#### Obtaining Summary Statistics with SPSS

Math 260

## Open the New York Travel Times data from Exercise 2.2

• File eg02-03.sav.

💼 eg02-	-03.sav [DataSet6] ·			
File	View Data Transfor	m Analyze	Graphs Ut	
🗁 🔲	兽 📴 🛧 🏓 🖁	= <b>!? /4</b>	▲ ■	
1 : minu	tes	10		
	minutes	var	var	
1	10			
2	30			
3	5			Your data should
4	25			have n=20 rows
5	40			
6	20			
7	10			
8	15			
9	30			
10	20			
11	15			
12	20			

#### **Explore Procedure**

• Select Analyze...Descriptive Statistics...Explore.



#### Complete the Explore Dialog Box



## The Statistics Dialog Box



## The Plots Dialog Box

"Factor levels together" allows comparing subgroups based on a factor.

conditions.



#### **Output: Case Processing Summary**

• Provides a quick check on sample size and missing values.

Case Processing Summary									
	Cases								
	Valid		Mis	sing	Total				
	N	Percent	Ν	Percent	Ν	Percent			
	20	100.0%	0	.0%	20	100.0%			

#### **Output: Descriptives**

Descriptives

	Statistic Std. Error
Mean	31.25 4.892
95% Confidence Lower Bound Interval for Mean Upper Bound	21.01 41.49 0ne-sample t- confidence ntervals will be used
5% Trimmed Mean	29.72 extensively later in
Median	22.50 the course.
Variance	478.618
Std. Deviation	21.877
Minimum	5
Maximum	85
Range	80
Interquartile Range <u>= IQR</u>	29
Skewness	1.040 .512
Kurtosis	.330 .992

#### **Output: Percentiles**

**Ignore:** This definition will often give percentiles markedly different from the definition in our text.

Percentiles

	Percentiles						
	5	10	25	50	75	90	95
Weighted Average(Definition 1)	5.25	10.00	15.00	22.50	43.75	64.50	84.00
Tukey's Hinges			15.00	22.50	42.50		

**Preferred:** Tukey's Hinges are very close to the Q1, median, and Q3 we have discussed in class.

**Footnote:** When *n* is odd, Tukey's hinges include the median as part of each half of the data set when finding the quartiles. Hence, Tukey's hinges may differ slightly from the quartiles as defined in our text when *n* is odd. They should agree when *n* is even. The median will always agree with our text's method.

#### **Output: Extreme Values**

	Extreme Values						
This just shows			Case Number	Value			
the most extreme	Highest	1	13	85			
at the high and		2	15	65			
are <b>not</b>		3	17	60			
necessarily		4	18	60			
outliers or event		5	20	45			
unusual values.	Lowest	1	3	5			
		2	7	10			
		3	1	10			
		4	16	15			
		5	14	15 <sup>a</sup>			

a. Only a partial list of cases with the value 15 are shown in the table of lower extremes.

**Note:** This section is used mostly to check for data errors. Use the boxplot to check for values that meet our definition of outliers (more than 1.5 IQRs from the box).

#### **Output: Stemplot**



**Note**: The way SPSS defines the stems and leaves is automatic. Unfortunately, this cannot be changed. In any case, stemplots should only be used for small data sets where they can reasonably be made by hand. For larger dataset, histograms are preferred.

#### Output: Boxplot



## Notes

- Histogram output is not shown here as it was discussed in the Intro to SPSS help sheet.
- Another example follows to illustrate subgroup analysis.
- Open the cars.sav file again and explore weights by country of origin.

# Find and Open the Cars.sav data set from my <u>SPSS data directory</u>

🖬 cars.s	av [DataSet	1] - SPSS Dai	ta Editor							×
File Edit	View Data	Transform Ar	halyze Graphs	: Utilities Ad	d-ons Window	/ Help				
🖻 🖬 d	9 🖽 🕈	🔶 🗽 🛛	斜情的	i 🖩 🗗 ا	🔊 🖗 🕽					
1 : mpg		18								
	mpg	engine	horse	weight	accel	year	origin	cylinder	filter_\$	~
1	18	307	130	3504	12	70	1	8	0	
2	15	350	165	3693	12	70	1	8	0	_
3	18	318	150	3436	11	70	1	8	0	
4	16	304	150	3433	12	70	1	8	0	
5	17	302	140	3449	11	70	1	8	0	
6	15	429	198	4341	10	70	1	8	0	
7	14	454	220	4354	9	70	1	8	0	
8	14	440	215	4312	9	70	1	8	0	
9	14	455	225	4425	10	70	1	8	0	
40	4.5	200	400	2050	0	70		0	0	

Click on the variable view to get more information about the variables from the labels.

#### Complete the Explore Dialog Boxes



#### **Explore** Output

Extensive output for each subgroup (not shown)

At right is the boxplot for weight by country of origin.

American cars were clearly heaviest.



## **Other Procedures**

- Analyze...Descriptive Statistics has other useful procedures for summary statistics:
  - Descriptives: extensive statistics if no subgroups or plots are needed.
  - Frequencies: frequency table and statistics, especially for discrete data (small number of possible values).
- Analyze...Compare Means...Means is also good for a concise summary of subgroups.
- Experiment and see what you prefer!
- Ask questions if you have problems.