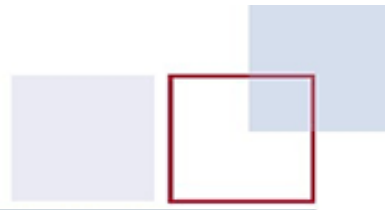


# Statistics



Roya Najafi Vosough  
Ph.D. in Biostatistics

# What is in this workshop

- Basics of Statistics
- How to use SPSS
- How to get descriptive statistics

# Statistics

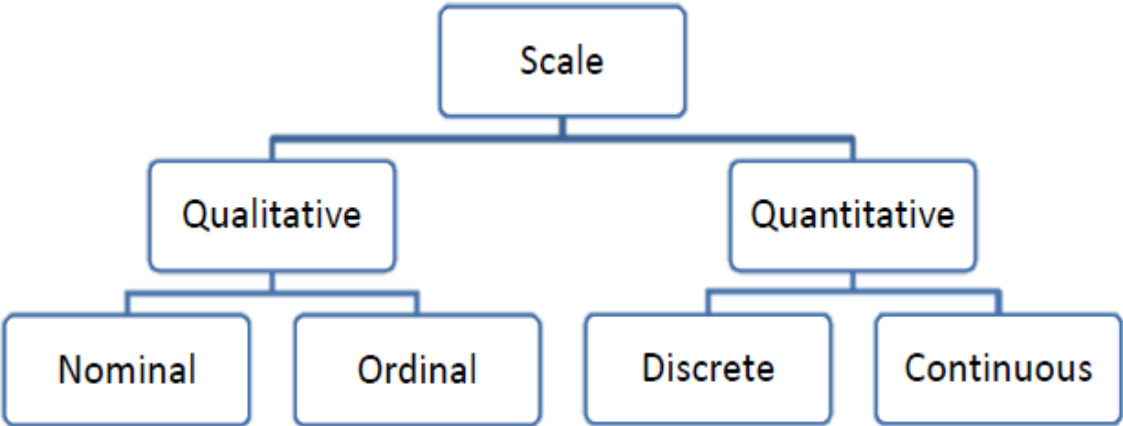
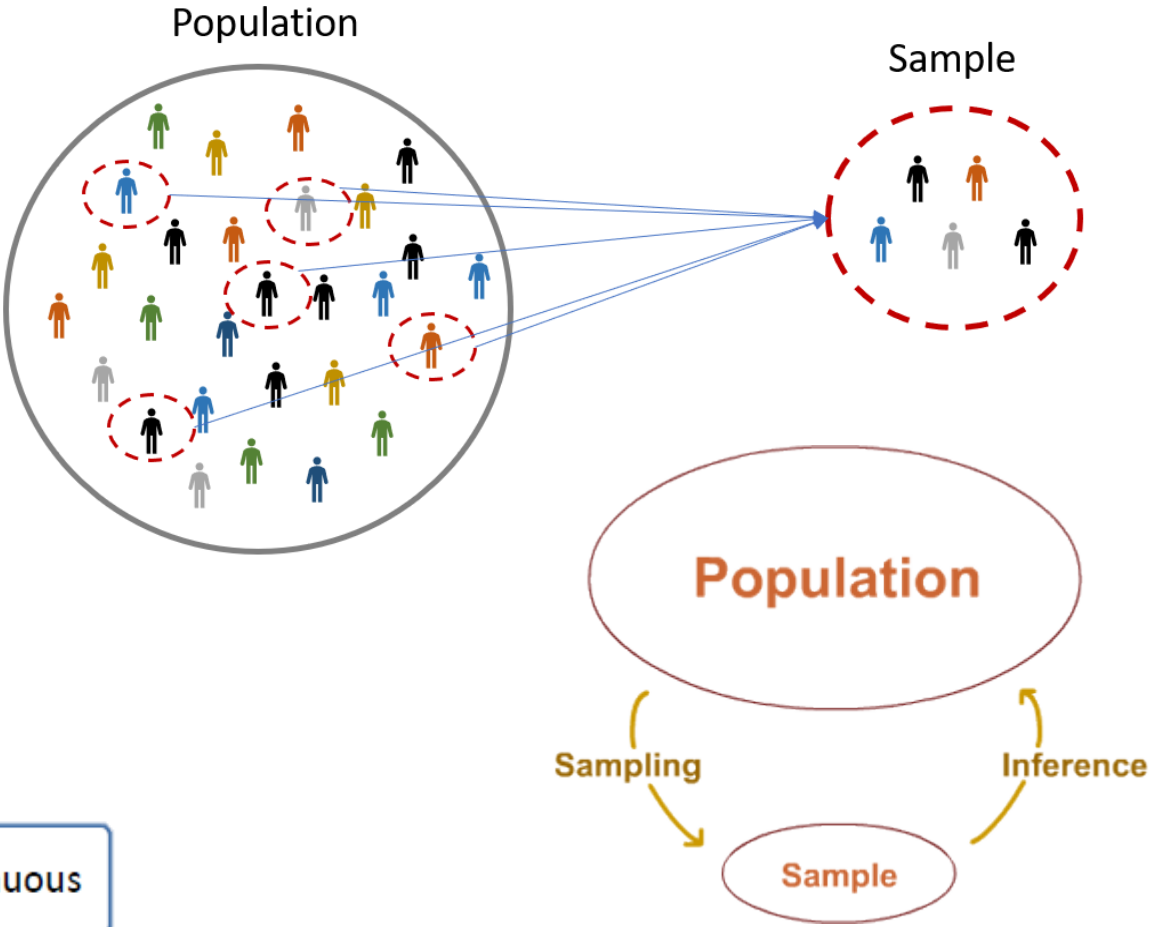
- The science of collection, organization, summarize, analysis, interpretation of data.

# Two areas of statistics:

- **Descriptive statistics**  
(which summarize some characteristic of a sample)
- **Inferential statistics**  
(which test for significant differences between groups and/or significant relationships among variables within the sample)

# Basic Terms

- Population
- Sample
- Variables

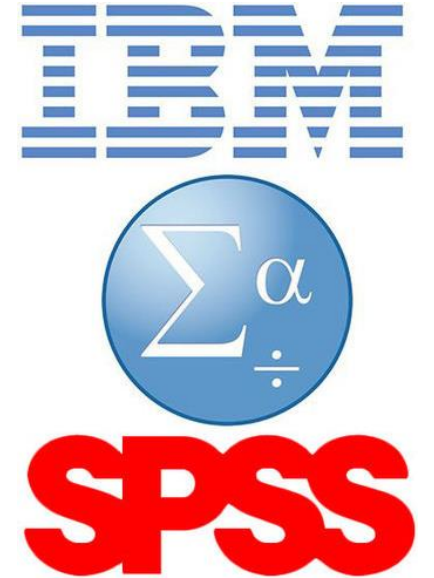


# Data Analytics Tools



# What is SPSS ?

- SPSS stands for Statistical Package for the Social Sciences
- SPSS was made to be easier to use than other statistical software like S-Plus, R, or SAS.
- The newest version of SPSS is SPSS 26.0.



# Uses for SPSS

- Data management
  - Defining variables
  - Coding values
  - Entering and editing data
  - Creating new variables
  - Recoding variables
  - Selecting cases
- Data analysis



# How to open SPSS

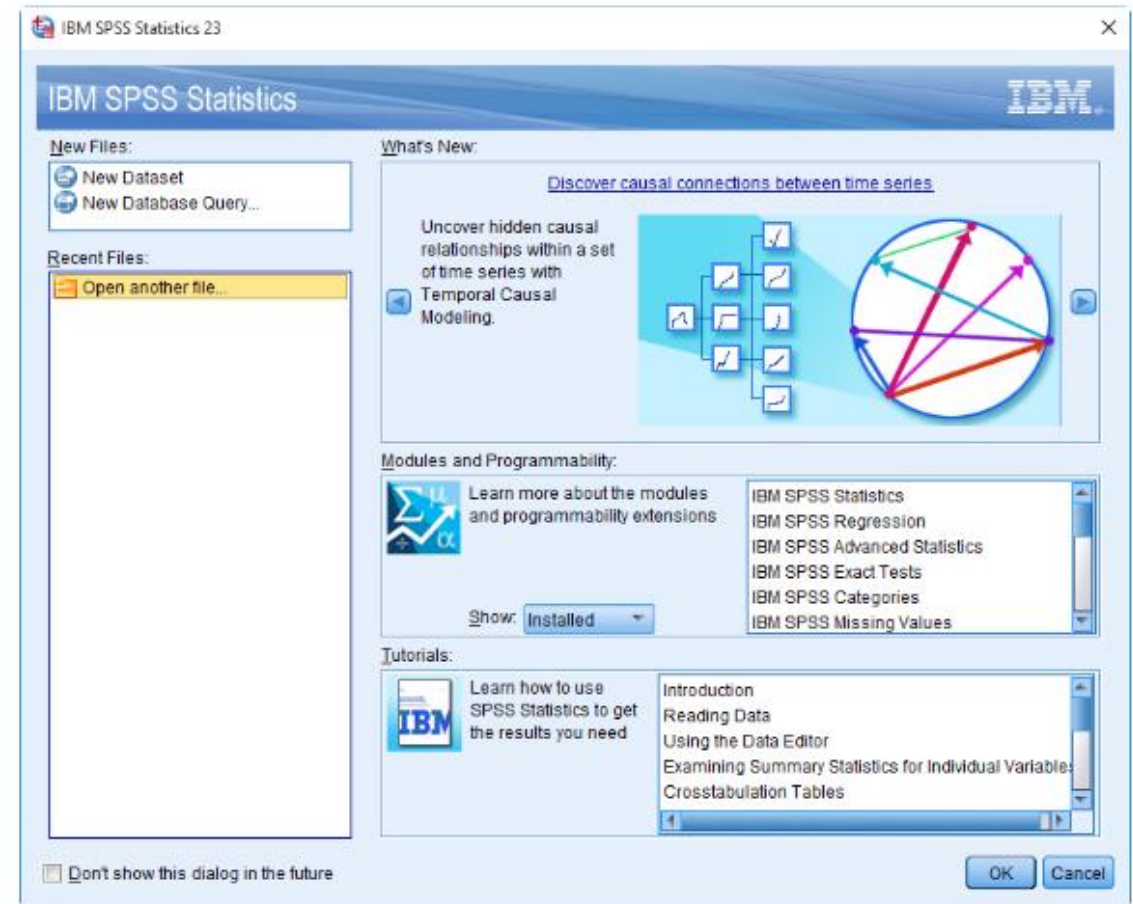
Double click the SPSS icon on the desktop

Start  
Menu

Programs

SPSS for  
windows

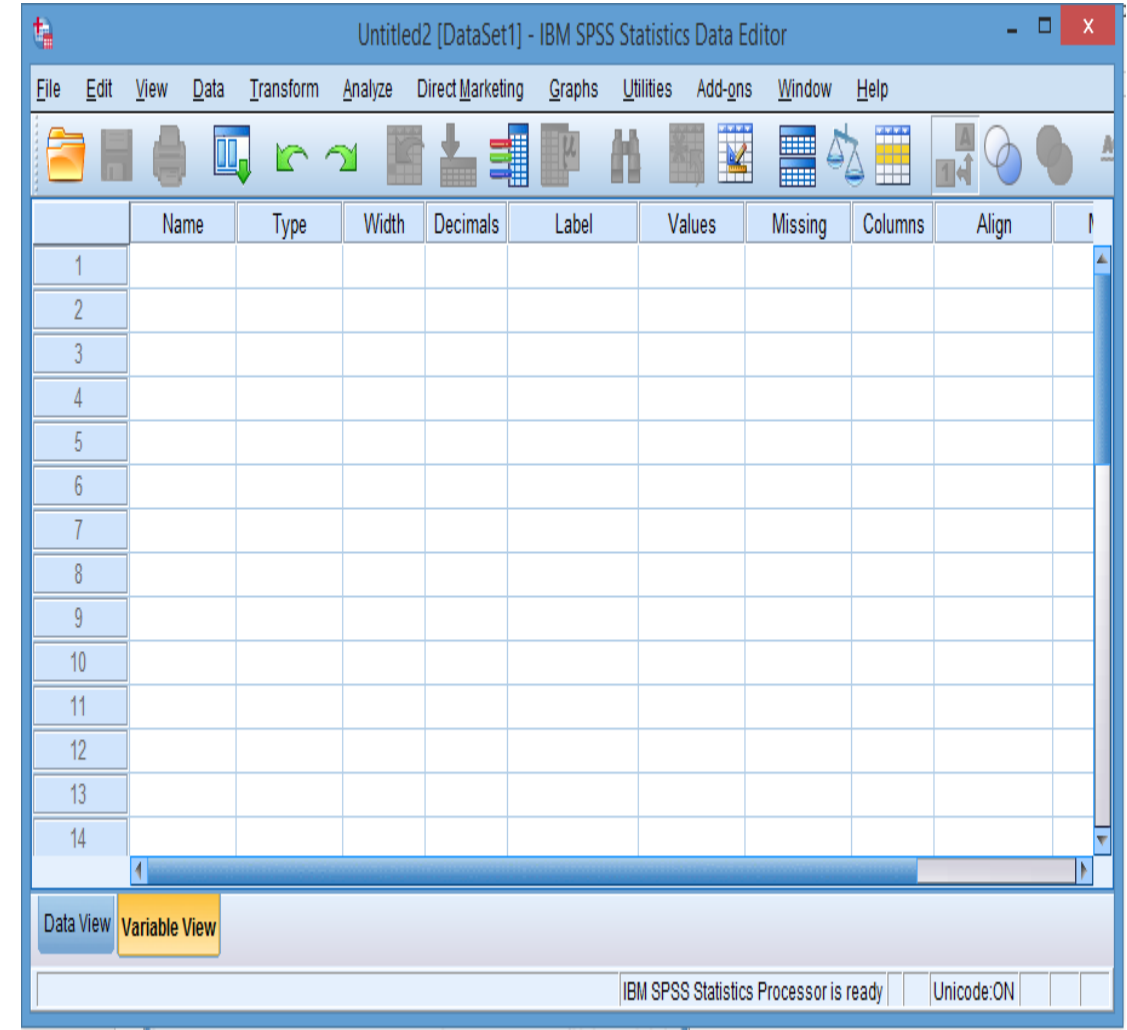
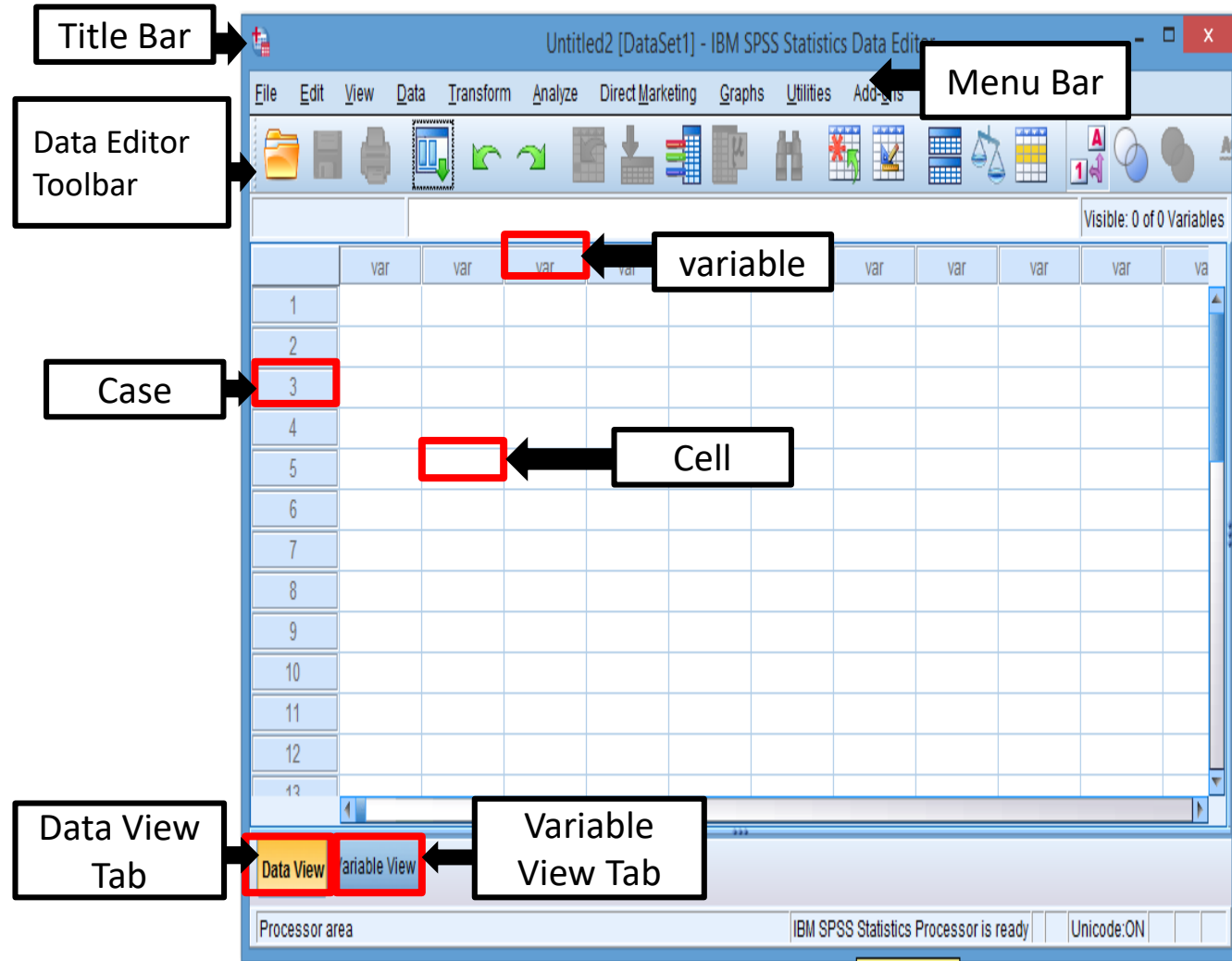
SPSS



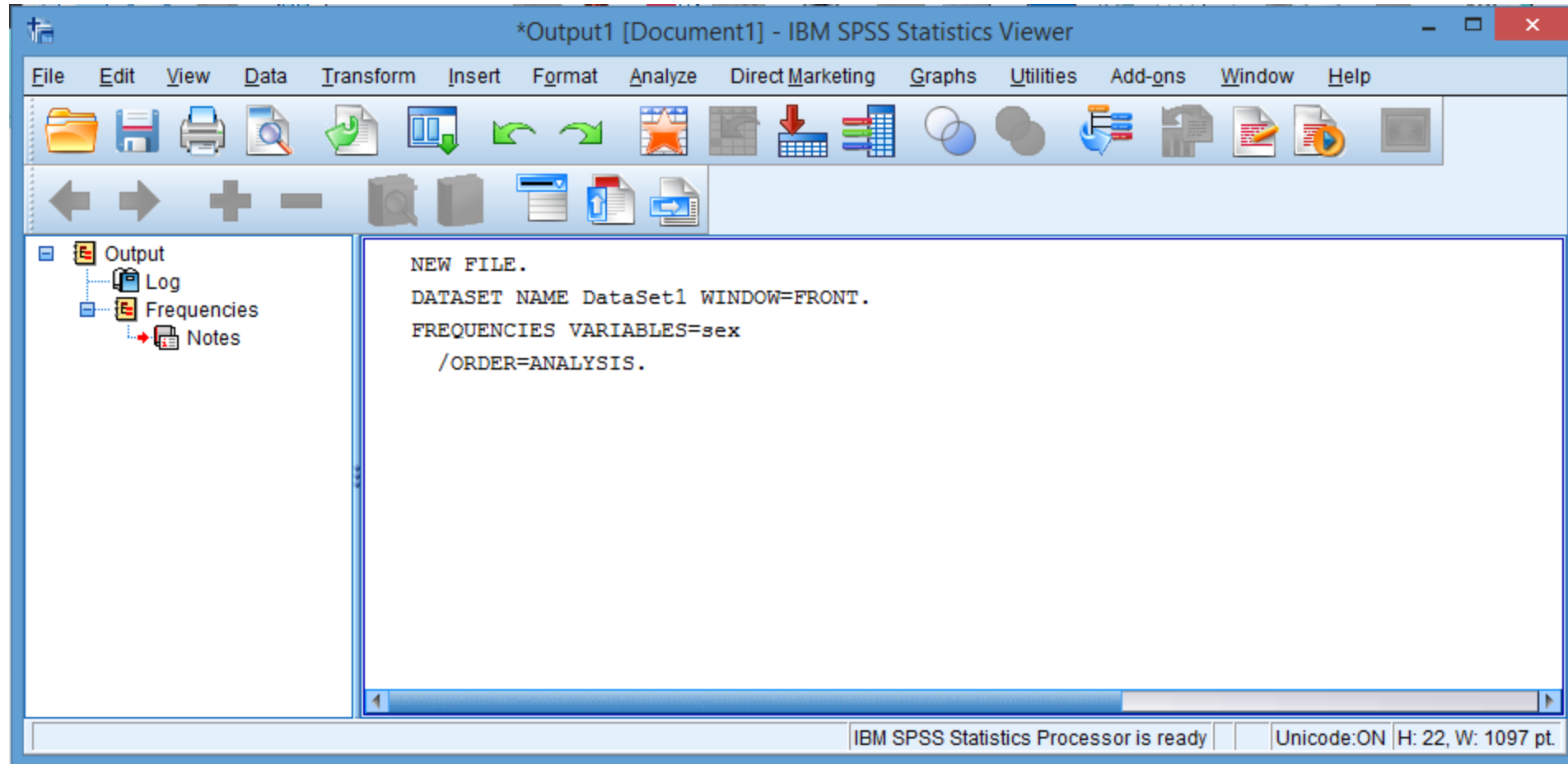
# The two main windows



# Data editor



# Output viewer



# Descriptive statistics

- ❑ Collection
- ❑ Organization
- ❑ Summarize data
  - Tables
  - Graphs
  - Measures of Central Tendency
  - Index of dispersion

# Example

## PATIENTS FORM

ID:

### DEMOGRAPHIC INFORMATION:

Sex:	1) Male <input type="checkbox"/>	2) Female <input type="checkbox"/>
Age (year):	<input type="text"/> <input type="text"/>	
Diabetes History	1) yes <input type="checkbox"/>	2) No <input type="checkbox"/>
Hypertension history	1) yes <input type="checkbox"/>	2) No <input type="checkbox"/>
Hyperlipidemia history	1) yes <input type="checkbox"/>	2) No <input type="checkbox"/>
Smoking history	1) yes <input type="checkbox"/>	2) No <input type="checkbox"/>

# Enter variables

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	sex	Numeric	8	2		None	None	8	Right	Scale
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										

1. Click Variable View
2. Type variable name under Name column (e.g. Sex).  
*NOTE: Variable name can be 64 bytes long, and the first character must be a letter or one of the characters @, #, or \$.*
3. Type: Numeric, string, etc.
4. Label: description of variables.
5. Measure: Nominal, Ordinal, Scale

# Enter variables

The screenshot shows the IBM SPSS Statistics Data Editor interface. The main window displays a variable definition table for a variable named 'sex'. The table has columns for Name, Type, Width, Decimals, Label, Values, Missing, Columns, and Align. The 'sex' variable is defined as Numeric, with a width of 8 and 2 decimal places. The 'Values' column shows 'None' with a dropdown arrow.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align
1	sex	Numeric	8	2		None	None	8	Right
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									

Two 'Value Labels' dialog boxes are overlaid on the main window. The top dialog box is empty, with 'Value:' and 'Label:' fields and an 'Add' button. The bottom dialog box shows the 'Value:' field set to '2' and the 'Label:' field set to 'female'. A list box below contains the entry '1.00 = "male"'. Both dialog boxes have 'Spelling...', 'Add', 'Change', 'Remove', 'OK', 'Cancel', and 'Help' buttons.



# Enter cases

\*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

10 : sex 1.00 Visible: 1 of 1 Variables

	sex	var	var	var	var	var	var	var	var	var	v
1	male										
2	female										
3	male										
4	male										
5	male										
6	male										
7	male										
8	female										
9	female										
10	male										
11											
12											
13											

1. One variables in the data set.  
2. This is: Sex: 1 = male, 2 = female

Under Data View

Data View Variable View

IBM SPSS Statistics Processor is ready Unicode:ON

# PRACTICE



How would you put the following information in to SPSS ??

Sex	Age	Diabetes History	Hypertension history	Hyperlipidemia history	Smoking history
Female	55	Yes	No	Yes	Yes
Male	53	Yes	Yes	Yes	Yes
Female	35	Yes	No	Yes	No
Female	45	No	No	Yes	No
Female	53	No	No	No	Yes
Male	47	Yes	Yes	Yes	Yes
Male	38	Yes	Yes	Yes	Yes
Female	46	No	Yes	No	Yes
Female	36	Yes	No	No	No
Male	51	No	Yes	Yes	Yes

# Data used in the workshop

## Variable:

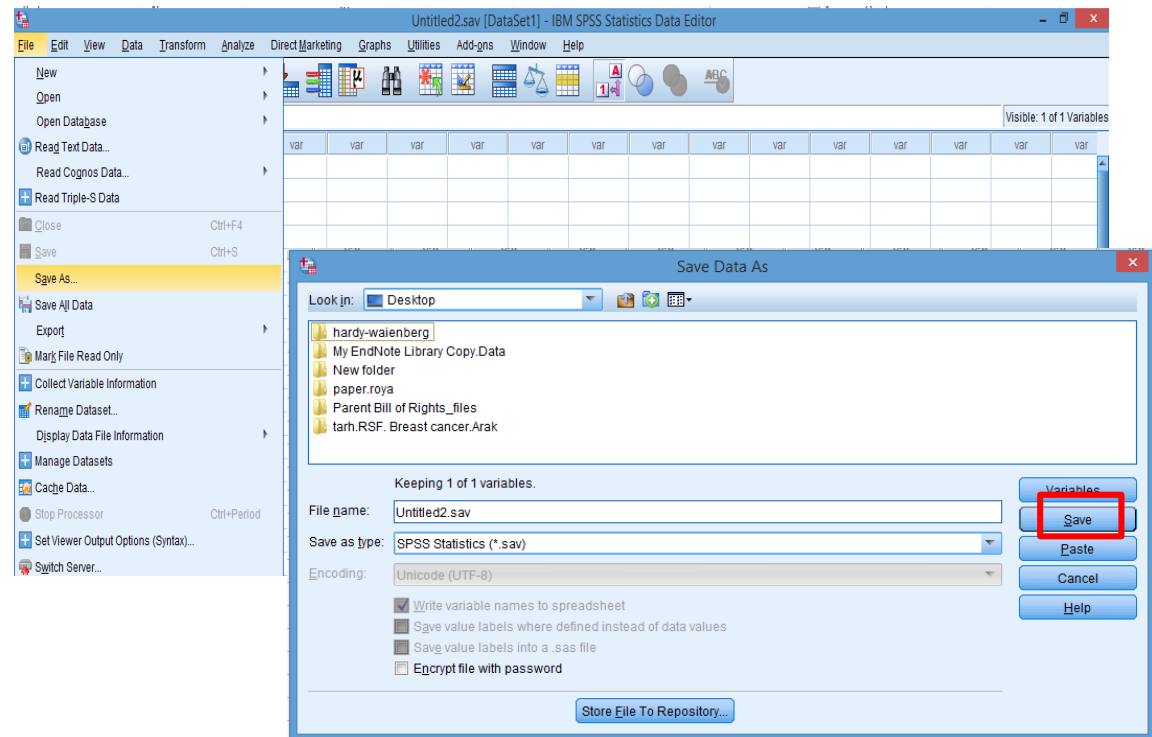
- Sex
- Age
- Diabetes History
- Hypertension history
- Hyperlipidemia history
- Smoking history

	ID	Sex	age	DM	HTN	HLP	HSmok	var	var	var
1	1	Female	55	Yes	No	Yes	Yes			
2	2	Male	53	Yes	Yes	Yes	Yes			
3	3	Female	35	Yes	No	Yes	No			
4	4	Female	45	No	No	Yes	No			
5	5	Female	53	No	No	No	Yes			
6	6	Male	47	Yes	Yes	Yes	Yes			
7	7	Male	38	Yes	Yes	Yes	Yes			
8	8	Female	46	No	Yes	No	Yes			
9	9	Female	36	Yes	No	No	No			
10	10	Male	51	No	Yes	Yes	Yes			
11										
12										
13										

# Saving the file

The file must always be saved in order to save the work that has been done to date:

- File/Save as
- Move to the target directory
- Enter a file name
- Save

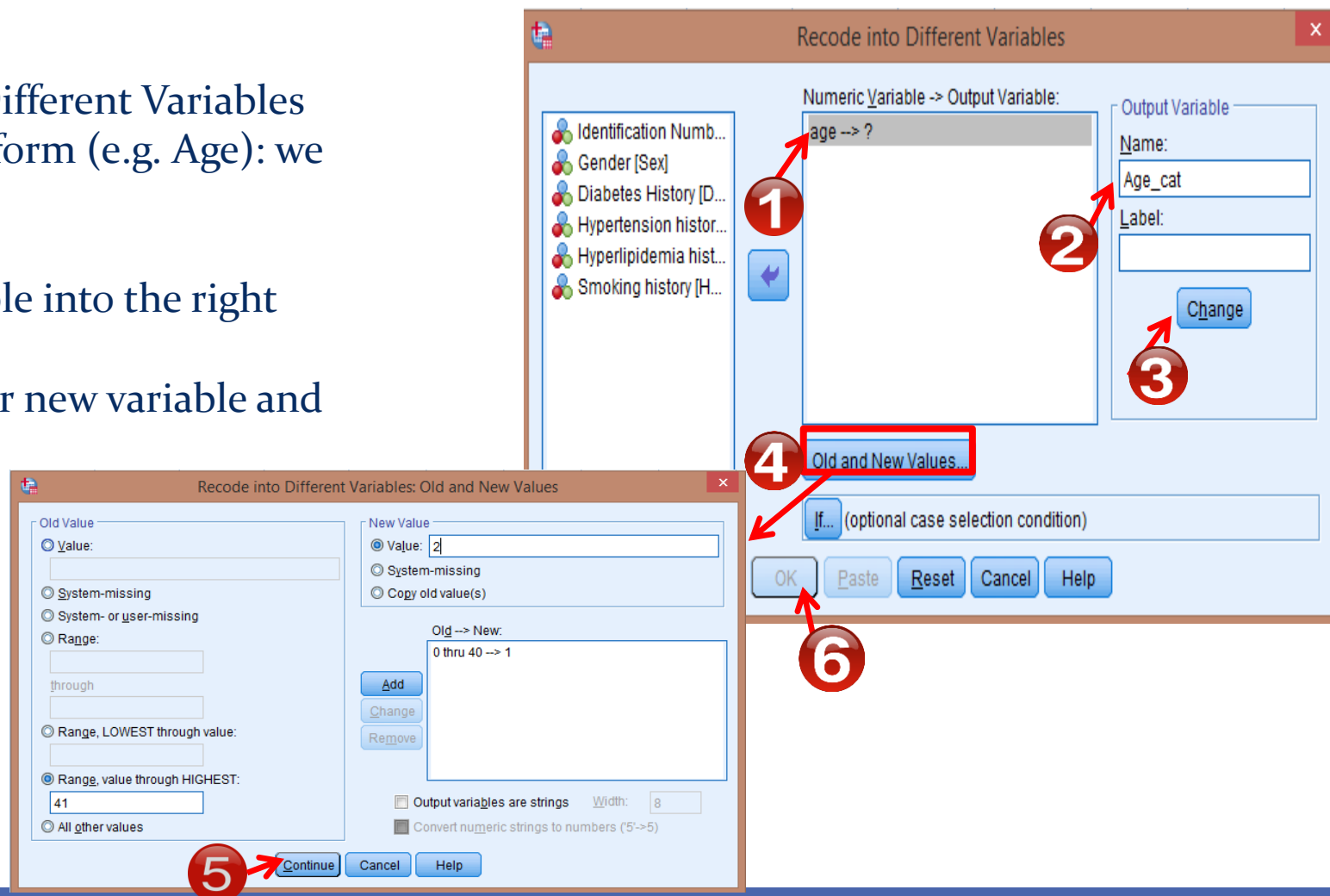


# Descriptive statistics

- Collection
- Organization
- Summarize data
  - Tables
  - Graphs
  - Measures of Central Tendency
  - Index of dispersion

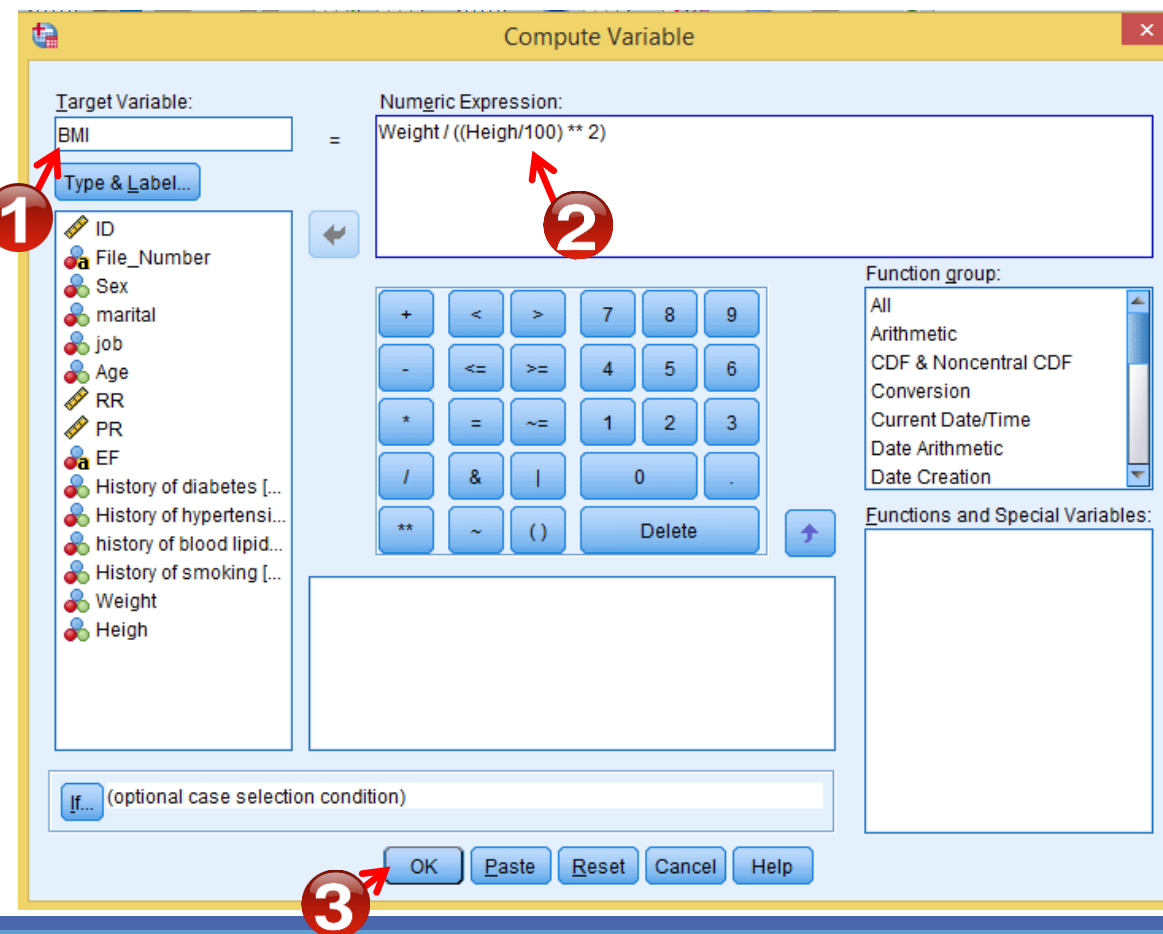
# Variable transformation

1. Select Transform → Recode into Different Variables
2. Select variable that you want to transform (e.g. Age): we want  
1:  $= < 40$  and 2:  $> 41$
3. Click Arrow button to put your variable into the right window
4. Under Output Variable: type name for new variable and label, then click Change
5. Click Old and New Values



# Variable transformation

- Compute variable
- Example 1. Create a new variable: BMI
- Use Weight, Height
- Go to Transform  $\longrightarrow$  Compute Variable **1**



# Sort and select cases

- Sort cases

- Sort cases by variables: Data  $\longrightarrow$  Sort Cases
- You can use Sort Cases to find missing.

\*Data Entry.sav [DataSet4] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

12: HSmok Visible: 8 of 8 Variables

	ID	Sex	age	DM	HTN	HLP	HSmok	Age_cat	var	var	var	var
1	1	Female	55	Yes	No	Yes	Yes	2.00				
2	2	Male	53	Yes	Y	Y	Y	0.00				
3	3	Female	35	Yes								
4	4	Female	45	No								
5	5	Female	53	No								
6	6	Male	47	Yes								
7	7	Male	38	Yes								
8	8	Female	46	No								
9	9	Female	36	Yes								
10	10	Male	51	No								
11												
12												
13												
14												
15												
16												
17												
18												
19												

Sort Cases

Sort by:

Gender [Sex]

Age of patients at di...

Diabetes History [D...

Hypertension histor...

Hyperlipidemia hist...

Smoking history [H...

Age\_cat

Sort Order

Ascending

Descending

Save Sorted Data

Save file with sorted data

File...

Create an index

OK Paste Reset Cancel Help



# Sort and select cases

- Select cases
  - Example 1. Select Females for analysis.
  - Go to Data → Select Cases
  - Under Select: check the second one
  - Click If button

\*Data Entry.sav [DataSet4] - IBM SPSS Statistics Data Editor

	ID	Sex	age	DM	HTN	HLP	HSmok	Age_cat	var	var	var
1	1	Female	55	Yes	No	Yes	Yes	2.00			
2	2	Male	53	Yes							
3	3	Female	35	Yes							
4	4	Female	45	No							
5	5	Female	53	No							
6	6	Male	47	Yes							
7	7	Male	38	Yes							
8	8	Female	46	No							
9	9	Female	36	Yes							

Select Cases

Select

All cases

If condition is satisfied

Random sample of cases

Based on time or case range

Use filter variable:

Output

Filter out unselected cases

Copy selected cases to a new dataset

Dataset name: \_\_\_\_\_

Delete unselected cases

not filter cases

OK Paste Reset Cancel Help

Select Cases: If

Sex=1

Function group:

All

Arithmetic

CDF & Noncentral CDF

Conversion

Current Date/Time

Date Arithmetic

Date Creation

Functions and Special Variables:

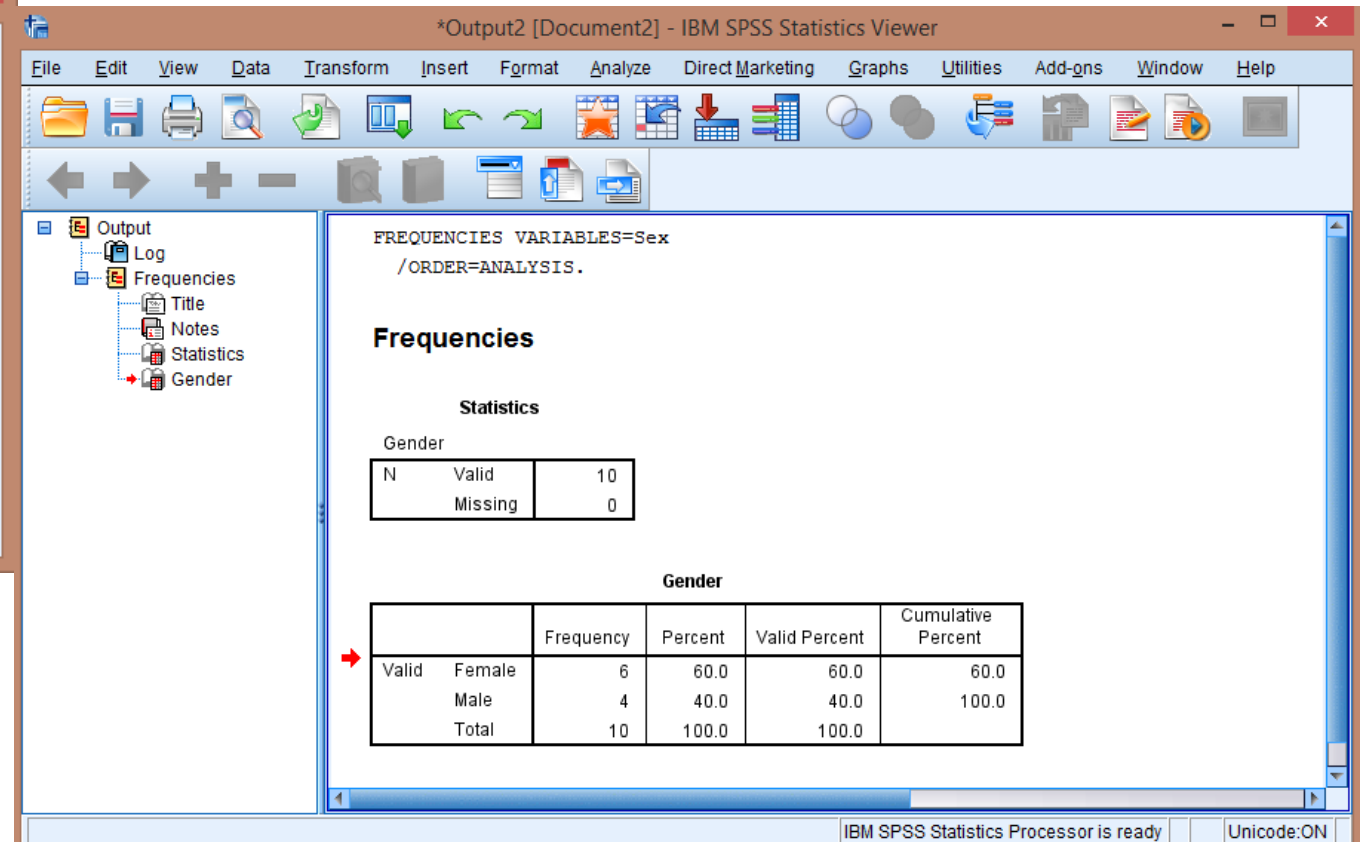
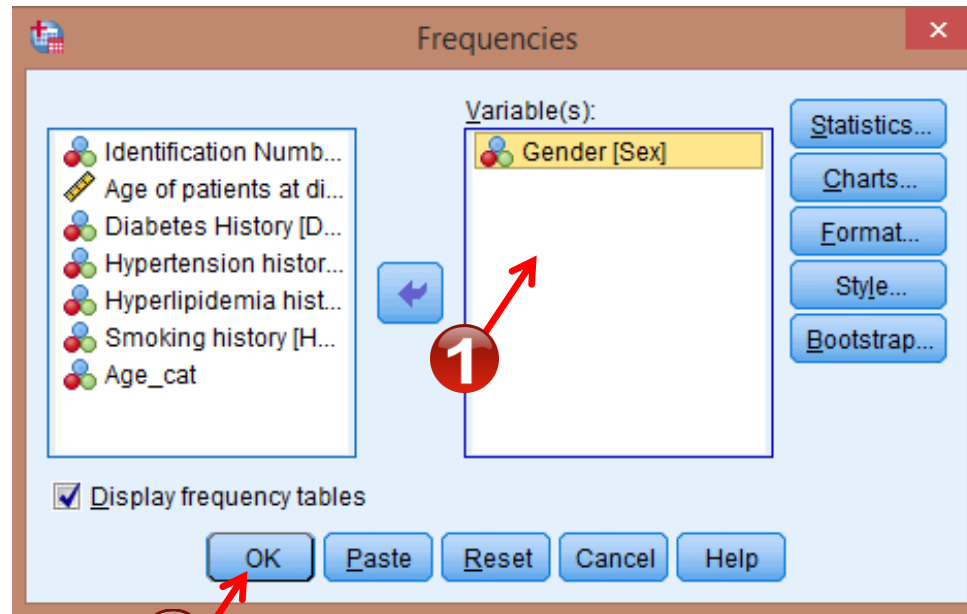
Continue Cancel Help

# Descriptive statistics

- Collection
- Organization
- Summarize data
  - Tables
  - Graphs
  - Measures of Central Tendency
  - Index of dispersion

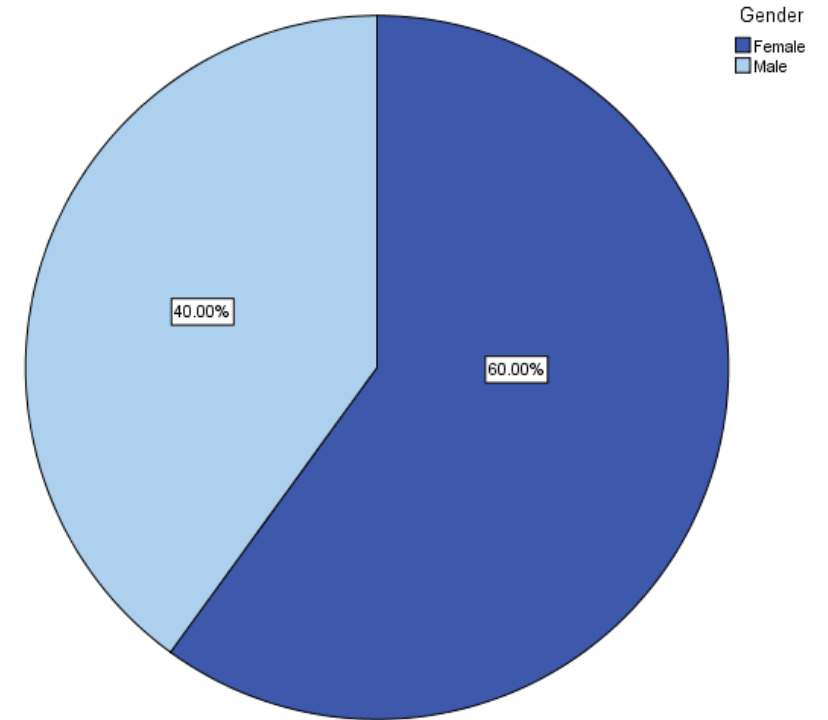
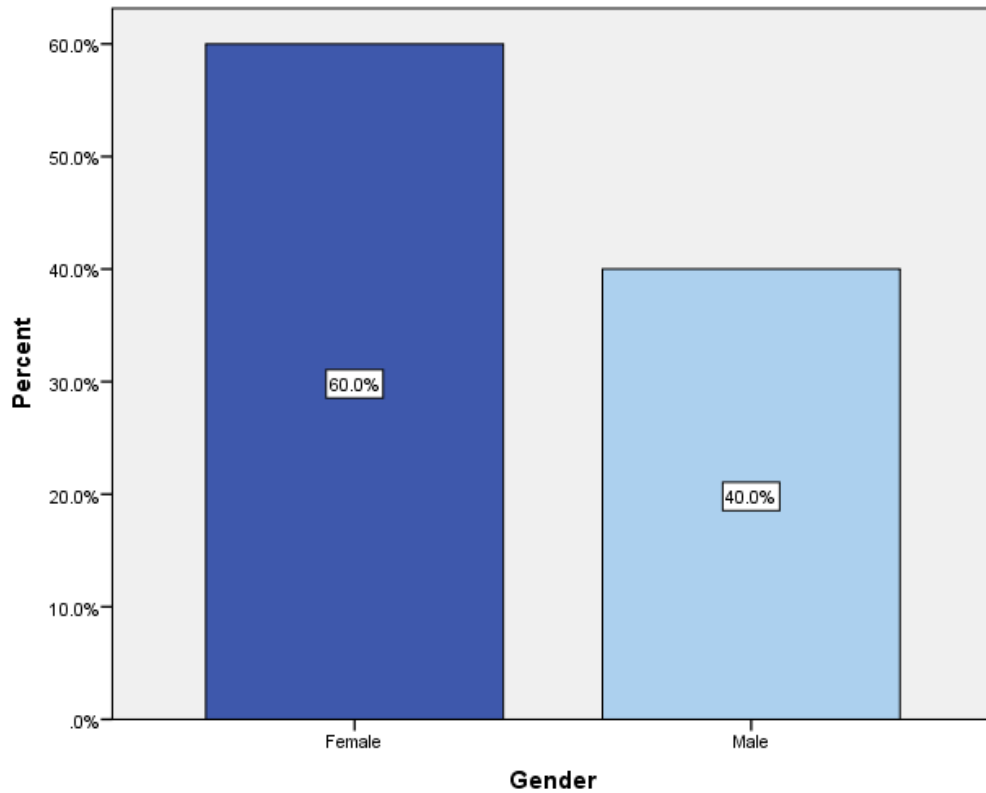
# Descriptive statistics: Tables

 Analyze → Descriptive statistics → Frequency



# Descriptive statistics: Graphs

$\sum_{+}^{\alpha}$  Graphs  $\longrightarrow$  Legacy Dialogs  $\longrightarrow$  Bar, Pie, ...



# Descriptive statistics: Measures of Central Tendency & Index of dispersion

$\Sigma \alpha$  Analyze → Descriptive statistics → Descriptive

1

2

3

→ Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Age of patients at diagnosis	10	35	55	45.90	7.385
Valid N (listwise)	10				

# Import data from Excel

- Select File → Open → Data
- Choose **Excel** as file type
- Select the file you want to import
- Then click Open

