Introduction to SPSS and data management

Αθανάσιος Κυργίδης

Παναγιώτης Μαυρίδης

Κωνσταντίνος Τζιόμαλος

Κωνσταντίνος Τουλής

Lecture Outline
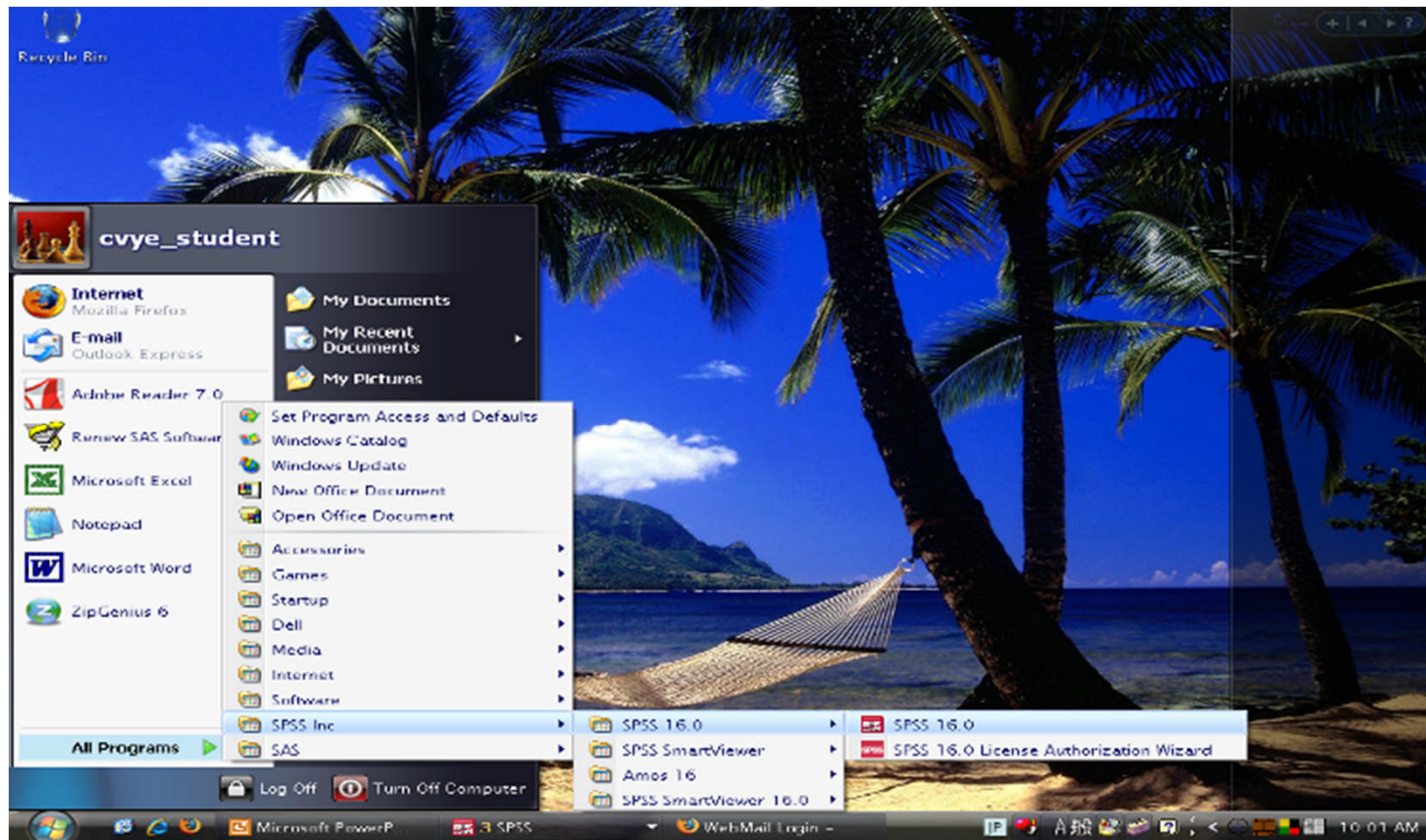
- Introduction
- Basics of managing data files
- Normality Explorations
- Basic Analysis
  - Descriptive Statistics
  - Analytical Statistics
- Exporting Output
- Hands On (Λυσάρι Ασκήσεων)

The basics of managing data files

Opening SPSS

- Start → All Programs → SPSS Inc→ SPSS 19.0 →
  SPSS 19.0

PASW Statistics

What would you like to do?

- ◉ Open an existing data source
  - More Files...

- ○ Open another type of file
  - More Files...

- ○ Run the tutorial

- ○ Type in data

- ○ Run an existing query
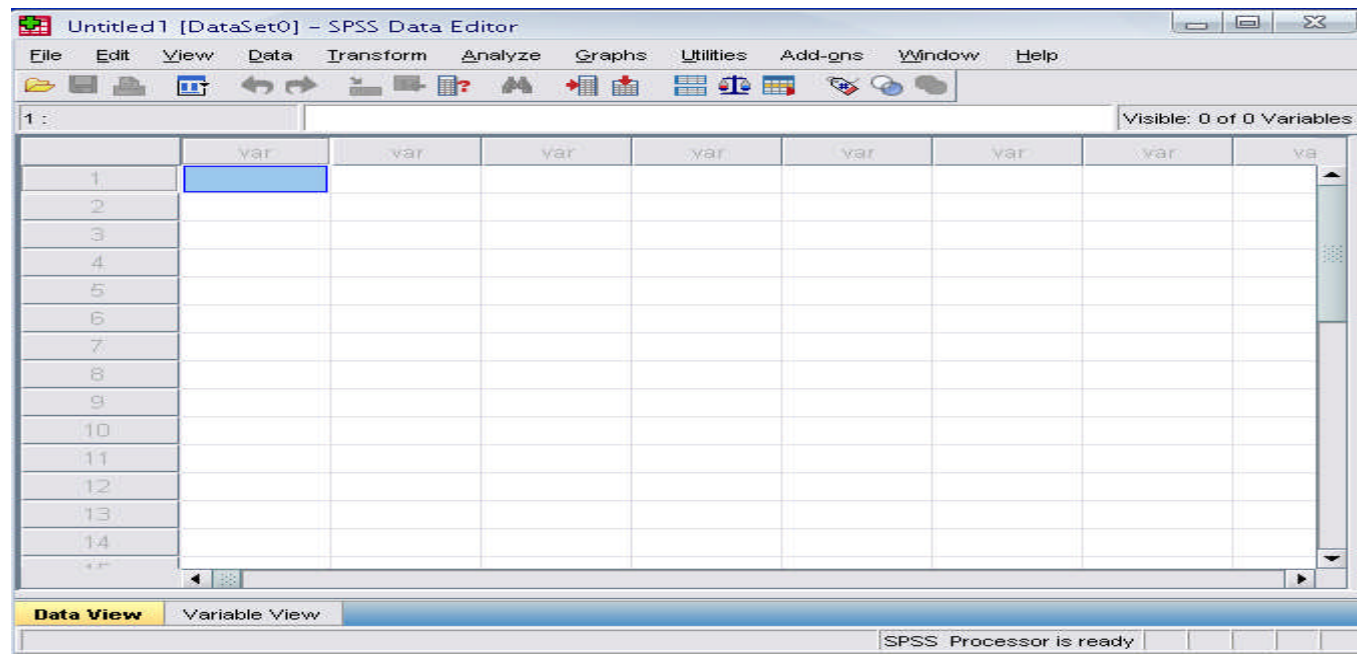
- ○ Create new query using Database Wizard

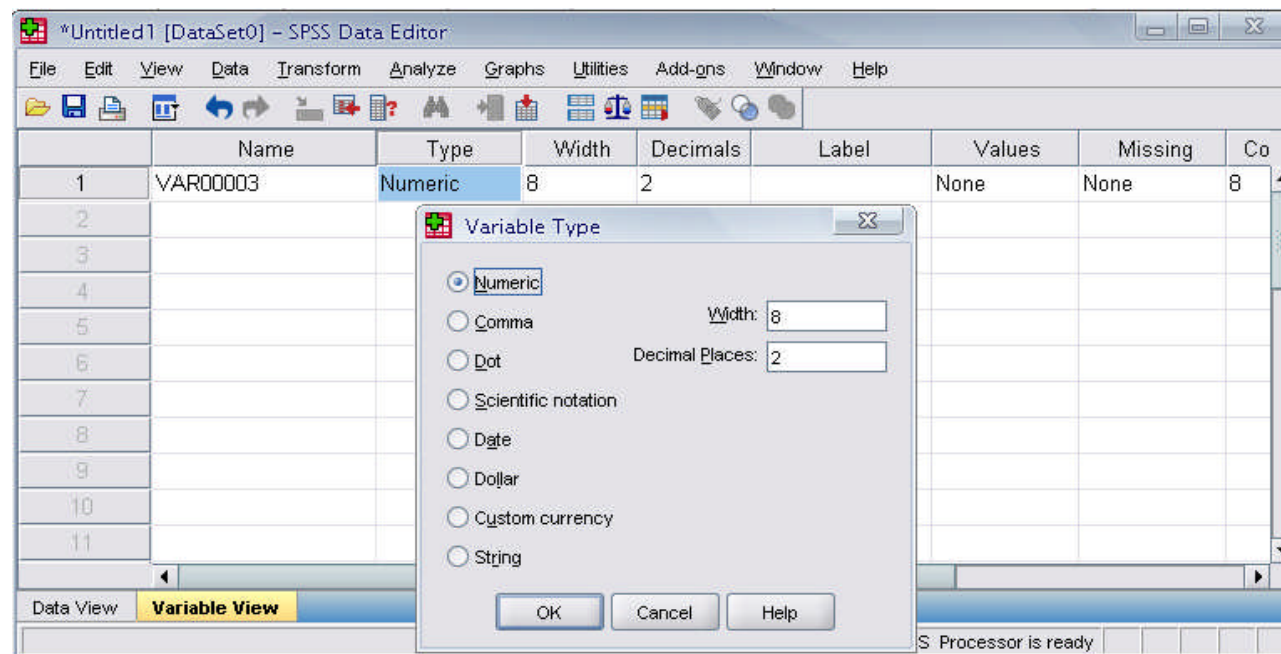☐ Don't show this dialog in the future

[OK] [Cancel]

Opening SPSS

- The default window will have the data editor
- There are two sheets in the window:

  1. Data view   2. Variable view

Variable View window: Type

- Type
  - Click on the 'type' box. The two basic types of variables that you will use are numeric and string. This column enables you to specify the type of variable.

Variable View window: Width

- Width
  - Width allows you to determine the number of characters SPSS will allow to be entered for the variable
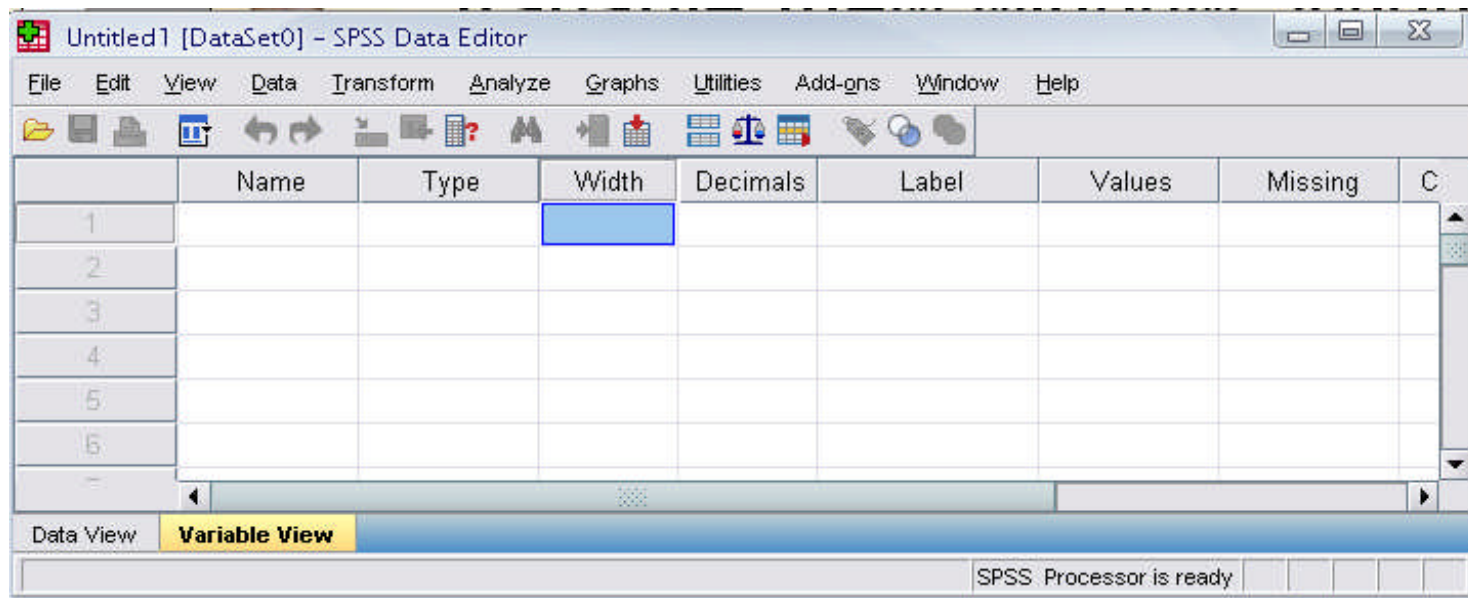
Variable View window: Decimals

- Decimals
  - Number of decimals
  - It has to be less than or equal to 16
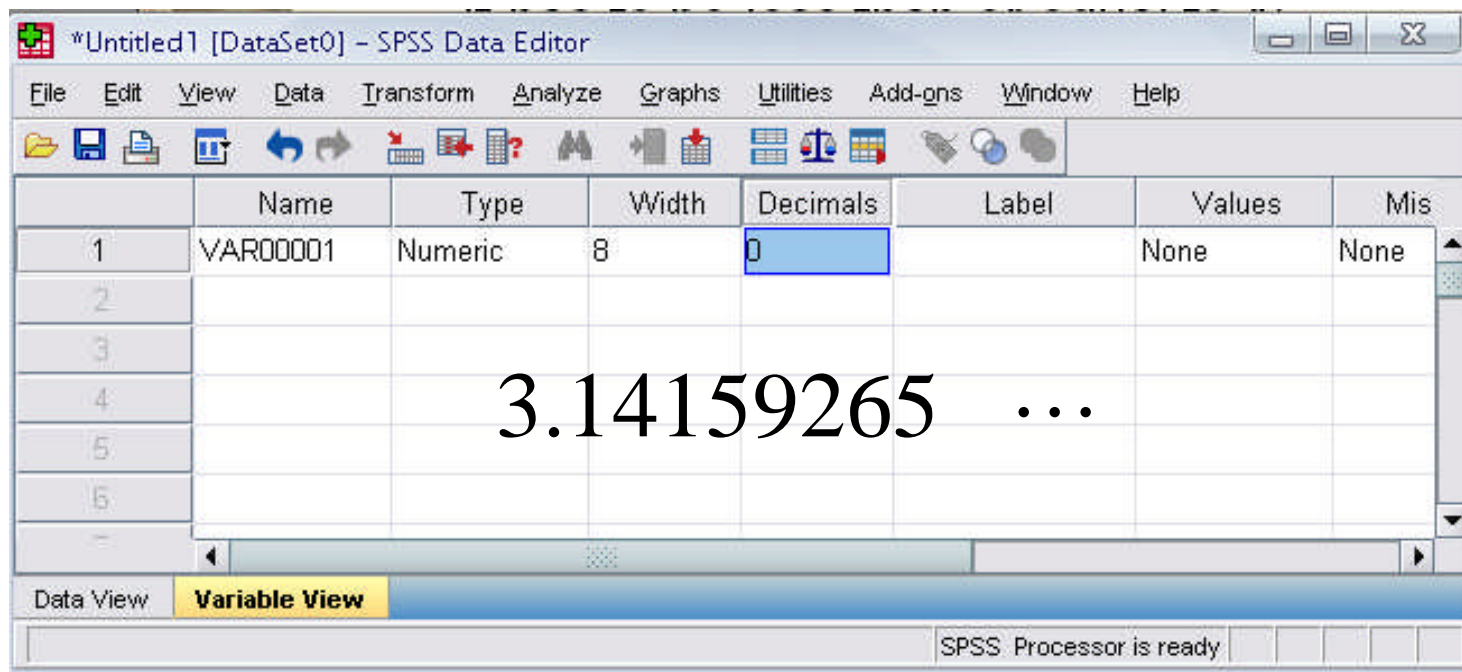
Variable View window: Label

- Label
  - You can specify the details of the variable
  - You can write characters with spaces up to 256 characters
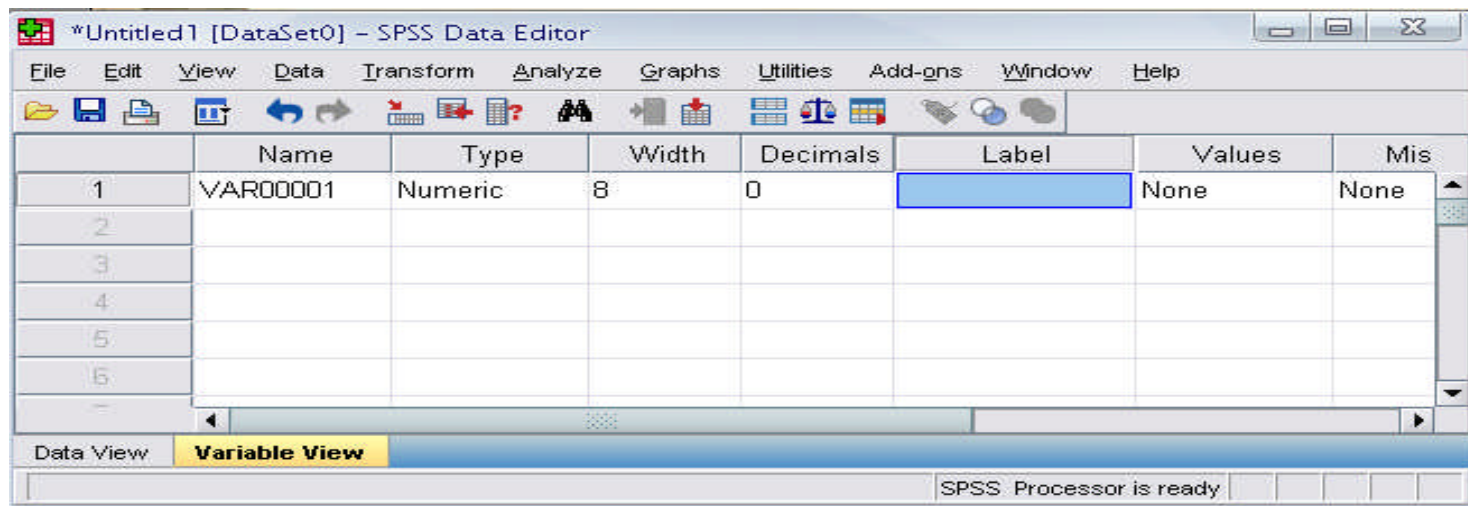
Variable View window: Values

- Values
  - This is used and to suggest which numbers represent which categories when the variable represents a category
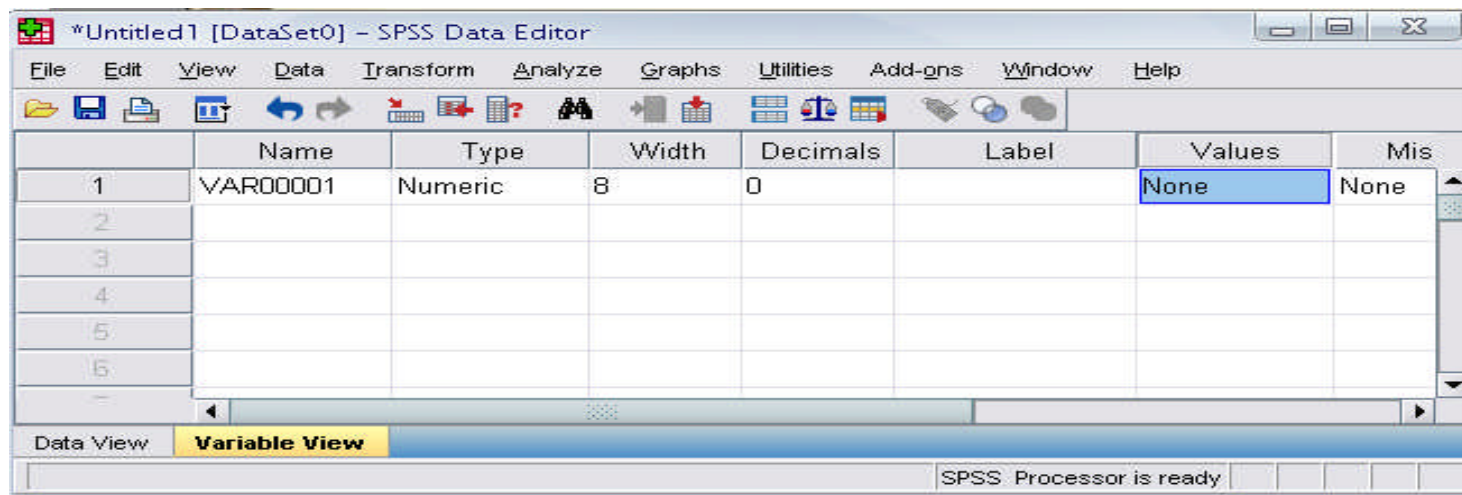
Defining the value labels

- Click the cell in the values column as shown below

- For the value, and the label, you can put up to 60 characters.

- After defining the values click add and then click OK.
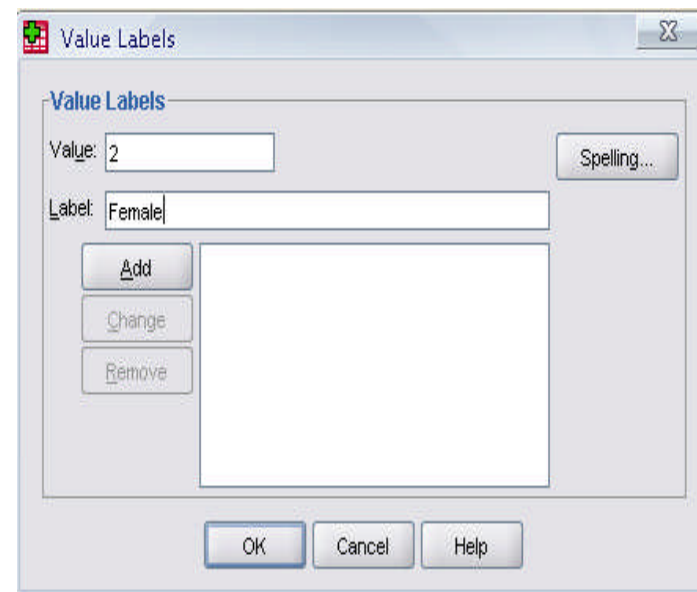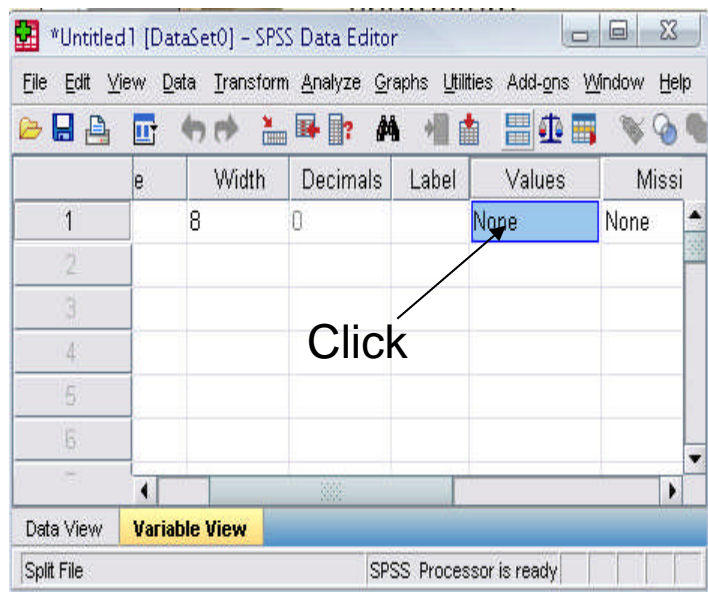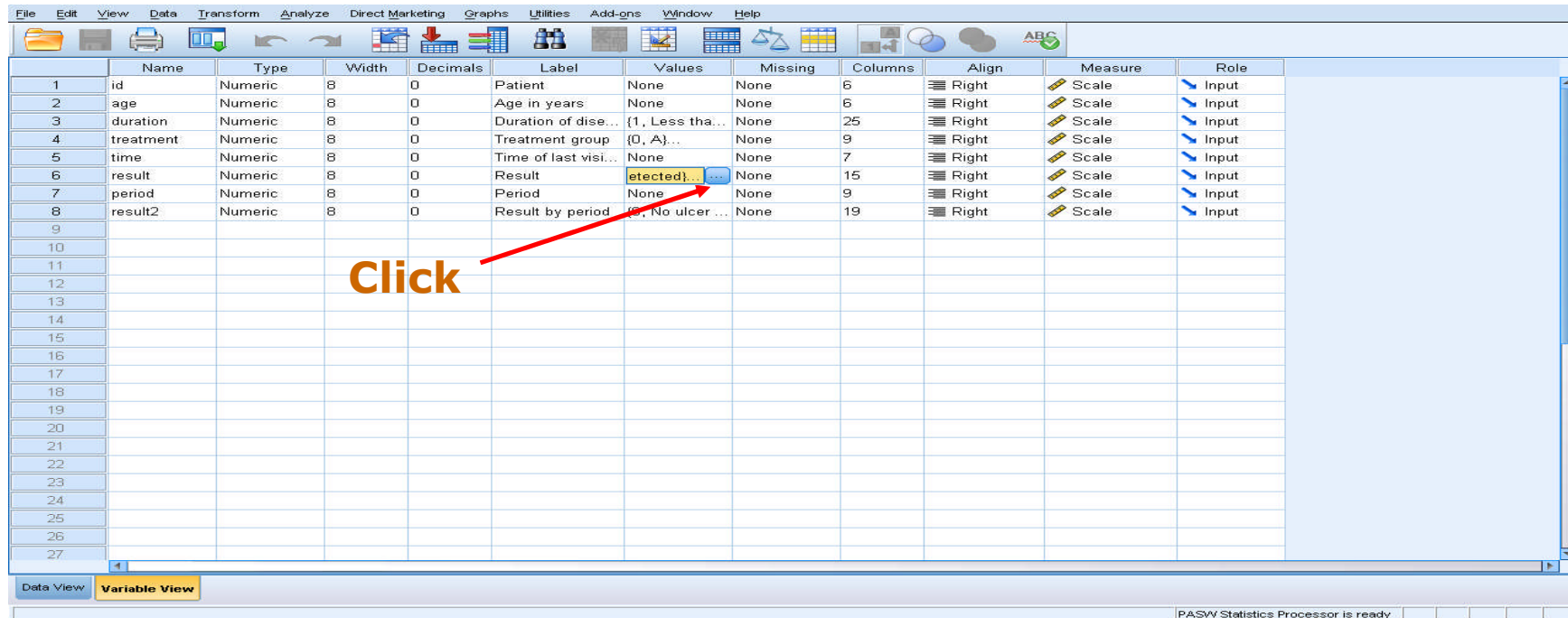
How to make a new data set?

- We will create now a data set on our own.

- Types of variables
  - String
  - Numeric

- Are you ready for this creation procedure?

Open "Άσκηση 3. ulcer_recurrence"
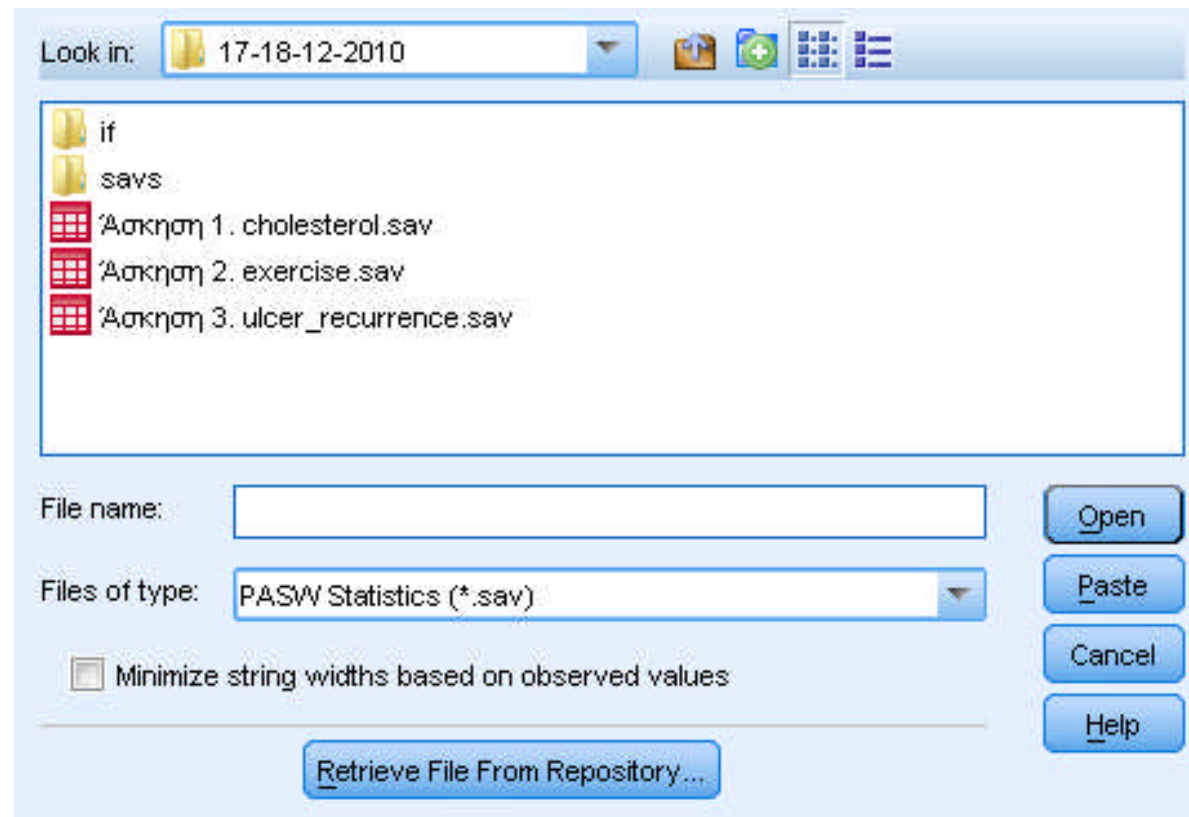
Example 1

Variable View:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	id	Numeric	8	0	Patient	None	None	6	Right	Scale	Input
2	age	Numeric	8	0	Age in years	None	None	6	Right	Scale	Input
3	duration	Numeric	8	0	Duration of dise...	{1, Less tha...	None	25	Right	Scale	Input
4	treatment	Numeric	8	0	Treatment group	{0, A}...	None	9	Right	Scale	Input
5	time	Numeric	8	0	Time of last visi...	None	None	7	Right	Scale	Input
6	result	Numeric	8	0	Result	etected}...	None	15	Right	Scale	Input
7	period	Numeric	8	0	Period	None	None	9	Right	Scale	Input
8	result2	Numeric	8	0	Result by period	{0, No ulcer ...	None	19	Right	Scale	Input

Data View   Variable View

Click

PASW Statistics Processor is ready

Data View:

1 : id   5   Visible: 8 of 8 Variables

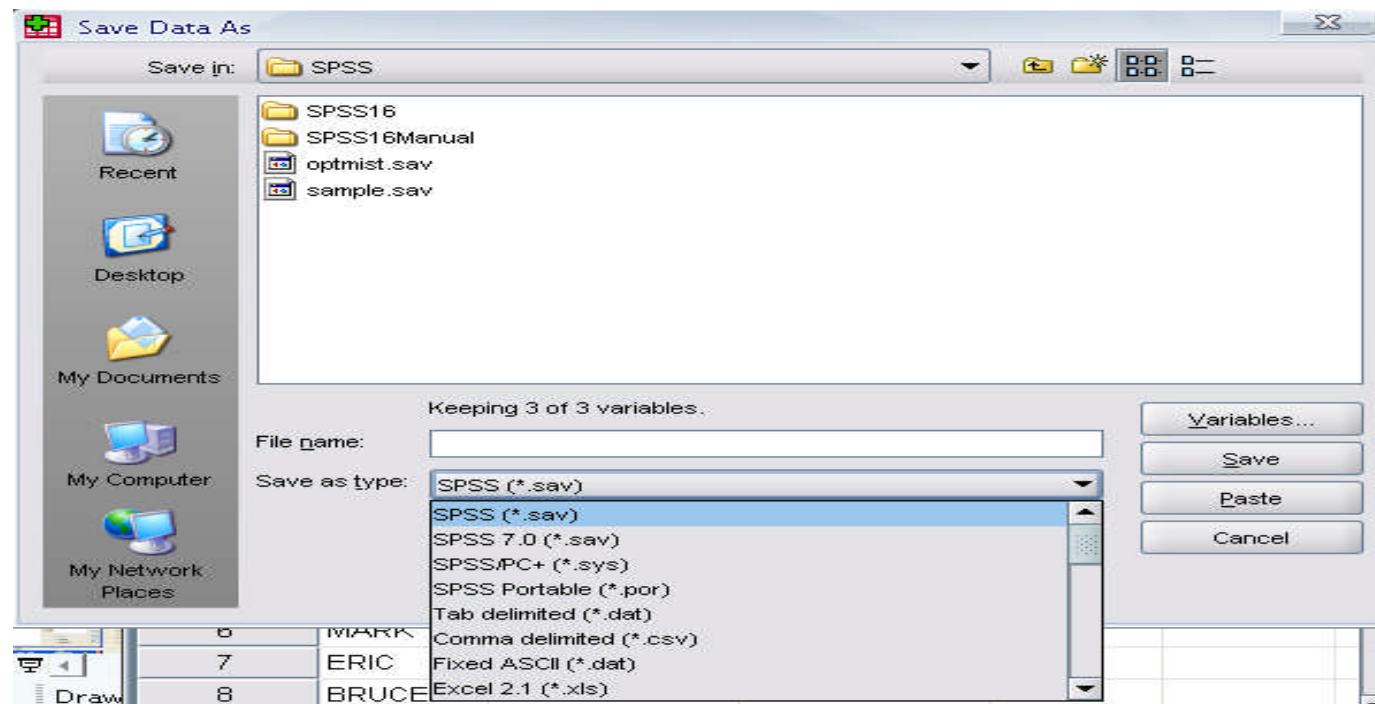|	id	age	duration	treatment	time	result	period	result2
1	5	56	Less than 5 years	A	12	No ulcer detected	1	No ulcer detected
2	8	41	Less than 5 years	A	12	No ulcer detected	1	No ulcer detected
3	14	27	Less than 5 years	A	6	Ulcer detected	1	Ulcer detected
4	16	54	Less than 5 years	A	6	No ulcer detected	1	No ulcer detected
5	21	25	Less than 5 years	A	12	No ulcer detected	1	No ulcer detected
6	22	58	Less than 5 years	A	12	No ulcer detected	1	No ulcer detected
7	24	41	Less than 5 years	A	12	No ulcer detected	1	No ulcer detected
8	26	58	Less than 5 years	A	3	Ulcer detected	1	Ulcer detected
9	29	72	Less than 5 years	A	12	No ulcer detected	1	No ulcer detected
10	33	76	Less than 5 years	A	12	No ulcer detected	1	No ulcer detected
11	37	44	Less than 5 years	A	12	Ulcer detected	1	No ulcer detected
12	40	23	Less than 5 years	A	12	No ulcer detected	1	No ulcer detected
13	42	61	Less than 5 years	A	12	No ulcer detected	1	No ulcer detected
14	2	73	Less than 5 years	B	12	No ulcer detected	1	No ulcer detected
15	3	54	Less than 5 years	B	12	No ulcer detected	1	No ulcer detected
16	7	71	Less than 5 years	B	12	No ulcer detected	1	No ulcer detected
17	9	23	Less than 5 years	B	12	No ulcer detected	1	No ulcer detected
18	10	37	Less than 5 years	B	5	Ulcer detected	1	Ulcer detected
19	11	38	Less than 5 years	B	12	No ulcer detected	1	No ulcer detected
20	15	47	Less than 5 years	B	6	Ulcer detected	1	Ulcer detected
21	17	38	Less than 5 years	B	10	Ulcer detected	1	No ulcer detected
22	20	75	Less than 5 years	B	12	No ulcer detected	1	No ulcer detected
23	23	63	Less than 5 years	B	12	No ulcer detected	1	No ulcer detected
24	25	47	Less than 5 years	B	12	No ulcer detected	1	No ulcer detected
25	30	59	Less than 5 years	B	12	Ulcer detected	1	No ulcer detected

Data View   Variable View

PASW Statistics Processor is ready

Saving the data

- To save the data file you created simply click 'file' and click 'save as.' You can save the file in different forms by clicking "Save as type."

Sorting the data
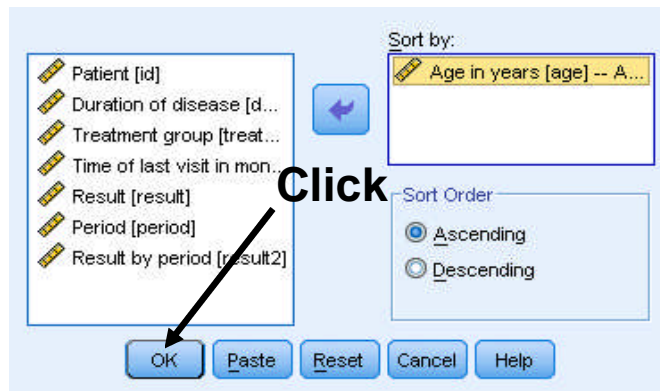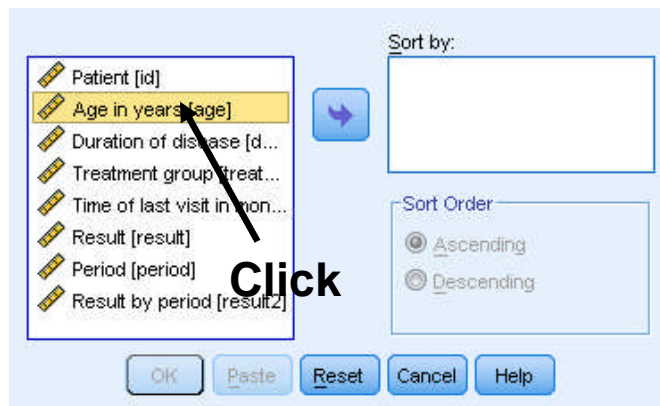
- Click 'Data' and then click Sort Cases

Sorting the data (cont'd)

- Double Click 'Name of the students.' Then click ok.

Example 2

- How would you sort the data by the time of their last visit of patients in descending order?

- Answer

  – Click data, sort cases, double click 'time of last visit in months,' click 'descending,' and finally click ok.

Transforming data

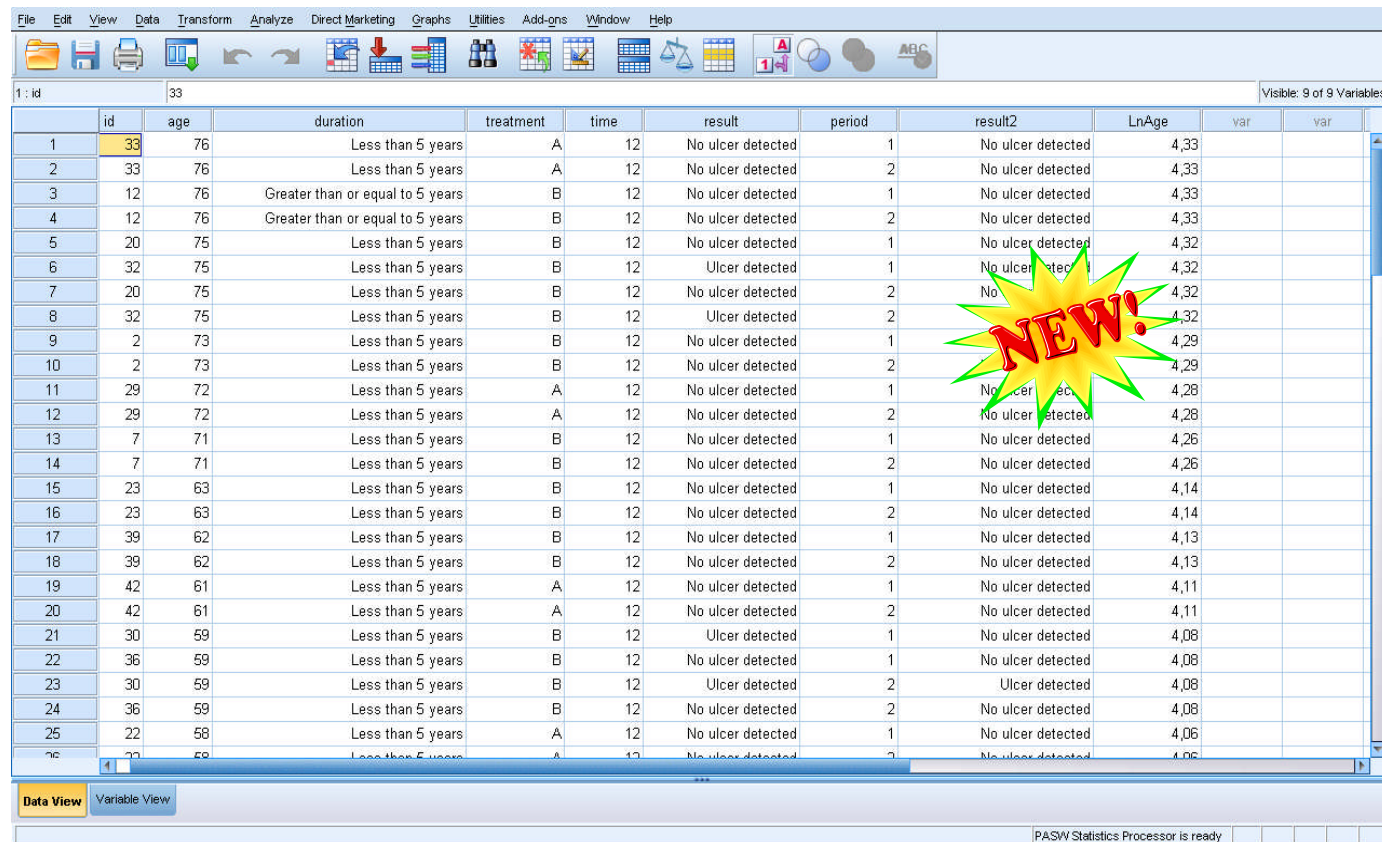- Click 'Transform' and then click 'Compute Variable…'

Transforming data (cont'd)

- Example: Adding a new variable named 'lnAge' which is the natural log of height
  - Type in lnAge in the 'Target Variable' box. Insert 'ln(Agen in years [age])' in the 'Numeric Expression' box. Click OK

Transforming data (cont'd)

- A new variable 'Inheight' is added to the table

Reverse Coding

- Transform > Recode > Into Different Variables
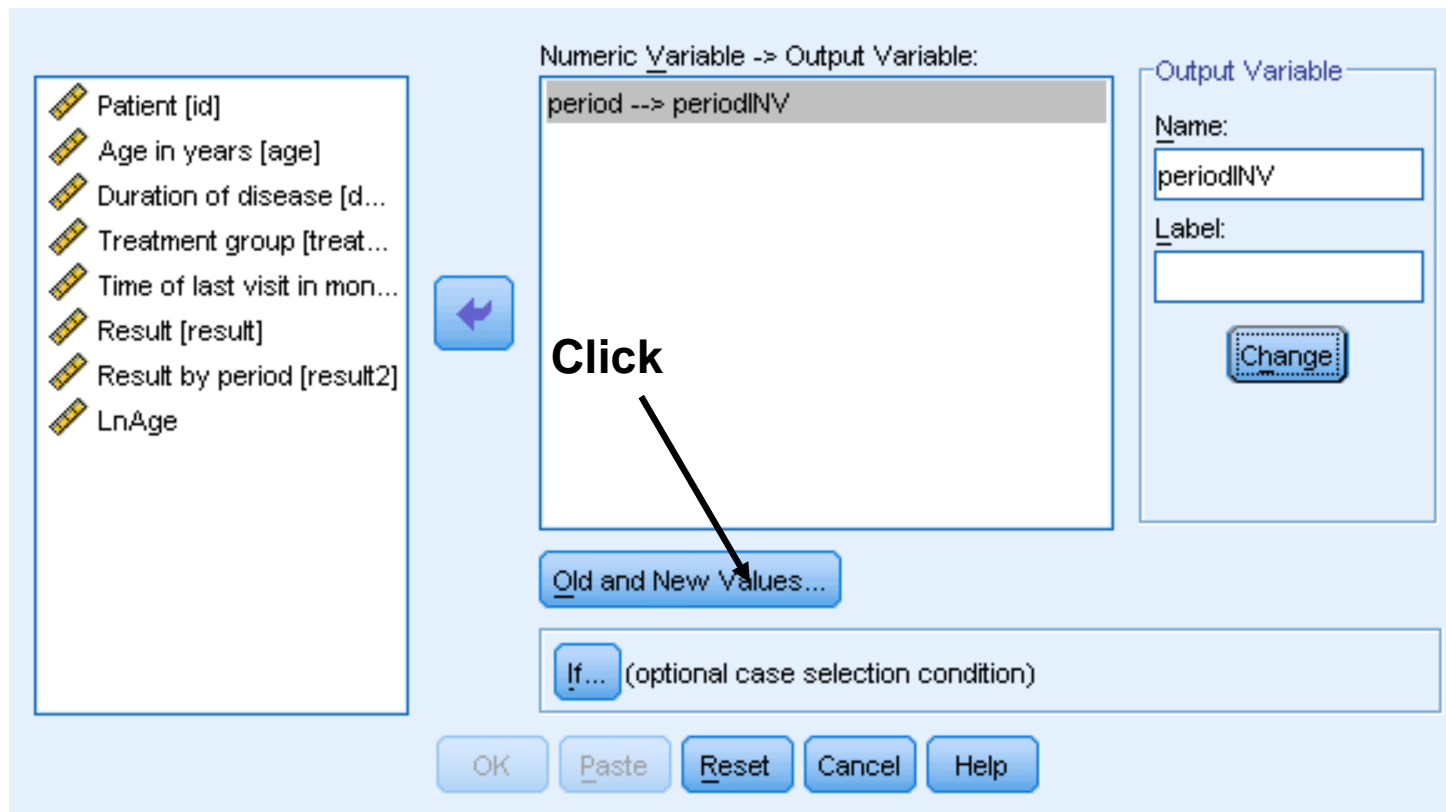
Reverse Coding

Example 3

- Create a new variable named "sqrtAge" which is the square root of Age

- Answer

Applications

- If the distribution turns out to be bimodal, a median or mean will be of little use.

- More often bimodal distributions give us a clear signal that we have lumped together two (or more) data sets which should have been separate.

  - Example: If women are paid less than men, income data for a given level of education turns to be bimodal.
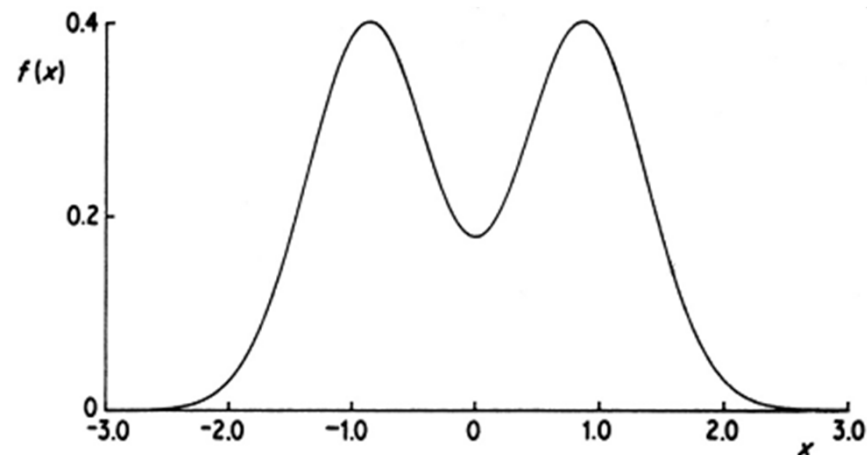
Normality explorations

Data Analysis

- Data analysis embraces both the problem of finding an appropriate model, on the one hand, and model estimation and testing, on the other.
- In this context normality assumption becomes important.
- In social sciences, it is hard to find typical bell shaped normal distribution.

Data Analysis

- Data analysis embraces both the problem of finding an appropriate model, on the one hand, and model estimation and testing, on the other.

- In this context normality assumption becomes important.

- In social sciences, it is hard to find typical bell shaped normal distribution.
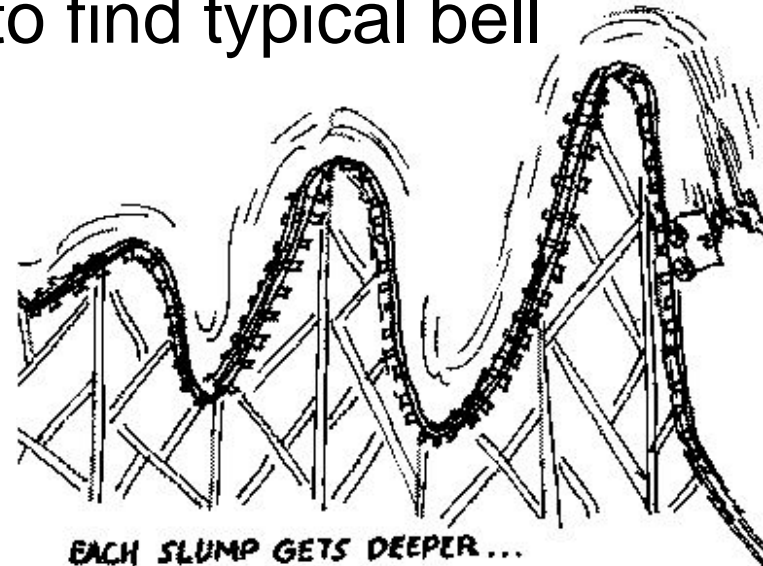
Data Analysis

- Data analysis embraces both the problem of finding an appropriate model, on the one hand, and model estimation and testing, on the other.
- In this context normality assumption becomes important.
- In social sciences, it is hard to find typical bell shaped normal distribution.

EACH SLUMP GETS DEEPER...

Open "Άσκηση 1. cholesterol"

Boxplots
- ○ Factor levels together
- ○ Dependents together
- ◉ None

Descriptive
- ☐ Stem-and-leaf
- ☐ Histogram

☑ Normality plots with tests

Spread vs Level with Levene Test
- ◉ None
- ○ Power estimation
- ○ Transformed   Power: Natural log
- ○ Untransformed

[Continue] [Cancel] [Help]

Tests of Normality

	Group	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Cholesterol	1	,063	75	,200[*]	,989	75	,740
	2	,126	30	,200[*]	,939	30	,088

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Patient
Group
Day

Dependent List:
Cholesterol

Factor List:

Label Cases by:

[Statistics...]
[Plots...]
[Options...]
[Bootstrap...]
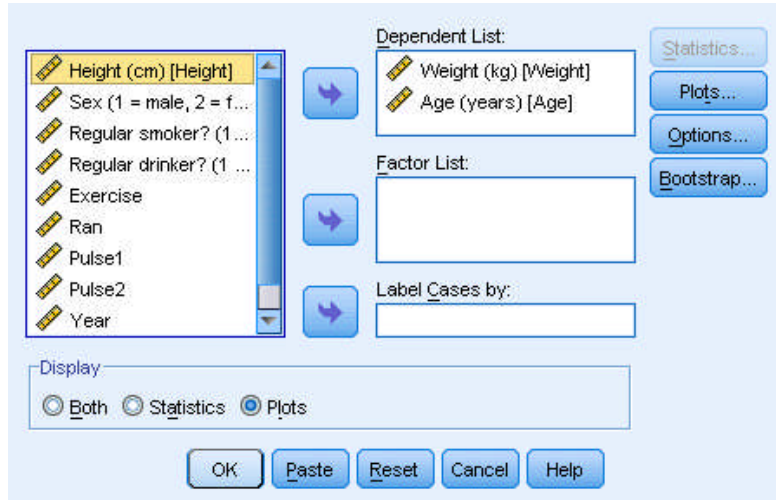
Display
- ○ Both ○ Statistics ◉ Plots

[OK] [Paste] [Reset] [Cancel] [Help]

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Cholesterol	,074	105	,193	,983	105	,198

a. Lilliefors Significance Correction

Open "Άσκηση 2. exercise"
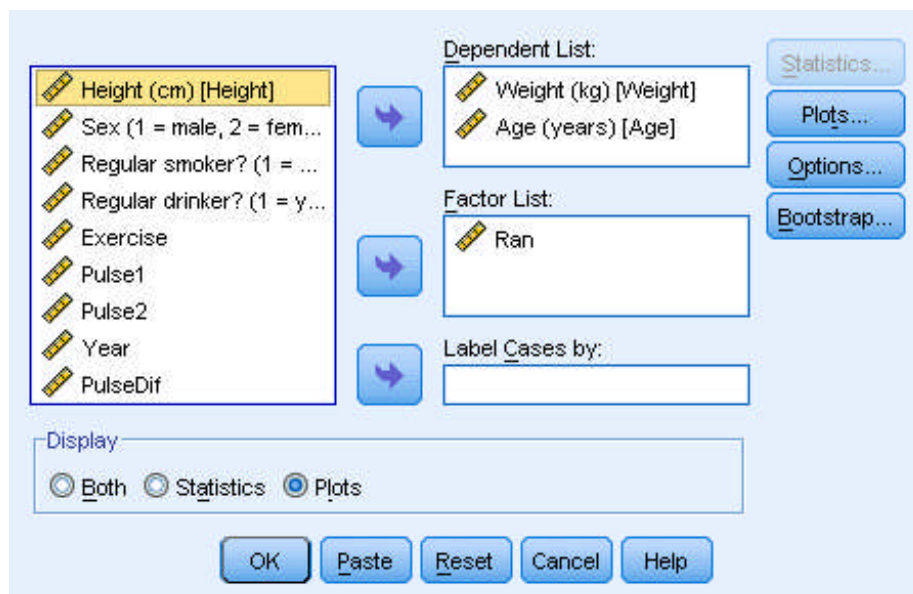

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Weight (kg)	,126	107	,000	,956	107	,001
Age (years)	,269	107	,000	,563	107	,000

a. Lilliefors Significance Correction

Open "Άσκηση 2. exercise"

Tests of Normality

	Ran	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Weight (kg)	1	,115	45	,166	,967	45	,230
	2	,141	62	,004	,922	62	,001
Age (years)	1	,325	45	,000	,638	45	,000
	2	,281	62	,000	,554	62	,000

a. Lilliefors Significance Correction