

# Calculating Advanced Statistics: Part II

**Collection Editor:**

National Council of Professors of Educational Administration



# Calculating Advanced Statistics: Part II

**Collection Editor:**

National Council of Professors of Educational Administration

**Authors:**

Ana Rojas-LeBouef

John R. Slate

**Online:**

< <http://cnx.org/content/col11348/1.2/> >

**C O N N E X I O N S**

**Rice University, Houston, Texas**





## Table of Contents

<b>1 Multivariate Analysis of Variance: Part I</b> .....	1
<b>2 Multivariate Analysis of Variance: Part II</b> .....	29
<b>3 Multiple Analysis of Variance: Part III</b> .....	43
<b>4 Discriminant Analysis: Assumptions</b> .....	55
<b>5 Discriminant Analysis: Part I</b> .....	63
<b>6 Discriminant Analysis: Part II</b> .....	73
<b>7 Discriminant Analysis: Part III</b> .....	85
<b>8 Multiple Regression: Assumptions</b> .....	95
<b>9 Multiple Regression: Part I</b> .....	105
<b>10 Multiple Regression: Part II</b> .....	115
<b>Attributions</b> .....	129



# Chapter 1

## Multivariate Analysis of Variance: Part I<sup>1</sup>



NOTE: This chapter is published by NCPEA Press<sup>2</sup> 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<sup>3</sup>) 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**

---

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

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

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

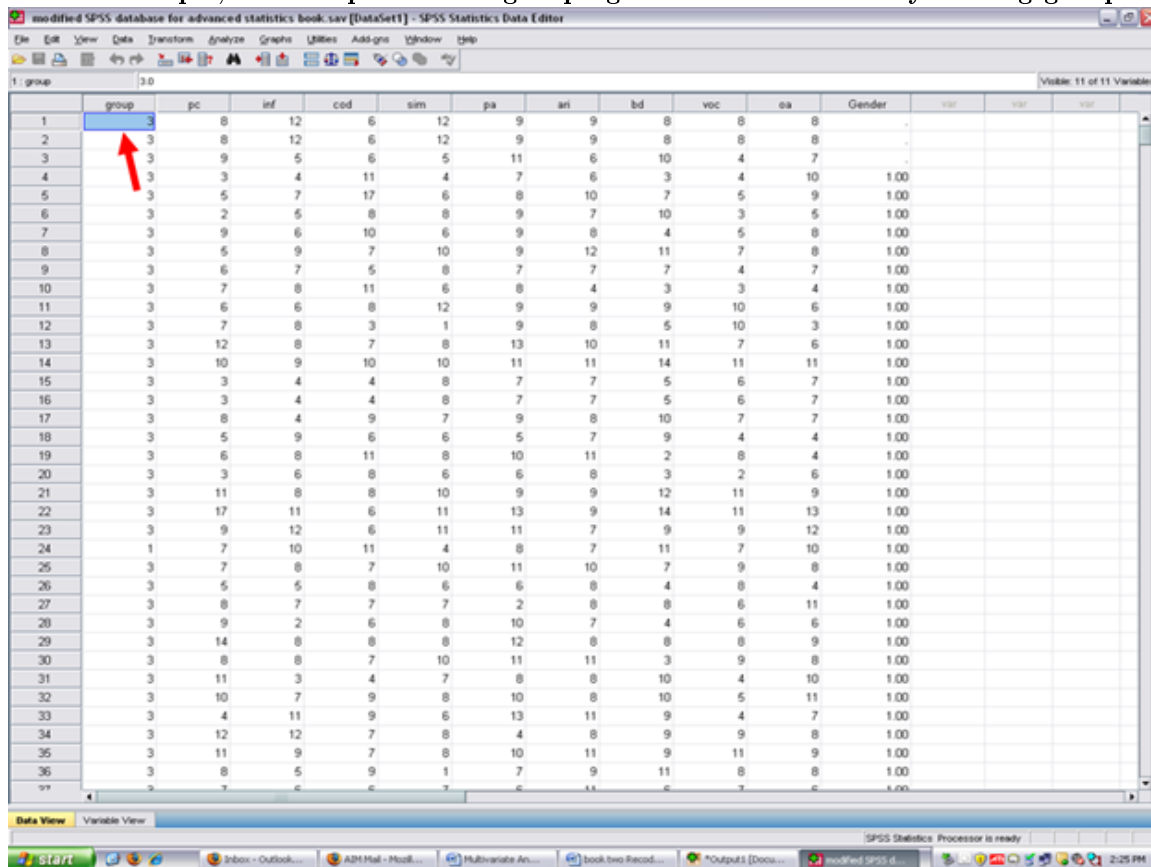
Research questions for which a MANOVA procedure is appropriate involve asking for differences in multiple dependent variables by group membership (i.e., more than two groups may be present). In addition to multiple dependent variables being present, multiple independent variables can be present as well. 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.

The screenshot displays the SPSS Statistics Data Editor window. The title bar reads "modified SPSS database for advanced statistics book.sav [DataSet1] - SPSS Statistics Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, and Help. The toolbar contains icons for file operations, editing, and analysis. The main data grid shows 36 rows and 11 variables: group, pc, inf, cod, sim, pa, ari, bd, voc, ea, and Gender. The 'group' variable is highlighted in blue. Red arrows point to the 'inf', 'sim', 'ari', and 'voc' variables. The bottom status bar indicates "SPSS Statistics Processor is ready".

	group	pc	inf	cod	sim	pa	ari	bd	voc	ea	Gender	vis1	vis2	vis3
1	3	8	12	6	12	9	9	8	8	8	-			
2	3	8	12	6	12	9	9	8	8	8	-			
3	3	9	5	6	5	11	6	10	4	7				
4	3	3	4	11	4	7	6	3	4	10	1.00			
5	3	5	7	17	6	8	10	7	5	9	1.00			
6	3	2	5	8	8	9	7	10	3	5	1.00			
7	3	9	6	10	6	9	8	4	5	8	1.00			
8	3	5	9	7	10	9	12	11	7	8	1.00			
9	3	6	7	5	8	7	7	7	4	7	1.00			
10	3	7	8	11	6	8	4	3	3	4	1.00			
11	3	6	6	8	12	9	9	9	10	6	1.00			
12	3	7	8	3	1	9	8	5	10	3	1.00			
13	3	12	8	7	8	13	10	11	7	6	1.00			
14	3	10	9	10	10	11	11	14	11	11	1.00			
15	3	3	4	4	8	7	7	5	6	7	1.00			
16	3	3	4	4	8	7	7	5	6	7	1.00			
17	3	8	4	9	7	9	8	10	7	7	1.00			
18	3	5	9	6	6	5	7	9	4	4	1.00			
19	3	6	8	11	8	10	11	2	8	4	1.00			
20	3	3	6	8	6	6	8	3	2	6	1.00			
21	3	11	8	8	10	9	9	12	11	9	1.00			
22	3	17	11	6	11	13	9	14	11	13	1.00			
23	3	9	12	6	11	11	7	9	9	12	1.00			
24	1	7	10	11	4	8	7	11	7	10	1.00			
25	3	7	8	7	10	11	10	7	9	8	1.00			
26	3	5	5	8	6	6	8	4	8	4	1.00			
27	3	8	7	7	7	2	8	8	6	11	1.00			
28	3	9	2	6	8	10	7	4	6	6	1.00			
29	3	14	8	8	8	12	8	8	8	9	1.00			
30	3	8	8	7	10	11	11	3	9	8	1.00			
31	3	11	3	4	7	8	8	10	4	10	1.00			
32	3	10	7	9	8	10	8	10	5	11	1.00			
33	3	4	11	9	6	13	11	9	4	7	1.00			
34	3	12	12	7	8	4	8	9	9	8	1.00			
35	3	11	9	7	8	10	11	9	11	9	1.00			
36	3	8	5	9	1	7	9	11	8	8	1.00			
37	3	7	6	6	7	6	14	6	7	6	1.00			

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

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



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

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.

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

modified SPSS database for advanced statistics book.sav [DataSet1] - SPSS Statistics Data Editor

1: group 3.0

Visible 11 of 11 Variables

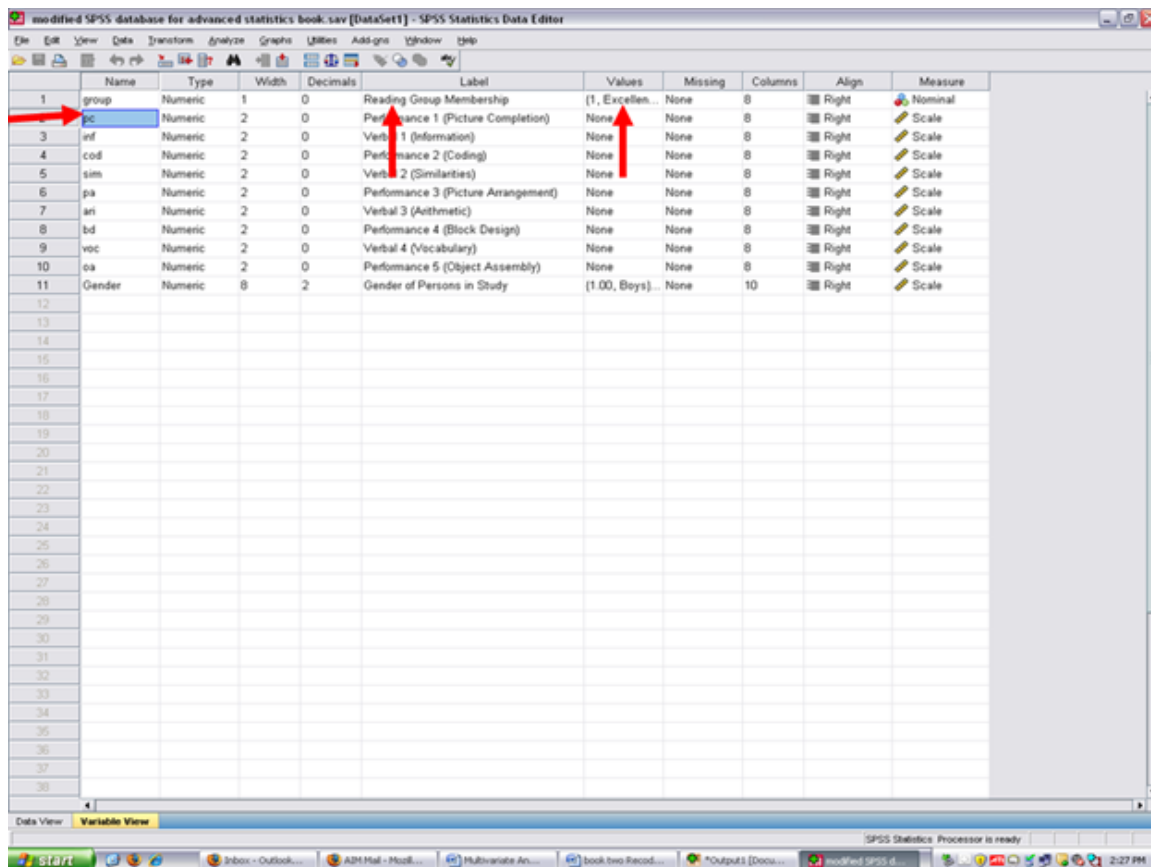
	group	pc	inf	cod	sim	pa	an	bd	voc	oa	Gender
1	3	8	12	6	12	9	9	8	8	8	1.00
2	3	8	12	6	12	9	9	8	8	8	1.00
3	3	9	5	6	5	11	6	10	4	7	1.00
4	3	3	4	11	4	7	6	3	4	10	1.00
5	3	5	7	17	6	8	10	7	5	9	1.00
6	3	2	5	8	8	9	7	10	3	5	1.00
7	3	9	6	10	6	9	8	4	5	8	1.00
8	3	5	9	7	10	9	12	11	7	8	1.00
9	3	6	7	5	8	7	7	7	4	7	1.00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

Data View Variable View

SPSS Statistics Processor is ready

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.

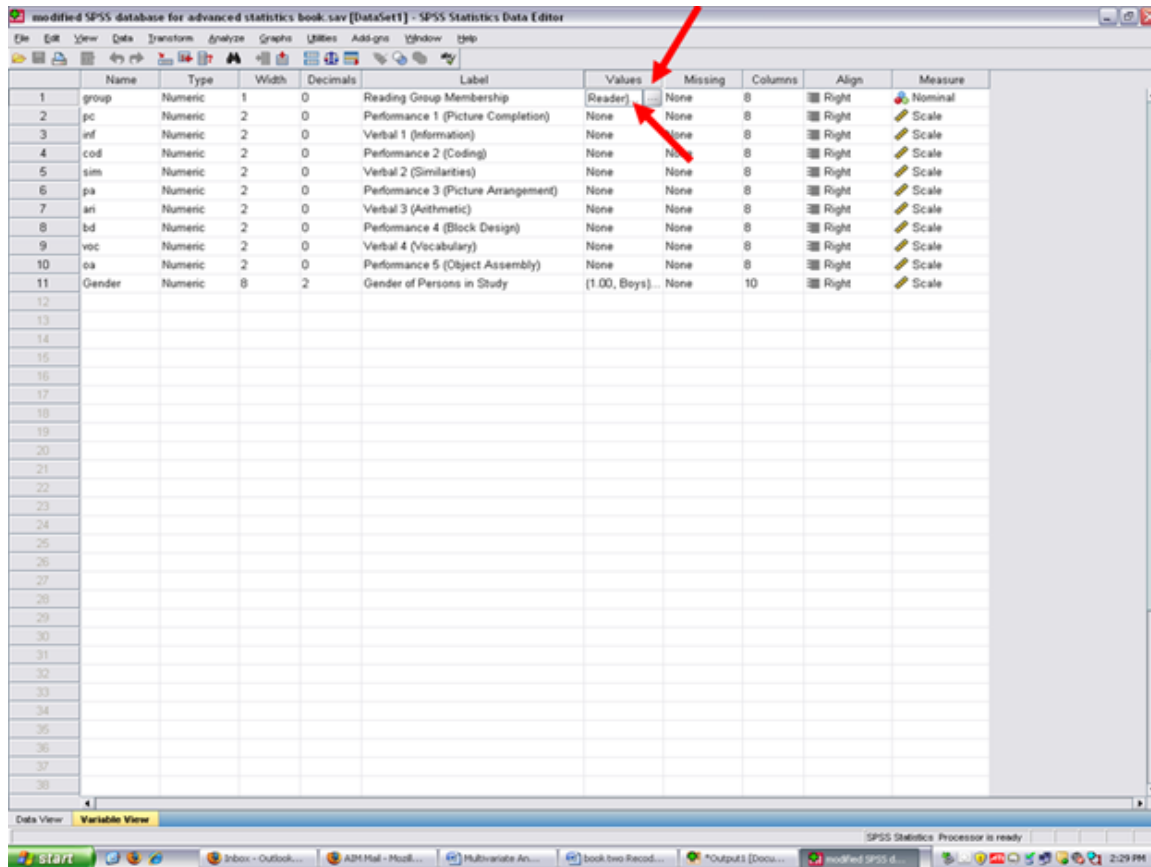
<sup>9</sup><http://cnx.org/content/m40728/latest/10.3.png/image>



10

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

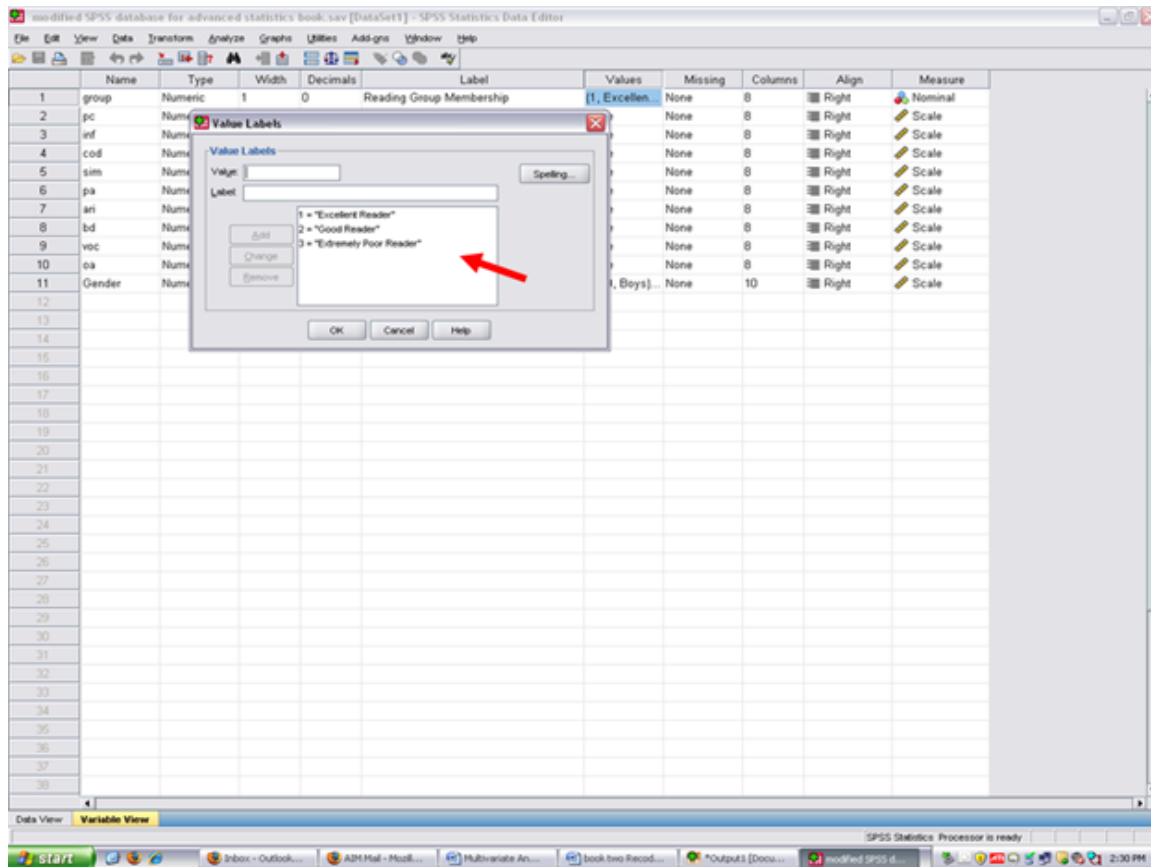
<sup>10</sup><http://cnx.org/content/m40728/latest/10.4.png/image>



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

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

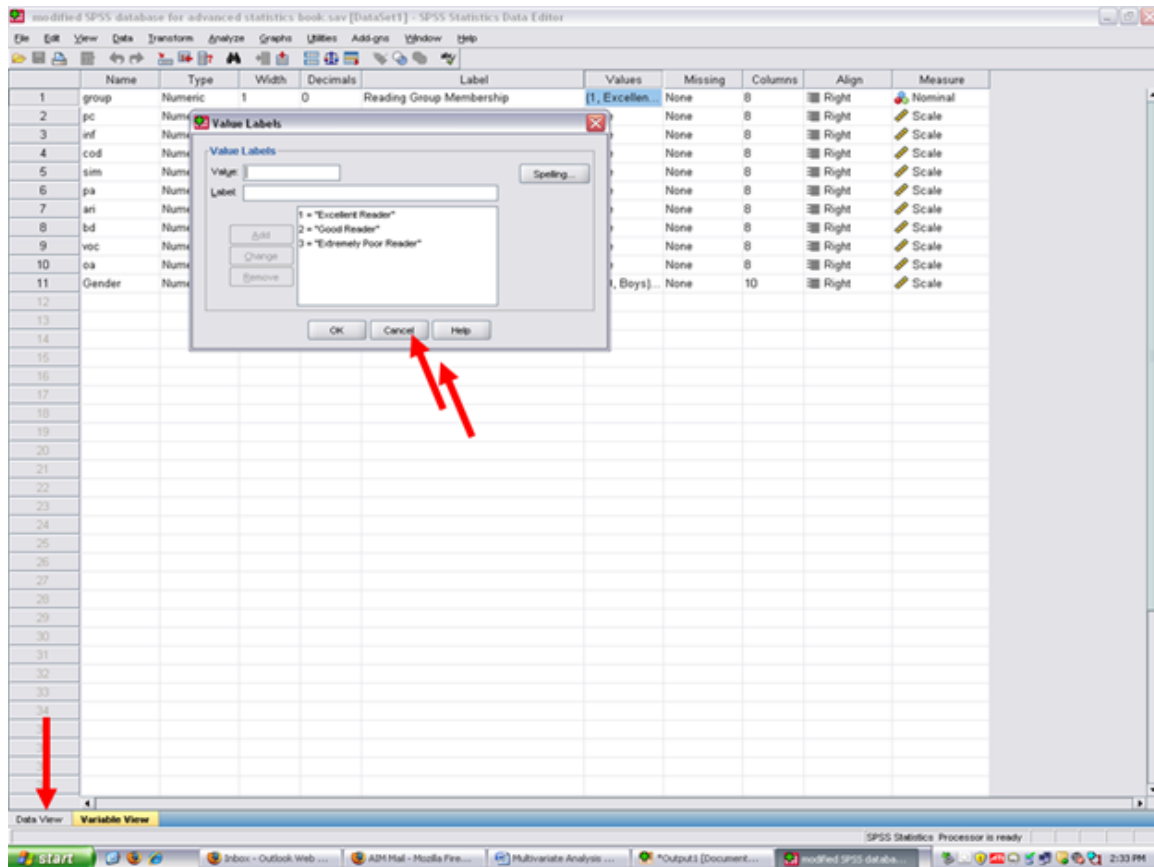




12

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.

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



Your screen should now look like the one below.

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

modified SPSS database for advanced statistics book.sav [DataSet1] - SPSS Statistics Data Editor

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

1: group 3.0 Visible: 11 of 11 Variables

	group	pc	inf	cod	sim	pa	an	bd	voc	oa	Gender	vis1	vis2	vis3
1	3	8	12	6	12	9	9	8	8	8	-			
2	3	8	12	6	12	9	9	8	8	8	-			
3	3	9	5	6	5	11	6	10	4	7	-			
4	3	3	4	11	4	7	6	3	4	10	1.00			
5	3	5	7	17	6	8	10	7	5	9	1.00			
6	3	2	5	8	8	9	7	10	3	5	1.00			
7	3	9	6	10	6	9	8	4	5	8	1.00			
8	3	5	9	7	10	9	12	11	7	8	1.00			
9	3	6	7	5	8	7	7	7	4	7	1.00			
10	3	7	8	11	6	8	4	3	3	4	1.00			
11	3	6	6	8	12	9	9	9	10	6	1.00			
12	3	7	8	3	1	9	8	5	10	3	1.00			
13	3	12	8	7	8	13	10	11	7	6	1.00			
14	3	10	9	10	10	11	11	14	11	11	1.00			
15	3	3	4	4	8	7	7	5	6	7	1.00			
16	3	3	4	4	8	7	7	5	6	7	1.00			
17	3	8	4	9	7	9	8	10	7	7	1.00			
18	3	5	9	6	6	5	7	9	4	4	1.00			
19	3	6	8	11	8	10	11	2	8	4	1.00			
20	3	3	6	8	6	6	8	3	2	6	1.00			
21	3	11	8	8	10	9	9	12	11	9	1.00			
22	3	17	11	6	11	13	9	14	11	13	1.00			
23	3	9	12	6	11	11	7	9	9	12	1.00			
24	1	7	10	11	4	8	7	11	7	10	1.00			
25	3	7	8	7	10	11	10	7	9	8	1.00			
26	3	5	5	8	6	6	8	4	8	4	1.00			
27	3	8	7	7	7	2	8	8	6	11	1.00			
28	3	9	2	6	8	10	7	4	6	6	1.00			
29	3	14	8	8	8	12	8	8	8	9	1.00			
30	3	8	8	7	10	11	11	3	9	8	1.00			
31	3	11	3	4	7	8	8	10	4	10	1.00			
32	3	10	7	9	8	10	8	10	5	11	1.00			
33	3	4	11	9	6	13	11	9	4	7	1.00			
34	3	12	12	7	8	4	8	9	9	8	1.00			
35	3	11	9	7	8	10	11	9	11	9	1.00			
36	3	8	5	9	1	7	9	11	8	8	1.00			
37	3	7	6	6	7	6	11	6	7	6	1.00			

Data View Variable View

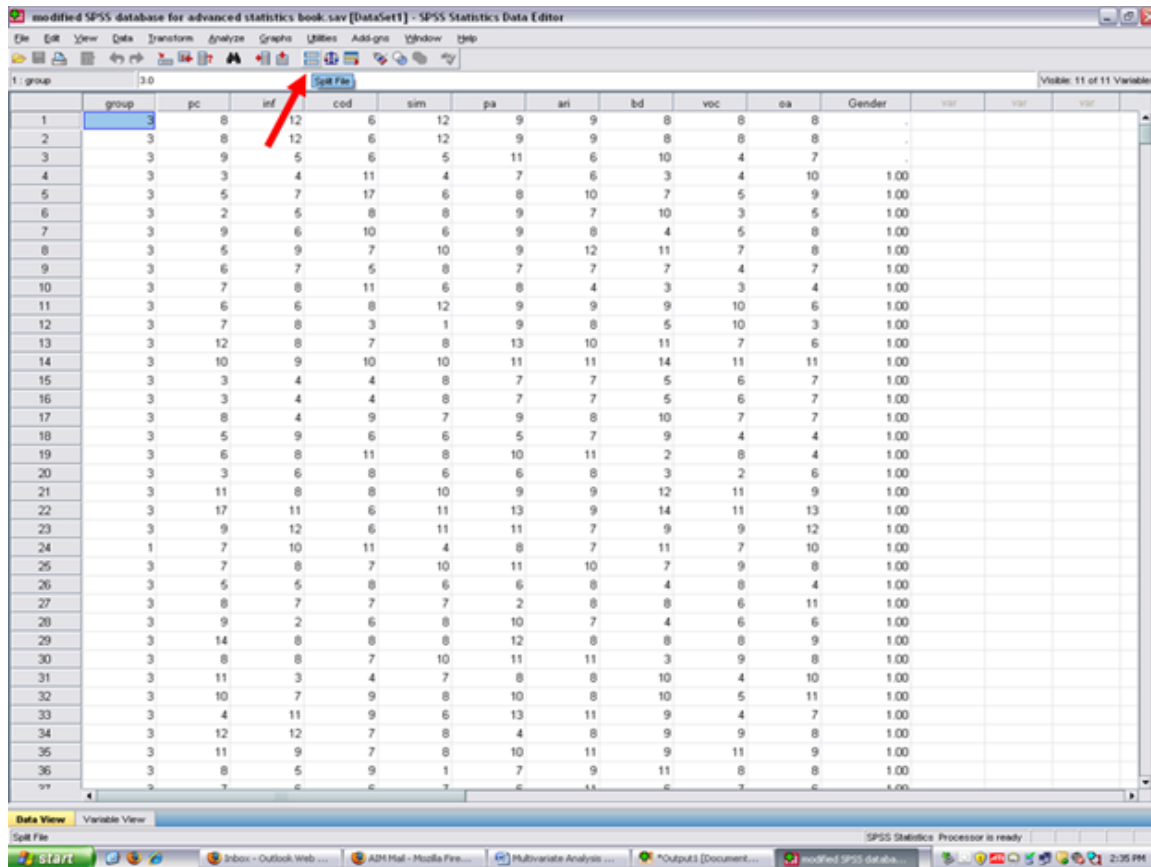
SPSS Statistics Processor is ready

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.

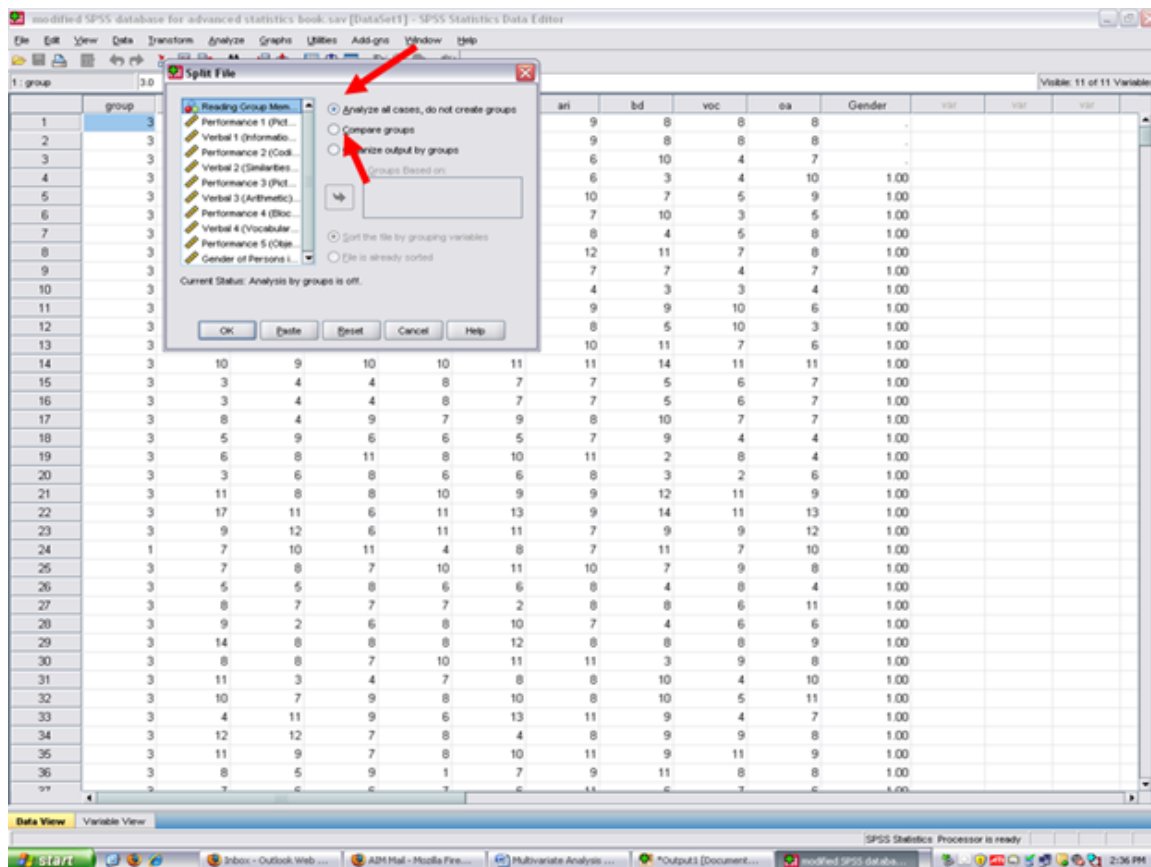
<sup>14</sup><http://cnx.org/content/m40728/latest/10.8.png/image>



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.

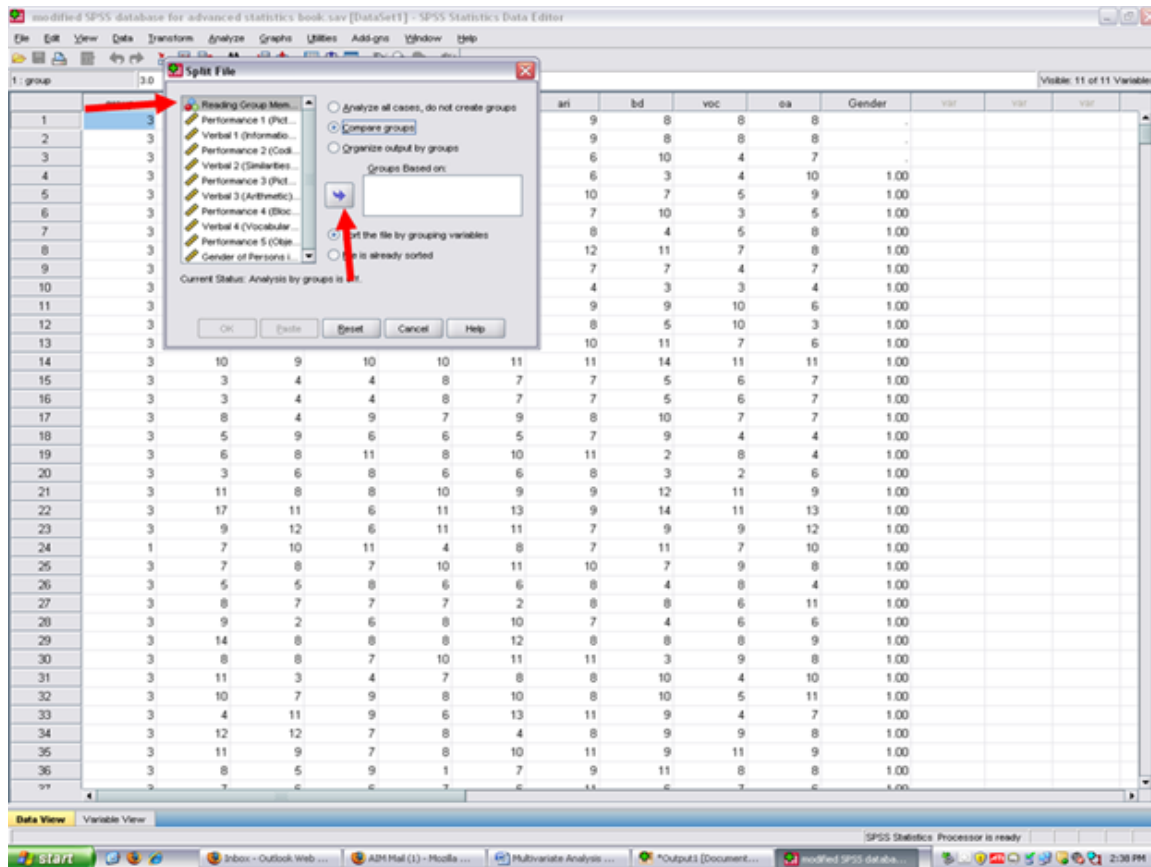
<sup>15</sup><http://cnx.org/content/m40728/latest/10.9.png/image>



16

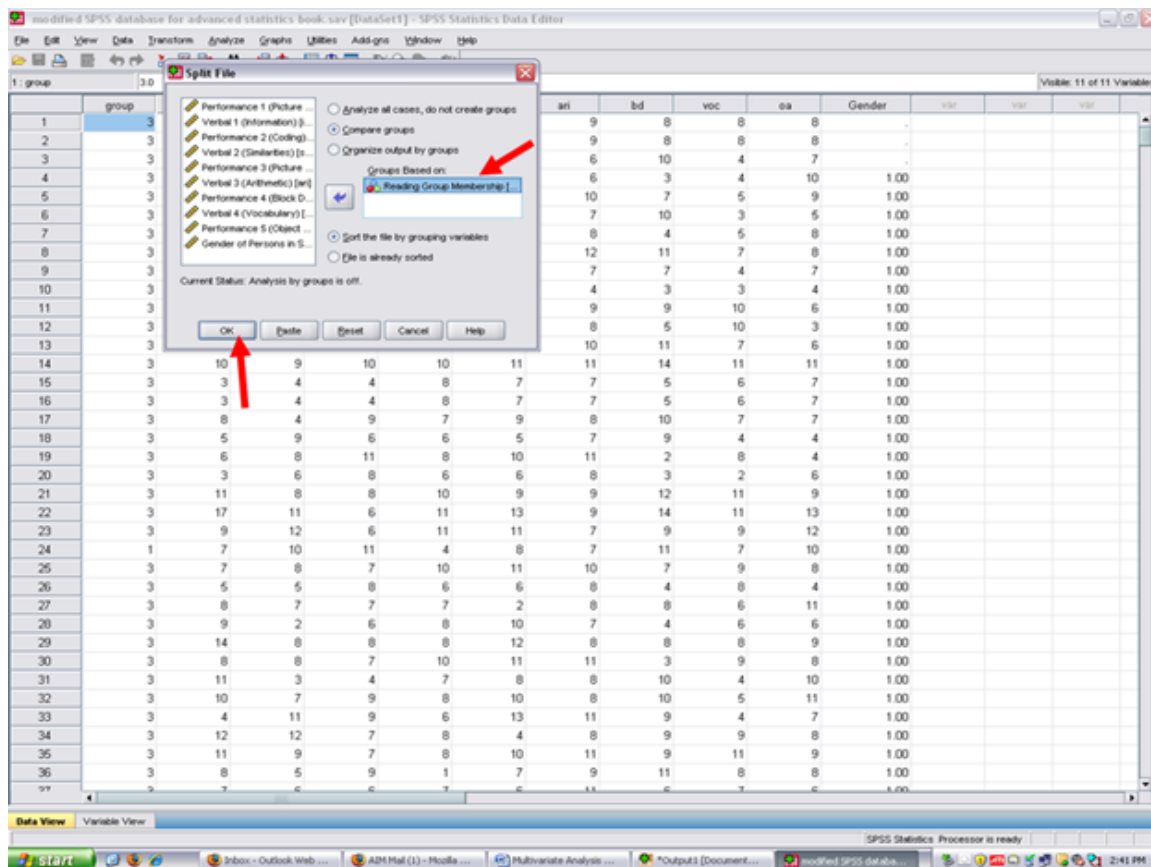
The screen below will appear after the Compare groups button has been clicked. Note that the Groups Based on rectangle has now become active. The independent (grouping) variable should be highlighted, as it already is, and then moved to the Groups Based on cell. After highlighting the Reading Group Membership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

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



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

<sup>17</sup><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.

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

SPSS Statistics Data Editor

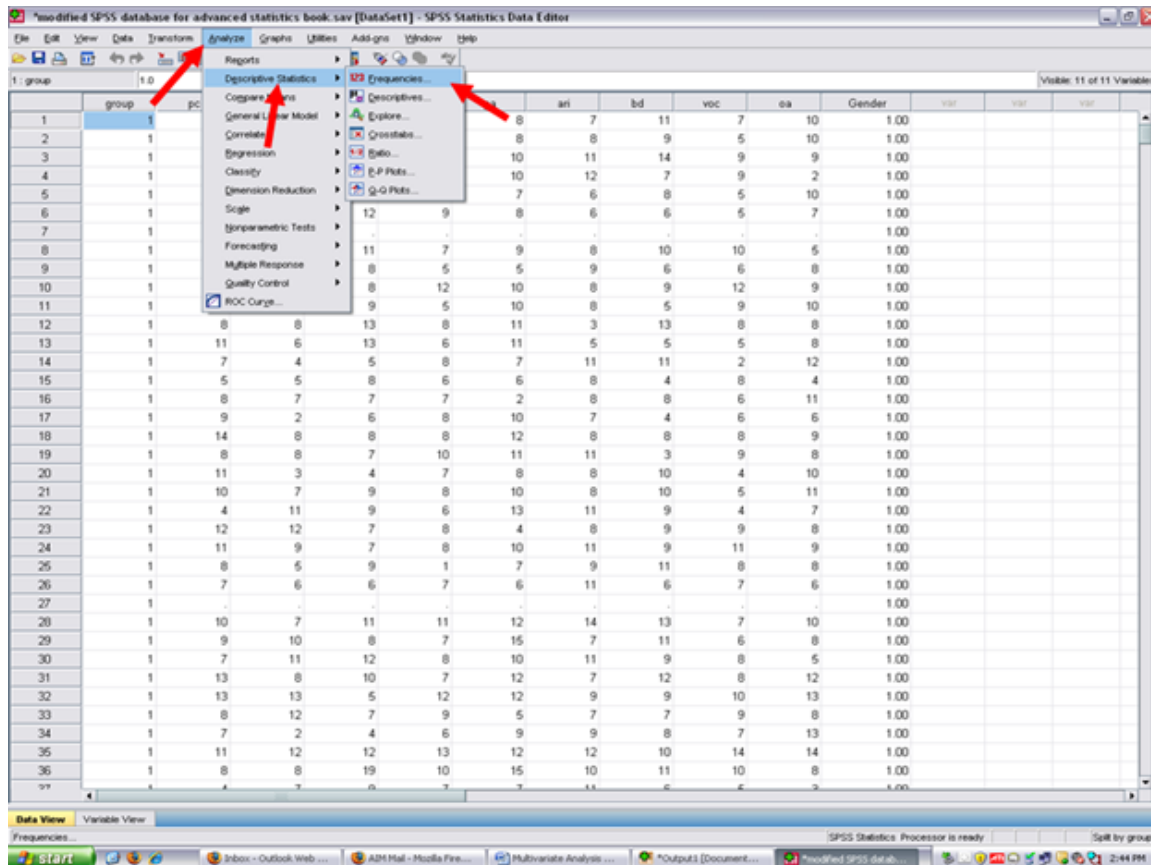
	group	pc	inf	cod	sim	pa	an	bd	voc	oa	Gender
1	1	7	10	11	4	8	7	11	7	10	1.00
2	1	9	6	10	7	8	8	9	5	10	1.00
3	1	8	6	8	8	10	11	14	9	9	1.00
4	1	2	7	14	8	10	12	7	9	2	1.00
5	1	6	4	6	4	7	6	8	5	10	1.00
6	1	6	9	12	9	8	6	6	5	7	1.00
7	1	.	.	.	.	.	.	.	.	.	1.00
8	1	10	11	11	7	9	8	10	10	5	1.00
9	1	9	4	8	5	5	9	6	6	8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

19

- ✓ Analyze
- \* Descriptive Statistics
- \* Frequencies

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

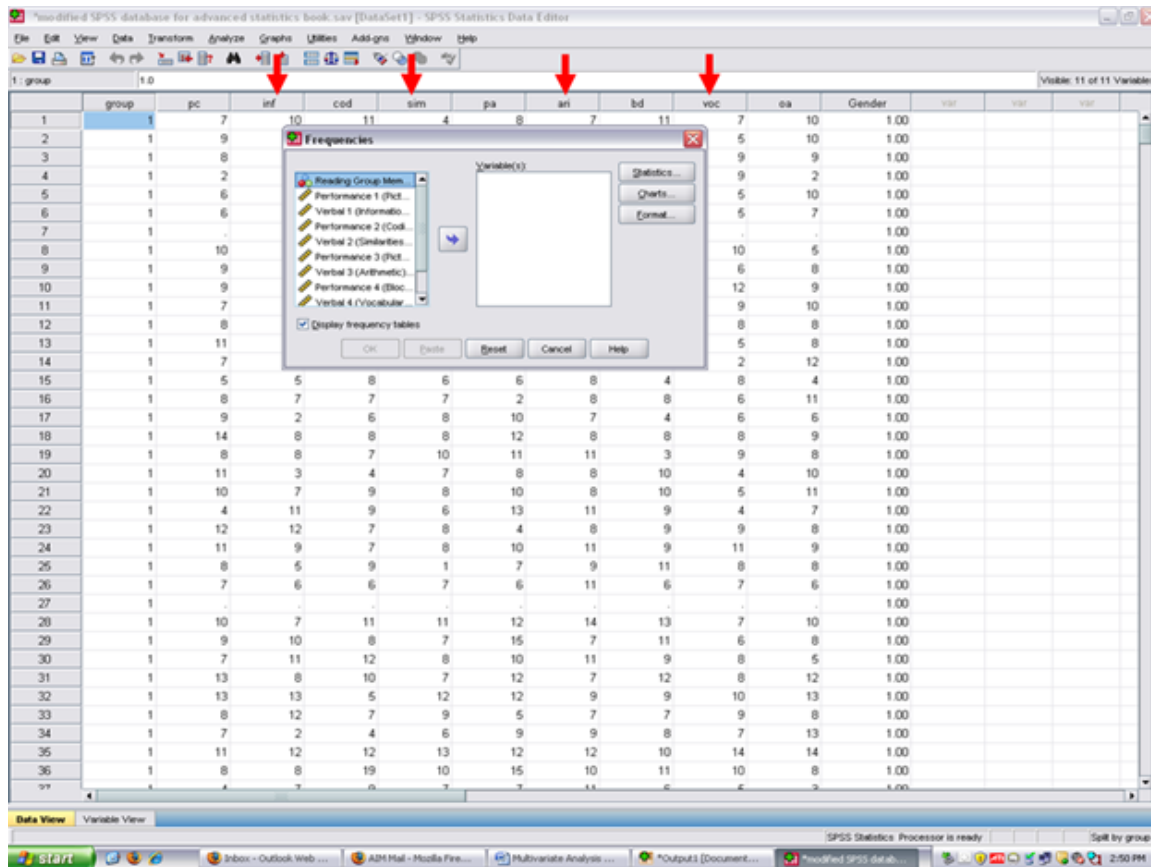




20

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)

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

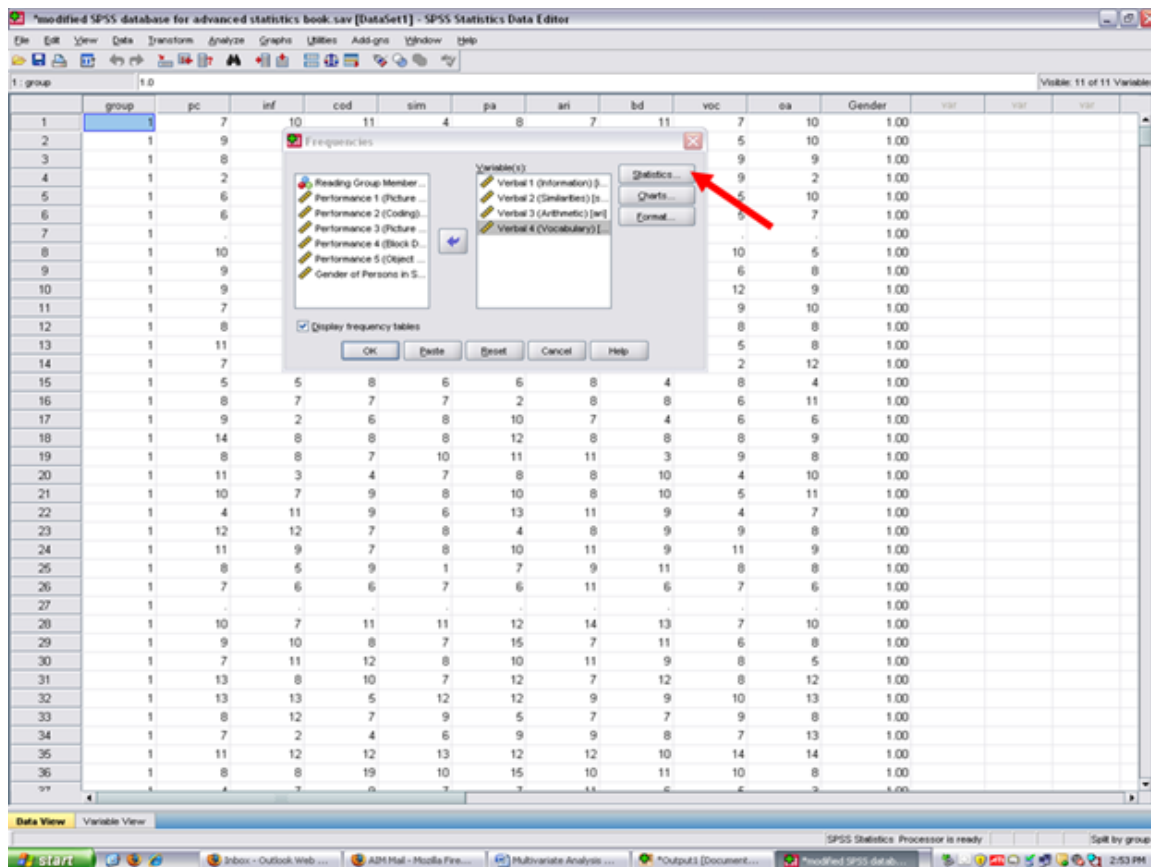


21

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

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

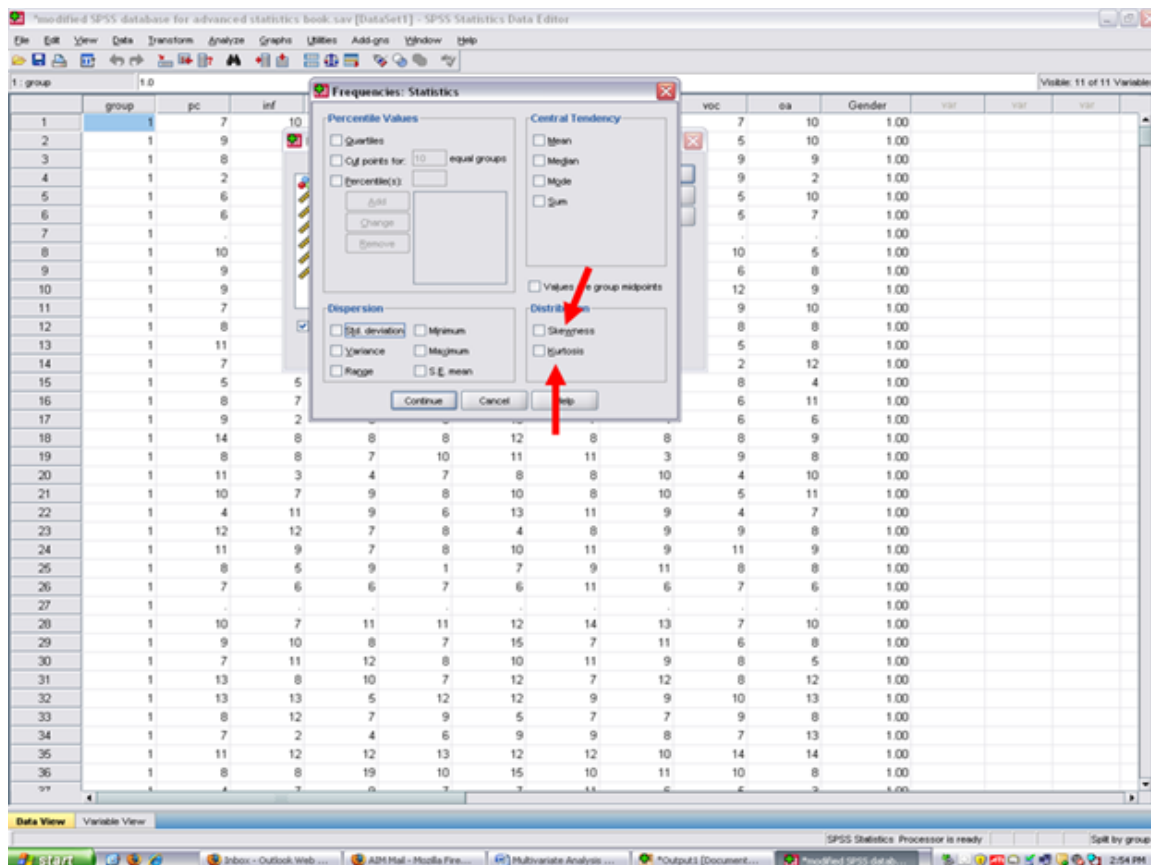




23

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.

<sup>23</sup><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](http://www.statistics.com/index.php?page=glossary&term_id=356)<sup>25</sup> and <http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb><sup>26</sup>

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](http://www.statistics.com/index.php?page=glossary&term_id=326)<sup>27</sup> and

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

<sup>25</sup>[http://www.statistics.com/index.php?page=glossary&term\\_id=356](http://www.statistics.com/index.php?page=glossary&term_id=356)

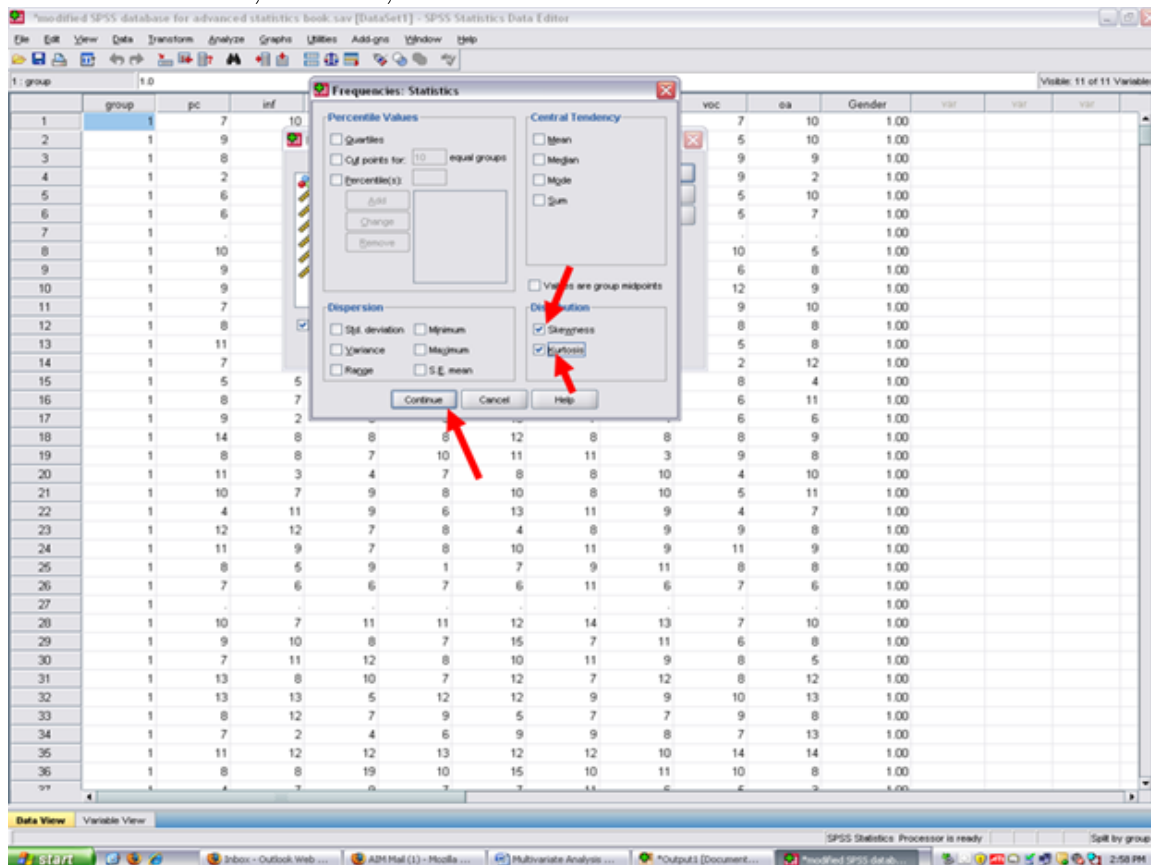
<sup>26</sup><http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb>

<sup>27</sup>[http://www.statistics.com/index.php?page=glossary&term\\_id=326](http://www.statistics.com/index.php?page=glossary&term_id=326)

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

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  $\pm 3$  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  $\pm 3$  range, the dataset is not normally distributed.

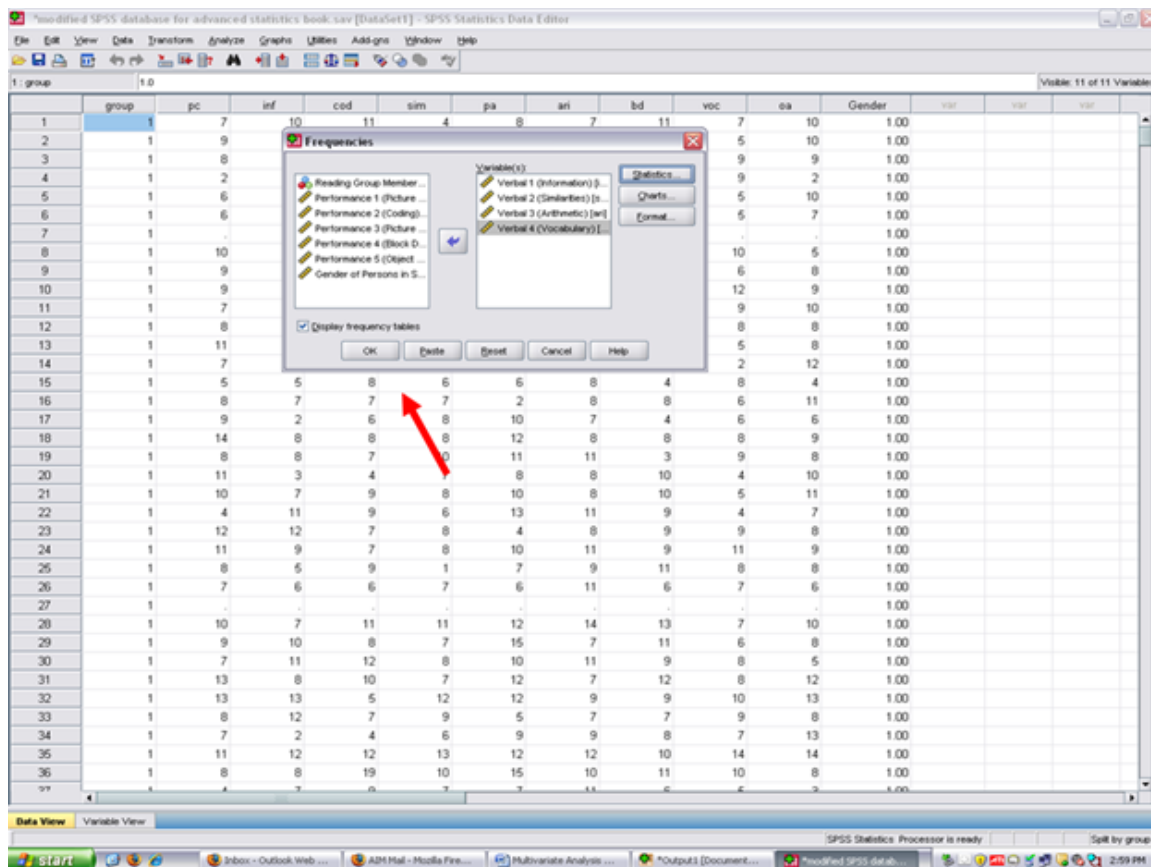
Click on Skewness, Kurtosis, and Continue



Now click on OK.

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

<sup>29</sup><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.

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

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

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

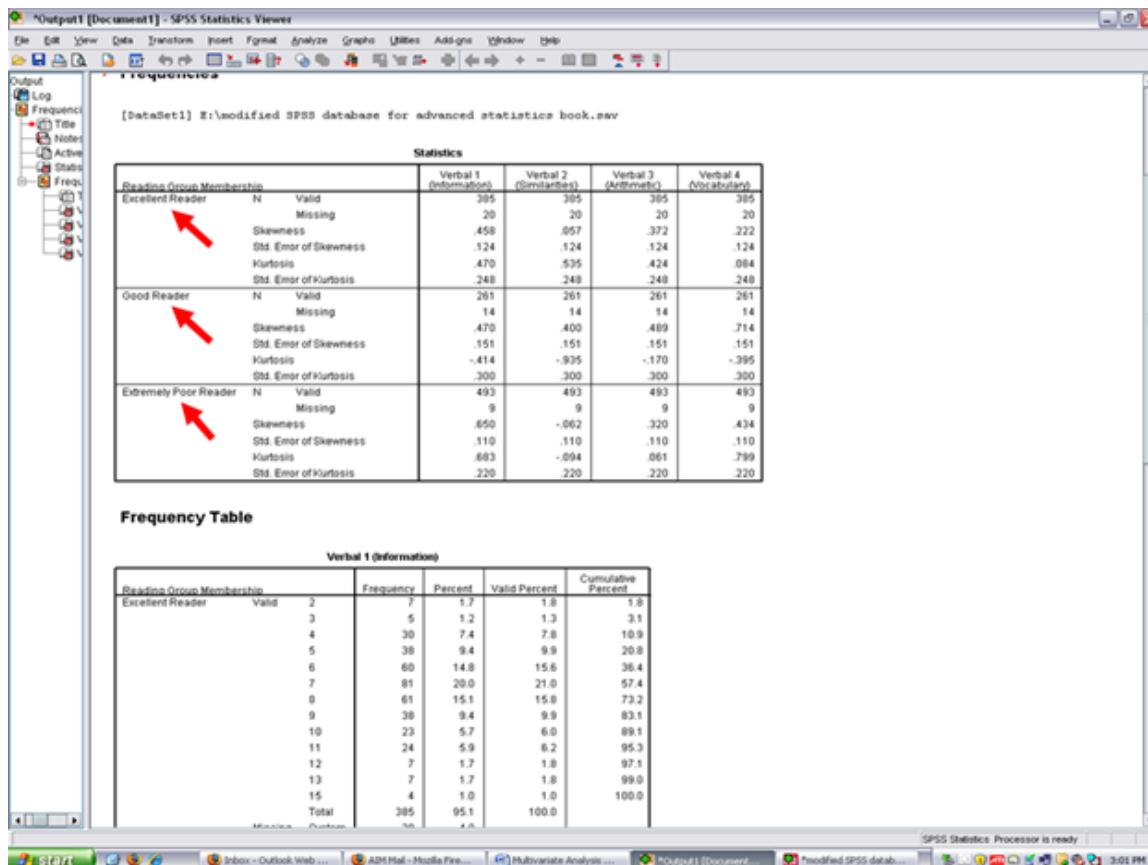
The taskbar at the bottom shows the 'Output' icon highlighted with a red arrow.

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.

<sup>31</sup><http://cnx.org/content/m40728/latest/10.21.png/image>

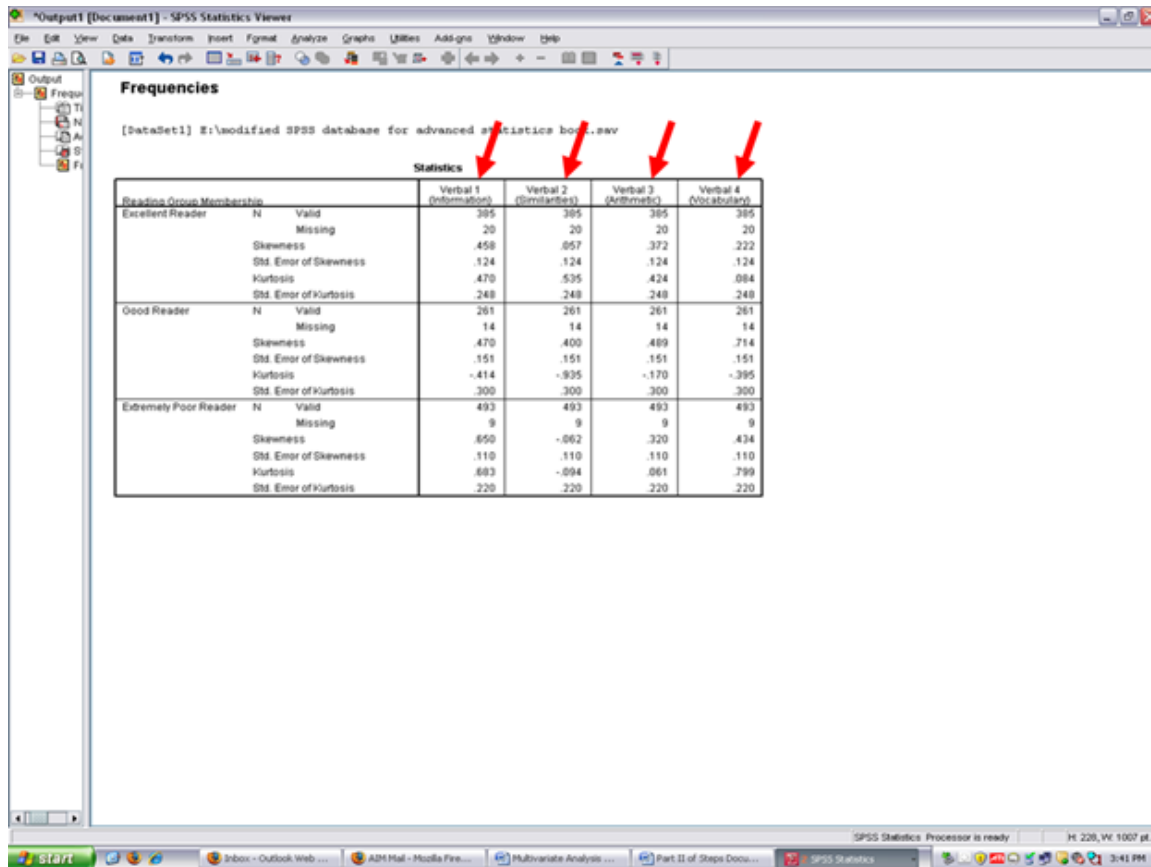




32

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

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



**Frequencies**

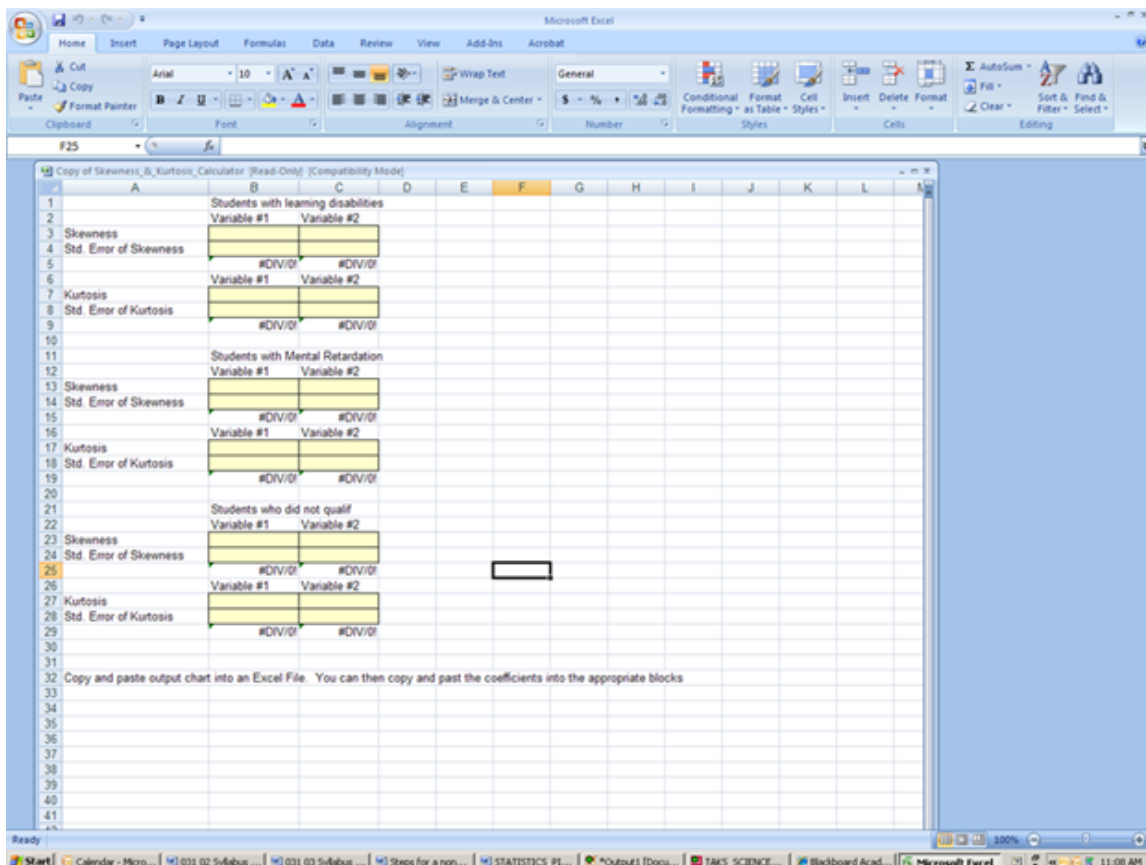
[DataSet1] E:\modified SPSS database for advanced statistics book.sav

		Statistics	Verbal 1 (Information)	Verbal 2 (Similarities)	Verbal 3 (Arithmetic)	Verbal 4 (Vocabulary)
Excellent Reader	N	Valid	395	395	395	395
		Missing	20	20	20	20
		Skewness	.458	.657	.372	.222
		Std. Error of Skewness	.124	.124	.124	.124
		Kurtosis	.470	.535	.424	.084
		Std. Error of Kurtosis	.248	.248	.248	.248
Good Reader	N	Valid	261	261	261	261
		Missing	14	14	14	14
		Skewness	.470	.400	.489	.714
		Std. Error of Skewness	.151	.151	.151	.151
		Kurtosis	-.414	-.935	-.170	-.395
		Std. Error of Kurtosis	.300	.300	.300	.300
Extremely Poor Reader	N	Valid	493	493	493	493
		Missing	9	9	9	9
		Skewness	.650	-.062	.320	.434
		Std. Error of Skewness	.110	.110	.110	.110
		Kurtosis	.683	-.094	.061	.799
		Std. Error of Kurtosis	.220	.220	.220	.220

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).

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



34

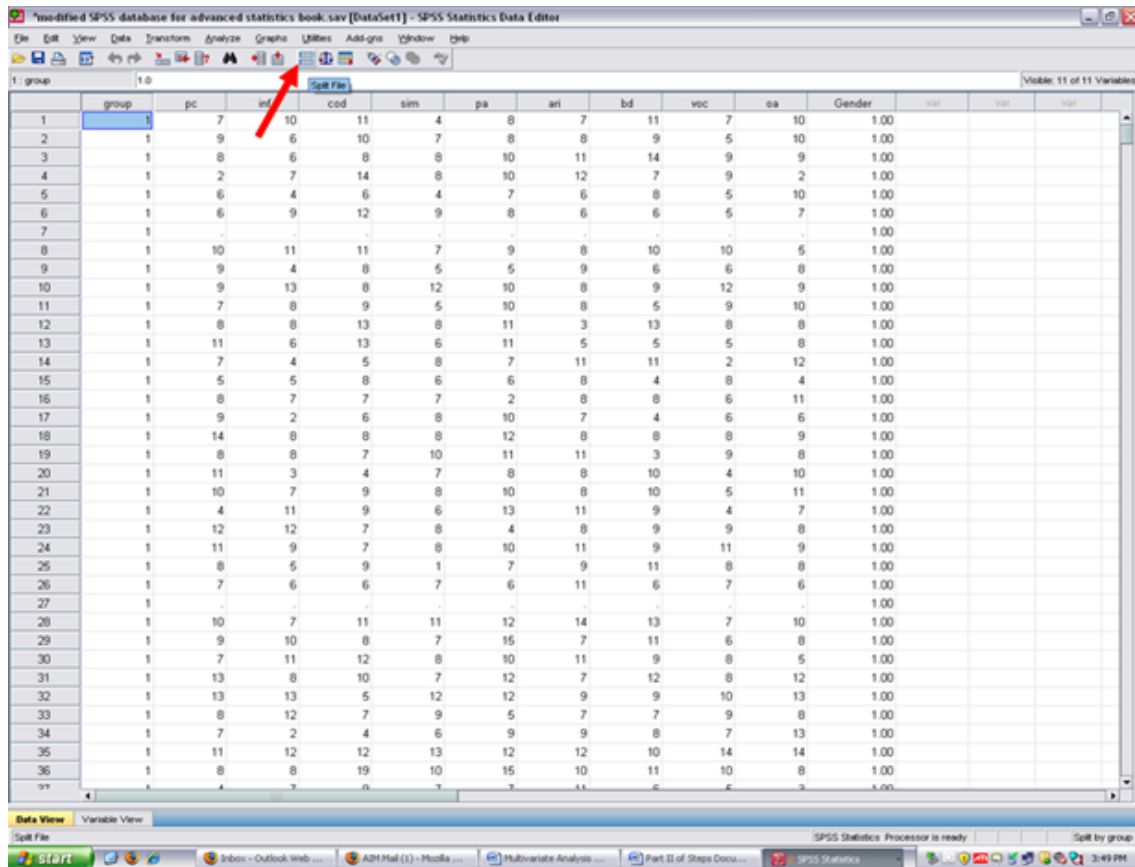
After calculating the standardized coefficients, record the number that are within  $\pm 3$  and the number that are outside of these boundaries. This information needs to be discussed in your results section where you document your checks of the assumptions underlying this particular statistical technique.

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)

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



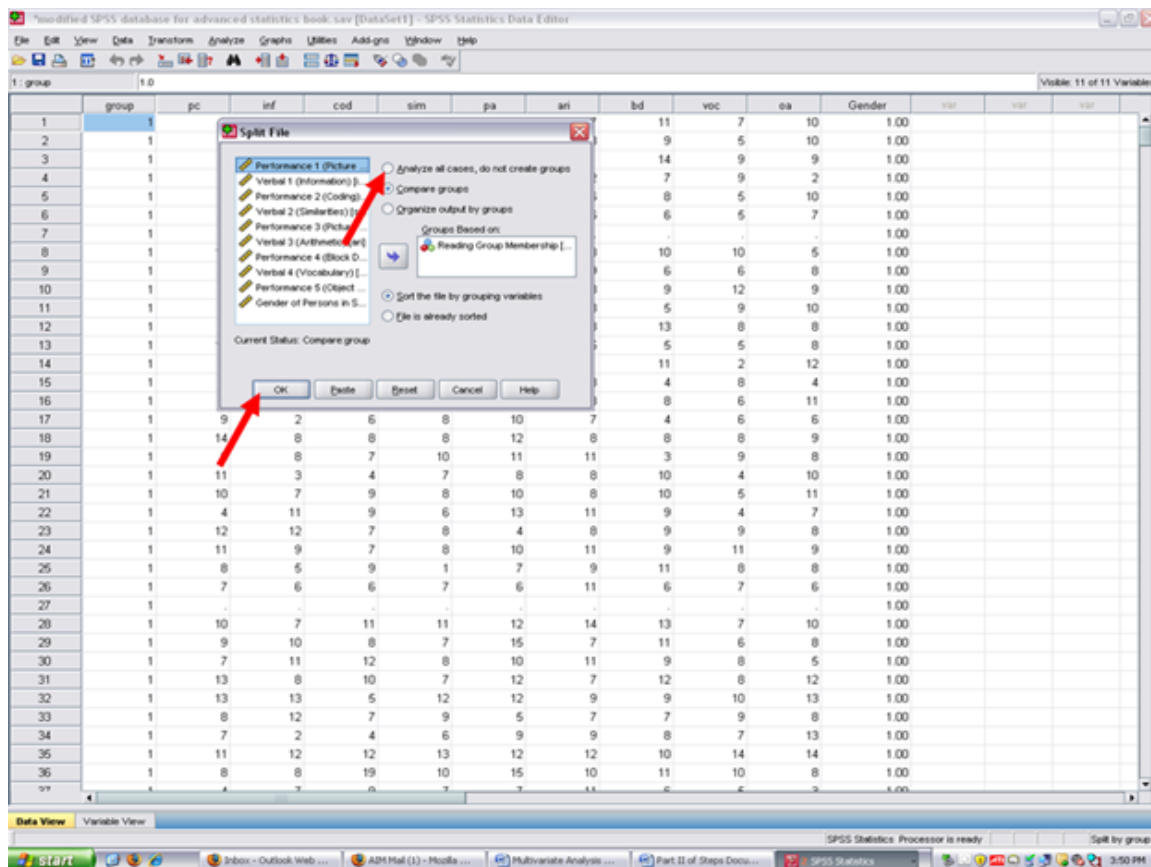
35

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

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



36

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

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

modified SPSS database for advanced statistics book.sav [DataSet1] - SPSS Statistics Data Editor

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

1: group 1.0 Visible: 11 of 11 Variables

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

Data View Variable View

SPSS Statistics Processor is ready Split by group

Start | Inbox - Outlook Web ... | AOL Mail - Mozilla Fire... | Multivariate Analysis ... | Part II of Steps Docu... | SPSS Statistics | 4:24 PM

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

## Chapter 2

# Multivariate Analysis of Variance: Part II<sup>1</sup>



NOTE: This chapter is published by NCPEA Press<sup>2</sup> 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<sup>3</sup>) 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.**

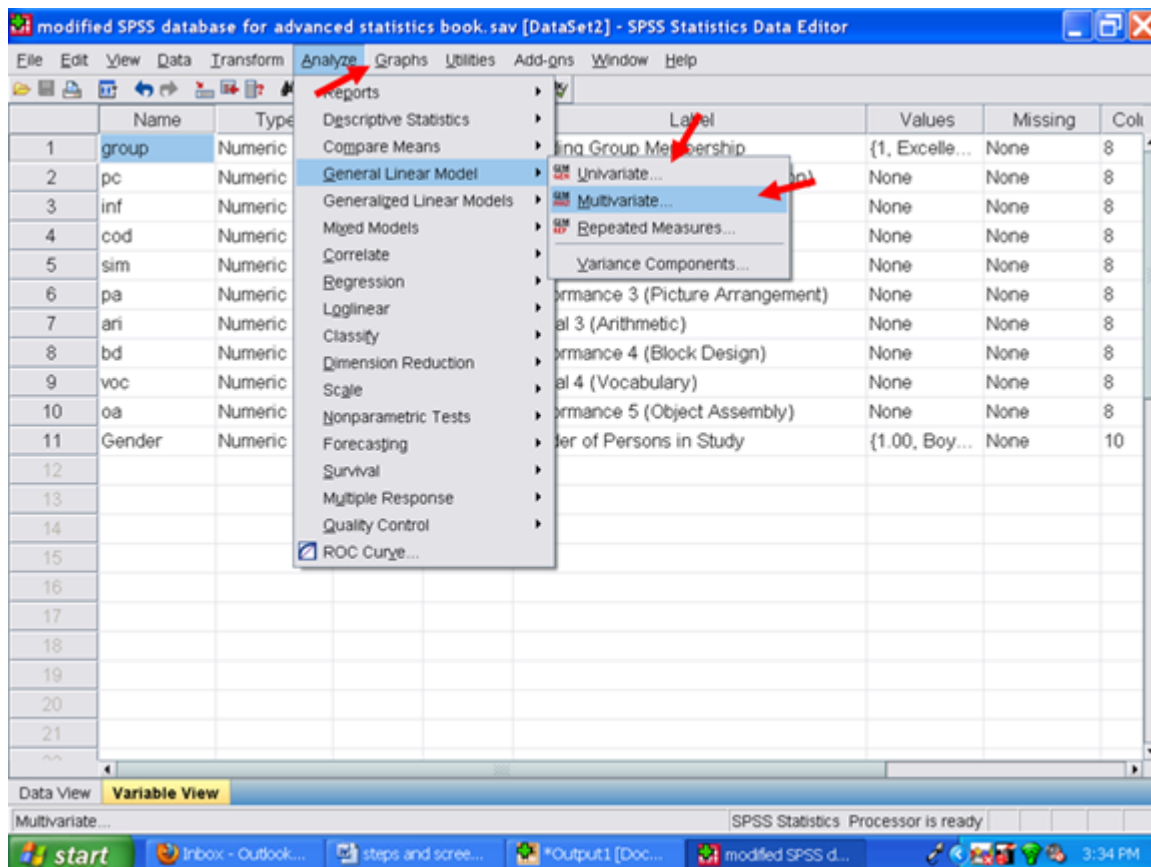
- ✓ **Analyze**
- ✓ **General Linear Model**
- ✓ **Multivariate**

---

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

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

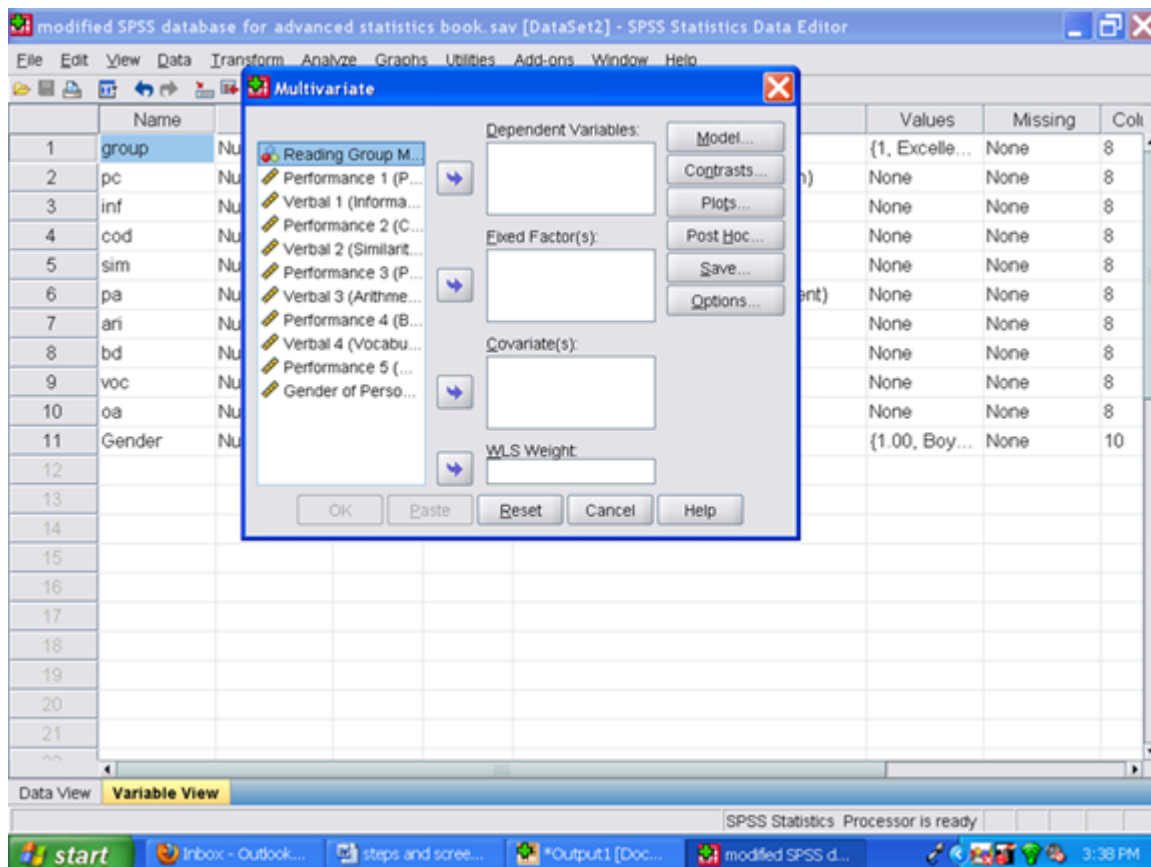
<sup>3</sup><http://cnx.org/content/m40729/latest/www.writingandstatisticalhelp>



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

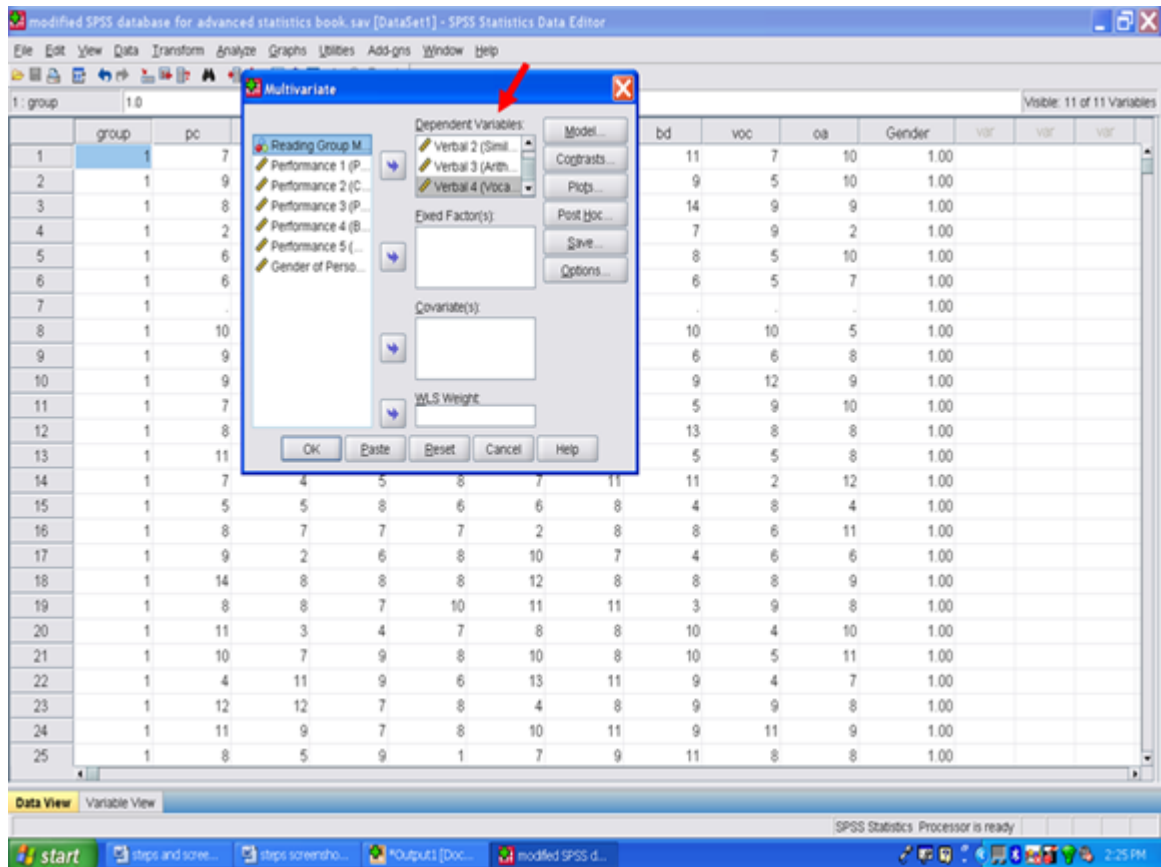
<sup>4</sup><http://cnx.org/content/m40729/latest/11.1.png/image>





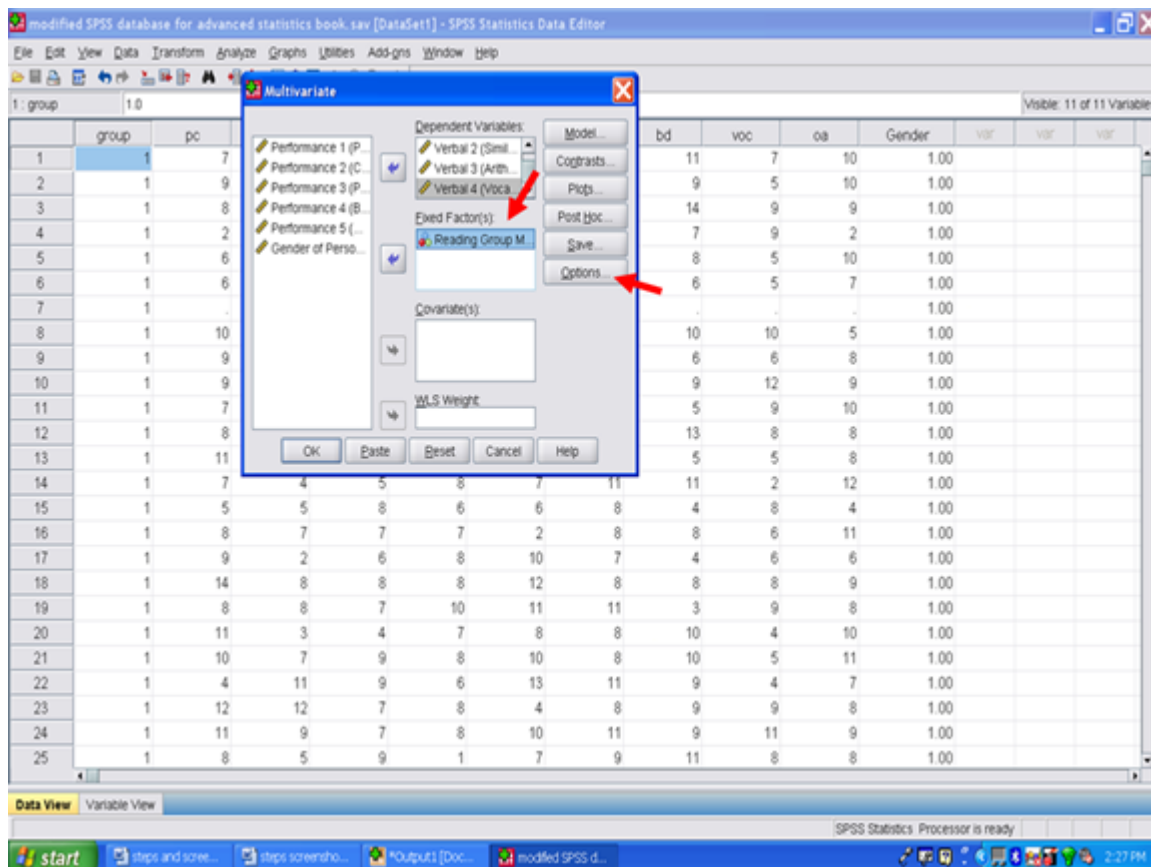
Now send the four dependent variables (i.e., Verbal 1 through Verbal 4) over one at a time to the Dependent Variables screen.

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



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

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

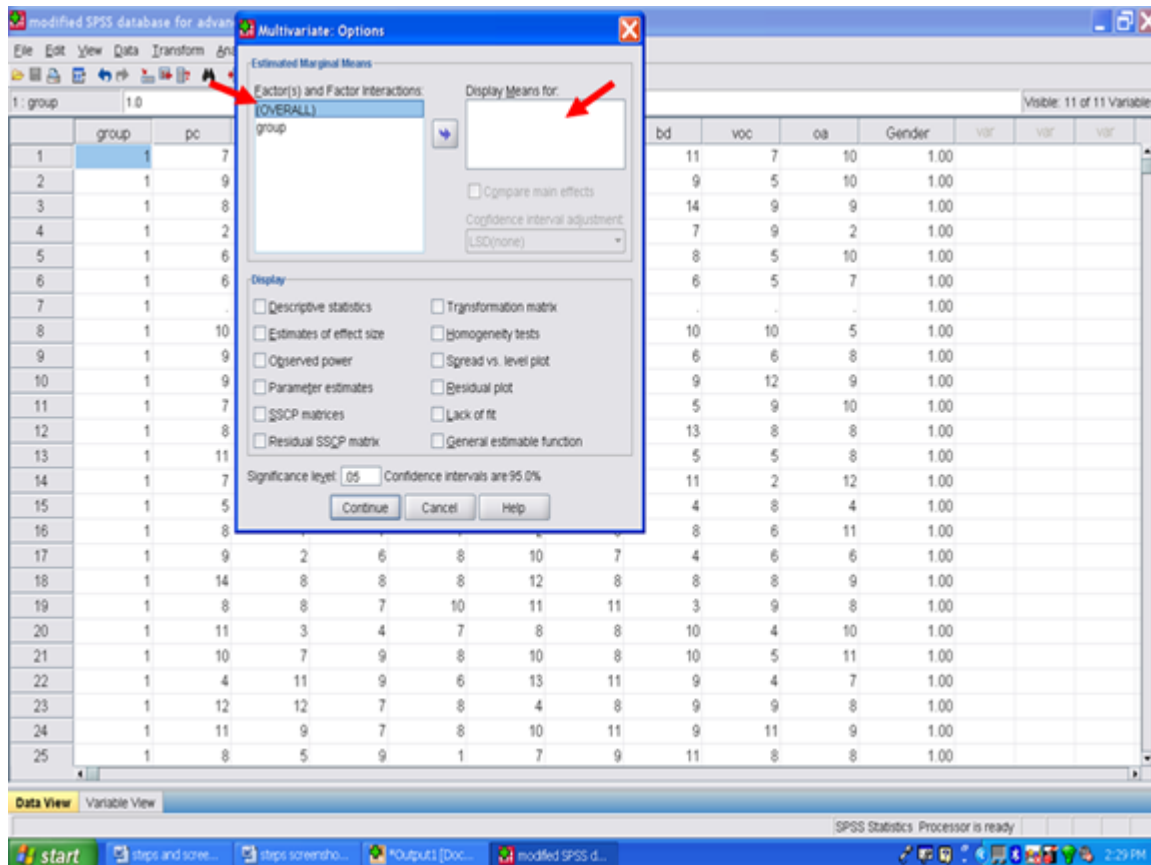


Clicking on options will then give you this screen.

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

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

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



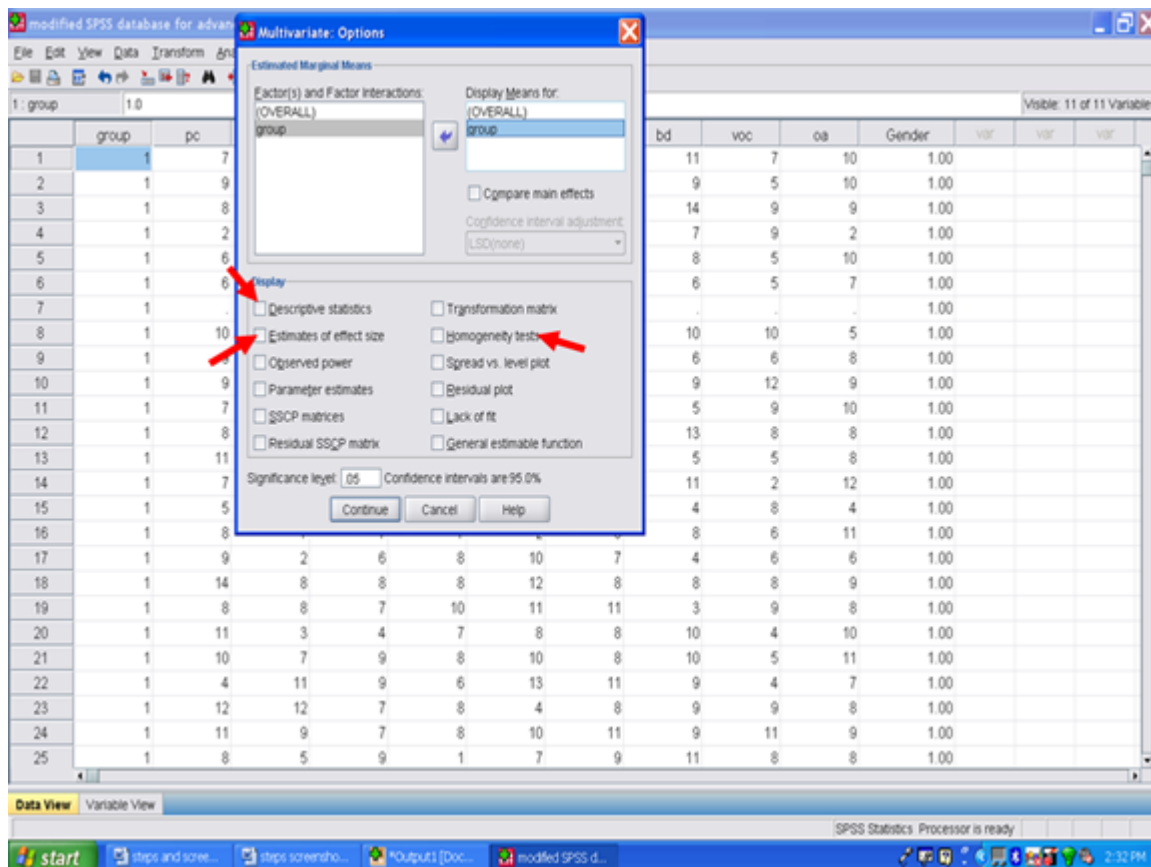
To obtain the information just mentioned, you will need to click on:

Descriptive Statistics

Estimates of Effect Size

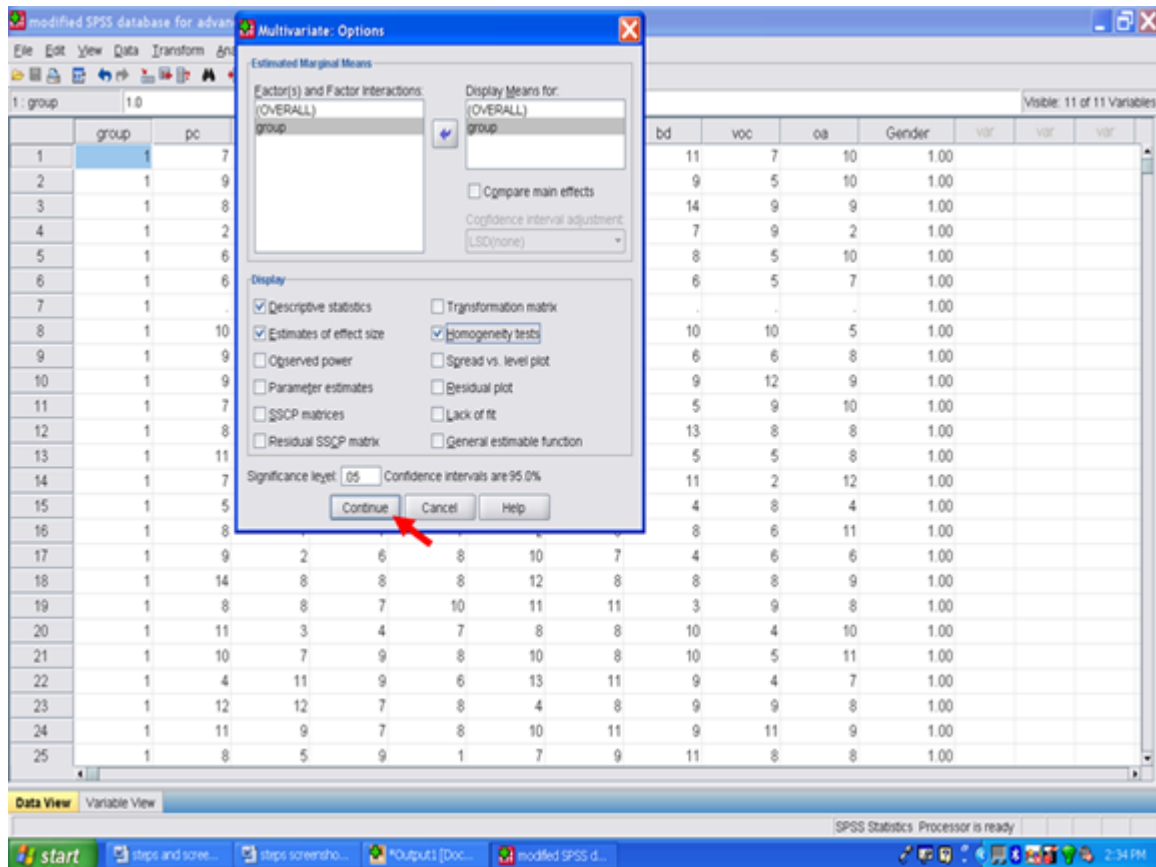
Homogeneity tests

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



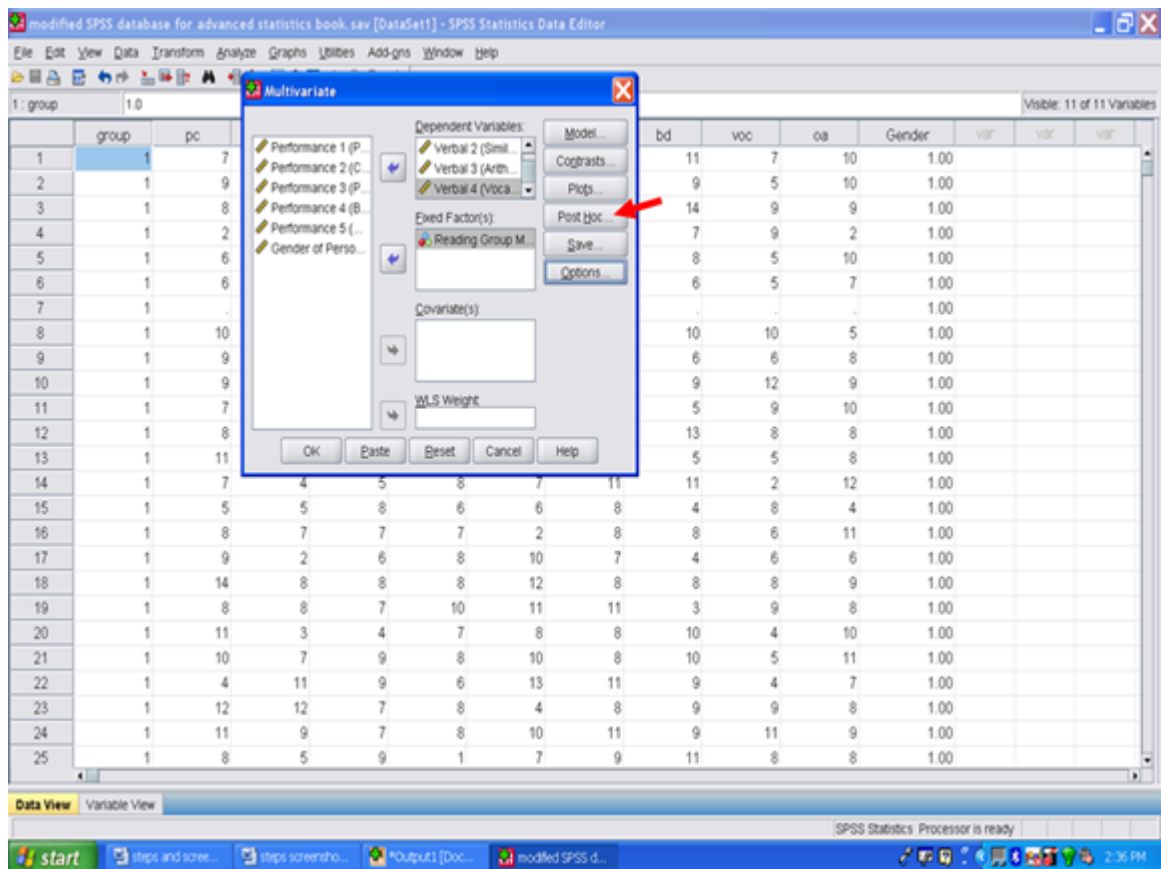
After checking the three boxes mentioned, then click on Continue

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



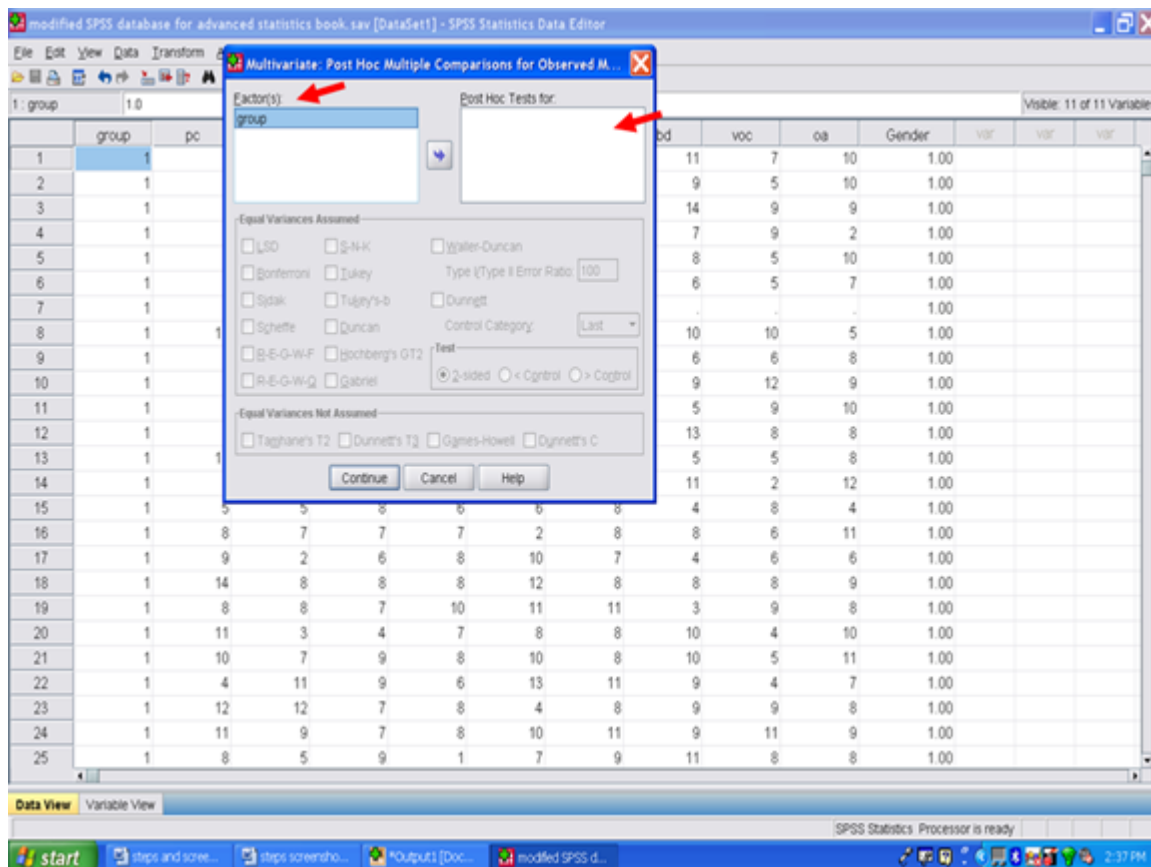
After clicking on Continue, the following screen will appear. Now click on Post Hoc so that pairwise analyses can be conducted.

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



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:

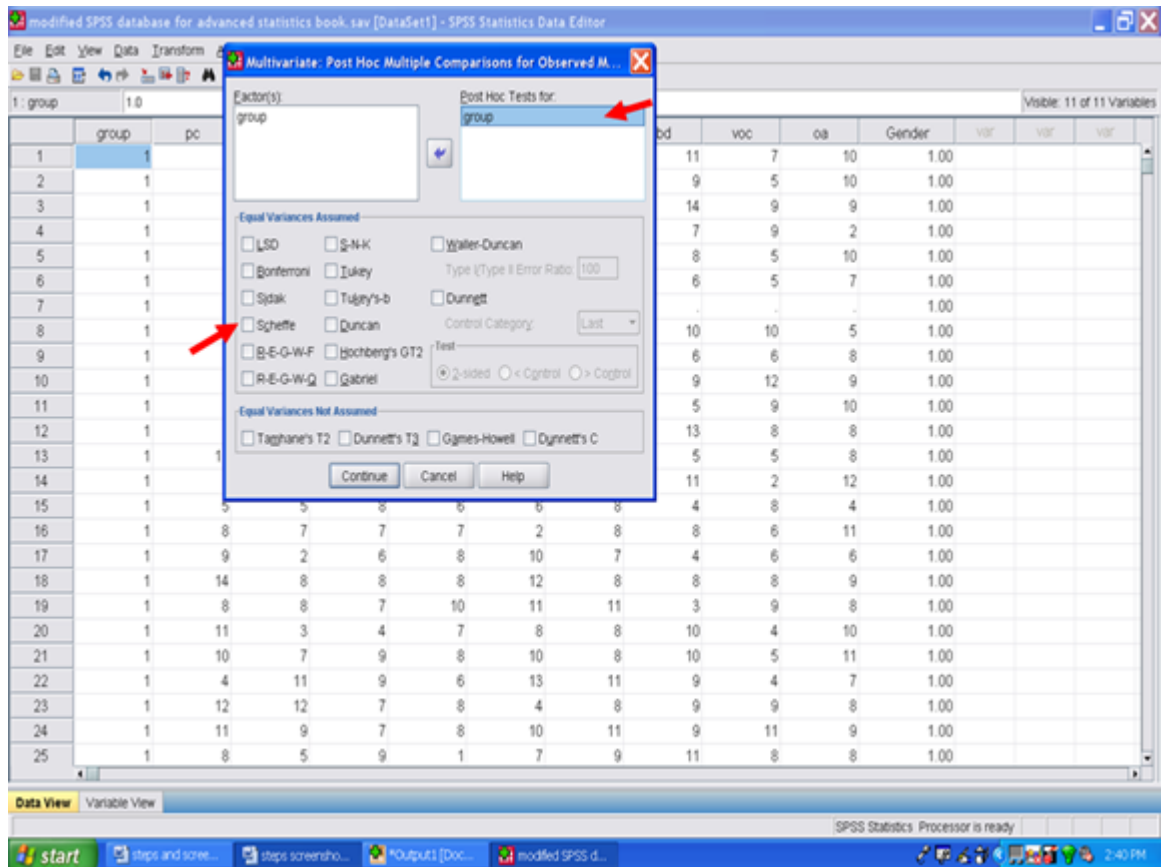
<sup>11</sup><http://cnx.org/content/m40729/latest/11.8.png/image>



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

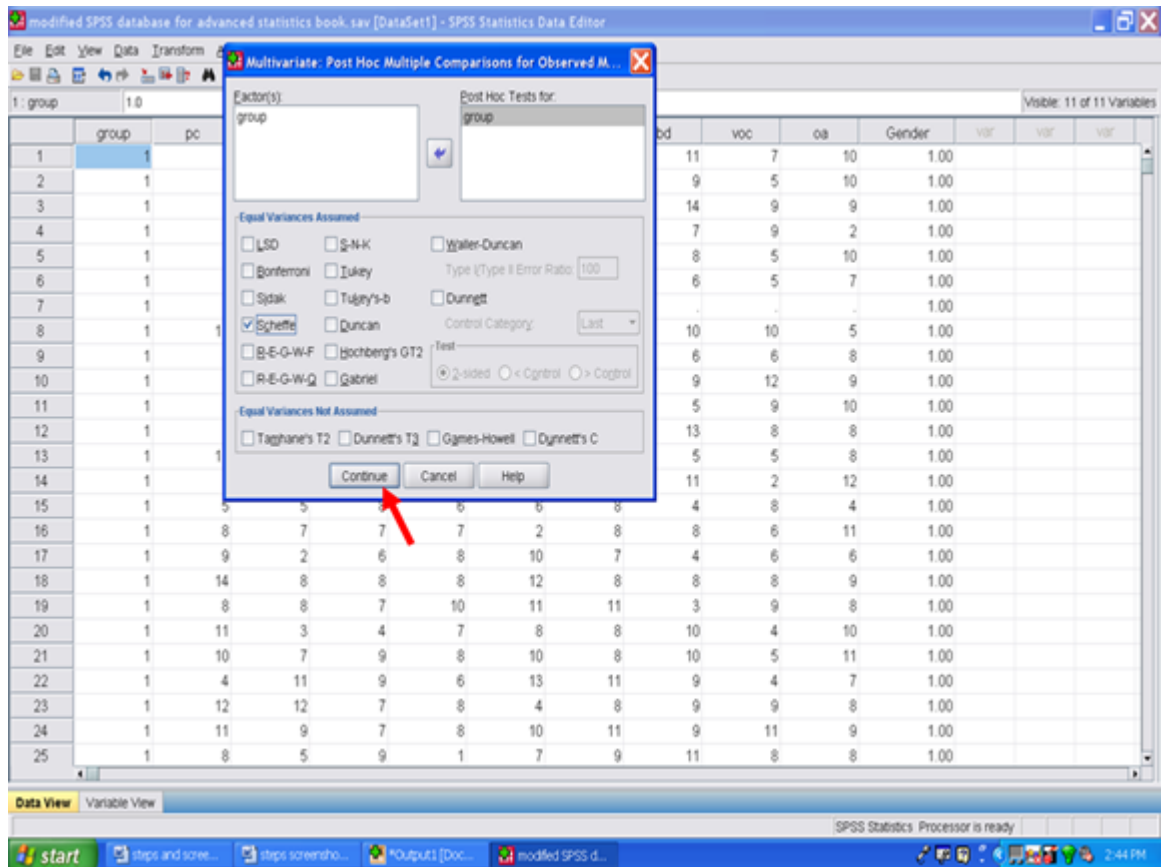
<sup>12</sup><http://cnx.org/content/m40729/latest/11.9.png/image>





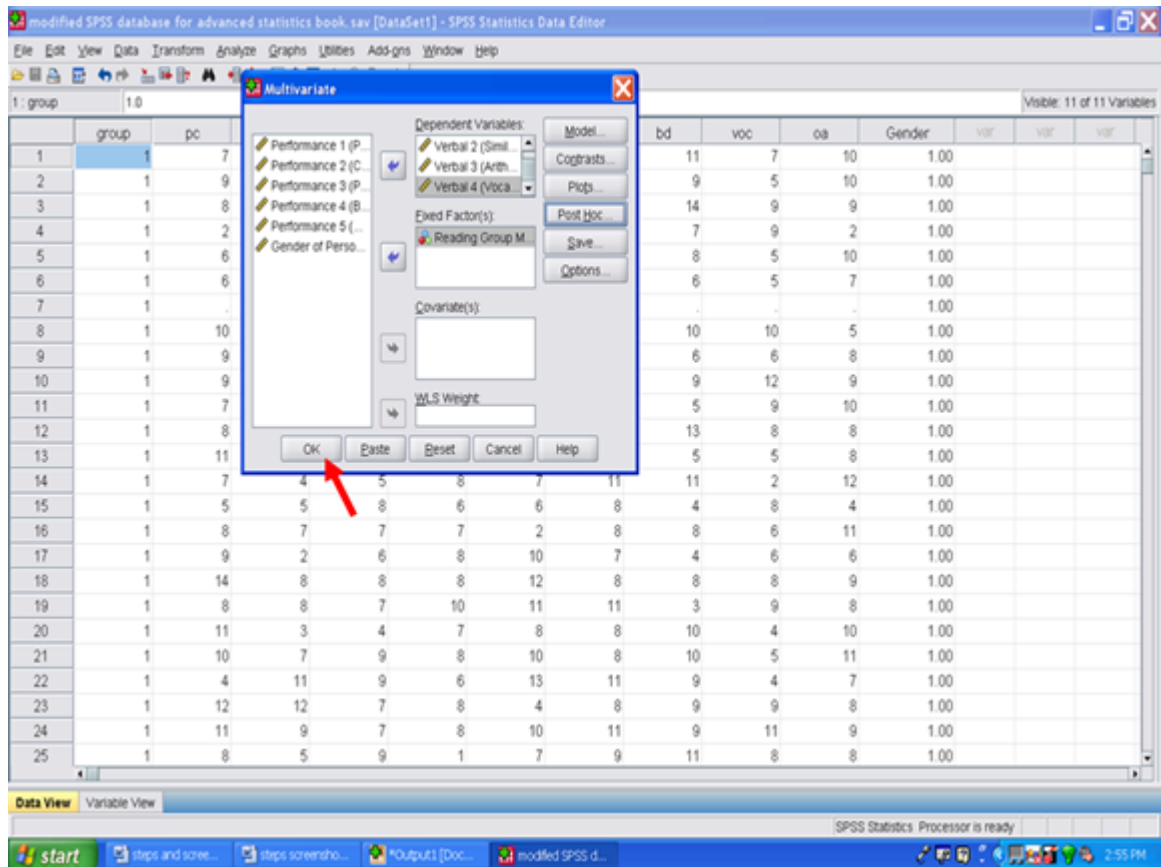
After clicking on the Scheffe' post hoc procedure, then click on Continue.

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



Then click on OK.

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



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.

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



## Chapter 3

# Multiple Analysis of Variance: Part III<sup>1</sup>



NOTE: This chapter is published by NCPEA Press<sup>2</sup> 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<sup>3</sup>) 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.

---

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

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

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

**General Linear Model**

**Between-Subjects Factors**

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

**Descriptive Statistics**

Reading Group Membership		Mean	Std. Deviation	N
Verbal 1 (Information)	Excellent Reader	7.34	2.383	395
	Good Reader	3.46	1.878	261
	Extremely Poor Reader	7.92	2.491	493
	Total	6.36	2.810	1139
Verbal 2 (Similarities)	Excellent Reader	7.44	2.486	395
	Good Reader	3.18	1.894	261
	Extremely Poor Reader	6.83	2.729	493
	Total	6.20	2.984	1139
Verbal 3 (Arithmetic)	Excellent Reader	7.59	2.214	395
	Good Reader	3.90	2.236	261
	Extremely Poor Reader	7.46	2.427	493
	Total	6.55	2.730	1139
Verbal 4 (Vocabulary)	Excellent Reader	6.47	2.555	395
	Good Reader	2.67	1.593	261
	Extremely Poor Reader	6.54	2.611	493
	Total	5.63	2.889	1139

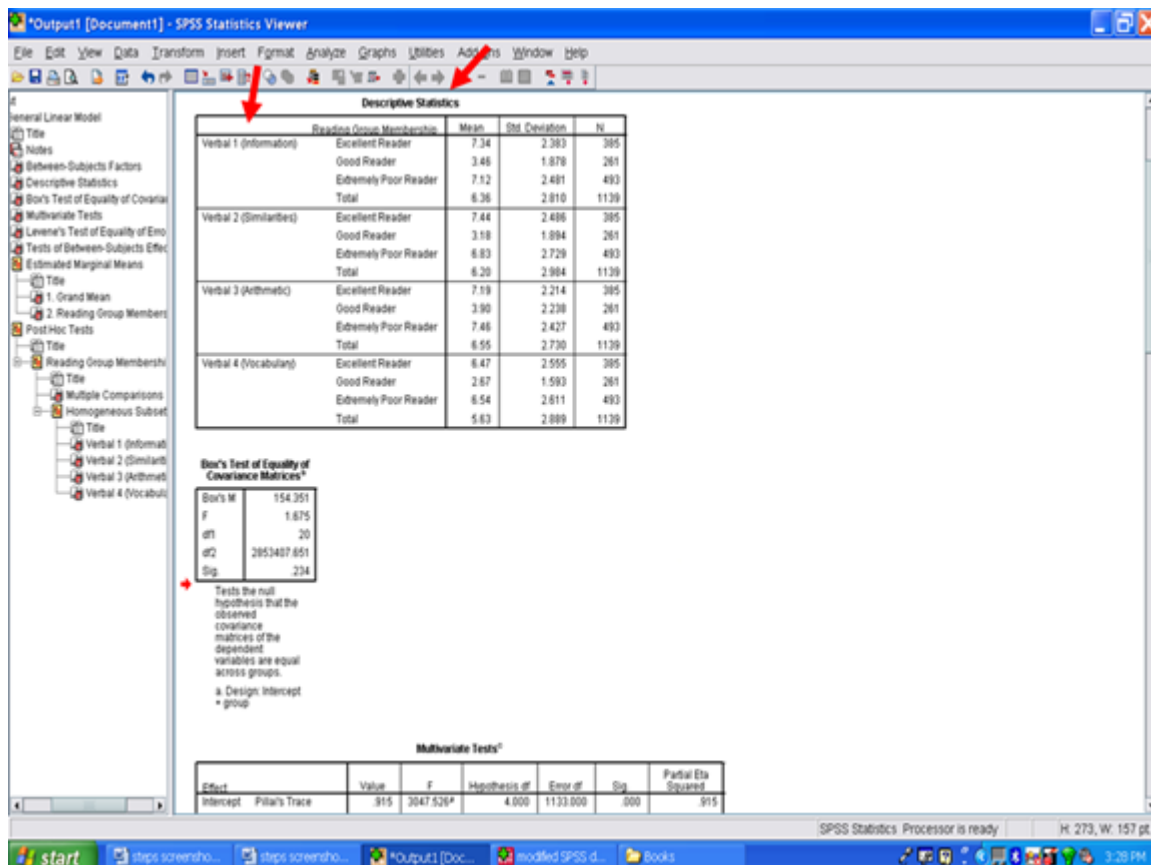
**Bor's Test of Equality of Covariance Matrices<sup>a</sup>**

Bor's M	154.351
F	7.675
df1	20
df2	2953487.651
Sig.	.000

Tests the null hypothesis that the covariance matrices are equal.

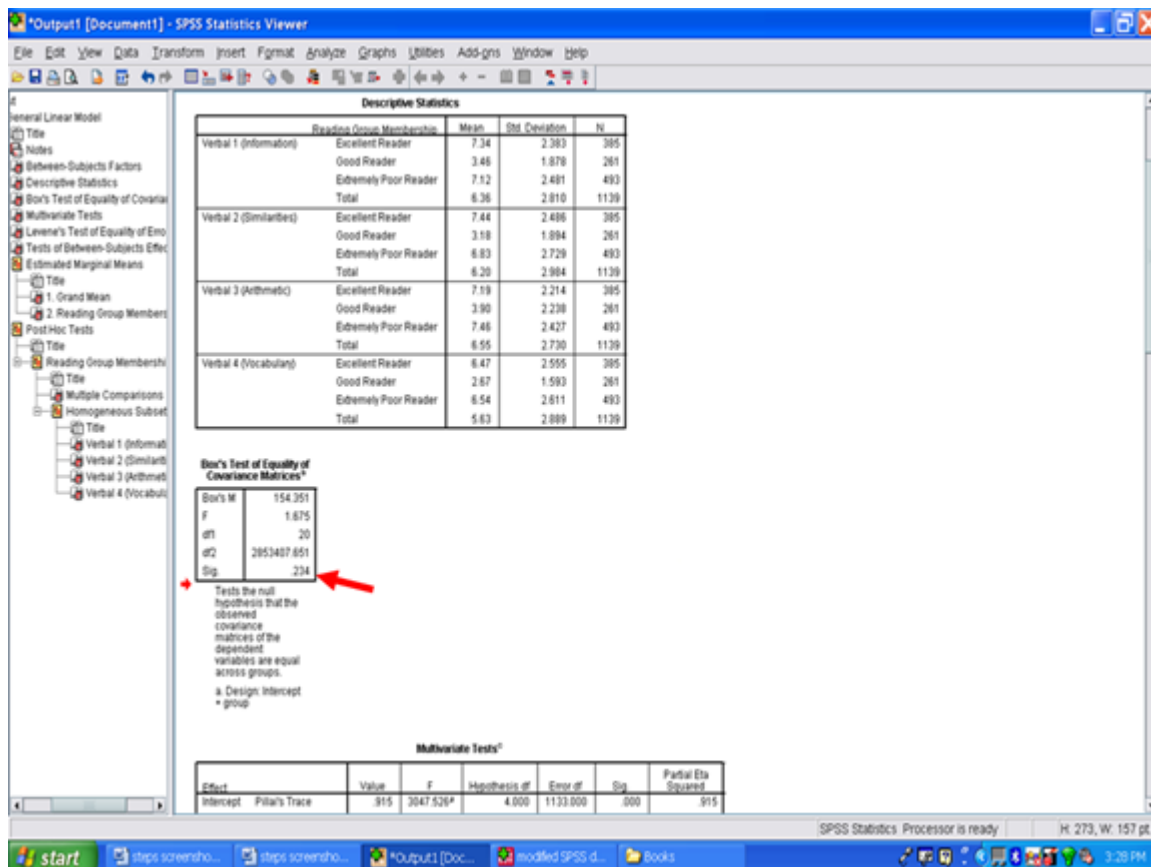
Focusing on the Descriptive Statistics table will show you a column for the *M*, for the *SD*, and for the sample size, *n*. This information has been provided to you for each of the four dependent variables for each of the three groups. This information will be used in your Results section and should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

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



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.

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



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.

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



**Multivariate Tests<sup>a</sup>**

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	
Intercept	Pillai's Trace	.915	3047.526 <sup>a</sup>	4.000	.1133.000	.000	.915
	Wilks' Lambda	.005	3047.526 <sup>a</sup>	4.000	.1133.000	.000	.915
	Hotelling's Trace	10.759	3047.526 <sup>a</sup>	4.000	.1133.000	.000	.915
	Roy's Largest Root	10.759	3047.526 <sup>a</sup>	4.000	.1133.000	.000	.915
group	Pillai's Trace	.469	86.881	8.000	.2268.000	.000	.235
	Wilks' Lambda	.542	101.858 <sup>a</sup>	8.000	.2268.000	.000	.264
	Hotelling's Trace	.827	117.002	8.000	.2264.000	.000	.293
	Roy's Largest Root	.802	227.447 <sup>b</sup>	4.000	.1134.000	.000	.445

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

**Levene's Test of Equality of Error Variances<sup>a</sup>**

	F	df1	df2	Sig.
Verbal 1 (Information)	.918	2	1136	.453
Verbal 2 (Similarities)	1.266	2	1136	.319
Verbal 3 (Arithmetic)	.981	2	1136	.651
Verbal 4 (Vocabulary)	1.894	2	1136	.249

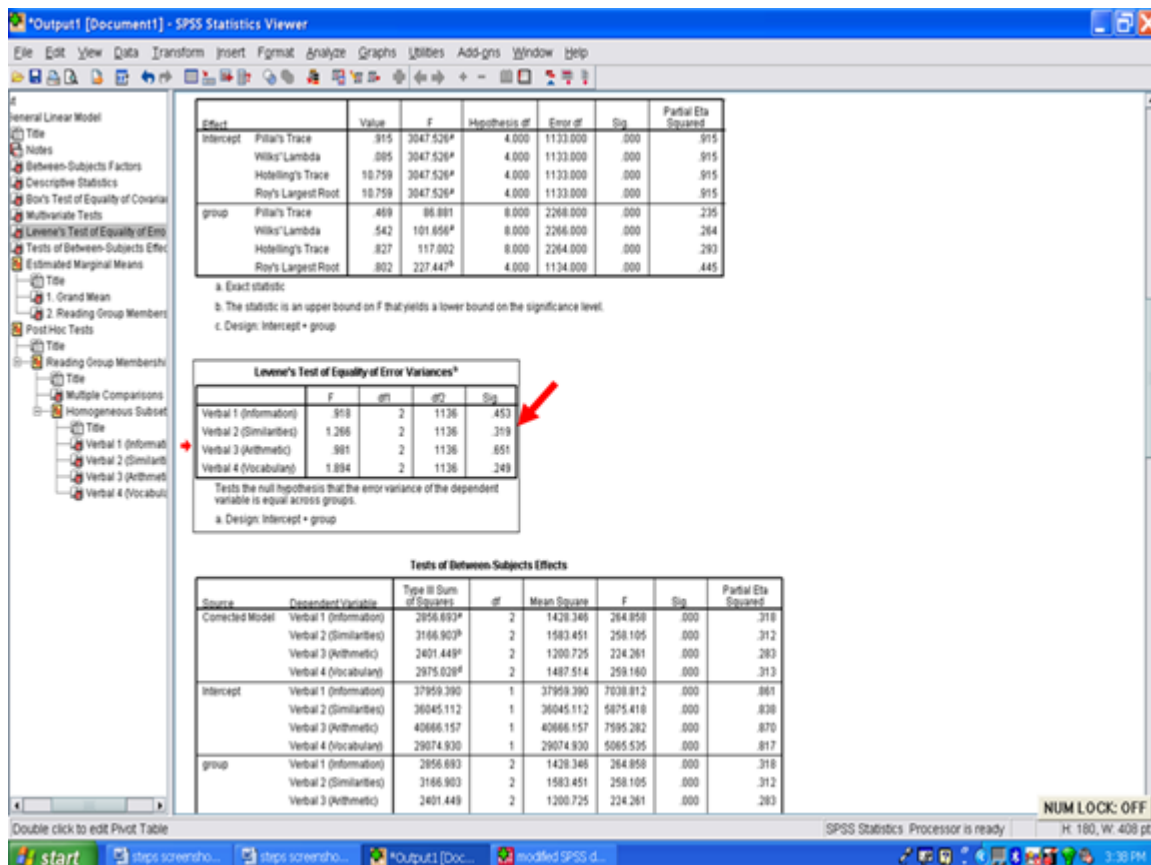
Tests the null hypothesis that the error variance of the dependent variable is equal across groups.  
a. Design: Intercept + group

**Tests of Between-Subjects Effects**

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Verbal 1 (Information)	2956.693 <sup>a</sup>	2	1478.346	264.858	.000	.318
	Verbal 2 (Similarities)	3166.903 <sup>a</sup>	2	1583.451	258.105	.000	.312
	Verbal 3 (Arithmetic)	2401.448 <sup>a</sup>	2	1200.725	224.261	.000	.283
	Verbal 4 (Vocabulary)	2975.028 <sup>a</sup>	2	1487.514	259.160	.000	.313
Intercept	Verbal 1 (Information)	37959.390	1	37959.390	7038.812	.000	.861
	Verbal 2 (Similarities)	36045.112	1	36045.112	5875.418	.000	.836
	Verbal 3 (Arithmetic)	40686.157	1	40686.157	7595.282	.000	.870
	Verbal 4 (Vocabulary)	29074.930	1	29074.930	5065.535	.000	.817
group	Verbal 1 (Information)	2956.693	2	1478.346	264.858	.000	.318
	Verbal 2 (Similarities)	3166.903	2	1583.451	258.105	.000	.312

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

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



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.

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

**Tests of Between-Subjects Effects**

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Verbal 1 (Information)	2956.893 <sup>a</sup>	2	1428.346	264.858	.000	.318
	Verbal 2 (Similarities)	3166.903 <sup>b</sup>	2	1583.451	258.105	.000	.312
	Verbal 3 (Arithmetic)	2401.449 <sup>c</sup>	2	1200.725	224.261	.000	.283
	Verbal 4 (Vocabulary)	2975.028 <sup>d</sup>	2	1487.514	258.160	.000	.313
Intercept	Verbal 1 (Information)	37959.390	1	37959.390	7038.812	.000	.861
	Verbal 2 (Similarities)	36045.112	1	36045.112	5875.418	.000	.838
	Verbal 3 (Arithmetic)	40666.157	1	40666.157	7595.282	.000	.870
	Verbal 4 (Vocabulary)	29074.930	1	29074.930	5065.535	.000	.817
group	Verbal 1 (Information)	2956.893	2	1428.346	264.858	.000	.318
	Verbal 2 (Similarities)	3166.903	2	1583.451	258.105	.000	.312
	Verbal 3 (Arithmetic)	2401.449	2	1200.725	224.261	.000	.283
	Verbal 4 (Vocabulary)	2975.028	2	1487.514	258.160	.000	.313
Error	Verbal 1 (Information)	6126.300	1136	5.393			
	Verbal 2 (Similarities)	6869.248	1136	6.135			
	Verbal 3 (Arithmetic)	6082.296	1136	5.354			
	Verbal 4 (Vocabulary)	8520.361	1136	5.740			
Total	Verbal 1 (Information)	54991.000	1139				
	Verbal 2 (Similarities)	53959.000	1139				
	Verbal 3 (Arithmetic)	57370.000	1139				
	Verbal 4 (Vocabulary)	45603.000	1139				
Corrected Total	Verbal 1 (Information)	8882.992	1138				
	Verbal 2 (Similarities)	10136.151	1138				
	Verbal 3 (Arithmetic)	8483.745	1138				
	Verbal 4 (Vocabulary)	9495.389	1138				

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

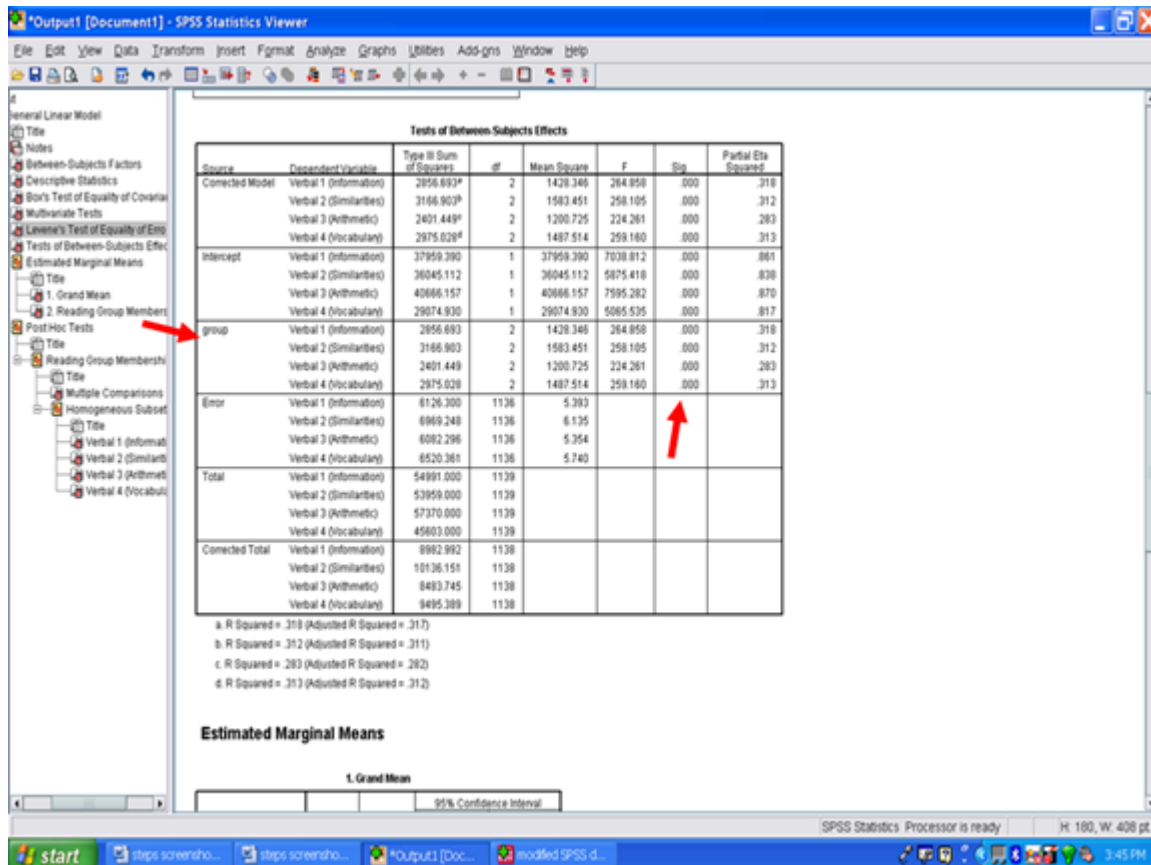
**Estimated Marginal Means**

1. Grand Mean

95% Confidence Interval

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.

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



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  $\eta^2$ , is the effect size metric provided to you in the SPSS output. To interpret whether it is small, moderate, or large, see Cohen (1988).

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

The screenshot displays the SPSS Statistics Viewer window for a General Linear Model. The 'Tests of Between-Subjects Effects' table is the primary focus. It lists various sources of variance (Corrected Model, Intercept, group, Error, Total, Corrected Total) and their corresponding dependent variables (Verbal 1, 2, 3, 4). The table includes columns for Type III Sum of Squares, df, Mean Square, F, Sig., and Partial Eta Squared. A red arrow points to the 'Error' row for 'Verbal 1 (Information)', which has a Sig. value of .000. Below the table, the 'Estimated Marginal Means' section is partially visible, showing the '1. Grand Mean'.

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Verbal 1 (Information)	2956.893 <sup>a</sup>	2	1428.346	264.858	.000	.318
	Verbal 2 (Similarities)	3166.903 <sup>a</sup>	2	1583.451	258.105	.000	.312
	Verbal 3 (Arithmetic)	2401.449 <sup>a</sup>	2	1200.725	224.261	.000	.283
	Verbal 4 (Vocabulary)	2975.028 <sup>a</sup>	2	1487.514	258.160	.000	.313
Intercept	Verbal 1 (Information)	37959.390	1	37959.390	7038.812	.000	.861
	Verbal 2 (Similarities)	36045.112	1	36045.112	5875.418	.000	.838
	Verbal 3 (Arithmetic)	40666.157	1	40666.157	7595.282	.000	.870
	Verbal 4 (Vocabulary)	29074.930	1	29074.930	5065.535	.000	.817
group	Verbal 1 (Information)	2956.893	2	1428.346	264.858	.000	.318
	Verbal 2 (Similarities)	3166.903	2	1583.451	258.105	.000	.312
	Verbal 3 (Arithmetic)	2401.449	2	1200.725	224.261	.000	.283
	Verbal 4 (Vocabulary)	2975.028	2	1487.514	258.160	.000	.313
Error	Verbal 1 (Information)	6126.300	1136	5.393			
	Verbal 2 (Similarities)	6869.248	1136	6.135			
	Verbal 3 (Arithmetic)	6082.296	1136	5.354			
	Verbal 4 (Vocabulary)	8520.361	1136	5.740			
Total	Verbal 1 (Information)	54991.000	1139				
	Verbal 2 (Similarities)	53959.000	1139				
	Verbal 3 (Arithmetic)	57370.000	1139				
	Verbal 4 (Vocabulary)	45603.000	1139				
Corrected Total	Verbal 1 (Information)	8882.992	1138				
	Verbal 2 (Similarities)	10136.151	1138				
	Verbal 3 (Arithmetic)	8483.745	1138				
	Verbal 4 (Vocabulary)	9495.389	1138				

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

**Estimated Marginal Means**

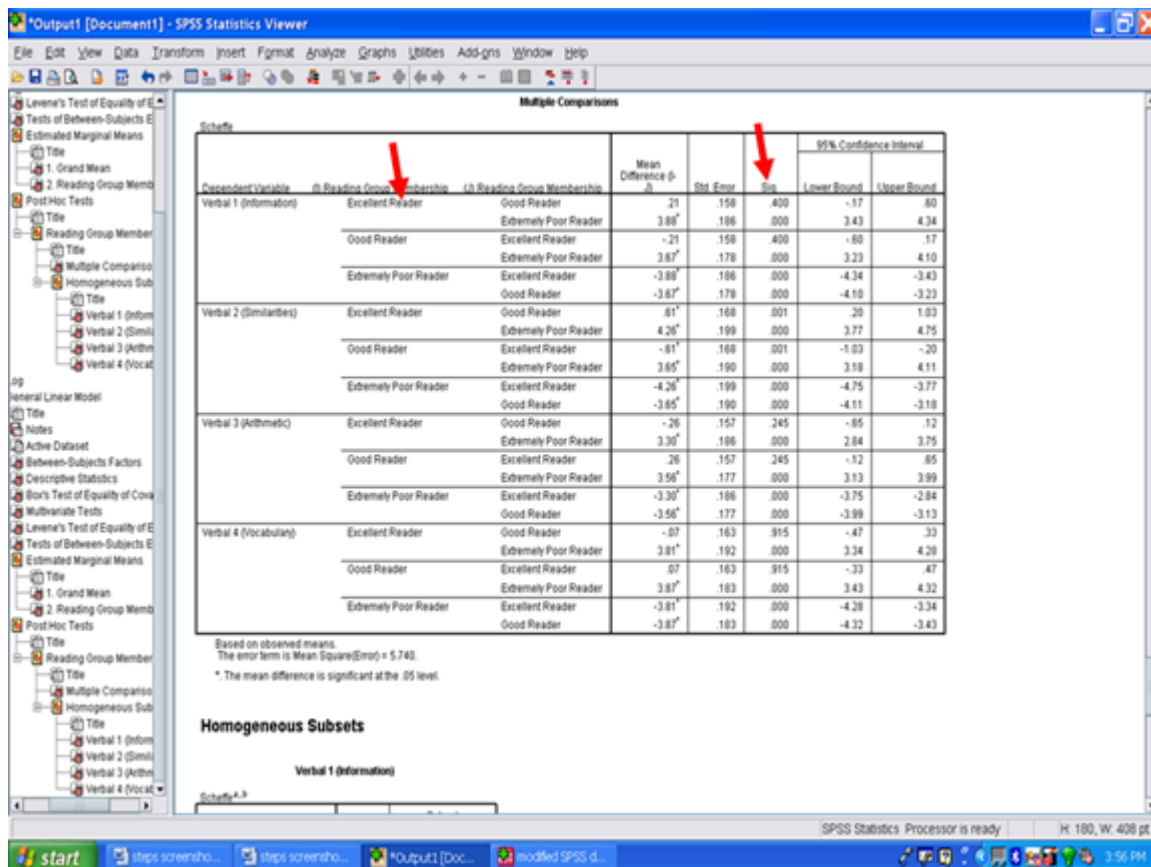
1. Grand Mean

95% Confidence Interval

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.

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





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.

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

**\*Output1 [Document1] - SPSS Statistics Viewer**

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

**Multiple Comparisons**

Schedule

Dependent Variable	Reading Group Membership	Reading Group Membership	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Verbal 1 (Information)	Excellent Reader	Good Reader	.21	.158	.400	-.17	.60
		Extremely Poor Reader	3.88 <sup>a</sup>	.186	.000	3.43	4.34
	Good Reader	Excellent Reader	-.21	.158	.400	-.60	.17
		Extremely Poor Reader	3.67 <sup>a</sup>	.170	.000	3.23	4.10
	Extremely Poor Reader	Excellent Reader	-3.88 <sup>a</sup>	.186	.000	-4.34	-3.43
		Good Reader	-3.67 <sup>a</sup>	.170	.000	-4.10	-3.23
Verbal 2 (Similarities)	Excellent Reader	Good Reader	.81 <sup>a</sup>	.168	.001	.28	1.33
		Extremely Poor Reader	4.26 <sup>a</sup>	.199	.000	3.77	4.75
	Good Reader	Excellent Reader	-.81 <sup>a</sup>	.168	.001	-1.33	-.28
		Extremely Poor Reader	3.65 <sup>a</sup>	.190	.000	3.18	4.11
	Extremely Poor Reader	Excellent Reader	-4.26 <sup>a</sup>	.199	.000	-4.75	-3.77
		Good Reader	-3.65 <sup>a</sup>	.190	.000	-4.11	-3.18
Verbal 3 (Arithmetic)	Excellent Reader	Good Reader	-.26	.157	.245	-.65	.12
		Extremely Poor Reader	3.30 <sup>a</sup>	.186	.000	2.84	3.75
	Good Reader	Excellent Reader	.26	.157	.245	-.12	.65
		Extremely Poor Reader	3.56 <sup>a</sup>	.177	.000	3.13	3.99
	Extremely Poor Reader	Excellent Reader	-3.30 <sup>a</sup>	.186	.000	-3.75	-2.84
		Good Reader	-3.56 <sup>a</sup>	.177	.000	-3.99	-3.13
Verbal 4 (Vocabulary)	Excellent Reader	Good Reader	-.67	.163	.915	-.47	.33
		Extremely Poor Reader	3.81 <sup>a</sup>	.192	.000	3.34	4.28
	Good Reader	Excellent Reader	.67	.163	.915	-.33	.47
		Extremely Poor Reader	3.87 <sup>a</sup>	.183	.000	3.43	4.32
	Extremely Poor Reader	Excellent Reader	-3.81 <sup>a</sup>	.192	.000	-4.28	-3.34
		Good Reader	-3.87 <sup>a</sup>	.183	.000	-4.32	-3.43

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

**Homogeneous Subsets**

Verbal 1 (Information)

Schedule

SPSS Statistics Processor is ready H: 180, W: 408 pt. 3:57 PM

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.

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



## Chapter 4

# Discriminant Analysis: Assumptions<sup>1</sup>



NOTE: This chapter is published by NCPEA Press<sup>2</sup> 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<sup>3</sup>) 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/><sup>4</sup>; to the *Electronic Statistics Textbook* (2011)

---

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

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

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

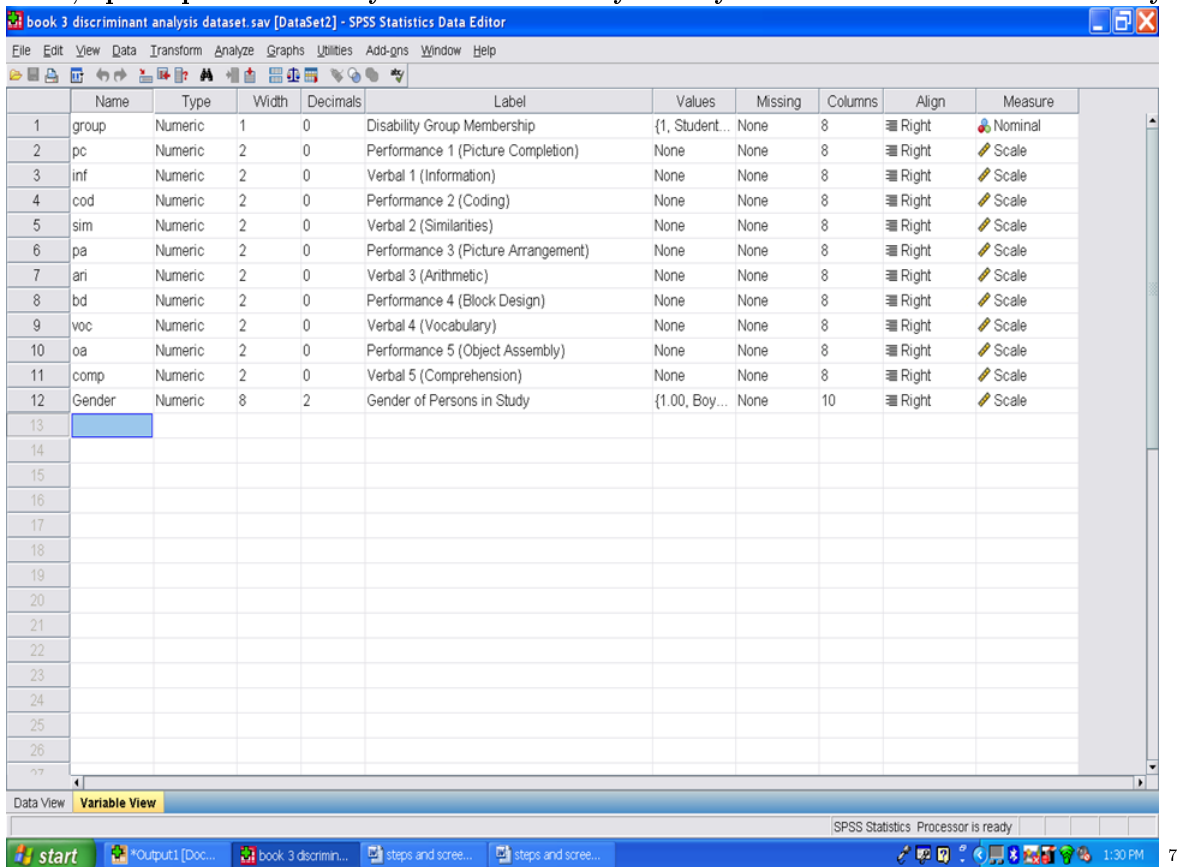
<sup>4</sup><http://davidmlane.com/hyperstat/>

at <http://www.statsoft.com/textbook/><sup>5</sup> ; 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](http://www.amazon.com/Discovering-Statistics-Introducing-Statistical-Method/dp/1847879071/ref=sr_1_1?s=books&ie=UTF8&qid=1304967862&sr=1-1)<sup>6</sup>

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

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

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

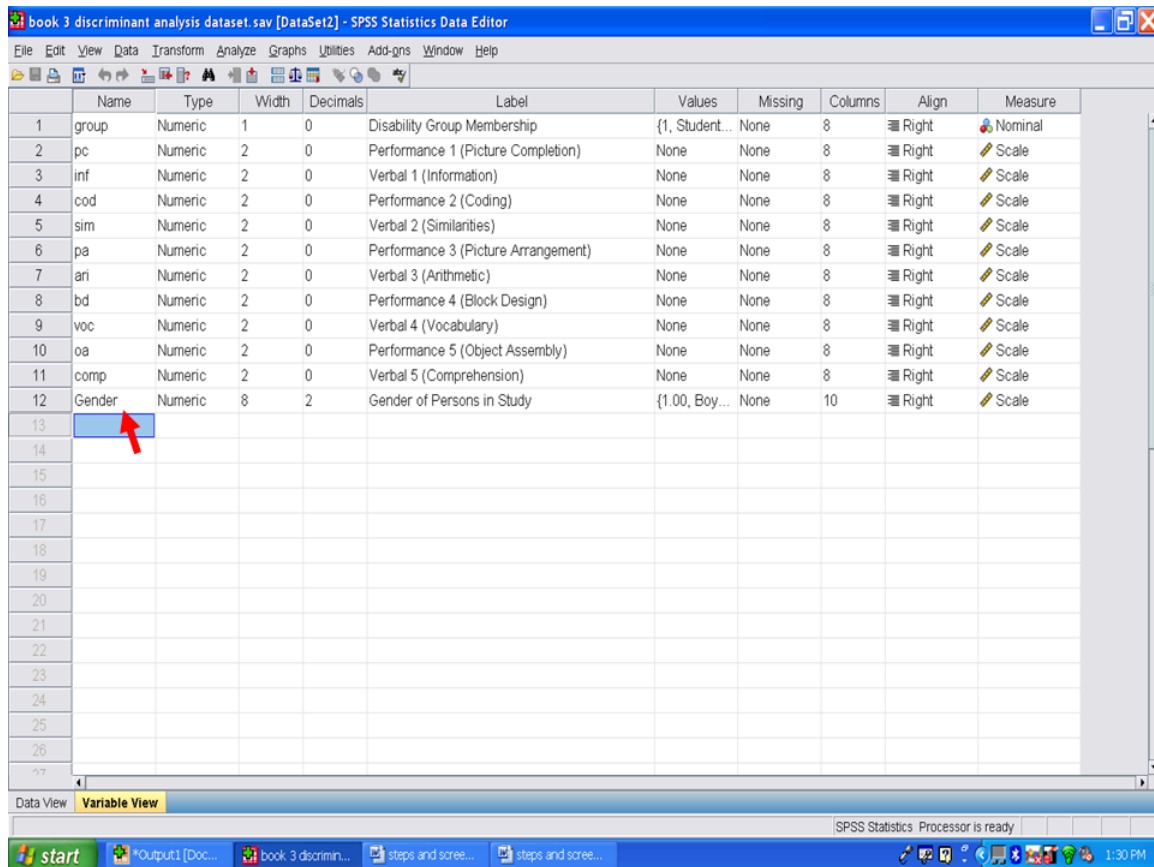


Our independent variable is gender. Boys are labeled as group 1 and girls are labeled as group 2.

<sup>5</sup><http://www.statsoft.com/textbook/>

<sup>6</sup>[http://www.amazon.com/Discovering-Statistics-Introducing-Statistical-Method/dp/1847879071/ref=sr\\_1\\_1?s=books&ie=UTF8&qid=1304967862&sr=1-1](http://www.amazon.com/Discovering-Statistics-Introducing-Statistical-Method/dp/1847879071/ref=sr_1_1?s=books&ie=UTF8&qid=1304967862&sr=1-1)

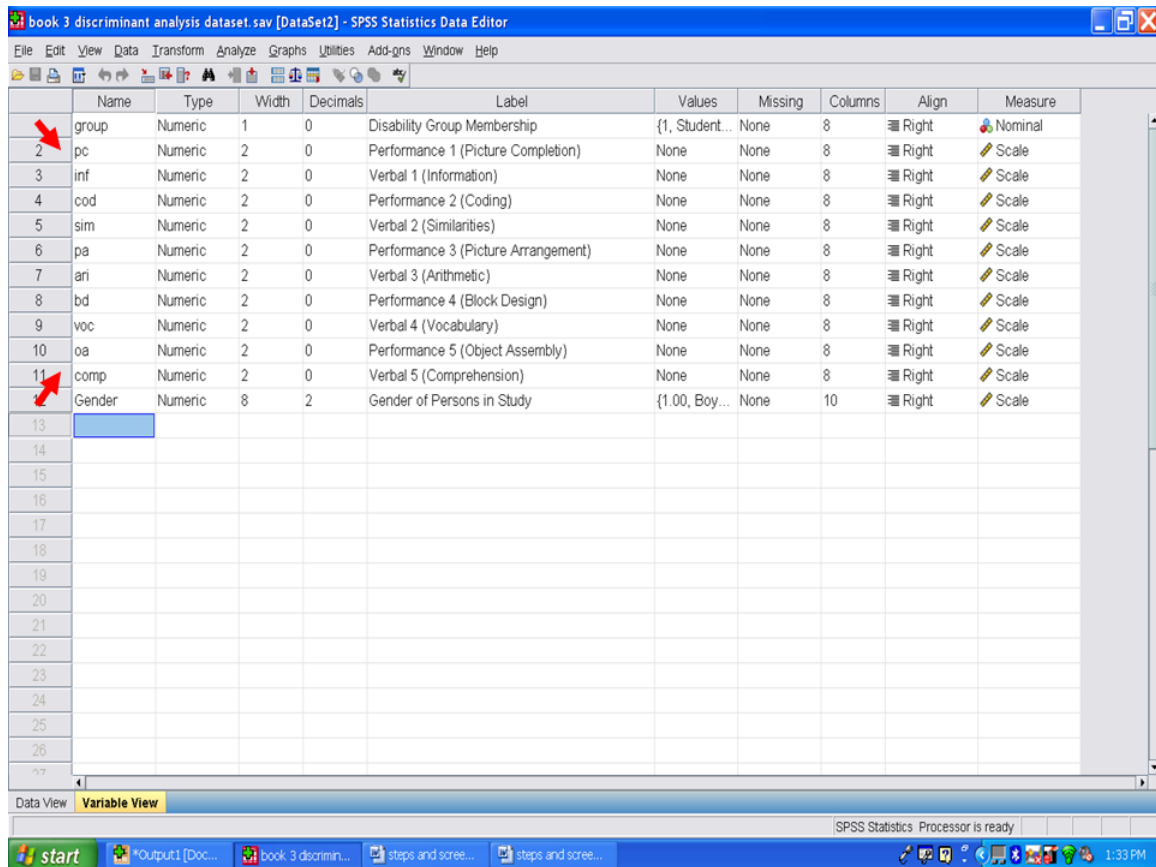
<sup>7</sup><http://cnx.org/content/m40733/latest/13.1.png/image>



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

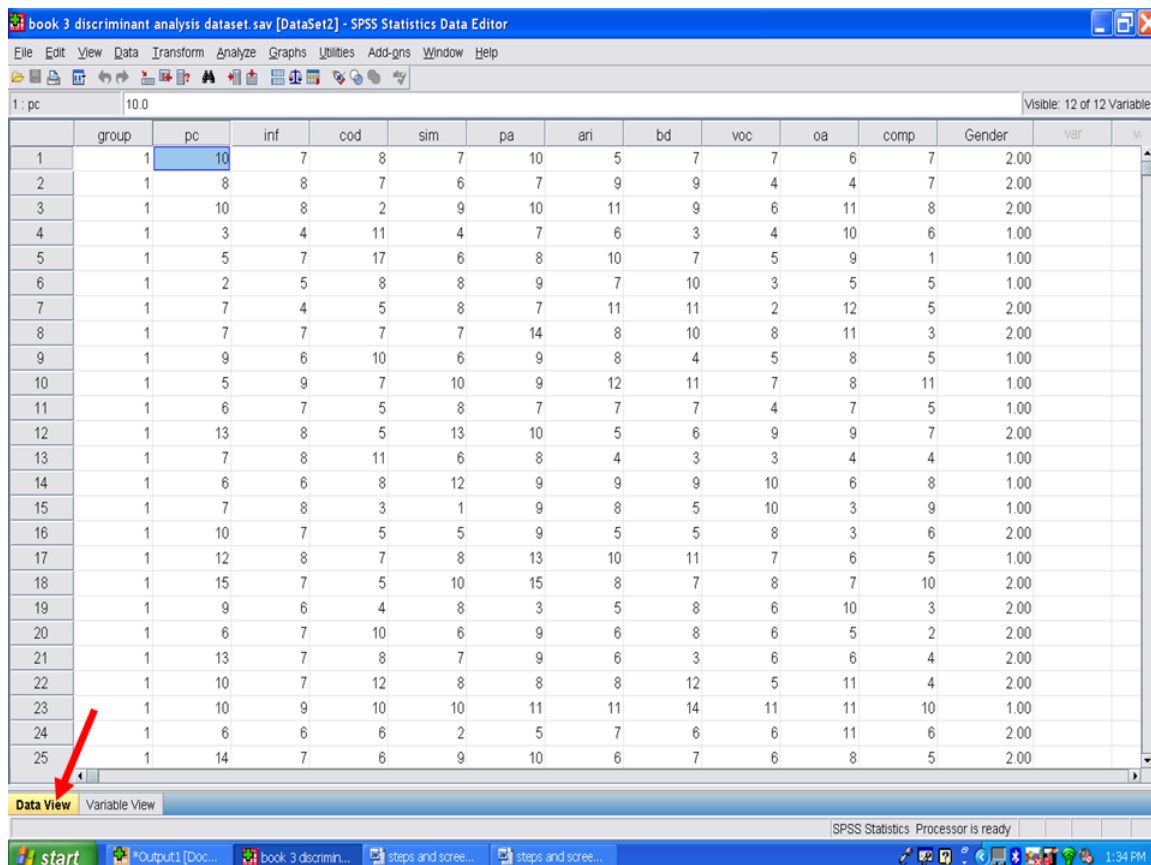
Our dependent variables, the ones we will use to differentiate boys from girls are 10 subscales from the Wechsler Intelligence Scale for Children-Third Edition: Picture Completion (pc), Information (inf), Coding (cod), Similarities (sim), Picture Arrangement (pa), Arithmetic (ari), Block Design (bd),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

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



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

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



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

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

1: pc 10.0 Visible: 12 of 12 Variables

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

Data View Variable View

SPSS Statistics Processor is ready

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

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

## 4.1 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><sup>11</sup>

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](http://www.statistics.com/index.php?page=glossary&term_id=356)<sup>12</sup> and <http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb><sup>13</sup>

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

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

<sup>12</sup>[http://www.statistics.com/index.php?page=glossary&term\\_id=356](http://www.statistics.com/index.php?page=glossary&term_id=356)

<sup>13</sup><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  $\pm 3$  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](http://www.statistics.com/index.php?page=glossary&term_id=326)<sup>14</sup> and <http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb><sup>15</sup>

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  $\pm 3$  range, the dataset is not normally distributed.

## 4.2 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><sup>16</sup>

## 4.3 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><sup>17</sup>

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.

## 4.4 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-

<sup>14</sup>[http://www.statistics.com/index.php?page=glossary&term\\_id=326](http://www.statistics.com/index.php?page=glossary&term_id=326)

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

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

<sup>17</sup><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><sup>18</sup>

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.

## 4.5 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><sup>19</sup>

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

---

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

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





## Chapter 5

# Discriminant Analysis: Part I<sup>1</sup>



NOTE: This chapter is published by NCPEA Press<sup>2</sup> 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<sup>3</sup>) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

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

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

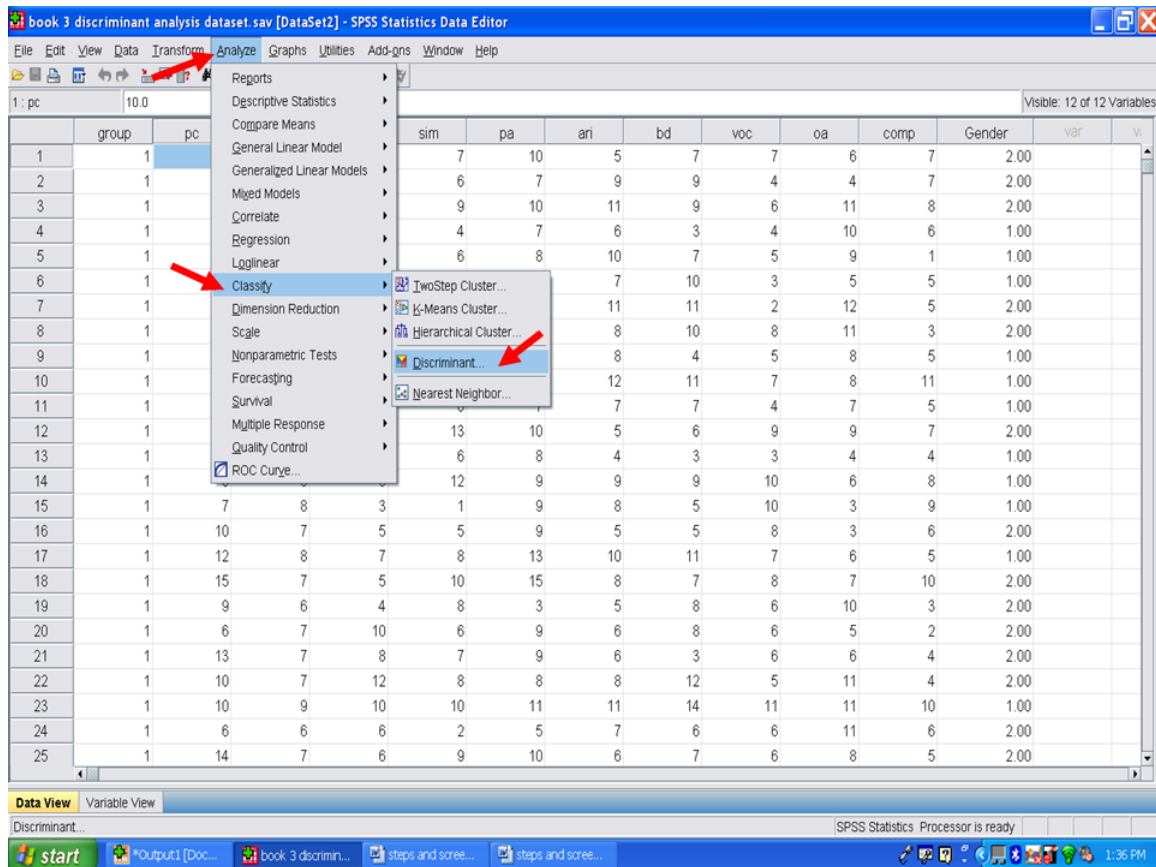
Click on Discriminant

---

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

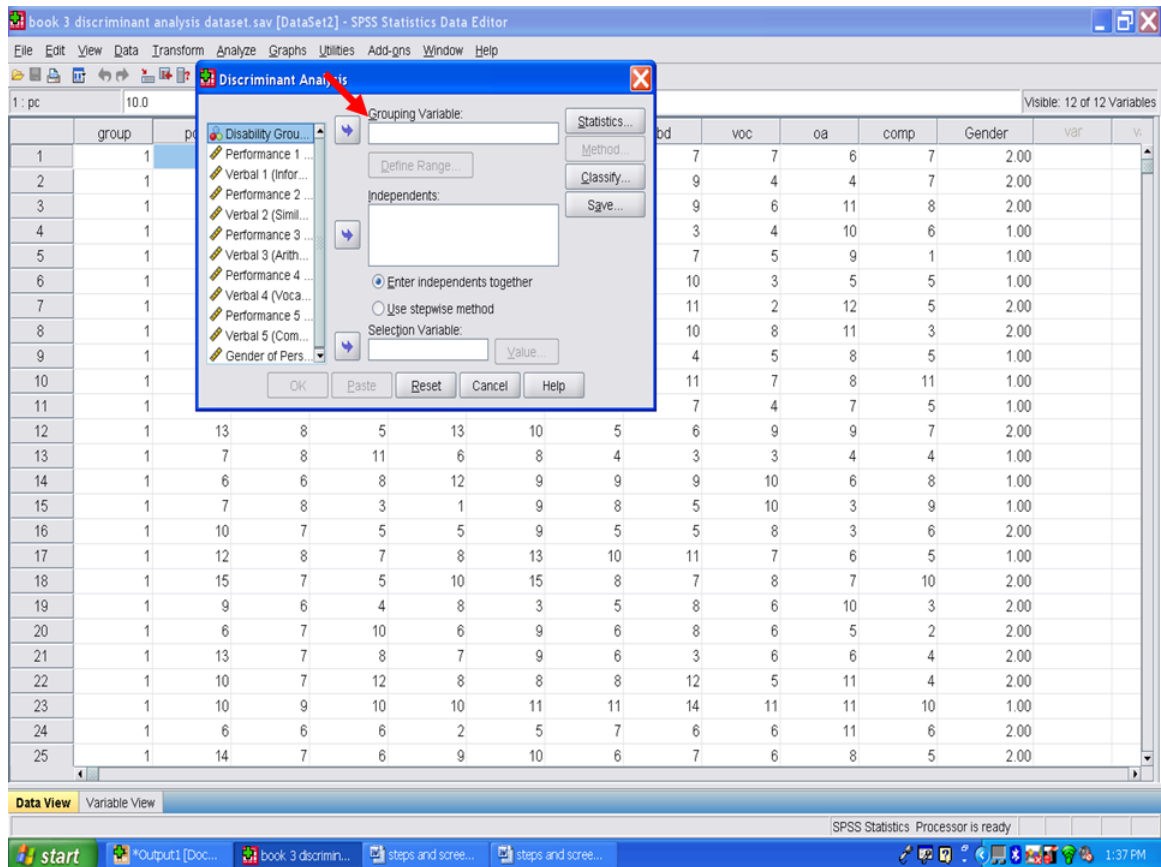
<sup>2</sup><http://www.ncpeapublications.org/books.html>

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



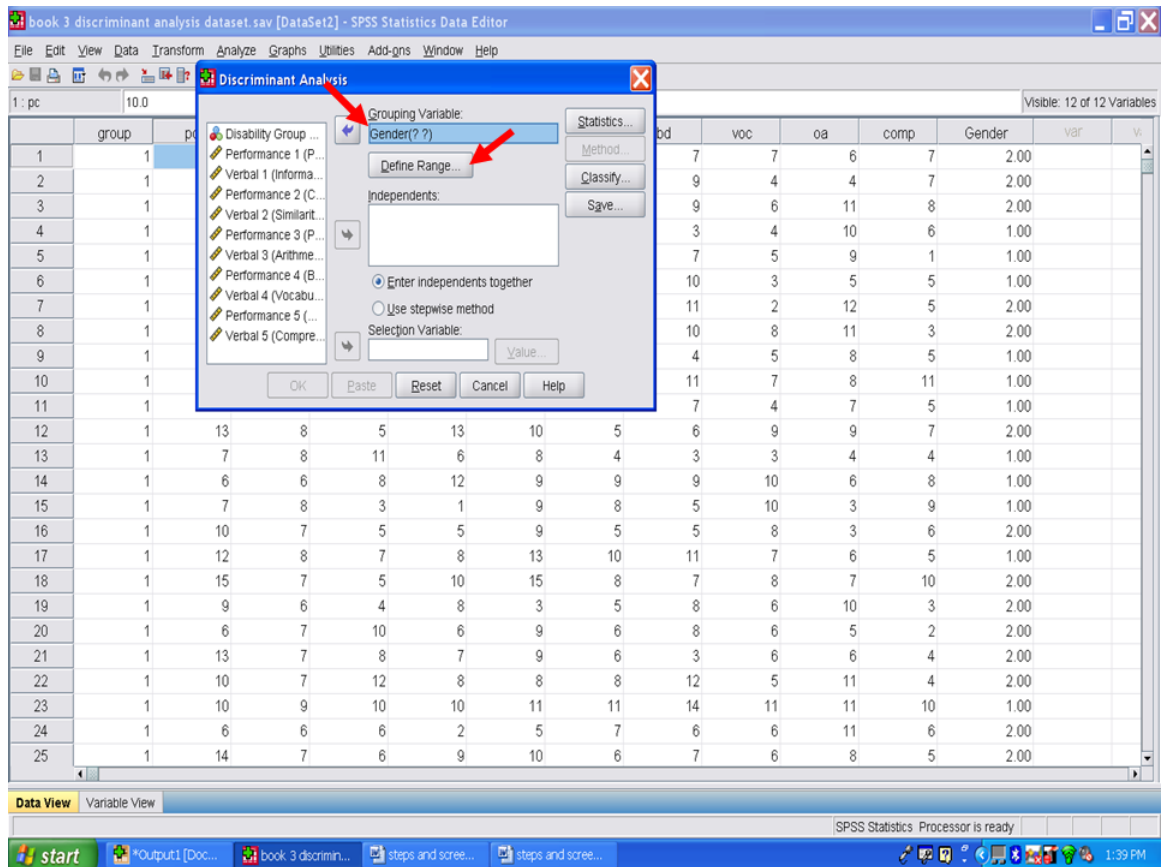
The following screen will then appear. Grouping variable is your independent variable. Remember that in our example that gender is our independent or grouping variable.

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



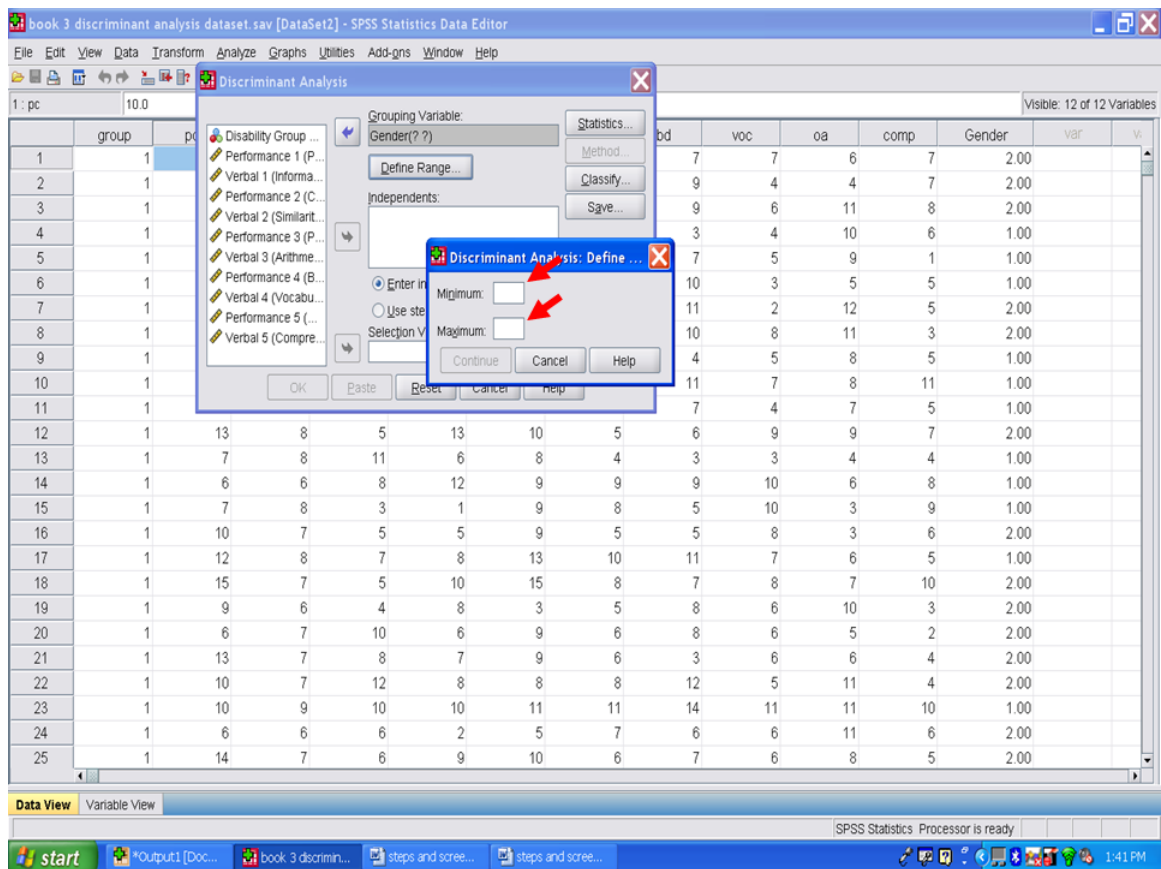
Highlight gender and send it into the Grouping variable cell. When you do so, you will note that after Gender is a set of ( ) with question marks. Click on Define Range.

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



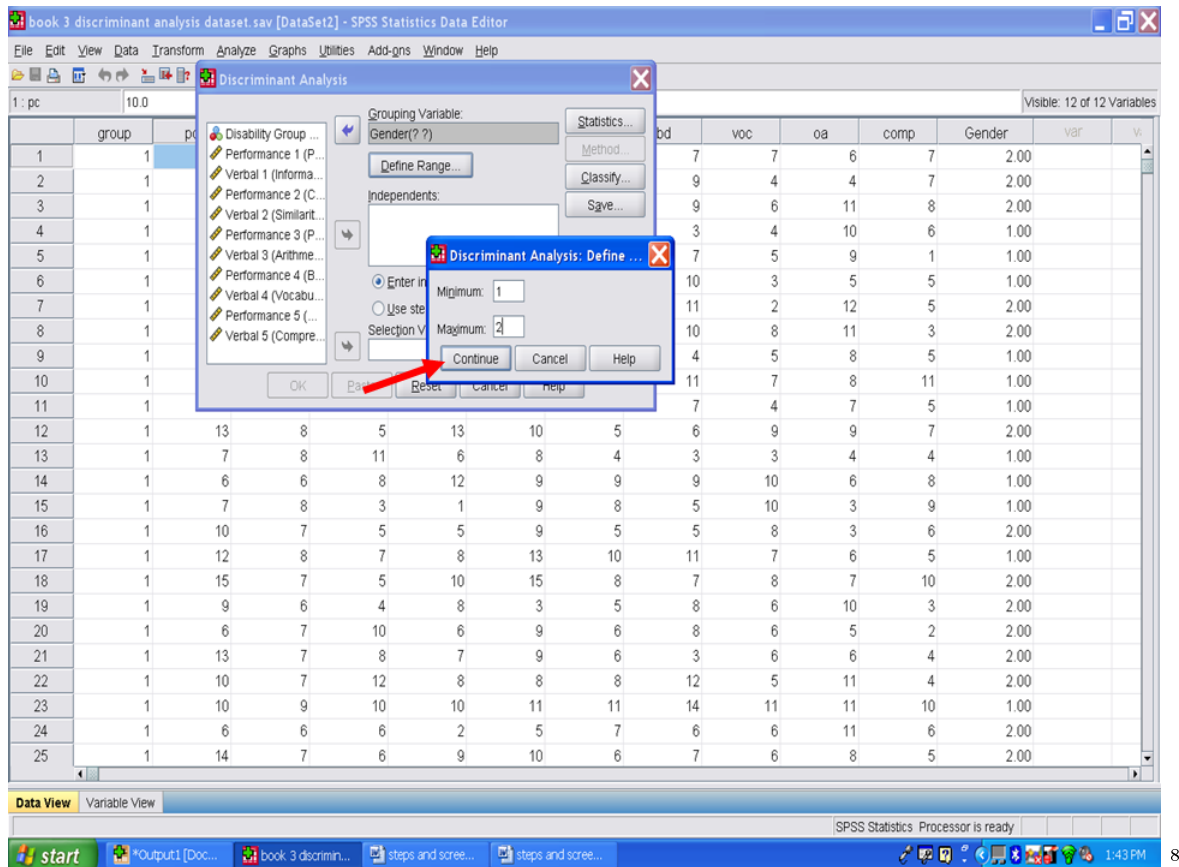
After clicking on Define Range, the following screen will appear: For the Minimum box, we will type in the number 1 because it represents boys. We will type in the number 2 in the Maximum box because it represents girls.

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



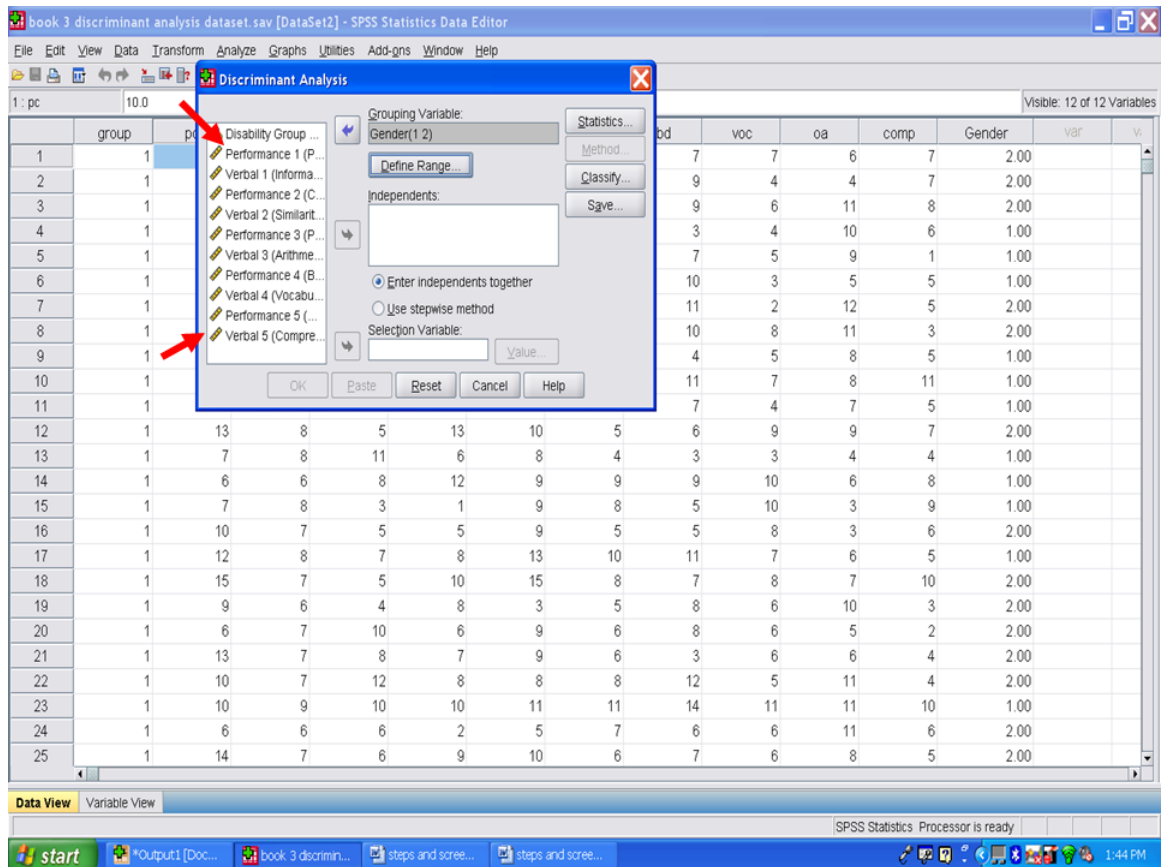
After typing in 1 and 2, then click on Continue.

<sup>7</sup> <http://cnx.org/content/m40735/latest/14.4.png/image>



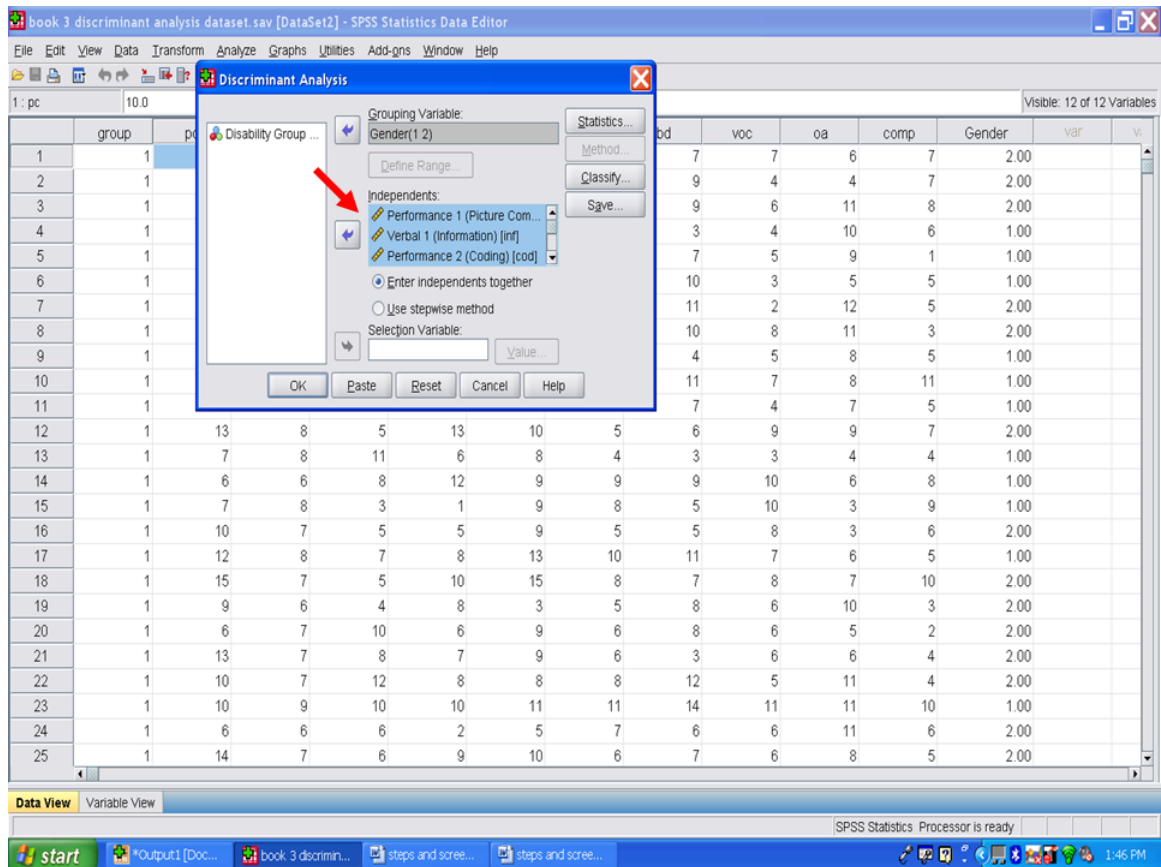
Next we will send over our 10 dependent variables, the 10 subscales, to the Independents box. Highlight Performance 1 through Verbal 5; click on the middle arrow; send them to the Independents box.

<sup>8</sup><http://cnx.org/content/m40735/latest/14.5.png/image>



The screen should now look like the one below:

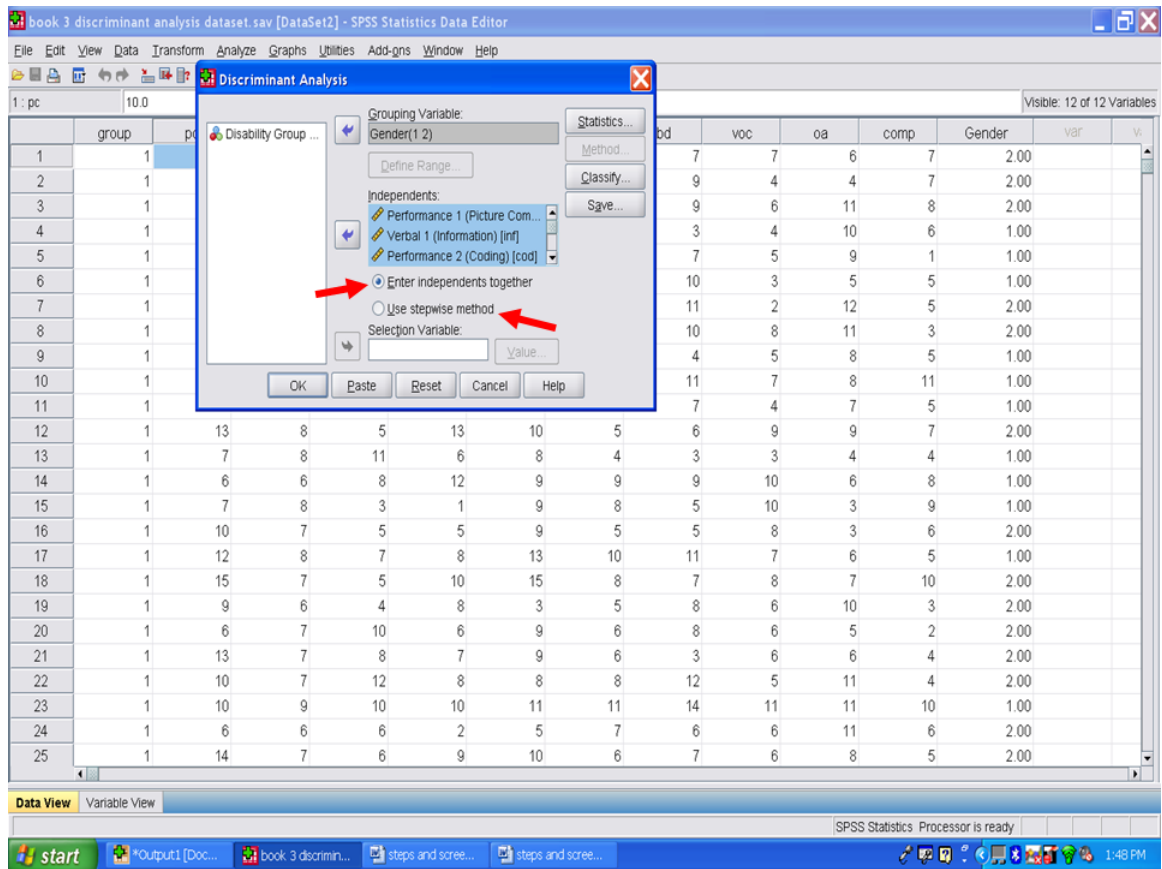
<sup>9</sup><http://cnx.org/content/m40735/latest/14.6.png/image>



Next we will change the specific type of discriminant analysis from the default of Enter Independents Together to Use Stepwise Method.

<sup>10</sup><http://cnx.org/content/m40735/latest/14.7.png/image>





<sup>11</sup> <http://cnx.org/content/m40735/latest/14.8.png/image>



## Chapter 6

# Discriminant Analysis: Part II<sup>1</sup>



NOTE: This chapter is published by NCPEA Press<sup>2</sup> 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<sup>3</sup>) 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.

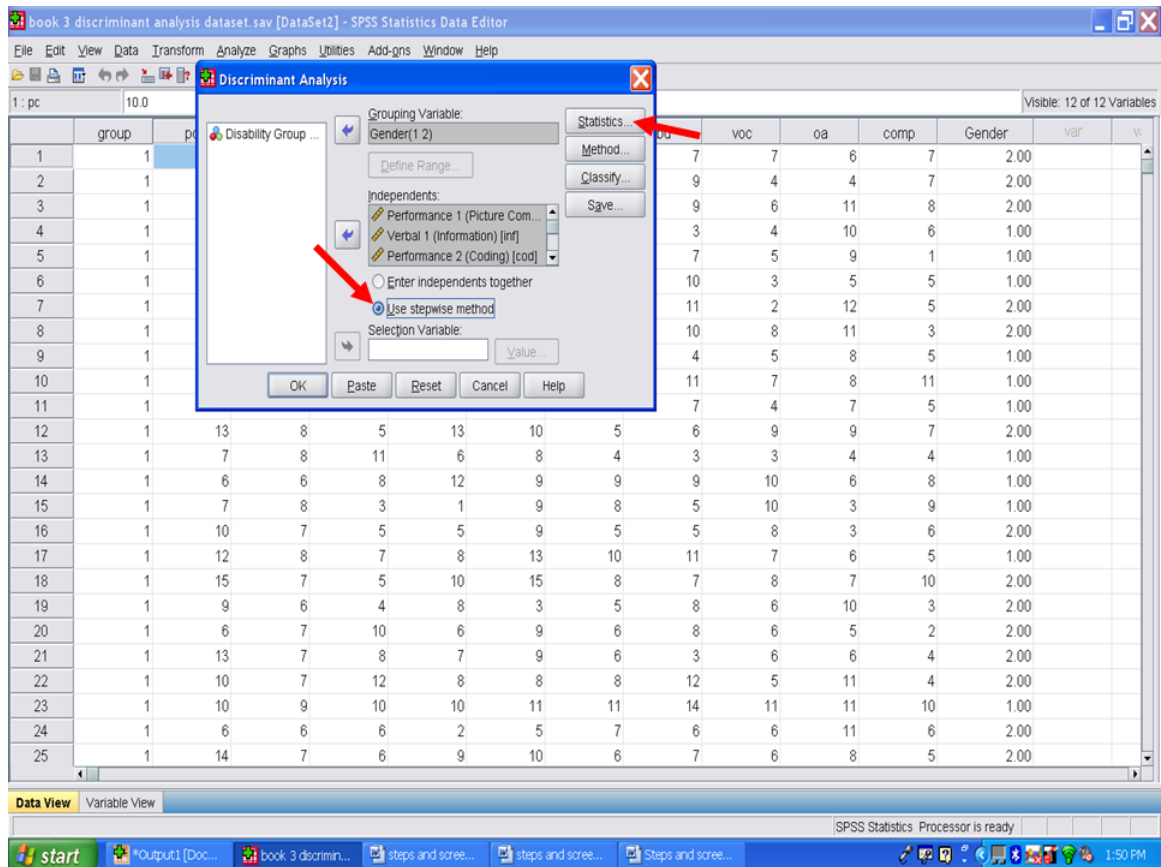
**Your screen should now look like the one below: Now click on Statistics.**

---

<sup>1</sup>This content is available online at <<http://cnx.org/content/m40736/1.2/>>.

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

<sup>3</sup><http://cnx.org/content/m40736/latest/www.writingandstatisticalhelp>

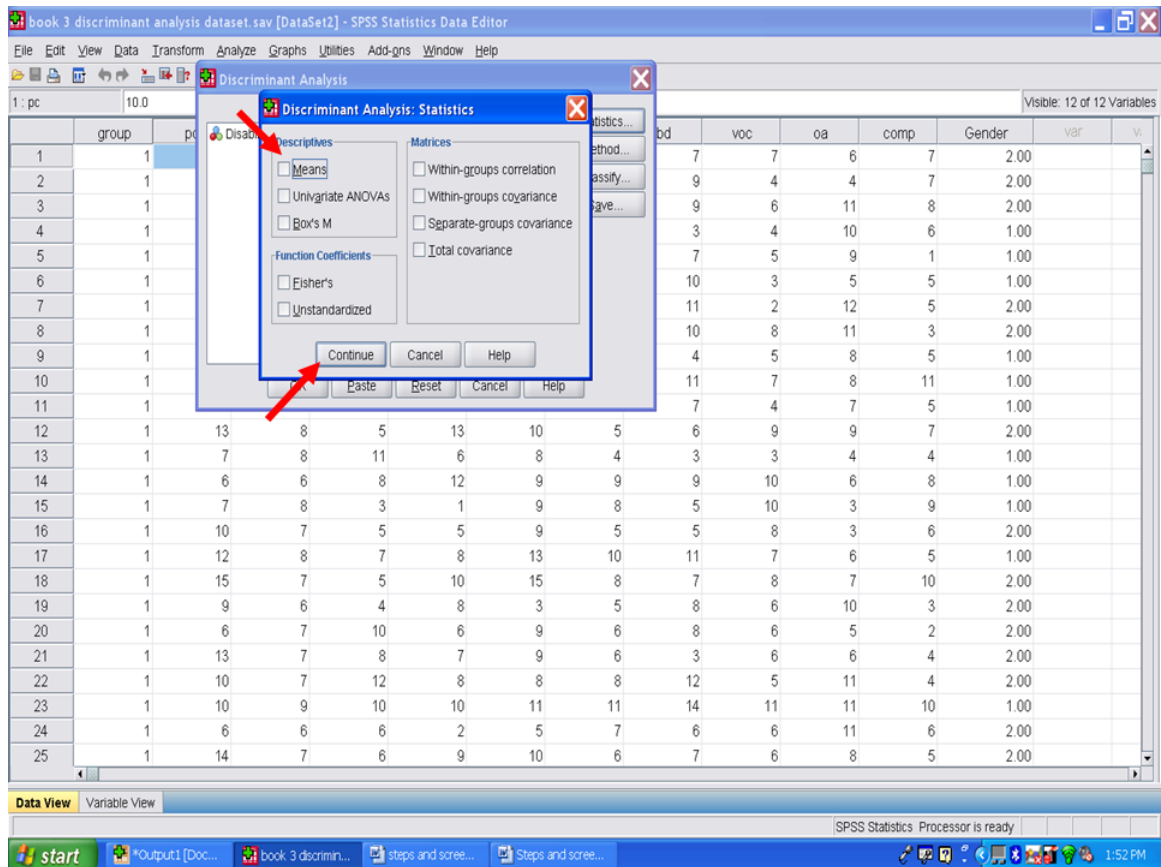


The following screen then appears.

Click on Means.

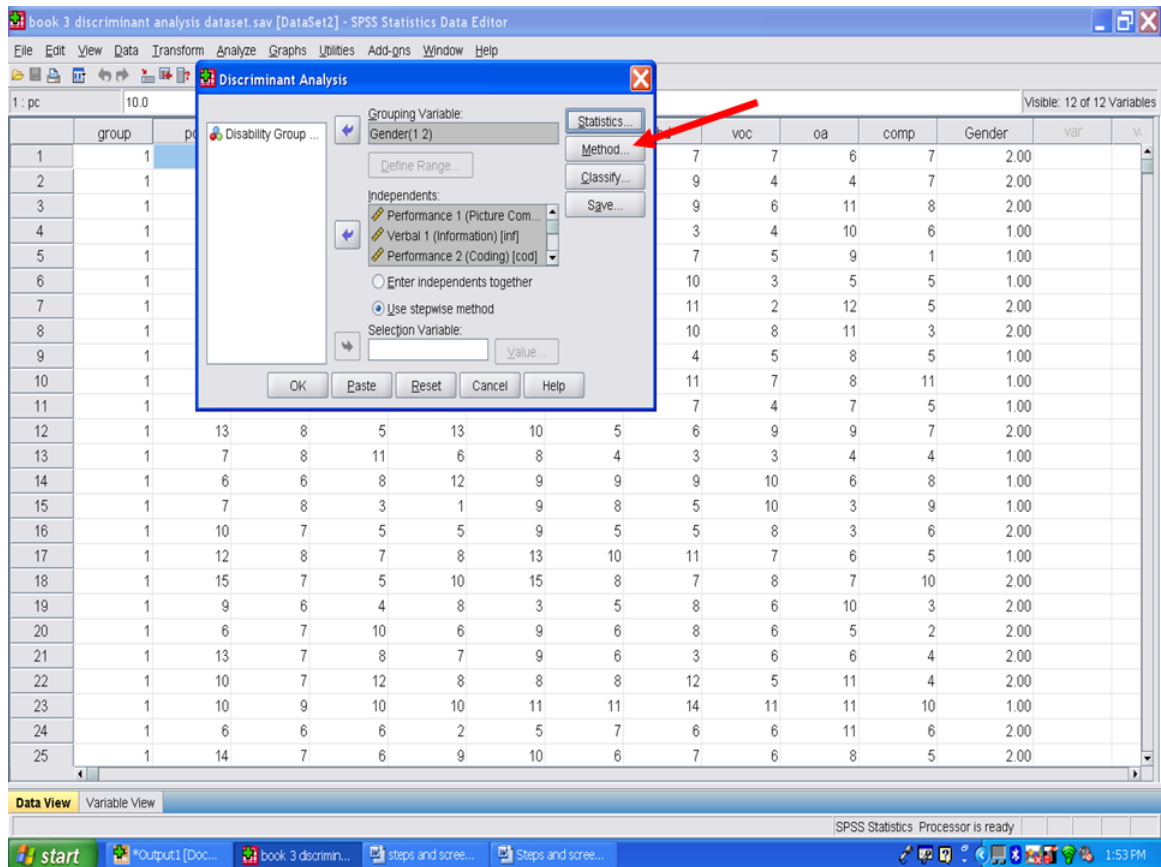
Click on Continue.

<sup>4</sup><http://cnx.org/content/m40736/latest/15.1.png/image>



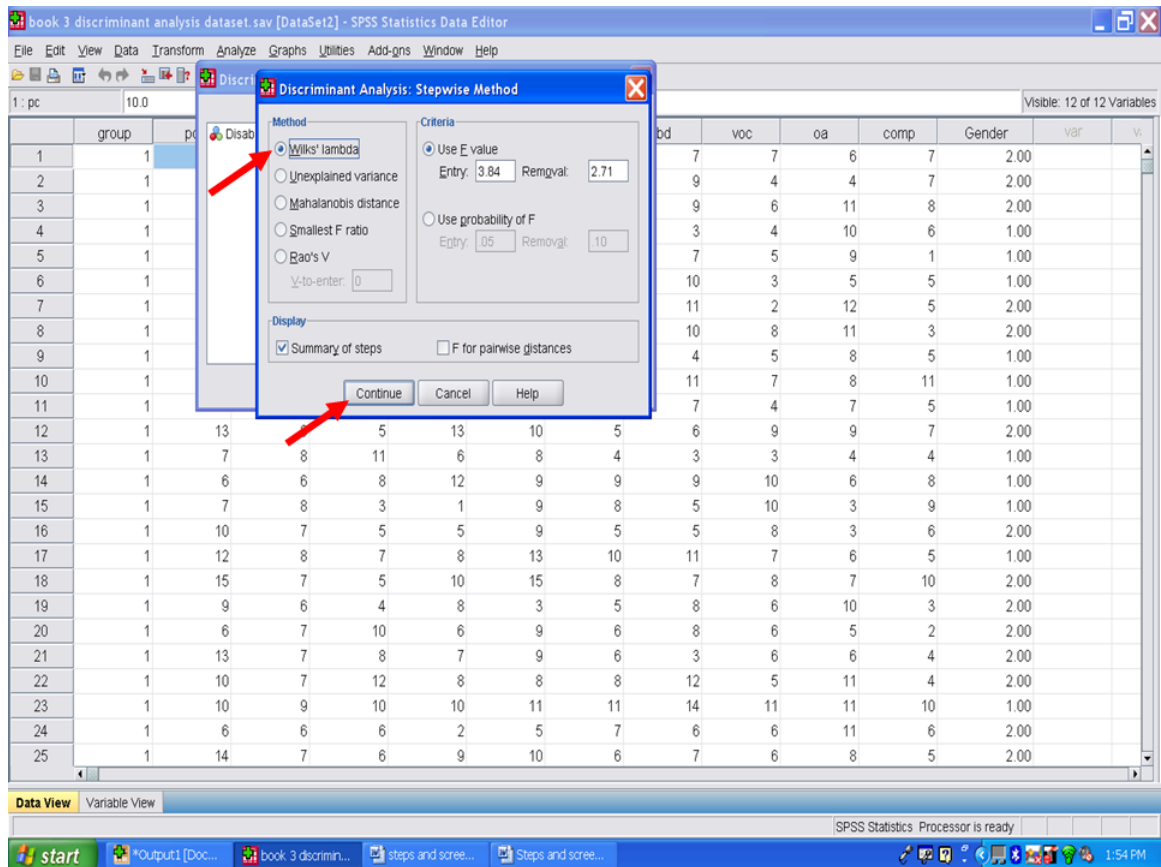
When the screen below appears, click on Method

<sup>5</sup><http://cnx.org/content/m40736/latest/15.2.png/image>



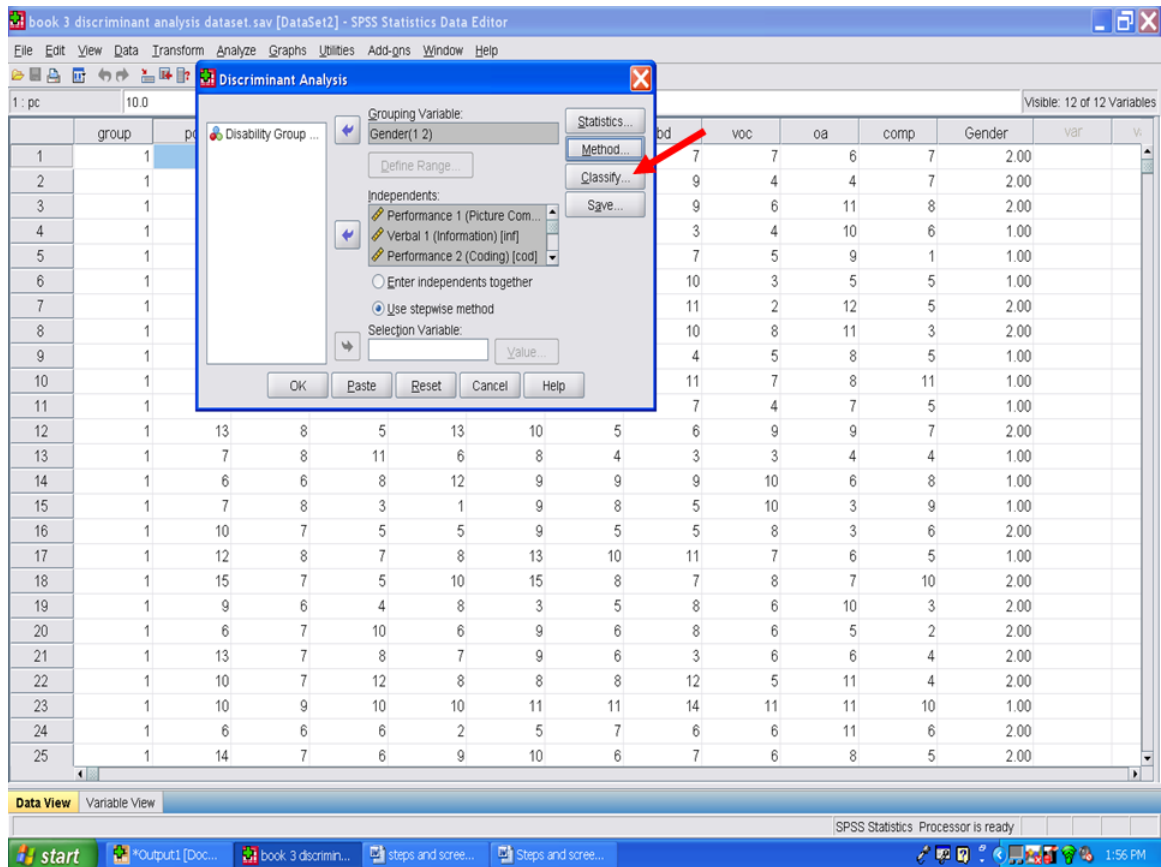
In the Method screen, we will use the default of Wilks' lambda.  
Click on Continue.

<sup>6</sup><http://cnx.org/content/m40736/latest/15.3.png/image>



Then click on Classify.

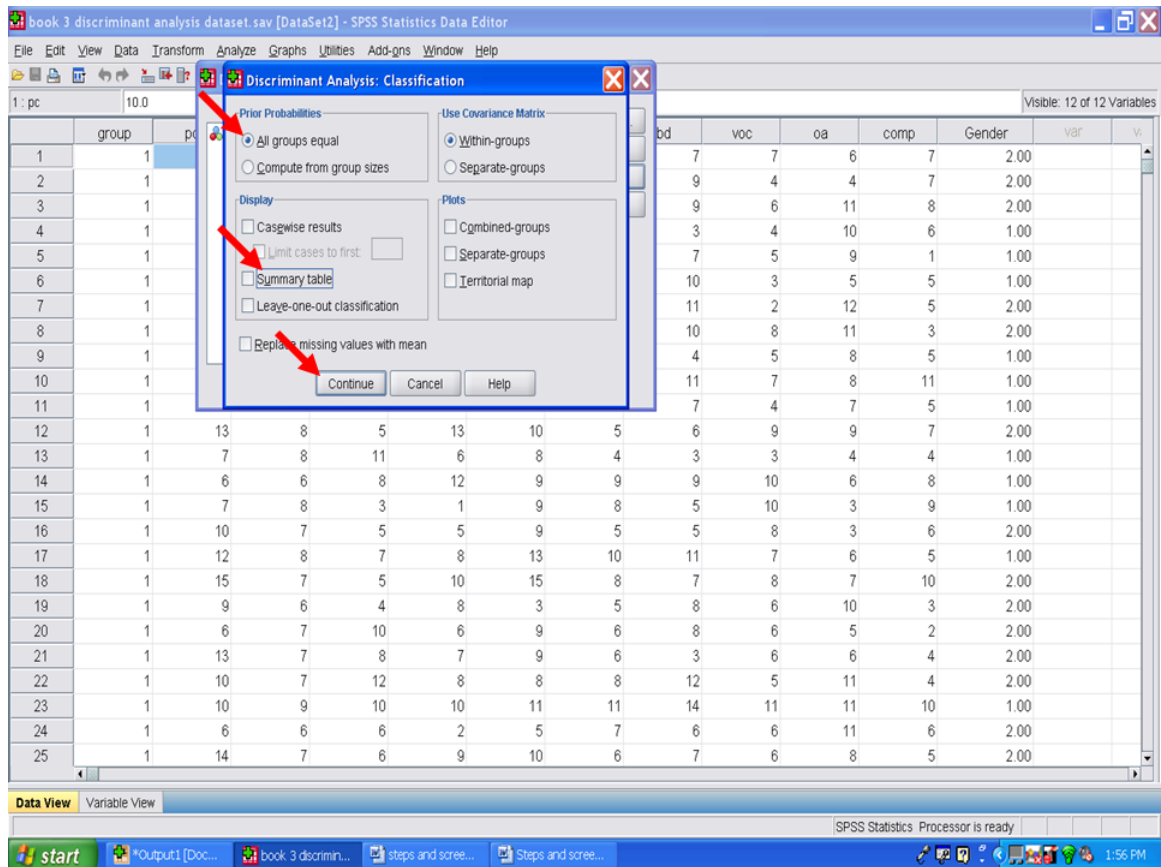
<sup>7</sup> <http://cnx.org/content/m40736/latest/15.4.png/image>



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

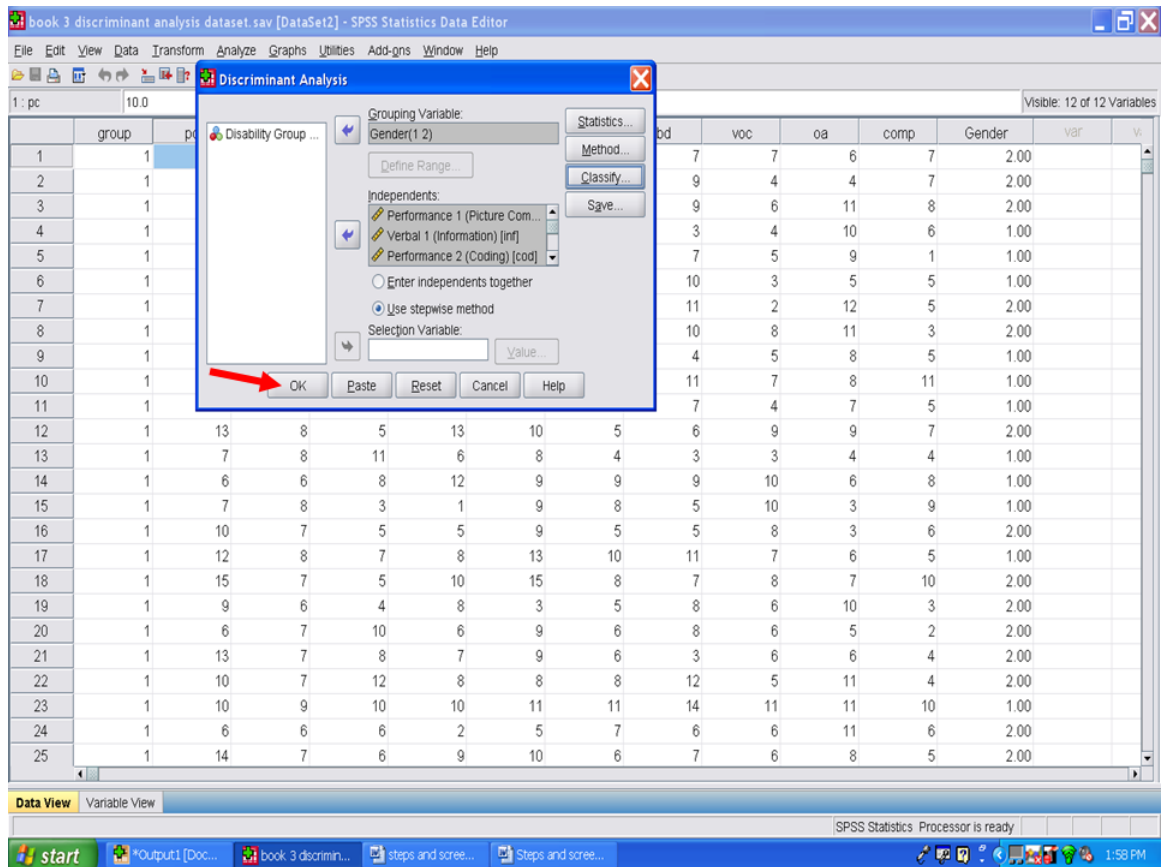
<sup>8</sup><http://cnx.org/content/m40736/latest/15.5.png/image>





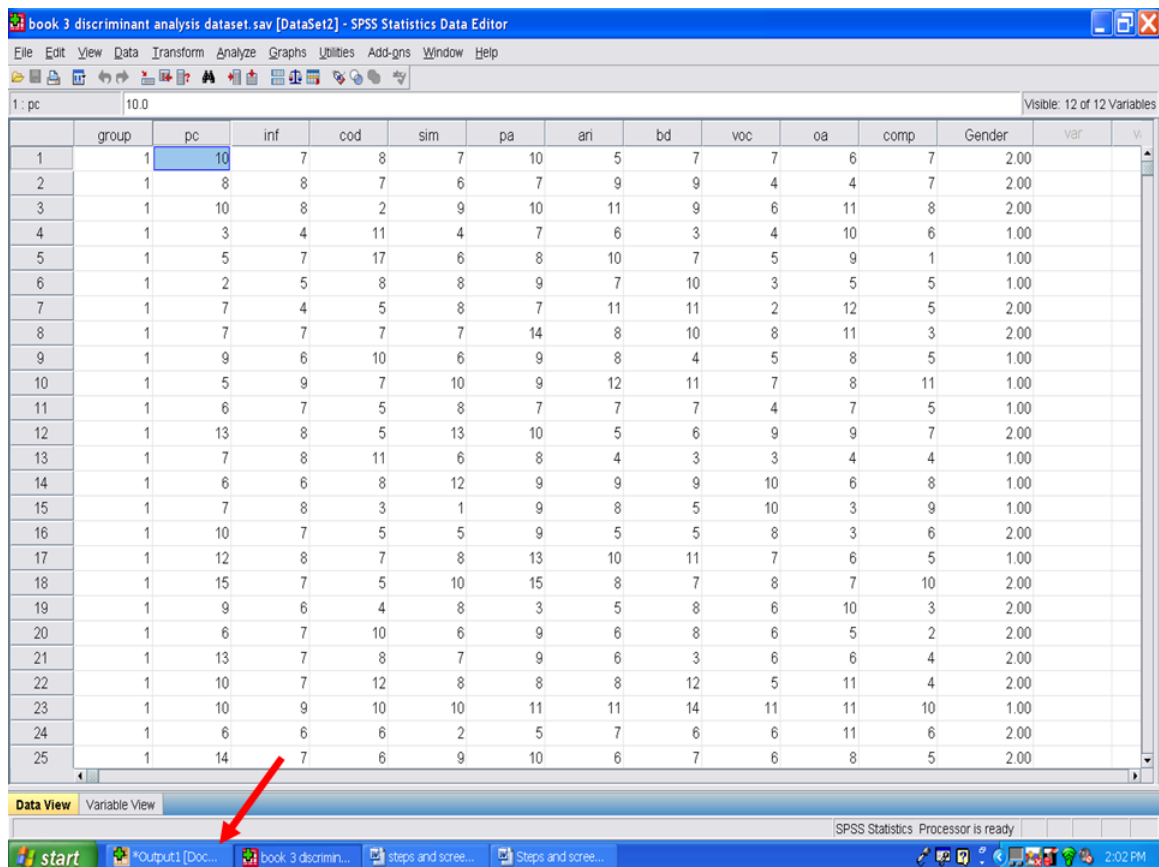
We are now ready to have SPSS calculate this procedure.  
Click on OK.

<sup>9</sup><http://cnx.org/content/m40736/latest/15.6.png/image>



After clicking on OK, SPSS should send you to your output file. If not, click on the Output icon at the bottom of your screen.

<sup>10</sup><http://cnx.org/content/m40736/latest/15.7.png/image>



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

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

1: pc 10.0 Visible: 12 of 12 Variables

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

Data View Variable View

SPSS Statistics Processor is ready

start \*Output1 [Doc... book 3 discrimin... steps and scree... Steps and scree...

2:02 PM

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.

<sup>11</sup><http://cnx.org/content/m40736/latest/15.8.png/image>

**Discriminant**

[DataSet2] C:\WINNT\profiles\shou\My Documents\John R Slate\Scholarly Works\Books\Book Two Advanced Statistics\book 3 discriminant analysis dataset

**Analysis Case Processing Summary**

Unweighted Cases	N	Percent
Valid	1130	95.6
Excluded	3	.3
Missing or out-of-range group codes		
At least one missing discriminating variable	49	4.1
Both missing or out-of-range group codes and at least one missing discriminating variable	0	.0
Total	52	4.4
Total	1182	100.0

**Group Statistics**

Gender of Persons in Study		Valid N (listwise)	
		Unweighted	Weighted
Boys	Performance 1 (Picture Completion)	571	571.000
	Verbal 1 (Information)	571	571.000
	Performance 2 (Coding)	571	571.000
	Verbal 2 (Similarities)	571	571.000
	Performance 3 (Picture Arrangement)	571	571.000
	Verbal 3 (Arithmetic)	571	571.000
	Performance 4 (Block Design)	571	571.000
	Verbal 4 (Vocabulary)	571	571.000
	Performance 5 (Object Assembly)	571	571.000
	Verbal 5 (Comprehension)	571	571.000
Girls	Performance 1 (Picture Completion)	559	559.000
	Verbal 1 (Information)	559	559.000
	Performance 2 (Coding)	559	559.000

The second table in the output tells us we have 571 boys in group one and 559 girls in group two.

<sup>12</sup><http://cnx.org/content/m40736/latest/15.9.png/image>

**\*Output1 [Document1] - SPSS Statistics Viewer**

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

Output

- Discriminant
  - Title
  - Notes
  - Active Dataset
  - Analysis Case Processing Summary
  - Group Statistics
  - Analysis 1
    - Title
    - Stepwise Statistics
      - Variables Entered/Removed
      - Variables in the Model
      - Variables Not in the Model
      - Wilks' Lambda
    - Summary of Canonical Discriminant Functions
    - Eigenvalues
    - Wilks' Lambda
    - Standardized Canonical Discriminant Function Coefficients
    - Structure Matrix
    - Functions at Group Level
    - Classification Statistics

**Discriminant**

[DataSet2] C:\WINNT\profiles\shou\My Documents\John R Slate\Scholarly Works\Books\Book Two Advanced Statistics\book 3 discriminant analysis dataset

**Analysis Case Processing Summary**

Unweighted Cases	N	Percent
Valid	1130	95.6
Excluded		
Missing or out-of-range group codes	3	.3
At least one missing discriminating variable	49	4.1
Both missing or out-of-range group codes and at least one missing discriminating variable	0	.0
Total	52	4.4
Total	1182	100.0

**Group Statistics**

Gender of Persons in Study		Valid N (listwise)	
		Unweighted	Weighted
Boys	Performance 1 (Picture Completion)	571	571.000
	Verbal 1 (Information)	571	571.000
	Performance 2 (Coding)	571	571.000
	Verbal 2 (Similarities)	571	571.000
	Performance 3 (Picture Arrangement)	571	571.000
	Verbal 3 (Arithmetic)	571	571.000
	Performance 4 (Block Design)	571	571.000
	Verbal 4 (Vocabulary)	571	571.000
	Performance 5 (Object Assembly)	571	571.000
	Verbal 5 (Comprehension)	571	571.000
Girls	Performance 1 (Picture Completion)	559	559.000
	Verbal 1 (Information)	559	559.000
	Performance 2 (Coding)	559	559.000

Information area

SPSS Statistics Processor is ready | H: 508, W: 1167 pt.

start \*Output1 [Doc... book 3 discrimin... steps and scree... Steps and scree...

2:18 PM



## Chapter 7

# Discriminant Analysis: Part III<sup>1</sup>



NOTE: This chapter is published by NCPEA Press<sup>2</sup> 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<sup>3</sup>) 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.**

---

<sup>1</sup>This content is available online at <<http://cnx.org/content/m40737/1.2/>>.

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

<sup>3</sup><http://cnx.org/content/m40737/latest/www.writingandstatisticalhelp>

**Analysis 1**

**Stepwise Statistics**

**Variables Entered/Removed<sup>a</sup>**

Step	Entered	Wilks' Lambda			Exact F			Sig.
		Statistic	df1	df2	Statistic	df1	df2	
1	Verbal 3 (Arithmetic)	.961	1	1128.000	46.270	1	1128.000	.000
2	Performance 1 (Picture Completion)	.940	2	1128.000	35.762	2	1127.000	.000
3	Performance 2 (Coding)	.919	3	1128.000	33.010	3	1126.000	.000
4	Verbal 4 (Vocabulary)	.904	4	1128.000	29.892	4	1125.000	.000
5	Verbal 1 (Information)	.894	5	1128.000	26.521	5	1124.000	.000

At each step, the variable that minimizes the overall Wilks' Lambda is entered.

a. Maximum number of steps is 20.  
 b. Minimum partial F to enter is 3.84.  
 c. Maximum partial F to remove is 2.71.  
 d. F level, tolerance, or VIN insufficient for further computation.

**Variables in the Analysis**

Step		Tolerance	F to Remove	Wilks' Lambda
1	Verbal 3 (Arithmetic)	1.000	46.270	.996
2	Verbal 3 (Arithmetic)	.866	66.451	.961
2	Performance 1 (Picture Completion)	.866	24.298	.953
3	Verbal 3 (Arithmetic)	.812	41.561	.944
3	Performance 1 (Picture Completion)	.852	30.085	.940
3	Performance 2 (Coding)	.891	25.924	.919
4	Verbal 3 (Arithmetic)	.710	19.124	.938
4	Performance 1 (Picture Completion)	.777	42.617	.922
4	Performance 2 (Coding)	.887	22.401	

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.

<sup>4</sup><http://cnx.org/content/m40737/latest/16.1.png/image>



**Variables in the Analysis**

Step		Tolerance	F to Remove	Wilks' Lambda
1	Verbal 3 (Arithmetic)	1.000	46.270	
2	Verbal 3 (Arithmetic)	.866	66.451	.996
	Performance 1 (Picture Completion)	.866	24.298	.961
3	Verbal 3 (Arithmetic)	.812	41.561	.953
	Performance 1 (Picture Completion)	.852	30.085	.944
	Performance 2 (Coding)	.891	25.924	.940
4	Verbal 3 (Arithmetic)	.710	19.124	.919
	Performance 1 (Picture Completion)	.777	42.617	.938
	Performance 2 (Coding)	.887	22.401	.922
	Verbal 4 (Vocabulary)	.699	18.959	.919
5	Verbal 3 (Arithmetic)	.657	26.455	.916
	Performance 1 (Picture Completion)	.751	32.684	.920
	Performance 2 (Coding)	.880	24.945	.914
	Verbal 4 (Vocabulary)	.525	30.298	.919
	Verbal 1 (Information)	.478	11.881	.904

**Variables Not in the Analysis**

Step		Tolerance	Min. Tolerance	F to Enter	Wilks' Lambda
0	Performance 1 (Picture Completion)	1.000	1.000	4.795	.996
	Verbal 1 (Information)	1.000	1.000	4.018	.996
	Performance 2 (Coding)	1.000	1.000	41.626	.964
	Verbal 2 (Similarities)	1.000	1.000	11.807	.990
	Performance 3 (Picture Arrangement)	1.000	1.000	5.937	.995
	Verbal 3 (Arithmetic)	1.000	1.000	46.270	.961
	Performance 4 (Block Design)	1.000	1.000	5.498	.995
	Verbal 4 (Vocabulary)	1.000	1.000	35.587	.969
	Performance 5 (Object Assembly)	1.000	1.000	.228	1.000
	Verbal 5 (Comprehension)	1.000	1.000	18.444	.984

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.

<sup>5</sup><http://cnx.org/content/m40737/latest/16.2.png/image>

**Variables in the Analysis**

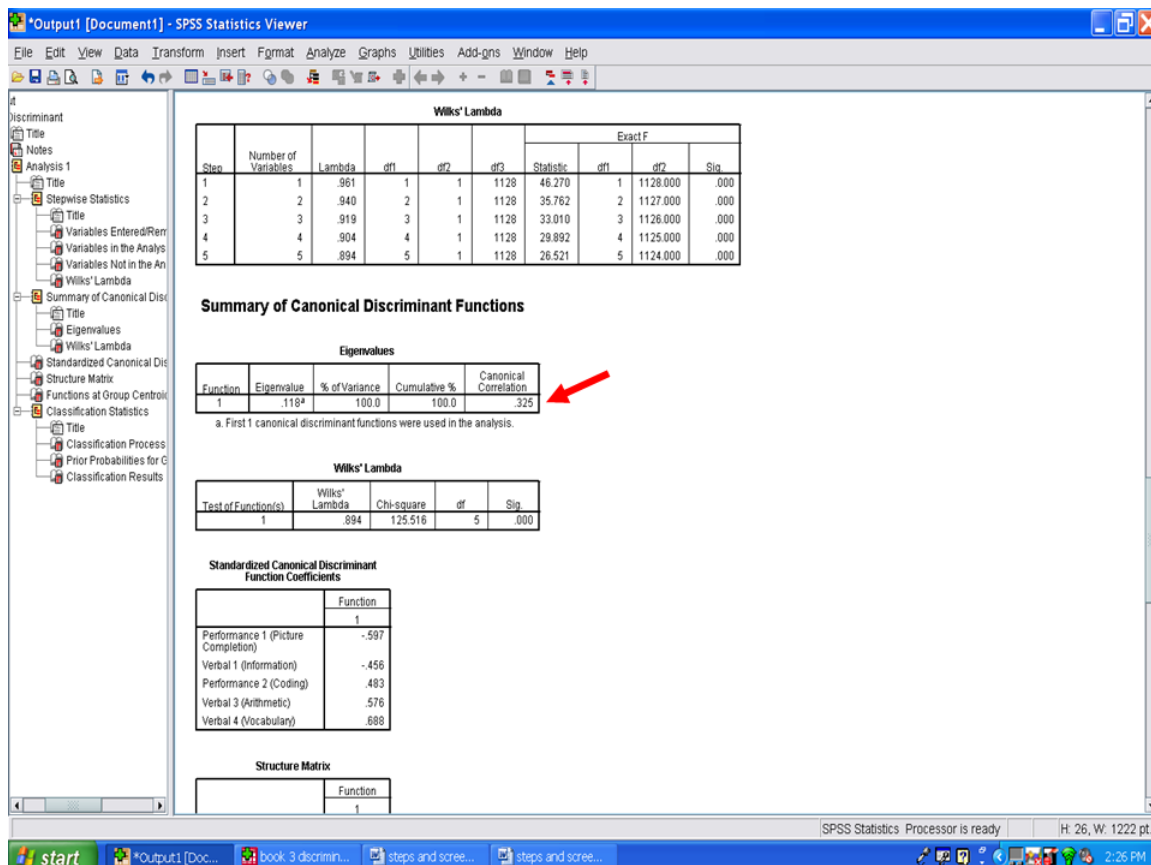
Step		Tolerance	F to Remove	Wilks' Lambda
1	Verbal 3 (Arithmetic)	1.000	46.270	
2	Verbal 3 (Arithmetic)	.866	66.451	.996
	Performance 1 (Picture Completion)	.866	24.298	.961
3	Verbal 3 (Arithmetic)	.812	41.561	.953
	Performance 1 (Picture Completion)	.852	30.085	.944
	Performance 2 (Coding)	.891	25.924	.940
4	Verbal 3 (Arithmetic)	.710	19.124	.919
	Performance 1 (Picture Completion)	.777	42.617	.938
	Performance 2 (Coding)	.887	22.401	.922
	Verbal 4 (Vocabulary)	.699	14.959	.919
5	Verbal 3 (Arithmetic)	.657	26.455	.916
	Performance 1 (Picture Completion)	.751	32.684	.920
	Performance 2 (Coding)	.880	24.945	.914
	Verbal 4 (Vocabulary)	.525	30.298	.919
	Verbal 1 (Information)	.478	11.881	.904

**Variables Not in the Analysis**

Step		Tolerance	Min. Tolerance	F to Enter	Wilks' Lambda
0	Performance 1 (Picture Completion)	1.000	1.000	4.795	.996
	Verbal 1 (Information)	1.000	1.000	4.018	.996
	Performance 2 (Coding)	1.000	1.000	41.626	.964
	Verbal 2 (Similarities)	1.000	1.000	11.807	.990
	Performance 3 (Picture Arrangement)	1.000	1.000	5.937	.995
	Verbal 3 (Arithmetic)	1.000	1.000	46.270	.961
	Performance 4 (Block Design)	1.000	1.000	5.498	.995
	Verbal 4 (Vocabulary)	1.000	1.000	35.587	.969
	Performance 5 (Object Assembly)	1.000	1.000	.228	1.000
	Verbal 5 (Comprehension)	1.000	1.000	18.444	.984

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

<sup>6</sup><http://cnx.org/content/m40737/latest/16.3.png/image>



**Wilks' Lambda**

Step	Number of Variables	Lambda	df1	df2	Statistic	df1	df2	Sig.
1	1	.961	1	1128	46.270	1	1128.000	.000
2	2	.940	2	1128	35.762	2	1127.000	.000
3	3	.919	3	1128	33.010	3	1126.000	.000
4	4	.904	4	1128	29.892	4	1125.000	.000
5	5	.894	5	1128	26.521	5	1124.000	.000

**Summary of Canonical Discriminant Functions**

**Eigenvalues**

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.116 <sup>a</sup>	100.0	100.0	.325

a. First 1 canonical discriminant functions were used in the analysis.

**Test of Function(s)**

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.894	125.516	5	.000

**Standardized Canonical Discriminant Function Coefficients**

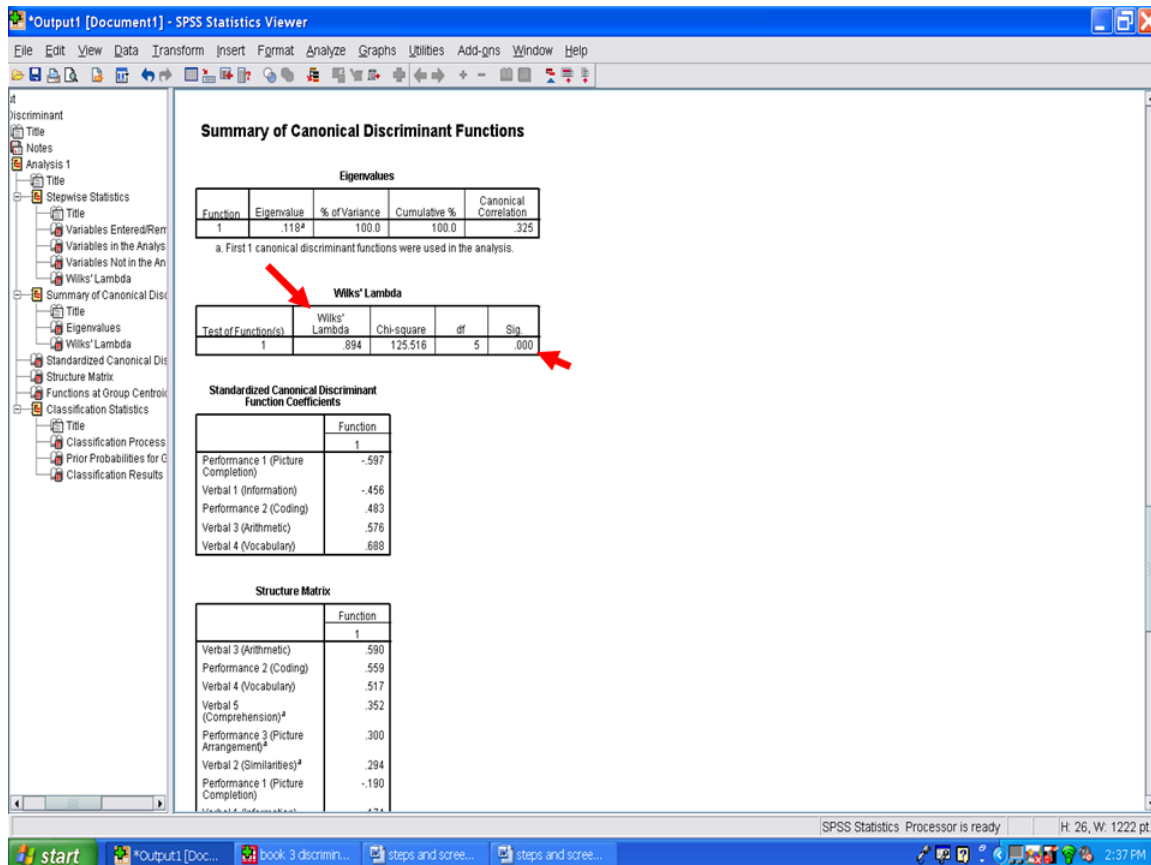
Function	1
Performance 1 (Picture Completion)	-.597
Verbal 1 (Information)	-.456
Performance 2 (Coding)	.483
Verbal 3 (Arithmetic)	.576
Verbal 4 (Vocabulary)	.688

**Structure Matrix**

Function	1
1	1

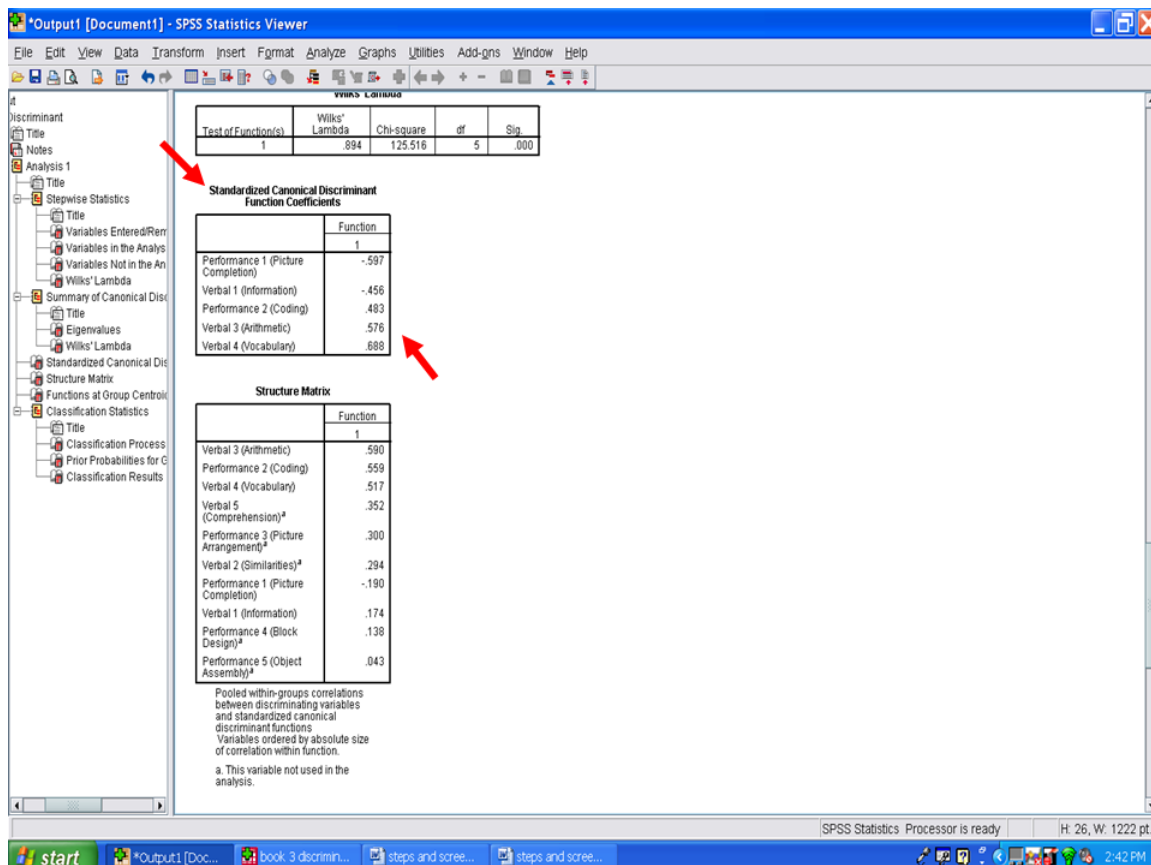
In the table below the Eigenvalues one is a table labeled Wilks' Lambda. This measure is what is used to determine whether the discriminant function is statistically significant in differentiating group membership. In our example, the Wilks' Lambda of .894 is statistically significant at the .001 level. 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.

<sup>7</sup> <http://cnx.org/content/m40737/latest/16.4.png/image>



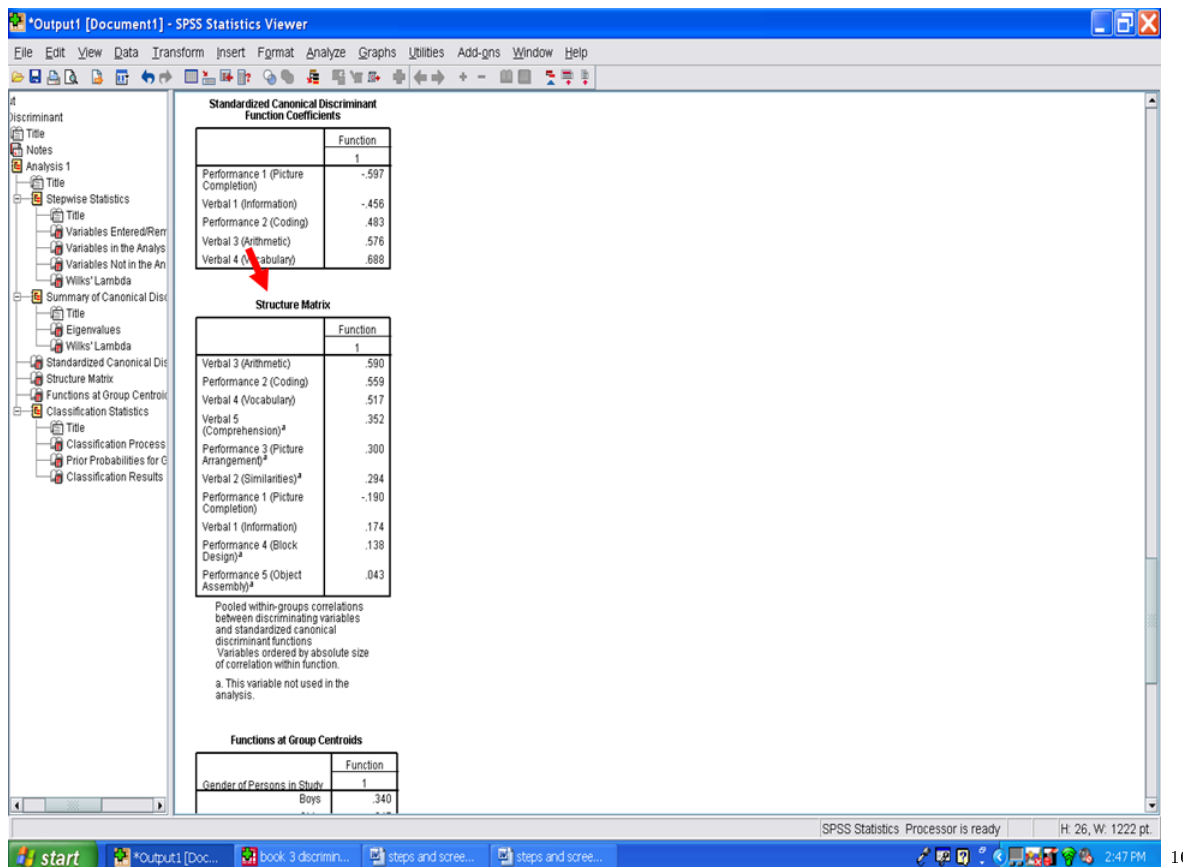
The table immediately below the Wilks' Lambda table is labeled Standardized Canonical Discriminant Function Coefficients. These values are also referred to as Standardized Discriminant Coefficients. Contained in this table are the values that depict the relative importance of each variable in differentiating group membership. These coefficients are comparable to Beta Weights in multiple regression. In this example, Verbal 4 (Vocabulary) is more important, with a coefficient of .688, than is Performance 2 (Coding), with a coefficient of .483, in differentiating boys from girls. These coefficients should be provided in your Results section.

<sup>8</sup><http://cnx.org/content/m40737/latest/16.5.png/image>



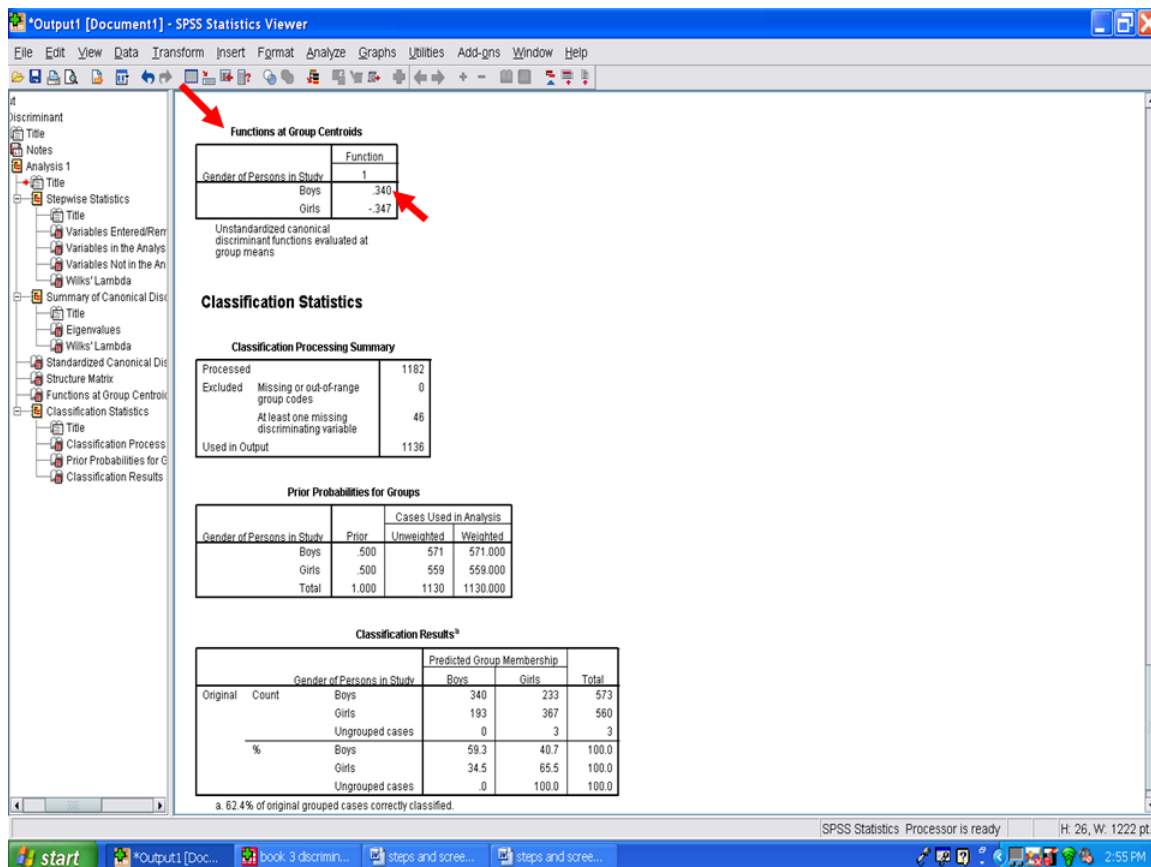
Next, we examine the table underneath the Standardized Canonical Discriminant Function Coefficients, the one labeled as Structure Matrix. Each of these values depicts the relationship of that variable with the discriminant function scores. Each value is analogous to a correlation coefficient. In the 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).

<sup>9</sup><http://cnx.org/content/m40737/latest/16.6.png/image>



Under this table is a table labeled Functions at Group Centroids. These values are z-scores, with a  $M$  of 0 and a  $SD$  of 1. The group centroid indicates the most typical location of any person from a particular group. For this example, the group centroid for boys is .34 and the group centroid for girls is -.347. Using these values we can compare how far apart boys and girls are differentiated by the statistically significant discriminant function in our example. The further apart these two means or centroids are, the better the separation or differentiation between the two groups. The centroids should be reported in the Results section.

<sup>10</sup><http://cnx.org/content/m40737/latest/16.7.png/image>



You have now successfully conducted a canonical discriminant analysis.

<sup>11</sup> <http://cnx.org/content/m40737/latest/16.8.png/image>





## Chapter 8

# Multiple Regression: Assumptions<sup>1</sup>



NOTE: This chapter is published by NCPEA Press<sup>2</sup> 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<sup>3</sup>) 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/><sup>4</sup>; to the *Electronic Statistics Textbook* (2011)

---

<sup>1</sup>This content is available online at <<http://cnx.org/content/m40738/1.2/>>.

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

<sup>3</sup><http://cnx.org/content/m40738/latest/www.writingandstatisticalhelp>

<sup>4</sup><http://davidmlane.com/hyperstat/>

at <http://www.statsoft.com/textbook/><sup>5</sup> ; 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](http://www.amazon.com/Discovering-Statistics-Introducing-Statistical-Method/dp/1847879071/ref=sr_1_1?s=books&ie=UTF8&qid=1304967862&sr=1-1)<sup>6</sup>

Research questions for which a multiple regression analysis is appropriate involve determining variables that predict a continuous variable. For example, if you want to predict the life expectancy of individuals and you have archival data available (e.g., health history, gender), then a multiple regression analysis procedure could be utilized. Such a procedure could identify specific variables that are predictive of a long or of a short life expectancy. As such, interventions could be developed and targeted toward the variables that were statistically significant predictors. Other sample research questions for which a multiple regression analysis might be appropriate: (a) What factors predict high scores on a scholastic aptitude measure?; and (b) What factors are predictive of high scores on a life satisfaction scale?

For the purposes of this chapter, our research question is: "What scholastic variables predict students' Full Scale IQ?"

Have your data set open in SPSS. In this dataset, we will determine what variables, if any, are predictive of students' Wechsler Full Scale IQ 3 (wfsiq). The 10 variables that we will use are: Picture Completion (pc), Information (inf), Coding (cod), Similarities (sim), Picture Arrangement (pa), Arithmetic (ari), Block Design (bd), Vocabulary (voc), Object Assembly (oa), and Comprehension (comp).

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	group	Numeric	1	0	Reading Group Membership	{1, Excelle...	None	8	Right	Nominal
2	age	Numeric	2	0	Age of Students in Study	None	None	8	Right	Ordinal
3	recodage	Numeric	8	2	Recoded Age Into Categories	{1.00, Age...	None	8	Right	Nominal
4	wfsiq	Numeric	3	0	Wechsler Full Scale IQ 3	None	None	8	Right	Scale
5	wvsiq	Numeric	3	0	Verbal IQ (Wechsler Verbal Intelligence 3)	None	None	8	Right	Scale
6	wpiq	Numeric	3	0	Performance IQ (Wechsler Performance Intell...	None	None	8	Right	Scale
7	pc	Numeric	2	0	Performance 1 (Picture Completion)	None	None	8	Right	Scale
8	inf	Numeric	2	0	Verbal 1 (Information)	None	None	8	Right	Scale
9	cod	Numeric	2	0	Performance 2 (Coding)	None	None	8	Right	Scale
10	sim	Numeric	2	0	Verbal 2 (Similarities)	None	None	8	Right	Scale
11	pa	Numeric	2	0	Performance 3 (Picture Arrangement)	None	None	8	Right	Scale
12	ari	Numeric	2	0	Verbal 3 (Arithmetic)	None	None	8	Right	Scale
13	bd	Numeric	2	0	Performance 4 (Block Design)	None	None	8	Right	Scale
14	voc	Numeric	2	0	Verbal 4 (Vocabulary)	None	None	8	Right	Scale
15	oa	Numeric	2	0	Performance 5 (Object Assembly)	None	None	8	Right	Scale
16	comp	Numeric	2	0	Verbal 5 (Comprehension)	None	None	8	Right	Scale
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										

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

<sup>5</sup><http://www.statsoft.com/textbook/>

<sup>6</sup>[http://www.amazon.com/Discovering-Statistics-Introducing-Statistical-Method/dp/1847879071/ref=sr\\_1\\_1?s=books&ie=UTF8&qid=1304967862&sr=1-1](http://www.amazon.com/Discovering-Statistics-Introducing-Statistical-Method/dp/1847879071/ref=sr_1_1?s=books&ie=UTF8&qid=1304967862&sr=1-1)

1

<sup>7</sup><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](http://www.statistics.com/index.php?page=glossary&term_id=356)<sup>8</sup> and <http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb><sup>9</sup>

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](http://www.statistics.com/index.php?page=glossary&term_id=326)<sup>10</sup> and <http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb><sup>11</sup>

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**

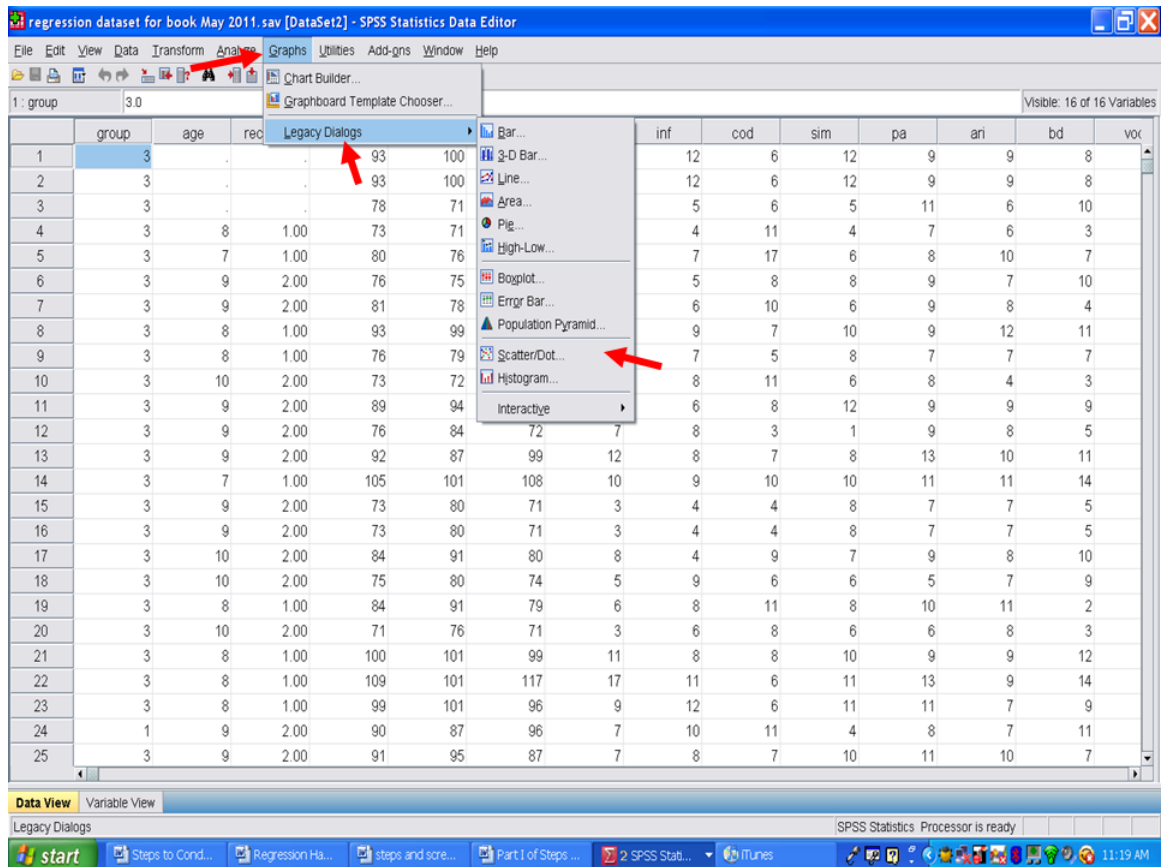
---

<sup>8</sup>[http://www.statistics.com/index.php?page=glossary&term\\_id=356](http://www.statistics.com/index.php?page=glossary&term_id=356)

<sup>9</sup><http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb>

<sup>10</sup>[http://www.statistics.com/index.php?page=glossary&term\\_id=326](http://www.statistics.com/index.php?page=glossary&term_id=326)

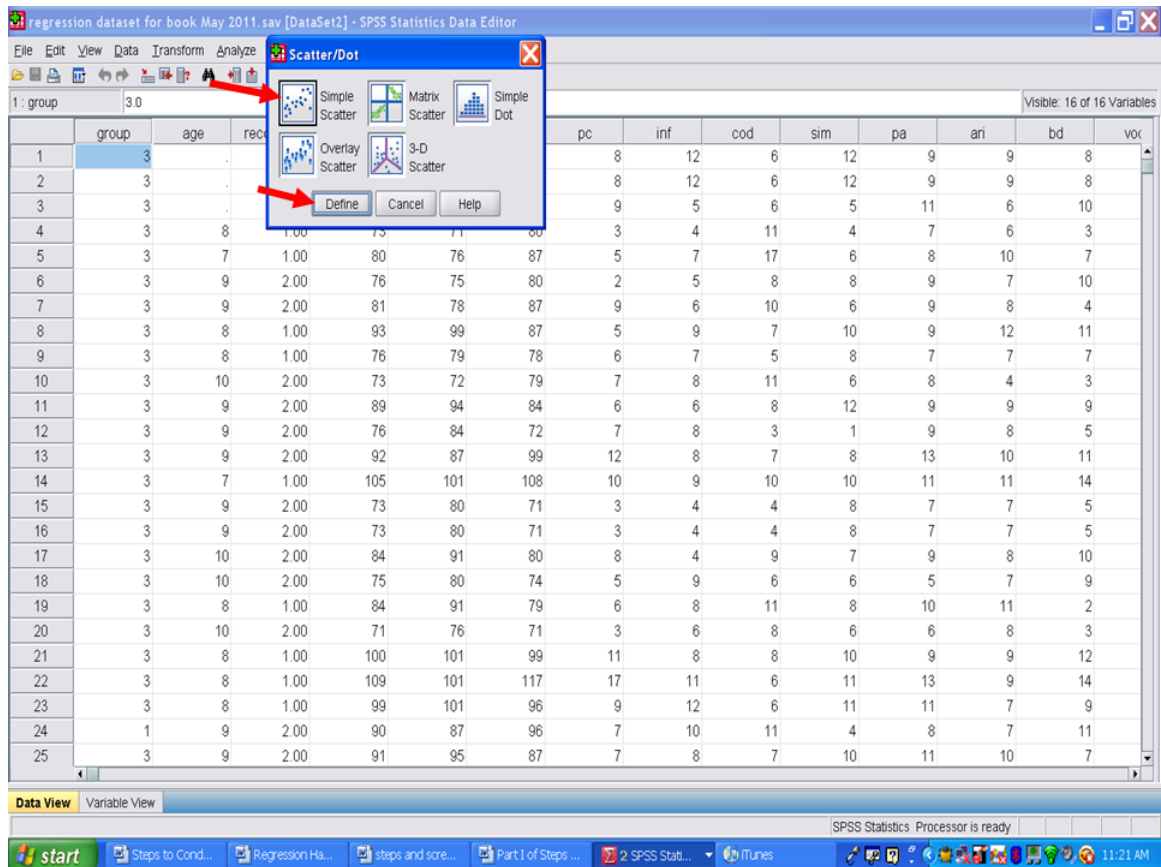
<sup>11</sup><http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb>



After clicking on Scatter/Dot, the following screen will appear. The Simple Scatter icon should be highlighted. If not, click on it.

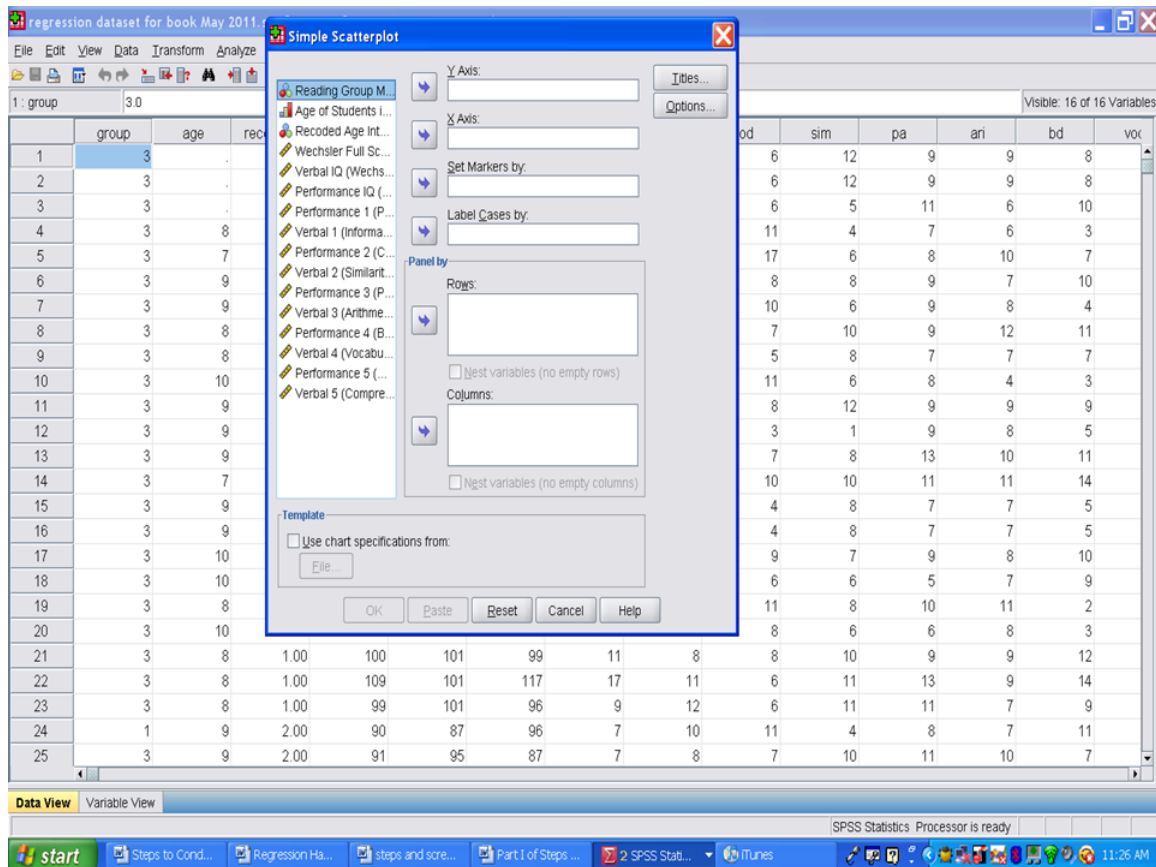
✓ Click on Define

<sup>12</sup><http://cnx.org/content/m40738/latest/17.2.png/image>



The following screen should now be present.

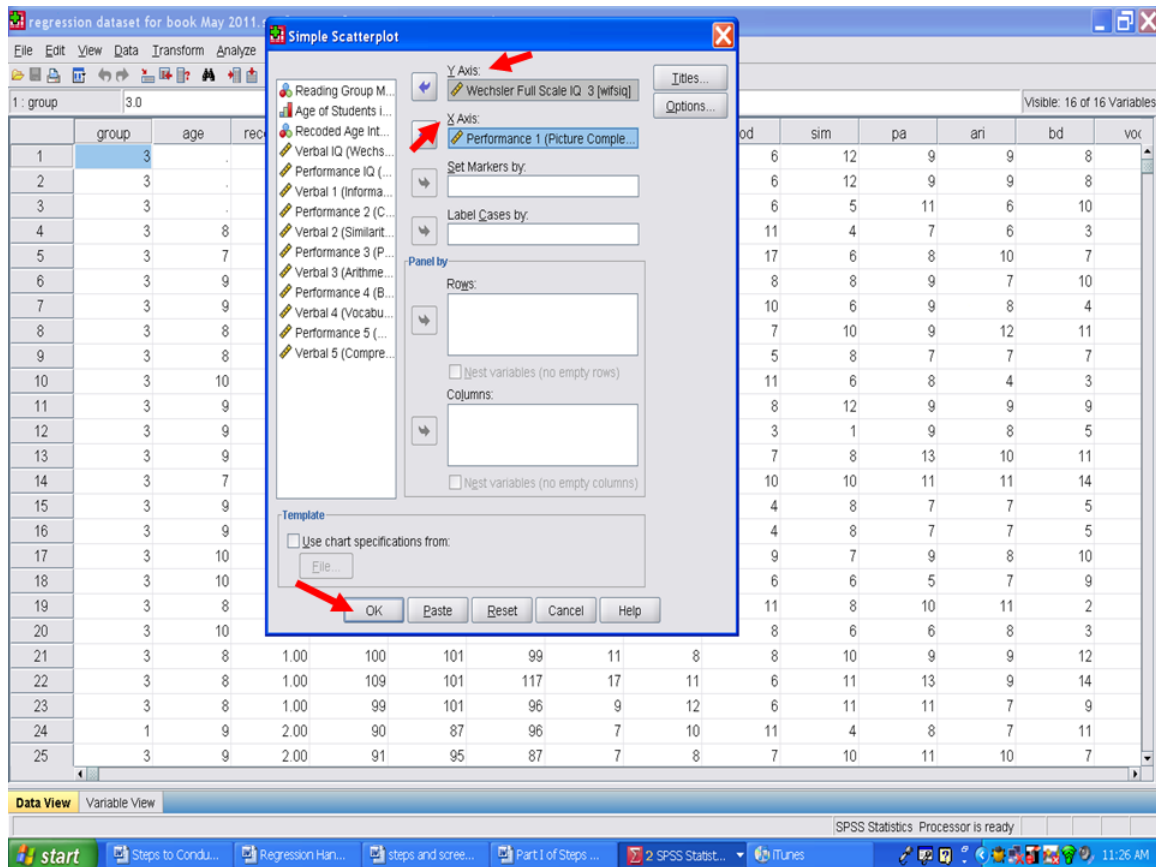
<sup>13</sup><http://cnx.org/content/m40738/latest/17.3.png/image>



✓ 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.

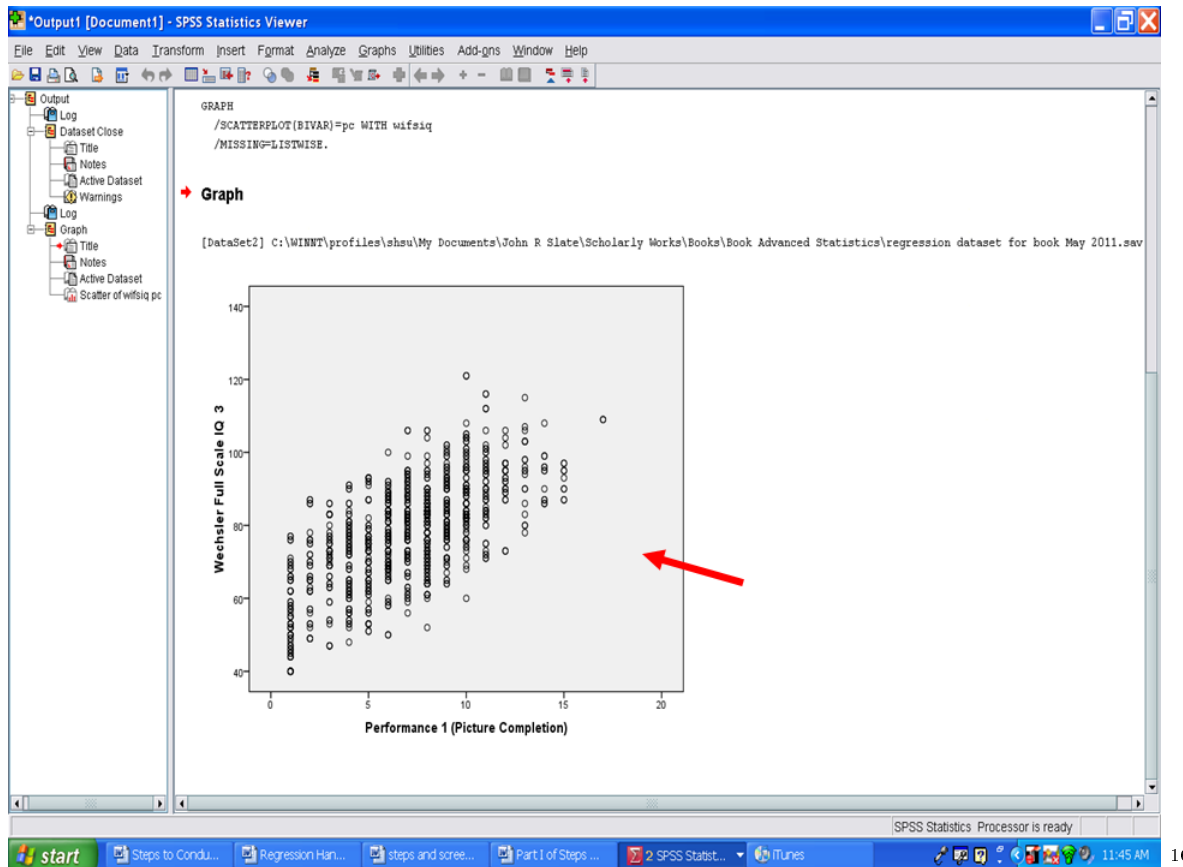
✓ Once you have a variable in each of the first two boxes, click on the OK tab on the bottom left hand corner of the screen.

<sup>14</sup><http://cnx.org/content/m40738/latest/17.4.png/image>



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

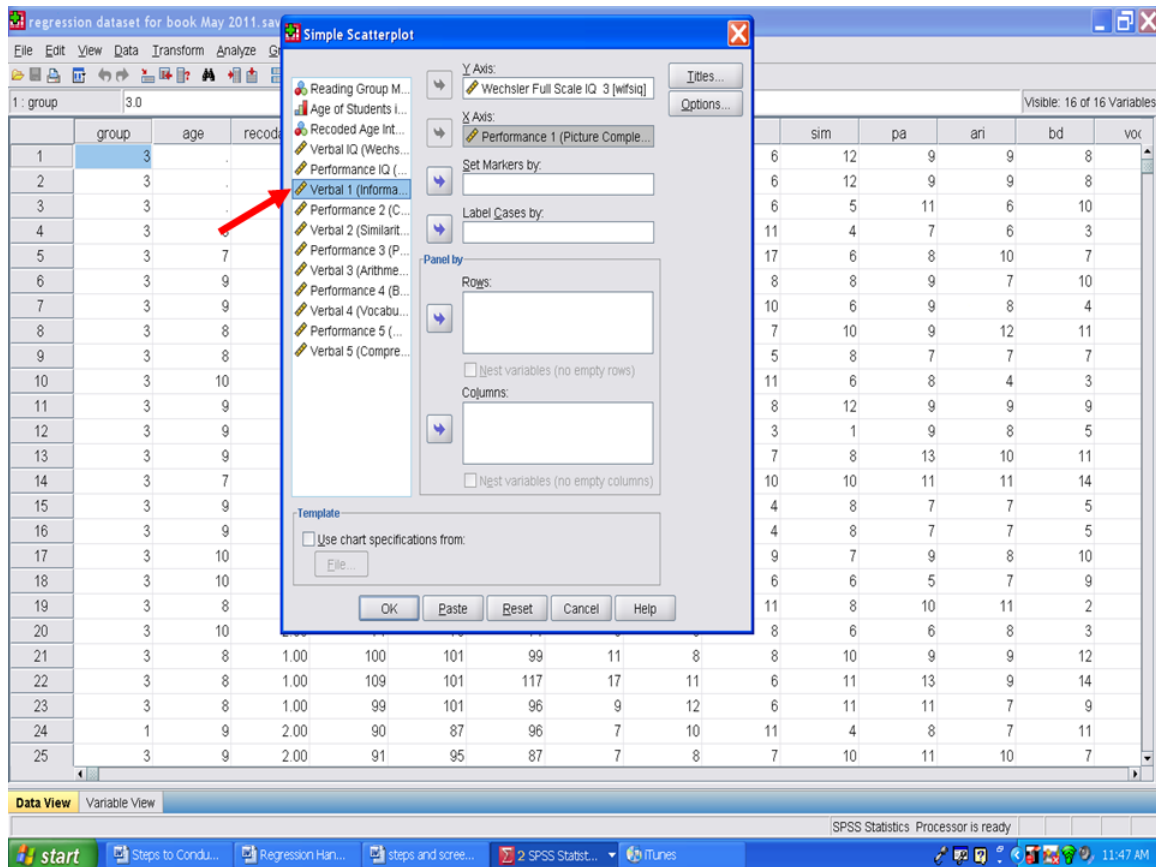
<sup>15</sup><http://cnx.org/content/m40738/latest/17.5.png/image>



You will need to repeat this process, for this example, nine more times. Leave the dependent variable of Wechsler Full Scale IQ 3 in the Y Axis box and replace the variable in the X Axis box with the next variable (i.e., Verbal 1). Then click on OK.

<sup>16</sup><http://cnx.org/content/m40738/latest/17.6.png/image>





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.

<sup>17</sup><http://cnx.org/content/m40738/latest/17.7.png/image>



## Chapter 9

# Multiple Regression: Part I<sup>1</sup>



NOTE: This chapter is published by NCPEA Press<sup>2</sup> 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<sup>3</sup>) to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

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

### From the data screen, click on

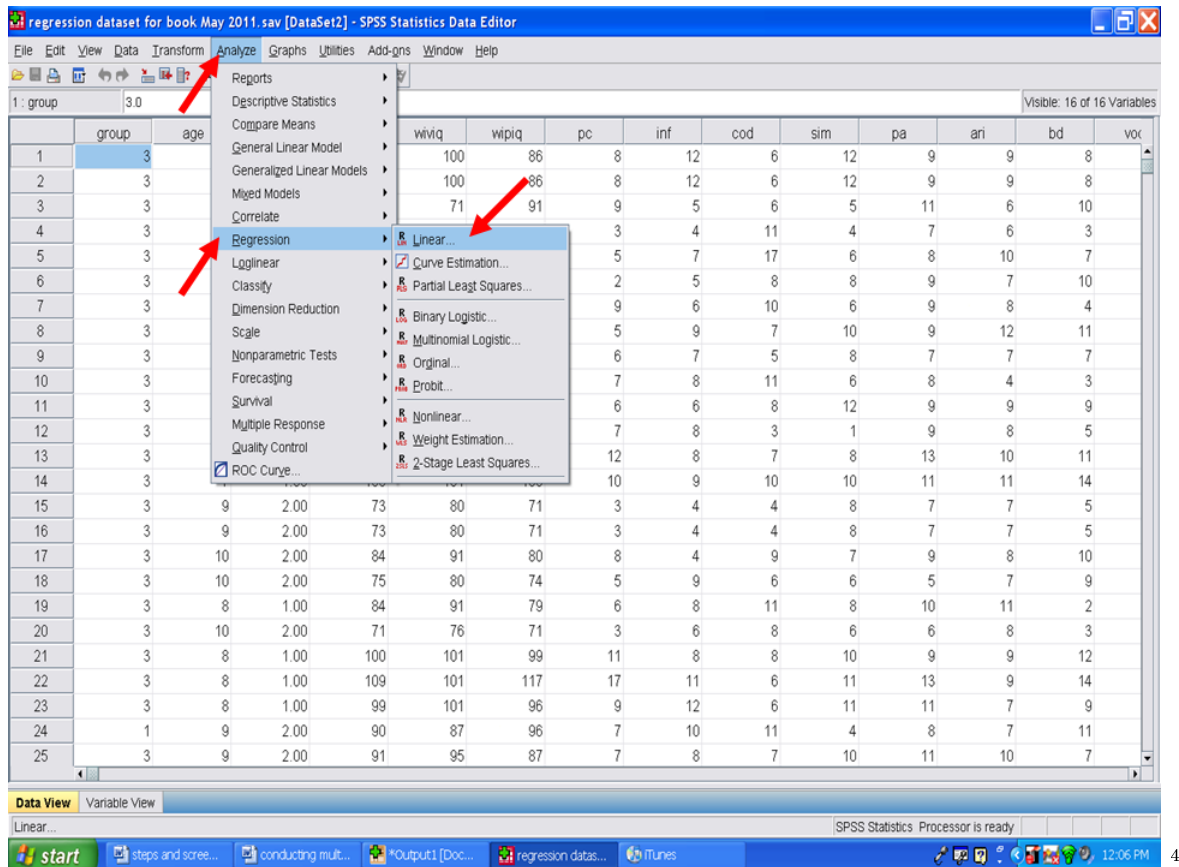
- ✓ Analyze
- ✓ Regression
- ✓ Linear

---

<sup>1</sup>This content is available online at <<http://cnx.org/content/m40739/1.2/>>.

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

<sup>3</sup><http://cnx.org/content/m40739/latest/www.writingandstatisticalhelp>



Your screen should now look like the one below. We will send the dependent variable, the one we are trying to predict, to the Dependent box. Click on Wechsler Full Scale IQ 3 and send it to the Dependent box.

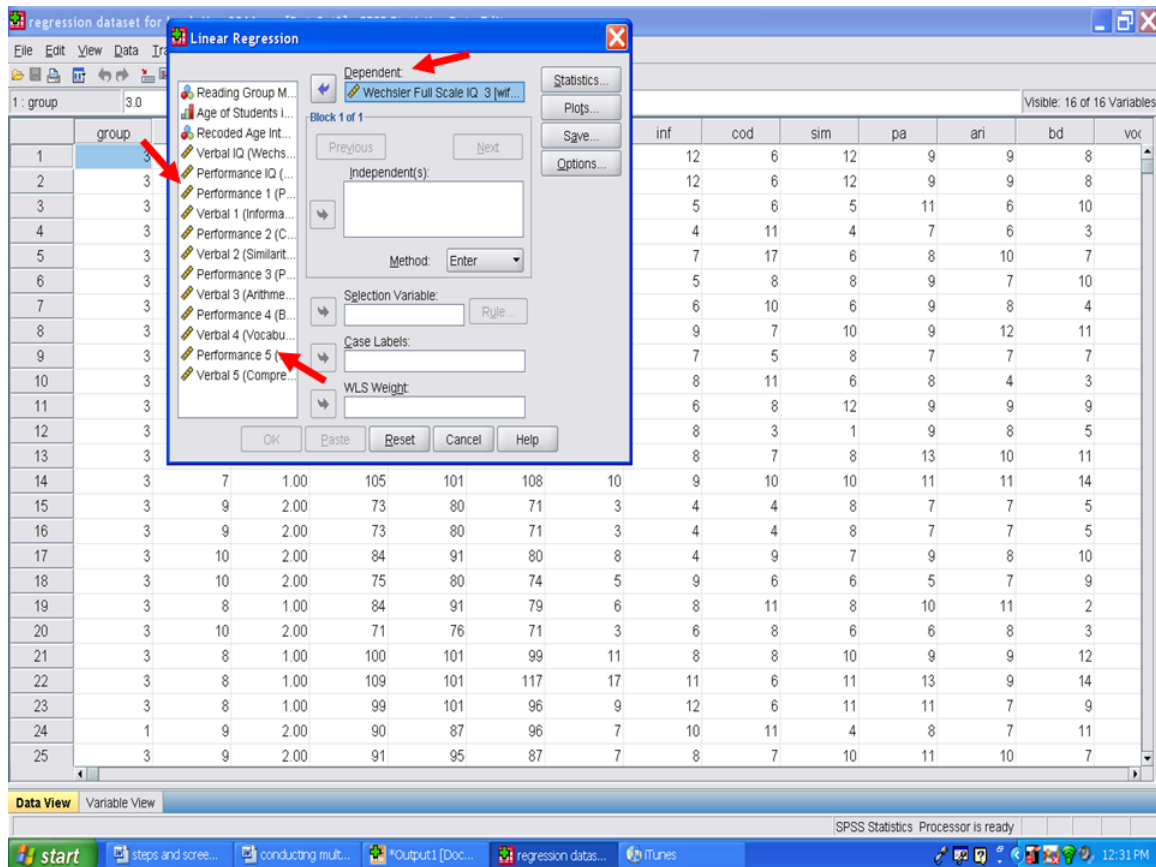
<sup>4</sup><http://cnx.org/content/m40739/latest/18.1.png/image>

The screenshot shows the SPSS Linear Regression dialog box open over a Data View window. The dialog box has 'Dependent' set to 'Reading Group M...' and 'Independent(s)' is empty. The 'Method' is set to 'Enter'. The 'Data View' window shows a table with 16 variables: group, inf, cod, sim, pa, ari, bd, vox, and 11 unlabeled variables. The 'group' variable is highlighted in the first column.

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

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

<sup>5</sup><http://cnx.org/content/m40739/latest/18.2.png/image>



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.

<sup>6</sup><http://cnx.org/content/m40739/latest/18.3.png/image>

**Linear Regression**

Dependent: Wechsler Full Scale IQ 3 [wif...]

Block 1 of 1

Independent(s):

- Performance 1 (Picture Co...)
- Verbal 1 (Information) [inf]
- Performance 2 (Coding) [c...

Method: Enter

Selection Variable: Rule

Case Labels:

WLS Weight:

OK Paste Reset Cancel Help

Statistics... Plots... Save... Options...

Visible: 16 of 16 Variables

inf	cod	sim	pa	ari	bd	vox
12	6	12	9	9	8	
12	6	12	9	9	8	
5	6	5	11	6	10	
4	11	4	7	6	3	
7	17	6	8	10	7	
5	8	8	9	7	10	
6	10	6	9	8	4	
9	7	10	9	12	11	
7	5	8	7	7	7	
8	11	6	8	4	3	
6	8	12	9	9	9	
8	3	1	9	8	5	
8	7	8	13	10	11	
9	10	10	11	11	14	
4	4	8	7	7	5	
4	4	8	7	7	5	
4	9	7	9	8	10	
9	6	6	5	7	9	
6	8	6	6	8	3	
8	8	10	9	9	12	
11	6	11	13	9	14	
12	6	11	11	7	9	
10	11	4	8	7	11	
8	7	10	11	10	7	

Data View Variable View

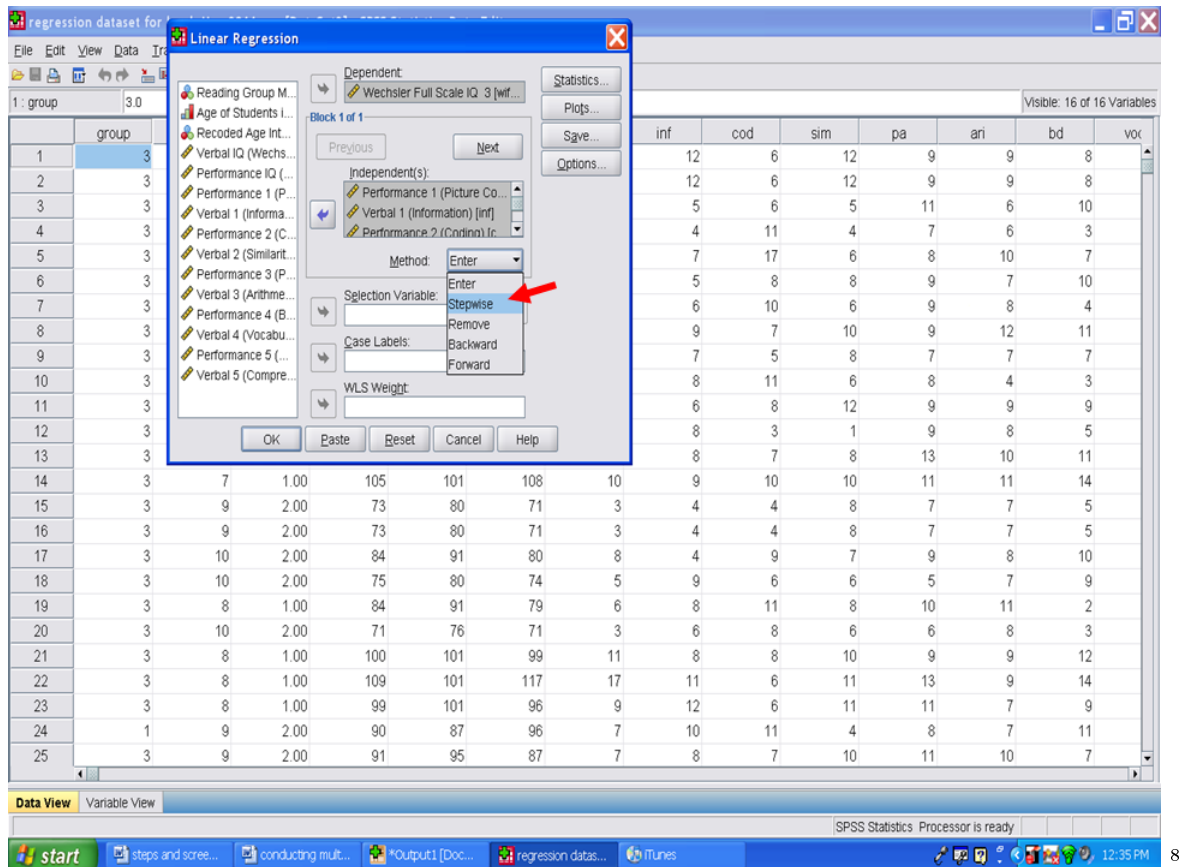
SPSS Statistics Processor is ready

start steps and scree... conducting mult... \*Output1 [Doc... regression datas... iTunes

12:34 PM

Click on Stepwise.

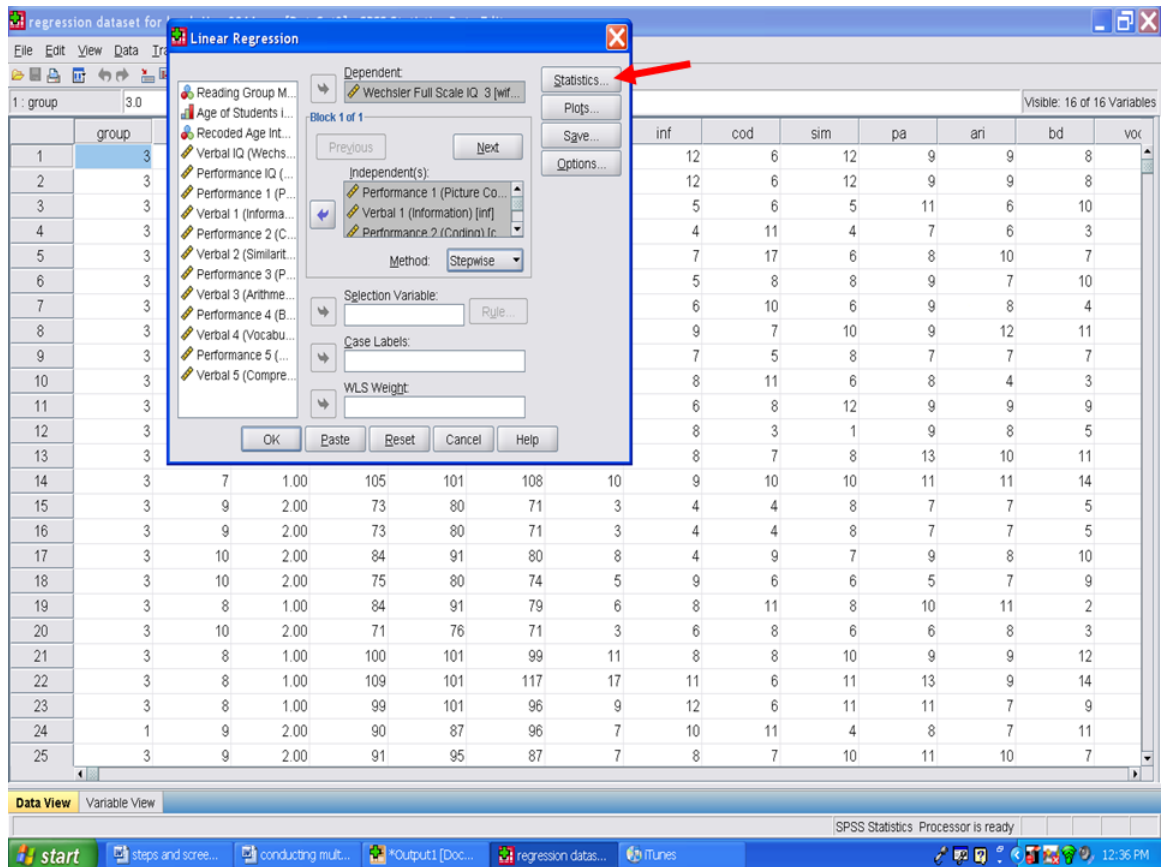
<sup>7</sup> <http://cnx.org/content/m40739/latest/18.4.png/image>



After clicking on Stepwise, your screen will look like the one below.  
Click on Statistics

<sup>8</sup><http://cnx.org/content/m40739/latest/18.5.png/image>

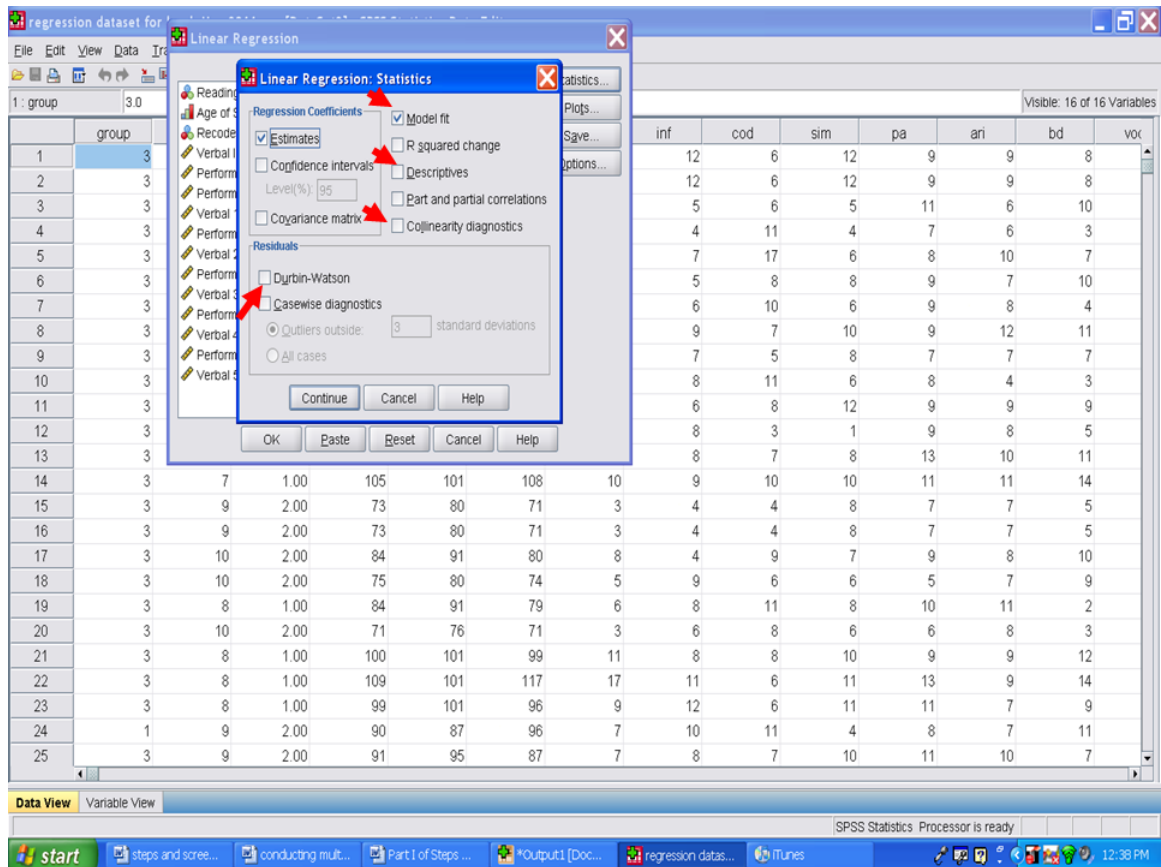




After clicking on Statistics, you will now

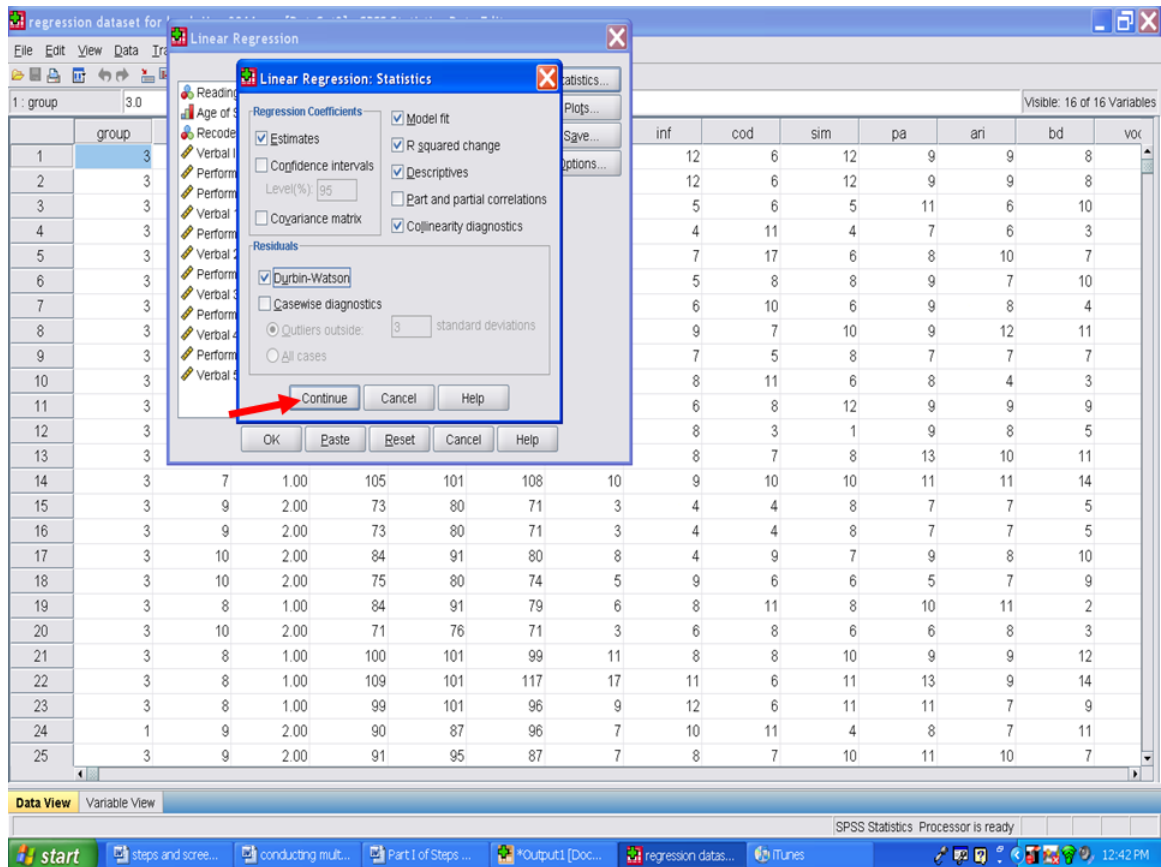
- ✓ Click on Estimates (default already checked)
- ✓ Model fit (default already checked)
- ✓ R squared change
- ✓ Descriptives
- ✓ Collinearity Diagnostics
- ✓ Durbin-Watson

<sup>9</sup><http://cnx.org/content/m40739/latest/18.6.png/image>



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

<sup>10</sup><http://cnx.org/content/m40739/latest/18.7.png/image>



Next click on

✓ OK

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

<sup>11</sup><http://cnx.org/content/m40739/latest/18.8.png/image>

The screenshot shows the SPSS Linear Regression dialog box overlaid on the Data View window. The dialog box is titled "Linear Regression" and has a "Dependent:" field containing "Wechsler Full Scale IQ 3 [wfr...". The "Independent(s):" field contains a list of variables: "Performance 1 (Picture Co...", "Verbal 1 (Information) [inf]", and "Performance 2 (Cognitive) [p...". The "Method:" dropdown is set to "Stepwise". The "Selection Variable:" field is empty. The "Case Labels:" field is empty. The "WLS Weight:" field is empty. The "OK" button is highlighted with a red arrow. The Data View window shows a table with 16 variables: inf, cod, sim, pa, ari, bd, vox, and 10 unnamed variables. The first 10 rows of data are visible.

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

<sup>12</sup><http://cnx.org/content/m40739/latest/18.89.png/image>

## Chapter 10

# Multiple Regression: Part II<sup>1</sup>



NOTE: This chapter is published by NCPEA Press<sup>2</sup> 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<sup>3</sup>) 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.

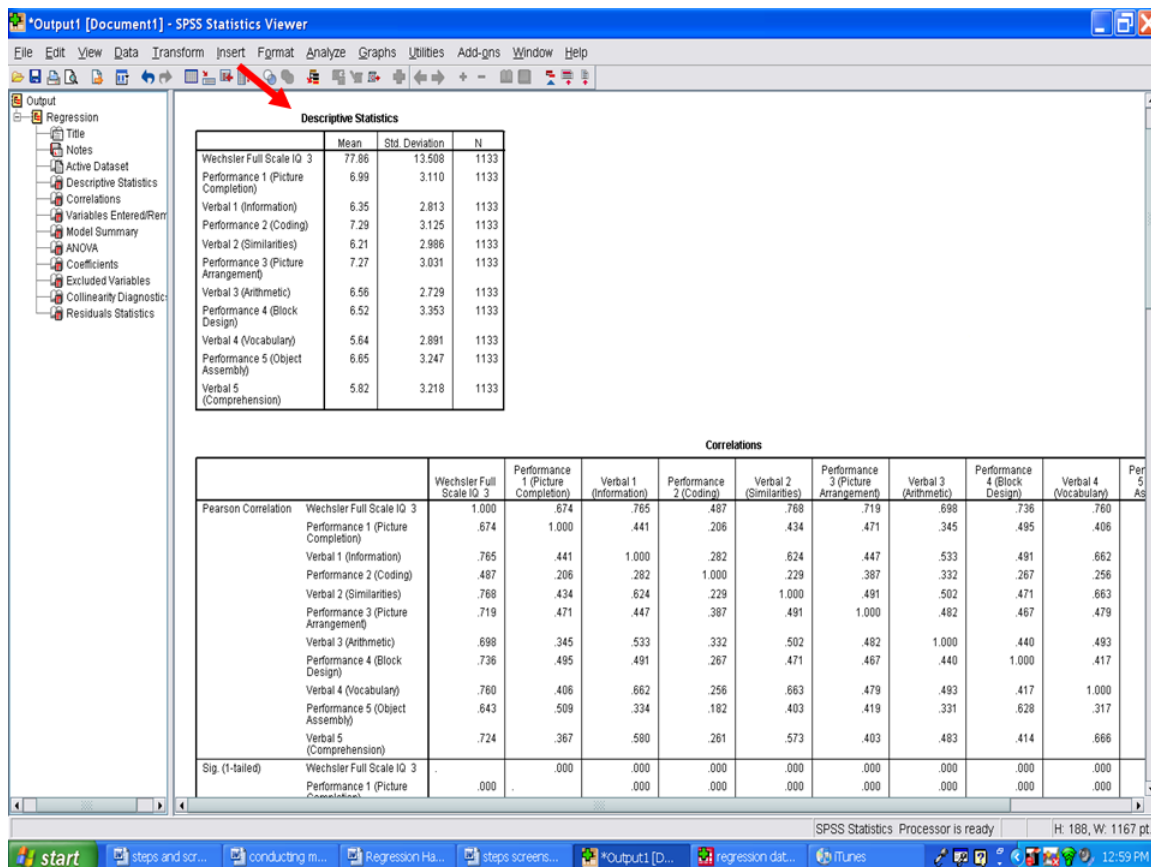
**The first table in your SPSS output file should be a Descriptive Statistics table. You will see a column for  $M$ ,  $SD$ , and  $n$ . This information should be used in your Results section.**

---

<sup>1</sup>This content is available online at <<http://cnx.org/content/m40742/1.2/>>.

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

<sup>3</sup><http://cnx.org/content/m40742/latest/www.writingandstatisticalhelp>



Underneath the Descriptive Statistics table is a table labeled Correlations. This table reflects the Pearson  $r$ s for each independent variable with the dependent variable, as well as the interrelationships among all of the variables.

Of the relationships of the independent variable with the Wechsler Full Scale IQ 3 (the dependent variable in this example), Verbal 2 (Similarities) has the highest correlation, .768. In a stepwise regression procedure, this variable should be the first statistically significant predictor.

<sup>4</sup><http://cnx.org/content/m40742/latest/19.1.png/image>

\*Output1 [Document1] - SPSS Statistics Viewer

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

Out Verbal 5 (Comprehension) 5.82 3.218 1133

**Correlations**

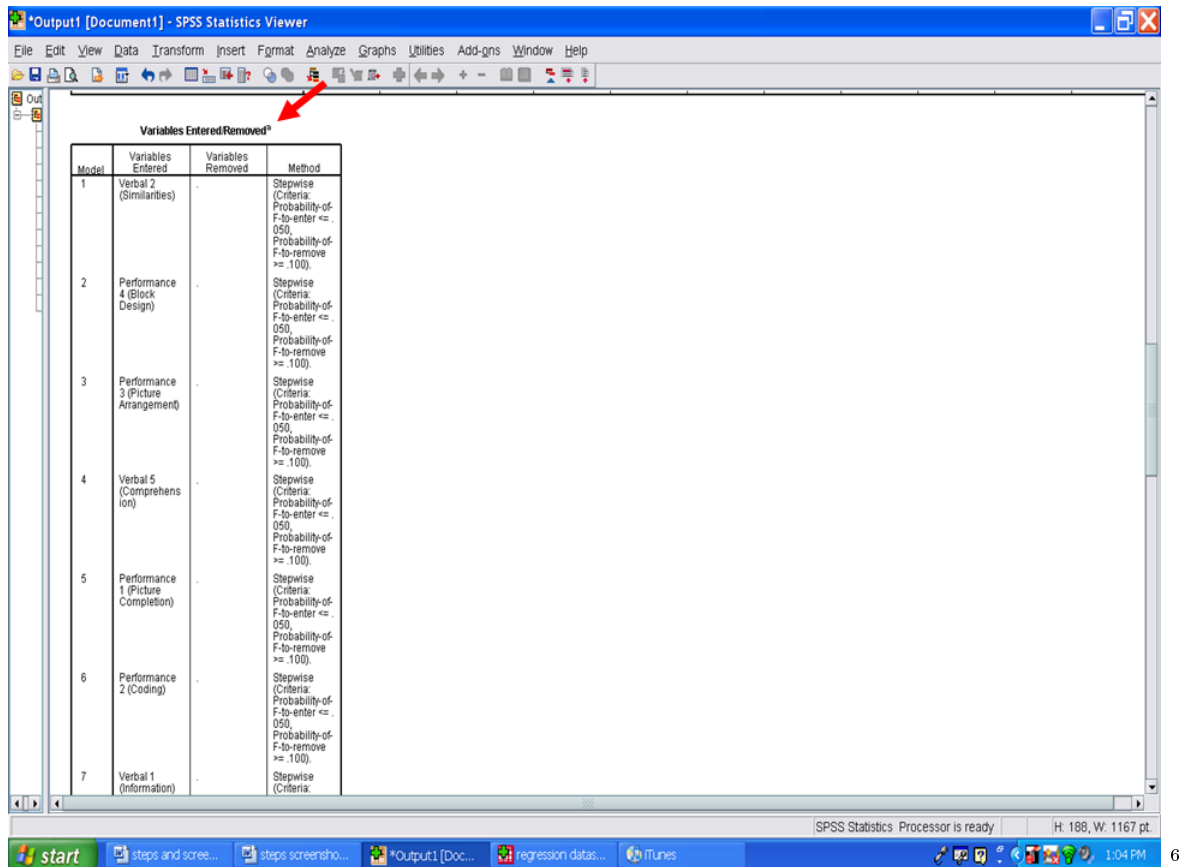
		Wechsler Full Scale IQ 3	Performance 1 (Picture Completion)	Verbal 1 (Information)	Performance 2 (Coding)	Verbal 2 (Similarities)	Performance 3 (Picture Arrangement)	Verbal 3 (Arithmetic)	Performance 4 (Block Design)	Verbal 4 (Vocabulary)	Performance 5 (Object Assembly)	Verbal 5 (Comprehension)
Pearson Correlation	Wechsler Full Scale IQ 3	1.000	.674	.765	.487	.768	.719	.698	.736	.760	.643	.724
	Performance 1 (Picture Completion)	.674	1.000	.441	.206	.434	.471	.345	.495	.406	.509	.367
	Verbal 1 (Information)	.765	.441	1.000	.262	.624	.447	.533	.491	.662	.334	.580
	Performance 2 (Coding)	.487	.206	.262	1.000	.229	.387	.332	.267	.256	.182	.261
	Verbal 2 (Similarities)	.768	.434	.624	.229	1.000	.491	.502	.471	.663	.403	.573
	Performance 3 (Picture Arrangement)	.719	.471	.447	.387	.491	1.000	.482	.467	.479	.419	.403
	Verbal 3 (Arithmetic)	.698	.345	.533	.332	.502	.482	1.000	.440	.493	.331	.483
	Performance 4 (Block Design)	.736	.495	.491	.267	.471	.467	.440	1.000	.417	.628	.414
	Verbal 4 (Vocabulary)	.760	.406	.662	.256	.663	.479	.493	.417	1.000	.317	.666
	Performance 5 (Object Assembly)	.643	.509	.334	.182	.403	.419	.331	.628	.317	1.000	.312
	Verbal 5 (Comprehension)	.724	.367	.580	.261	.573	.403	.483	.414	.666	.312	1.000
Sig. (1-tailed)	Wechsler Full Scale IQ 3	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Performance 1 (Picture Completion)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Verbal 1 (Information)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Performance 2 (Coding)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Verbal 2 (Similarities)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Performance 3 (Picture Arrangement)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Verbal 3 (Arithmetic)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Performance 4 (Block Design)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Verbal 4 (Vocabulary)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Performance 5 (Object Assembly)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Verbal 5 (Comprehension)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
N	Wechsler Full Scale IQ 3	1133	1133	1133	1133	1133	1133	1133	1133	1133	1133	1133
	Performance 1 (Picture Completion)	1133	1133	1133	1133	1133	1133	1133	1133	1133	1133	1133
	Verbal 1 (Information)	1133	1133	1133	1133	1133	1133	1133	1133	1133	1133	1133

SPSS Statistics Processor is ready H 188, W 1167 pt.

start steps and scree... steps screenho... \*Output1 [Doc... regression datas... iTunes 1:01 PM

Next, you will see the Variables Entered/Removed table. We will not use the information in this table.

<sup>5</sup> <http://cnx.org/content/m40742/latest/19.2.png/image>



The screenshot shows the SPSS Statistics Viewer window with the 'Variables Entered/Removed\*' table. A red arrow points to the 'Variables Entered' column. The table lists seven models, each with a specific variable entered and a stepwise method used. The 'Variables Removed' column is empty for all models.

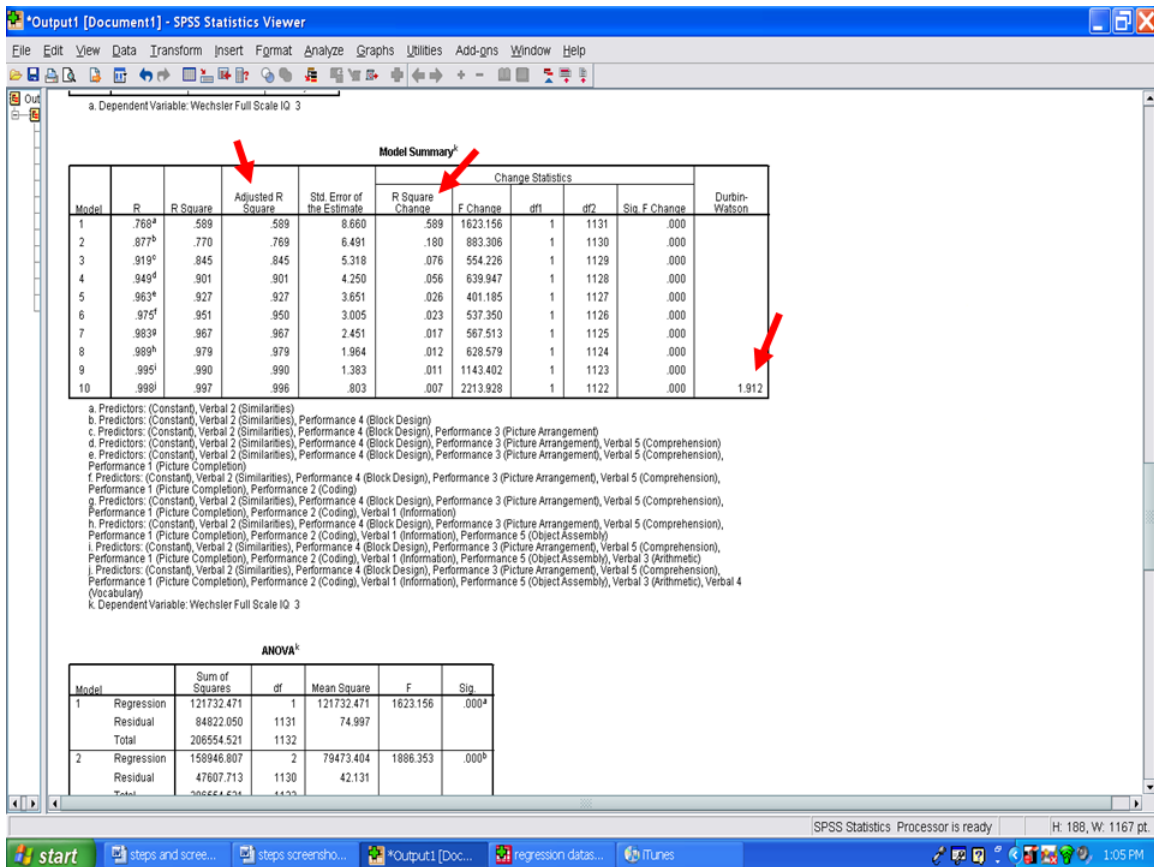
Model	Variables Entered	Variables Removed	Method
1	Verbal 2 (Similarities)		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	Performance 4 (Block Design)		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	Performance 3 (Picture Arrangement)		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
4	Verbal 5 (Comprehension)		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
5	Performance 1 (Picture Completion)		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
6	Performance 2 (Coding)		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
7	Verbal 1 (Information)		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

The table labeled Model Summary is an important table. This table is a summary of all of the steps entered/removed. Important columns in this table are:

Adjusted R Square  
R Square Change  
Durbin-Watson

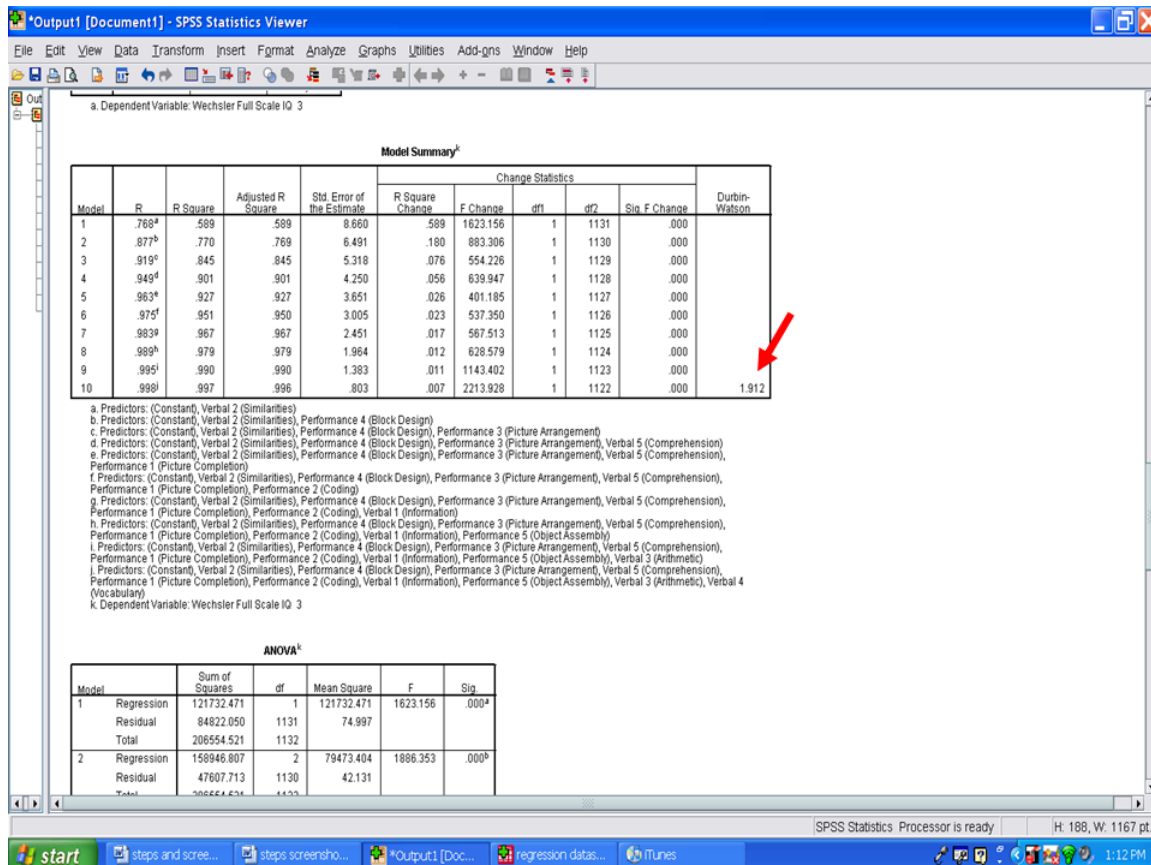
<sup>6</sup> <http://cnx.org/content/m40742/latest/19.3.png/image>





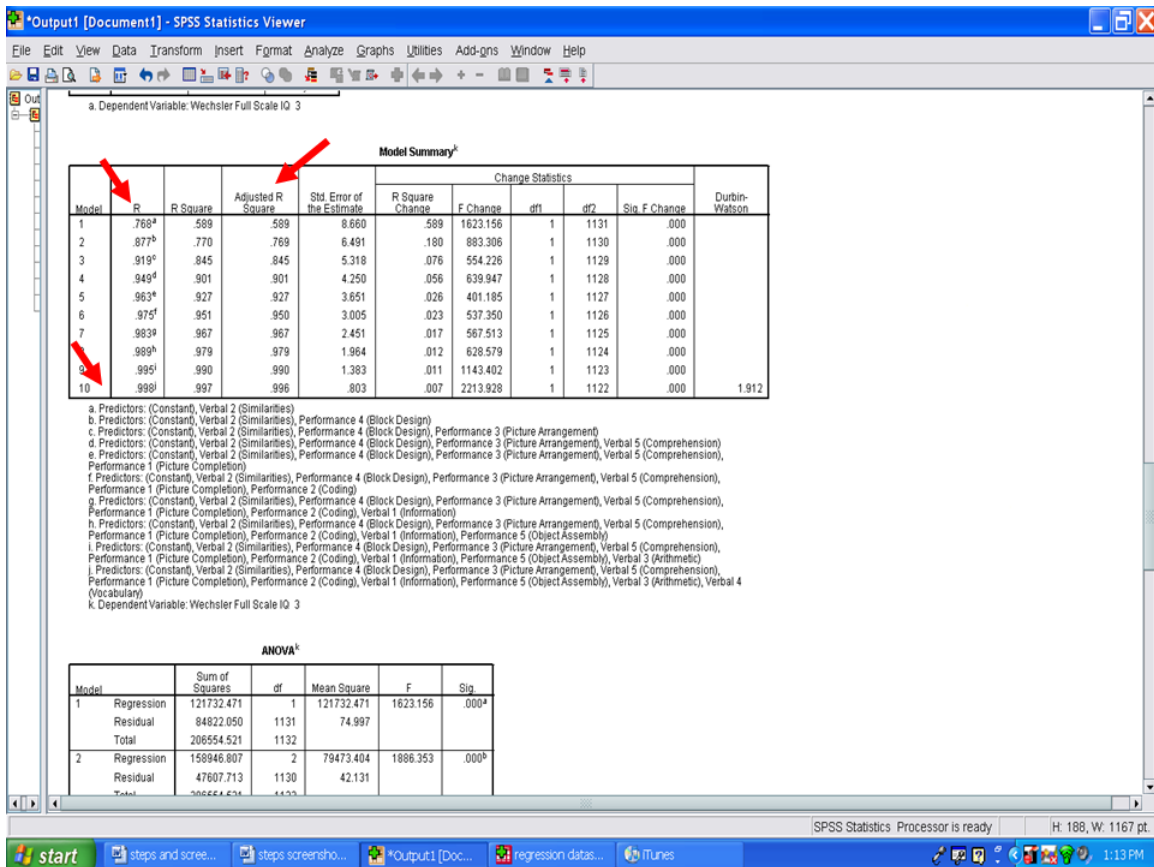
The Durbin-Watson statistic is a measure for significant residual autocorrelation. Ideally it should be close to 2. In our example, the Durbin-Watson statistic is 1.912. Therefore, this assumption has not been violated.

<sup>7</sup><http://cnx.org/content/m40742/latest/19.4.png/image>



The Adjusted R Square column indicates the amount of variance that each model explains in the dependent variable (i.e., Wechsler Full Scale IQ 3). The first model has the letter a next to it. This model contains a single independent variable, Verbal 2 (Similarities). If you recall, this variable has the highest Pearson r with the Wechsler Full Scale IQ 3. In this example, Verbal 2 (Similarities) accounts for 58.9% of the variance in the Wechsler Full Scale IQ 3.

<sup>8</sup><http://cnx.org/content/m40742/latest/19.5.png/image>



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.

<sup>9</sup><http://cnx.org/content/m40742/latest/19.6.png/image>

**Model Summary<sup>a</sup>**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.766 <sup>a</sup>	.589	.589	8.660	.589	1623.156	1	1131	.000	
2	.877 <sup>b</sup>	.770	.769	6.491	.180	883.306	1	1130	.000	
3	.919 <sup>c</sup>	.845	.845	5.318	.076	554.226	1	1128	.000	
4	.949 <sup>d</sup>	.901	.901	4.250	.056	639.947	1	1128	.000	
5	.963 <sup>e</sup>	.927	.927	3.651	.026	401.185	1	1127	.000	
6	.975 <sup>f</sup>	.951	.950	3.005	.023	537.350	1	1126	.000	
7	.983 <sup>g</sup>	.967	.967	2.451	.017	567.513	1	1125	.000	
8	.989 <sup>h</sup>	.979	.979	1.964	.012	628.579	1	1124	.000	
9	.995 <sup>i</sup>	.990	.990	1.383	.011	1143.402	1	1123	.000	
10	.998 <sup>j</sup>	.997	.996	.803	.007	2213.928	1	1122	.000	1.912

<sup>a</sup> Predictors: (Constant), Verbal 2 (Similarities)  
<sup>b</sup> Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design)  
<sup>c</sup> Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement)  
<sup>d</sup> Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension)  
<sup>e</sup> Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion)  
<sup>f</sup> Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding)  
<sup>g</sup> Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information)  
<sup>h</sup> Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information), Performance 5 (Object Assembly)  
<sup>i</sup> Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information), Performance 5 (Object Assembly), Verbal 3 (Arithmetic)  
<sup>j</sup> Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1 (Picture Completion), Performance 2 (Coding), Verbal 1 (Information), Performance 5 (Object Assembly), Verbal 3 (Arithmetic), Verbal 4 (Vocabulary)  
<sup>k</sup> Dependent Variable: Wechsler Full Scale IQ 3

**ANOVA<sup>a</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	121732.471	1	121732.471	1623.156	.000 <sup>a</sup>
	Residual	84822.050	1131	74.997		
	Total	206554.521	1132			
2	Regression	158946.807	2	79473.404	1886.353	.000 <sup>b</sup>
	Residual	47607.713	1130	42.131		
	Total	206554.521	1132			

SPSS Statistics Processor is ready | H 188, W 1167 pt | 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.

<sup>10</sup><http://cnx.org/content/m40742/latest/19.7.png/image>

SPSS Statistics Processor is ready | H 188, W 1167 pt

1:19 PM

start | steps and scree... | steps screenshot... | \*Output1 [Doc... | regression data...

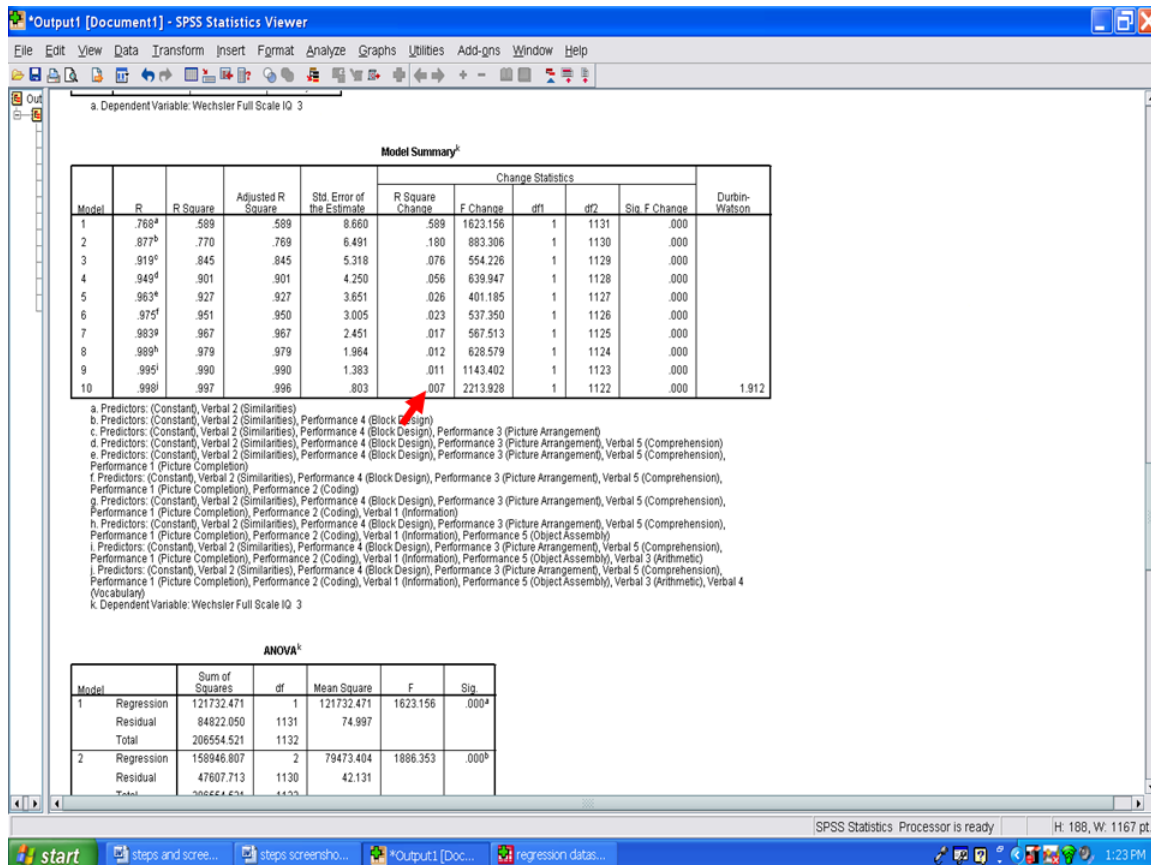
SPSS Statistics Processor is ready | H 188, W 1167 pt

1:19 PM

11

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

11 <http://cnx.org/content/m40742/latest/19.8.png/image>



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

<sup>12</sup><http://cnx.org/content/m40742/latest/19.9.png/image>

SPSS Statistics Viewer

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

ANOVA<sup>a</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	121732.471	1	121732.471	1623.156	.000 <sup>a</sup>
	Residual	84822.050	1131	74.997		
	Total	206554.521	1132			
2	Regression	158946.807	2	79473.404	1886.353	.000 <sup>b</sup>
	Residual	47607.713	1130	42.131		
	Total	206554.521	1132			
3	Regression	174622.319	3	58207.440	2057.991	.000 <sup>c</sup>
	Residual	31932.202	1129	28.284		
	Total	206554.521	1132			
4	Regression	186180.877	4	46545.219	2577.006	.000 <sup>d</sup>
	Residual	20373.644	1128	18.062		
	Total	206554.521	1132			
5	Regression	191529.444	5	38305.889	2873.246	.000 <sup>e</sup>
	Residual	15025.077	1127	13.332		
	Total	206554.521	1132			
6	Regression	196383.338	6	32730.556	3623.434	.000 <sup>f</sup>
	Residual	10171.182	1126	9.033		
	Total	206554.521	1132			
7	Regression	199793.816	7	28541.974	4749.463	.000 <sup>g</sup>
	Residual	6760.705	1125	6.010		
	Total	206554.521	1132			
8	Regression	202218.608	8	25277.326	6552.648	.000 <sup>h</sup>
	Residual	4335.913	1124	3.858		
	Total	206554.521	1132			
9	Regression	204406.080	9	22711.787	11871.558	.000 <sup>i</sup>
	Residual	2148.440	1123	1.913		
	Total	206554.521	1132			
10	Regression	205831.918	10	20583.192	31959.947	.000 <sup>j</sup>
	Residual	722.603	1122	.644		
	Total	206554.521	1132			

a. Predictors: (Constant), Verbal 2 (Similarities)  
b. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design)  
c. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement)  
d. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension)  
e. Predictors: (Constant), Verbal 2 (Similarities), Performance 4 (Block Design), Performance 3 (Picture Arrangement), Verbal 5 (Comprehension), Performance 1

Double click to edit Pivot Table

SPSS Statistics Processor is ready | H: 980, W: 537 pt

start | steps and scree... | steps screenho... | \*Output1 [Doc... | regression datas...

1:24 PM

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.

<sup>13</sup><http://cnx.org/content/m40742/latest/19.10.png/image>

SPSS Statistics Processor ready. H: 980, W: 537 pt.

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.

<sup>14</sup><http://cnx.org/content/m40742/latest/19.11.png/image>



**\*Output1 [Document1] - SPSS Statistics Viewer**

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

Regress: Total: Not: Act: Des: Cor: Var: Mo: AN: Co: Exc: Col: Res:

	Performance 1 (Picture Completion)	.661	.017	.152	38.965	.000	.608	1.645
	Performance 2 (Coding)	.628	.015	.145	42.978	.000	.811	1.233
	Verbal 1 (Information)	.814	.021	.169	38.209	.000	.471	2.123
	Performance 5 (Object Assembly)	.606	.017	.146	35.134	.000	.539	1.854
	Verbal 3 (Arithmetic)	.665	.020	.134	33.814	.000	.58	1.704
10	(Constant)	36.456	.083		437.681	.000		
	Verbal 2 (Similarities)	.647	.012	.143	54.007	.000	.444	2.251
	Performance 4 (Block Design)	.626	.010	.155	61.016	.000	.482	2.076
	Performance 3 (Picture Arrangement)	.619	.011	.139	58.888	.000	.560	1.785
	Verbal 5 (Comprehension)	.640	.011	.152	60.551	.000	.492	2.033
	Performance 1 (Picture Completion)	.645	.010	.149	65.543	.000	.607	1.647
	Performance 2 (Coding)	.631	.008	.146	74.406	.000	.811	1.233
	Verbal 1 (Information)	.645	.013	.134	50.123	.000	.434	2.302
	Performance 5 (Object Assembly)	.620	.010	.149	62.011	.000	.539	1.856
	Verbal 3 (Arithmetic)	.645	.011	.130	56.513	.000	.586	1.706
	Verbal 4 (Vocabulary)	.625	.013	.134	47.052	.000	.386	2.592

a. Dependent Variable: Wechsler Full Scale IQ\_3

**Excluded Variables<sup>1</sup>**

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	Performance 1 (Picture Completion)	.419 <sup>a</sup>	24.509	.000	.589	.811	1.233	.811
	Verbal 1 (Information)	.468 <sup>a</sup>	23.391	.000	.571	.611	1.636	.611
	Performance 2 (Coding)	.328 <sup>a</sup>	19.338	.000	.499	.948	1.055	.948
	Performance 3 (Picture Arrangement)	.450 <sup>a</sup>	26.023	.000	.612	.759	1.318	.759
	Verbal 3 (Arithmetic)	.417 <sup>a</sup>	22.916	.000	.563	.748	1.337	.748
	Performance 4 (Block Design)	.481 <sup>a</sup>	29.720	.000	.662	.778	1.285	.778
	Verbal 4 (Vocabulary)	.447 <sup>a</sup>	20.607	.000	.523	.561	1.783	.561
	Performance 5 (Object Assembly)	.398 <sup>a</sup>	23.207	.000	.568	.837	1.194	.837
	Verbal 5 (Comprehension)	.423 <sup>a</sup>	21.610	.000	.541	.672	1.488	.672

SPSS Statistics Processor is ready | H: 980, W: 537 pt.

start | steps and scree... | steps screenho... | \*Output1 [Doc... | regression datas... | 1:30 PM

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.

<sup>15</sup><http://cnx.org/content/m40742/latest/19.12.png/image>

**\*Output1 [Document1] - SPSS Statistics Viewer**

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

Performance 1 (Picture Completion) .661 .017 .152 38.965 .000 .608 1.645

Performance 2 (Coding) .628 .015 .145 42.978 .000 .811 1.233

Verbal 1 (Information) .814 .021 .169 38.209 .000 .471 2.123

Performance 5 (Object Assembly) .606 .017 .146 35.134 .000 .539 1.854

Verbal 3 (Arithmetic) .665 .020 .134 33.814 .000 .587 1.704

10 (Constant) 36.456 .083 .437 681 .000 .444 2.251

Verbal 2 (Similarities) .647 .012 .143 54.007 .000 .482 2.076

Performance 4 (Block Design) .626 .010 .155 61.016 .000 .560 1.785

Performance 3 (Picture Arrangement) .619 .011 .139 58.888 .000 .492 2.033

Verbal 5 (Comprehension) .640 .011 .152 60.551 .000 .607 1.647

Performance 1 (Picture Completion) .645 .010 .149 65.543 .000 .811 1.233

Performance 2 (Coding) .631 .008 .146 74.406 .000 .434 2.302

Verbal 1 (Information) .645 .013 .134 50.123 .000 .539 1.856

Performance 5 (Object Assembly) .620 .010 .149 62.011 .000 .586 1.706

Verbal 3 (Arithmetic) .645 .011 .130 56.513 .000 .386 2.592

Verbal 4 (Vocabulary) .625 .013 .134 47.052 .000

a. Dependent Variable: Wechsler Full Scale IQ-3

**Excluded Variables<sup>a</sup>**

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	Performance 1 (Picture Completion)	.419 <sup>a</sup>	24.509	.000	.589	.811	1.233	.811
	Verbal 1 (Information)	.468 <sup>a</sup>	23.391	.000	.571	.811	1.636	.611
	Performance 2 (Coding)	.328 <sup>a</sup>	19.339	.000	.499	.948	1.055	.948
	Performance 3 (Picture Arrangement)	.450 <sup>a</sup>	26.023	.000	.612	.759	1.318	.759
	Verbal 3 (Arithmetic)	.417 <sup>a</sup>	22.916	.000	.563	.748	1.337	.748
	Performance 4 (Block Design)	.481 <sup>a</sup>	29.720	.000	.662	.778	1.285	.778
	Verbal 4 (Vocabulary)	.447 <sup>a</sup>	20.607	.000	.523	.561	1.783	.561
	Performance 5 (Object Assembly)	.398 <sup>a</sup>	23.207	.000	.568	.837	1.194	.837
	Verbal 5 (Comprehension)	.423 <sup>a</sup>	21.810	.000	.541	.672	1.488	.672

SPSS Statistics Processor is ready H: 980, W: 537 pt

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.

<sup>16</sup><http://cnx.org/content/m40742/latest/19.13.png/image>

## Attributions

Collection: *Calculating Advanced Statistics: Part II*  
Edited by: National Council of Professors of Educational Administration  
URL: <http://cnx.org/content/col11348/1.2/>  
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Multivariate Analysis of Variance: Part I"  
By: John R. Slate, Ana Rojas-LeBouef  
URL: <http://cnx.org/content/m40728/1.2/>  
Pages: 1-28  
Copyright: John R. Slate, Ana Rojas-LeBouef  
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Multivariate Analysis of Variance: Part II"  
By: John R. Slate, Ana Rojas-LeBouef  
URL: <http://cnx.org/content/m40729/1.2/>  
Pages: 29-41  
Copyright: John R. Slate, Ana Rojas-LeBouef  
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Multiple Analysis of Variance: Part III"  
By: John R. Slate, Ana Rojas-LeBouef  
URL: <http://cnx.org/content/m40731/1.2/>  
Pages: 43-54  
Copyright: John R. Slate, Ana Rojas-LeBouef  
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Discriminant Analysis: Assumptions"  
By: John R. Slate, Ana Rojas-LeBouef  
URL: <http://cnx.org/content/m40733/1.2/>  
Pages: 55-61  
Copyright: John R. Slate, Ana Rojas-LeBouef  
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Discriminant Analysis: Part I"  
By: John R. Slate, Ana Rojas-LeBouef  
URL: <http://cnx.org/content/m40735/1.2/>  
Pages: 63-71  
Copyright: John R. Slate, Ana Rojas-LeBouef  
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Discriminant Analysis: Part II"  
By: John R. Slate, Ana Rojas-LeBouef  
URL: <http://cnx.org/content/m40736/1.2/>  
Pages: 73-83  
Copyright: John R. Slate, Ana Rojas-LeBouef  
License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Discriminant Analysis: Part III"

By: John R. Slate, Ana Rojas-LeBouef

URL: <http://cnx.org/content/m40737/1.2/>

Pages: 85-93

Copyright: John R. Slate, Ana Rojas-LeBouef

License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Multiple Regression: Assumptions"

By: John R. Slate, Ana Rojas-LeBouef

URL: <http://cnx.org/content/m40738/1.2/>

Pages: 95-103

Copyright: John R. Slate, Ana Rojas-LeBouef

License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Multiple Regression: Part I"

By: John R. Slate, Ana Rojas-LeBouef

URL: <http://cnx.org/content/m40739/1.2/>

Pages: 105-114

Copyright: John R. Slate, Ana Rojas-LeBouef

License: <http://creativecommons.org/licenses/by/3.0/>

Module: "Multiple Regression: Part II"

By: John R. Slate, Ana Rojas-LeBouef

URL: <http://cnx.org/content/m40742/1.2/>

Pages: 115-128

Copyright: John R. Slate, Ana Rojas-LeBouef

License: <http://creativecommons.org/licenses/by/3.0/>

## About Connexions

Since 1999, Connexions has been pioneering a global system where anyone can create course materials and make them fully accessible and easily reusable free of charge. We are a Web-based authoring, teaching and learning environment open to anyone interested in education, including students, teachers, professors and lifelong learners. We connect ideas and facilitate educational communities.

Connexions's modular, interactive courses are in use worldwide by universities, community colleges, K-12 schools, distance learners, and lifelong learners. Connexions materials are in many languages, including English, Spanish, Chinese, Japanese, Italian, Vietnamese, French, Portuguese, and Thai. Connexions is part of an exciting new information distribution system that allows for **Print on Demand Books**. Connexions has partnered with innovative on-demand publisher QOOP to accelerate the delivery of printed course materials and textbooks into classrooms worldwide at lower prices than traditional academic publishers.