

# Anesthesia dept

## SPSS-T-1

SPSS-tutorial-1

Abdulmajeed ahmed

2021-2022

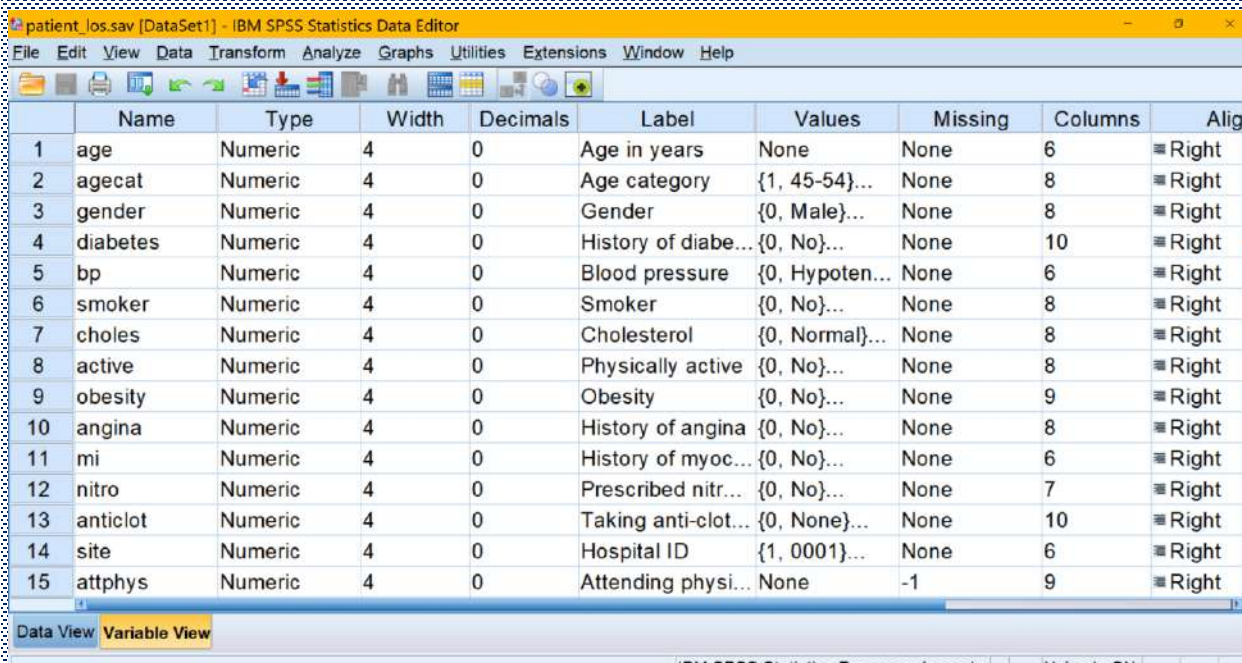


# Tutorial - patient\_los.sav

The data consists of drug information collected on 1000 patients used to perform frequency and descriptive statistics.

Variables in the data set are:

Age : age of each case  
Agecat : age category  
Gender : male or female  
Diabetes : is the case diabetes patient?  
Bp : Blood Pressure  
Smoker :  
Choles :cholesterol  
Bleed :does the patient bleeding  
Result :operation result  
Cost :operation cost  
There are more columns inside database.



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align
1	age	Numeric	4	0	Age in years	None	None	6	Right
2	agecat	Numeric	4	0	Age category	{1, 45-54}...	None	8	Right
3	gender	Numeric	4	0	Gender	{0, Male}...	None	8	Right
4	diabetes	Numeric	4	0	History of diabe...	{0, No}...	None	10	Right
5	bp	Numeric	4	0	Blood pressure	{0, Hypoten...	None	6	Right
6	smoker	Numeric	4	0	Smoker	{0, No}...	None	8	Right
7	choles	Numeric	4	0	Cholesterol	{0, Normal}...	None	8	Right
8	active	Numeric	4	0	Physically active	{0, No}...	None	8	Right
9	obesity	Numeric	4	0	Obesity	{0, No}...	None	9	Right
10	angina	Numeric	4	0	History of angina	{0, No}...	None	8	Right
11	mi	Numeric	4	0	History of myoc...	{0, No}...	None	6	Right
12	nitro	Numeric	4	0	Prescribed nitr...	{0, No}...	None	7	Right
13	anticlot	Numeric	4	0	Taking anti-clot...	{0, None}...	None	10	Right
14	site	Numeric	4	0	Hospital ID	{1, 0001}...	None	6	Right
15	attphys	Numeric	4	0	Attending physi...	None	-1	9	Right

# Problem 1

## **Which is more (male or Female), depending on statistics metrics? how?**

- To find the most mentioned one (male or female), we can use MODE to decide which is more.
- As we know normally “gender” variable is “string”, but it is not possible to find mode for string.
- We can define gender as numeric, but coded numeric by using “values parameter” (0 for male and 1 for female as example).
- From analyze → descriptive statistics → frequency we can find mode by checking mode from statistics

# Problem 1

**Which is more (Male or Female), depending on statistics metrics? how?**

## Statistics

Gender

N	Valid	10000
	Missing	0

## Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	female	4971	49.7	49.7	49.7
	male	5029	50.3	50.3	100.0
	Total	10000	100.0	100.0	

Results for “string” gender  
Note that there is no Mode.

## Statistics

Gender

N	Valid	10000
	Missing	0
Mode		0

## Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	5029	50.3	50.3	50.3
	Female	4971	49.7	49.7	100.0
	Total	10000	100.0	100.0	

Results for “coded numeric” gender  
Note that there is Mode. With same results

## Problem 2

**For (AGEcat) variable which category is more frequent, in other word which age stage has more cases?**

- AGEcat is already numeric variable, so we can get Mode.
- AGEcat is modified to be “ordinal” numeric (not scale), what does that mean?
- From analyze → descriptive statistics → frequency  
we can find mode by checking mode from statistics

# Problem 2

- From first table we note that is most of age category mentioned in our study is which (coded with number “2”).
- If we look back to the “values” table 2= 55-64

## → Frequencies

### Statistics

Age category

N	Valid	10000
	Missing	0
Mode		2

### Age category

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	45-54	2195	22.0	22.0	22.0
	55-64	3878	38.8	38.8	60.7
	65-74	2861	28.6	28.6	89.3
	75+	1066	10.7	10.7	100.0
Total		10000	100.0	100.0	



# Anesthesia dept

## SPSS-TH-8

Compare mean

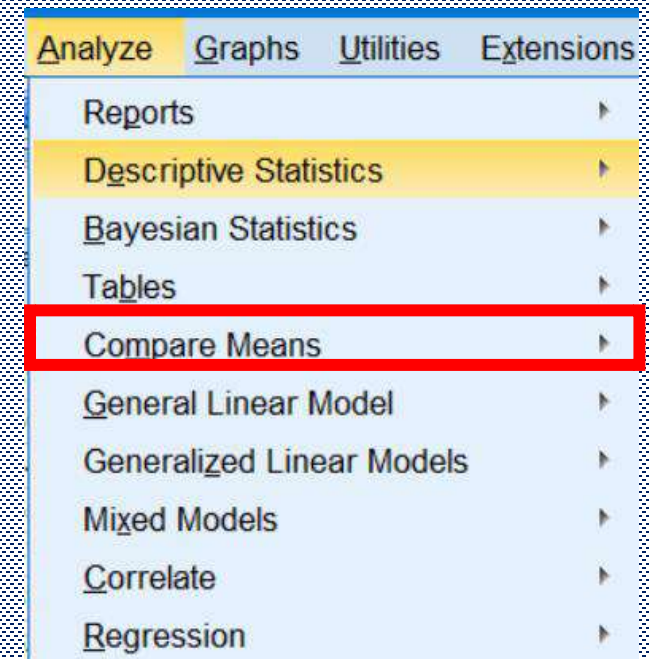
Abdulmajeed ahmed

2021-2022



# Compare means

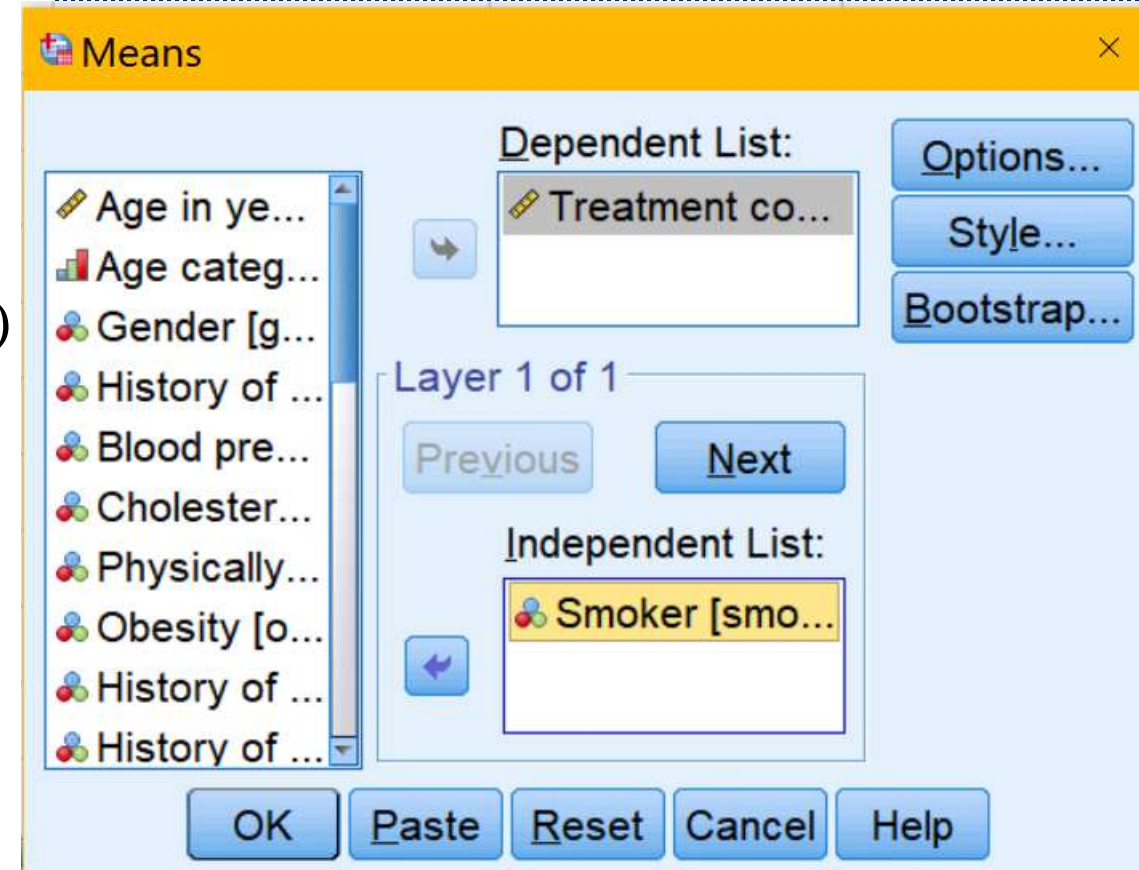
- It also uses to compare between differences in descriptive statistics across one or more factors, or categorical variables.
- To open the Compare Means procedure, click **Analyze > Compare Means > Means.**
- it have multiple layer property, that is used to compare more than one variable at the same time.





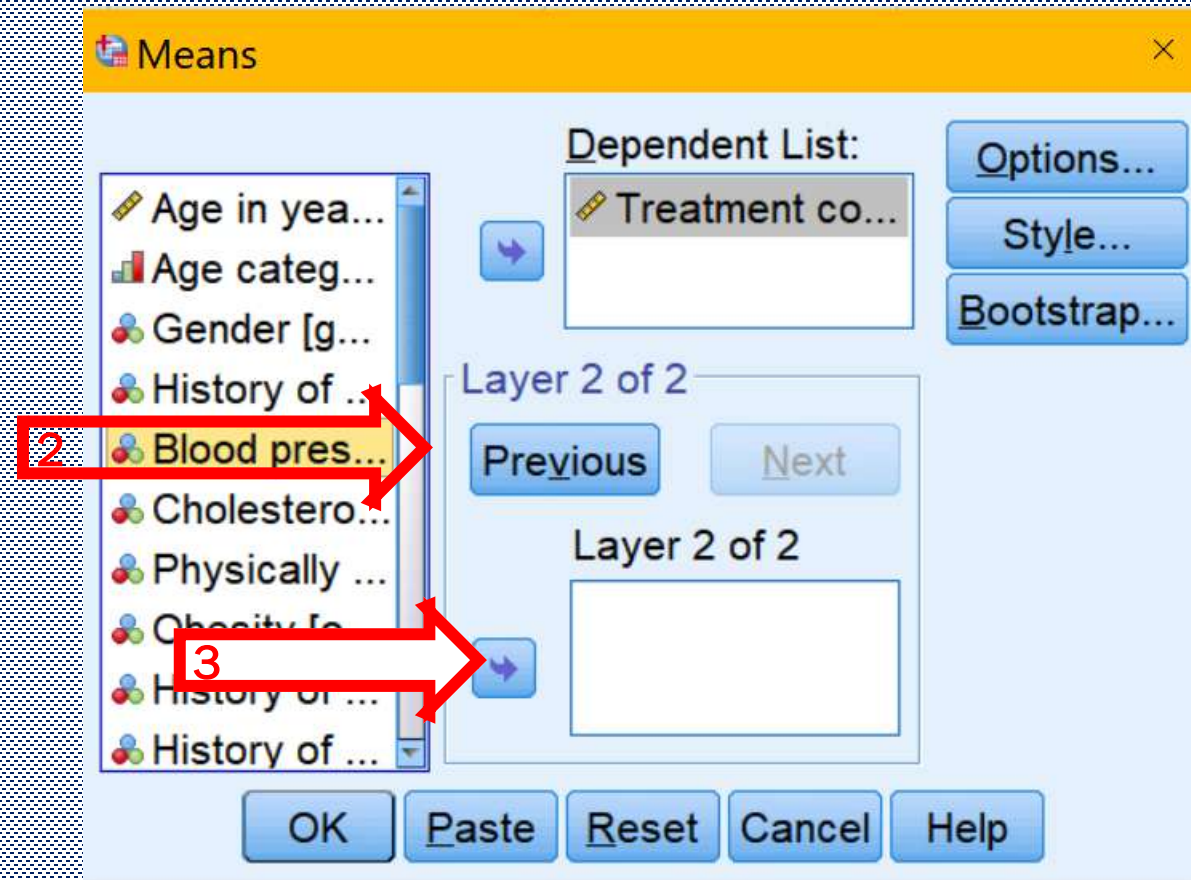
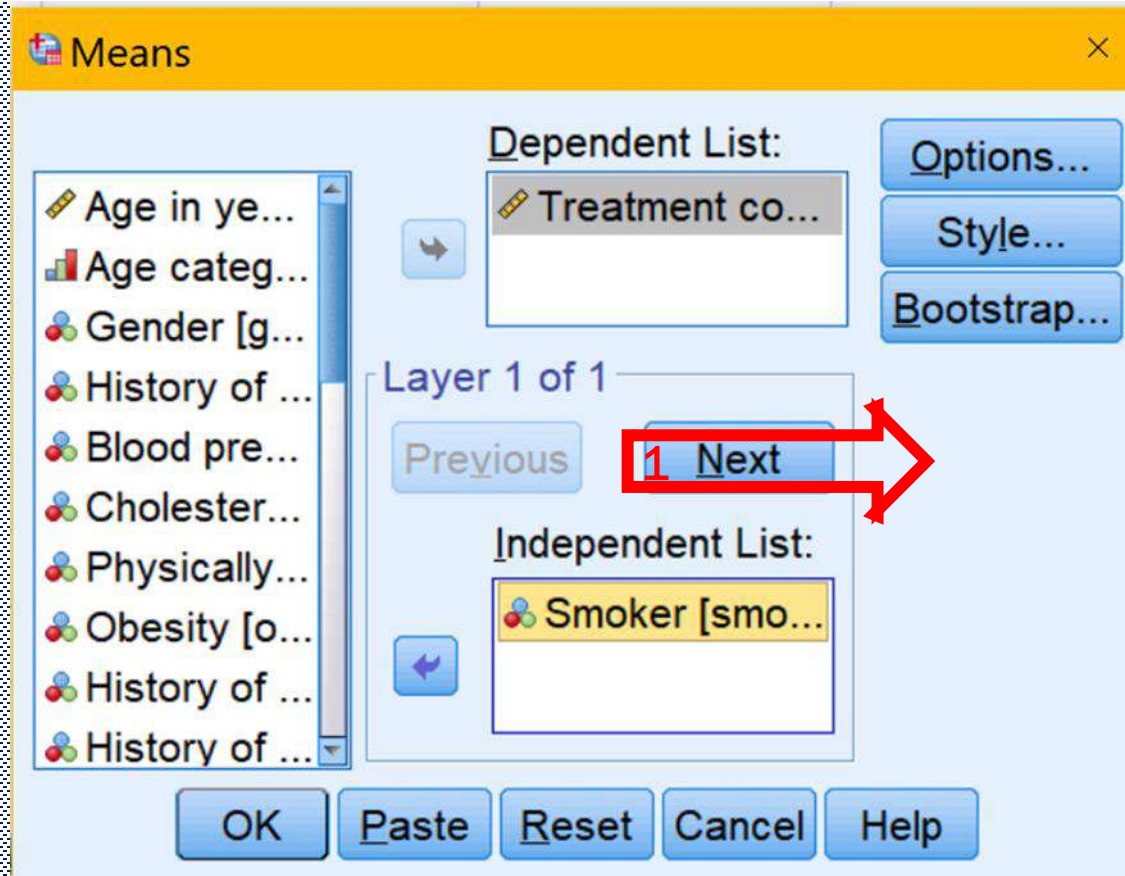
# Compare means

- *Dependent List:* The continuous **numeric** variables to be analyzed. It is **necessary** to fill this list.
- *Independent List:* The categorical variable(s) that will be **used to subset the dependent variables**. Specifying multiple values in the "Layer 1 of 1" box will produce several tables, each with one layer variable.
  - this will produce a table that looks like a **hybrid of a crosstab and the Descriptive procedure**.



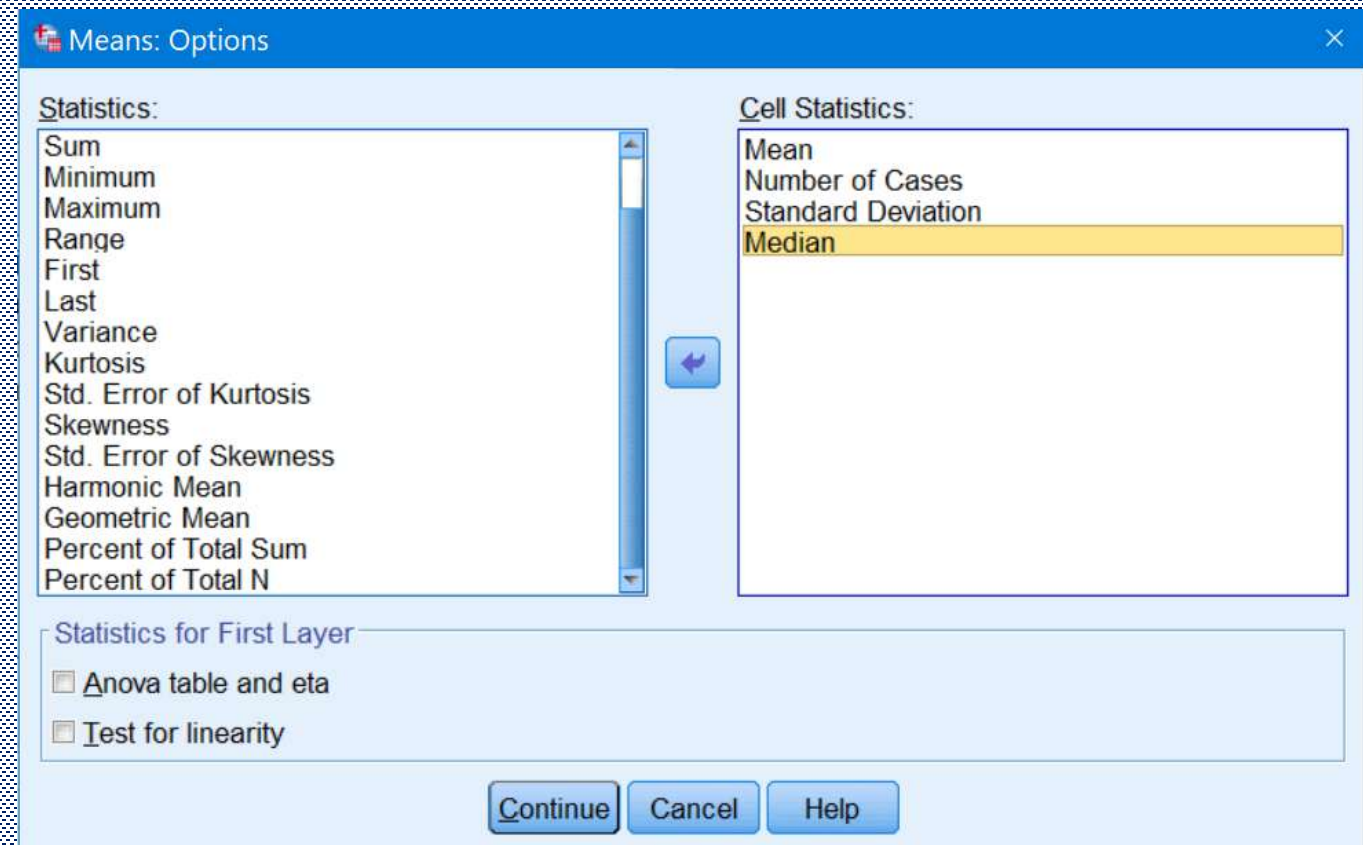
# Compare means

- If we need to add another variable “layer”, in layer 1 of 1 click “next”:



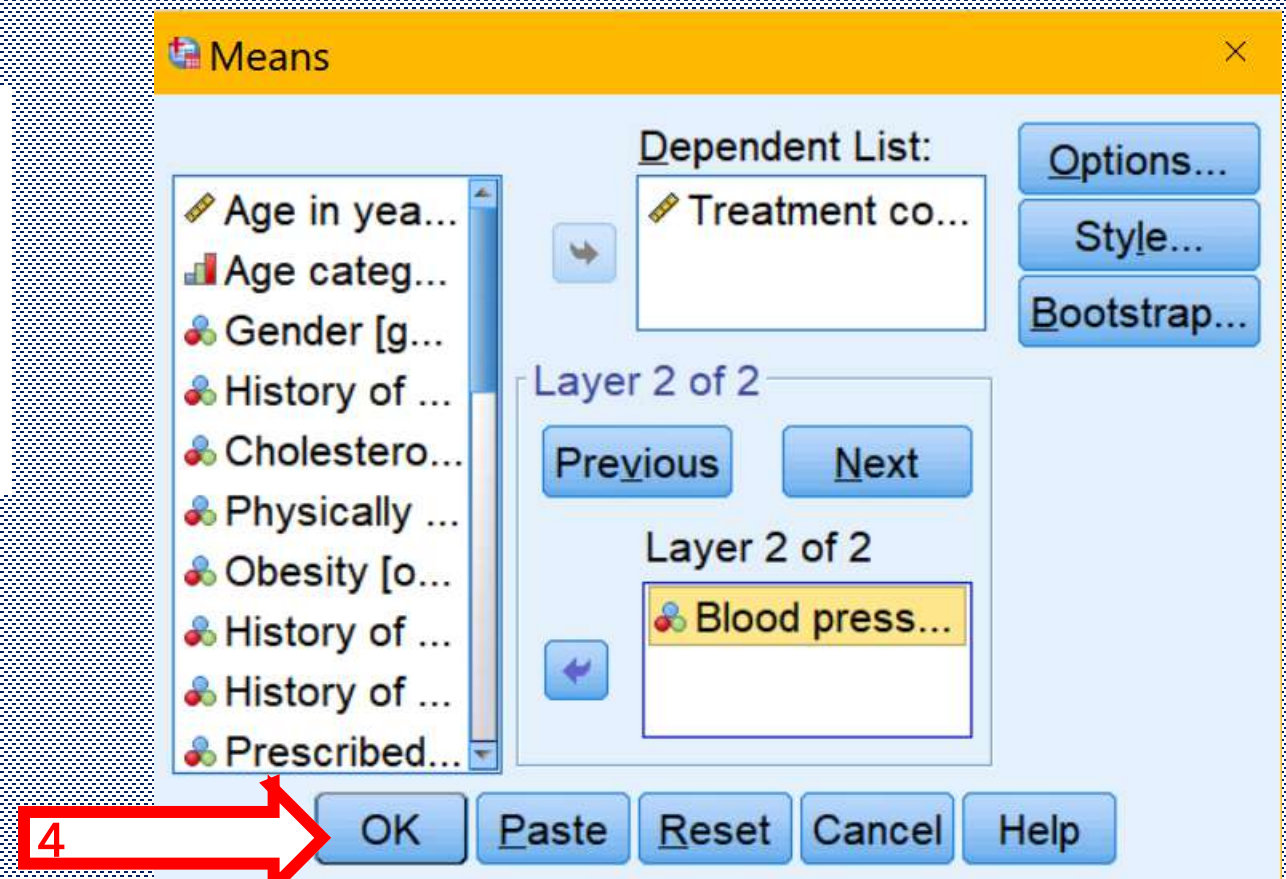
# Compare means options

- To modify descriptive that we need to compare data we have.
- The most common descriptive to be compared is “mean, number of cases, std- deviation, median”.



# Compare means options

- By clicking OK we will get a detailed table (**report**) that describes “treatment costs” for different cases (6 cases).



# Compare means

- As we can see (in general ) the smokers with hypertension had higher treatment costs, but the non-smoker cases with normal blood pressure had paid less (in general too :).
- By doing these steps we can get good reports for our databases.

Report				
Treatment costs				
Blood pressure	Smoker	Mean	N	Std. Deviation
Hypotension	No	16.6002	833	16.58711
	Yes	26.7780	217	16.38853
	Total	18.7036	1050	17.04469
Normal	No	15.3652	4212	16.30801
	Yes	23.6839	1139	16.84875
	Total	17.1359	5351	16.77236
Hypertension	No	22.5950	1730	17.05704
	Yes	27.2646	498	15.29837
	Total	23.6388	2228	16.78985
Total	No	17.3632	6775	16.82033
	Yes	25.0079	1854	16.46925
	Total	19.0057	8629	17.03644



# Anesthesia dept

## SPSS-TH-15

Chi-square

Abdulmajeed ahmed





2021-2022





# What is chi-square

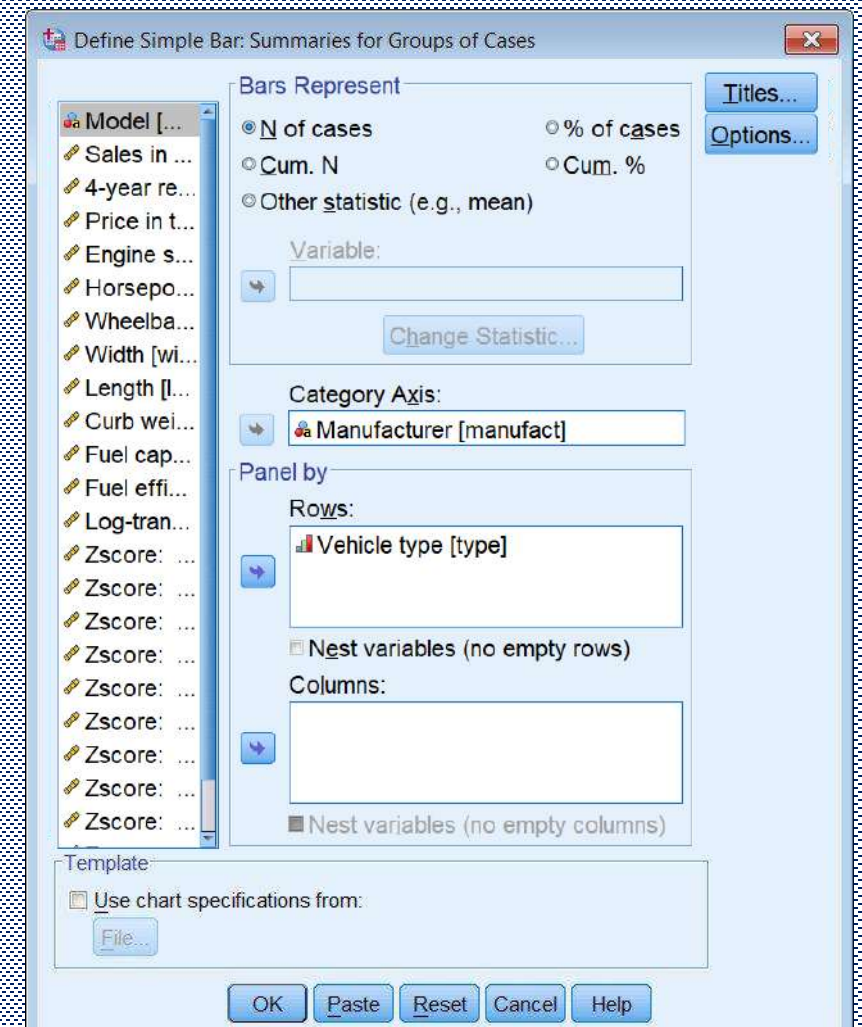
1. Test of Independence determines whether there is an association between categorical variables.
2. It is non-parametric\one sample test.
3. Comparisons must be between two categoric variables “only”.
4. Uses contingency table (also known as a cross-tabulation, crosstab, or two-way table) is an arrangement in which data is classified according to two categorical variables.
5. The categories for one variable appear in the rows, and the categories for the other variable appear in columns

	Variable A		Total
Variable B	 Latte	 Vanilla	Afernoorn Tea
	 Cinnamon	 De-caf	Bedtime
Total	Special Treat	Tea Break	A good Cuppa!

The Chai-Squared Test

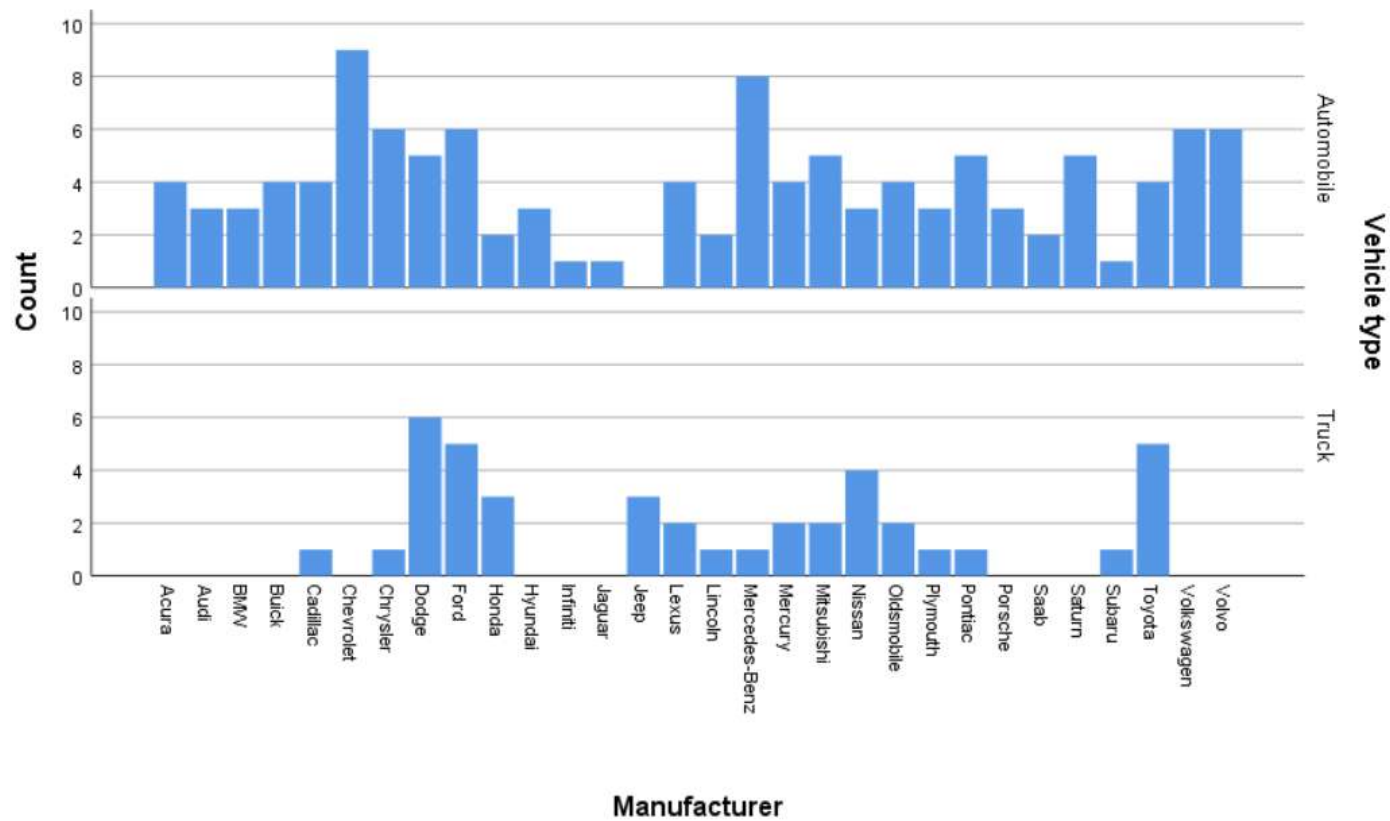
# Before chi-square

1. We are using “car\_sales.sav” for our implementations.
2. Before having chi-test it is better to understand how two variables is associated to each other
3. To do that :  
graphs → legacy dialogs → bars → simple and click define



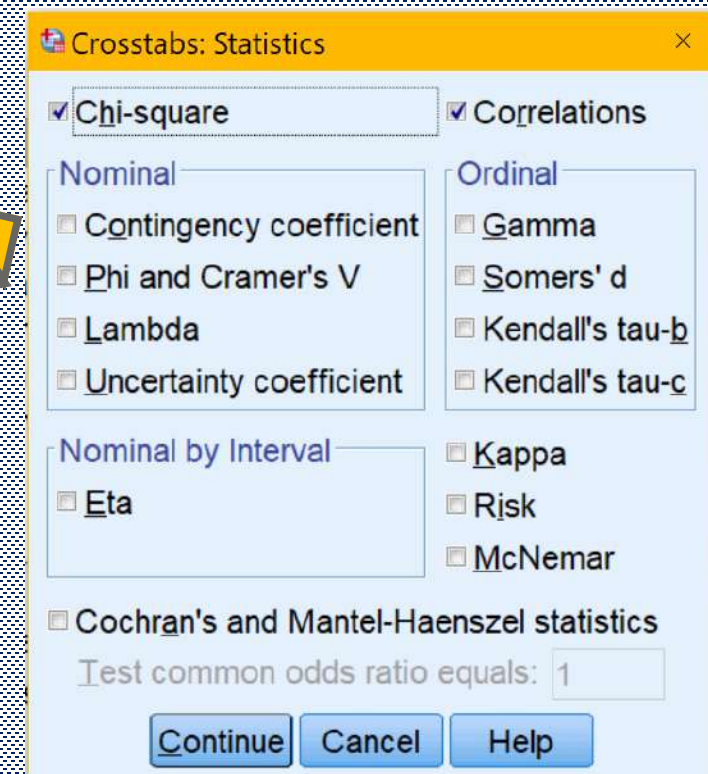
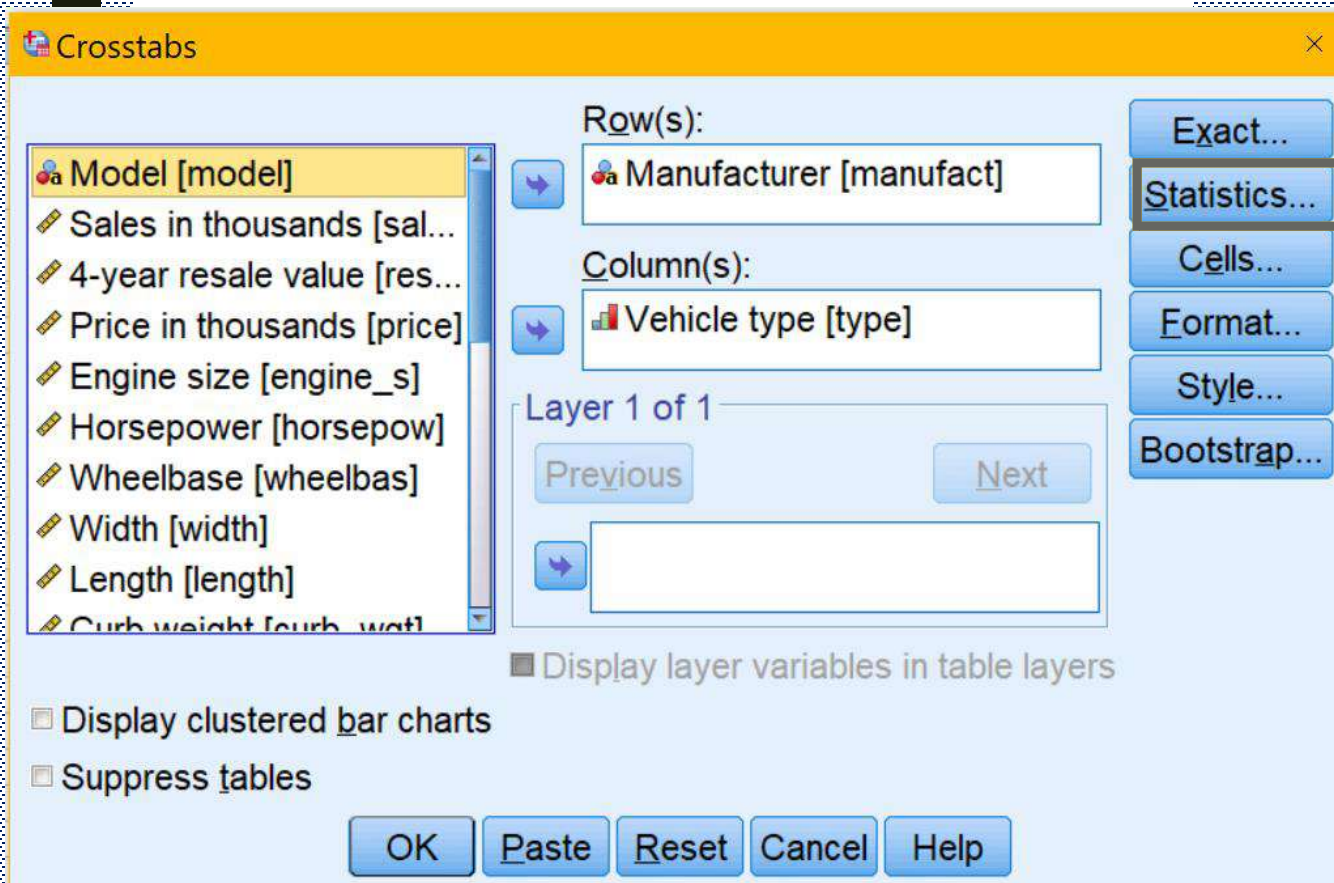
# Before chi-square

By this way we can get a visualized Idea about which companies produce trucks beside automobile.



# chi-square test

To get our numerical result for our study :  
Analyze → descriptive statistics → cross tab  
That means that we will have cross table



We must check “chi-square” box to have our result.  
Our results will be three tables as follows:

# chi-square test

The first table shows us :

- Number of samples
- Number of valid data
- Number of missing data

Case Processing Summary						
	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
Manufacturer * Vehicle type	157	100.0%	0	0.0%	157	100.0%

# chi-square test

The second table shows us :  
Each company production in details with numbers,

Instead of converting the data type from string to numeric data and give numeric symbol for each category, we can get all results we need from “chi-square test”.

Manufacturer * Vehicle type Crosstabulation				
Count	Manufacturer	Vehicle type		Total
		Automobile	Truck	
	Acura	4	0	4
	Audi	3	0	3
	BMW	3	0	3
	Buick	4	0	4
	Cadillac	4	1	5
	Chevrolet	9	0	9
	Chrysler	6	1	7
	Dodge	5	6	11
	Ford	6	5	11
	Honda	2	3	5
	Hyundai	3	0	3
	Infiniti	1	0	1
	Jaguar	1	0	1
	Jeep	0	3	3
	Lexus	4	2	6
	Lincoln	2	1	3
	Mercedes-Benz	8	1	9
	Mercury	4	2	6
	Mitsubishi	5	2	7
	Nissan	3	4	7
	Oldsmobile	4	2	6
	Plymouth	3	1	4
	Pontiac	5	1	6
	Porsche	3	0	3
	Saab	2	0	2
	Saturn	5	0	5
	Subaru	1	1	2
	Toyota	4	5	9
	Volkswagen	6	0	6
	Volvo	6	0	6
	Total	116	41	157



# chi-square test

The third table shows us :

The P value (probability) which must be smaller than 0.05 to reject null hypothesis.

In our case “Asymptotic Significance (2-sided)”, which represents P is equal to 0.021

$$0.021 < 0.05$$

Where 0.05 is our  $\alpha$

So we reject null hypothesis, in other words there is NO association between two variables.

Chi-Square Tests						
	Value	df	Asymptotic Significance (2-sided)	Significance	Monte Carlo Sig. (2-sided) 99% Confidence Interval	
					Lower Bound	Upper Bound
Pearson Chi-Square	46.527 <sup>a</sup>	29	.021	.013 <sup>b</sup>	.010	.015
Likelihood Ratio	56.531	29	.002	.007 <sup>b</sup>	.005	.009
Fisher's Exact Test	39.881			.017 <sup>b</sup>	.013	.020
N of Valid Cases	157					

a. 52 cells (86.7%) have expected count less than 5. The minimum expected count is .26.

b. Based on 10000 sampled tables with starting seed 112562564.

# Anesthesia dept

## SPSS-TH-7

Explore command

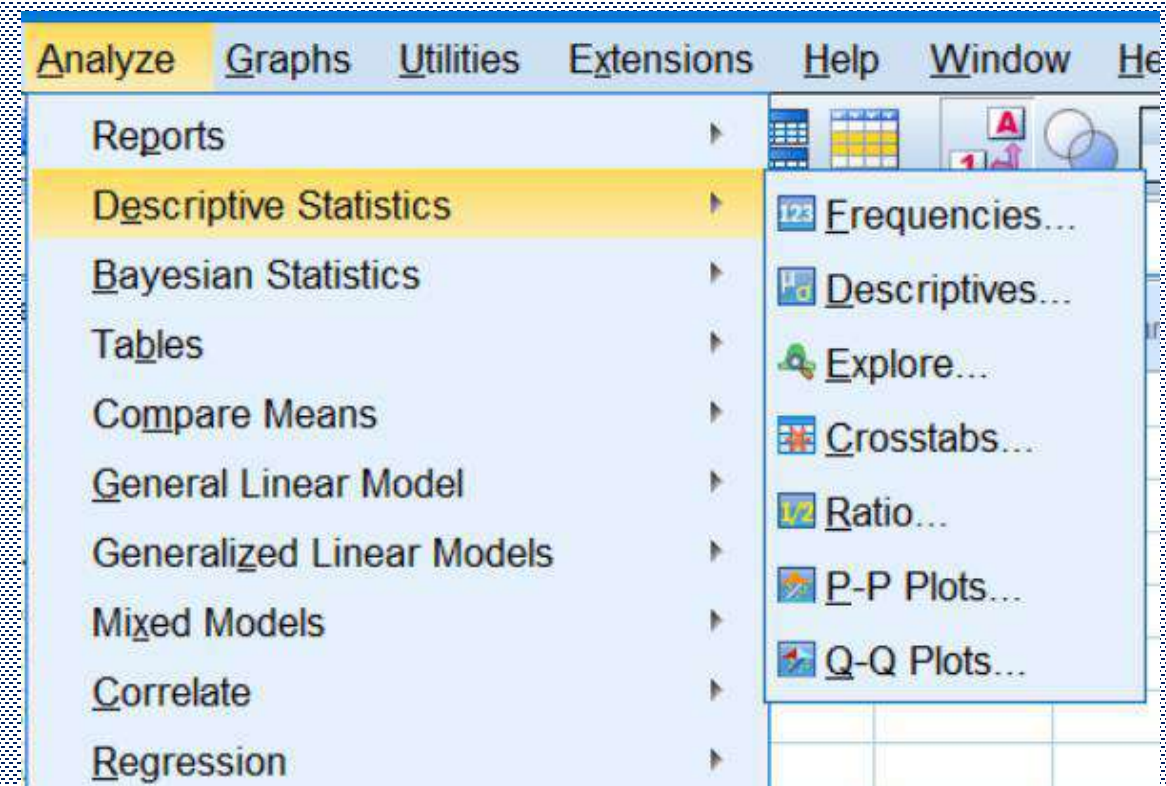
Abdulmajeed ahmed

2021-2022



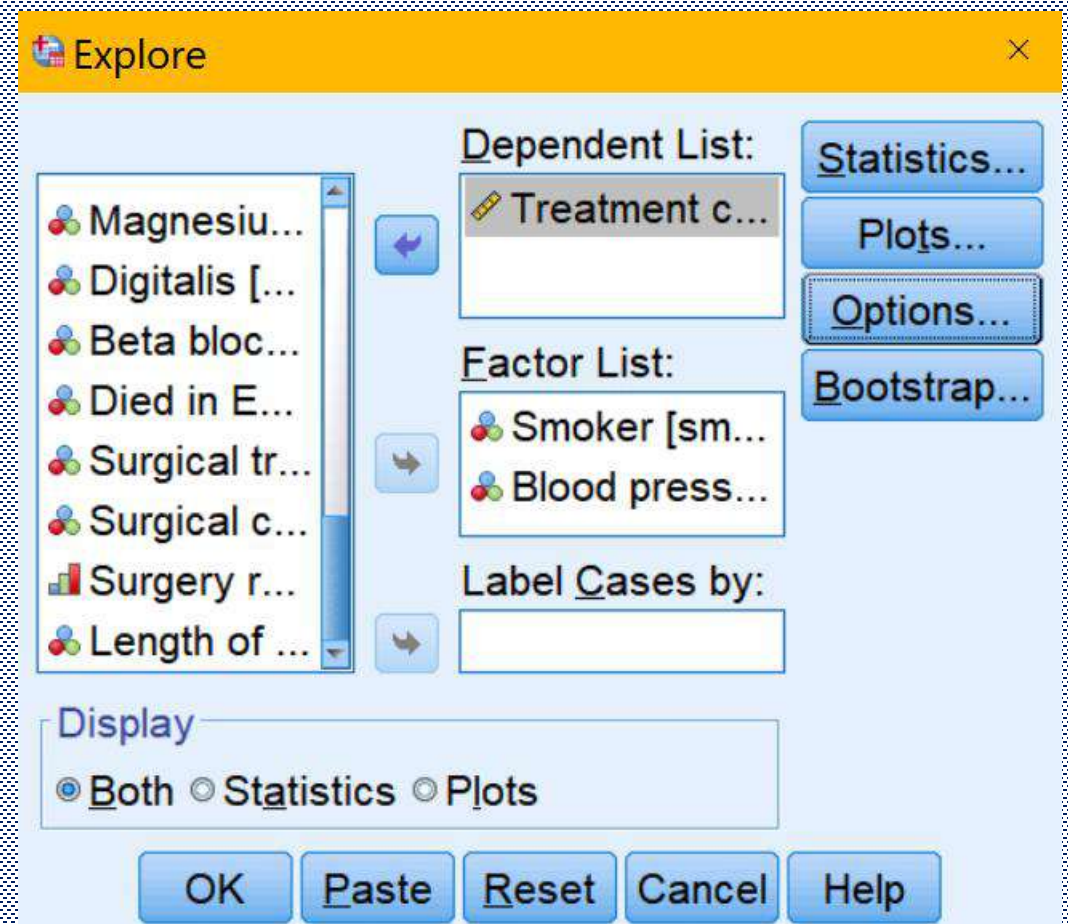
# Compare means

- allows the users to **generate an array of descriptive statistics** for a particular.
- The users can **compare two or more than two groups of variables in different descriptive statistics.**
- The explore command has the ability to **generate a summary** of the statistics for all the valid cases in each comparison group.
- It can also give separated study for **different cases.**



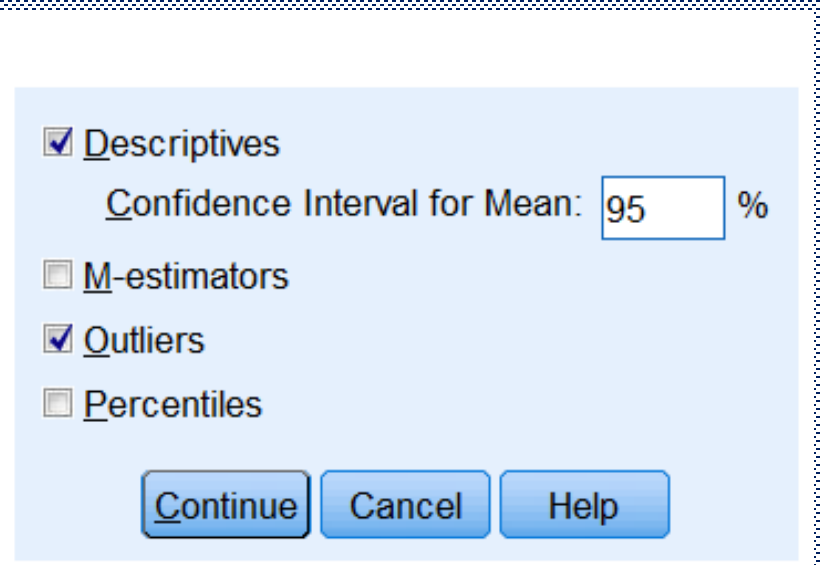
# Compare means

- In this lect. we will also use “patient\_los.spv” in (explore).
  - ❖ In “dependent list”, the variable that we want to analyze must be better **“numeric”**, note that it is **necessary** to select variable in this field.
  - ❖ In “factor list”, variable here is used to **divide the “analyzed” data**, so the **comparison** would be clear, “smoker & none smoker”, or “blood pressure” or both , in our example, this field is **optional**.
  - ❖ “label list” is also **optional** list that used to **label data by variable** “we can use cases name as example”.



# explore command

- By entering to statistics:
  - ❖ Set descriptive to 95%.
  - ❖ Check outliers to view “extreme values”, that shows cases with the highest & lowest records.
  - ❖ Click continue



The image shows the 'Statistics' dialog box for the SPSS Explore command. It has a light blue background. There are four checked options: 'Descriptives', 'Outliers', and 'Percentiles'. The 'Confidence Interval for Mean' is set to 95%. There are three buttons at the bottom: 'Continue', 'Cancel', and 'Help'.

☒ Descriptives  
Confidence Interval for Mean: 95 %

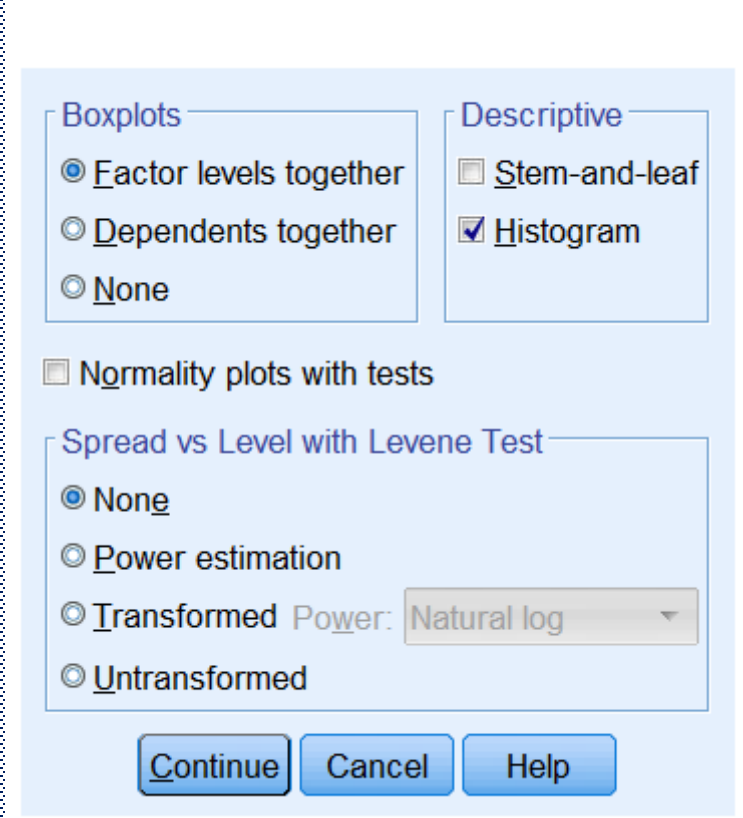
☐ M-estimators

☒ Outliers

☐ Percentiles

# explore command

- To edit “plots” that you can get from “explore”:
  - ❖ in “boxplot”, check (factors level together), so we can compare between the our factors.
  - ❖ In descriptive check “histogram.
  - ❖ And set (spread vs level with tense test) to none.
  - ❖ Press continue.

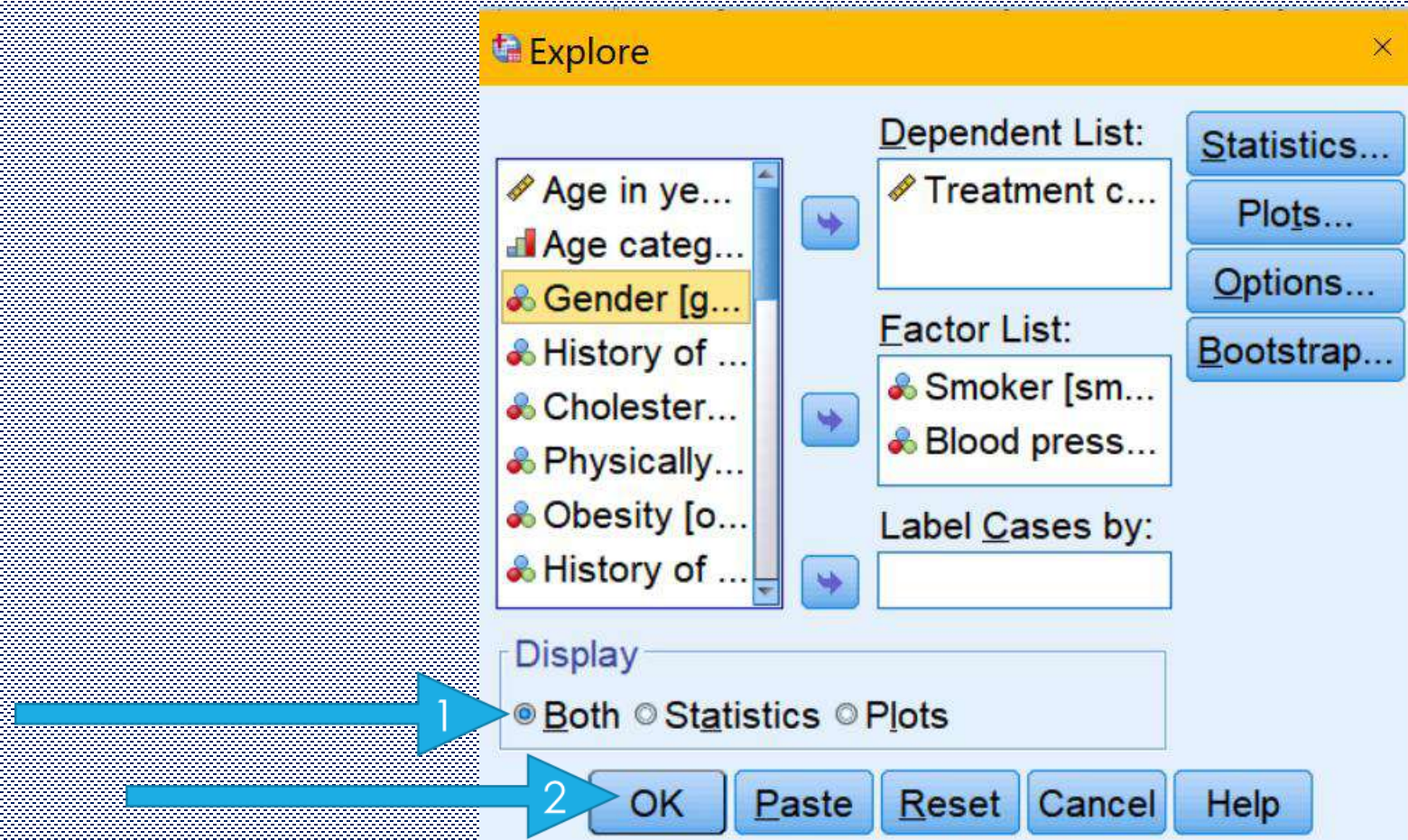


The image shows the 'Plots' sub-dialog box of the SPSS 'Explore' command. It is divided into two main sections: 'Boxplots' and 'Descriptive'. In the 'Boxplots' section, the 'Factor levels together' radio button is selected. In the 'Descriptive' section, the 'Histogram' checkbox is checked. Below these sections, the 'Normality plots with tests' checkbox is unchecked. The 'Spread vs Level with Levene Test' section has the 'None' radio button selected. At the bottom, there are three buttons: 'Continue', 'Cancel', and 'Help'.

Section	Option	Selected	
Boxplots	Factor levels together	Yes	
	Dependents together	No	
	None	No	
Descriptive	Stem-and-leaf	No	
	Histogram	Yes	
Normality plots with tests			No
Spread vs Level with Levene Test	None	Yes	
	Power estimation	No	
	Transformed	No	Natural log
	Untransformed	No	



# explore command



# Results in explore command

## ➤ Descriptive table:

- The first table we see is the “descriptive table”, unlike “summary and descriptive commands”, explore command provides statistics for “smoker” and “non-smoker” separately.
- From (mean, median, standard deviation and other statistics) we can say (smokers are spending more money ☹️💀).

Descriptives				
Treatment costs	Smoker		Statistic	Std. Error
	No	Mean	17.3632	.20435
		95% Confidence Interval for Mean	Lower Bound	16.9626
			Upper Bound	17.7638
		5% Trimmed Mean	16.3954	
		Median	6.5400	
		Variance	282.923	
		Std. Deviation	16.82033	
		Minimum	1.63	
		Maximum	80.30	
		Range	78.67	
		Interquartile Range	28.01	
		Skewness	.559	.030
		Kurtosis	-1.093	.059
	Yes	Mean	25.0079	.38249
		95% Confidence Interval for Mean	Lower Bound	24.2577
			Upper Bound	25.7580
		5% Trimmed Mean	24.6515	
		Median	28.0250	
		Variance	271.236	
		Std. Deviation	16.46925	
		Minimum	1.66	
		Maximum	75.09	
		Range	73.43	
		Interquartile Range	33.44	
		Skewness	-.076	.057
		Kurtosis	-1.043	.114

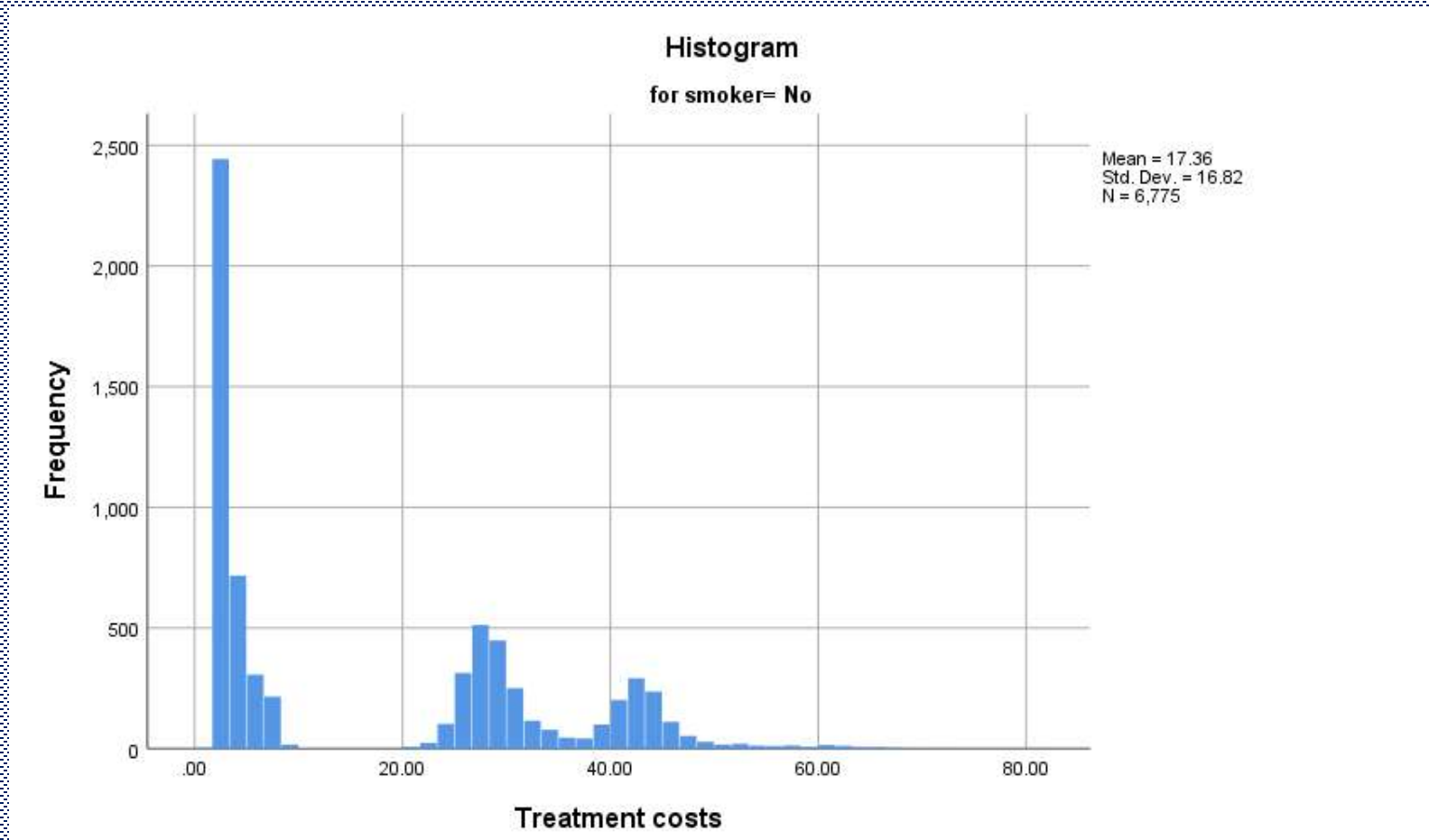
# Results in explore command

## ➤ Extreme values tables

This table shows biggest values (5 values), and smallest values(5 values) for both (smoker & non-smoker ) **separately**.

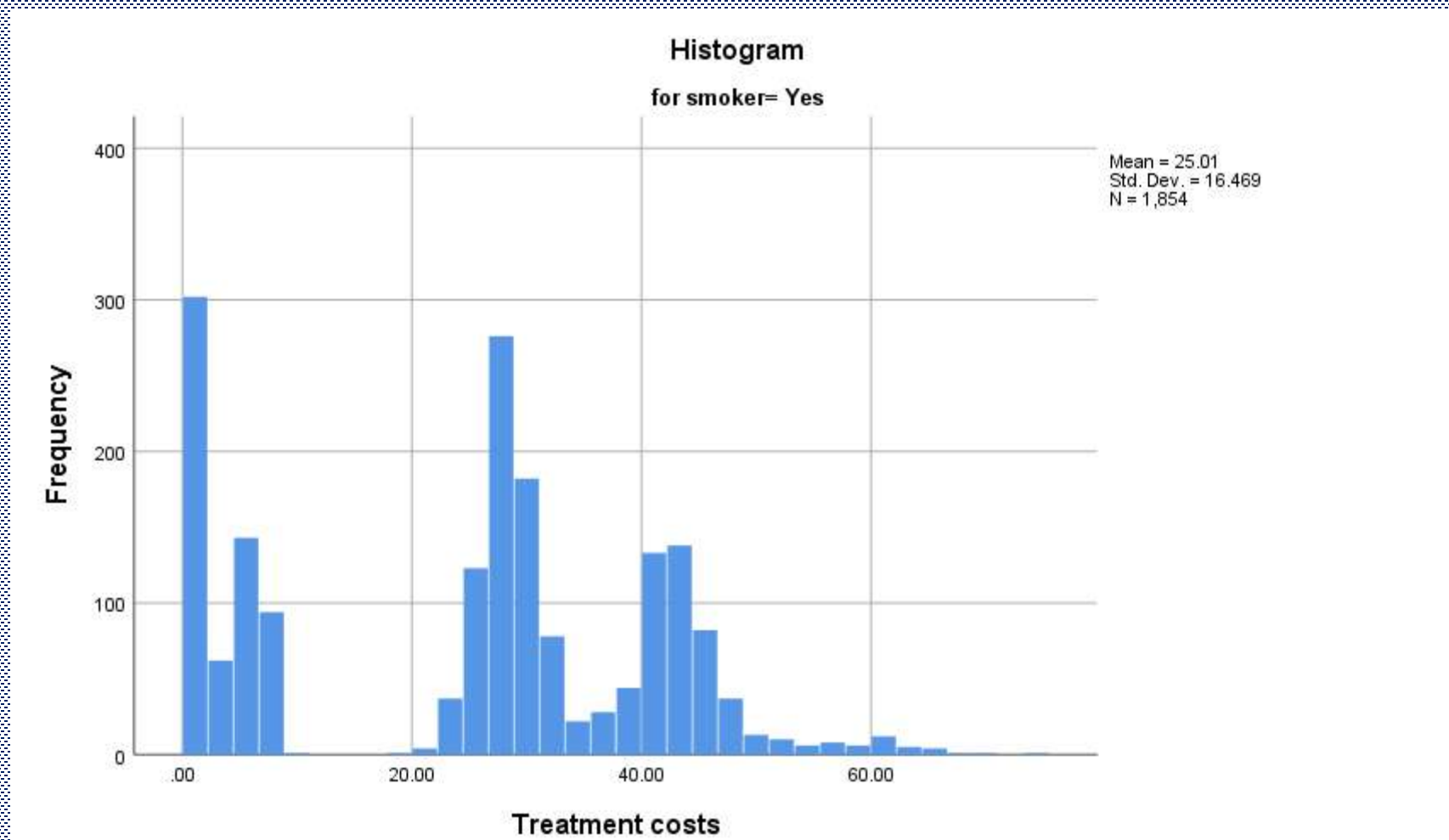
Extreme Values					
Treatment costs	Smoker			Case Number	Value
	No	Highest			
		Highest	1	5687	80.30
			2	521	71.90
			3	8870	67.05
			4	2322	66.82
			5	3545	66.78
		Lowest	1	3128	1.63
			2	2265	1.64
			3	757	1.65
			4	9840	1.67
			5	7402	1.67
	Yes	Highest	1	3302	75.09
			2	2122	70.31
			3	3001	67.67
			4	322	66.00
			5	5986	65.61
		Lowest	1	3399	1.66
			2	5931	1.68
			3	9535	1.71
			4	5377	1.71
			5	6529	1.74

# Results in explore command



Note that most of the non-smokers paid less than 20 dollar

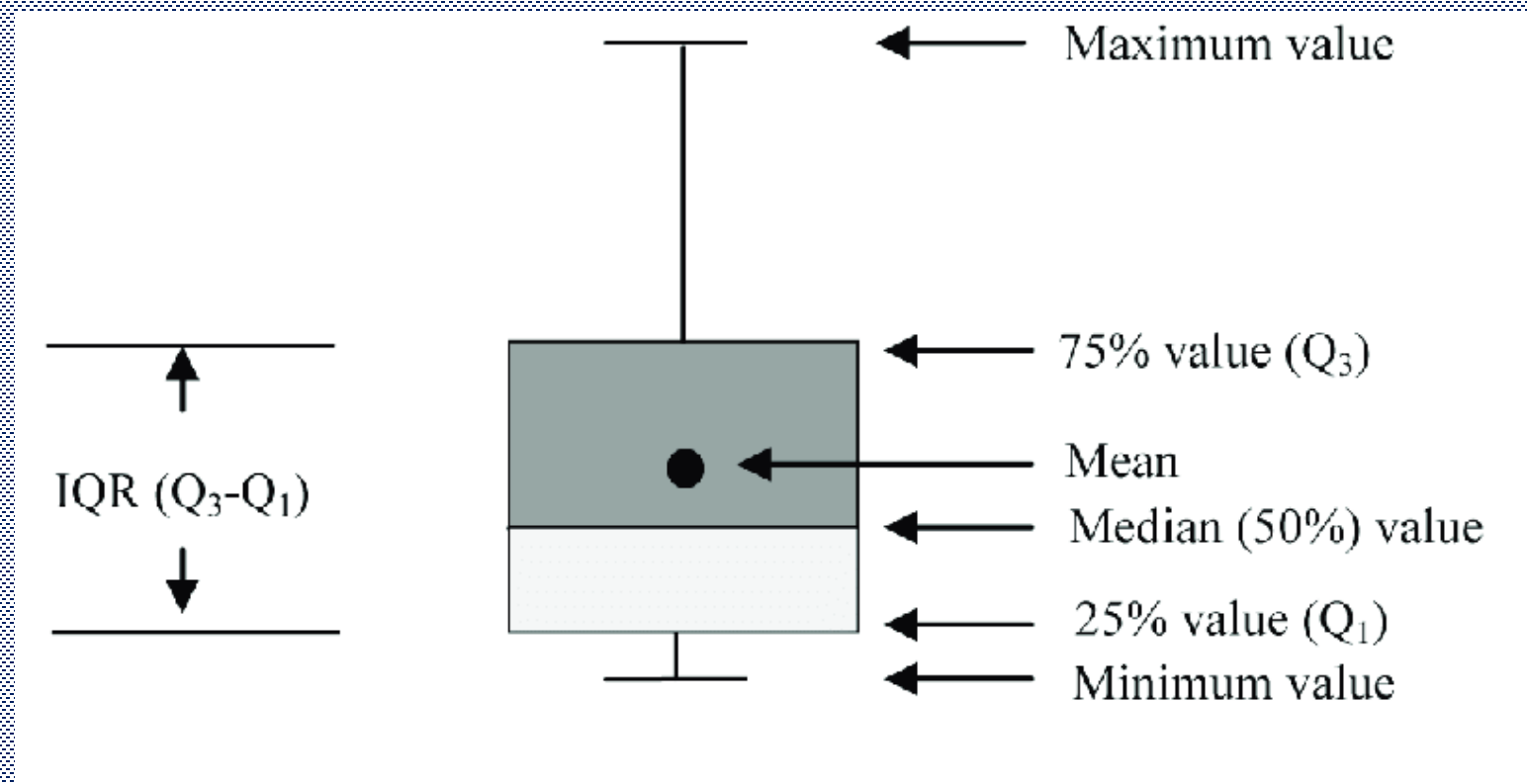
# Results in explore command



Note that most of the smokers paid more than 20 dollar

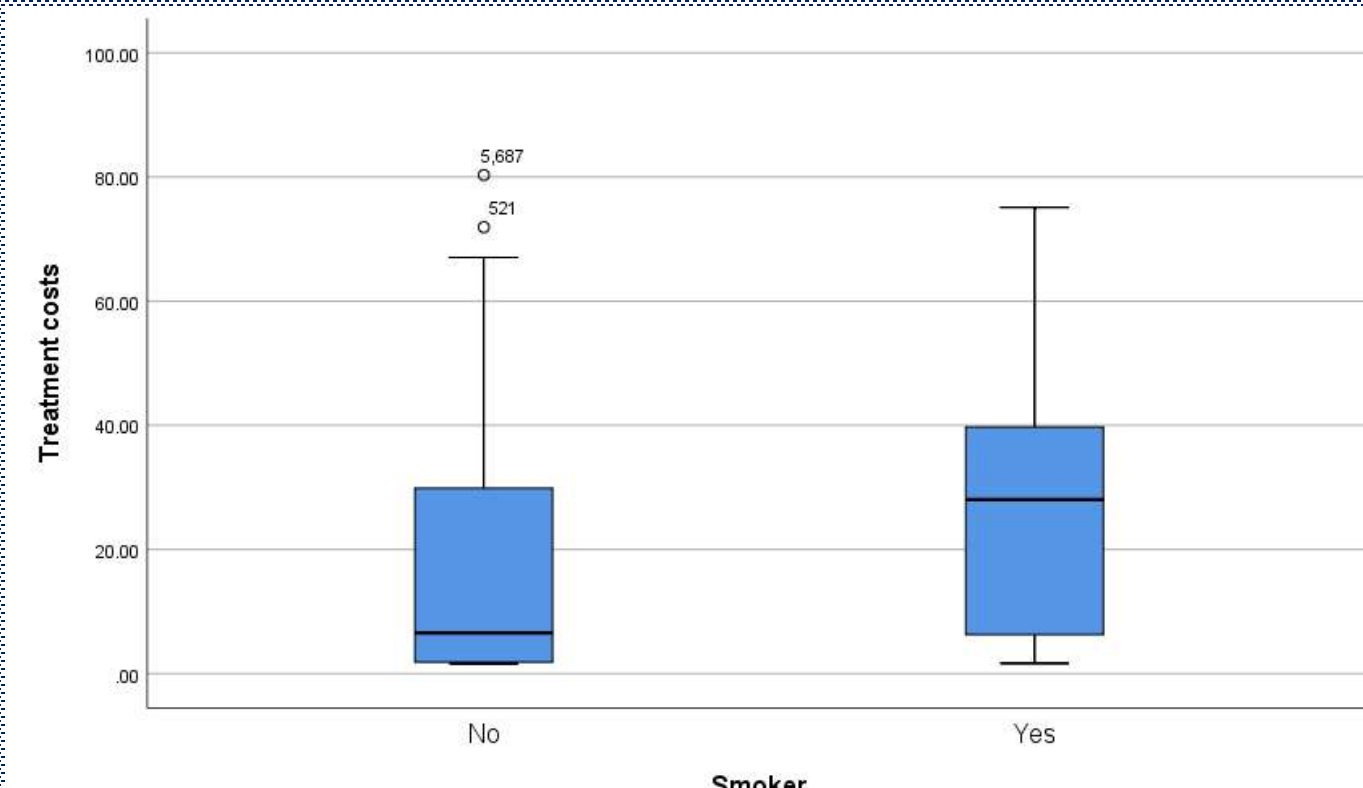
# Results in explore command

- Boxplots: can give summary of descriptive statistics in one graph.





# Results in explore command



Boxplot reading (approximations):

- Maximum(non-smokers=68 smokers=77).
- minimum(non-smokers=0 smokers=5).
- Mean(middle of box) (non-smokers=17, smokers=20).
- Median (line in the box) (non-smokers=5, smokers=28).