This tutorial will show you how to use SPSS version 12.0 to perform exploratory data analysis and descriptive statistics. You will use SPSS to create histograms, frequency distributions, stem and leaf plots, Tukey box plots, calculate the standard measures of central tendency (mean, median, and mode), calculate the standard measures of dispersion (range, semi-interquartile range, and standard deviation / variance), and calculate measures of kurtosis and skewness. This tutorial assumes that you have:

- Downloaded the standard class <u>data set</u> (click on the link and save the data file)
- Started SPSS (click on Start | Programs | SPSS for Windows | SPSS 12.0 for Windows)
- Loaded the standard data set

The Frequency Command

The frequencies command can be used to determine quartiles, percentiles, measures of central tendency (mean, median, and mode), measures of dispersion (range, standard deviation, variance, minimum and maximum), measures of kurtosis and skewness, and create histograms. The command is found at Analyze | Descriptive Statistics | Frequencies (this is shorthand for clicking on the Analyze menu item at the top of the window, and then clicking on Descriptive Statistics from the drop down menu, and Frequencies from the pop up menu.):

🗰 216d	ata.sav - SPS	SS Data Ed	litor					
File Edit	View Data	Transform	Analyze	Graphs	Utilities	Add-ons	: Wind	ow ⊦
			Report Descri	ts ptive Stat	istics 🕨	Frequ	പെറ്റം encies	
10	older 2	younger	Genera Mixed Correla	al Linear N Models ate	1odel ►	Explor Cross Ratio.	e tabs	
11 12 13	4 1 0		Regres Logline Classif	ssion ear fv	+ + +	0 ENG 5 HIST	LISH	DO(
14 15	0		Data F Scale	, Reduction))	0 ENG	LISH H	DO(DO(
16 17 18	0		Surviv	al e Respon	se +	0 MAT	H	DOC DOC DOC



The frequencies dialog box will appear:

Select the variable(s) that you want to analyze by clicking on it in the left hand pane of the frequencies dialog box. Then click on the arrow button to move the variable into the Variables pane:

Frequencies			×
 Number of Younger Grade Point Averaç Predicted Points in Major If Not Psychc Pet [pet] Area of Interest [are Intend To Get PhD Section [section] Display frequency tables 		<u>V</u> ariable(s):	OK <u>P</u> aste <u>R</u> eset Cancel Help
	Statistic	s <u>C</u> harts <u>F</u> ormat	

Be sure to select "Display frequency tables" if you want a frequency distribution. Specify which statistics you want to perform by clicking on the Statistics button. The Statistics dialog box will appear:

Frequencies: Statistics	
Percentile Values Quartiles Quartiles Cut points for: 10 equal groups Percentile(s): Add Change Remove	Central Tendency <u>Mean</u> Median <u>Mode</u> <u>S</u> um Values are group midpoints
Dispersion Std. deviation Minimum Variance Maximum Range S.E. mean	Distribution Ske <u>w</u> ness Kurtosis

From the statistics dialog box, click on the desired statistics that you want to perform. To calculate a given percentile, click in the box to the left of percentile(s). Type in the desired percentile and click on the Add button. When you have selected all the desired statistics (e.g. mean, median, mode, standard deviation, variance, ragne, etc.), click on the Continue button.

Specify which chart you want to display by clicking on the Chart button. The chart dialog box will appear:

Frequencies: Charts	
Chart Type None Bar charts Pie charts Histograms: With normal curve	Continue Cancel Help
Chart Values C Erequencies C Pe	r <u>c</u> entages

Click on the desired chart (usually Histogram) and click on the Continue button. Click on OK in the frequencies dialog box. The SPSS Output Viewer will appear. In the SPSS Output Viewer, you will see the requested statistics and chart. This is what the Statistics output looks like. It lists the requested measures of central tendency, measures of dispersion, measures of skewness and kurtosis, and the quartiles and percentiles.

	Statistics	
Number of Older Sibling	gs	
N	Valid	46
	Missing	0
Mean		1.26
Median		1.00
Mode		0
Std. Deviation		1.255
Variance		1.575
Skewness		.678
Std. Error of Skewness		.350
Kurtosis		543
Std. Error of Kurtosis		.688
Range		4
Percentiles	25	.00
	50	1.00
	75	2.00

The output has two columns. The left column names the statistic and the right column gives the value of the statistic. For example, the mean of this data is 1.26 (since your data set may be different, you may get a different value.)

The skewness measure is greater than 0 when the distribution is skewed. The kurtosis measure is 0 for a normal distribution. Positive values imply a leptokurtic distribution, while negative values imply a platykurtic distribution.

If you scroll down, you will see the frequency distributions. Number of Older Siblings

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	17	37.0	37.0	37.0
	1	11	23.9	23.9	60.9
	2	10	21.7	21.7	82.6
	3	5	10.9	10.9	93.5
	4	3	6.5	6.5	100.0
	Total	46	100.0	100.0	

If you scroll down, you will see the histogram (or whatever chart you requested.) Histogram



The Descriptives Command

The descriptives command can be used to determine measures of central tendency (mean), measures of dispersion (range, standard deviation, variance, minimum and maximum), and measures of kurtosis and skewness. The command is found at Analyze | Descriptive Statistics | Descriptives (this is shorthand for clicking on the Analyze menu item at the top of the window, and then clicking on Descriptive Statistics from the drop down menu, and Descriptives from the pop up menu.):

iiii 216a	lata.sav - SP	SS Data Ed	litor					
File Edit	: View Data	Transform	Analyze	Graphs	Utilities	Add-ons	Wind	low H
		la 🔚	Repor	ts		ki ee L	പര	1
		Descri	ptive Stat	istics 🔸	Frequ	encies	.	
1 : older 0			Compa	are Means	• •	Descri	iptives	
	older	vounder	Gener	al Linear N	1odel 🕨	Explor	e	
10	2	youngoi	Mixed	Mixed Models			tabs	
	2		Correl	Correlate				
	4		Regre	ssion) `			
12	! 1		Logline	ar	+	0 ENG	LISH	DOC
13	0		Classif	v	+	5 HIST	ORY	DOC
14	. 0		Data F	eduction	+	0 ENG	LISH	DOC
15	0		Scale		+	I0 MAT	Ή	DOC
16	i 0		Nonpa	rametric 1	Fests 🕨	0 MAT	Ή	DOC
17	0		Surviv	al	•	0 ART		DOC
18	3		Multip	e Respon	se 🕨 🕨	0 MAT	Ή	DOC

The descriptives dialog box will appear:

Descriptives	
Number of Older Sib Number of Younger Grade Point Average Predicted Points in (Section [section] I would rather stay a One of my favorite p I live a fast paced lif	OK <u>P</u> aste <u>R</u> eset Cancel Help
Save standardized values as variables	Options

Select the variable(s) that you want to analyze by clicking on it in the left hand pane of the descriptives dialog box. Then click on the arrow button to move the variable into the Variables pane:

Descriptives	
 Number of Younger Grade Point Average Predicted Points in (Section [section] I would rather stay a One of my favorite p I live a fast paced lif I hardly ever sit arou 	OK <u>P</u> aste <u>R</u> eset Cancel Help
Save standardized values as variables	Options

Specify which statistics you want to perform by clicking on the Options button. The Options dialog box will appear:

Descriptives: Op	otions	
	□ <u>S</u> um	Continue
Std. deviation	Minimum	Cancel Help
☐ <u>V</u> anance ☐ <u>R</u> ange	I♥ Ma <u>xi</u> mum □ S. <u>E</u> . mean	
Distribution		
Kurtosis	Ske <u>w</u> ness	
Display Order		
Variable list		
C <u>Alphabetic</u>		
C Ascending mea	ans	
C Descending me	eans	

Select the statistics that you want by clicking on them (e.g. mean, standard deviation, variance, range, minimum, etc.). Then click on the Continue button. Click on the OK button in the Descriptives dialog box. The SPSS Output Viewer will appear with your results in it. The following is an example of the output: Descriptive Statistics

	N	Range	Minimu	Maximu	Mean	Std.	Varianc	Skev	vness	Kur	tosis
	Statistic	Std. Error	Statistic	Std. Error							
Number of Older Siblings	46	4	0	4	1.26	1.255	1.575	.678	.350	543	.688
Valid N (listwise)	46										

The output gives the values of the requested statistics.

The Explore Command

The explore command can be used to determine measures of central tendency (mean and median), measures of dispersion (range, interquartile range, standard deviation, variance, minimum and maximum), measures of kurtosis and skewness, and prepare histograms, stem and leaf plots, and Tukey box plots. The command is found at Analyze | Descriptive Statistics | Explore:

🗰 216data.say - SPSS Data Editor								
File Edit	View Data	Transform	Analyze	Graphs	Utilities	Add-ons	Windo	ow ⊦
□ □ □ □ □ □ □ 1 : older			Report Descri Compa	ts ptive Stat are Means	istics 🕨	Frequer Descrip	ncies	
10	older 2	younger	Gener Mixed Correl	al Linear N Models ate	4odel ► ►	Crossta Ratio	bs	
12 13	1		Regres Logline Classif	ssion ear fy	• • •	0 ENGL	ISH DRY	DO(
14 15	0		Data R Scale	Reduction)	0 ENGL 10 MATH	ISH	DO(DO(
16 17 18	0		Nonpa Surviv Multipl	rametric 1 al e Respon	rests + + se +	0 MATH 0 ART 0 MATH		DO(DO(DO(

The explore dialog box will appear:

Explore		X
Number of Older Si Number of Younger Grade Point Average Predicted Points in Maior If Not Psychol	Dependent List:	OK <u>B</u> aste <u>R</u> eset
Area of Interest [are Action of Content of		Cancel Help
Display	Label <u>C</u> ases by:	
	Statistics Plots Options	-

Select the variable(s) that you want to analyze by clicking on it in the left hand pane of the explore dialog box. Then click on the top arrow button to move the variable into the Dependent List:

Explore Explore		
Number of Younger	Dependent List:	ок
Grade Point Average Predicted Points in		<u>P</u> aste
Ajor If Not Psych		Reset
A Pet [pet]	Factor List:	Cancel
Area of Interest [are		
Section [section]		Help
I would rather stay :	Label Cases by:	
🛊 One of my favorite 🔽		
Display		
Both C Statistics C Plots	Statistics Plots Options	

Specify which plots you want to prepare by clicking on the Plots button. The Plots dialog box will appear:

Explore: Plots		
Boxplots • Eactor levels together • Dependents together • None	Descriptive <u>Stem-and-leaf</u> <u>Histogram</u>	Continue Cancel Help
Nomality plots with tests Spread vs. Level with Leve Nong Power estimation Iransformed Power: Untransformed	ne Test Natural log	

Select the plots that you want by clicking on them (e.g. Stem-and-leaf and histogram). Then click on the Continue button. Click on the OK button in the Explore dialog box. The SPSS Output Viewer will appear with your results in it. The following is an example of the output for the descriptive statistics:

Descriptives							
			Statistic	Std. Error			
Number of Older Siblings Mean 95% Confidence Interval for Mean 5% Trimmed Mean Median	Mean		1.26	.185			
	95% Confidence	Lower Bound	.89				
	Interval for Mean	Upper Bound	1.63				
	5% Trimmed Mean		1.18				
	Median		1.00				
Variance		1.575					
Std. Mini Maxi Ran Inter Skev Kurt	Std. Deviation		1.255				
	Minimum		0				
	Maximum		4				
	Range		4				
	Interquartile Range		2				
	Skewness		.678	.350			
	Kurtosis		543	.688			

The output gives the values of the requested statistics. If you scroll down, you will see the requested plots: **Histogram**



Number of Older Siblings Stem-and-Leaf Plot

Stem & Leaf Frequency 17.00 ο. .00 ο. 11.00 1. 000000000000 .00 1. 2 . 000000000 10.00 2. .00 3. 00000 5.00 .00 з. 3.00 4. 000 Stem width: 1 Each leaf: 1 case(s)



Number of Older Siblings

The Tukey box plot shows the first (bottom of box) and third (top of box) quartiles (equivalently the 25th and 75th percentiles), the median (the horizontal line in the box), the range (excluding outliers and extreme scores) (the "whiskers" or lines that extend from the box show the range), outliers (a circle represents each outlier -- the number next to the outlier is the observation number.) An outlier is defined as a score that is between 1.5 and 3 box lengths away from the upper or lower edge of the box (remember the box represents the middle 50 percent of the scores). An extreme score is defined as a score that is greater than 3 box lengths away from the upper or lower edge of the box.