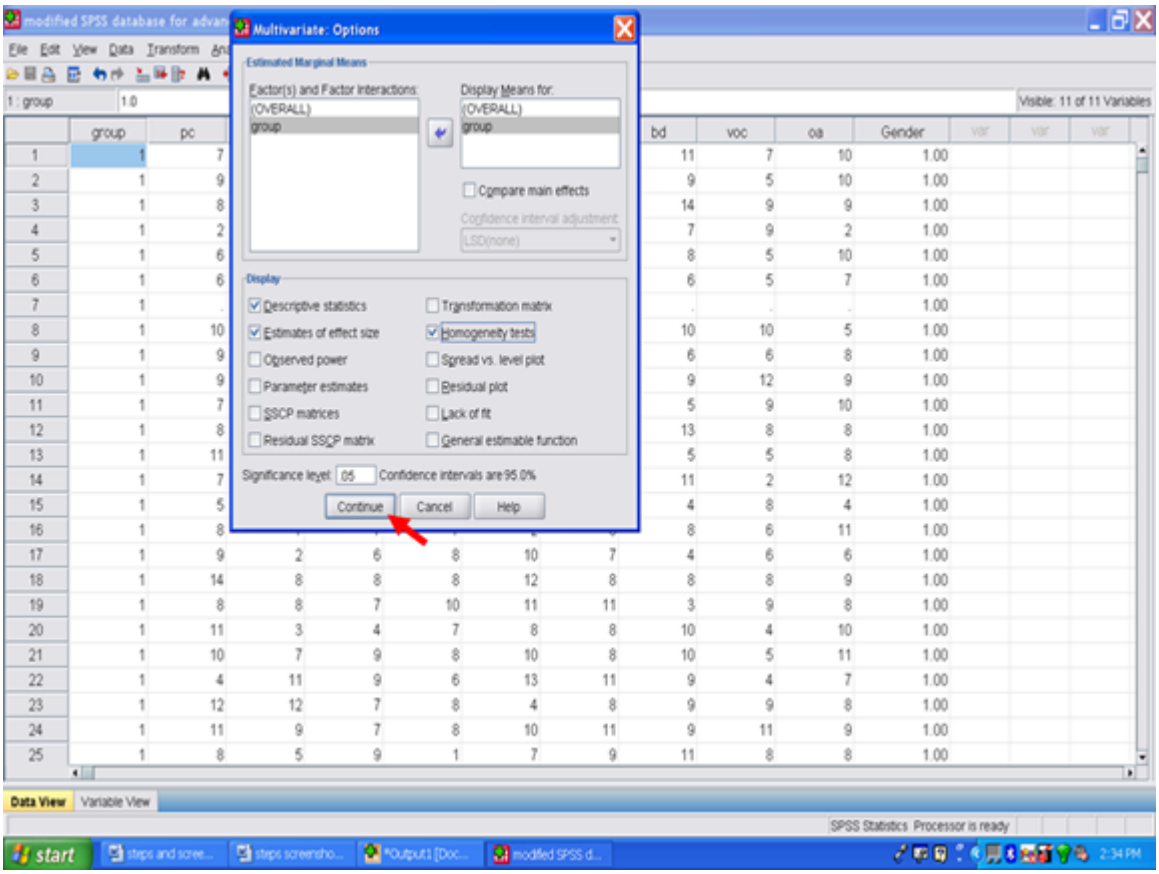
[8] http://cnx.org/content/m40729/latest/11.5.png/image



[9]

After checking the three boxes mentioned, then click on Continue

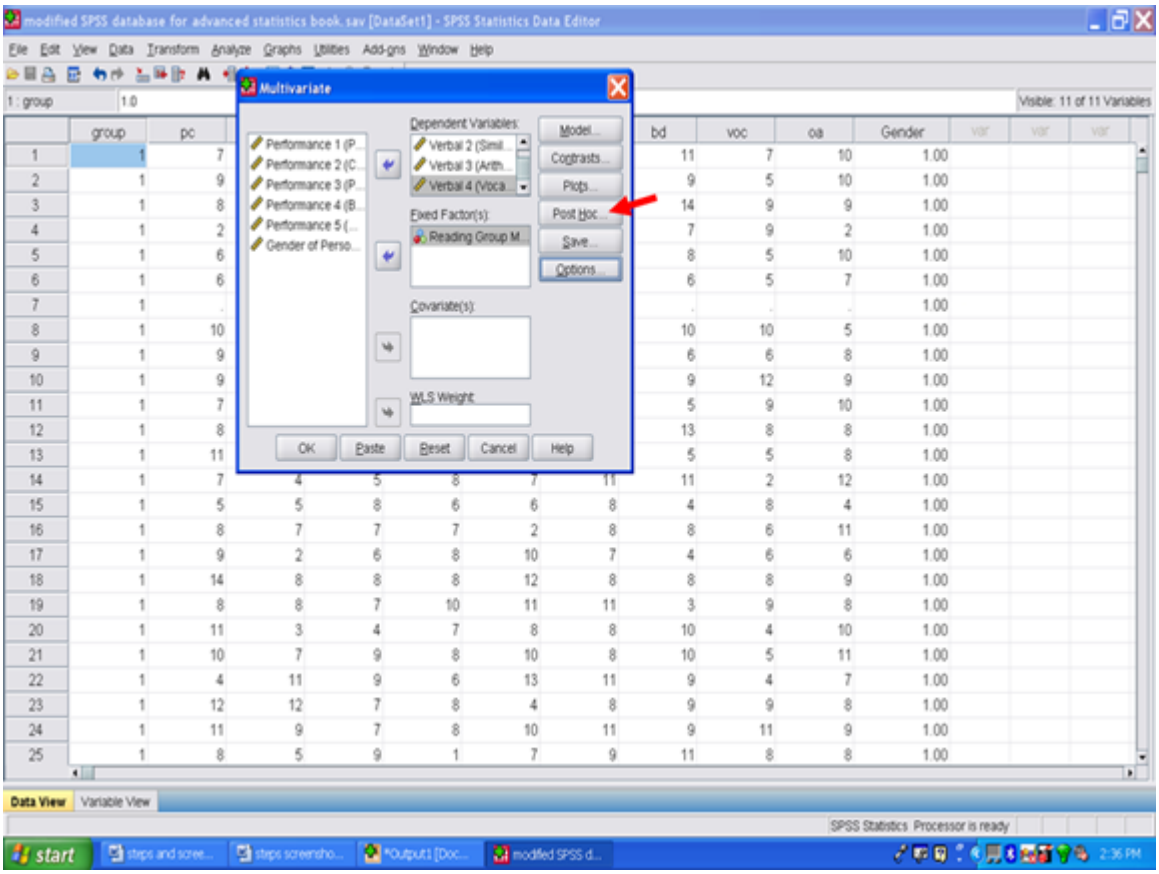
[9] http://cnx.org/content/m40729/latest/11.6.png/image



[10]

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

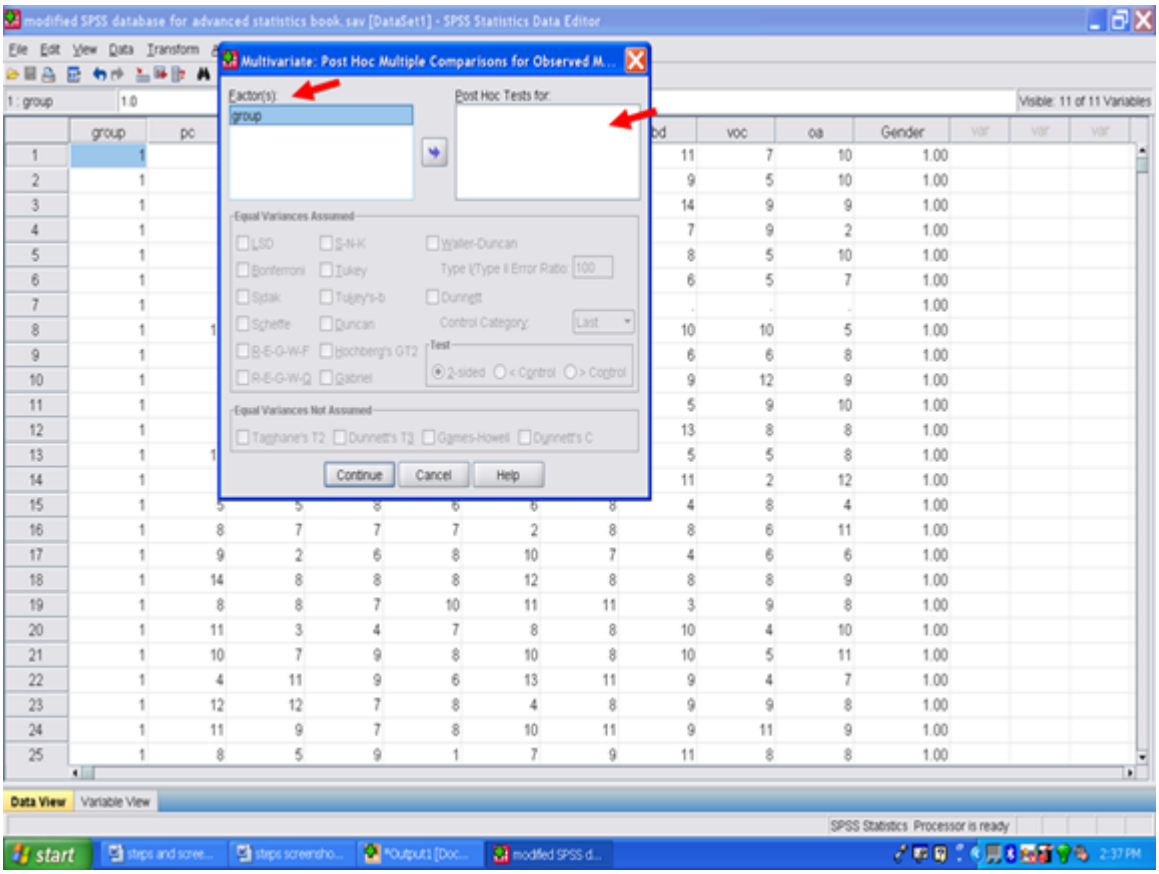
[10] http://cnx.org/content/m40729/latest/11.7.png/image



[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:

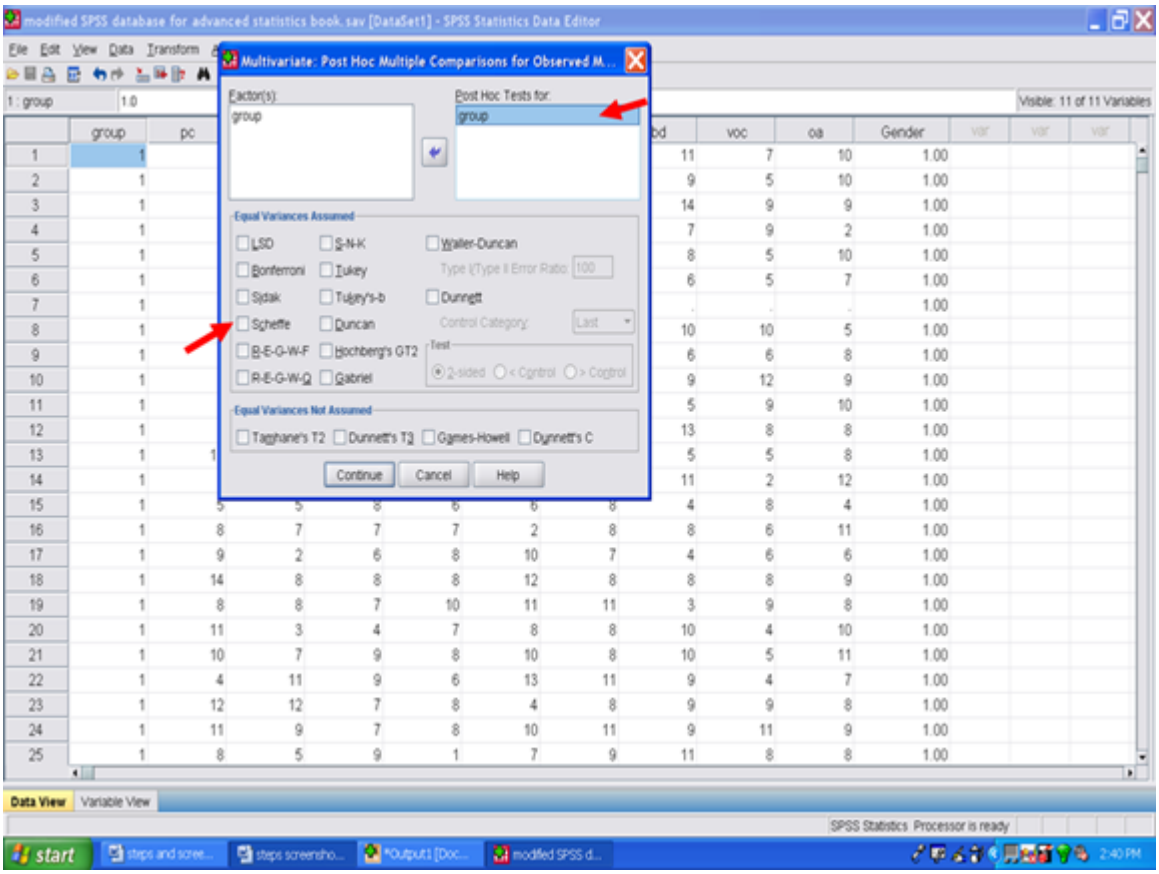
[11] http://cnx.org/content/m40729/latest/11.8.png/image



[12]

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.

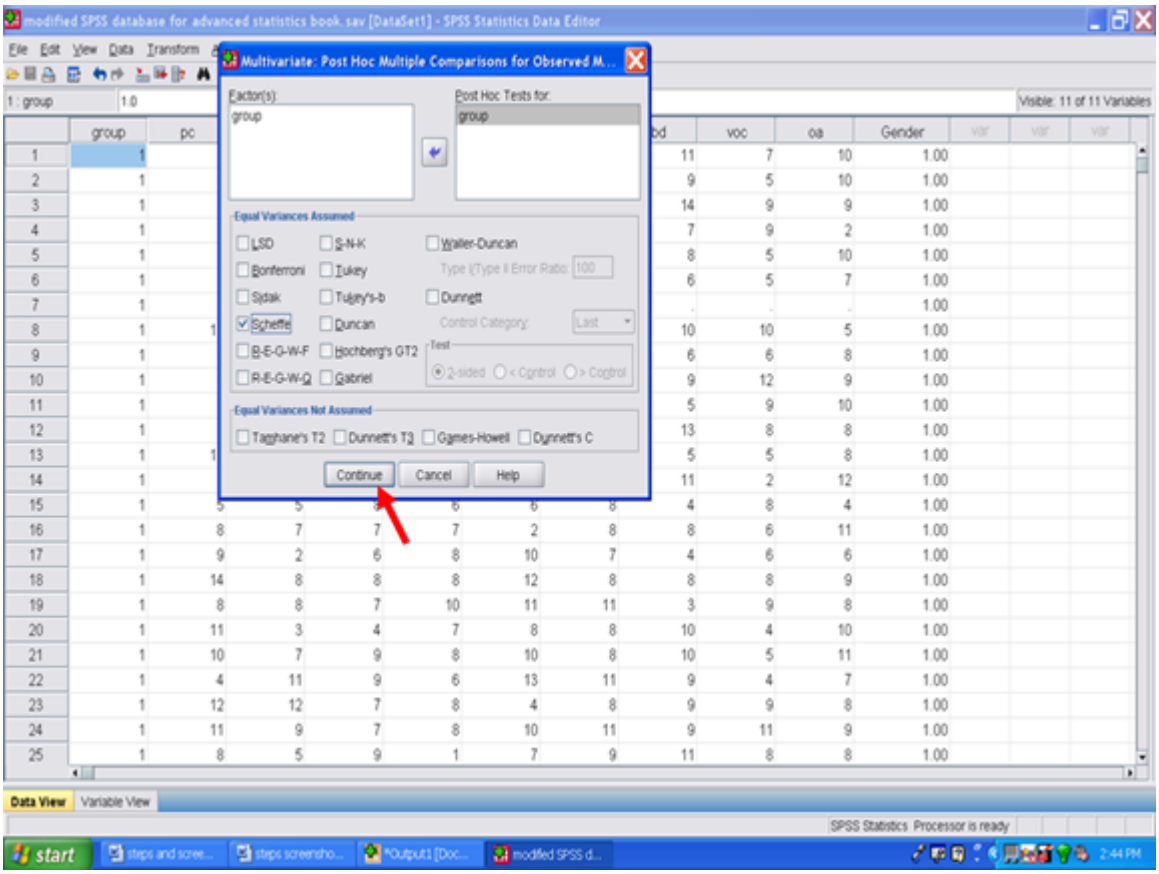
[12] http://cnx.org/content/m40729/latest/11.9.png/image



13

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

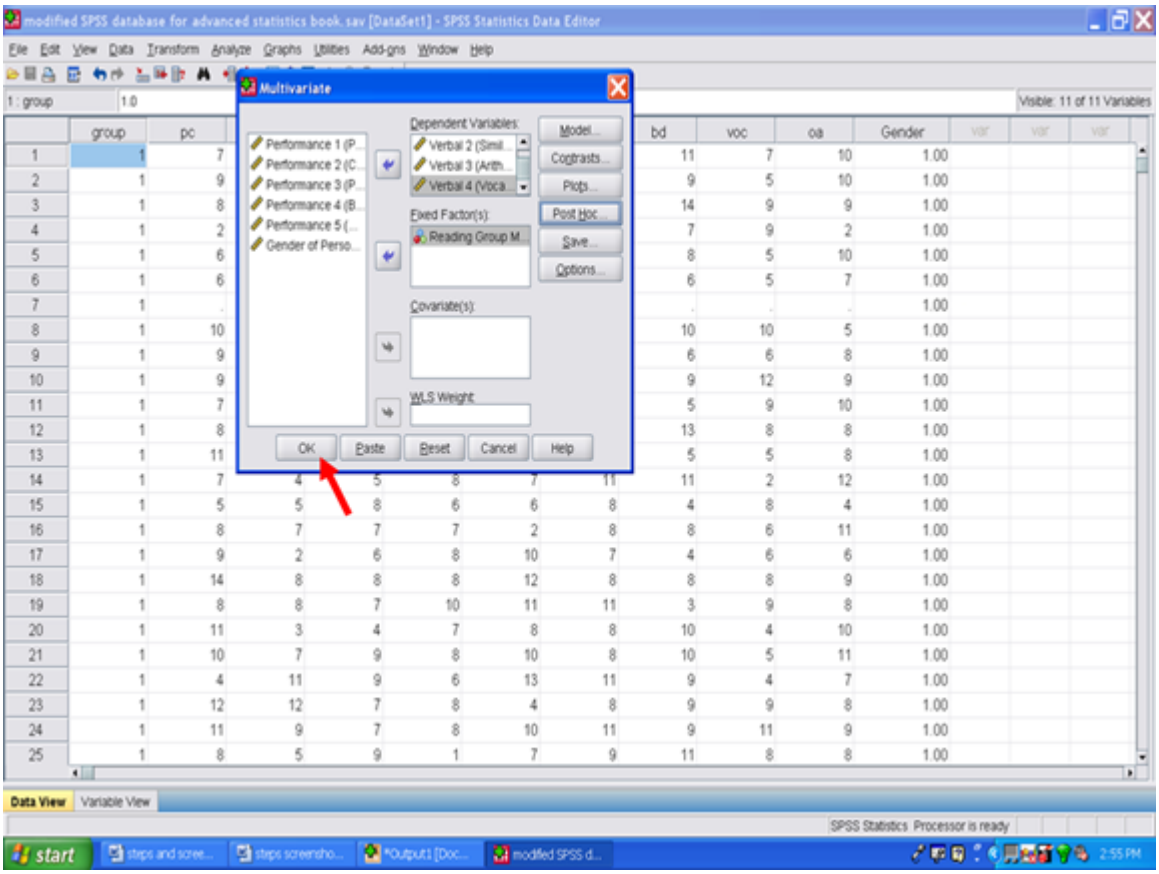
[13] http://cnx.org/content/m40729/latest/11.10.png/image



14

Then click on OK.

[14] http://cnx.org/content/m40729/latest/11.11.png/image



[15]

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 III[1]

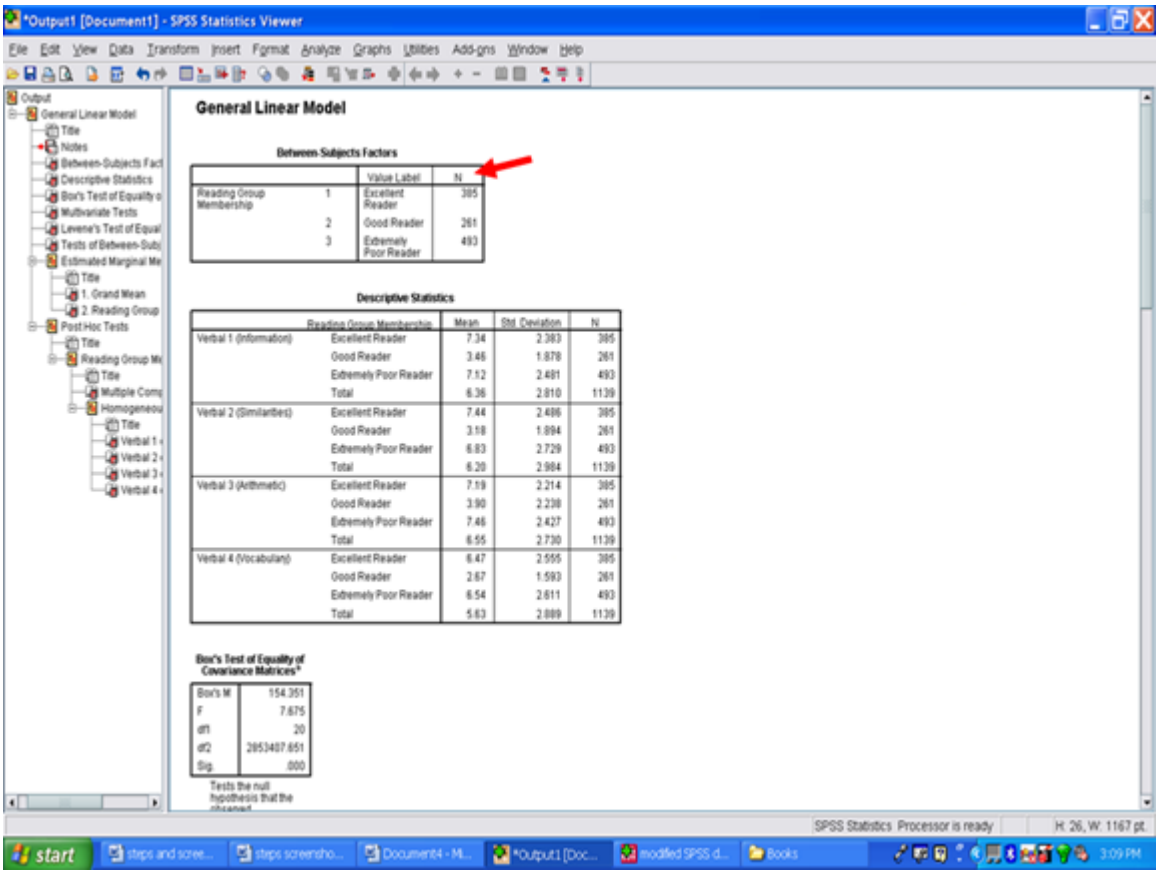
NOTE: This chapter is published by NCPEA Press[2] and is presented as an NCPEA/Connexions publication "print on demand book." Each chapter has been peer-reviewed, accepted, and endorsed by the National Council of Professors of Educational Administration (NCPEA) as a significant contribution to the scholarship 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 Help[3]) 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.

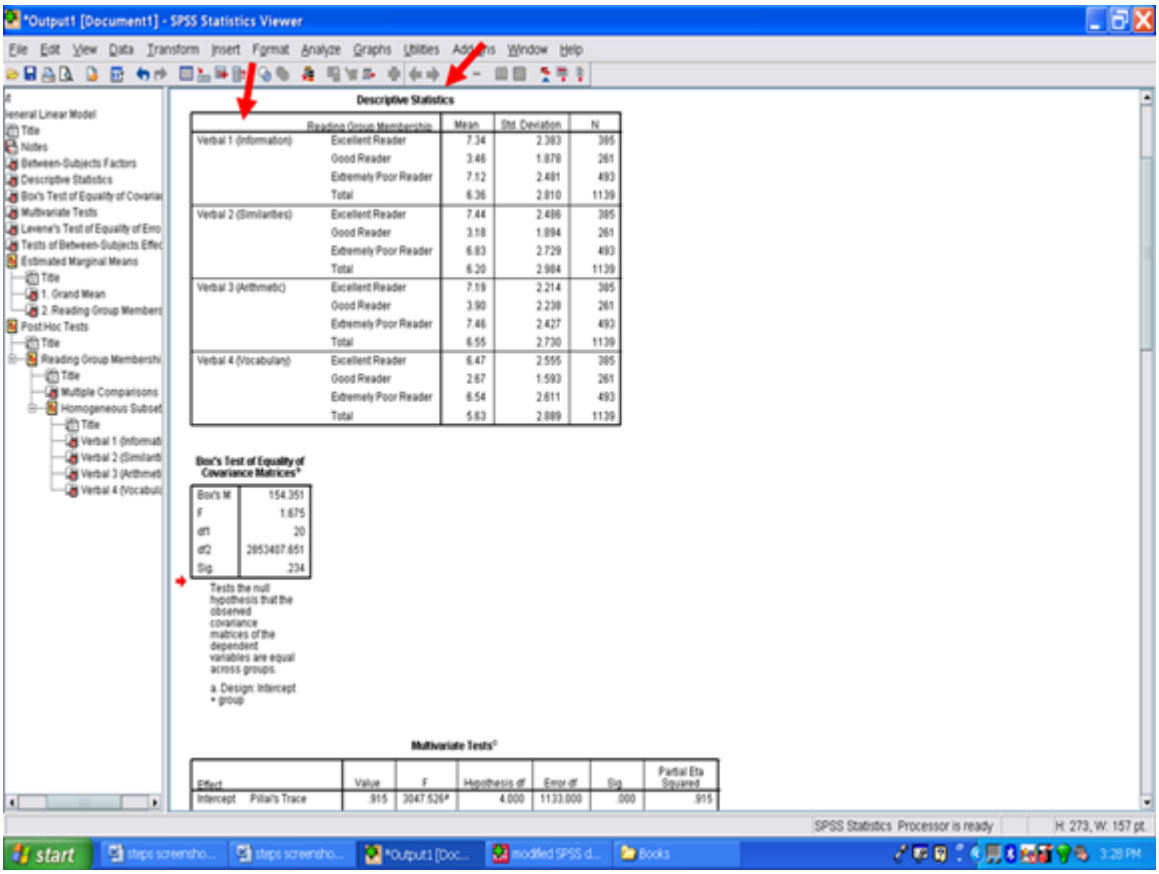
[1] This content is available online at <http://cnx.org/content/m40731/1.2/>.
[2] http://www.ncpeapublications.org/books.html
[3] http://cnx.org/content/m40731/latest/www.writingandstatisticalhelp



[4]

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.

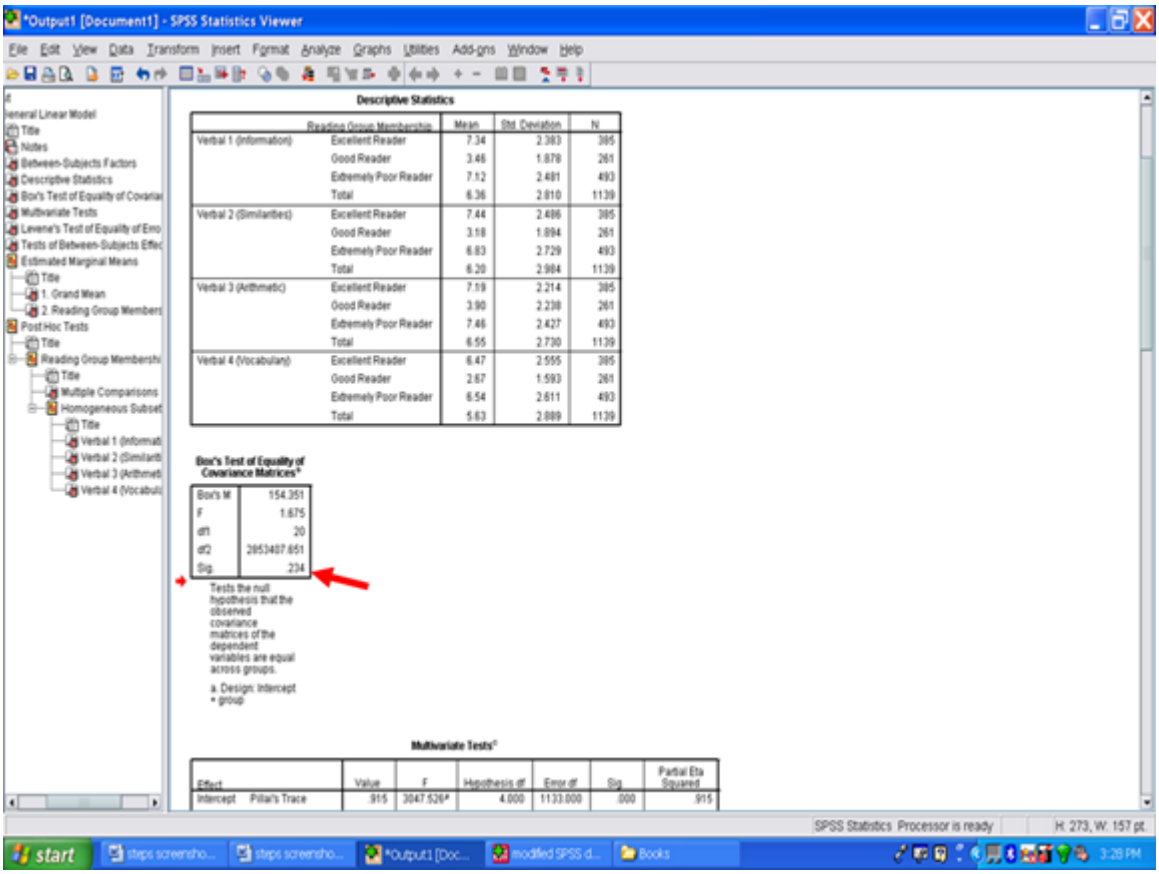
[4] http://cnx.org/content/m40731/latest/12.1.png/image



[5]

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.

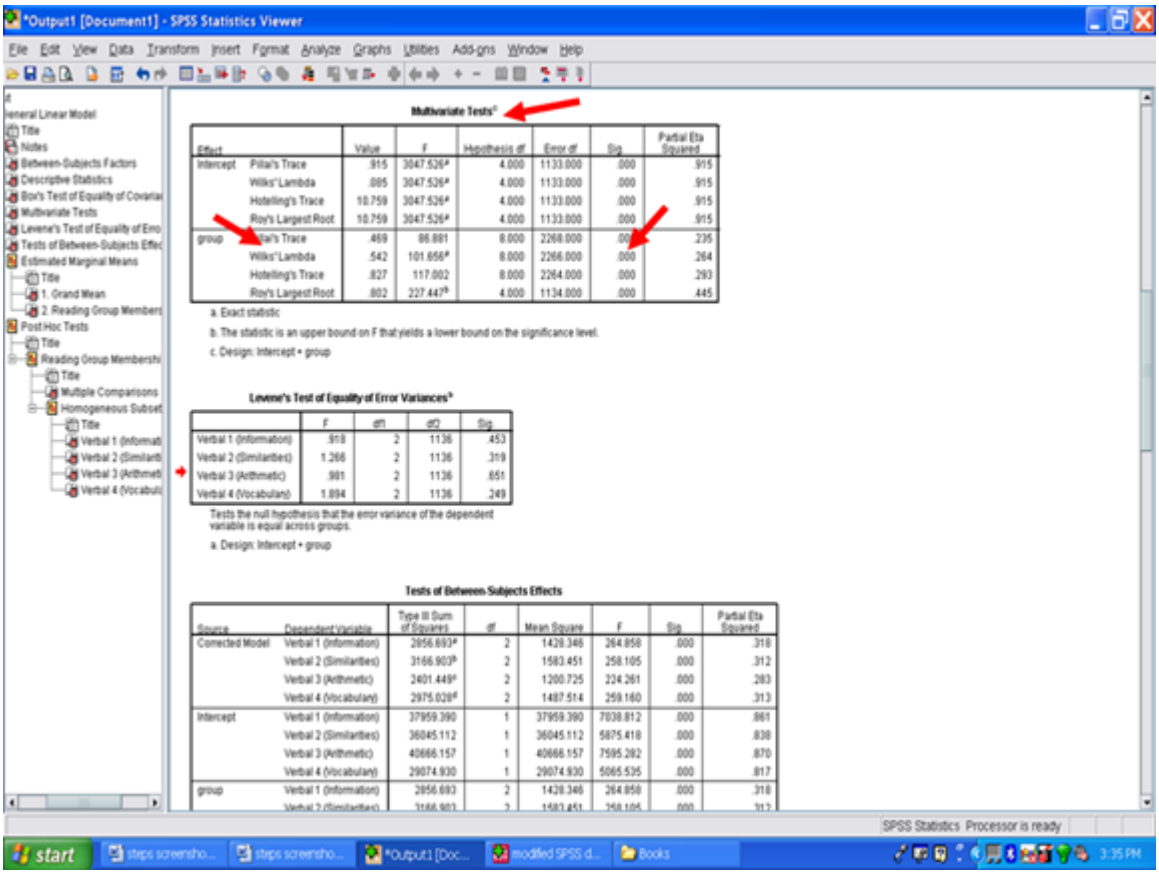
[5] http://cnx.org/content/m40731/latest/12.2.png/image



[6]

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.

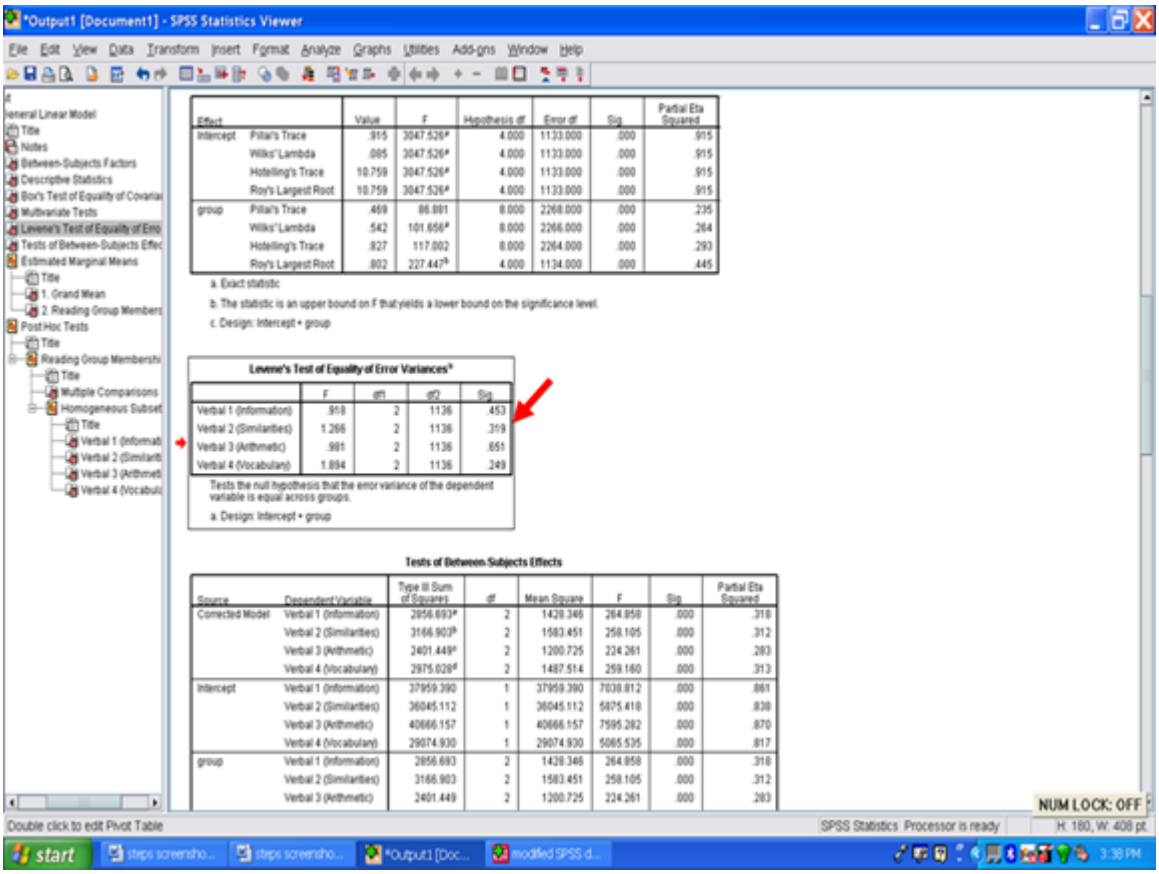
[6] http://cnx.org/content/m40731/latest/12.3.png/image



[7]

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.

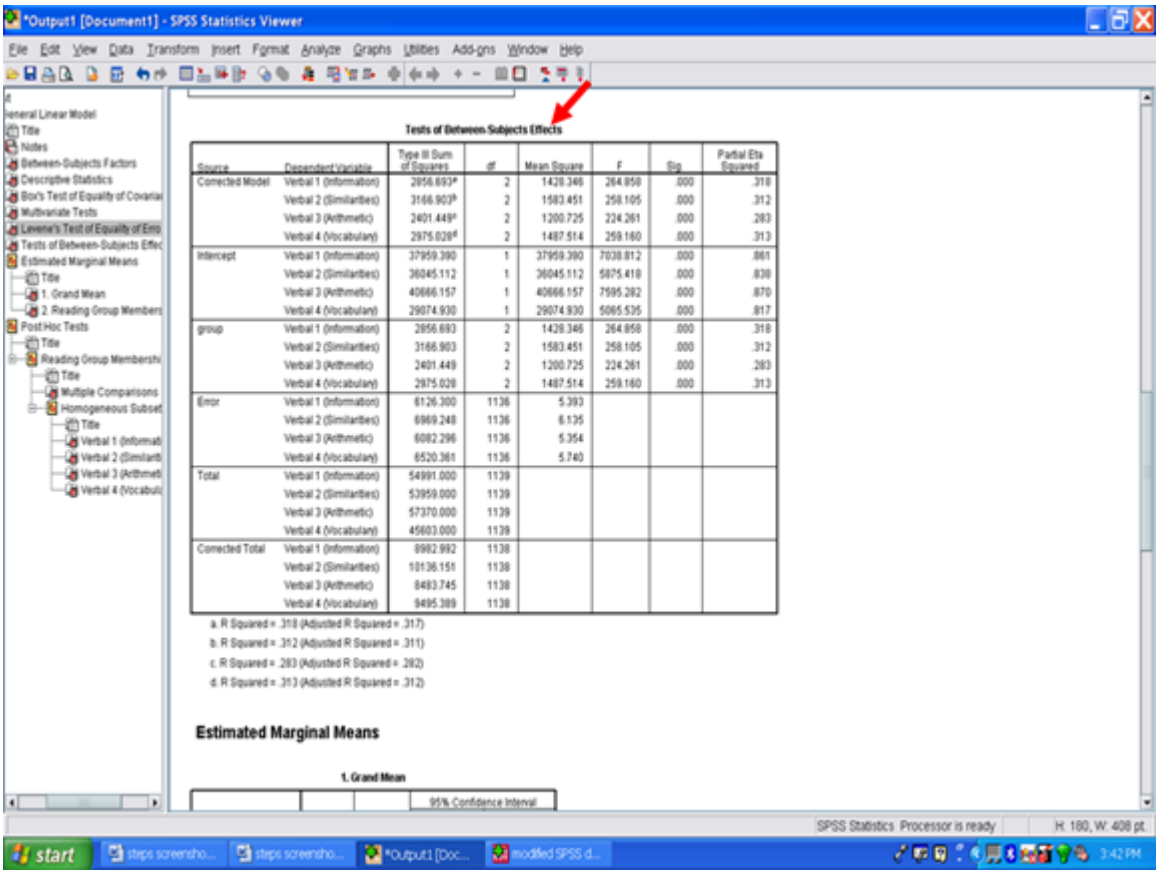
[7] http://cnx.org/content/m40731/latest/12.4.png/image



[8]

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.

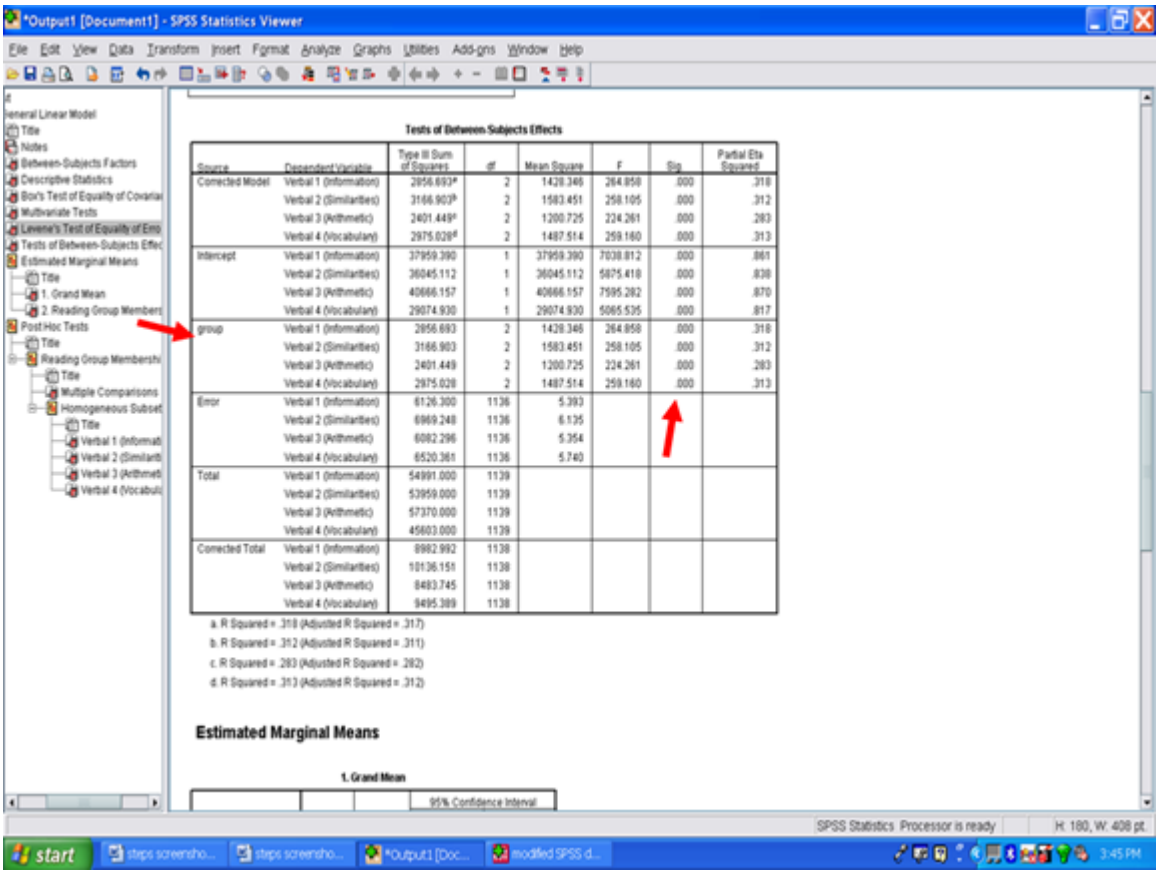
[8] http://cnx.org/content/m40731/latest/12.5.png/image



[9]

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.

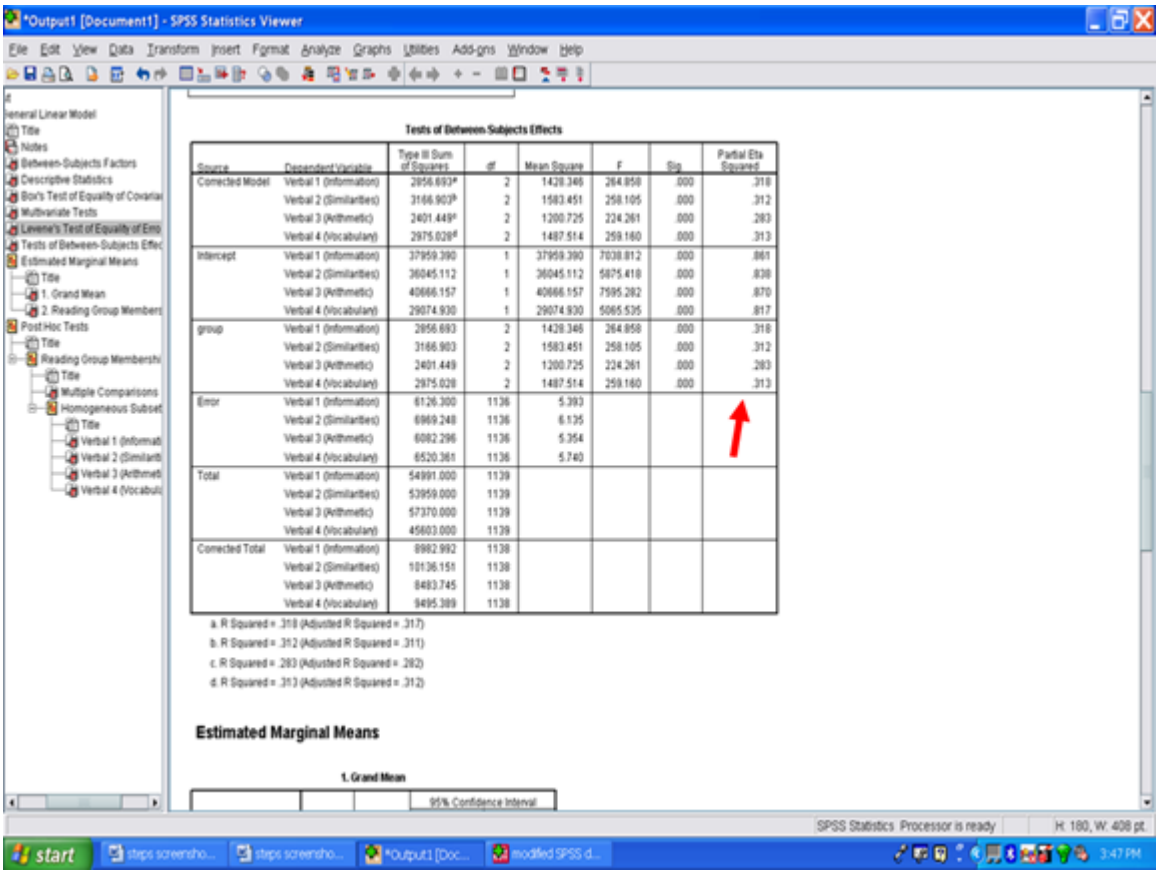
[9] http://cnx.org/content/m40731/latest/12.6.png/image



10

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 n^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).

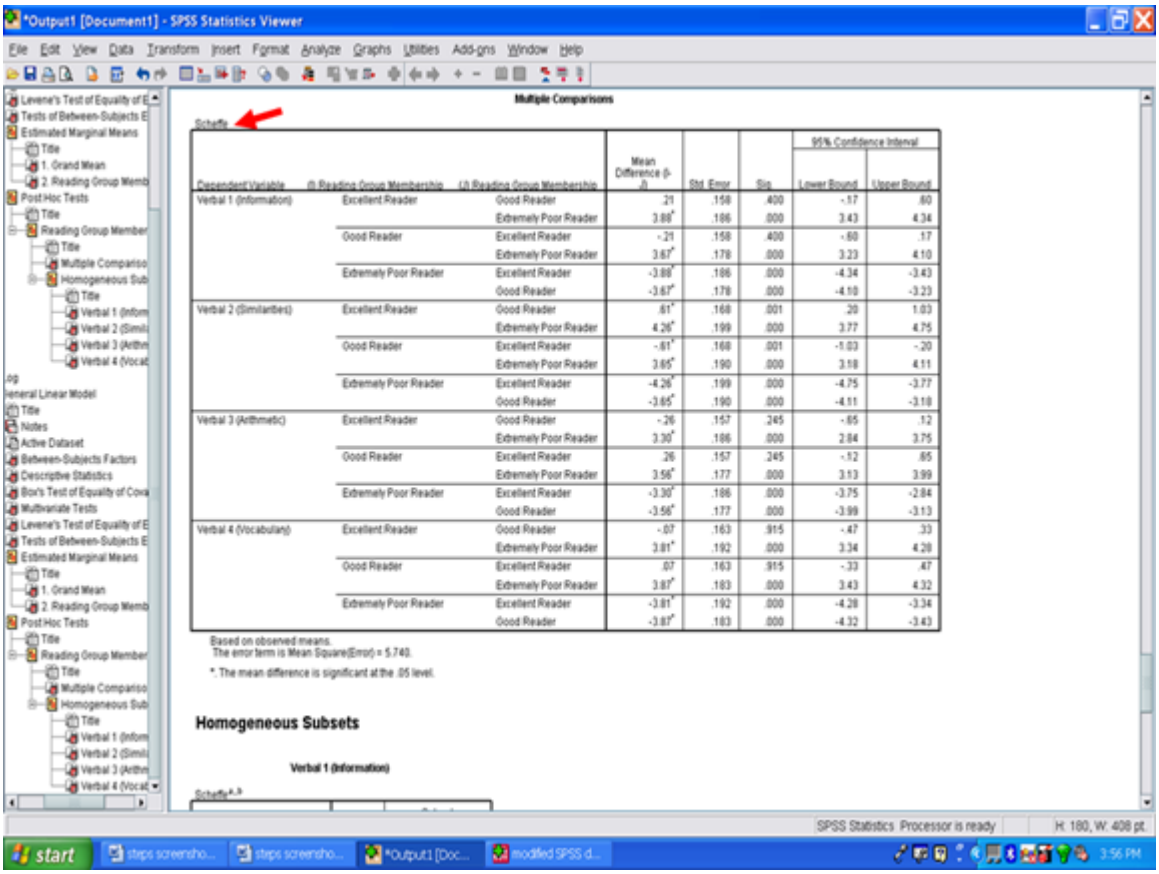
[10] http://cnx.org/content/m40731/latest/12.7.png/image



[11]

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.

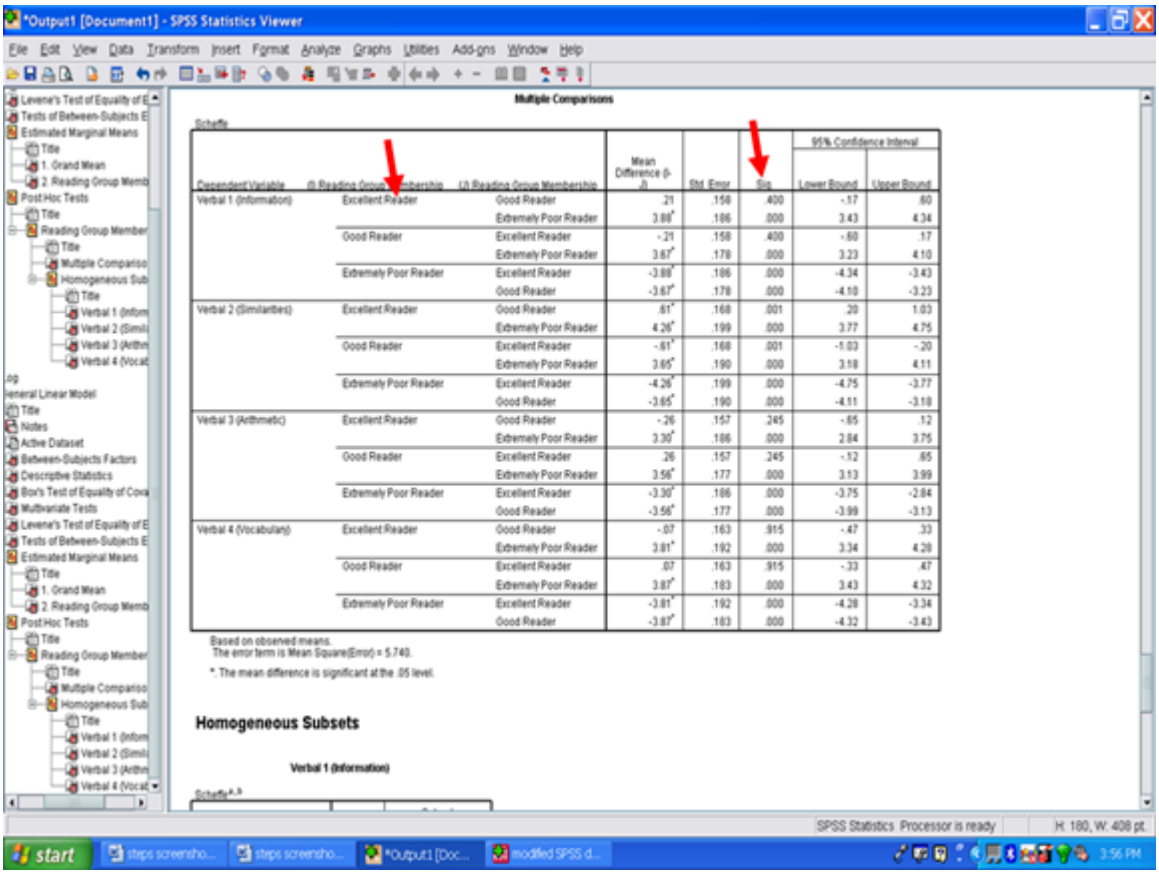
[11] http://cnx.org/content/m40731/latest/12.8.png/image



[12]

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.

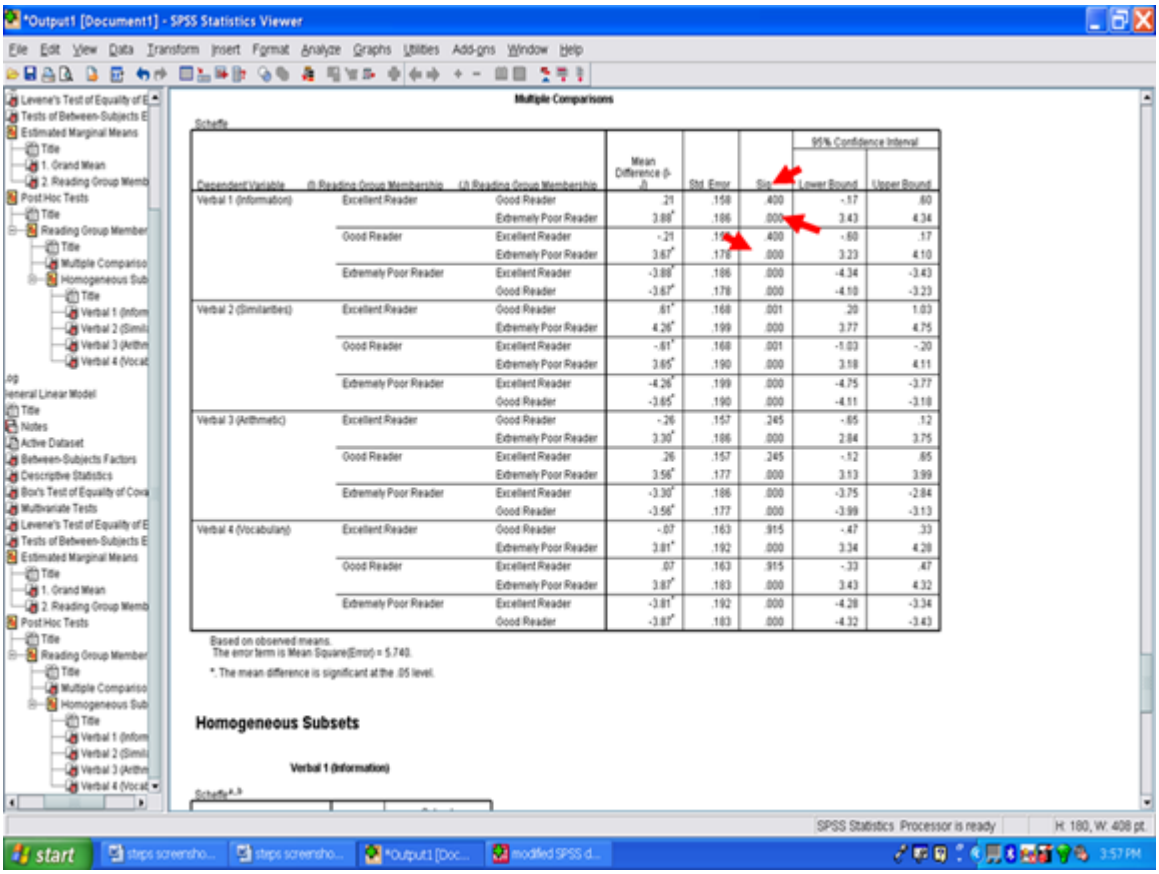
[12] http://cnx.org/content/m40731/latest/12.9.png/image



[13]

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



[14]

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: Assumptions[1]

NOTE: This chapter is published by NCPEA Press[2] 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 Help[3]) 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)

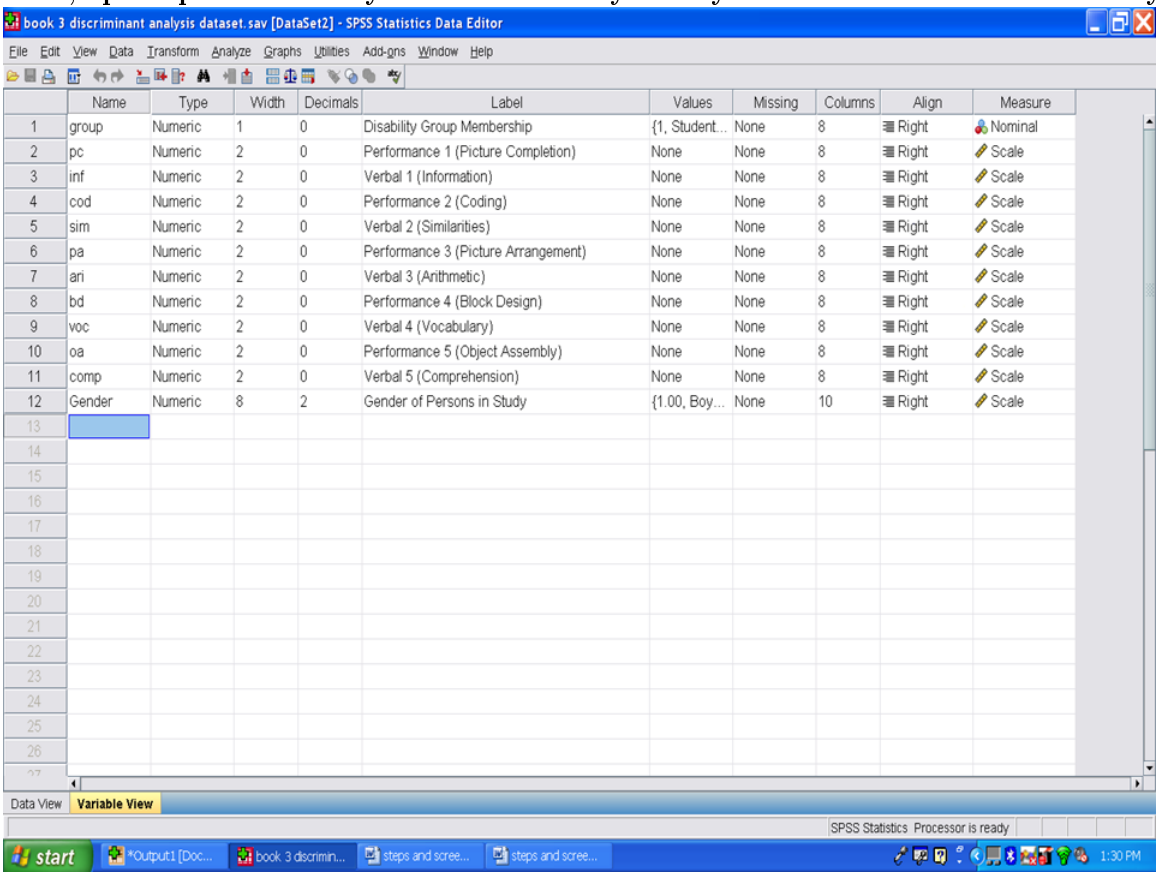
[1] This content is available online at <http://cnx.org/content/m40733/1.2/>.
[2] http://www.ncpeapublications.org/books.html
[3] http://cnx.org/content/m40733/latest/www.writingandstatisticalhelp
[4] http://davidmlane.com/hyperstat/

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=1-1[6]

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.



[7]

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

[5] http://www.statsoft.com/textbook/

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

[7] http://cnx.org/content/m40733/latest/13.1.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

	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

Data View Variable View

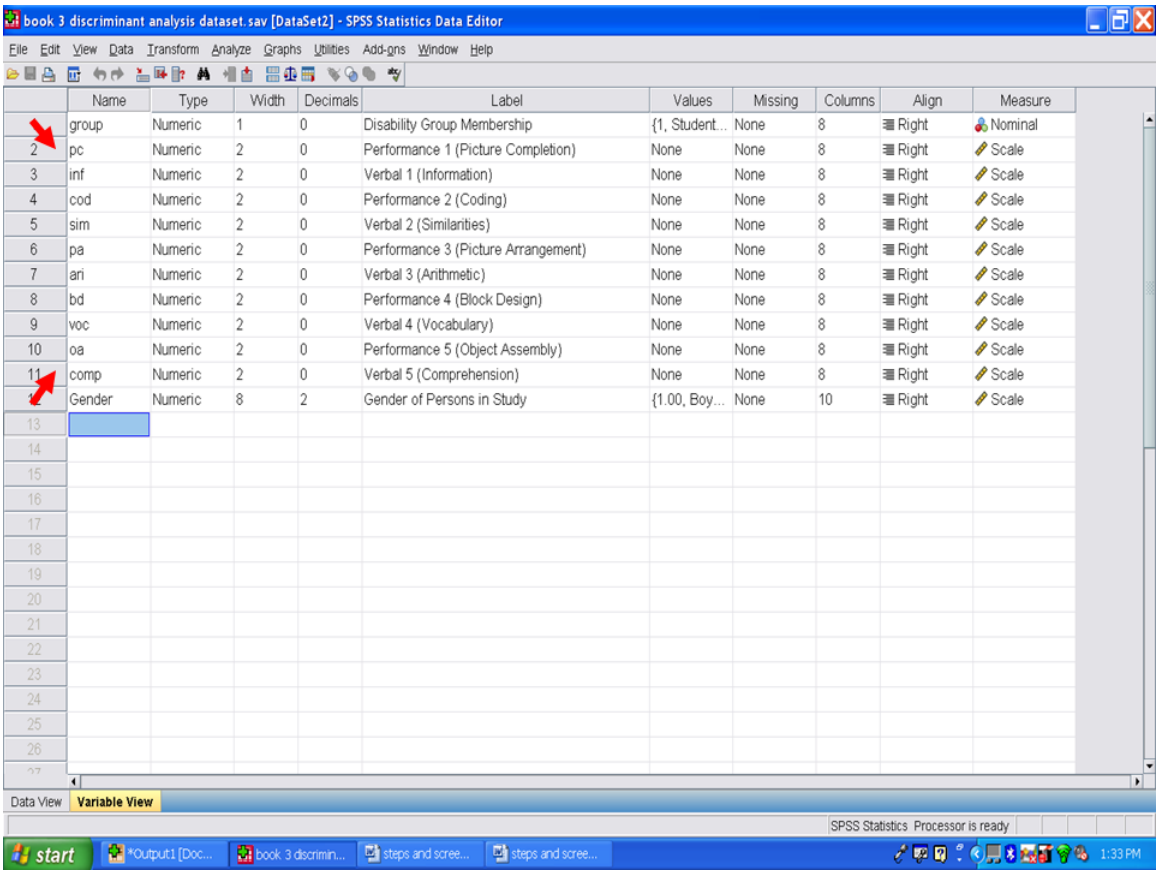
SPSS Statistics Processor is ready

start *Output1 [Doc... book 3 discrimin... steps and scree... steps and scree... 1:30 PM

[8]

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).

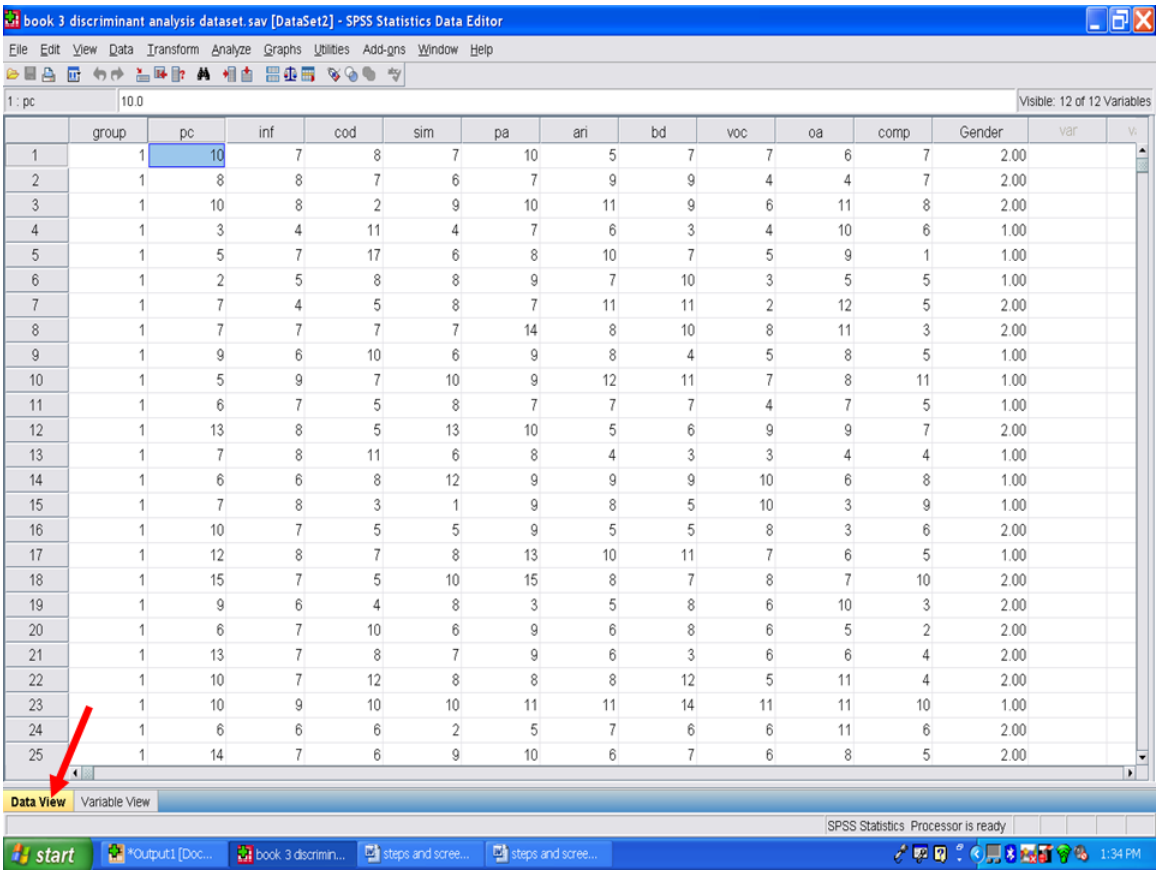
[8] http://cnx.org/content/m40733/latest/13.2.png/image



[9]

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



[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=356[12] 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
[12] http://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=326[14] 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-

[14] http://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 I[1]

NOTE: This chapter is published by NCPEA Press[2] 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

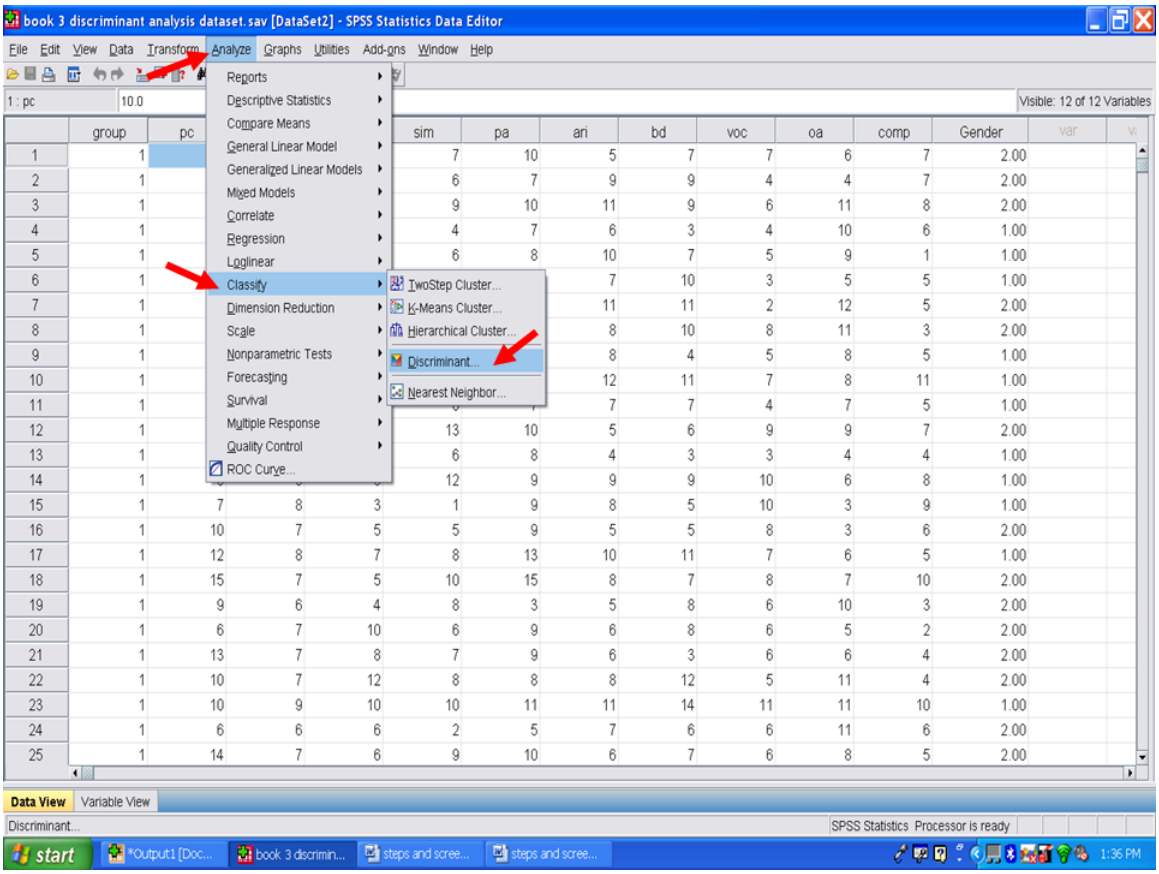
John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, 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 Help[3]) 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

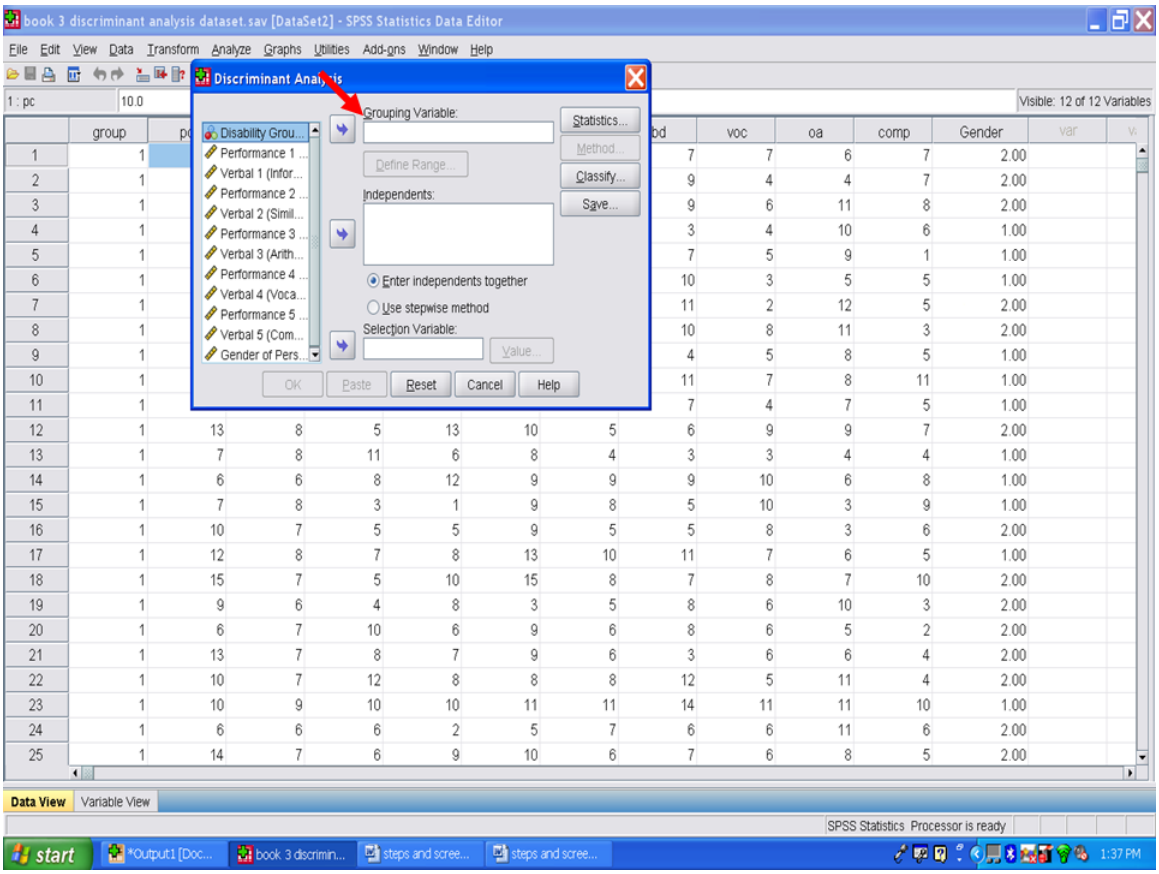
[1] This content is available online at <http://cnx.org/content/m40735/1.2/>.
[2] http://www.ncpeapublications.org/books.html
[3] http://cnx.org/content/m40735/latest/www.writingandstatisticalhelp



[4]

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.

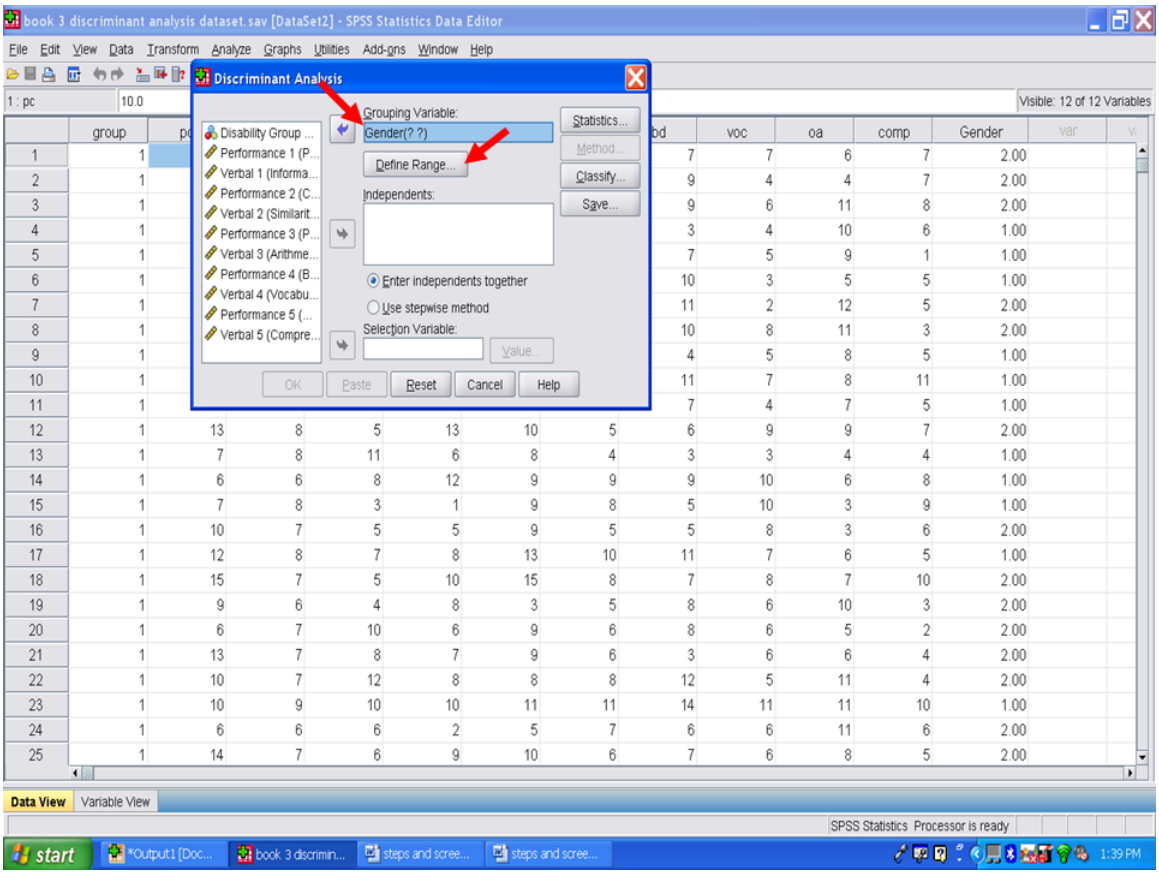
[4] http://cnx.org/content/m40735/latest/14.1.png/image



[5]

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.

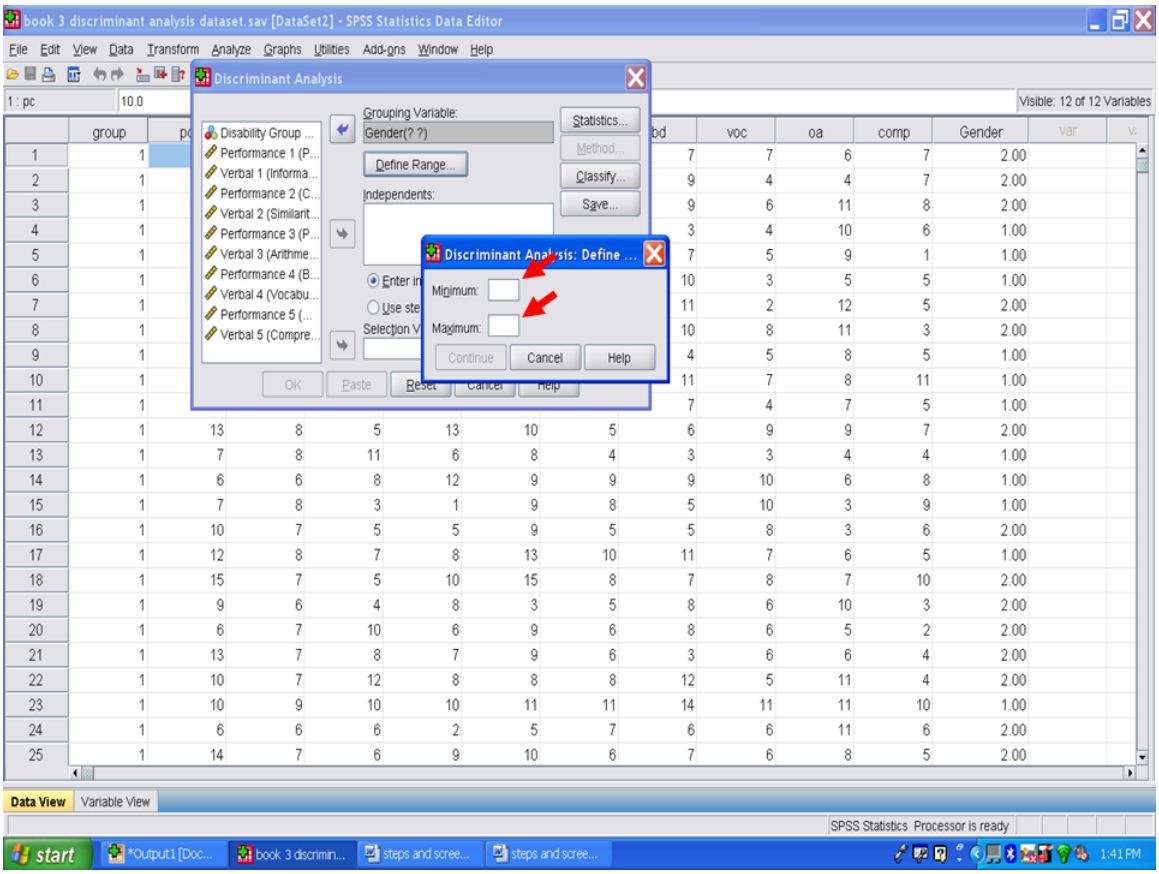
[5] http://cnx.org/content/m40735/latest/14.2.png/image



6

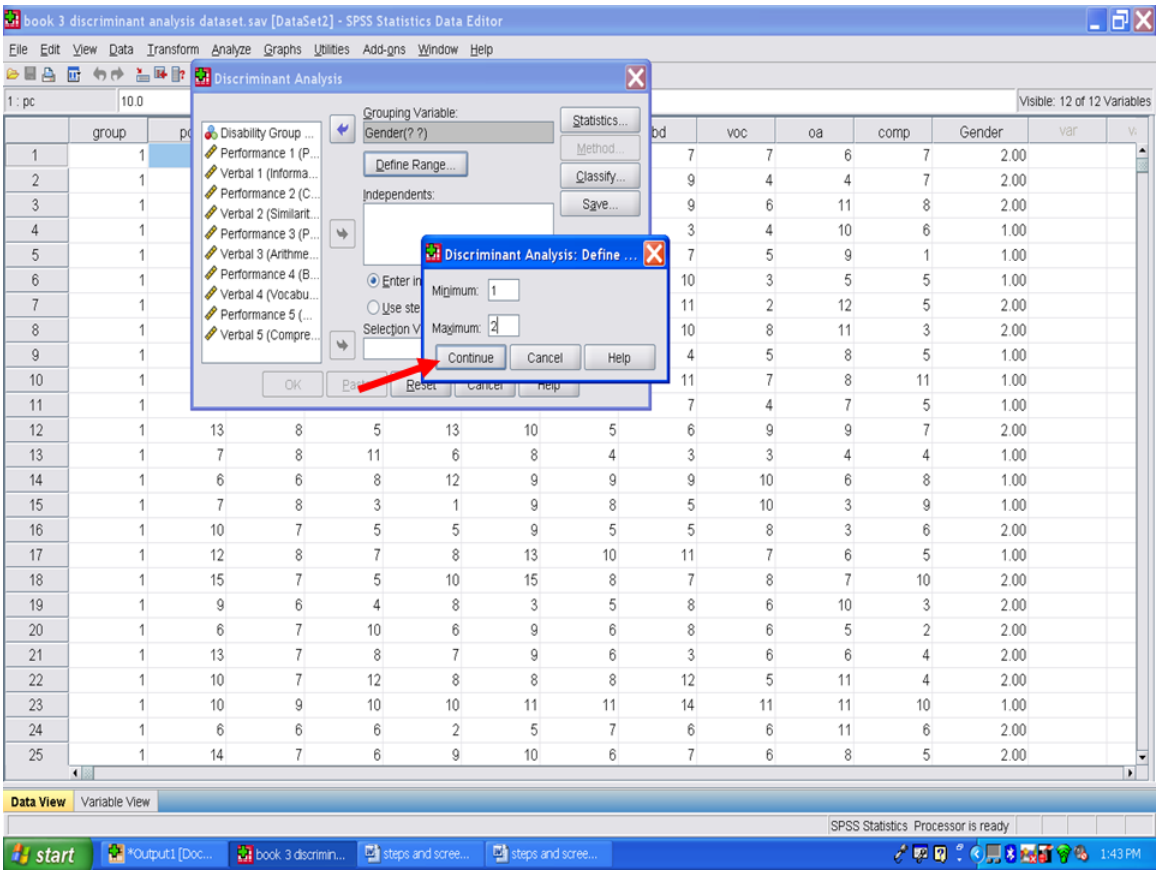
After clicking on Define Range, the following screen will appear: For the Minimum box, we will type in the number 1 because it represents boys. We will type in the number 2 in the Maximum box because it represents girls.

[6] http://cnx.org/content/m40735/latest/14.3.png/image



After typing in 1 and 2, then click on Continue.

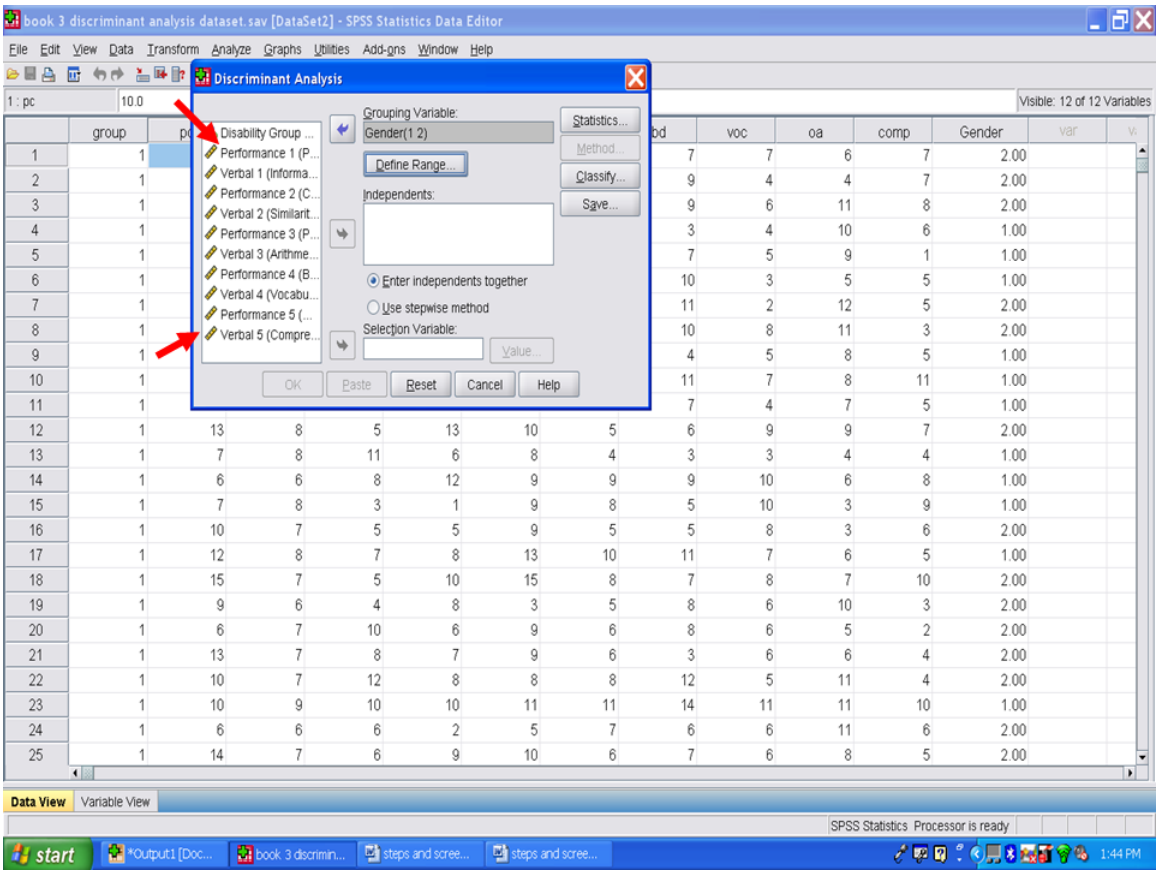
[7] http://cnx.org/content/m40735/latest/14.4.png/image



[8]

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.

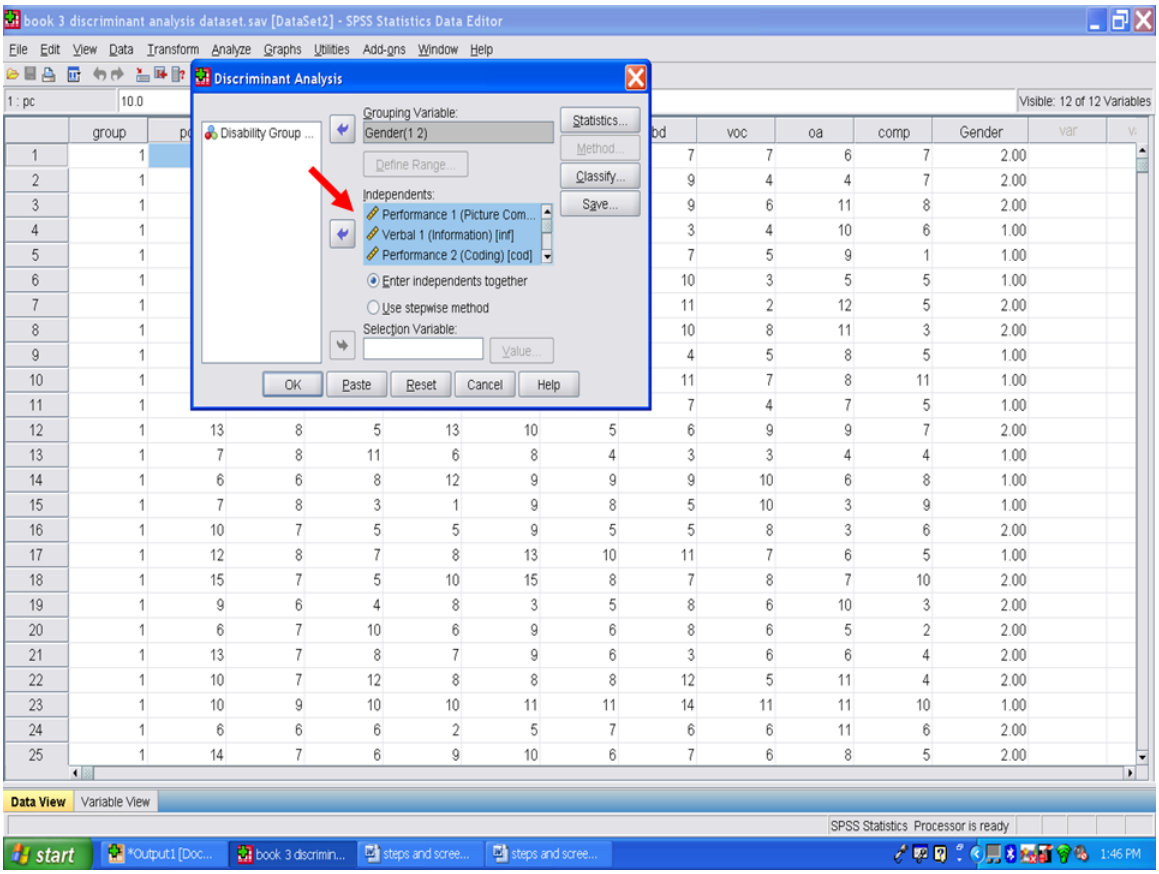
[8] http://cnx.org/content/m40735/latest/14.5.png/image



[9]

The screen should now look like the one below:

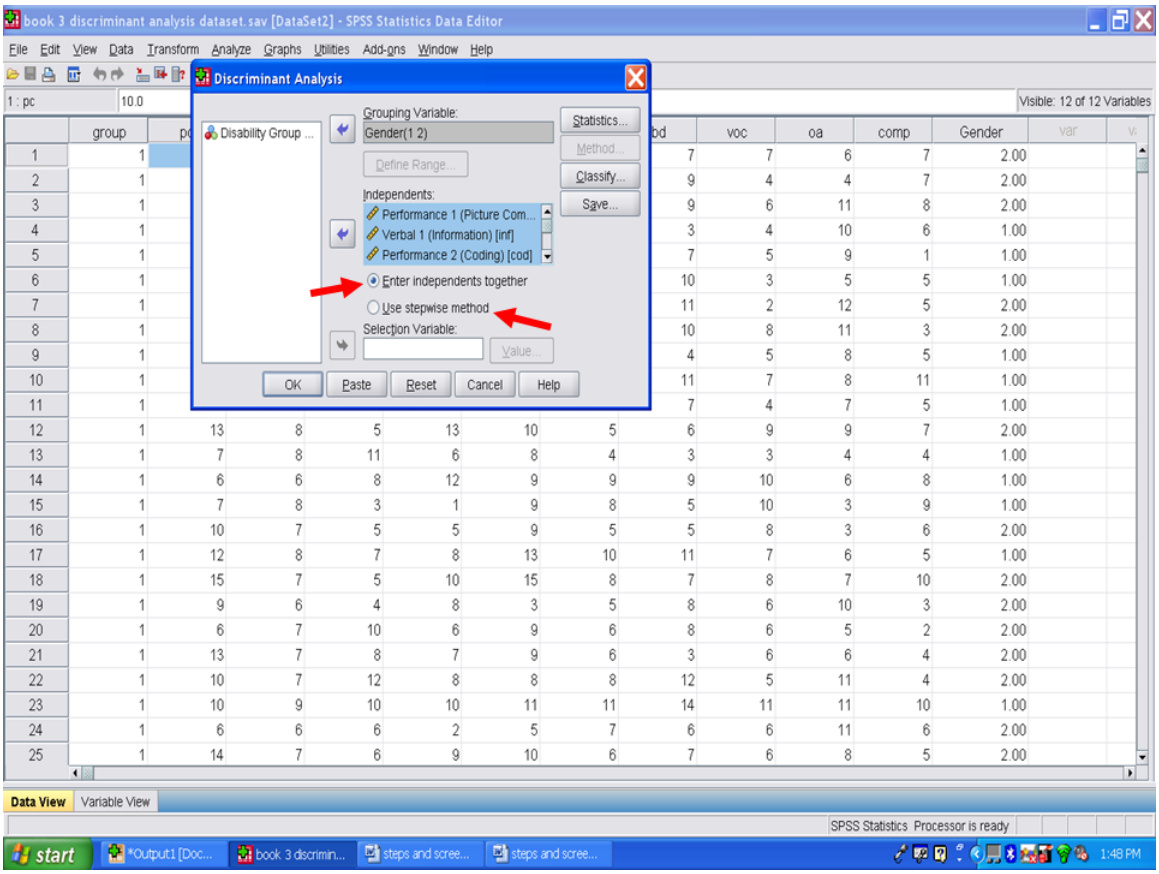
[9] http://cnx.org/content/m40735/latest/14.6.png/image



[10]

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



[11]

[11] http://cnx.org/content/m40735/latest/14.8.png/image

Chapter 16

16. Discriminant Analysis: Part II[1]

NOTE: This chapter is published by NCPEA Press[2] 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.

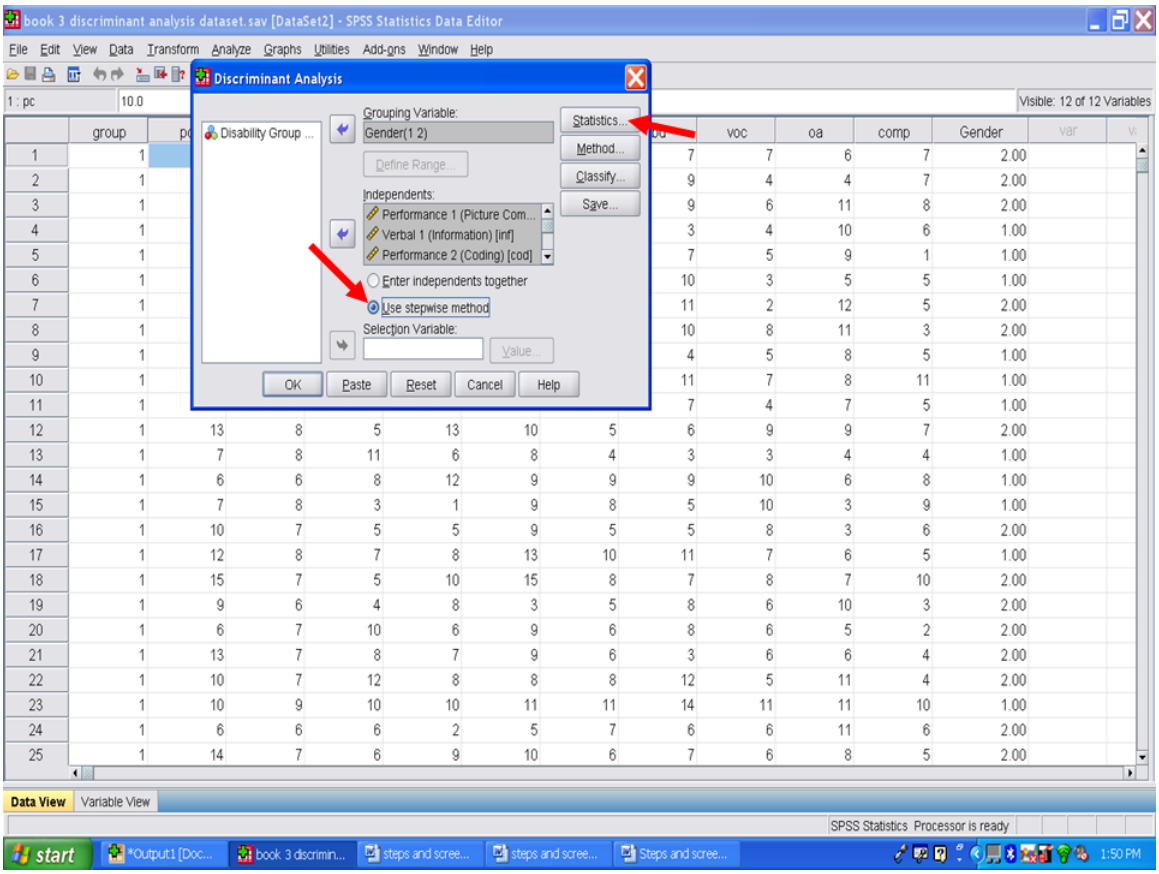
About the Authors

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 Help[3]) 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.

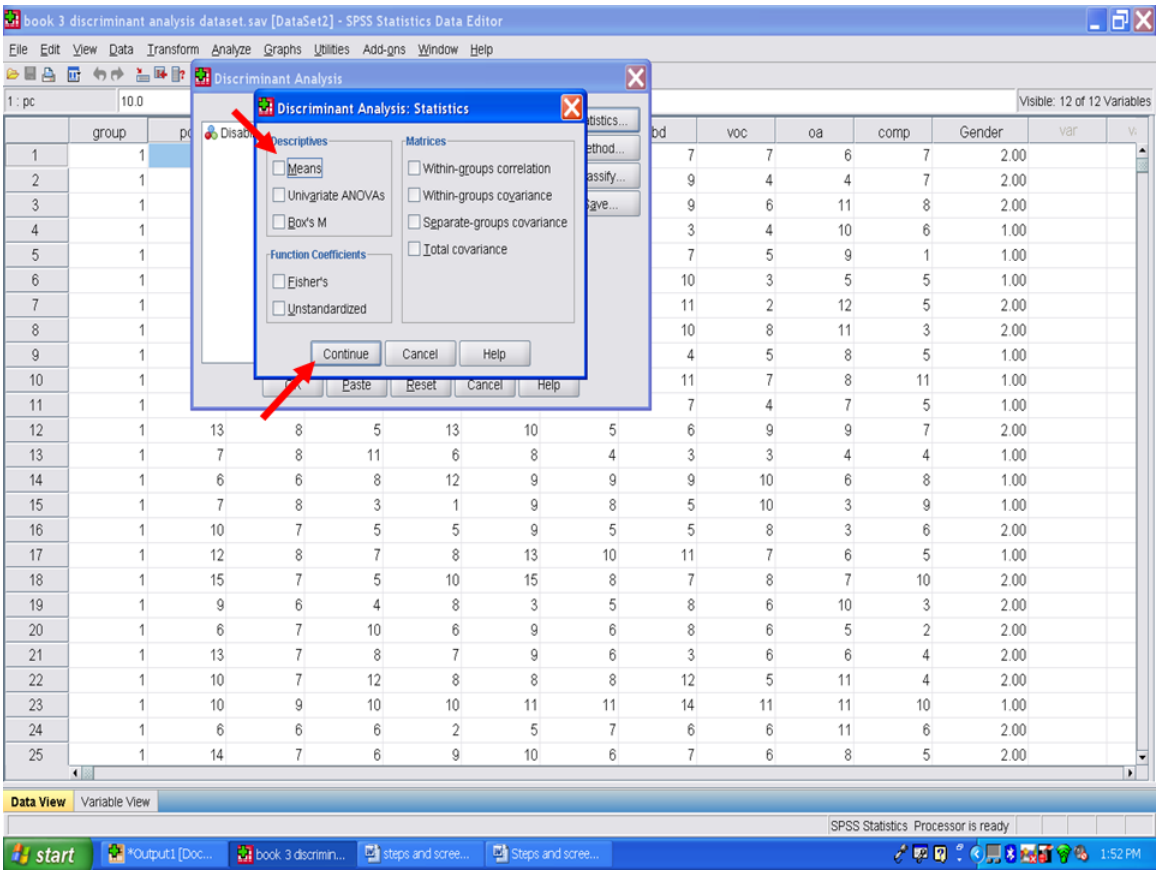
[1] This content is available online at <http://cnx.org/content/m40736/1.2/>.
[2] http://www.ncpeapublications.org/books.html
[3] http://cnx.org/content/m40736/latest/www.writingandstatisticalhelp



[4]

The following screen then appears.
Click on Means.
Click on Continue.

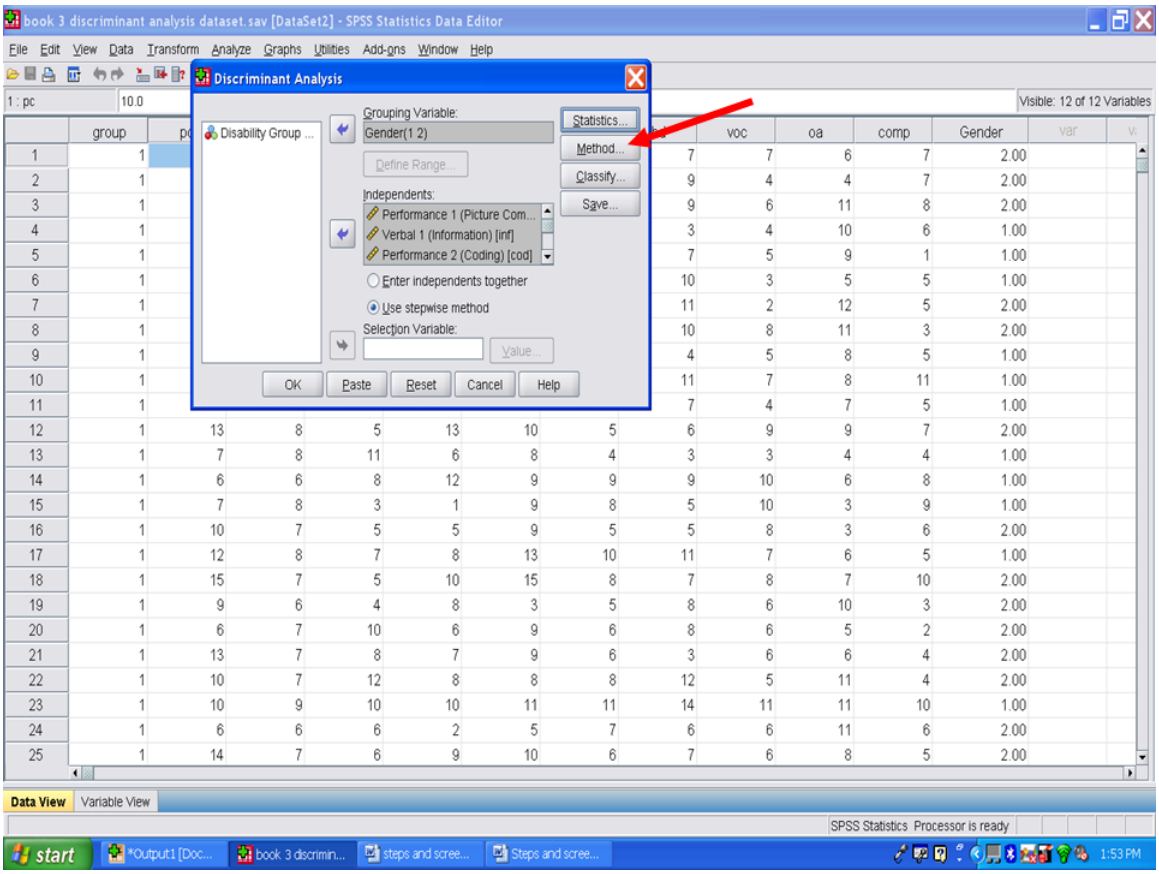
[4] http://cnx.org/content/m40736/latest/15.1.png/image



[5]

When the screen below appears, click on Method

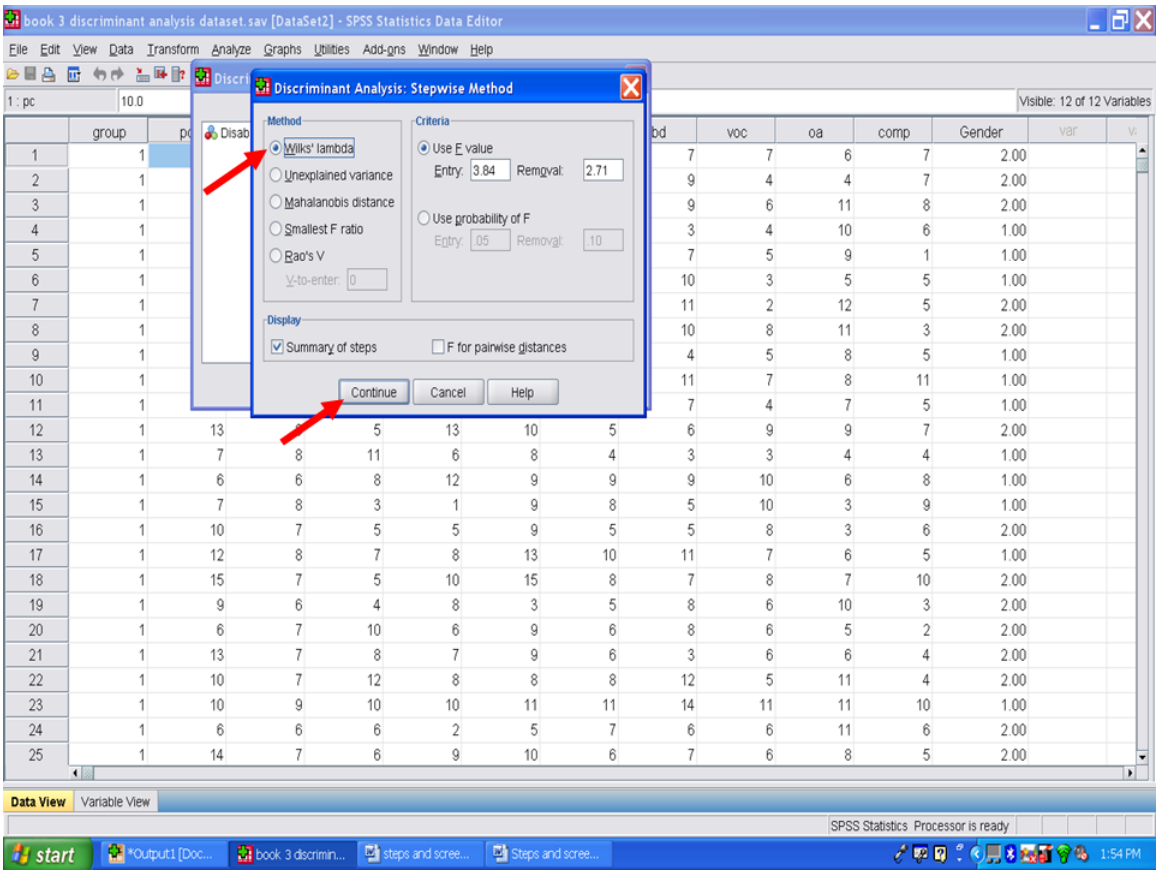
[5] http://cnx.org/content/m40736/latest/15.2.png/image



[6]

In the Method screen, we will use the default of Wilks' lambda. Click on Continue.

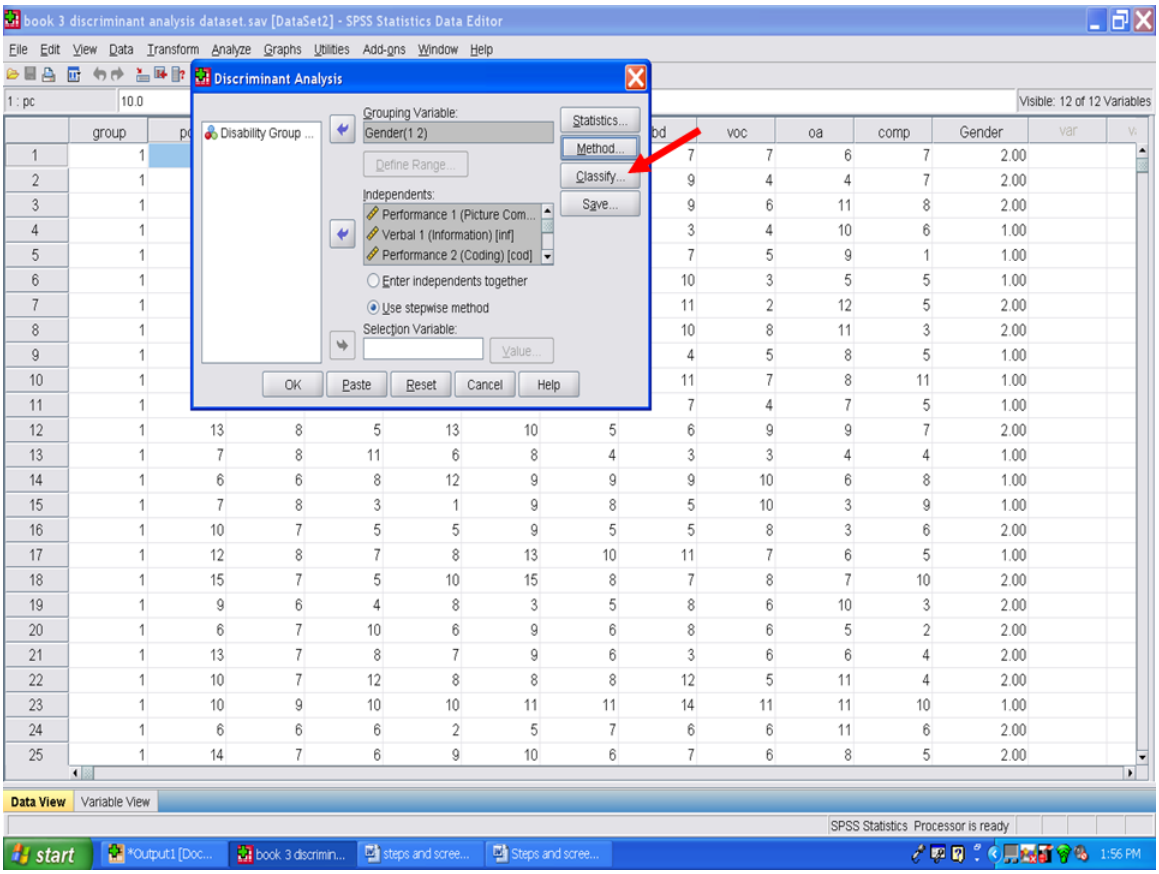
[6] http://cnx.org/content/m40736/latest/15.3.png/image



[7]

Then click on Classify.

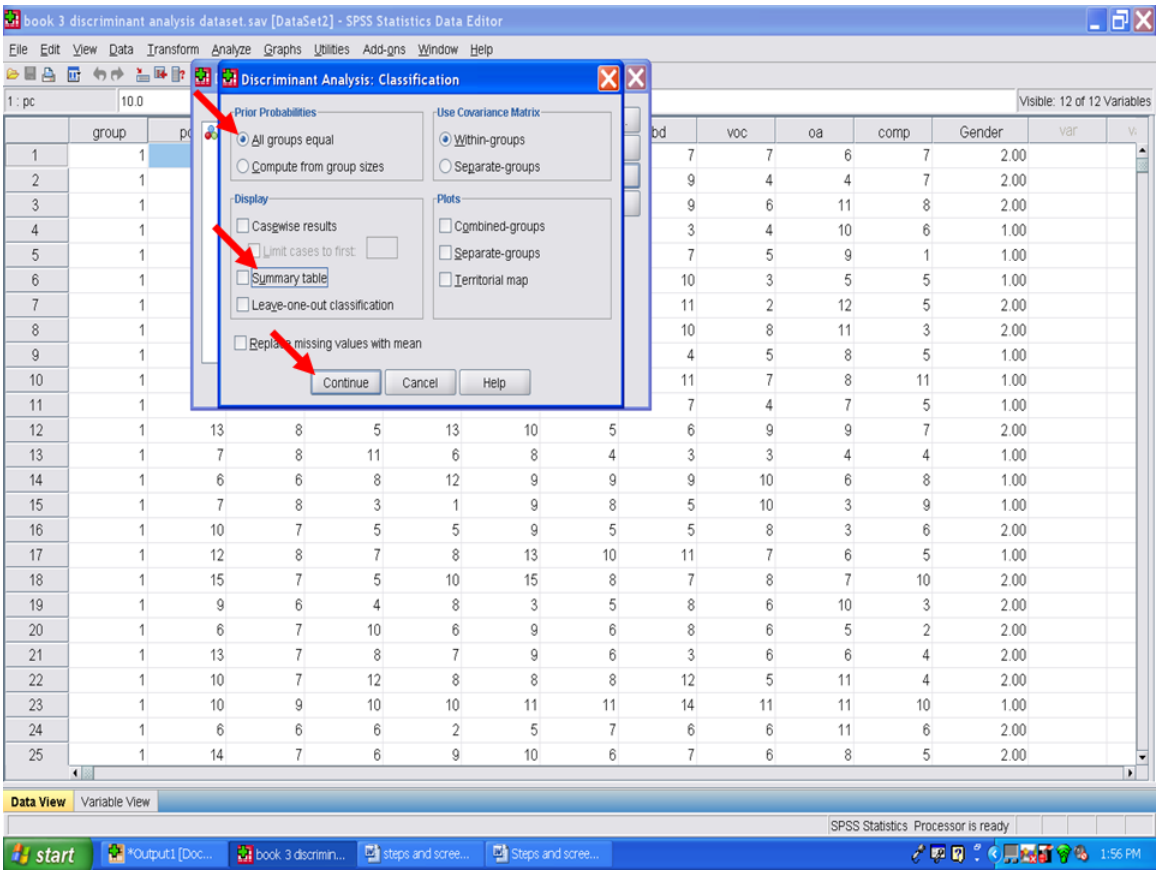
[7] http://cnx.org/content/m40736/latest/15.4.png/image



[8]

In the screen that appears, the All groups equal button is already clicked.
Click on Summary table under the Display.
Then click on Continue.

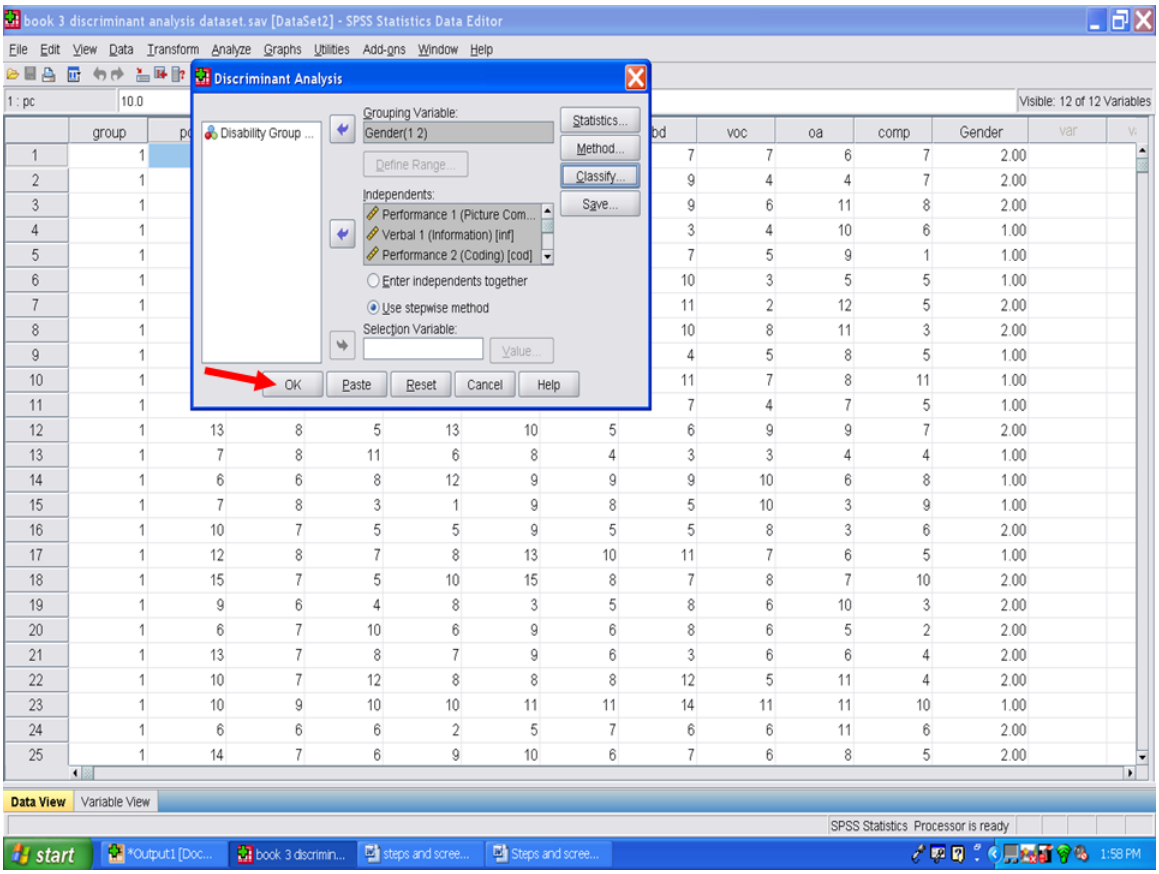
[8] http://cnx.org/content/m40736/latest/15.5.png/image



[9]

We are now ready to have SPSS calculate this procedure. Click on OK.

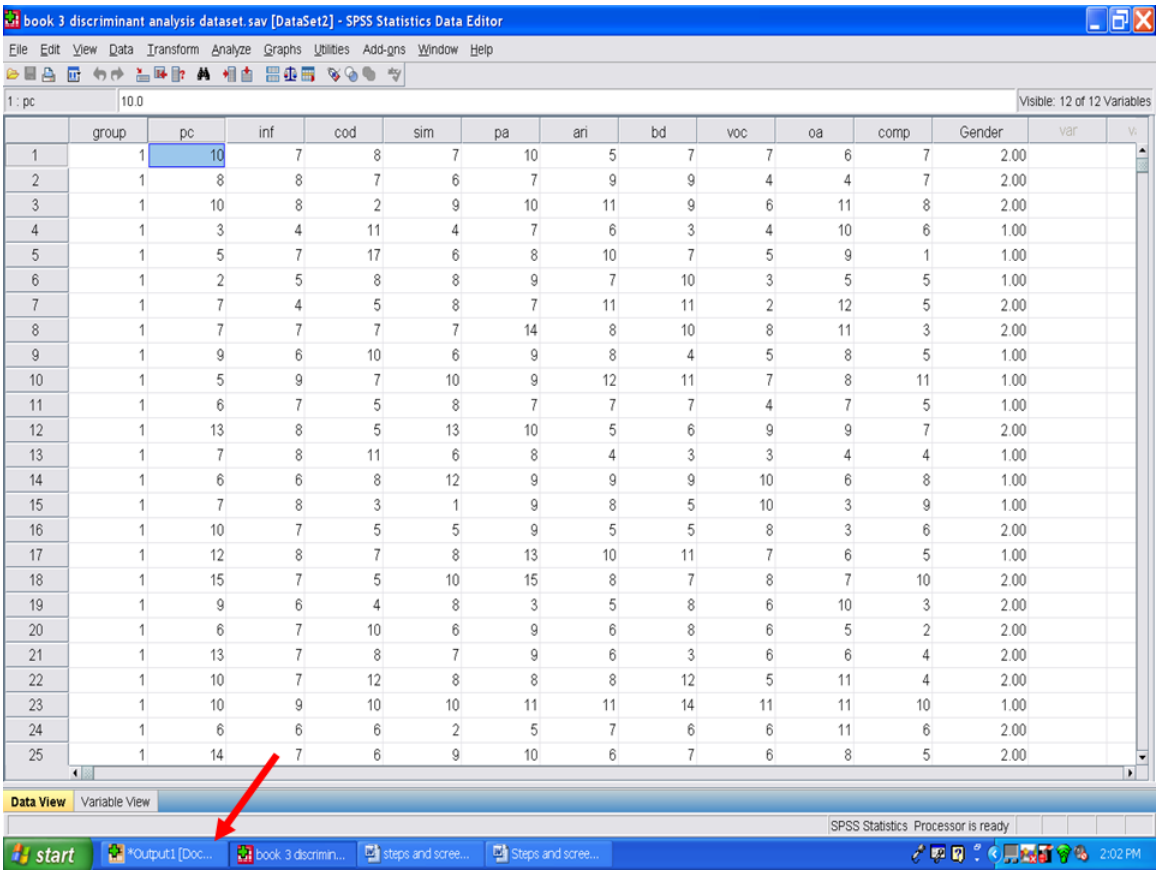
[9] http://cnx.org/content/m40736/latest/15.6.png/image



[10]

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

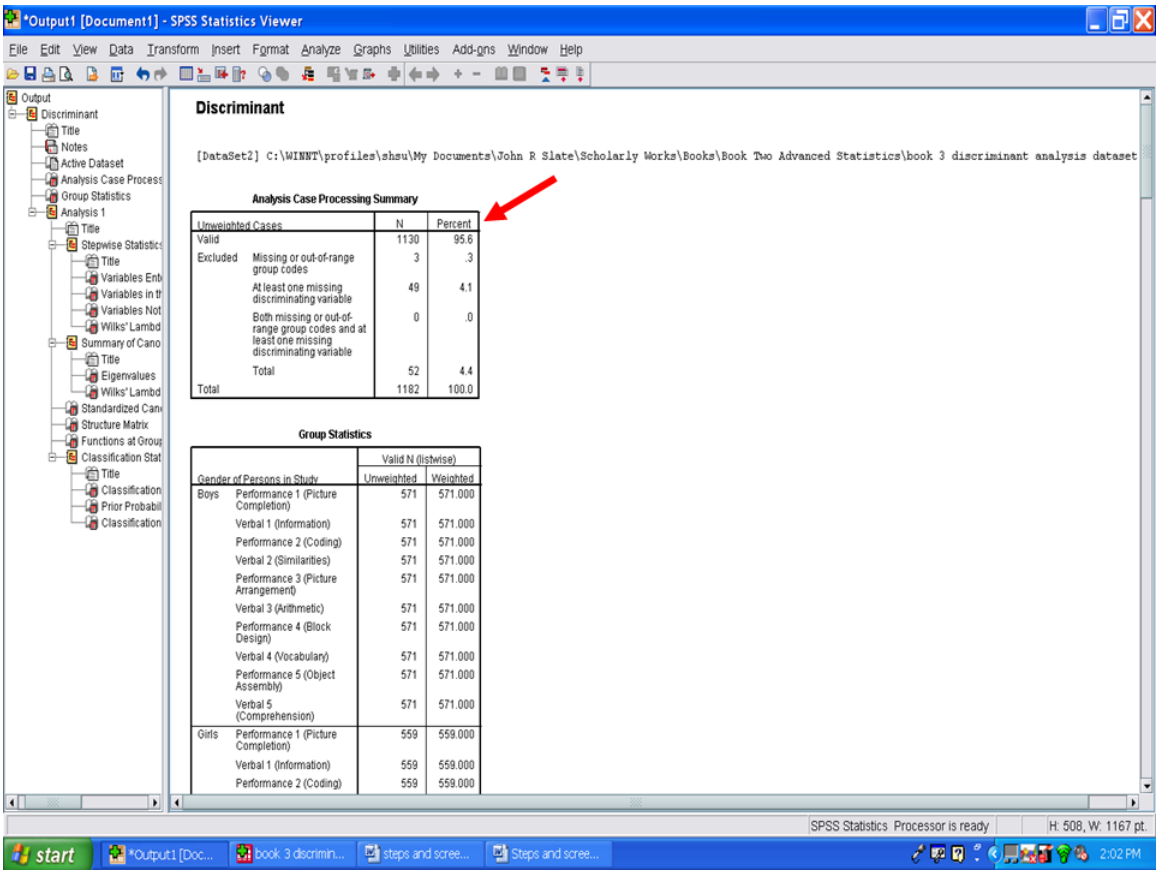


[11]

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.

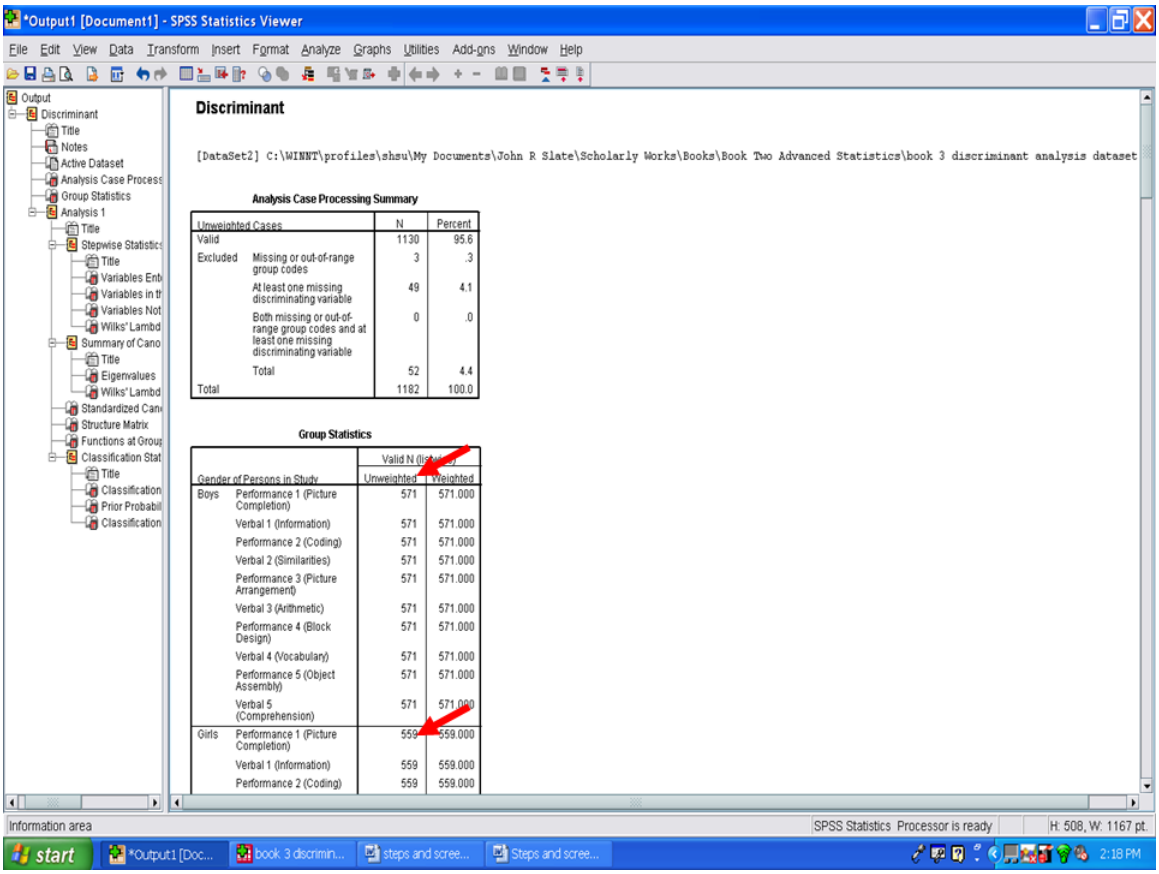
[11] http://cnx.org/content/m40736/latest/15.8.png/image



12

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



13

[13] http://cnx.org/content/m40736/latest/15.10.png/image

Chapter 17

17. Discriminant Analysis: Part III[1]

NOTE: This chapter is published by NCPEA Press[2] 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.

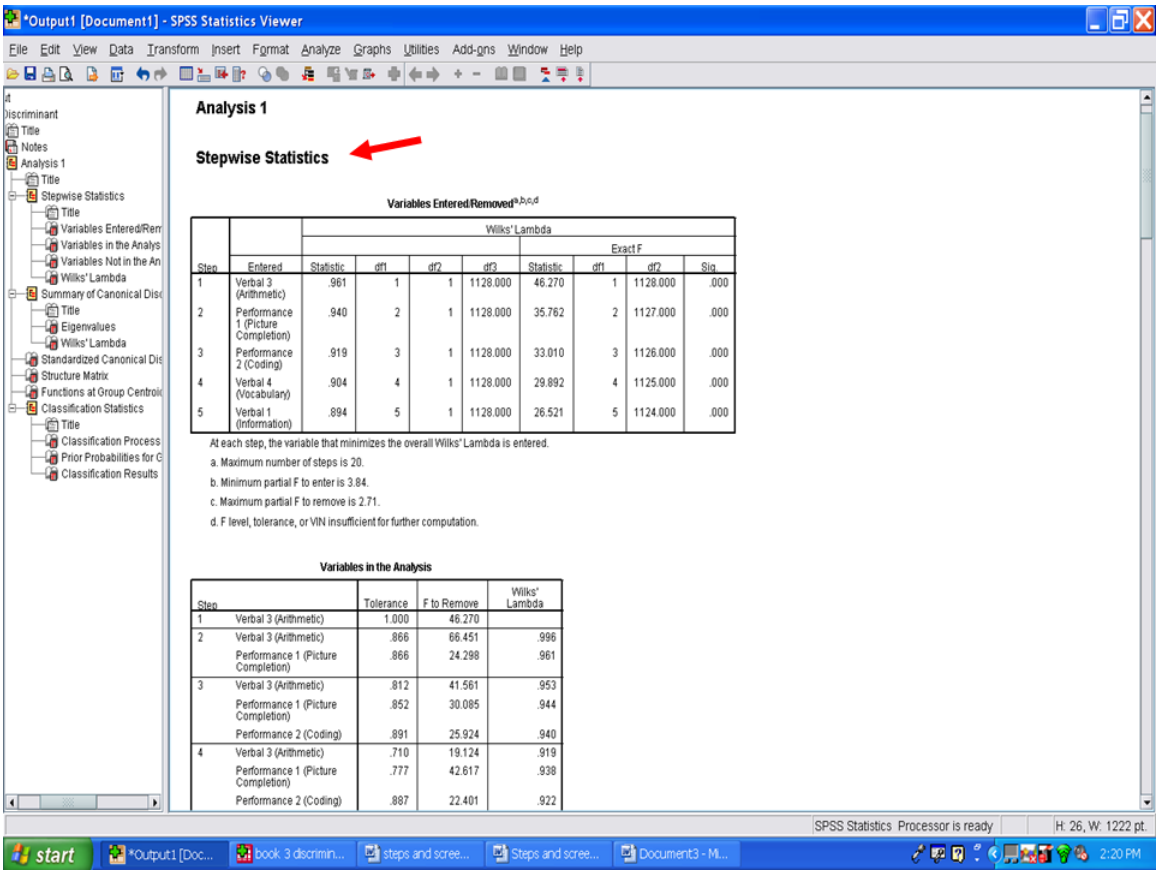
About the Authors

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 Help[3]) 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 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.

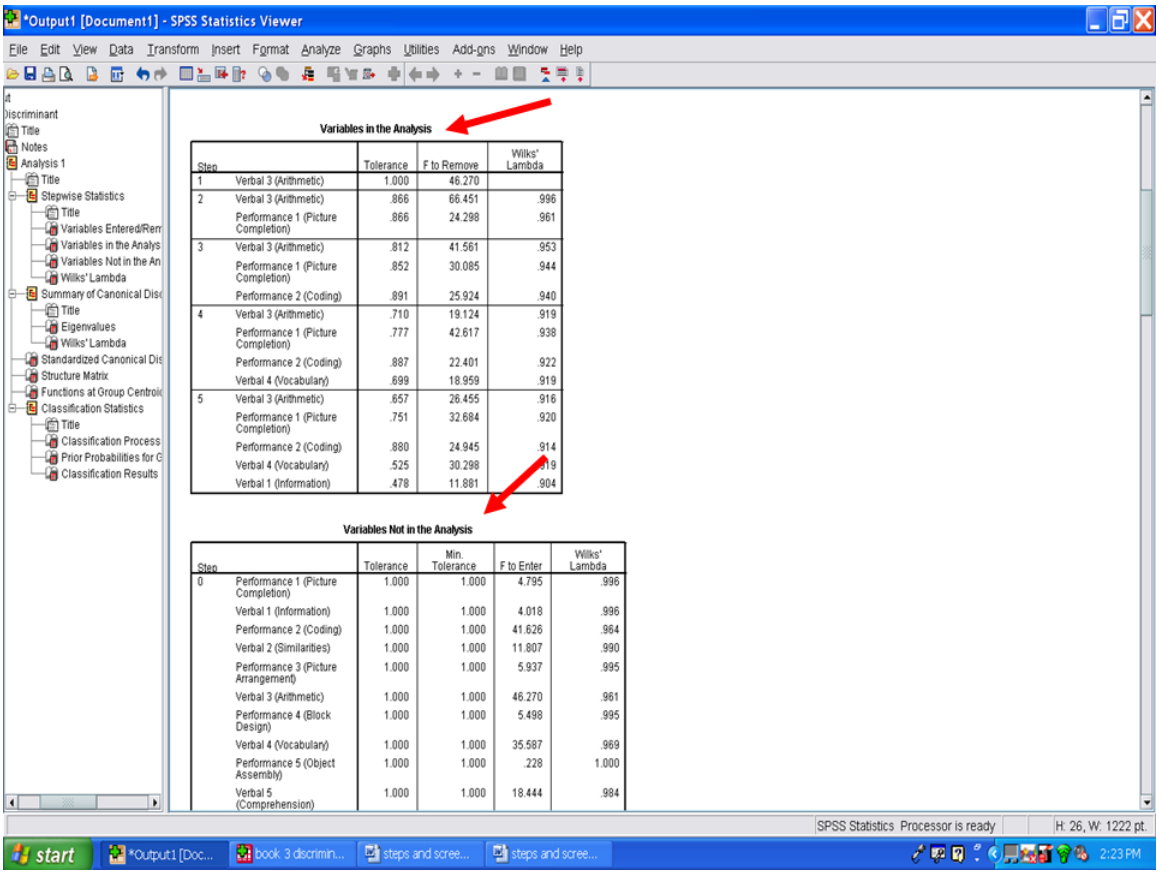
[1] This content is available online at <http://cnx.org/content/m40737/1.2/>.
[2] http://www.ncpeapublications.org/books.html
[3] http://cnx.org/content/m40737/latest/www.writingandstatisticalhelp



[4]

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.

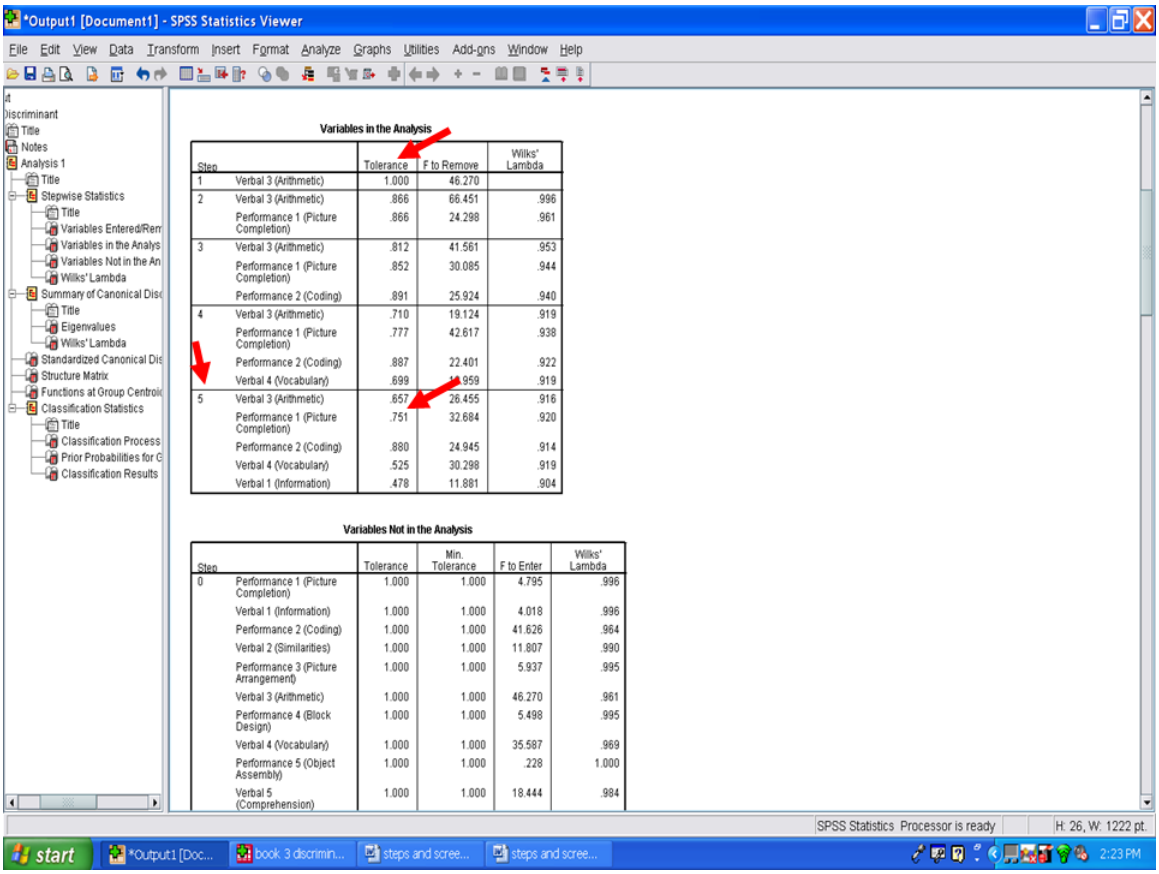
[4] http://cnx.org/content/m40737/latest/16.1.png/image



[5]

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.

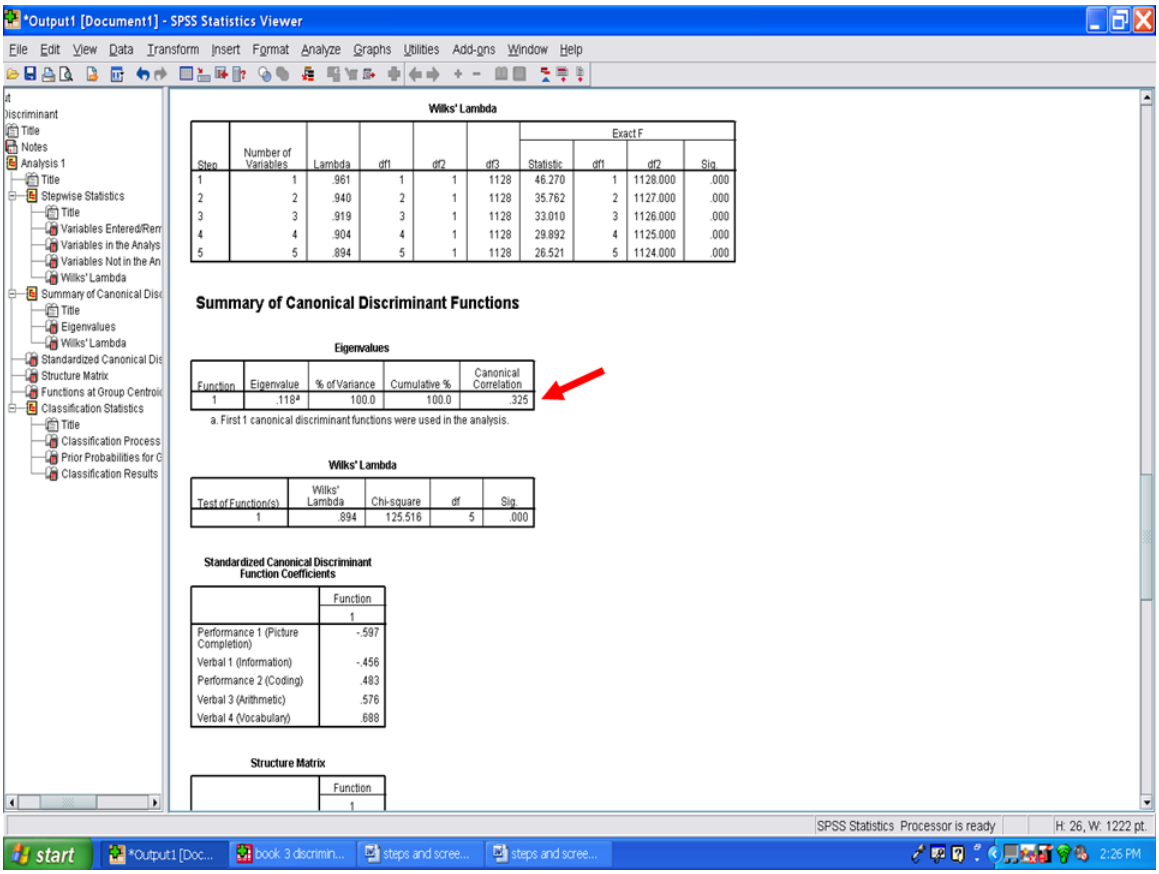
[5] http://cnx.org/content/m40737/latest/16.2.png/image



[6]

Under the Wilks' lambda table is a Summary of Canonical Discriminant Functions table labeled Eigenvalues. In this table is the canonical correlation, or *Rc*, which in this example is .325. This *Rc* value reflects the correlation between the groups and the discriminant function formed by the five variables that were statistically significant. The *Rc* may be interpreted as a Pearson correlation of the discriminant scores with the grouping variable. This information will be reported in your Results section.

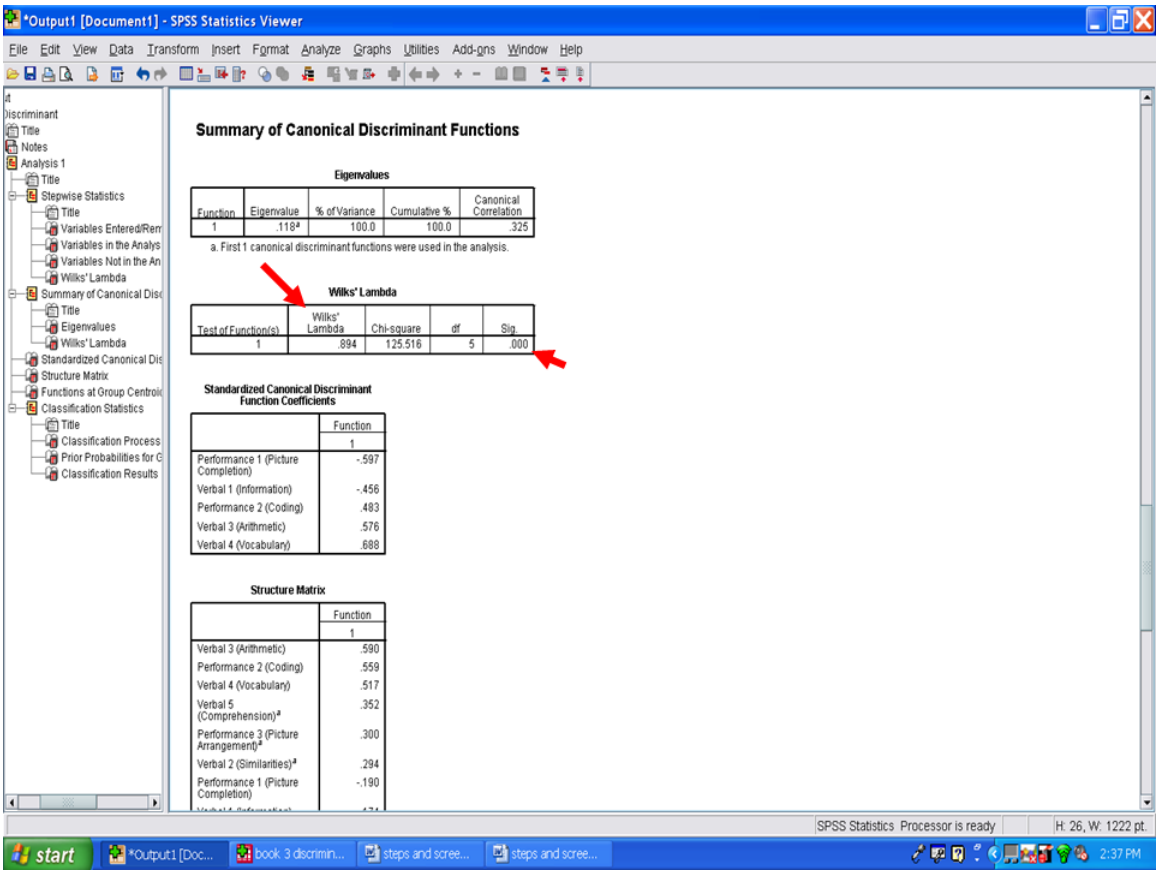
[6] http://cnx.org/content/m40737/latest/16.3.png/image



[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.

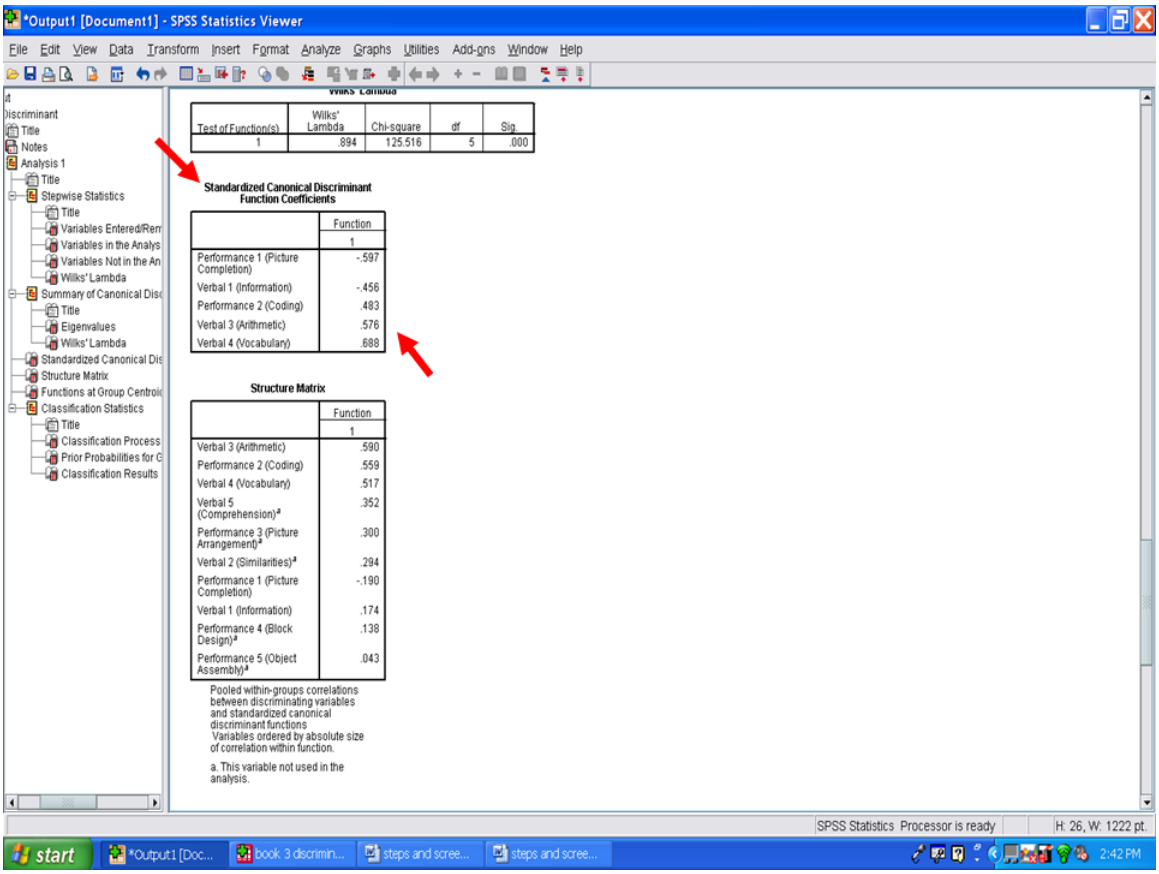
[7] http://cnx.org/content/m40737/latest/16.4.png/image



[8]

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.

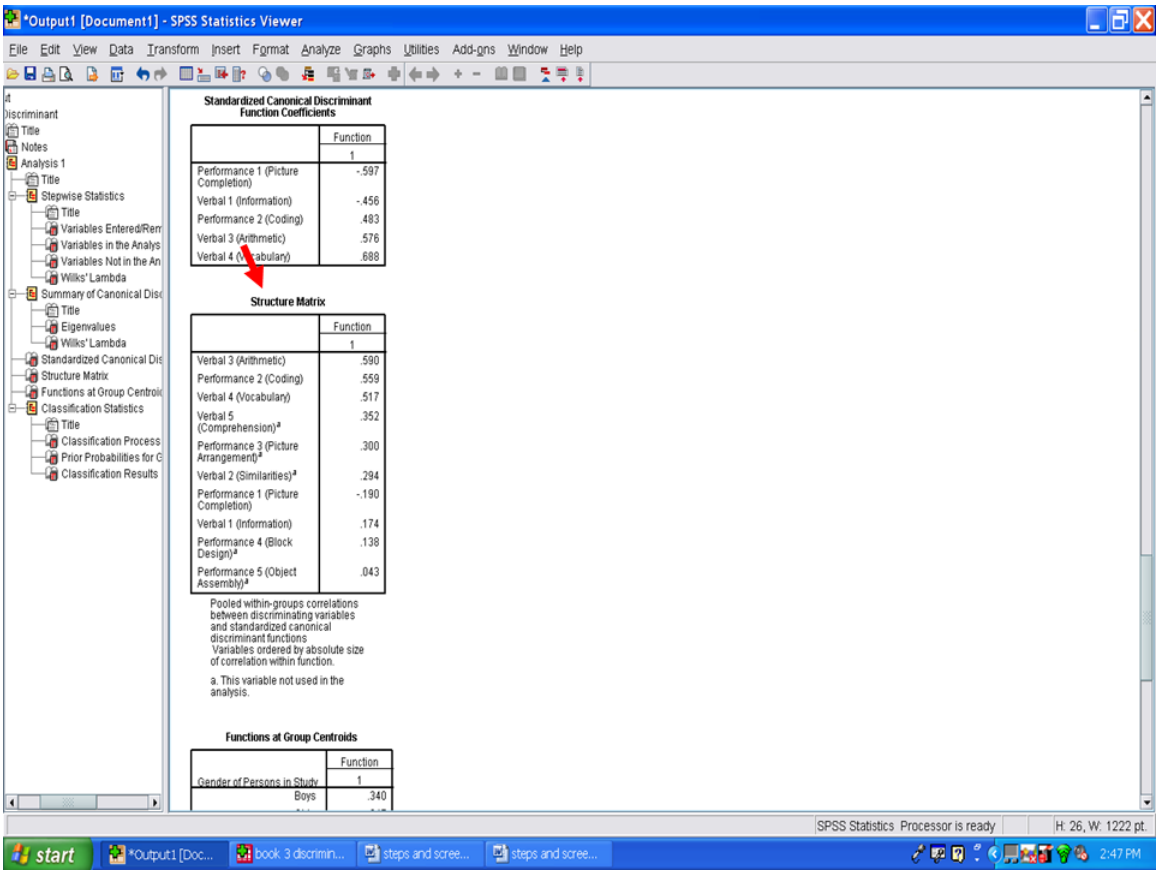
[8] http://cnx.org/content/m40737/latest/16.5.png/image



[9]

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).

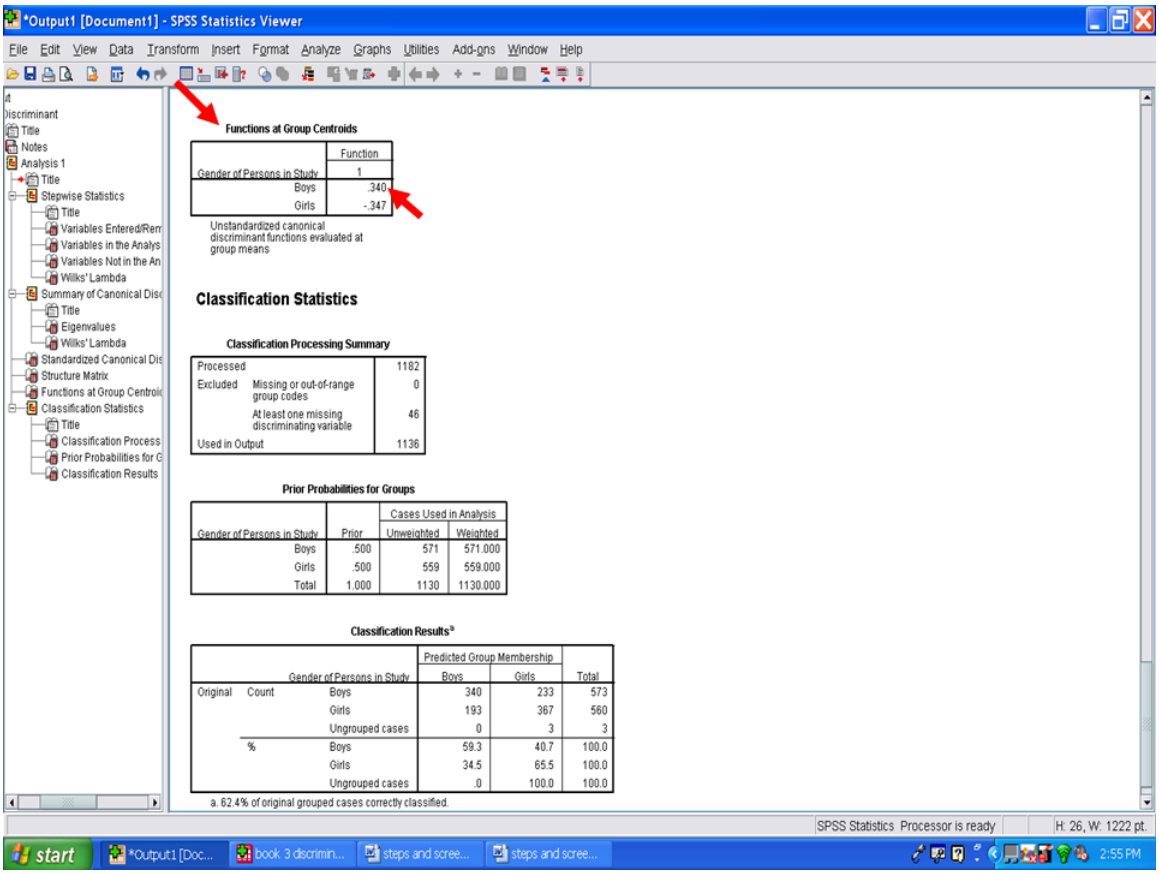
[9] http://cnx.org/content/m40737/latest/16.6.png/image



10

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



[11]

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: Assumptions[1]

NOTE: This chapter is published by NCPEA Press[2] 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 Help[3]) 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)

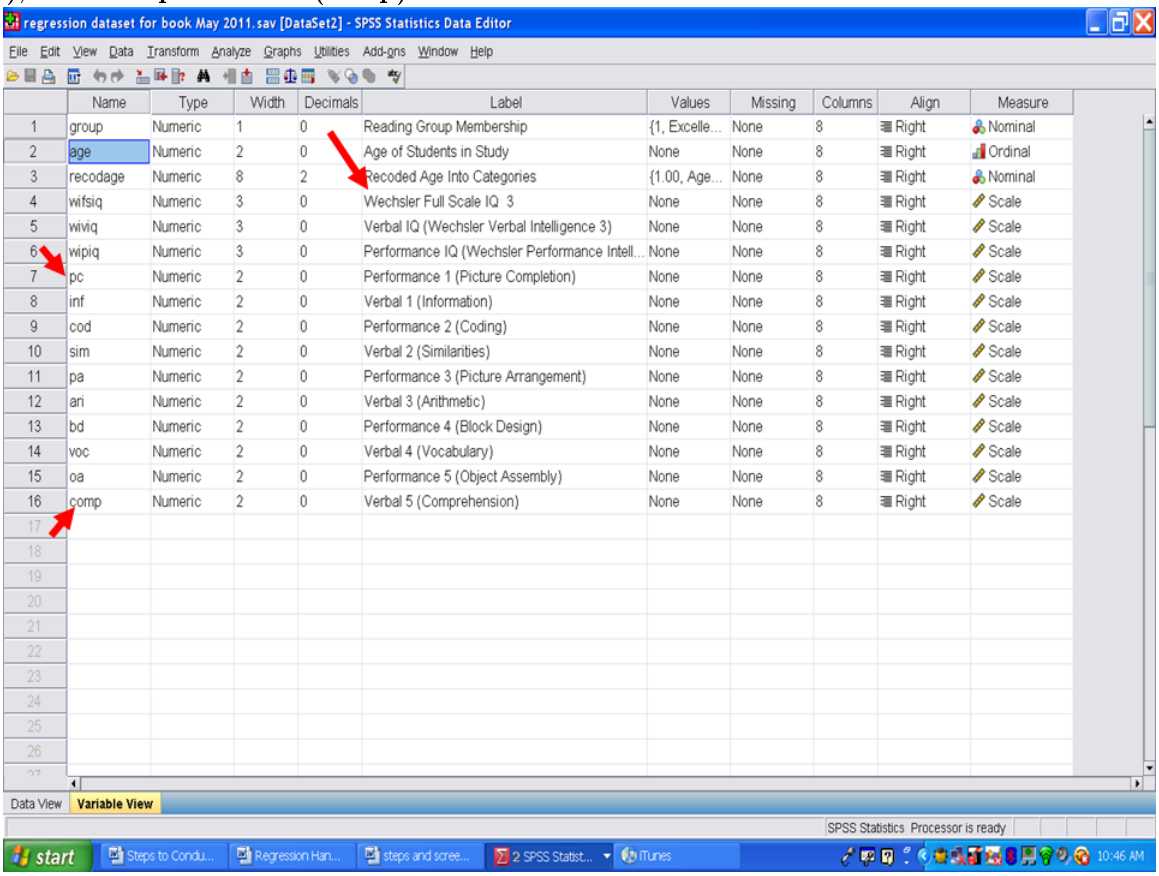
[1] This content is available online at <http://cnx.org/content/m40738/1.2/>.
[2] http://www.ncpeapublications.org/books.html
[3] http://cnx.org/content/m40738/latest/www.writingandstatisticalhelp
[4] http://davidmlane.com/hyperstat/

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=1-1[6]

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 (wifsiq). 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).



[7]

As with every statistical procedure, we need to check the underlying assumptions. One

[5] http://www.statsoft.com/textbook/

[6] http://www.amazon.com/Discovering-Statistics-Introducing-Statistical-Method/dp/1847879071/ref=sr_1_1?s=books&ie=UTF8&qid=1304961

[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=356[8] 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.

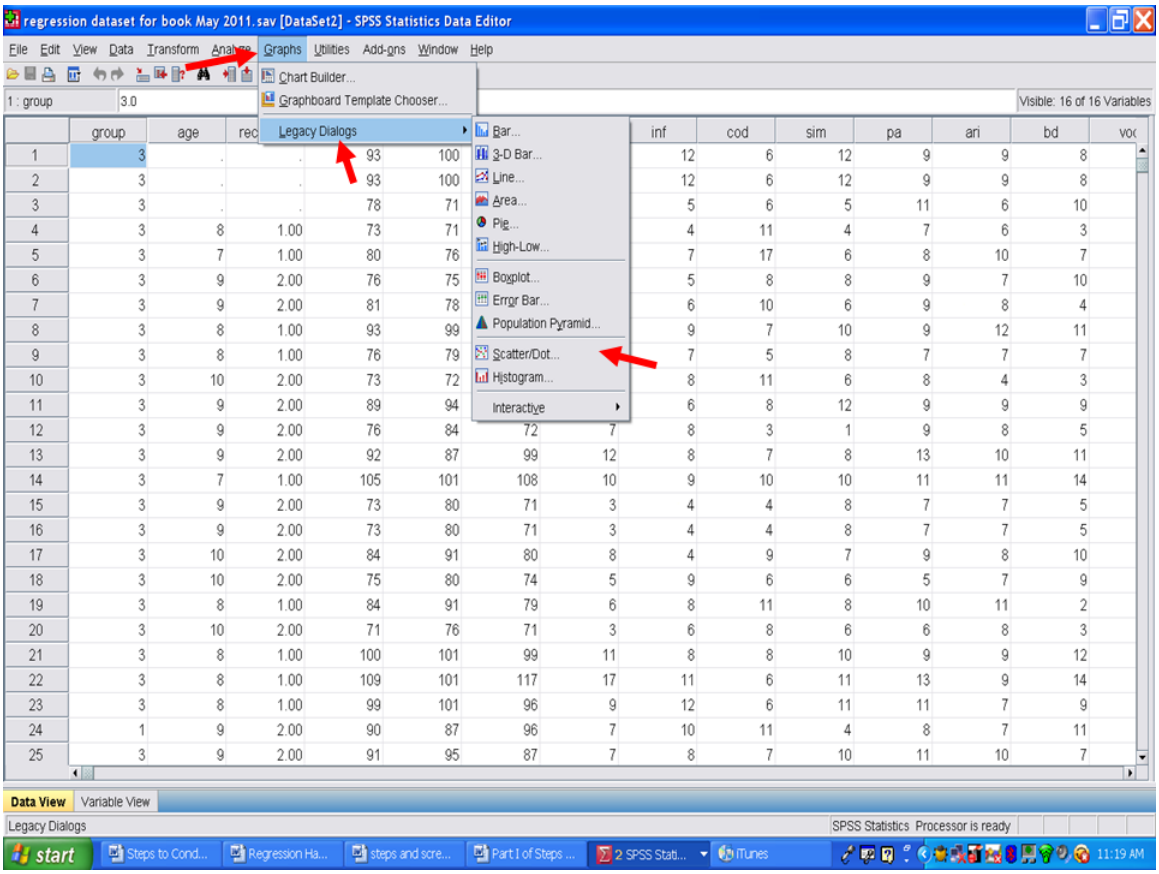
*** Kurtosis [Note. Kurtosis also refers to the extent to which the data are normally distributed around the mean. This time, the data are piled up higher than normal around the mean or piled up higher than normal at the ends of the distribution.] Readers are referred to the following sources for a more detailed definition of kurtosis: http://www.statistics.com/index.php?page=glossary&term_id=326[10] 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

[8] http://www.statistics.com/index.php?page=glossary&term_id=356
[9] http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb
[10] http://www.statistics.com/index.php?page=glossary&term_id=326
[11] http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb

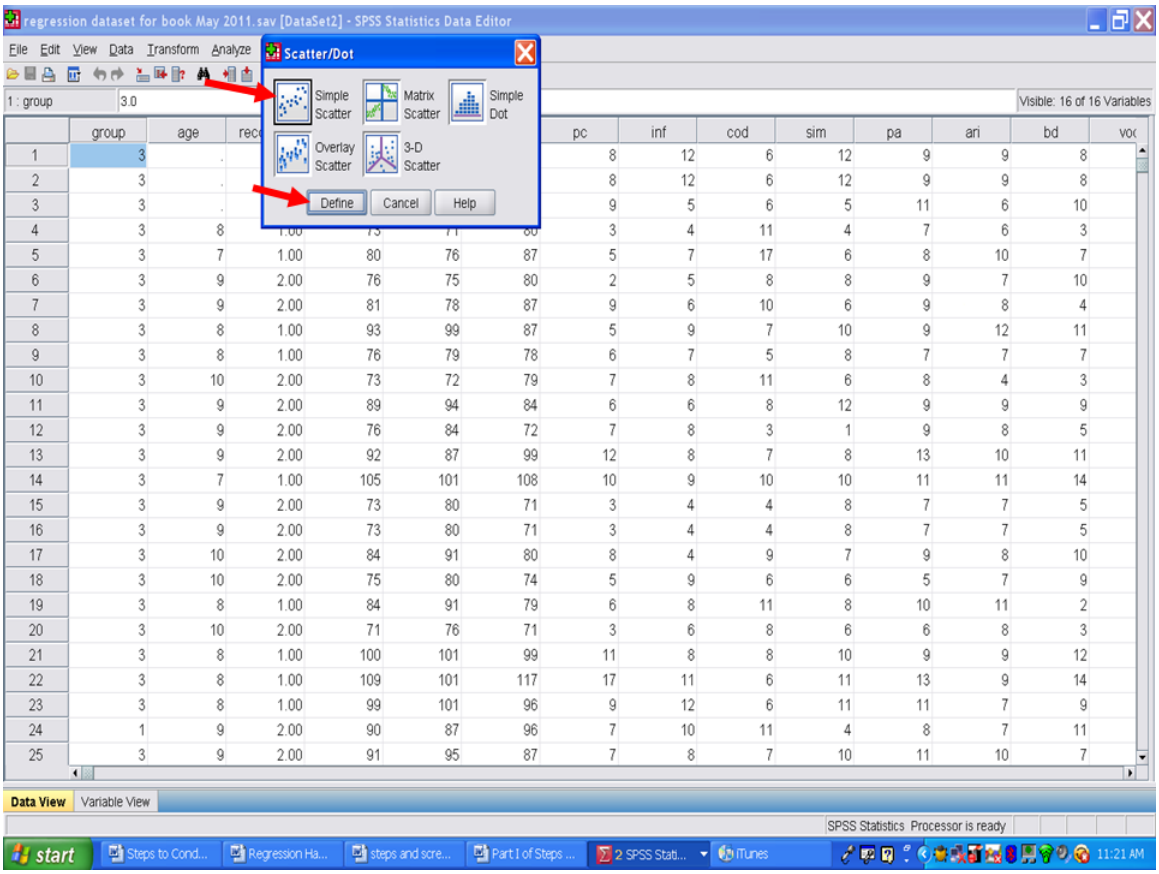


12

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

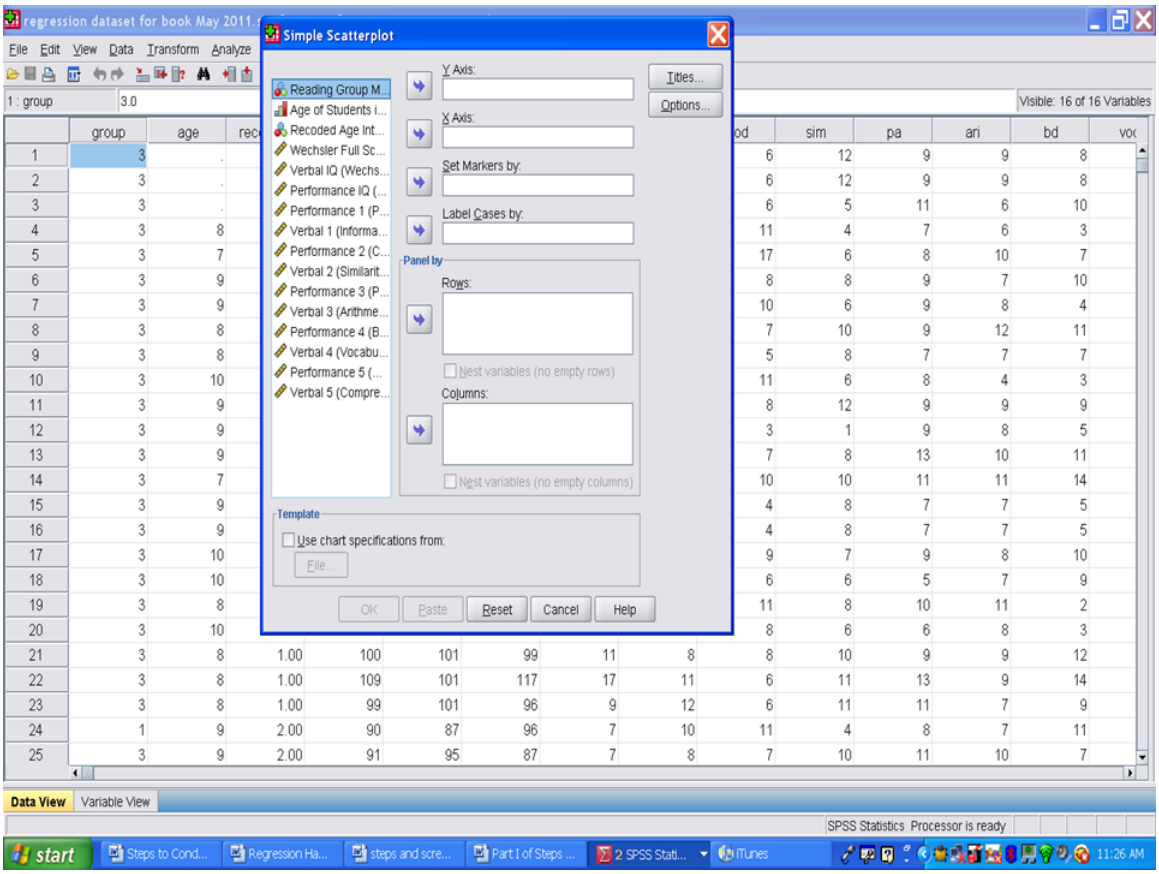
[12] http://cnx.org/content/m40738/latest/17.2.png/image



13

The following screen should now be present.

[13] http://cnx.org/content/m40738/latest/17.3.png/image

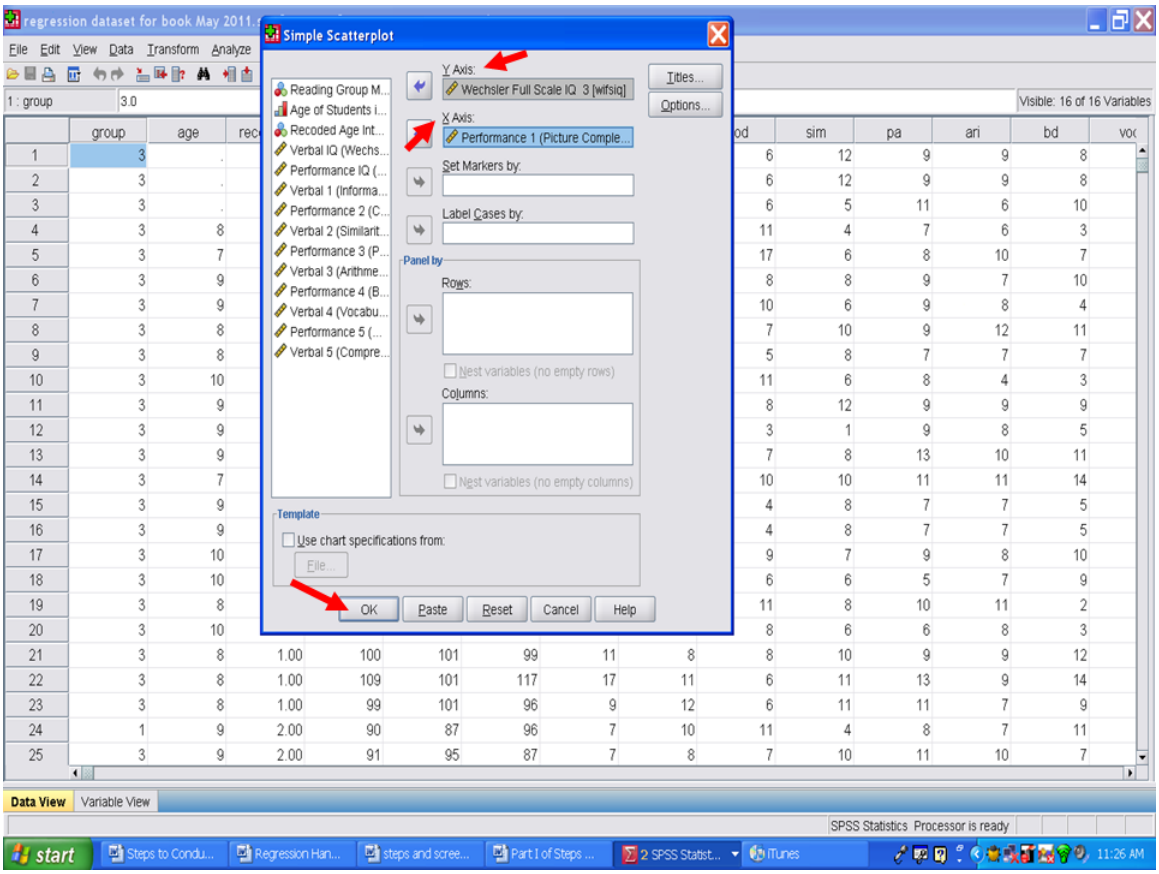


14

√ **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.**

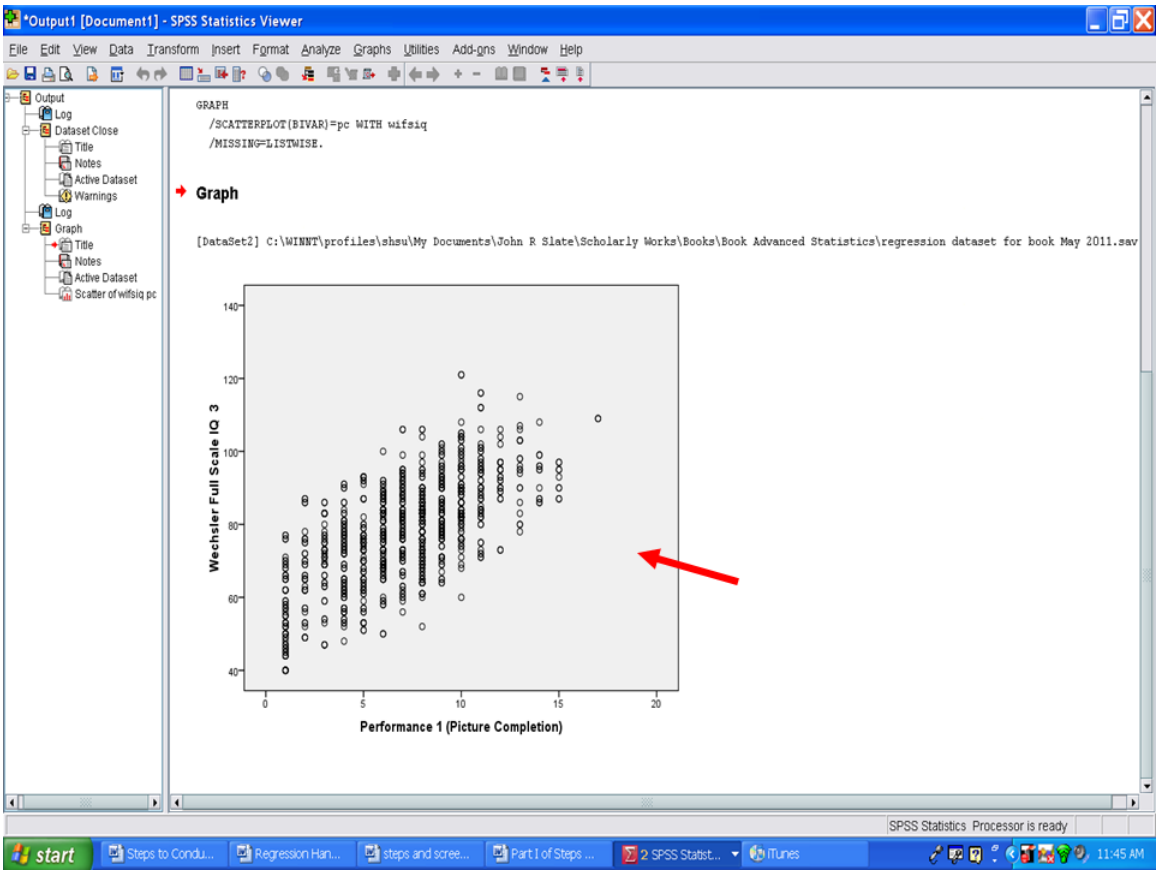
[14] http://cnx.org/content/m40738/latest/17.4.png/image



15

√ **Look at the scatterplot to determine whether a linear relationship is present. In the screenshot below, the relationship is very clearly linear.**

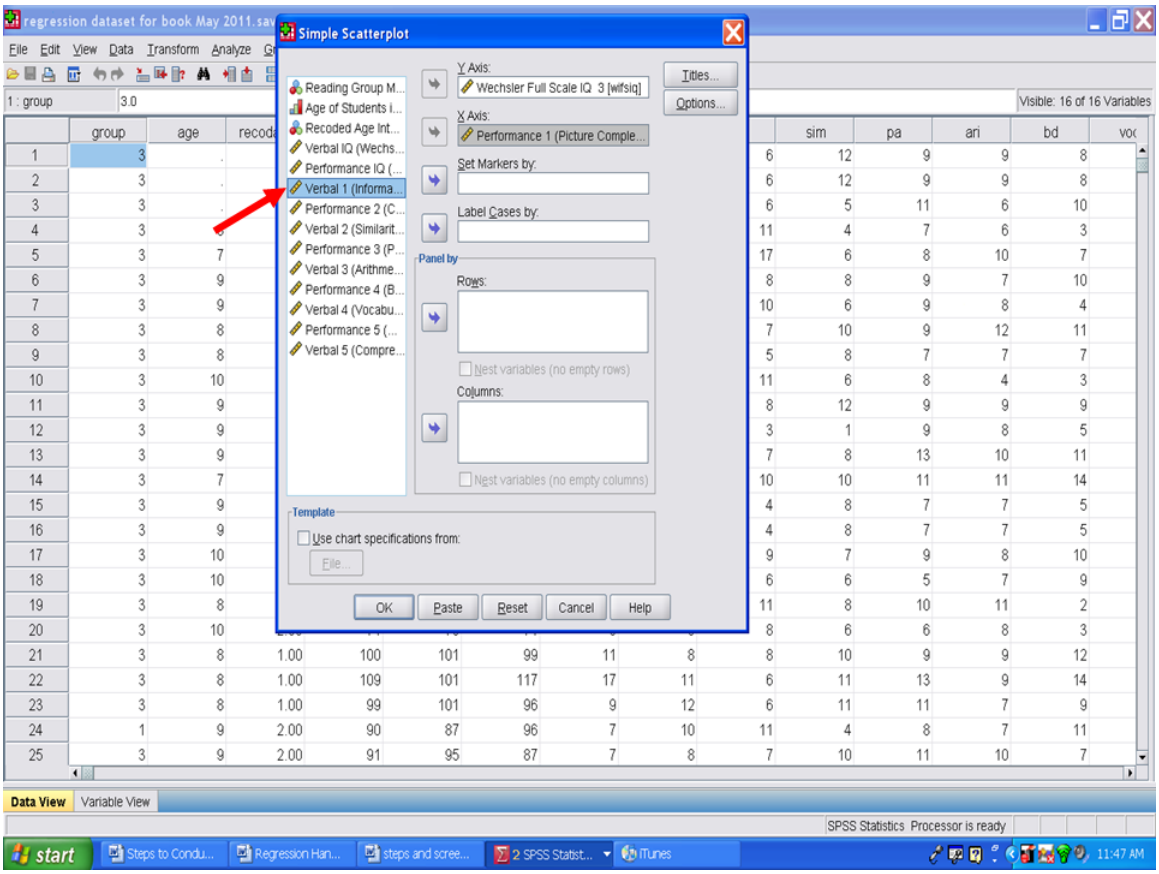
[15] http://cnx.org/content/m40738/latest/17.5.png/image



16

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



17

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 I[1]

NOTE: This chapter is published by NCPEA Press[2] 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

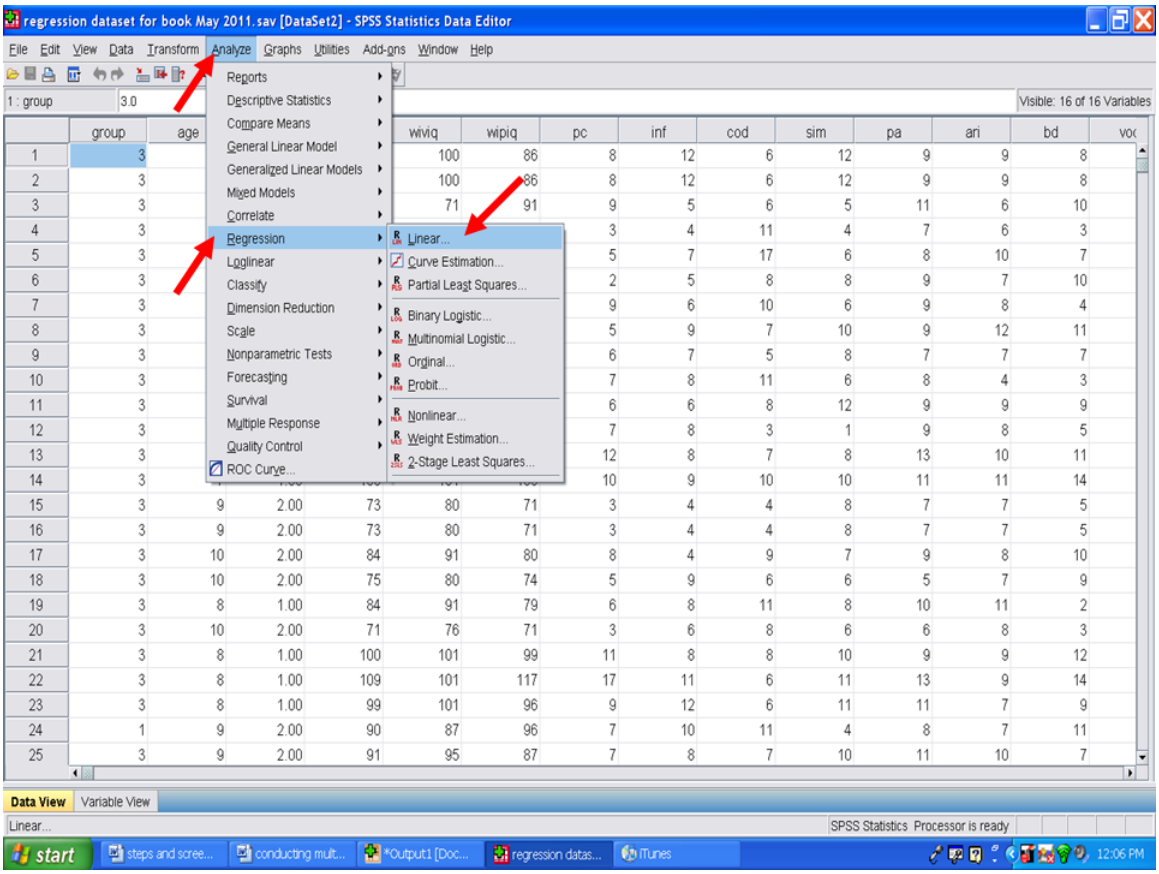
John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, 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 Help[3]) 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**

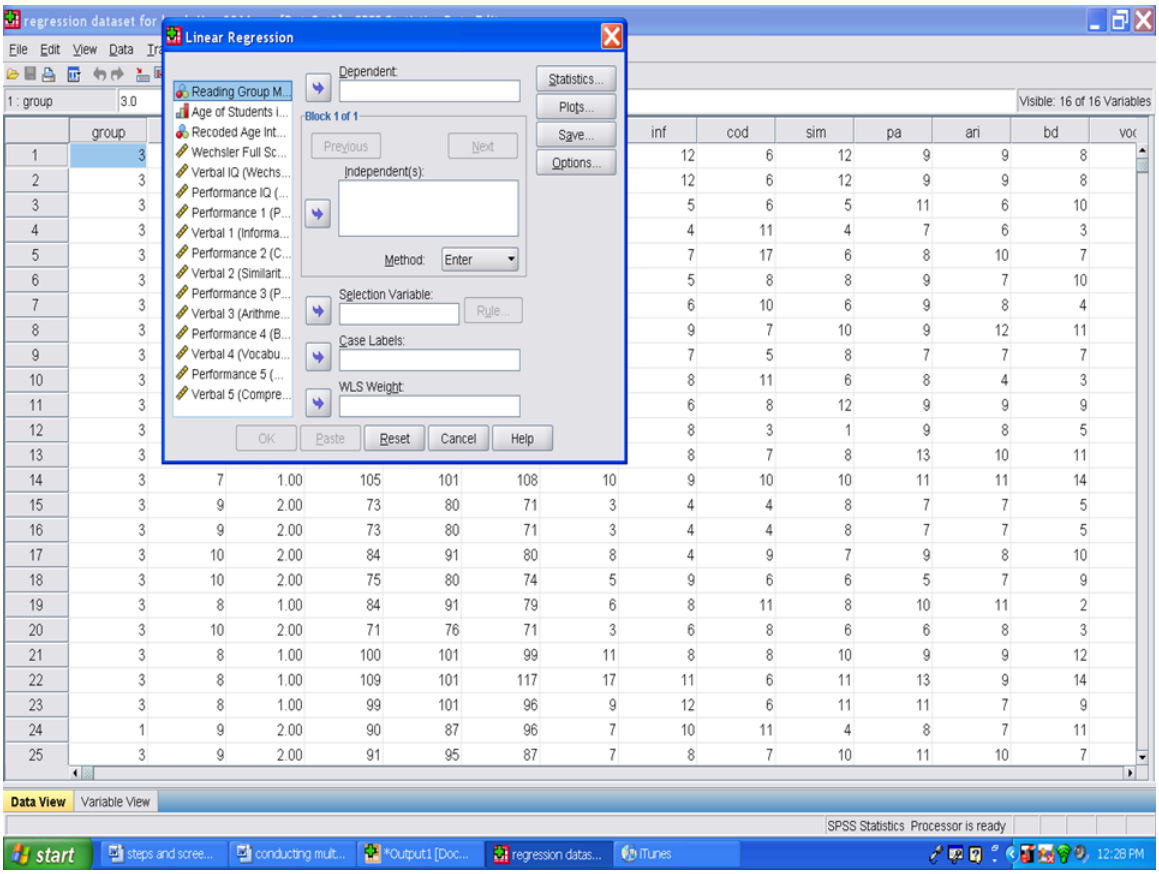
[1] This content is available online at <http://cnx.org/content/m40739/1.2/>.
[2] http://www.ncpeapublications.org/books.html
[3] http://cnx.org/content/m40739/latest/www.writingandstatisticalhelp



[4]

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.

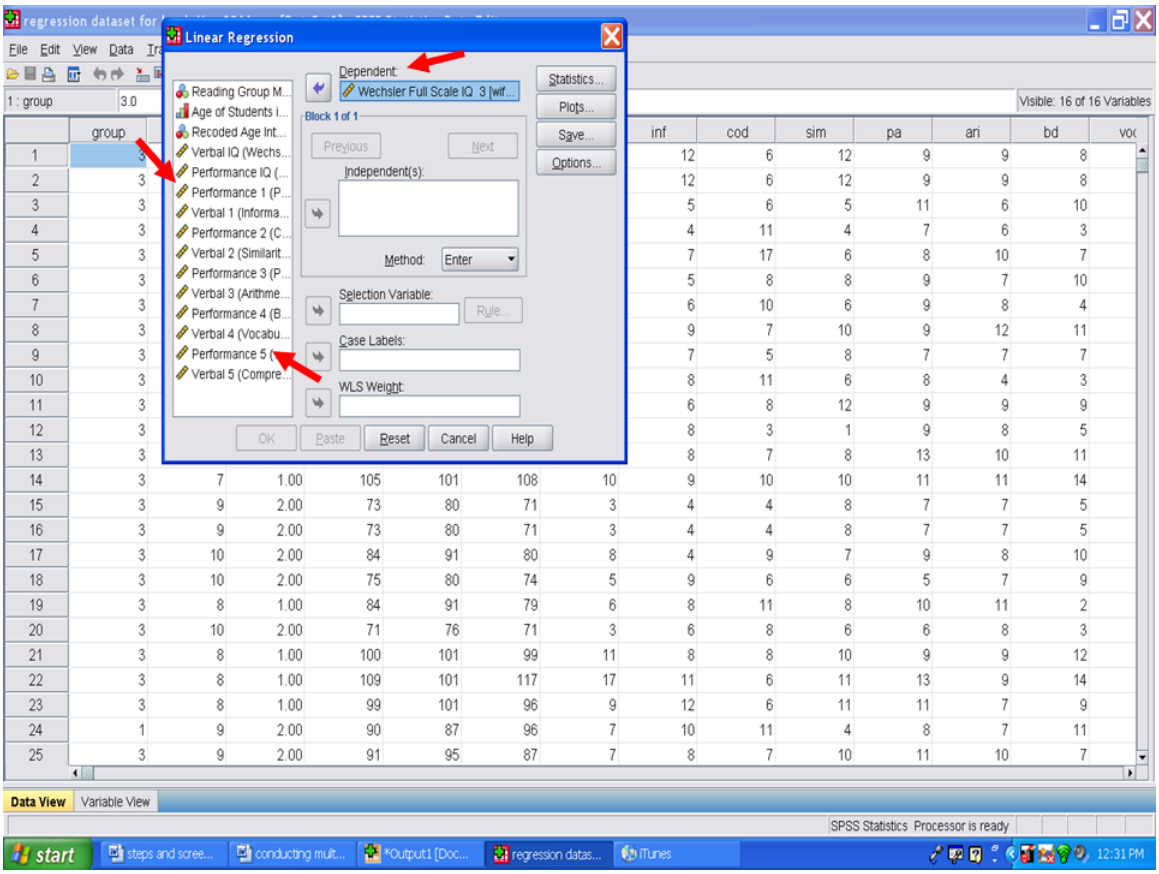
[4] http://cnx.org/content/m40739/latest/18.1.png/image



[5]

Your screen should now look like the following one. Next we will place all of the independent variables in the Independent(s) box. I this example, remember that we are using Performance 1 through Verbal 5.

[5] http://cnx.org/content/m40739/latest/18.2.png/image

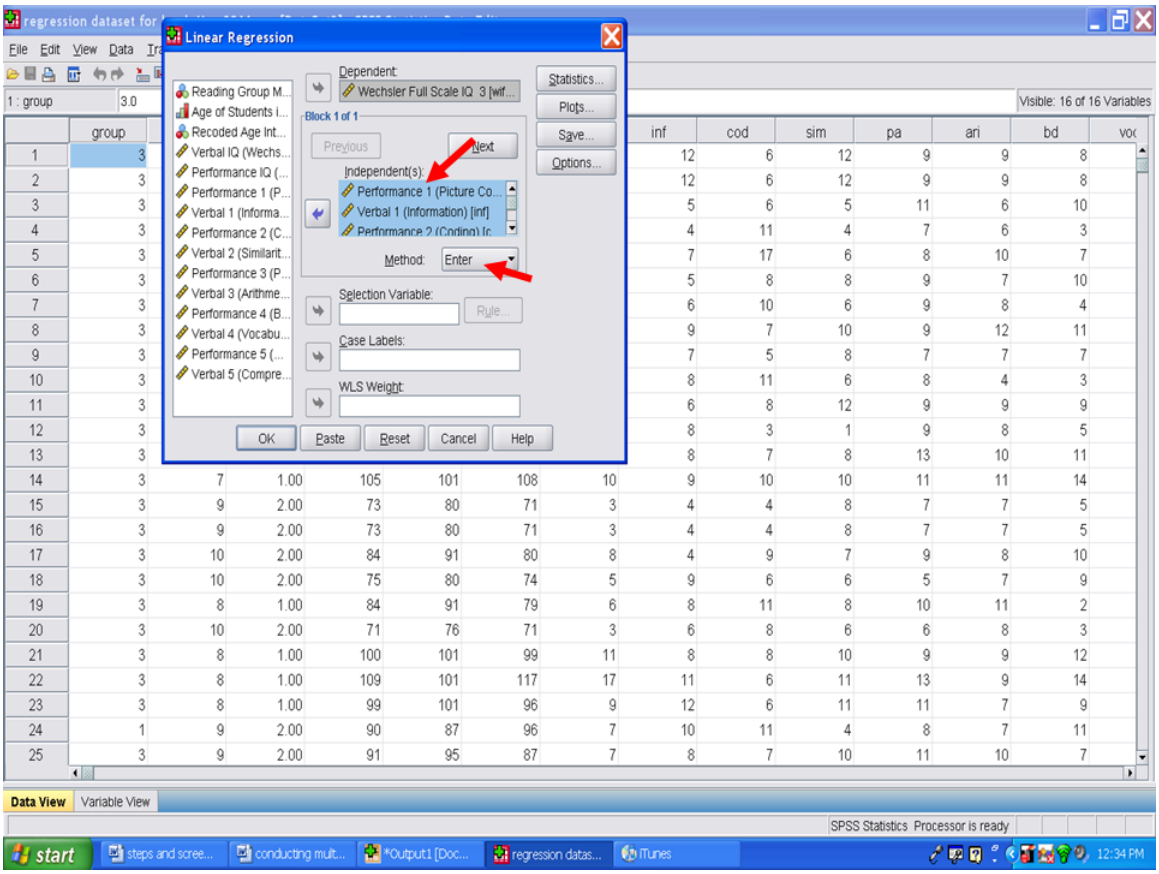


[6]

After you have placed all 10 of these variables in the Independent(s) box, your screen will look like the following one.

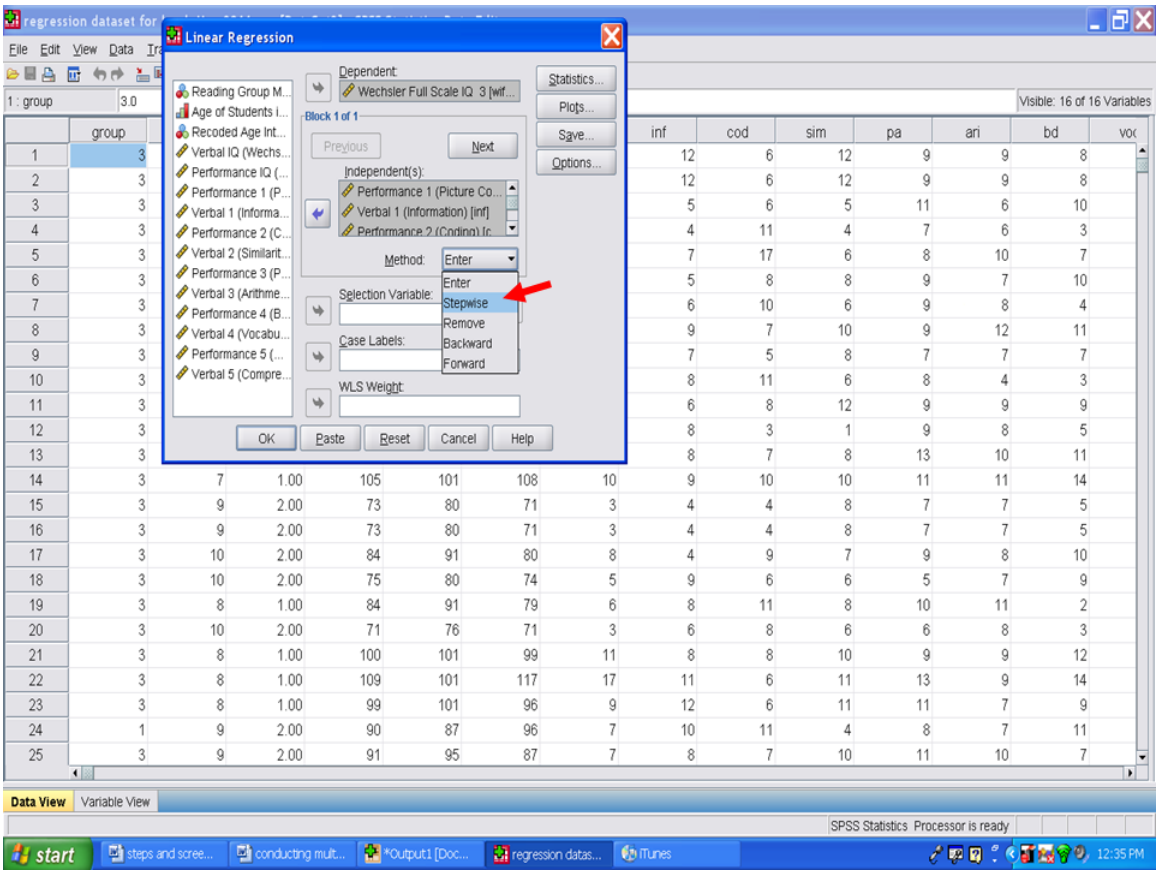
Next we will click on Method and change it from the default of Enter to Stepwise.

[6] http://cnx.org/content/m40739/latest/18.3.png/image



Click on Stepwise.

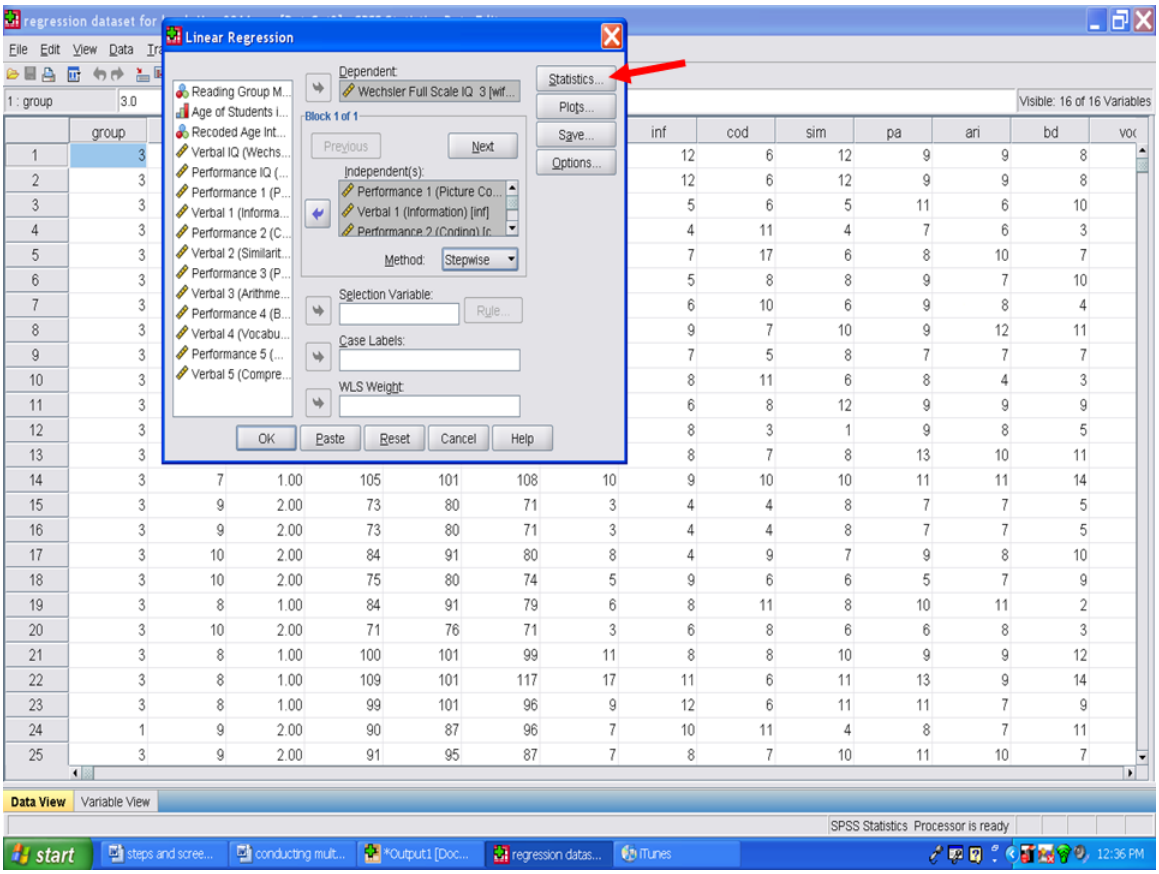
[7] http://cnx.org/content/m40739/latest/18.4.png/image



[8]

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

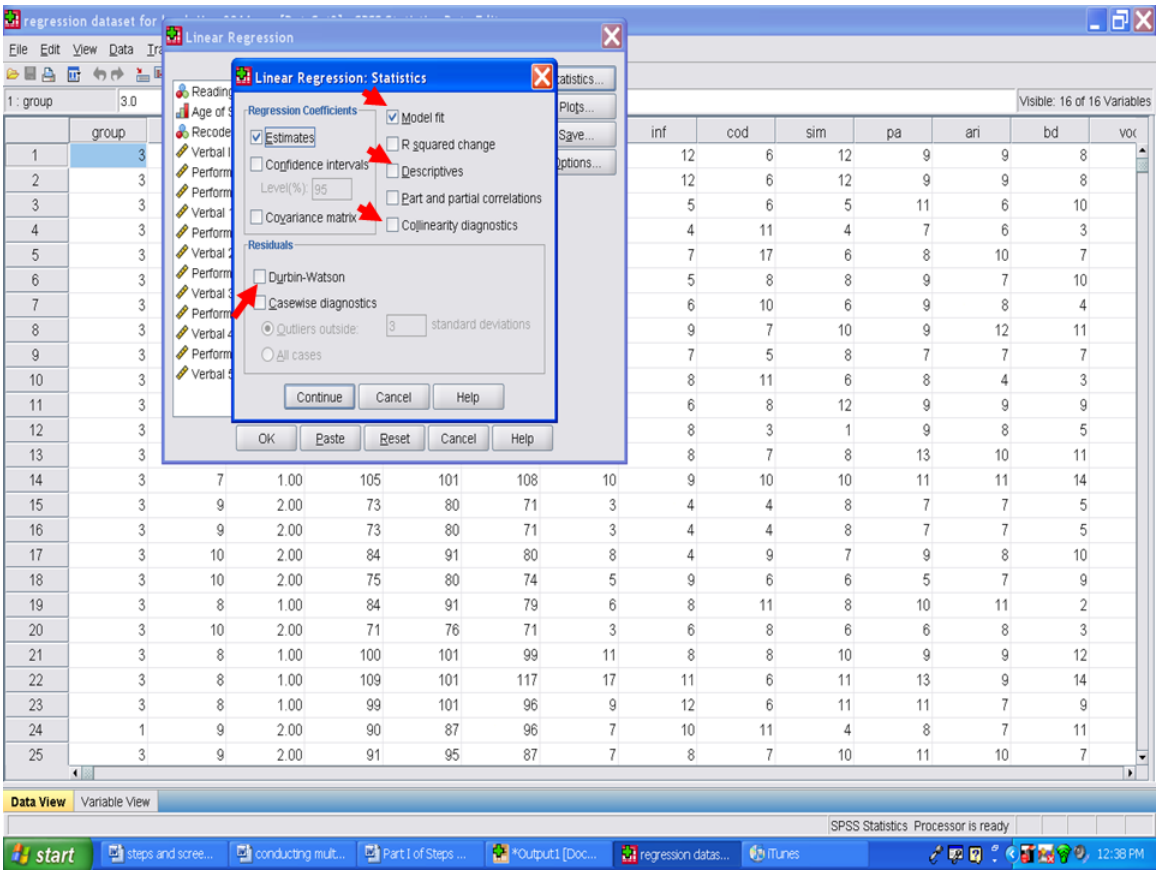


[9]

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

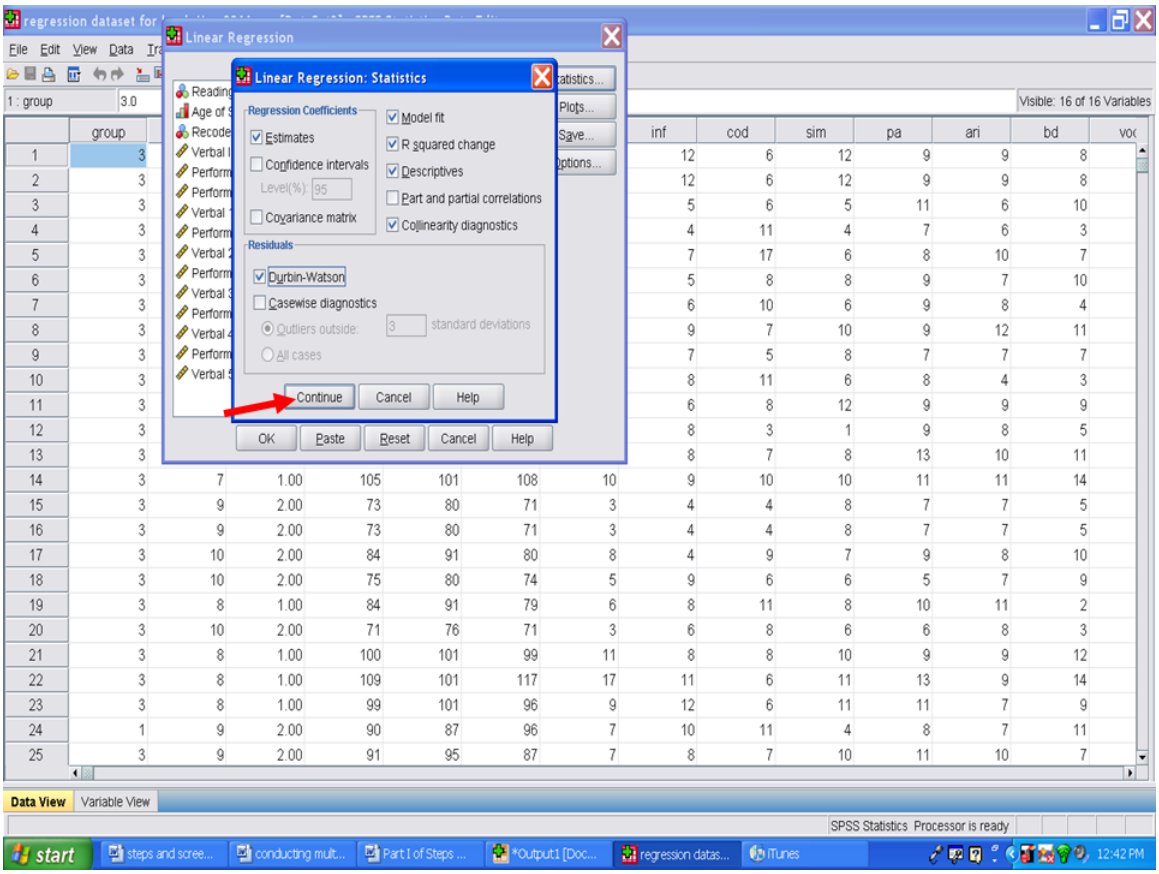
[9] http://cnx.org/content/m40739/latest/18.6.png/image



10

Your screen should like the one below. Now click on Continue.

[10] http://cnx.org/content/m40739/latest/18.7.png/image



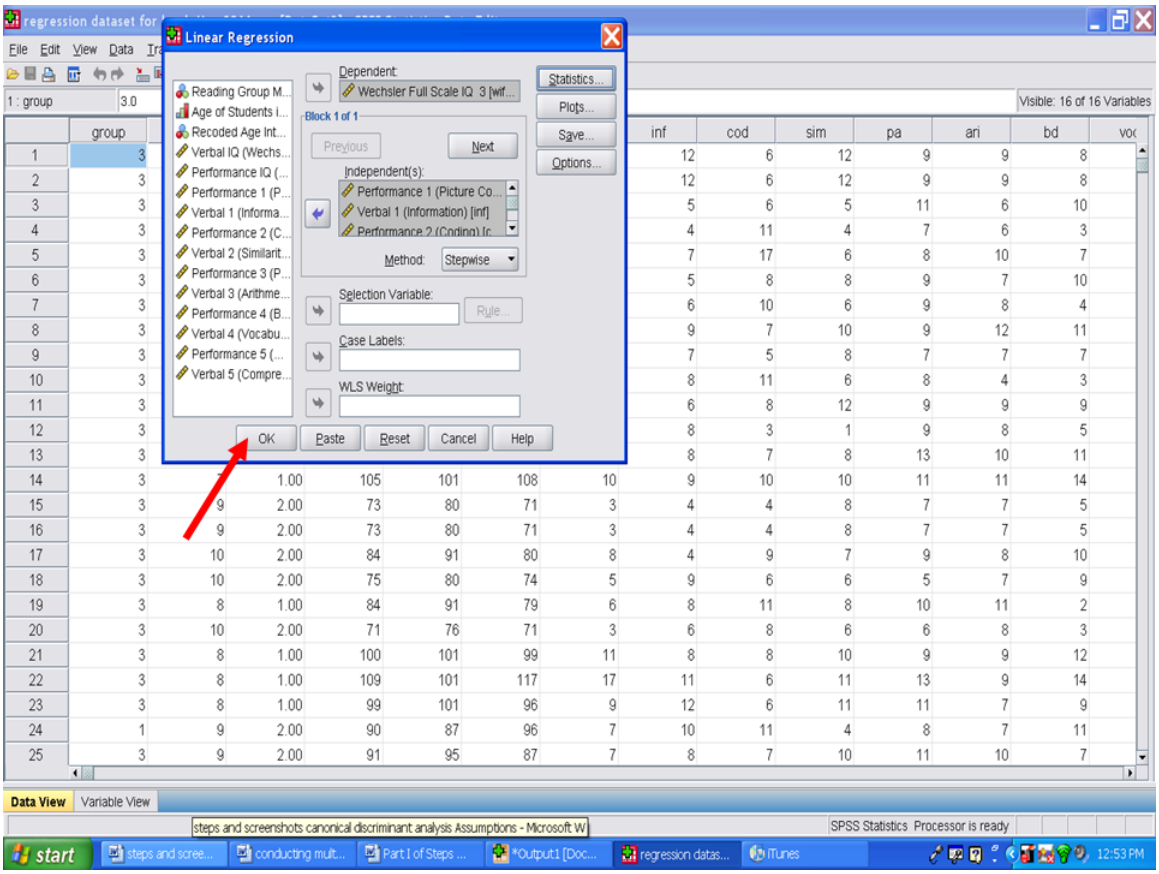
[11]

Next click on

√ OK

Go to the Output file if SPSS does not auomatically send you there.

[11] http://cnx.org/content/m40739/latest/18.8.png/image



[12]

[12] http://cnx.org/content/m40739/latest/18.89.png/image

Chapter 20

20. Multiple Regression: Part II[1]

NOTE: This chapter is published by NCPEA Press[2] 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.

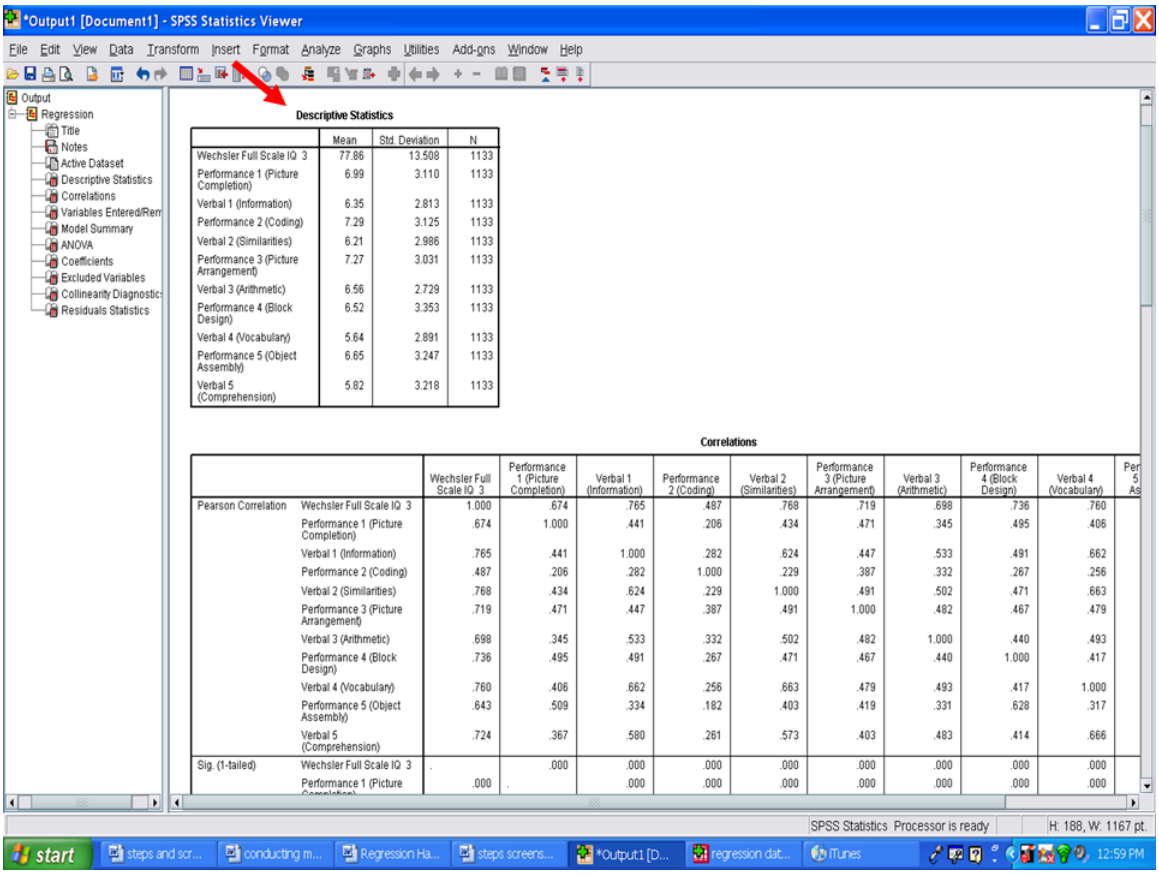
About the Authors

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 Help[3]) 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.

[1] This 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

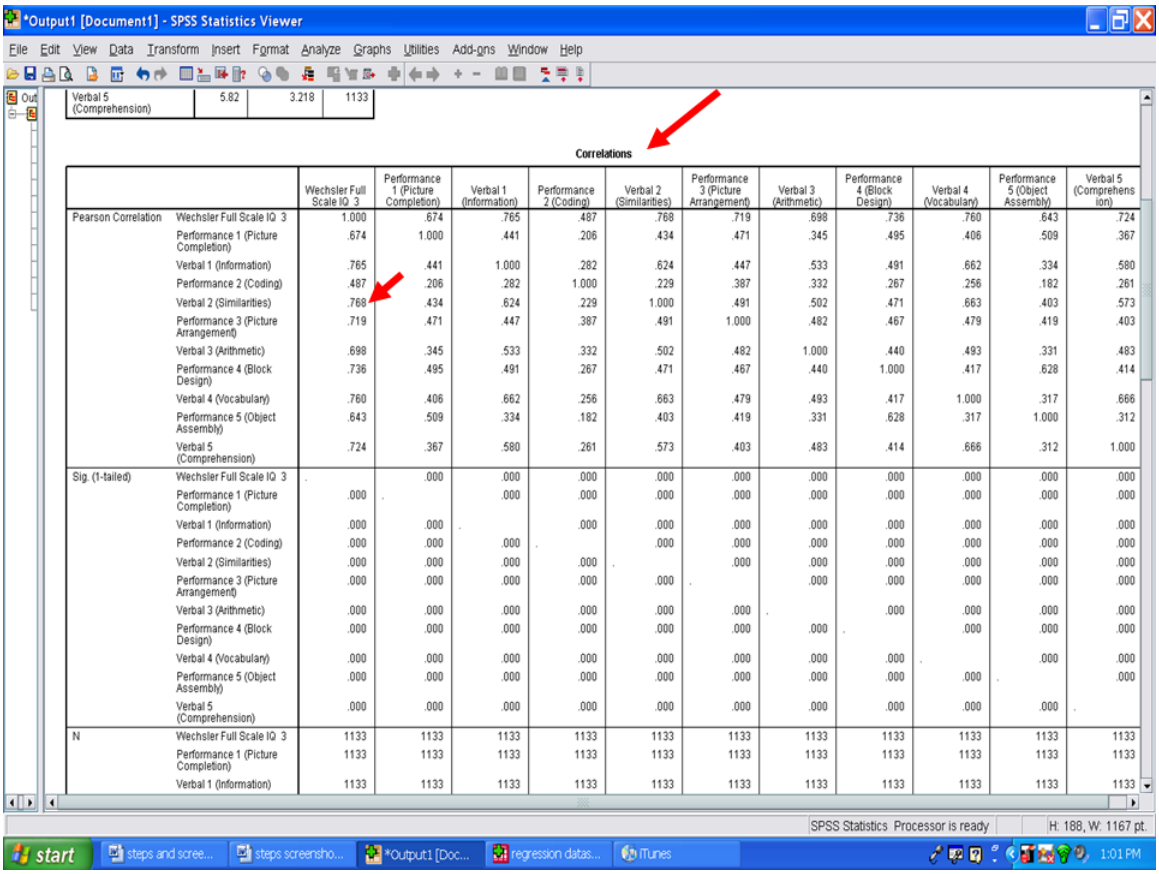


[4]

Underneath the Descriptive Statistics table is a table labeled Correlations. This table reflects the Pearson rs 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.

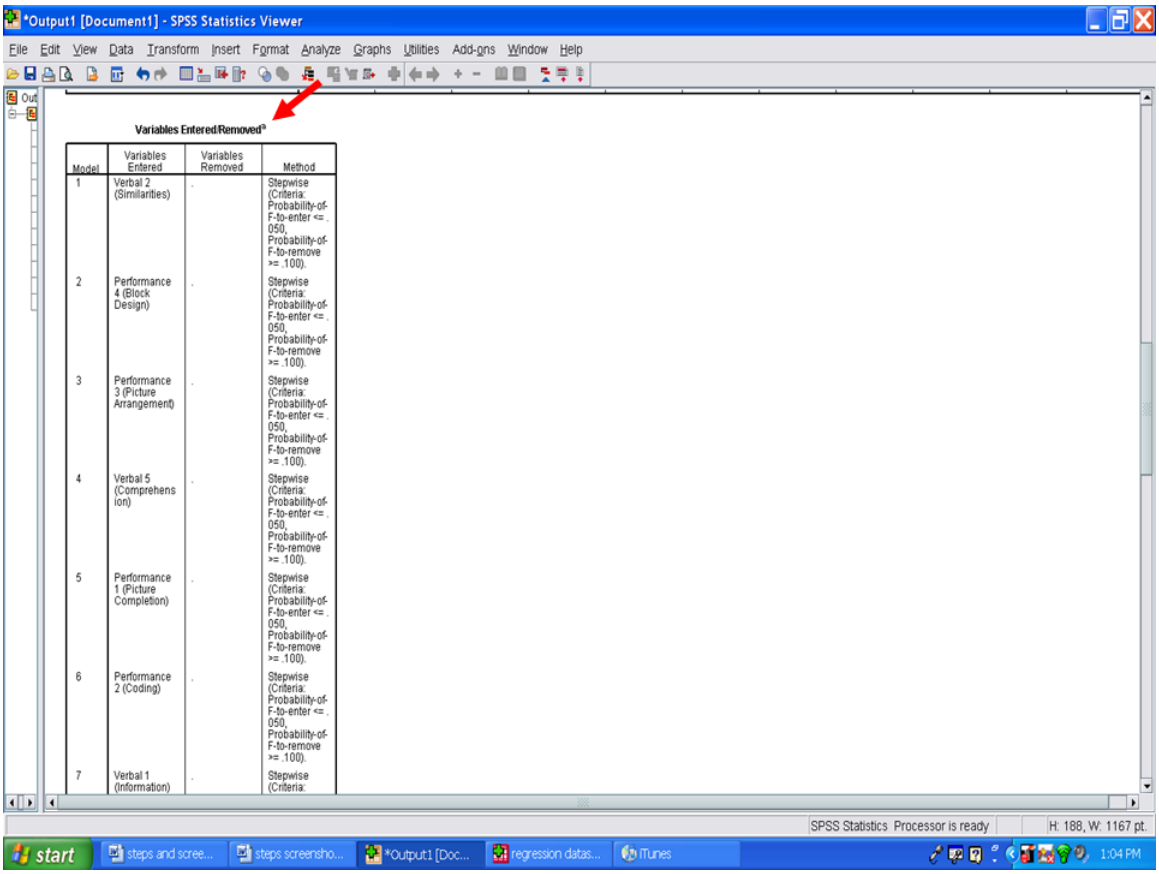
[4] http://cnx.org/content/m40742/latest/19.1.png/image



[5]

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

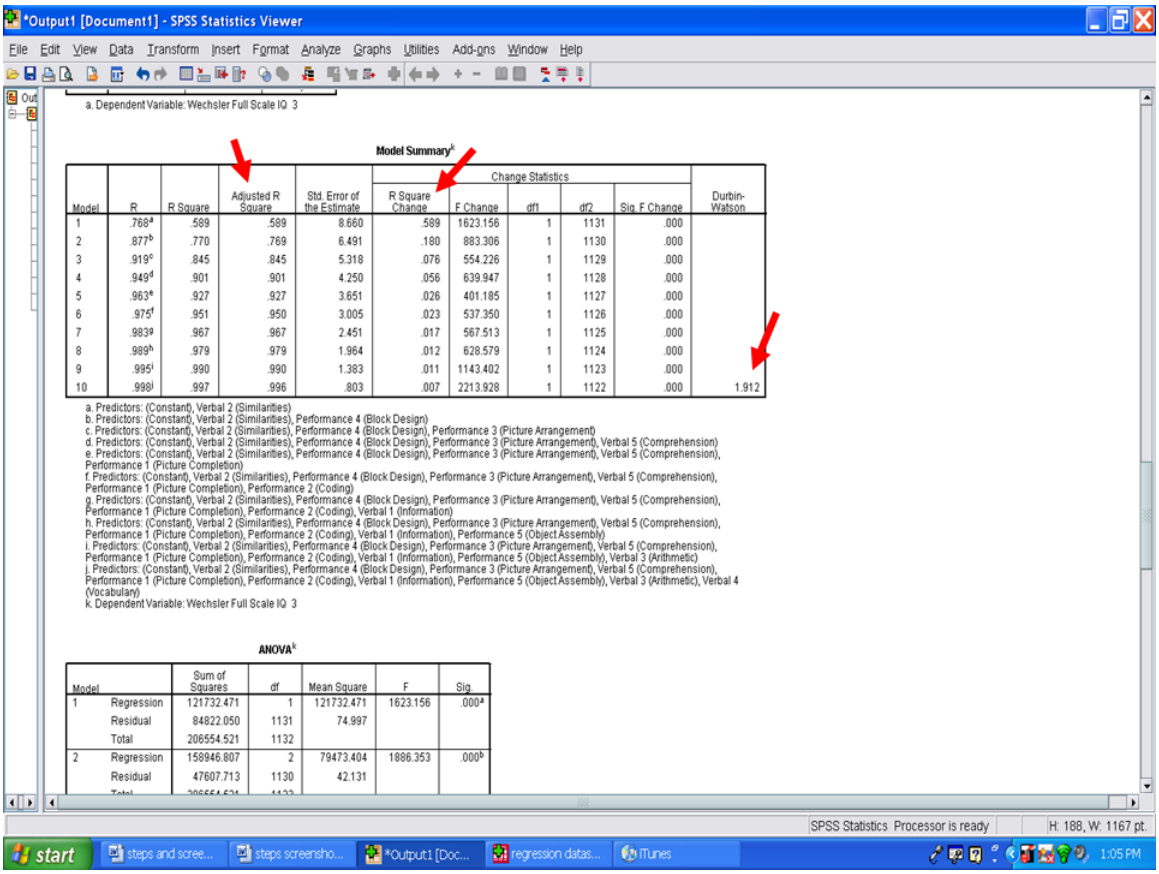


[6]

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

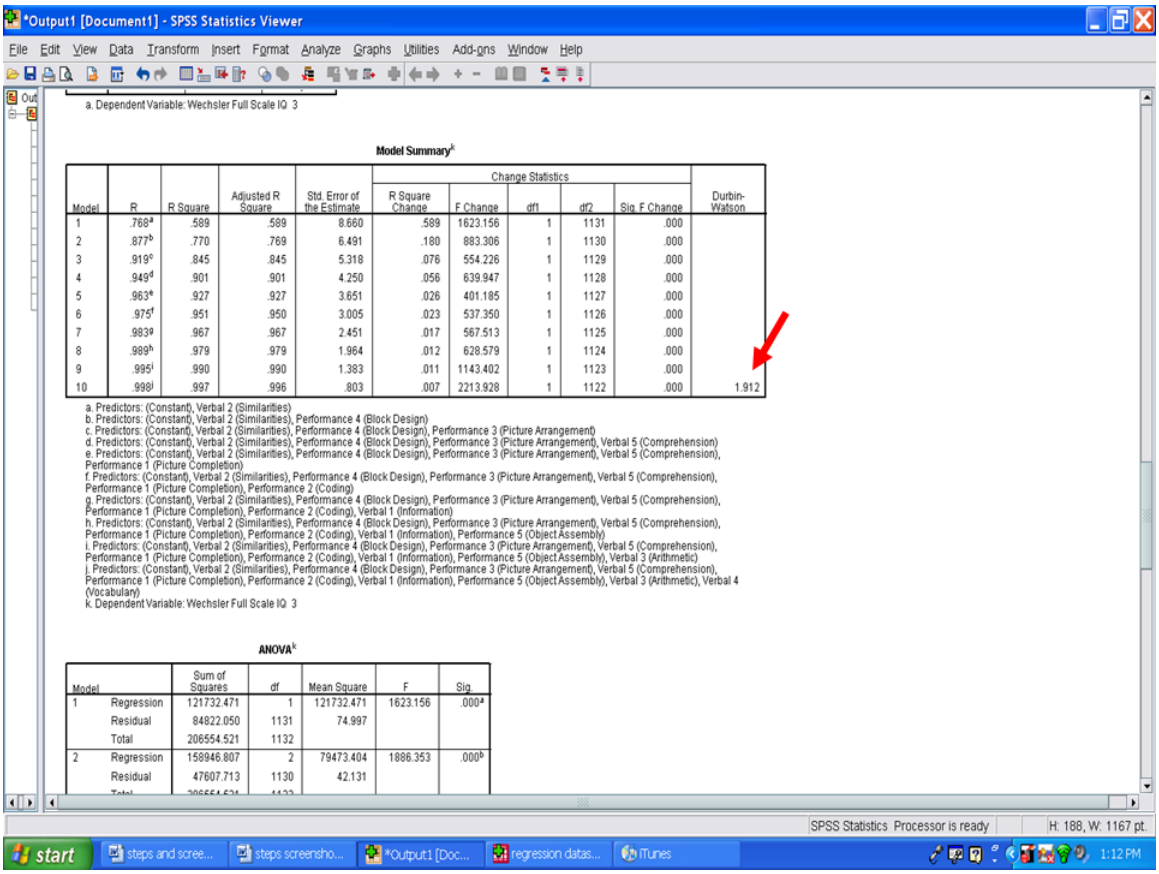
[6] http://cnx.org/content/m40742/latest/19.3.png/image



[7]

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.

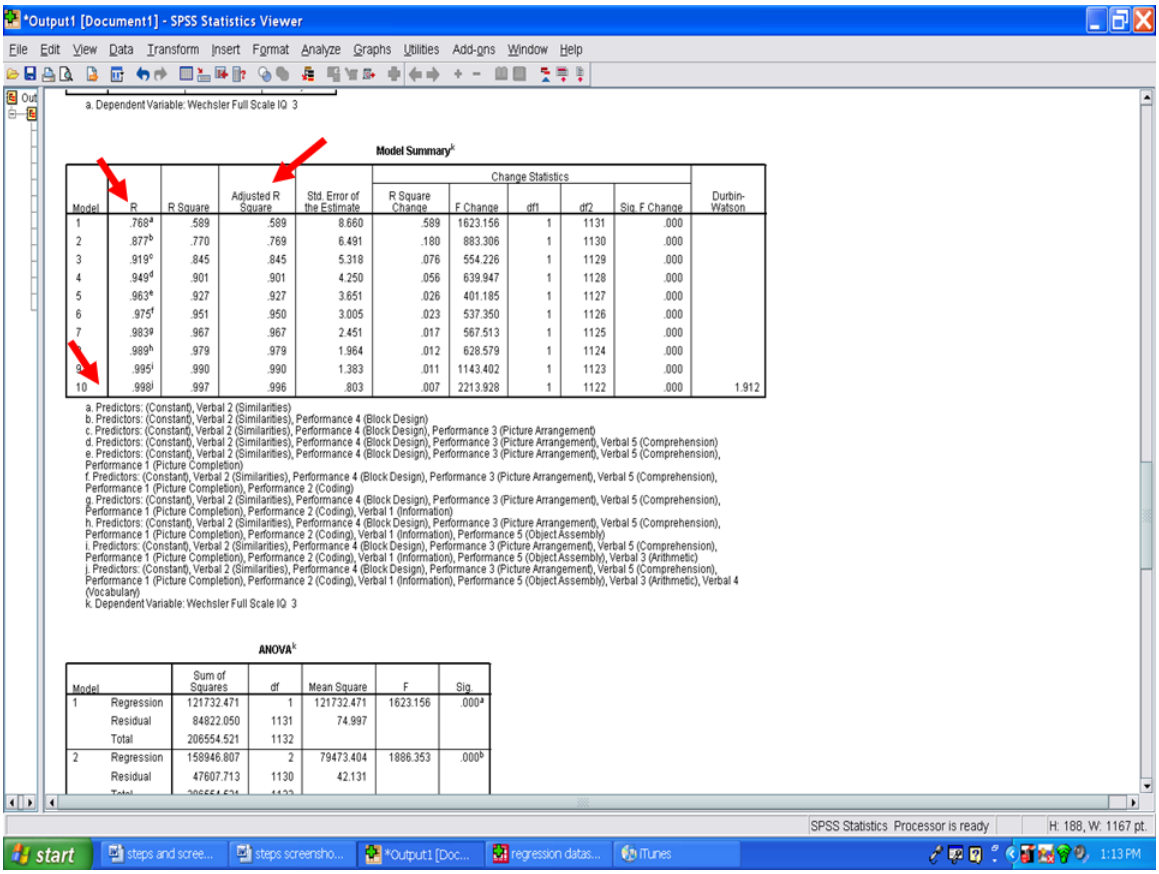
[7] http://cnx.org/content/m40742/latest/19.4.png/image



[8]

The Adjusted R Square column indicates the amount of variance that each model explains in the dependent variable (i.e., Wechsler Fll 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.

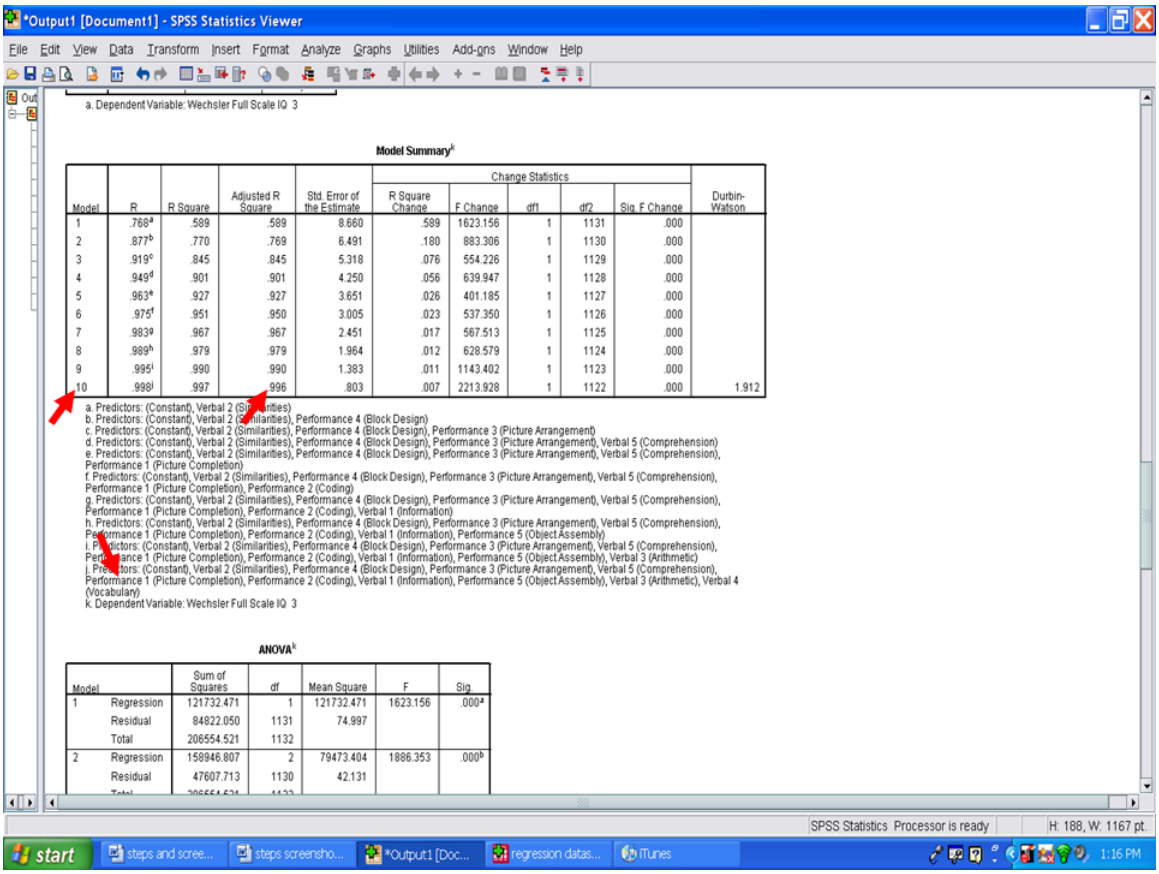
[8] http://cnx.org/content/m40742/latest/19.5.png/image



[9]

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.

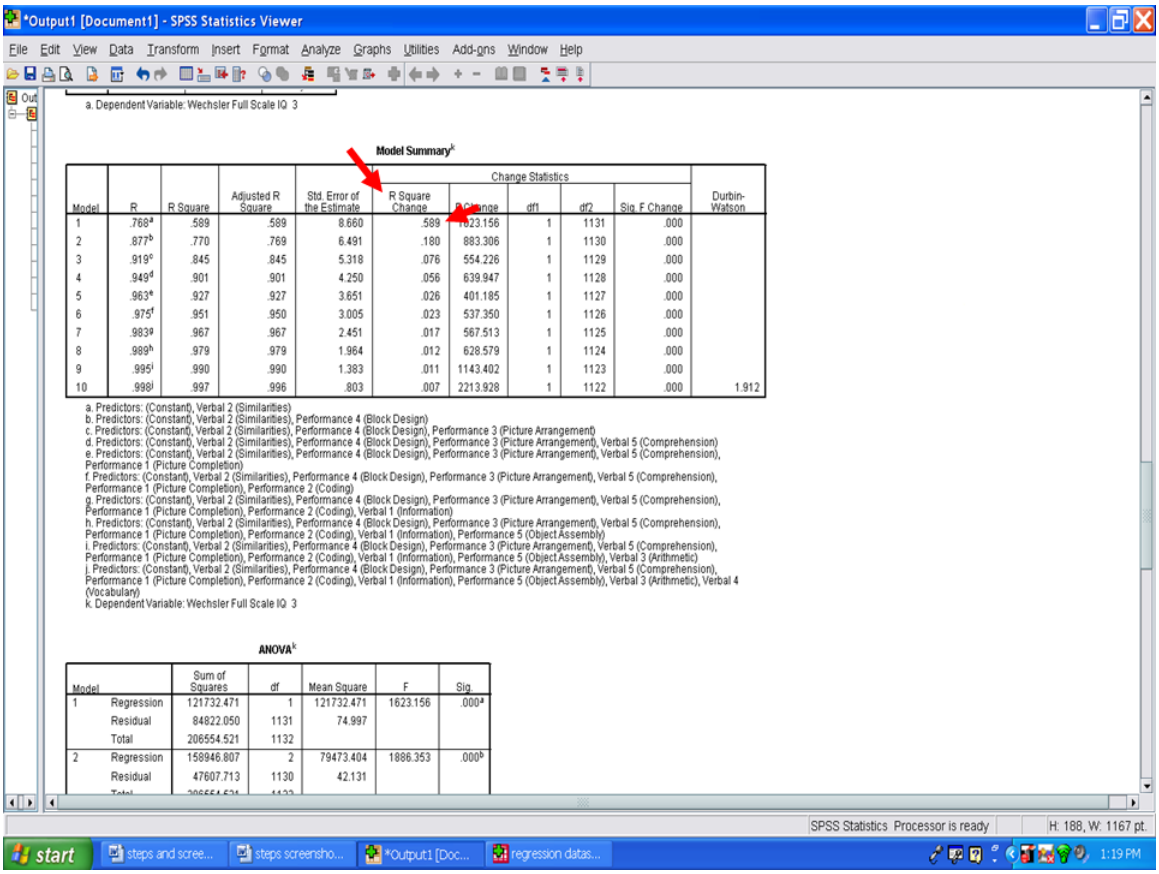
[9] http://cnx.org/content/m40742/latest/19.6.png/image



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.

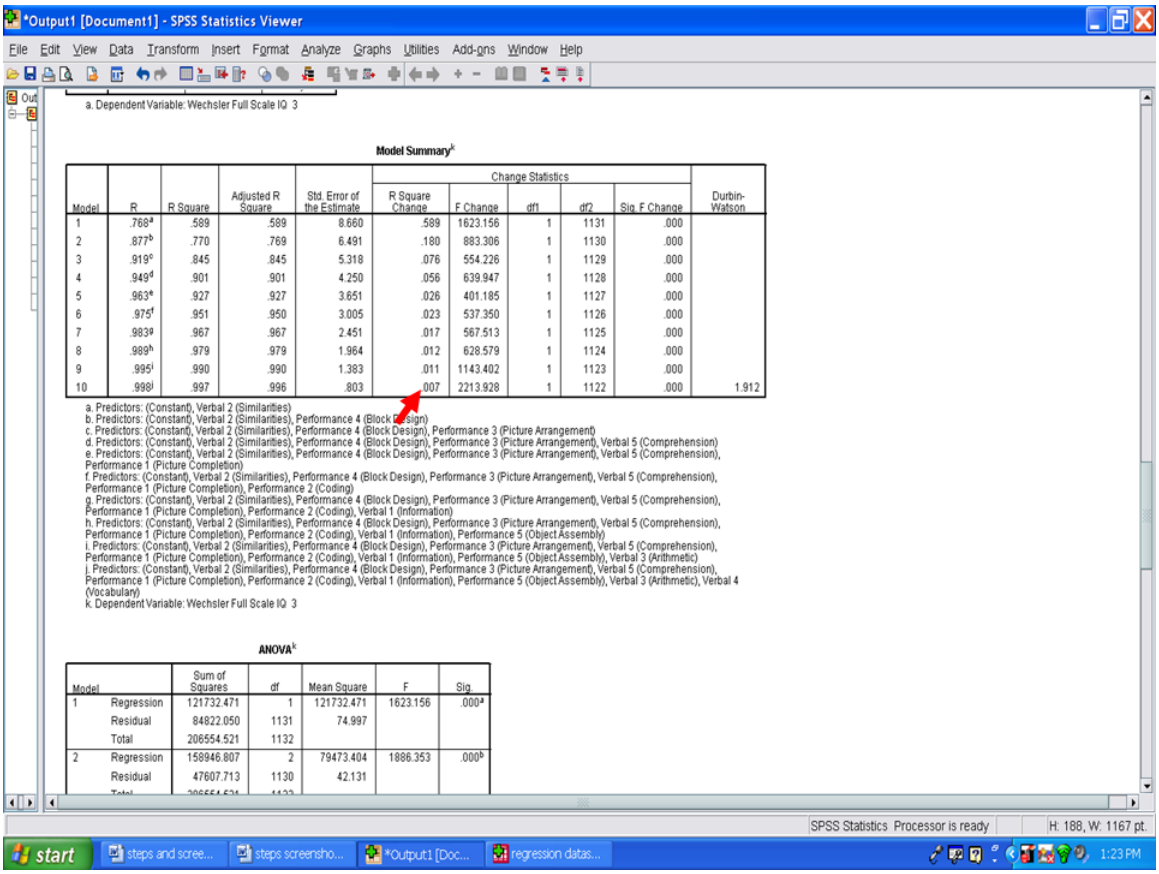
[10] http://cnx.org/content/m40742/latest/19.7.png/image



[11]

The final model, 10, only added 0.7% of additional variance explained.

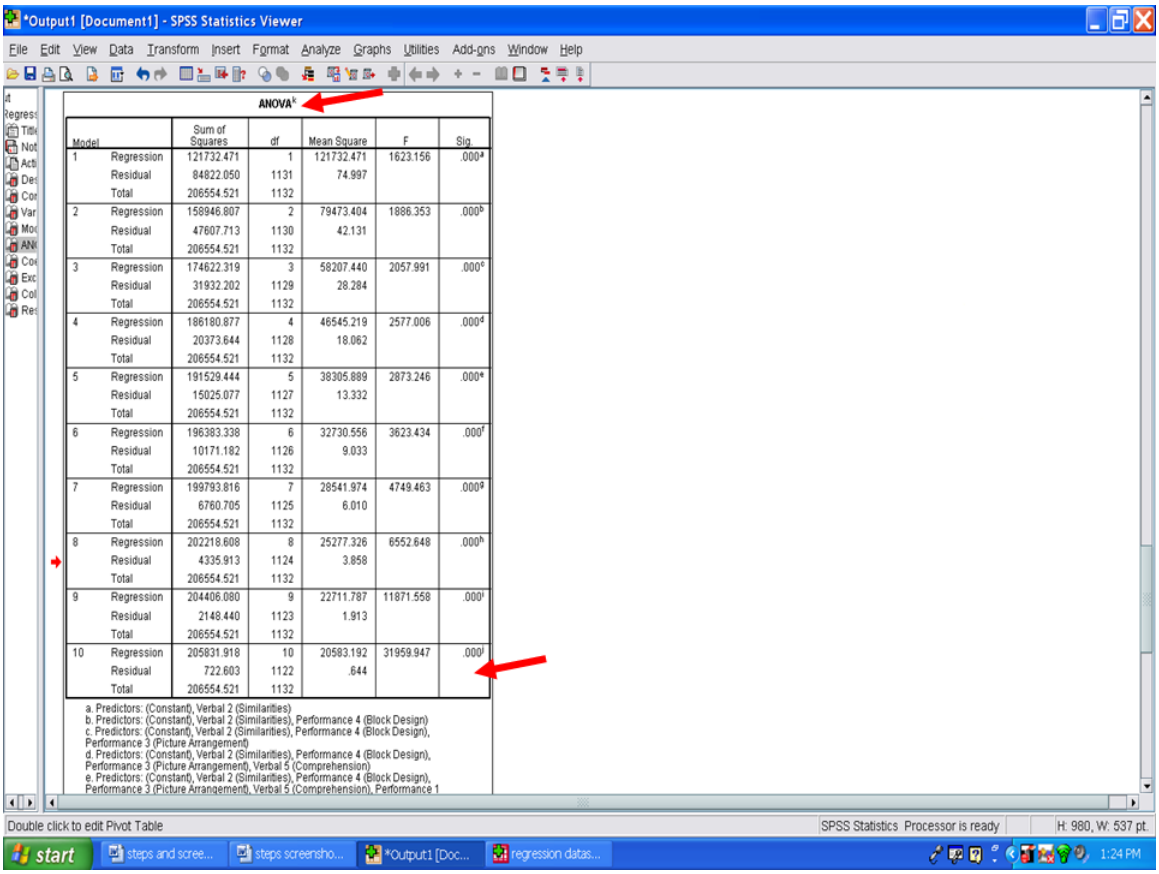
[11] http://cnx.org/content/m40742/latest/19.8.png/image



12

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, $F(10, 1122) = 31959.947$, $p < .001$.

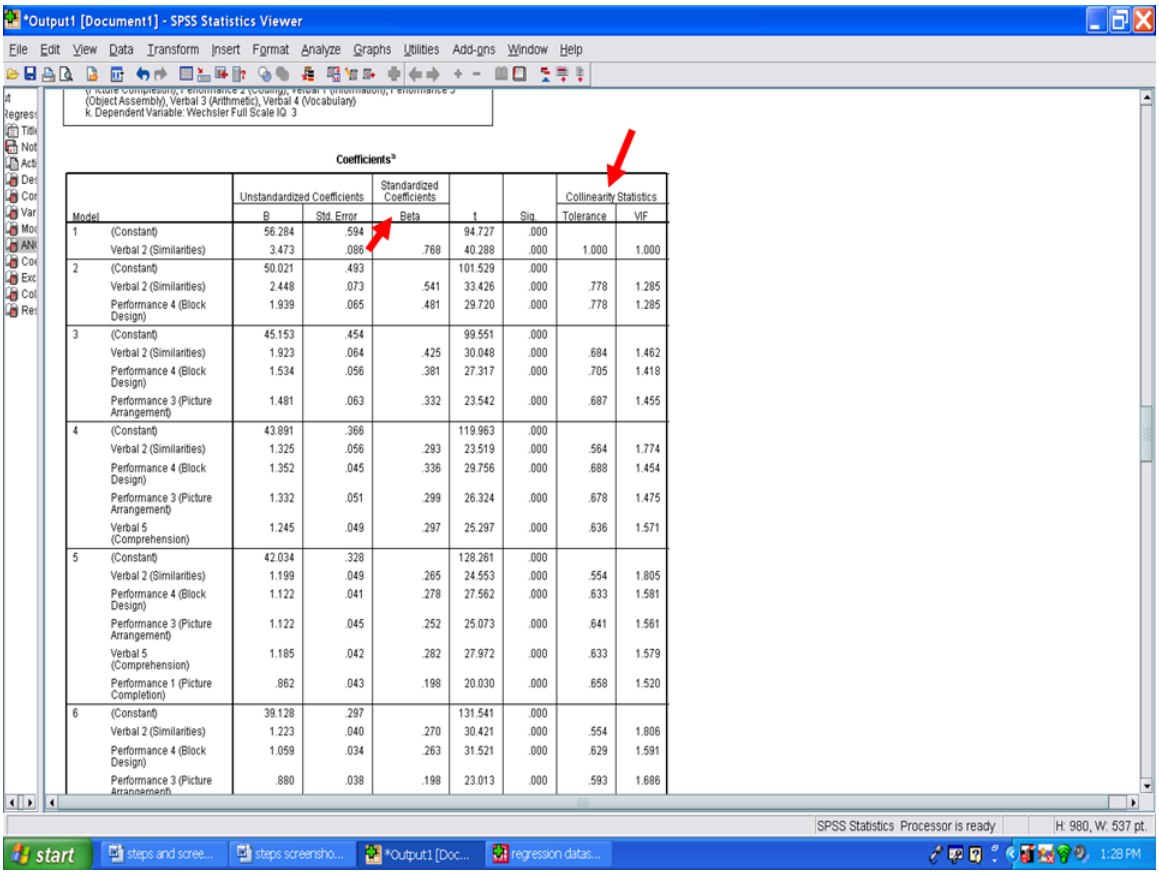
[12] http://cnx.org/content/m40742/latest/19.9.png/image



13

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.

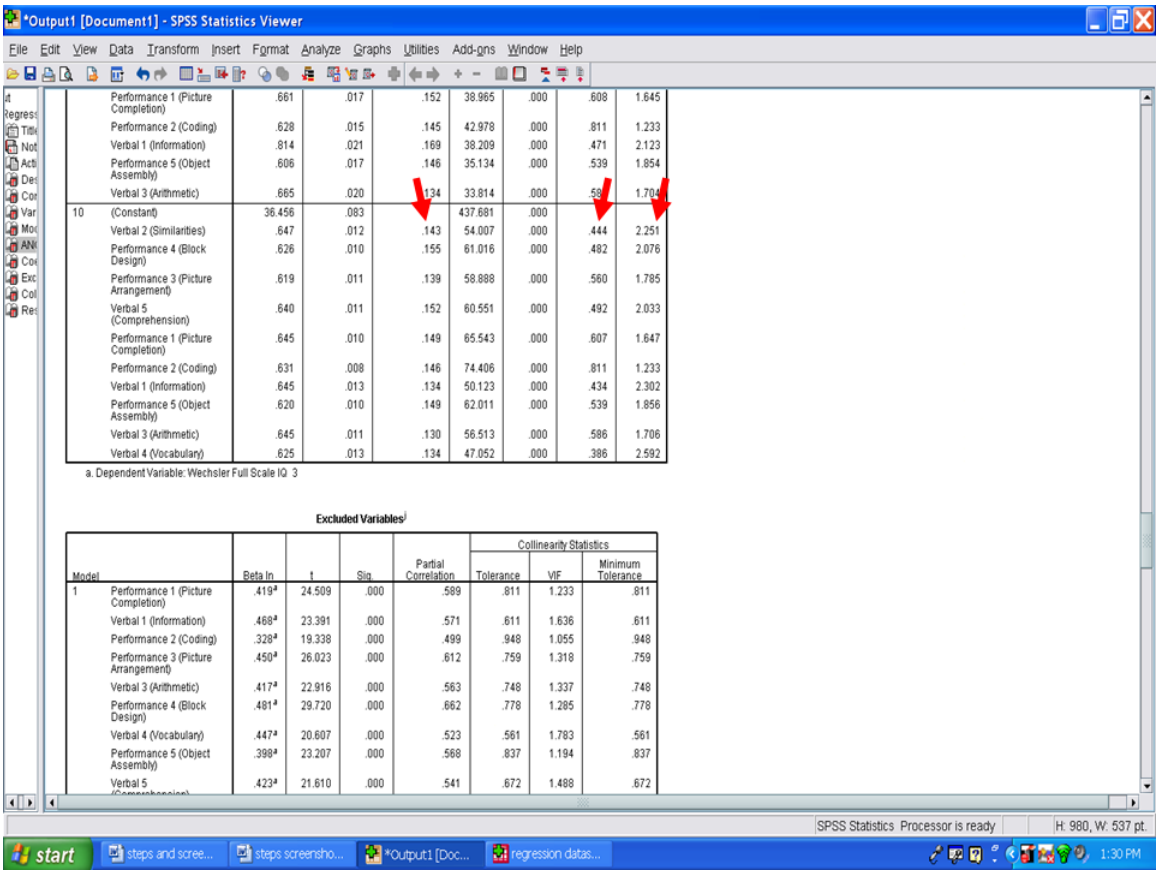
[13] http://cnx.org/content/m40742/latest/19.10.png/image



[14]

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.

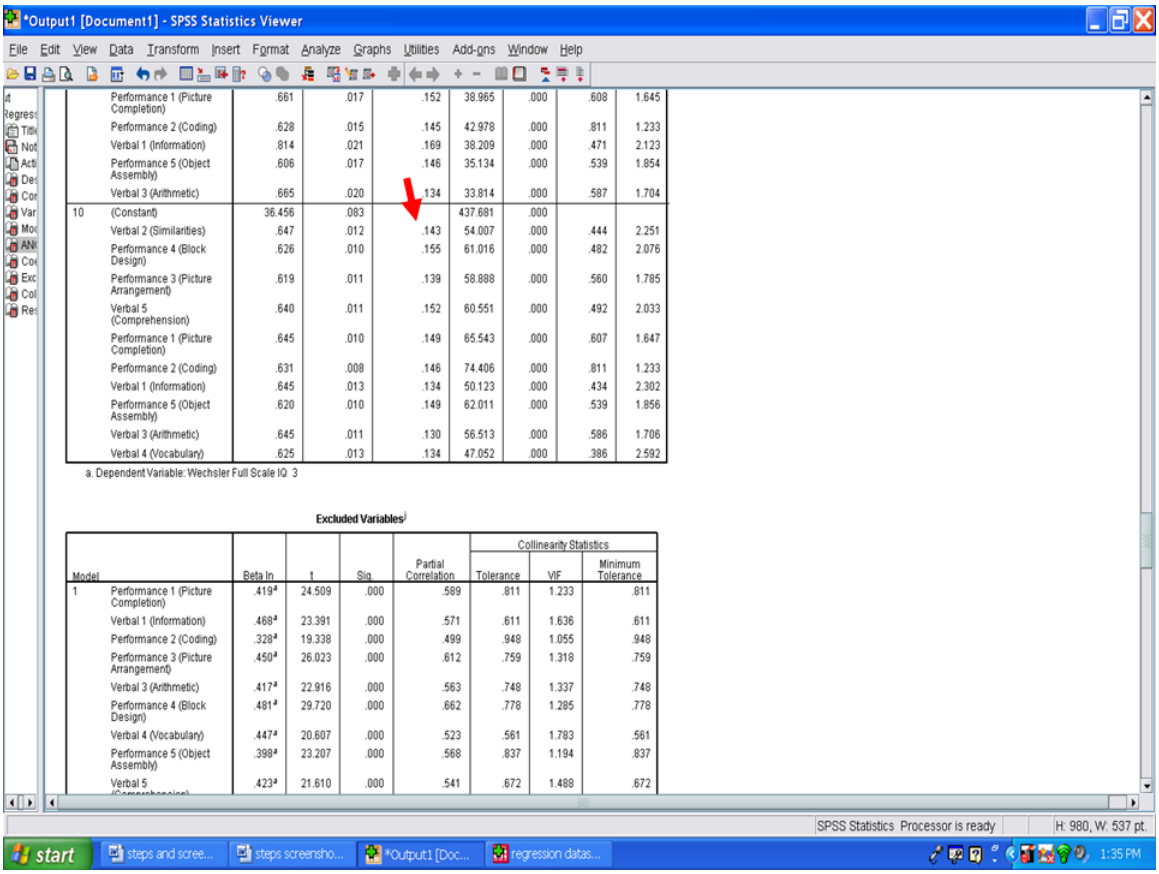
[14] http://cnx.org/content/m40742/latest/19.11.png/image



[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



16

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/col11346/1.1/
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

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

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

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

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

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

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

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

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

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.