Descriptive Statistics in SPSS

- For simple description of interval or ratio variables (items measured on a scale) use the descriptives command
- For more complex description of interval or ratio variables use the explore command

✓ For interval or ratio variables (i.e., "scales"), use the descriptives sub-command

✓ on the "Analyze > Descriptive Statistics" menu:

٦	Analyze	Direct <u>M</u> arketing	Graphs	Utilities	Add- <u>o</u> ns	Win	
~	Rep	orts		A4 1			
	D <u>e</u> s	criptive Statistics	*	123 Freque	ncies	7	
	Ta <u>b</u>	les		Descri	ptives		
siz	Con	npare Means	•	A Explore			
9	Gen	ieral Linear Model	•	Crosst	abs	5-7	
9	Gen	ierali <u>z</u> ed Linear Mod	iels 🕨	Ratio		5-6	
2	Mixe	ed Models	- F.			p-1	
9	Con	relate	•		ots	p-5	
9	Reg	ression		🛃 <u>Q</u> -Q PI	ots	5-5	

- This opens a dialog box where you can select the variable that you want descriptive statistics on.
- Make sure that what you are selecting is actually an interval or ratio variable.
- Because all of the data are entered as numbers, SPSS will actually calculate descriptive statistics on any of these variables; but these results are only meaningful for interval or ratio variables

- This opens a dialog box where you can select the variable that you want descriptive statistics on.
- Make sure that what you are selecting is actually an interval or ratio variable.
- Because all of the data are entered as numbers, SPSS will actually calculate descriptive statistics on any of these variables; but these results are only meaningful for interval or ratio variables



- As usual, select the variable(s) that you are interested in from the left-hand column, and move them to the right-hand column.
- ✓ Use the "Options" button to select the specific descriptive statistics that you are interested in.
- Usually, good choices include the mean, standard deviation, maximum, and minimum.
- If you are concerned about the impact of outliers on your data, a measure of skewness may also be appropriate.



✓ The output from this procedure looks like this:

Descriptive Statistics							
	N Minimum Maximum Mean Std. Deviation Skewness				ness		
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
Age in years	1048	45	86	62.46	9.113	.301	.076
Total treatment and rehabilitation costs in thousands	1048	.00	200.51	33.7641	27.31266	1.692	.076
Length of stay for rehabilitation	787	0	36	16.39	12.565	.340	.087
Valid N (listwise)	787						

 This shows you the specific results for each variable that you entered into the analysis.

 It is possible to get a table that gives you basic descriptive statistics for many variables simultaneously, just by moving them all at once from the left-hand to the right-hand list in the main dialog box.

 Finally, to get more complex descriptive results for an interval or ratio variable, use the explore command in the "Analyze/Descriptive Statistics" sub-menu.

<u>A</u> nalyze	Direct <u>M</u> arketing	Graphs	<u>U</u> tilities	Add-ons W		
Rep	orts	•	44 0			
Des	criptive Statistics	•	123 Frequencies			
Ta <u>b</u>	les	•	Descriptives			
Con	npare Means		A Explore			
<u>G</u> en	eral Linear Model	*	Crosstabs			
Gen	eralized Linear Mode	ls 🕨				
Mi <u>x</u> e	d Models	•				
Con	relate		<u>P</u> -P Plo	ots		
Regression			<u>Q</u> -Q Pl	ots		

Descriptive statistics

✓ The following dialog box will appear:

Q	Dependent List:	Statistics
Hospital ID [hosp]	*	Plots
Patient ID [patid]	Eactor List	Options
Age in years [age]		Bootstrap
Age category [ag		
Physically active [Label <u>C</u> ases by:	
🂑 Obesity [obesity] 🛛 🔽		
Display		
Both O Statistics O P	lots	

 The variable that you are interested in goes into the "dependent" list.

✓ For now, we won't use the other lists, but we will use the "plots" button to select additional options.

Explore	Explore: Plots Boxplots Exclore levels together	Descriptive	Atatistics
Rankir Rankir	© <u>D</u> ependents together ◎ <u>N</u> one	🗾 <u>H</u> istogram	Options
Recod Recod Treatm	Normality plots with tests Spread vs Level with Leven	e Test	lootstrap
Length	 None Power estimation 		
Display <u> B</u> oth	© <u>Transformed</u> Power: N © <u>Untransformed</u>	latural log 🔹 👻	
	Continue	Help	

 The "plots" dialog box lets you select options for graphical display of the data, including a stem-and-leaf plot like this:

Age in years	Stem-and	l-Leaf Plot
Frequency	Stem &	Leaf
3.00	4.	555
23.00	4.	6666666666777777777777
35.00	4.	999999999999999999999999999999999999999
58.00	5.	000000000000000000000000001111111111111
94.00	5.	222222222222222222222222222222222222222
58.00	5.	444444444444444444444444444455555555555
81.00	5.	666666666666666666666666666666666777777
75.00	5.	999999999999999999999999999999999999999
79.00	6.	000000000000000000000000000000001111111
88.00	6.	222222222222222222222222222222222222223333
74.00	6.	444444444444444444444444444444455555555
76.00	6.	666666666666666666666666666666666666666
67.00	6.	8883838383838383838383838383838383838383
52.00	7.	000000000000000000001111111111111111111
40.00	7.	222222222222222223333333333333333333333
43.00	7.	444444444444444555555555555555555555555
32.00	7.	6666666666666666666777777777777
30.00	7.	999999999999999999999999999999999999999
15.00	8.	00000011111111
17.00	8.	22222222333333
6.00	8.	444455
2.00	8.	66
Stem width:	10	
Each leaf:	1 0	case (s)

- ✓ Or a boxplot like this:
- The box plot is a standardized way of displaying the distribution of data based on the five number summary:
- > Minimum
- First quartile
- Median
- > Third quartile
- Maximum.



✓ A histogram like this:



Age in years

Descriptive statistics

Or a plot to test whether the data are normally distributed, like this:

 This result is showing a slight deviation from normality, based on the difference between the actual data points and the theoretical line they should fit.



 The "explore" command also gives you descriptive statistics for each variable, with more extensive results (median, interquartile range, etc.) than the "descriptives" command.

	Descriptives		
		Statistic	Std. Error
Age in years	Mean	62.46	.281
	95% Confidence Interval Lower Bound	61.91	
	for Mean Upper Bound	63.02	6
	5% Trimmed Mean	62.25	
	Median	62.00	
	Variance	83.042	
	Std. Deviation	9.113	ан. С
	Minimum	45	
	Maximum	86	5
	Range	41	
	Interquartile Range	14	
	Skewness	.301	.076
	Kurtosis	659	.151

- One additional feature of the "explore" command is the ability to get separate descriptive results for different sub-groups within your dataset.
- The way to do this is by grouping your data based on some additional variable, which is treated as a "factor" for breaking down the analyses.
- Go back to the main dialog box to add one of your nominal variables as a "factor" (the grouping variable always has to be nominal).



Descriptive statistics

 After you add this factor, the results will show you two separate sets of statistics and plots.

Age i

Descriptives

	Smoke	r		Statistic	Std. Error
n years	No	Mean		62.42	.319
		95% Confidence Interval for Mean	Lower Bound	61.79	
			Upper Bound	63.04	2 0
		5% Trimmed Mean		62.20	
	8	Median		62.00	
	24	Variance		83.822	
	8	Std. Deviation		9.155	
	04	Minimum		45	
	8	Maximum		86	6 3
	24	Range		41	
		Interquartile Range		14	
		Skewness		.305	.085
		Kurtosis		670	.170
	Yes	Mean		62.64	.602
	8	95% Confidence Interval for Mean	Lower Bound	61.45	6
			Upper Bound	63.83	
		5% Trimmed Mean		62.43	
		Median		62.00	
	2	Variance		80.467	
	с.,	Std. Deviation		8.970	
	8	Minimum		46	
	94	Maximum		84	
	8	Range		38	
	94	Interquartile Range		13	
	8	Skewness		.290	.163
	0.5	Kurtosis		602	.325

Descriptive statistics

Histograms



Age in years Stem-and-Leaf Plot for



Stem-and-Leaf Plots

smoker= No		
Frequency	Stem &	Leaf
.00	4.	
50.00	4.	566677777888888888999999
138.00	5.	000000000111111111112222222222222222222
151.00	5.	555555555555666666666666666666677777777
154.00	6.	000000000000111111111111122222222222222
150.00	6.	5555555555555566666666666666666666777777
78.00	7.	0000000111111112222222333333344444
75.00	7.	55555555566666666677777888888999999
26.00	8.	01111222334
4.00	8.	56
Stem width:	10	
Each leaf:	2 c	ase(s)

