# Descriptive Statistics in SPSS

# Running an Analysis

 An icon next to each variable provides information about data type and level of measurement.

	Numeric	String	Date	Time
Scale (Continuous)		n/a		
Ordinal				
Nominal		a		

# 3

# Level of Measurement

 Different summary measures are appropriate for different types of data, depending on the level of measurement:

CategoricalScale

# Categorical

 Data with a limited number of distinct values or categories (for example, gender or marital status).

✓ Also referred to as qualitative data.

 Categorical variables can be string (alphanumeric) data or numeric variables that use numeric codes to represent categories

(for example, 0 =Unmarried and 1 =Married).

# Categorical

✓ There are two basic types of categorical data:

 Nominal. Categorical data where there is no inherent order to the categories.

For example, a job category of sales is not higher or lower than a job category of marketing or research.

# Categorical

✓ There are two basic types of categorical data:

- Ordinal. Categorical data where there is a meaningful order of categories, but there is not a measurable distance between categories.
- For example, there is an order to the values high, medium, and low, but the "distance" between the values cannot be calculated.

# Scale

- Data measured on an interval or ratio scale, where the data values indicate both the order of values and the distance between values.
- For example, a salary of \$72,195 is higher than a salary of \$52,398, and the distance between the two values is \$19,797. Also referred to as quantitative or continuous data.

- For categorical data, the most typical summary measure is the number or percentage of cases in each category. The mode is the category with the greatest number of cases.
- For ordinal data, the median (the value at which half of the cases fall above and below) may also be a useful summary measure if there is a large number of categories.

 The Frequencies procedure produces frequency tables that display both the number and percentage of cases for each observed value of a variable.

From the menus choose:
 Analyze > Descriptive Statistics > Frequencies...

- Right-click any variable and choose Variable Information.
- Click the down arrow on the Value labels
   drop-down list.
   All of the defined
   value labels for the
   variable are displayed.



✓ Select Owns PDA [ownpda] and Owns TV [owntv] and move them into the Variable(s) list.



- ✓ Select Owns PDA [ownpda] and Owns TV [owntv] and move them into the Variable(s) list.
- Click OK to run the procedure.

📅 Output1 - Viewer 📃 🔍											
<u>File E</u> dit <u>V</u> iew <u>E</u>	<u>)</u> ata	<u>T</u> ransform	Inser	rt F <u>o</u> rmat	<u>A</u> nalyze	<u>G</u> raphs	<u>U</u> tilities	Add- <u>o</u> ns	s <u>W</u> indow	<u>H</u> elp	
tput Frequencies Ittle Notes	F	requer	ıcy ˈ	Table	Own	is PDA					~
Title				Frequency	Per	cent	Valid Pe	rcent	Cumulativ Percent	'e	
	LE	Valid N	lo	5093	;	79.6		79.6	79	3.6	
	Ш	Y	es	1307	,	20.4		20.4	100	0.0	
	L L	Т	otal	6400	ı ,	100.0	1	00.0			
					Ow	ns TV					
	Ш			Frequency	/ Per	cent	Valid Per	rcent	Cumulativ Percent	'e	
	۱ľ	Valid N	lo	63	;	1.0		1.0	1	1.0	
		Y	es	6337	,	99.0		99.0	100	0.0	
	Ιl	Т	otal	6400	ı -	100.0	1	00.0			
<	<									>	•]

- The frequency tables are displayed in the Viewer window.
- The frequency tables reveal that only 20.4% of the people own PDAs, but almost everybody owns a TV (99.0%).
- These might not be interesting revelations, although it might be interesting to find out more about the small group of people who do not own televisions.

# Charts for Categorical Data

- ✓ You can graphically display the information in a frequency table with a bar chart or pie chart.
- Open the Frequencies dialog box again. (The two variables should still be selected.)
- ✓ You can use the Dialog Recall button on the toolbar to quickly return to recently used procedures.



# Charts for Categorical Data

#### ✓ Click Charts.

- Select Bar charts and then click Continue.
- Click OK in the main dialog box to run the procedure.

ÿ	🛿 Frequencies: Charts 🛛 🛛 🔀
	-Chart Type
	© None
	◙ <u>B</u> ar charts
	© <u>P</u> ie charts
	© <u>H</u> istograms:
	Show normal curve on histogram
	-Chart Values
	Interpretation ● Percentages
	Continue Cancel Help

# Charts for Categorical Data



There are many summary measures available for scale variables, including:

- Measures of central tendency. The most common measures of central tendency are the mean (arithmetic average) and median (value at which half the cases fall above and below).
- Measures of dispersion. Statistics that measure the amount of variation or spread in the data include the standard deviation, minimum, and maximum.

Open the Frequencies dialog box again.
Click Reset to clear any previous settings.

 Select Household income in thousands [income] and move it into the Variable(s) list

🔛 Frequencies 🛛 🔀
Income category in         Income category in         Income category in         Job satisfaction [job         Income category in ca

### ✓ Click Statistics.

Select Mean, Median,
 Std. deviation, Minimum,
 and Maximum.

🛃 Frequencies: Statistics	$\mathbf{X}$
Percentile Values	Central Tendency
🧾 Quartiles	📝 Mean
Cut points for: 10 equal groups	📝 Me <u>d</u> ian
Percentile(s):	Mode
Add	🔝 Sum
Change	
Remove	
	🔝 Values are group midpoints
Dispersion	Distribution
🗹 St̪d. deviation 🔽 Mi̯nimum	📃 Ske <u>w</u> ness
🔲 Variance 🛛 🗹 Maximum	📃 Kurtosis
🔲 Range 📃 S. <u>E</u> . mean	
Continue	Help

- ✓ Click Continue.
- Deselect Display frequency tables in the main dialog box.
- Frequency tables are usually not useful for scale variables since there may be almost as many distinct values as there are cases in the data file.

✓ Click OK to run the procedure.

 The Frequencies Statistics table is displayed in the Viewer window.

<b>11</b>	Dutpu	it1 - V	iewe	ſ								- 0	$\mathbf{X}$
Eile	<u>E</u> dit	<u>V</u> iew	<u>D</u> ata	Transform	Insert	F <u>o</u> rmat	<u>A</u> nalyze	<u>G</u> raphs	<u>U</u> tilities	Add- <u>o</u> ns	<u>W</u> indow	<u>H</u> elp	
tput Fre @@@@@@ Fre @@@@	Edit quencio Title Notes Statist Freque Title Notes Notes	view es ency Ta le wns PD wns TV es		Transform Housel N Mean Median Std. De Minimu Maximu	Insert nold inc viation m	Statist ome in th Valid Missir	<u>Analyze</u> tics nousands	69.43 69.43 45.00 78.718 9 1116	Utilities	Add- <u>o</u> ns	Window	Helb	
· III · IIII · IIII	Statist Histog	ics ram	<									>	>

✓ In this example, there is a large difference between the mean and the median.

The mean is almost 25,000 greater than the median, indicating that the values are not normally distributed. You can visually check the distribution with a histogram.

# Histograms for Scale Variables

- ✓ Open the Frequencies dialog box again.
- ✓ Click Charts.
- Select Histograms and With normal curve.
- Click Continue, and then click OK in the main dialog box to run the procedure.



## Histograms for Scale Variables



### Histograms for Scale Variables

- ✓ The majority of cases are clustered at the lower end of the scale, with most falling below 100,000.
- There are, however, a few cases in the 500,000 range and beyond (too few to even be visible without modifying the histogram).
- These high values for only a few cases have a significant effect on the mean but little or no effect on the median.
- So the median a better indicator of central tendency in this example.

# Descriptive statistics

- For more complex description of nominal variables use crosstabs.
- 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 a more detailed breakdown of nominal data, use the crosstabs command in the "Analyze/Descriptive Statistics" sub-menu.

This is appropriate if you want to consider two variables in combination.

- ✓ The "Crosstabs" dialog box lets you enter one variable (or more) as "rows" in a frequency table, and another variable as the "columns" in the same table.
- Use the "Cells" command to get percentages for each row and column:



		Case Process	sing Sumn	пагу		
		3	Cas	ses		
	Va	lid	Miss	sing	Total	
	Ν	Percent	N	Percent	Ν	Percent
Obesity * Gender	1048	100.0%	0	0.0%	1048	100.0%

#### **Obesity \* Gender Crosstabulation**

			Ger	e internet i	
			Male	Female	Total
Obesity	No	Count	406	394	800
		% within Obesity	50.8%	49.3%	100.0%
		% within Gender	76.7%	75.9%	76.3%
		% of Total	38.7%	37.6%	76.3%
-	Yes	Count	123	125	248
		% within Obesity	49.6%	50.4%	100.0%
		% within Gender	23.3%	24.1%	23.7%
		% of Total	11.7%	11.9%	23.7%
Total		Count	529	519	1048
		% within Obesity	50.5%	49.5%	100.0%
		% within Gender	100.0%	100.0%	100.0%
		% of Total	50.5%	49.5%	100.0%