Calculating Descriptive Statistics*

John R. Slate
Ana Rojas-LeBouef

This work is produced by The Connexions Project and licensed under the
Creative Commons Attribution License †

Abstract

Calculating Descriptive Statistics is Chapter 1 of Calculating Basic Statistical Procedures in SPSS: A Self-Help and Practical Guide to Preparing Theses, Dissertations, and Manuscripts, authored by John R. Slate and Ana Rojas-LeBouef from Sam Houston State University. This book is written to assist graduate students and faculty members, as well as undergraduate students, in their use of the Statistical Package of the Social Sciences-PC (SPSS-PC) versions 15-19. Specifically, we have generated a set of steps and screenshots to depict each important step in conducting basic statistical analyses. We believe that this book supplements existing statistical texts in which readers are informed about the statistical underpinnings of basic statistical procedures and in which definitions of terms are provided. Accordingly, other than providing a few basic definitions, we assume that dissertation chairs/thesis directors, students, and/or faculty will obtain their own definition of terms. We hope you find this set of steps and screenshots to be helpful as you use SPSS-PC in conducting basic statistical analyses.

NOTE: This 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. Formatted and edited in Connexions by Theodore Creighton and Brad Bizzell, Virginia Tech, Janet Tareilo, Stephen F. Austin State University, and Thomas Kersten, Roosevelt University.

1

This chapter is part of a larger Collection (Book) and is available at: *Calculating Basic Statistical Procedures in SPSS: A Self-Help and Practical Guide to Preparing Theses, Dissertations, and Manuscripts*[1]

NOTE: Slate and LeBouef have written a "companion book" which is available at: *Preparing and Presenting Your Statistical Findings: Model Write Ups*[2]

*Version 1.9: Apr 28, 2011 7:56 am -0500
†http://creativecommons.org/licenses/by/3.0/
[1]http://my.qoop.com/store/NCPEA-Publications-1781472103076212/
[2]http://my.qoop.com/store/NCPEA-Publications-1781472103076212/

Authors Information

John R. Slate is a Professor at Sam Houston State University where he teaches Basic and Advanced Statistics courses, as well as professional writing, to doctoral students in Educational Leadership and Counseling. His research interests lie in the use of educational databases, both state and national, to reform school practices. To date, he has chaired and/or served over 100 doctoral student dissertation committees. Recently, Dr. Slate created a website, Writing and Statistical Help[3] to assist students and faculty with both statistical assistance and in editing/writing their dissertations/theses and manuscripts.

Ana Rojas-LeBouef is a Literacy Specialist at the Reading Center at Sam Houston State University where she teaches developmental reading courses. She 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 website, Writing and Statistical Help[4]

Editors Information

Theodore B. Creighton, is a Professor at Virginia Tech and the Publications Director for NCPEA Publications[5] , the Founding Editor of Education Leadership Review,[6] and the Senior Editor of the NCPEA Connexions Project.

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

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

Thomas Kersten is a Professor at Roosevelt University in Chicago. Dr. Kersten is widely published and an experienced editor and is the author of Taking the Mystery Out of Illinois School Finance[7] , a Connexions Print on Demand publication. He is also serving as Editor in Residence for this book by Slate and LeBouef.

2 Calculating Descriptive Statistics

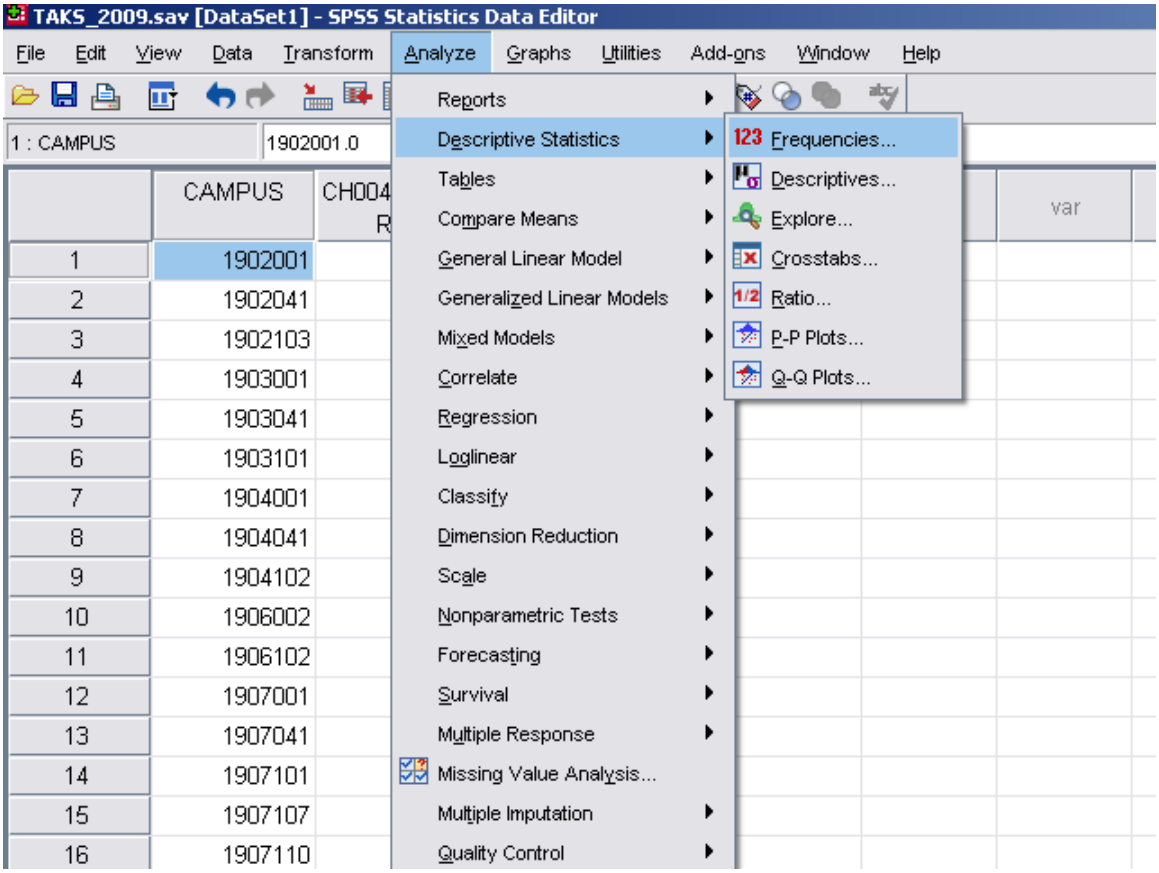
In this set of steps, readers are provided with directions on calculating basic measures of central tendency (i.e., mean, median, and mode), measures of dispersion (i.e., standard deviation, variance, and range), and measures of normality (i.e., skewness and kurtosis). For detailed information regarding the advantages and limitations of each of the measures cited, readers are referred to the Hyperstats Online Statistics Textbook at http://davidmlane.com/hyperstat/[8] or to the *Electronic Statistics Textbook* (2011) at http://www.statsoft.com/textbook/[9]

Step One

First check the accuracy of your dataset.

[3] http://cnx.org/content/m37276/latest/www.writingandstatisticalhelp.com
[4] http://cnx.org/content/m37276/latest/www.writingandstatisticalhelp.com
[5] http://www.ncpeapublications.org
[6] http://ncpeapublications.org/about-elr.html
[7] http://cnx.org/content/col10606/latest/
[8] http://davidmlane.com/hyperstat/
[9] http://www.statsoft.com/textbook/

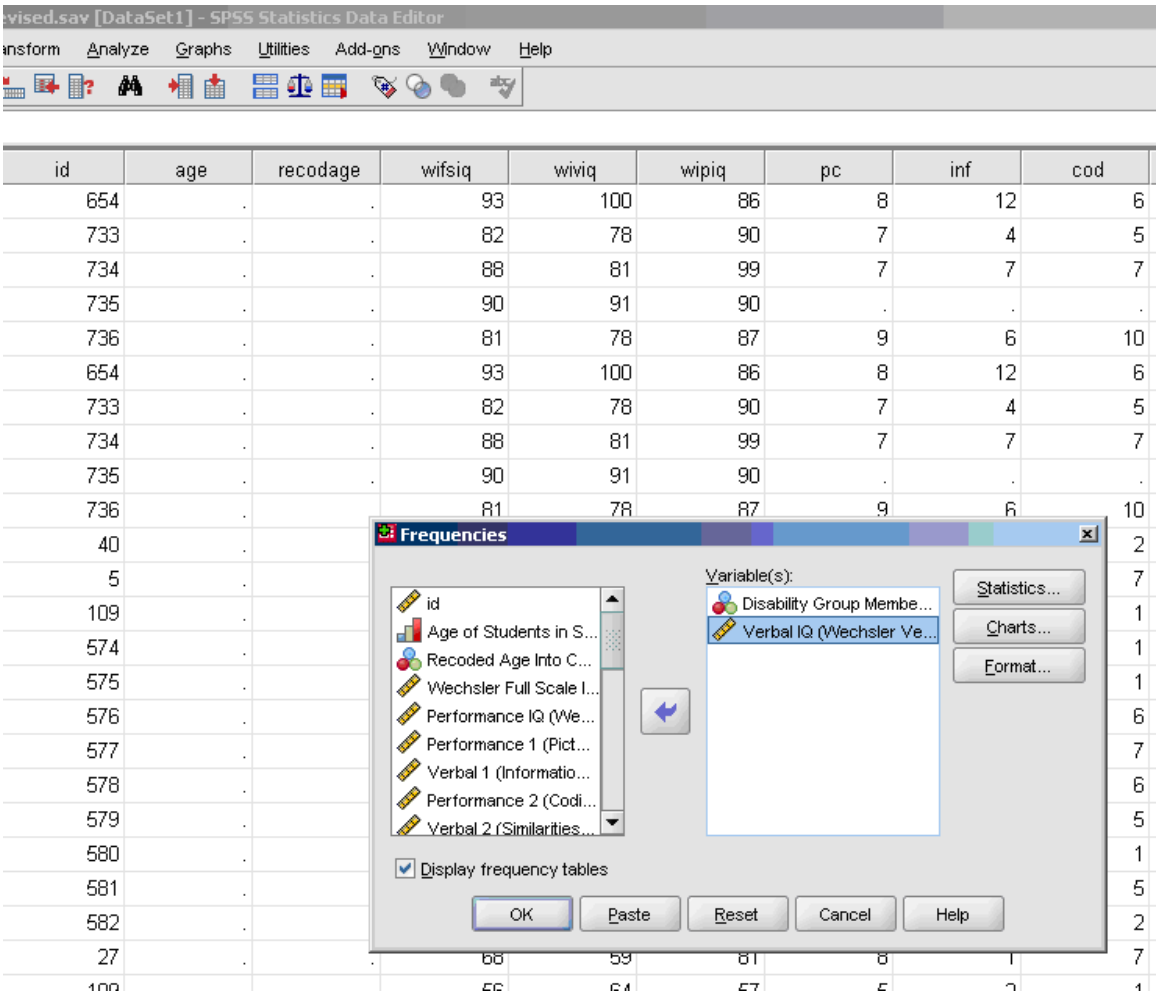
√ Analyze
* Descriptive Statistics
* Frequencies



[10]

√ Move over the independent variable/s
√ Move over the dependent variable/s
√ OK

[10] http://cnx.org/content/m37276/latest/figure1.1-thumb.png/image



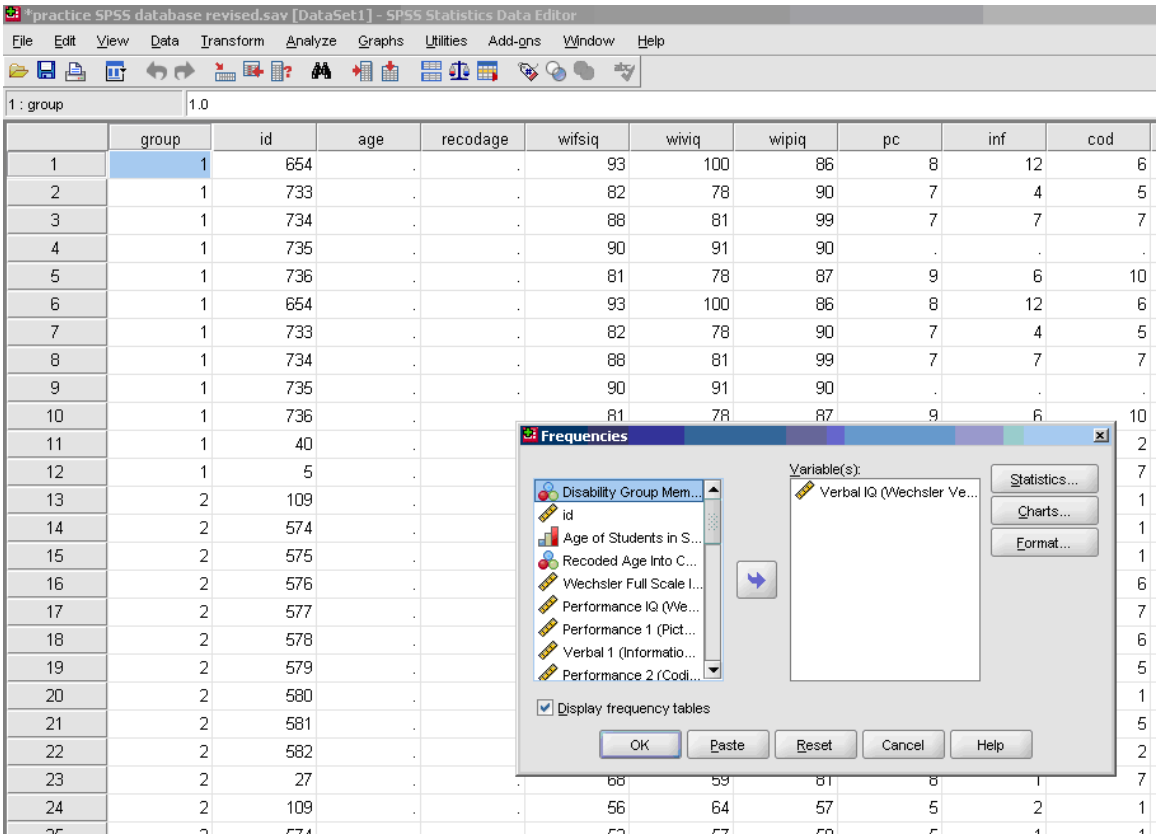
[11]

* Uncheck the display "frequency tables" so that you are not provided with the frequencies of your data every time descriptive statistics are obtained.

Now check your output to see that the values for each of the variables is within the possible limits (e.g., 1 and 2 for gender). If your dataset is inaccurate, correct any inaccuracies before calculating any statistics.

To calculate descriptive statistics:
√ Analyze
* Descriptive Statistics
* Frequencies
* Move over the dependent variable/s
* Do **NOT** move over the independent variable/s or any string variables
* Statistics

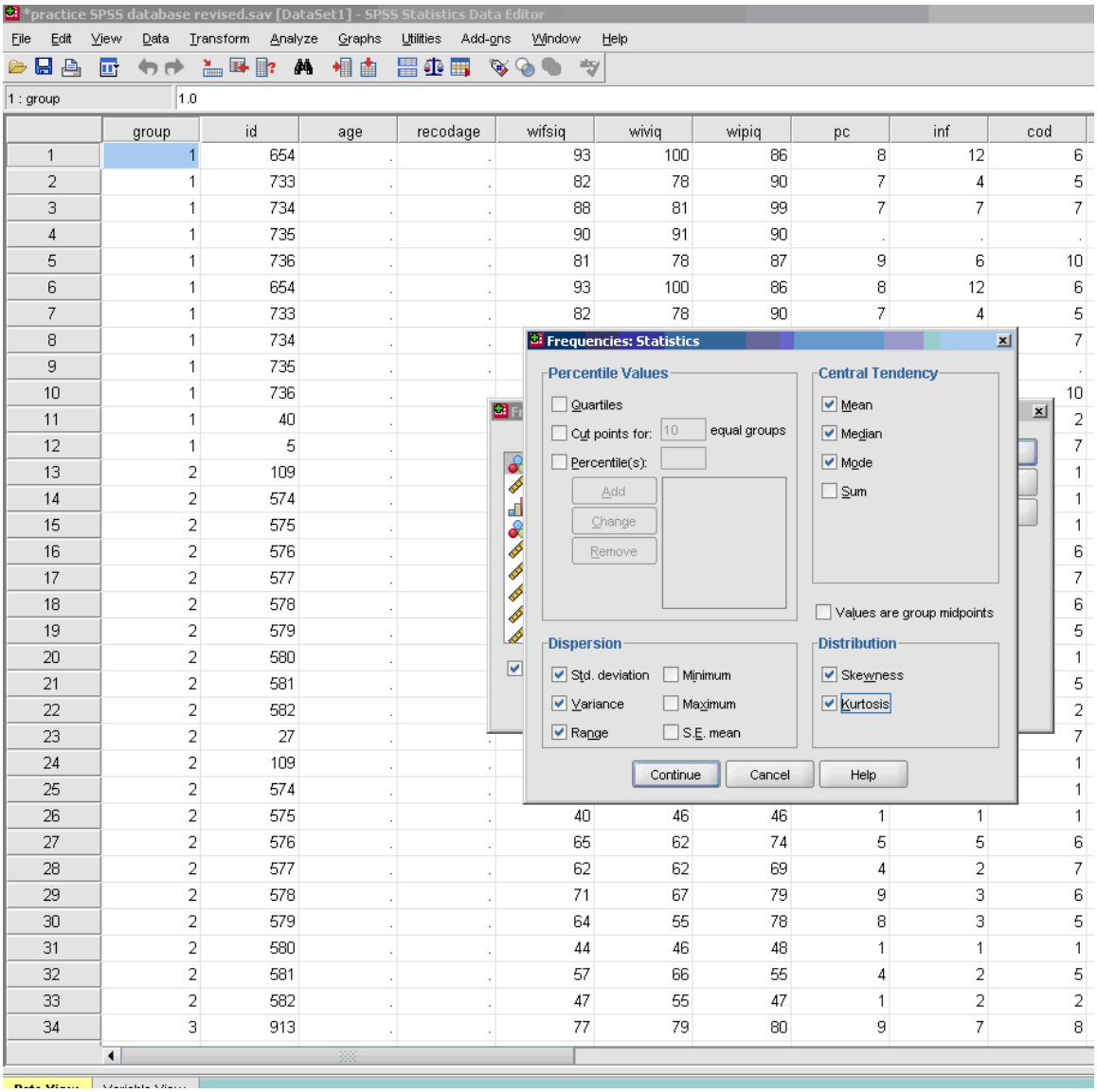
[11] http://cnx.org/content/m37276/latest/figure1.2.PNG/image



[12]

* Three basic measures of central tendency, upper right part of screen: mean, median, and mode.
* Three basic measures of variability, bottom left part of screen: variance, Standard Deviation, and range.
* 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[13] and http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb[14]
* 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[15] and http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb[16]

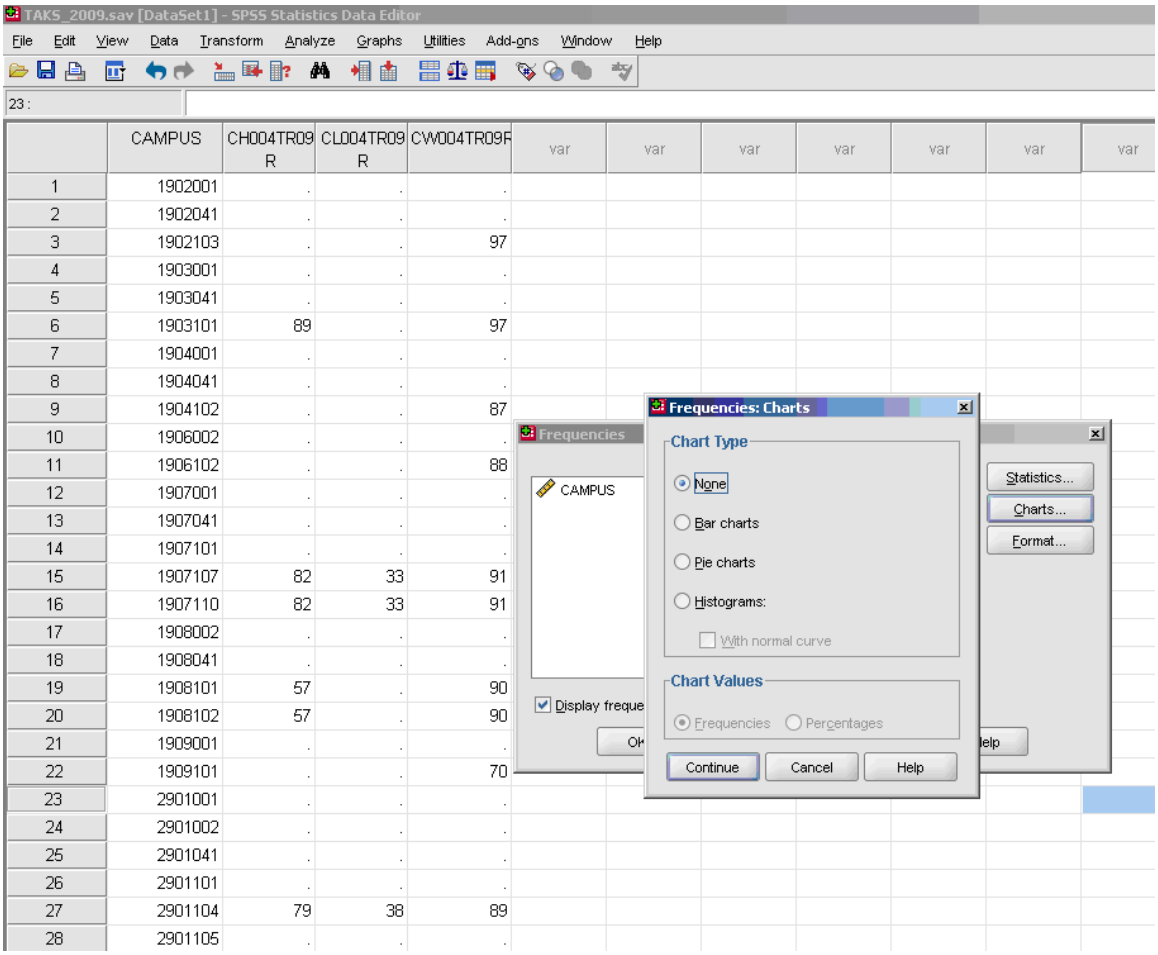
[12] http://cnx.org/content/m37276/latest/figure1.3.PNG/image
[13] http://www.statistics.com/index.php?page=glossary&term_id=356
[14] http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb
[15] http://www.statistics.com/index.php?page=glossary&term_id=326
[16] http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb



[17]

* Charts (optional, use only if you want a visual depiction of your data)
* Histograms (optional, use only if you want a visual depiction of your data)with normal curve

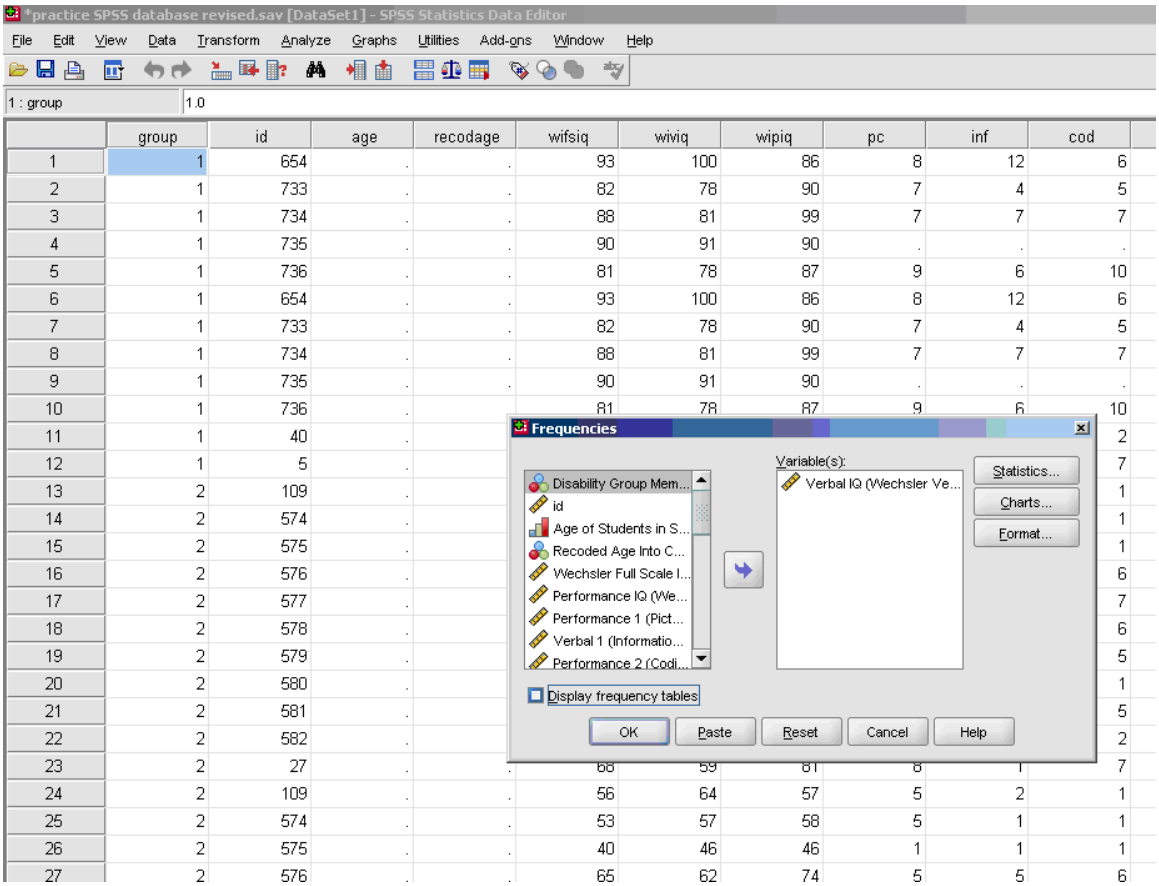
[17] http://cnx.org/content/m37276/latest/figure1.4.PNG/image



[18]

* Uncheck the display frequency tables so that you are not provided with the frequencies of your data every time descriptive statistics are obtained.
* OK

[18] http://cnx.org/content/m37276/latest/figure1.5.PNG/image

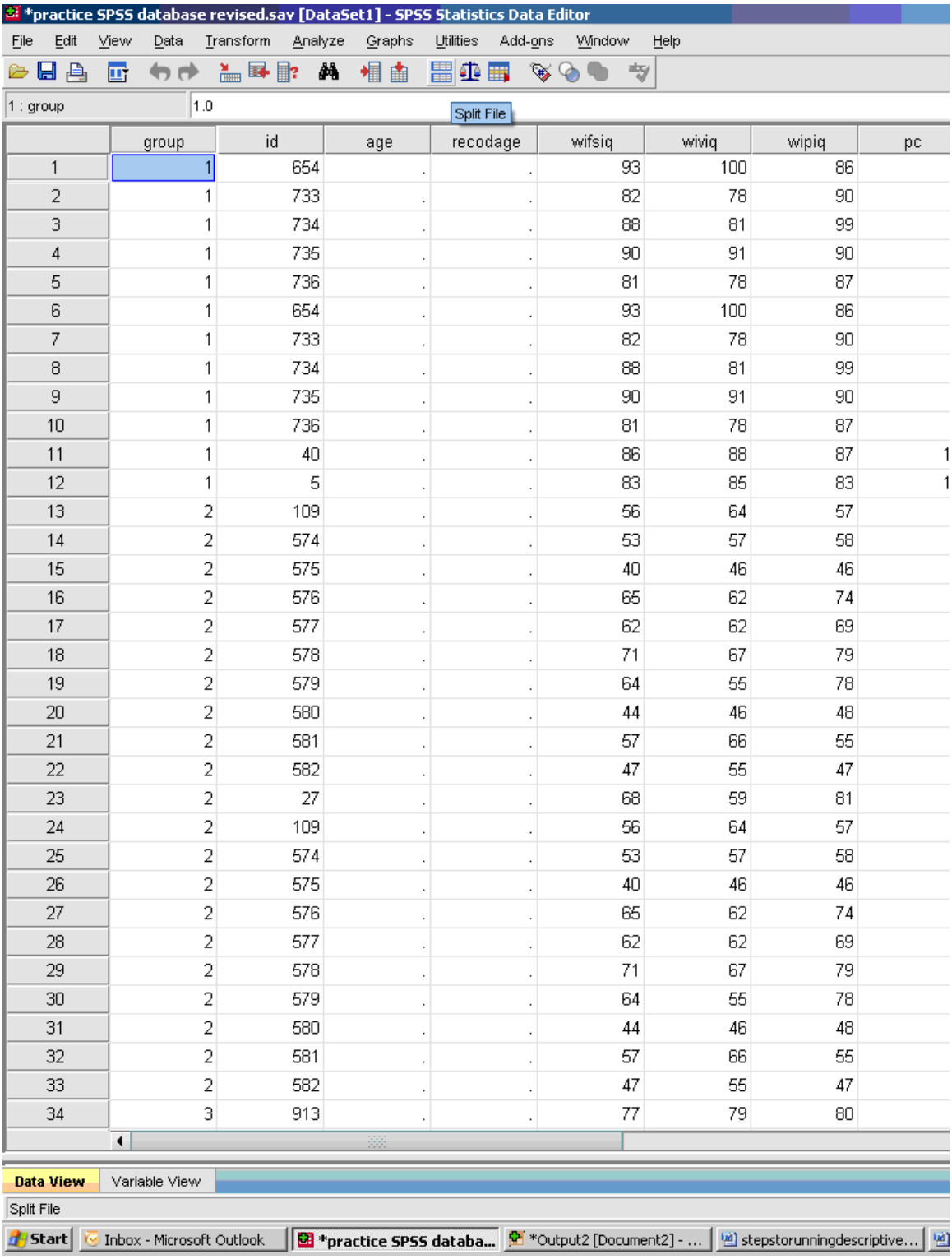


[19]

To obtain descriptive statistics for subgroups, do the following:

* Split File (icon middle top of screen next to the scales)

[19] http://cnx.org/content/m37276/latest/figure1.6.PNG/image



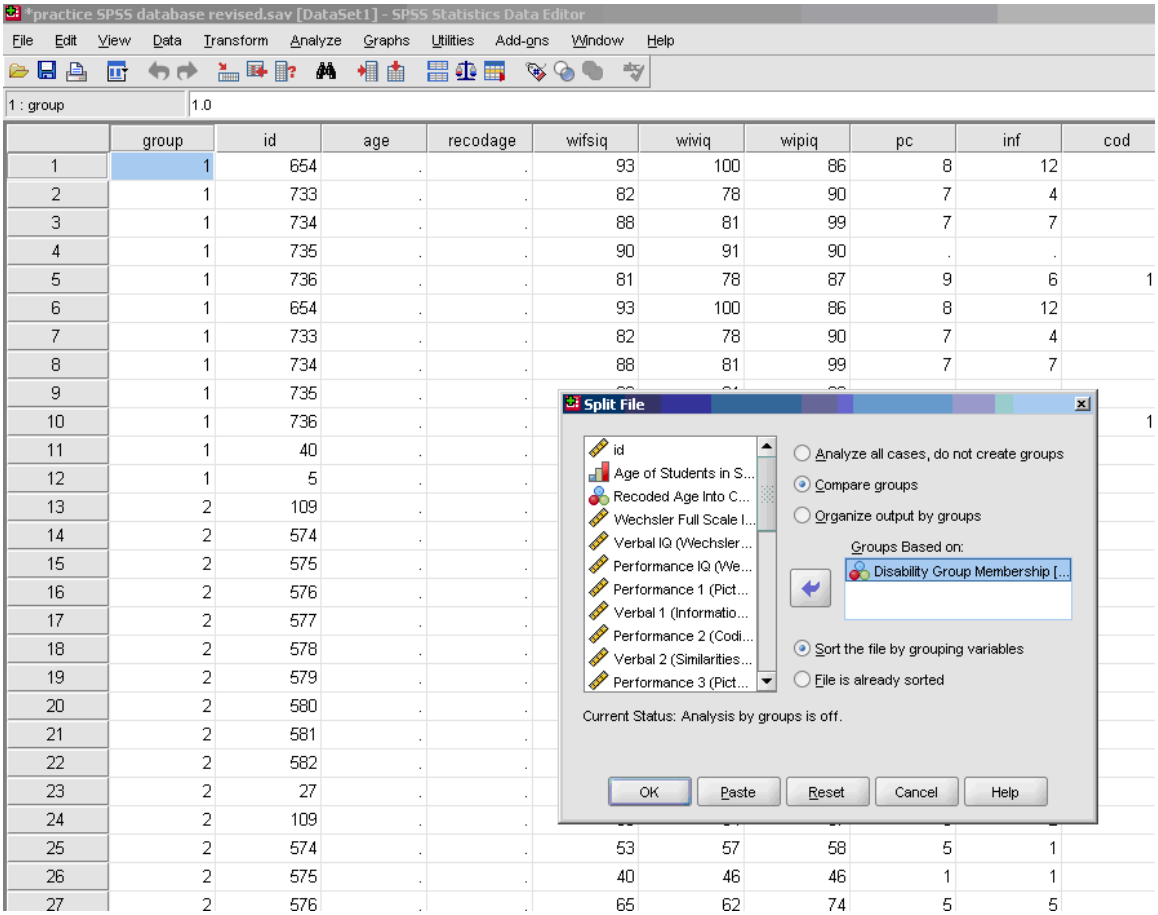
	group	id	age	recodage	wifsiq	wiviq	wipiq	pc
1	1	654	.	.	93	100	86	
2	1	733	.	.	82	78	90	
3	1	734	.	.	88	81	99	
4	1	735	.	.	90	91	90	
5	1	736	.	.	81	78	87	
6	1	654	.	.	93	100	86	
7	1	733	.	.	82	78	90	
8	1	734	.	.	88	81	99	
9	1	735	.	.	90	91	90	
10	1	736	.	.	81	78	87	
11	1	40	.	.	86	88	87	1
12	1	5	.	.	83	85	83	1
13	2	109	.	.	56	64	57	
14	2	574	.	.	53	57	58	
15	2	575	.	.	40	46	46	
16	2	576	.	.	65	62	74	
17	2	577	.	.	62	62	69	
18	2	578	.	.	71	67	79	
19	2	579	.	.	64	55	78	
20	2	580	.	.	44	46	48	
21	2	581	.	.	57	66	55	
22	2	582	.	.	47	55	47	
23	2	27	.	.	68	59	81	
24	2	109	.	.	56	64	57	
25	2	574	.	.	53	57	58	
26	2	575	.	.	40	46	46	
27	2	576	.	.	65	62	74	
28	2	577	.	.	62	62	69	
29	2	578	.	.	71	67	79	
30	2	579	.	.	64	55	78	
31	2	580	.	.	44	46	48	
32	2	581	.	.	57	66	55	
33	2	582	.	.	47	55	47	
34	3	913	.	.	77	79	80	

[20]

* Compare Groups

[20] http://cnx.org/content/m37276/latest/figure1.7.PNG/image

* Click on group (typically dichotomous in nature) and move to empty cell.

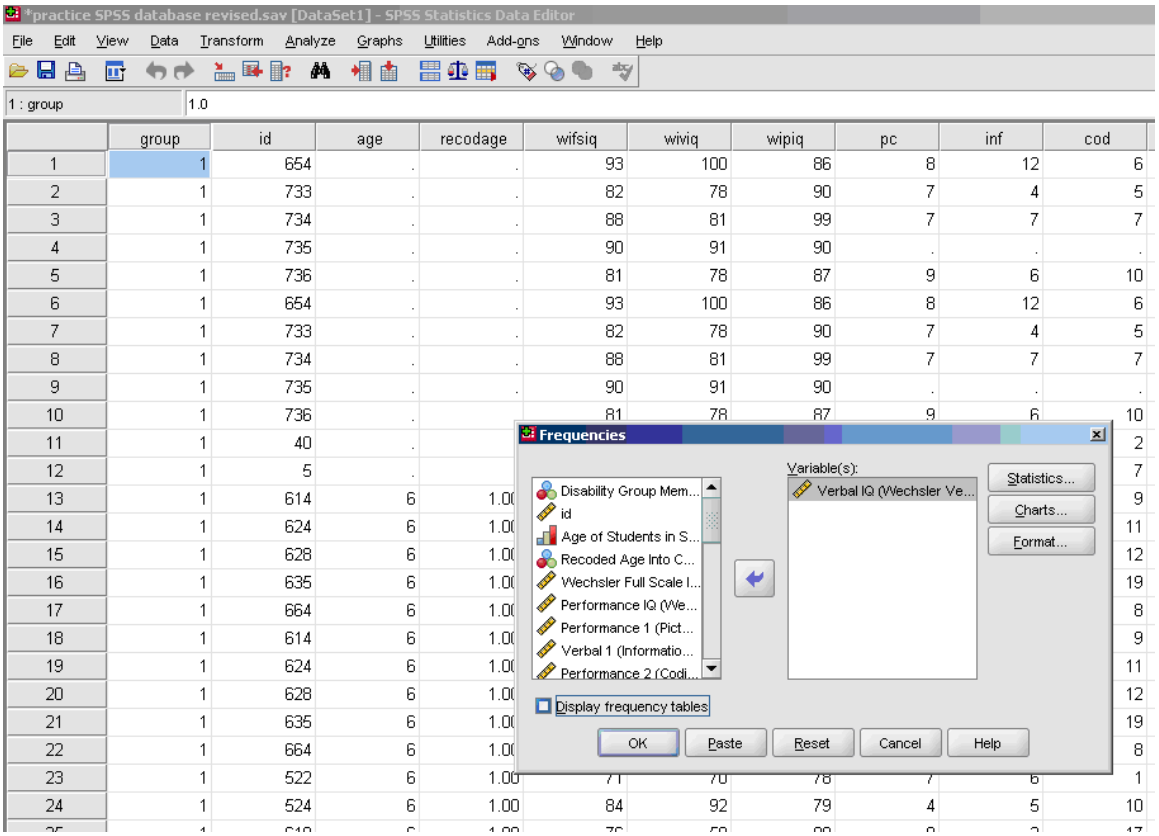


[21]

* OK

√ Analyze
* Descriptive Statistics
* Frequencies
* Move over the dependent variable/s
* Do **NOT** move over the independent variable/s or any string variables

[21] http://cnx.org/content/m37276/latest/figure1.8.PNG/image



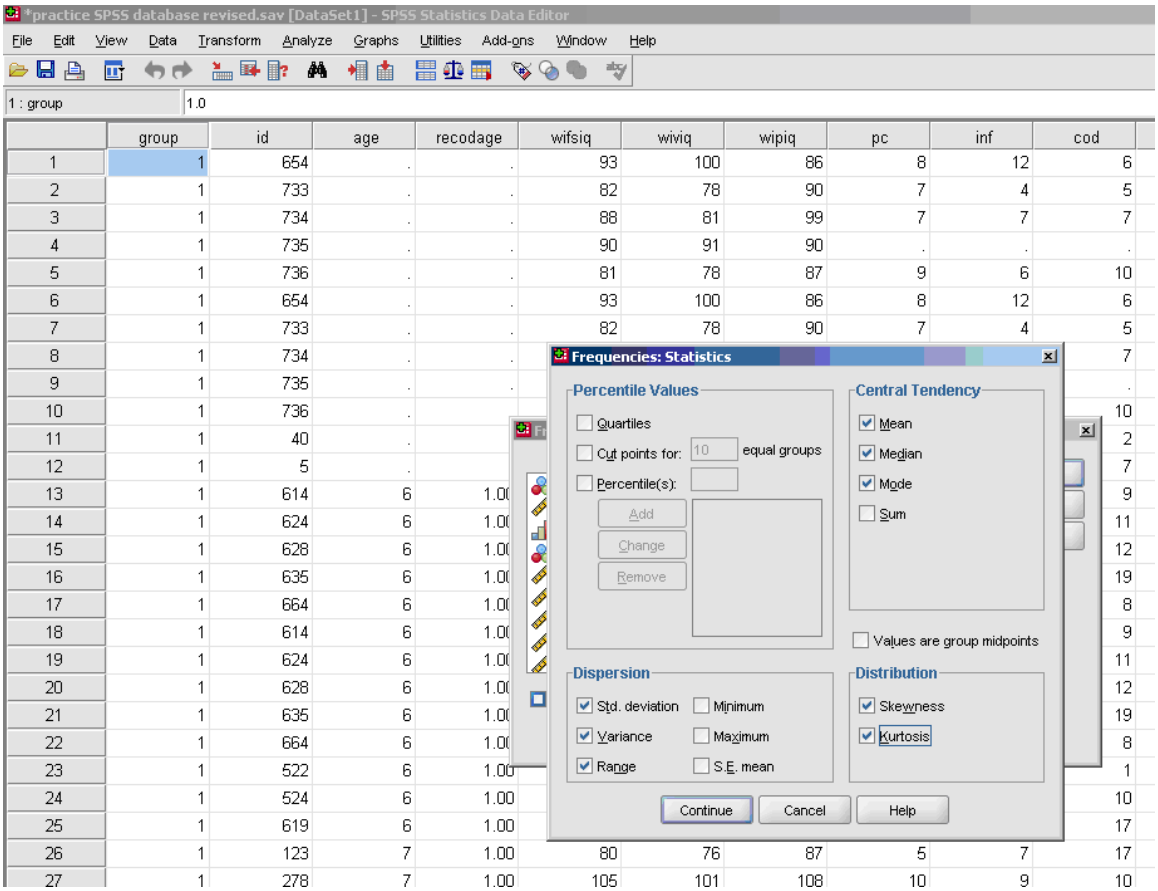
[22]

√ Statistics

* Three basic measures of central tendency, upper right part of screen: mean, median, and mode.
* Three basic measures of variability, bottom left part of screen: variance, Standard Deviation, and range.
* 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[23] and http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb[24]
* 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[25] and http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb[26]

* Continue

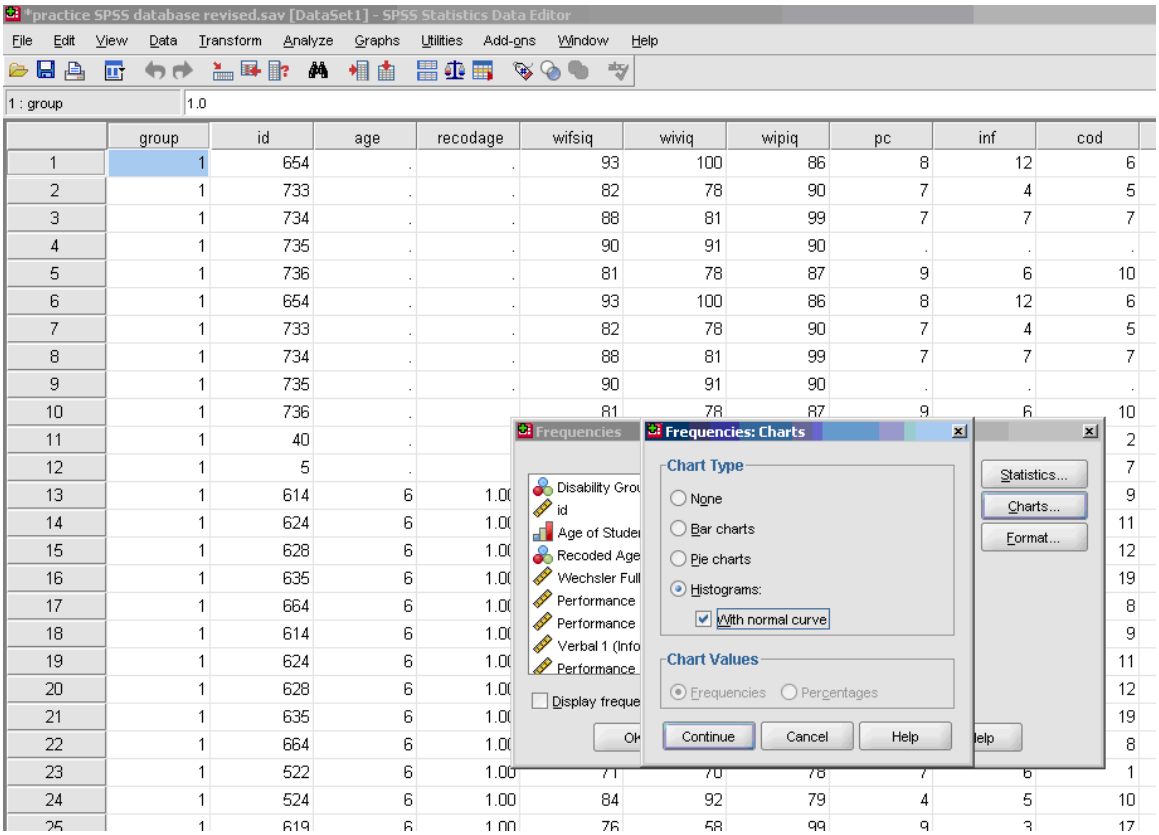
[22] http://cnx.org/content/m37276/latest/figure1.9.PNG/image
[23] http://www.statistics.com/index.php?page=glossary&term_id=356
[24] http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb
[25] http://www.statistics.com/index.php?page=glossary&term_id=326
[26] http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb



[27]

* Charts (optional, use only if you want a visual depiction of your data)
* Histograms (optional, use only if you want a visual depiction of your data)with normal curve

[27] http://cnx.org/content/m37276/latest/figure1.10.PNG/image

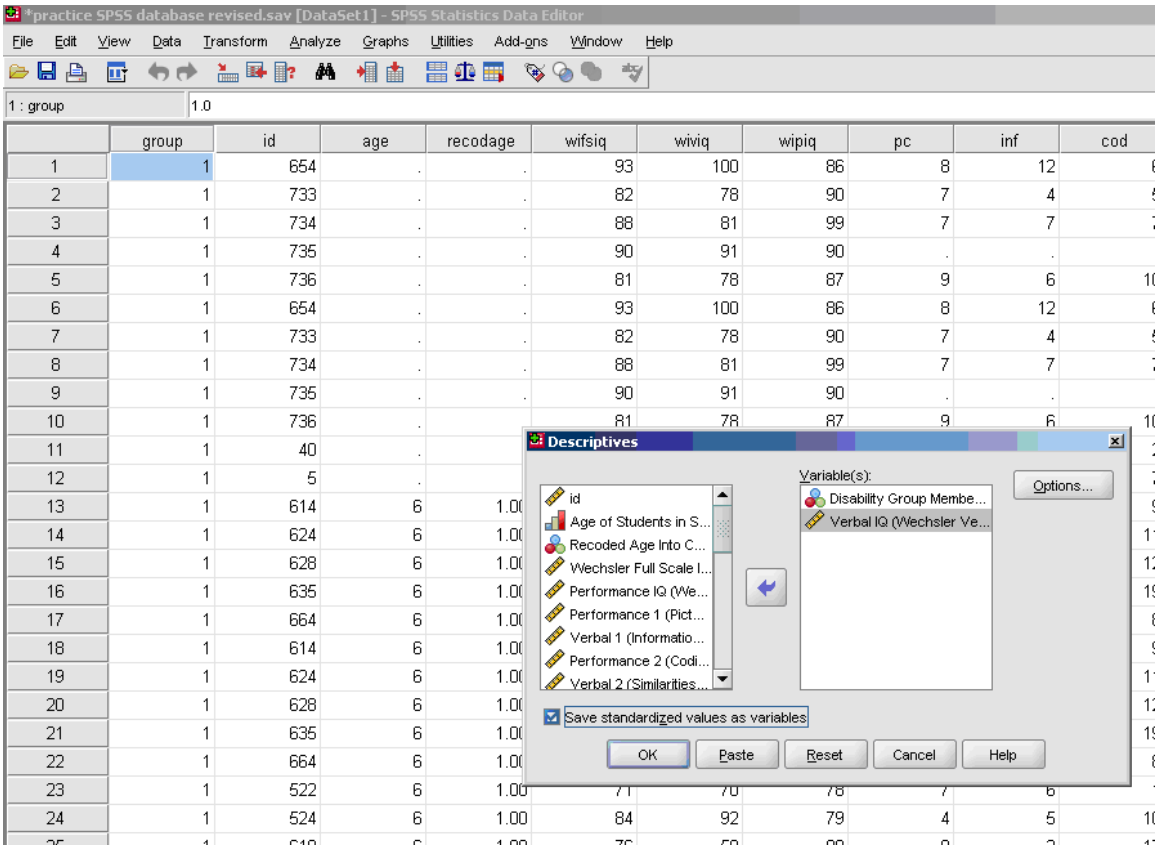


[28]

* OK

To calculate a z-score for any continuous variable:
√ Analyze
* Descriptive Statistics
* Descriptives
* Send variable on which you want z-scores to be calculated to empty cell
* Check box for Save standardized values as variables

[28] http://cnx.org/content/m37276/latest/figure1.11.PNG/image



[29]

* OK
* You will be sent to the output window, as shown in Table 1. [Note. In some versions of SPSS, you will not be sent to the output window, but will remain in the data window.] The information in the output window is not relevant for your purposes. To see the variable that was just created, go to the SPSS data. The far right column should now be the new z-score variable that was created.

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Verbal IQ (Wechsler Verbal Intelligence 3)	1182	46	129	77.97	13.661
Valid N (listwise)	1182				

Table 1: Descriptive Statistics

* A new variable/s will have been generated for you in the data window

To get this information in a usable output form, do the following:

√ Analyze
* Descriptive Statistics
* Frequencies
* Move over the newly created z-score variable(s) (z-scores will generally appear at the bottom of your list with the words: "Zscore: Verbal IQ (Wechsler Verbal Intelligence 3)

[29] http://cnx.org/content/m37276/latest/figure1.12.PNG/image

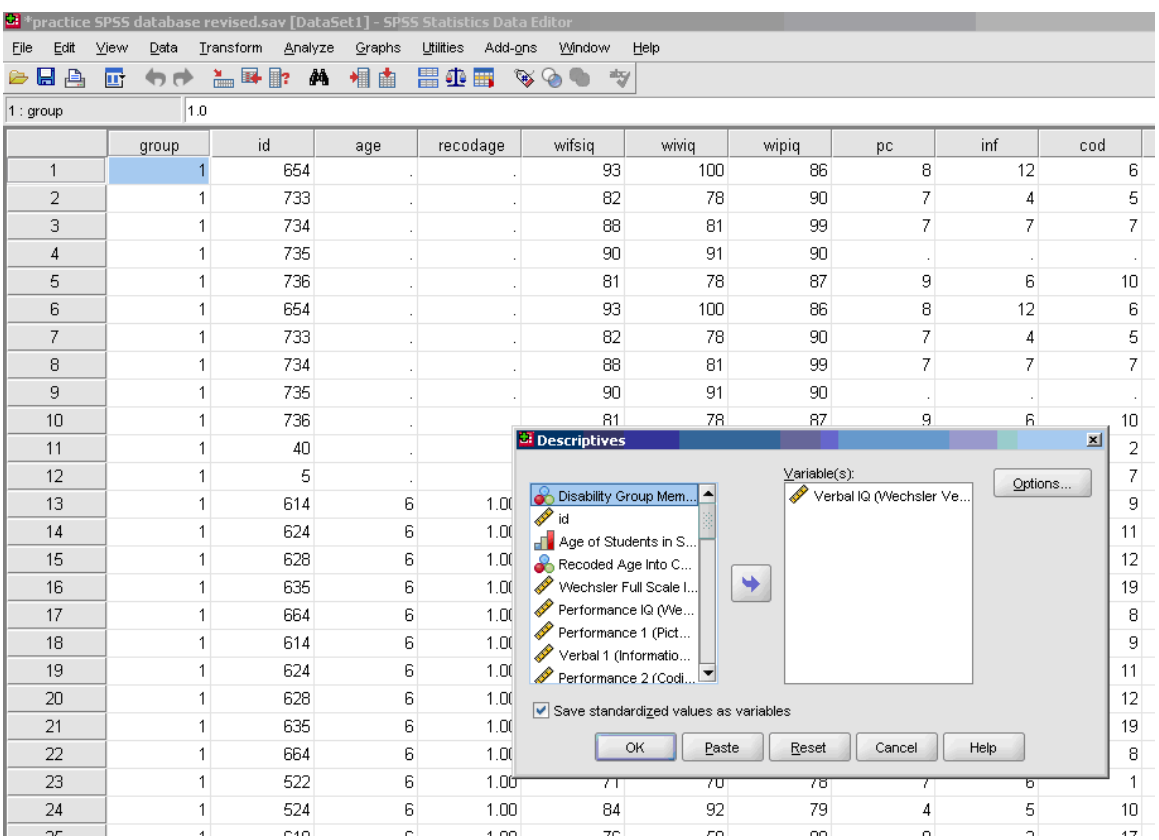
* Make sure the frequencies box is checked
* OK

* Copy or cut the frequency table for this z-score variable and carry it into WORD. Delete any irrelevant information.

		Zscore: Verbal IQ (Wechsler Verbal Intelligence 3)	Zscore(wiviq) Verbal IQ (Wechsler Verbal Intelligence 3)
N	Valid	1182	1182
	Missing	0	0

Table 2: Z Scores

To calculate a T-score for any continuous variable:
√ Analyze
* Descriptive Statistics
* Descriptives
* Send variable on which you want T scores to be calculated to empty cell
* Check box for Save standardized values as variables



30

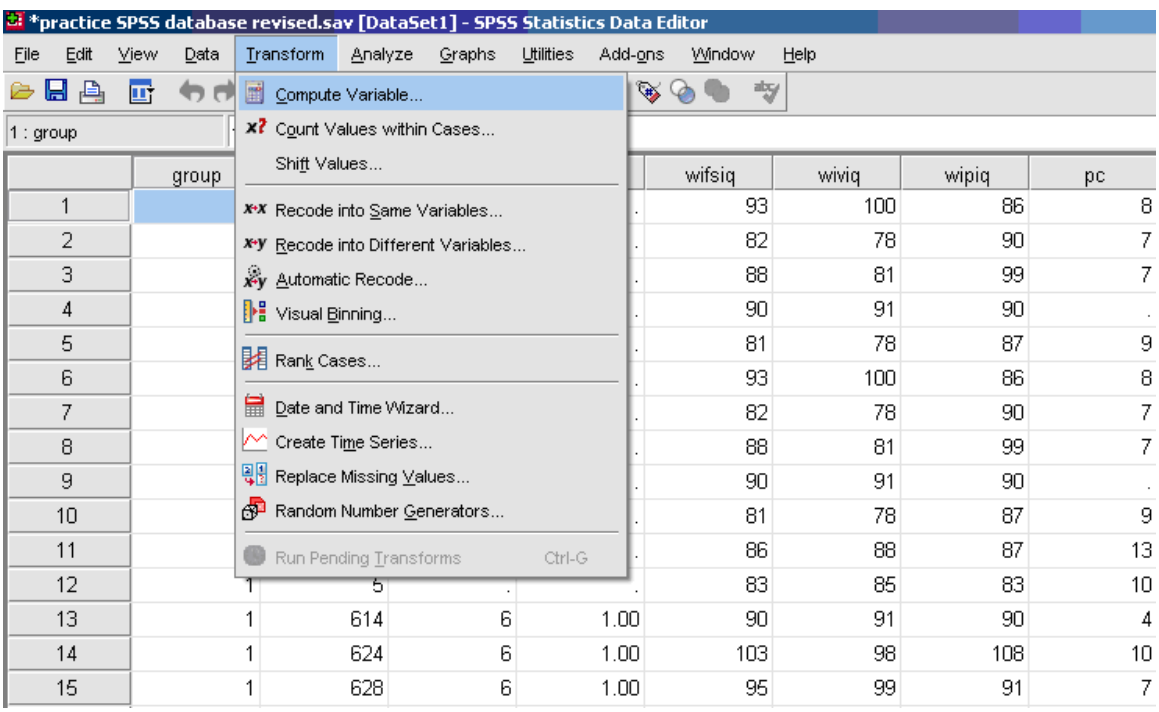
[30] http://cnx.org/content/m37276/latest/figure1.13.PNG/image

* OK
* You will be sent to the output window. Nothing in the output window is helpful. Go to the SPSS data screen by clicking on the data button bottom of screen. A new variable(s) will have been generated for you. This variable will be inserted into a formula so that you can have T scores.
* Variable view window

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Verbal IQ (Wechsler Verbal Intelligence 3)	1182	46	129	77.97	13.661
Valid N (listwise)	1182				

Table 3: Descriptive Statistics

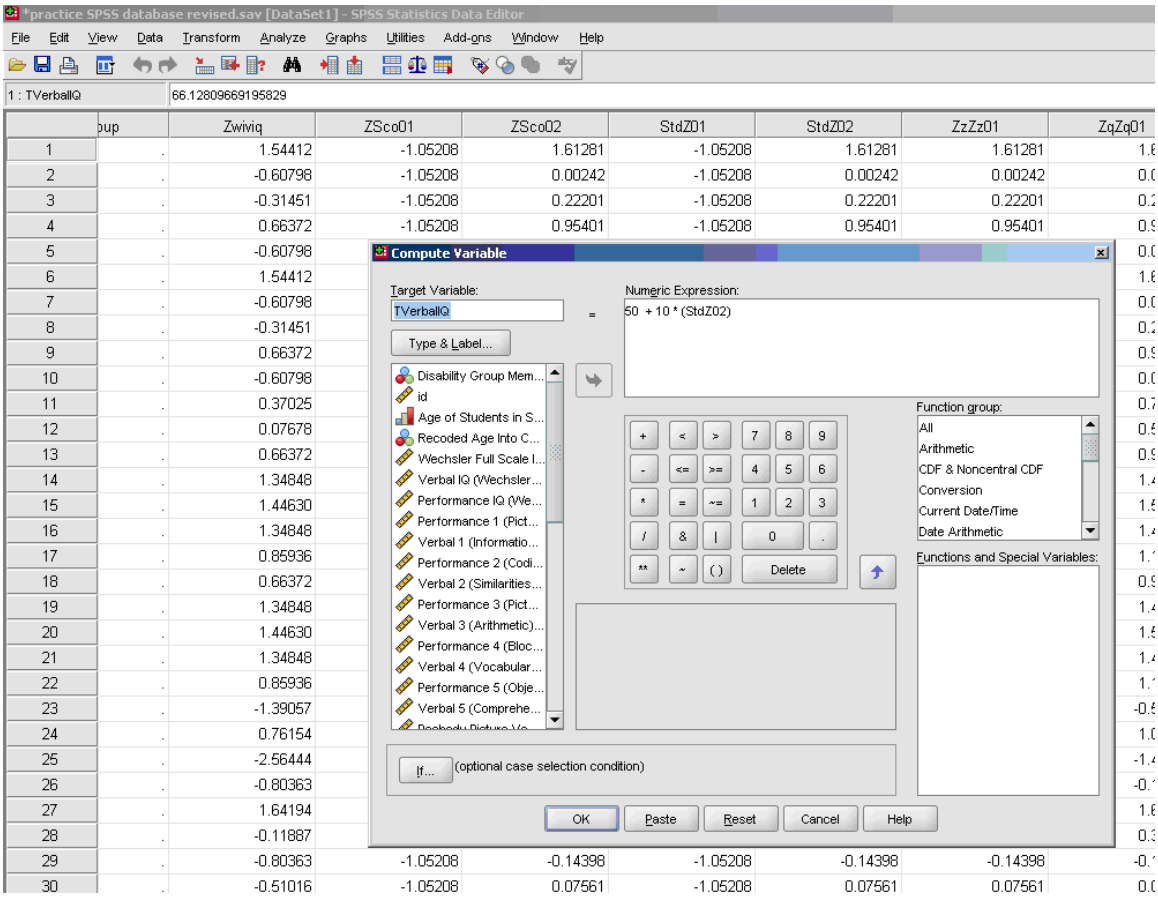
√ Create a new variable for your T score variable
* Data view window
* Transform
* Compute Variable



[31]

* Name your target variable the name you just generated for your T score variable
* In the numeric expression window, type:
* 50 + (10 x [name of the z-score variable generated by the computer earlier])

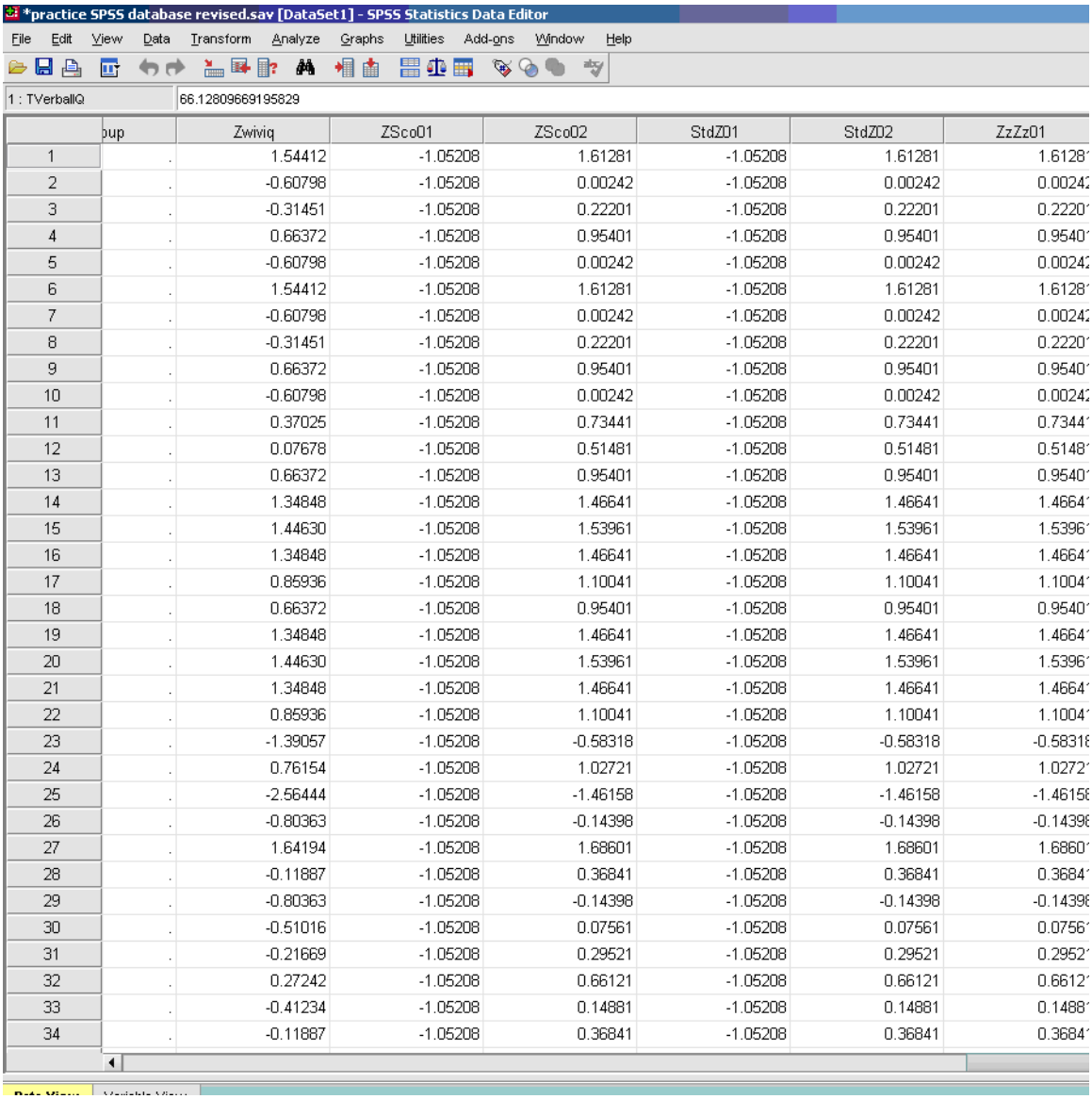
[31] http://cnx.org/content/m37276/latest/figure1.14.PNG/image



[32]

* OK

[32] http://cnx.org/content/m37276/latest/figure1.15.PNG/image



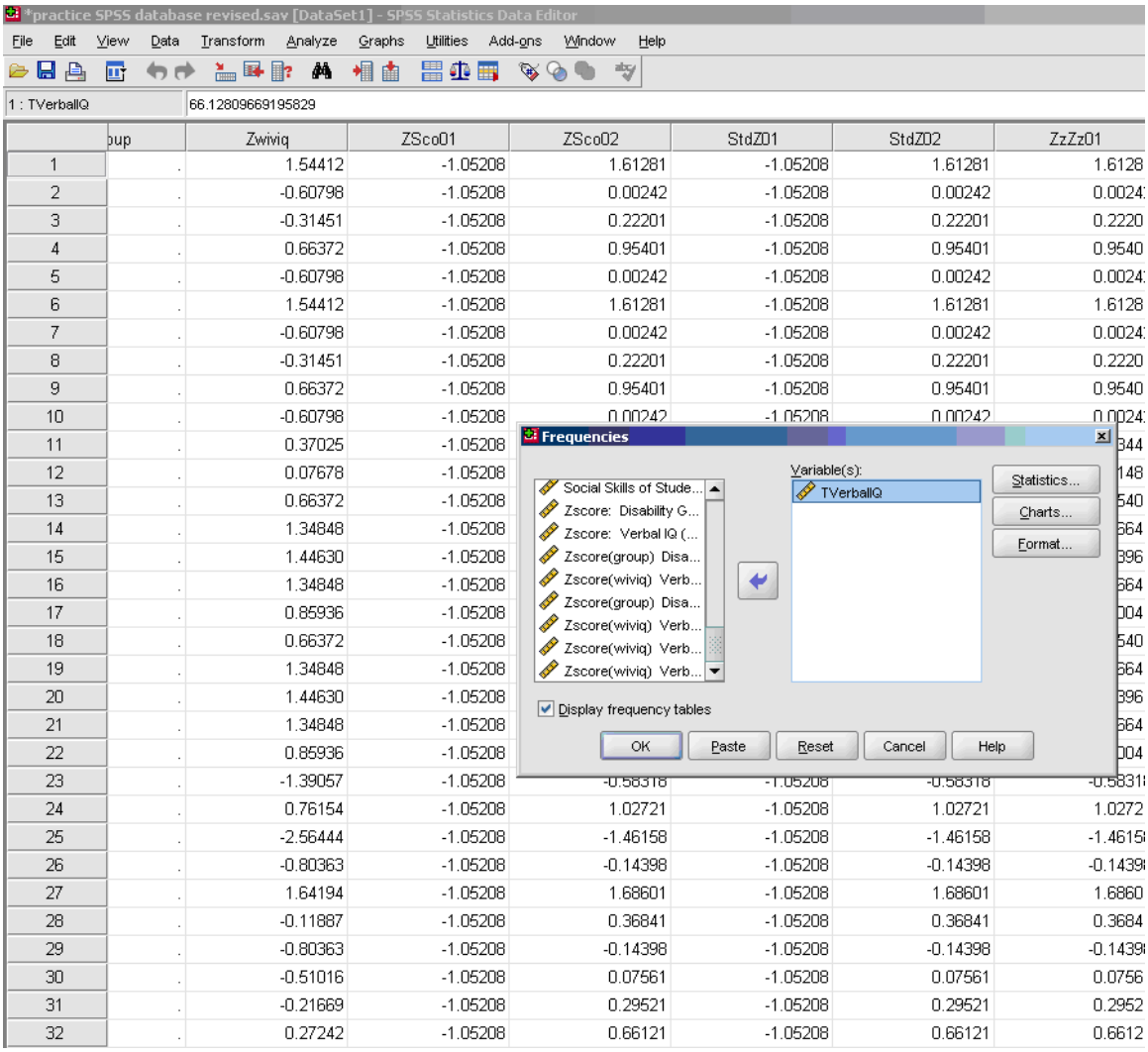
	oup	Zwiviq	ZSco01	ZSco02	StdZ01	StdZ02	ZzZz01
1	.	1.54412	-1.05208	1.61281	-1.05208	1.61281	1.6128
2	.	-0.60798	-1.05208	0.00242	-1.05208	0.00242	0.0024
3	.	-0.31451	-1.05208	0.22201	-1.05208	0.22201	0.2220
4	.	0.66372	-1.05208	0.95401	-1.05208	0.95401	0.9540
5	.	-0.60798	-1.05208	0.00242	-1.05208	0.00242	0.0024
6	.	1.54412	-1.05208	1.61281	-1.05208	1.61281	1.6128
7	.	-0.60798	-1.05208	0.00242	-1.05208	0.00242	0.0024
8	.	-0.31451	-1.05208	0.22201	-1.05208	0.22201	0.2220
9	.	0.66372	-1.05208	0.95401	-1.05208	0.95401	0.9540
10	.	-0.60798	-1.05208	0.00242	-1.05208	0.00242	0.0024
11	.	0.37025	-1.05208	0.73441	-1.05208	0.73441	0.7344
12	.	0.07678	-1.05208	0.51481	-1.05208	0.51481	0.5148
13	.	0.66372	-1.05208	0.95401	-1.05208	0.95401	0.9540
14	.	1.34848	-1.05208	1.46641	-1.05208	1.46641	1.4664
15	.	1.44630	-1.05208	1.53961	-1.05208	1.53961	1.5396
16	.	1.34848	-1.05208	1.46641	-1.05208	1.46641	1.4664
17	.	0.85936	-1.05208	1.10041	-1.05208	1.10041	1.1004
18	.	0.66372	-1.05208	0.95401	-1.05208	0.95401	0.9540
19	.	1.34848	-1.05208	1.46641	-1.05208	1.46641	1.4664
20	.	1.44630	-1.05208	1.53961	-1.05208	1.53961	1.5396
21	.	1.34848	-1.05208	1.46641	-1.05208	1.46641	1.4664
22	.	0.85936	-1.05208	1.10041	-1.05208	1.10041	1.1004
23	.	-1.39057	-1.05208	-0.58318	-1.05208	-0.58318	-0.5831
24	.	0.76154	-1.05208	1.02721	-1.05208	1.02721	1.0272
25	.	-2.56444	-1.05208	-1.46158	-1.05208	-1.46158	-1.4615
26	.	-0.80363	-1.05208	-0.14398	-1.05208	-0.14398	-0.1439
27	.	1.64194	-1.05208	1.68601	-1.05208	1.68601	1.6860
28	.	-0.11887	-1.05208	0.36841	-1.05208	0.36841	0.3684
29	.	-0.80363	-1.05208	-0.14398	-1.05208	-0.14398	-0.1439
30	.	-0.51016	-1.05208	0.07561	-1.05208	0.07561	0.0756
31	.	-0.21669	-1.05208	0.29521	-1.05208	0.29521	0.2952
32	.	0.27242	-1.05208	0.66121	-1.05208	0.66121	0.6612
33	.	-0.41234	-1.05208	0.14881	-1.05208	0.14881	0.1488
34	.	-0.11887	-1.05208	0.36841	-1.05208	0.36841	0.3684

[33]

* Respond yes to change existing variable
* You may be sent to the output screen. Nothing there is helpful.
* Go to data button and view your new T score variable.
* To get this information in a usable output form, do the following:

√ Analyze
* Descriptive Statistics
* Frequencies
* Move over the newly created T score variable
* Make sure the frequencies box is checked.

[33] http://cnx.org/content/m37276/latest/figure1.16.PNG/image



34

* OK
* Copy or cut the frequency table for this T score variable and carry it into WORD. Delete any irrelevant information.

3 Writing Up Your Statistics

So, how do you "write up" your Research Questions and your Results? Schuler W. Huck (2000) in his seminal book entitled, *Reading Statistics and Research*, points to the importance of your audience understanding and making sense of your research in written form. Huck further states:

3.1

This book is designed to help people *decipher* what researchers are trying to communicate in the written or oral summaries of their investigations. Here, the goal is simply to distill meaning from the words, symbols, tables, and figures included in the research report. To be competent in this arena, one must not only be

[34] http://cnx.org/content/m37276/latest/figure1.17.PNG/image

able to decipher what's presented but also to "fill in the holes"; this is the case because researchers typically assume that those receiving the research report are familiar with unmentioned details of the research process and statistical treatment of data.

A Note from the Editors

Researchers and Professors John Slate and Ana Rojas-LeBouef understand this critical issue, so often neglected or not addressed by other authors and researchers. They point to the importance of doctoral students "writing up their statistics" in a way that others can understand your reporting and as importantly, interpret the meaning of your significant findings and implications for the preparation and practice of educational leadership. Slate and LeBouef provide you with a model for "writing up your descriptive statistics."

Click here to view: *Writing Up Your Descriptive Staistics*[35]

4 References

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erbaum

Hyperstats Online Statistics Textbook. (n.d.) Retrieved from http://davidmlane.com/hyperstat/[36]

Kurtosis. (n.d.). Definition. Retrieved from http://www.statistics.com/index.php?page=glossary&term_id=326[37]

Kurtosis. (n.d.). *Definition of normality*. Retrieved from http://www.statsoft.com/textbook/basic-statistics/#Descriptive%

Onwuegbuzie, A. J., & Daniel, L. G. (2002). Uses and misuses of the correlation coefficient. *Research in the Schools, 9*(1), 73-90.

Skewness. (n.d.) Retrieved from http://www.statistics.com/index.php?page=glossary&term_id=356[39]

Skewness. (n.d.). *Definition of normality*. Retrieved from http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb[40]

StatSoft, Inc. (2011). *Electronic statistics textbook*. Tulsa, OK: StatSoft. WEB: http://www.statsoft.com/textbook/[41]

[35] http://cnx.org/content/col11299/latest/
[36] http://davidmlane.com/hyperstat/
[37] http://www.statistics.com/index.php?page=glossary&term_id=326
[38] http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb
[39] http://www.statistics.com/index.php?page=glossary&term_id=356
[40] http://www.statsoft.com/textbook/basic-statistics/#Descriptive%20statisticsb
[41] http://www.statsoft.com/textbook/