Introduction to SPSS

Object of the Course

- About the four-windows in SPSS
- The basics of managing data files
- The basic analysis in SPSS

Introduction: What is SPSS?

- Originally it is an acronym of Statistical Package for the Social Science but now it stands for Statistical Product and Service Solutions

- One of the most popular statistical packages which can perform highly complex data manipulation and analysis with simple instructions

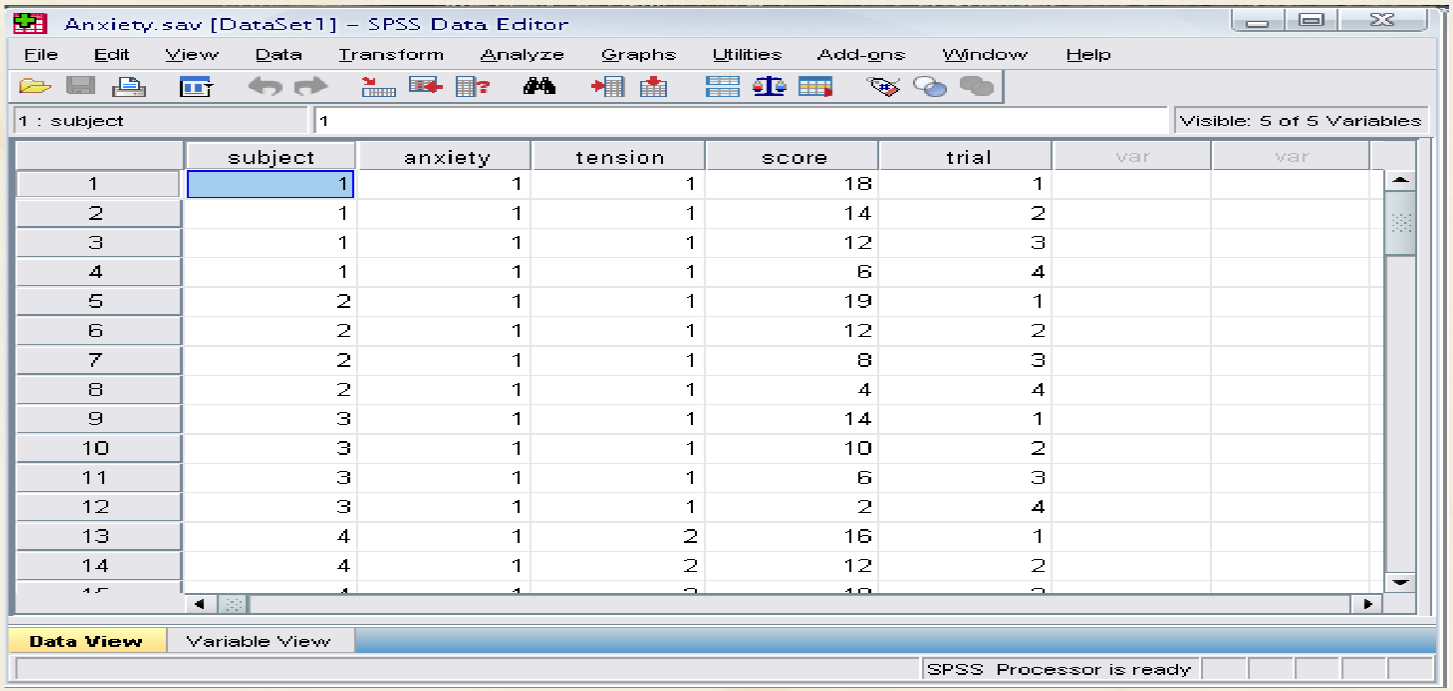
The Four Windows:

- Data editor
- Output viewer
- Syntax editor
- Script window

The Four Windows: Data Editor

- Data Editor

  Spreadsheet-like system for defining, entering, editing, and displaying data. Extension of the saved file will be "sav."

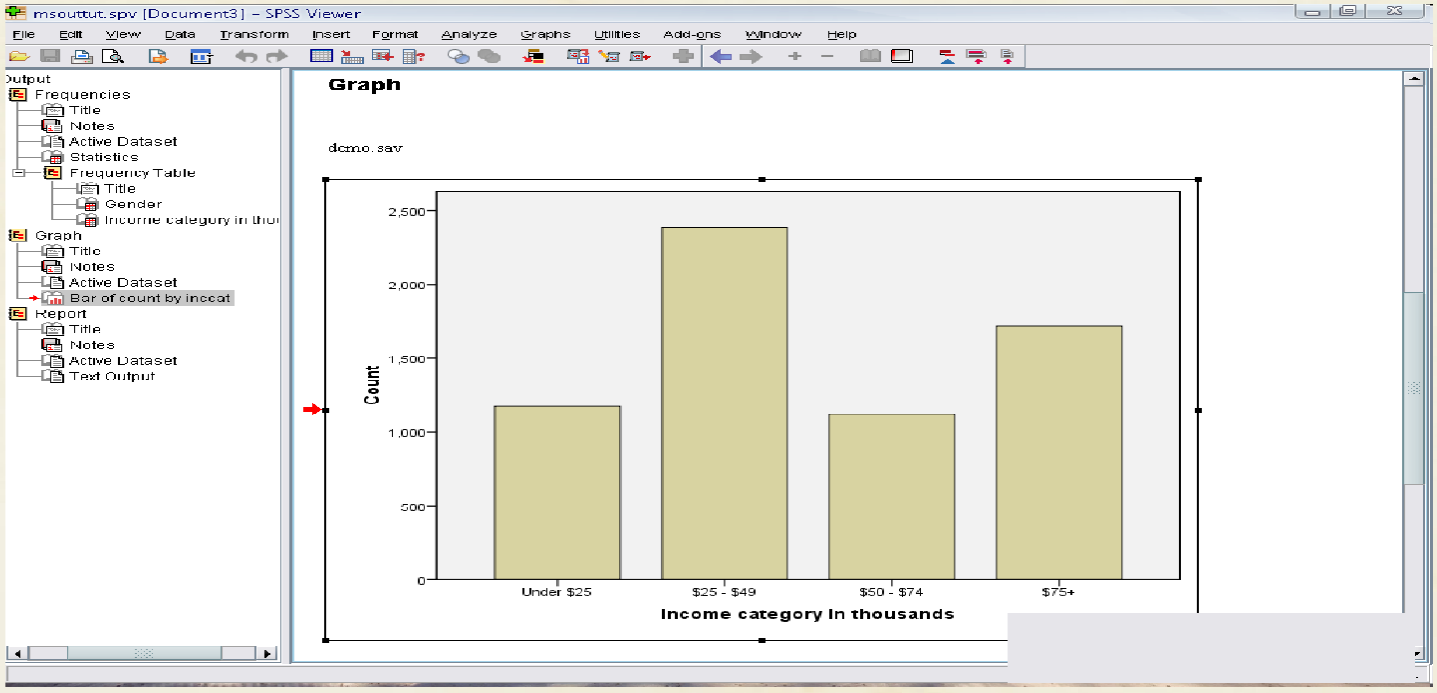


	subject	anxiety	tension	score	trial	var	var
1	1	1	1	18	1		
2	1	1	1	14	2		
3	1	1	1	12	3		
4	1	1	1	6	4		
5	2	1	1	19	1		
6	2	1	1	12	2		
7	2	1	1	8	3		
8	2	1	1	4	4		
9	3	1	1	14	1		
10	3	1	1	10	2		
11	3	1	1	6	3		
12	3	1	1	2	4		
13	4	1	2	16	1		
14	4	1	2	12	2		

The Four Windows: Output Viewer

- Output Viewer

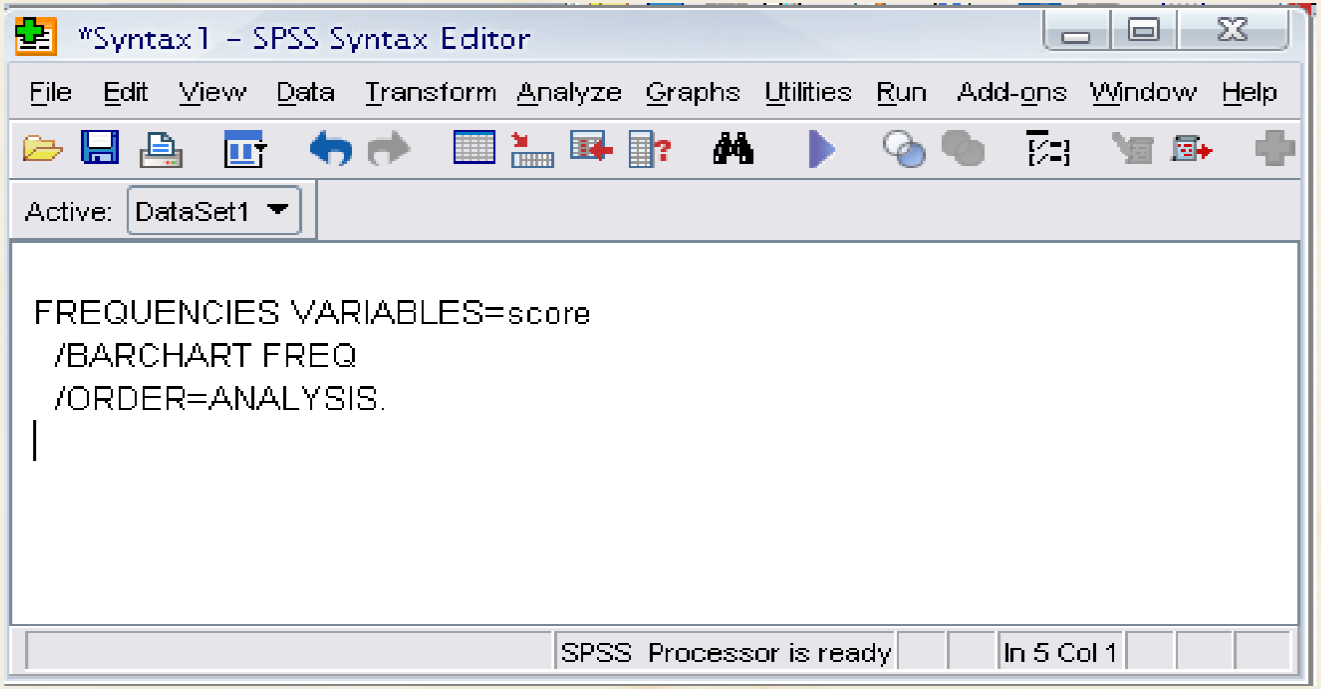
  Displays output and errors. Extension of the saved file will be “spv.”



The Four Windows: Syntax editor

- Syntax Editor

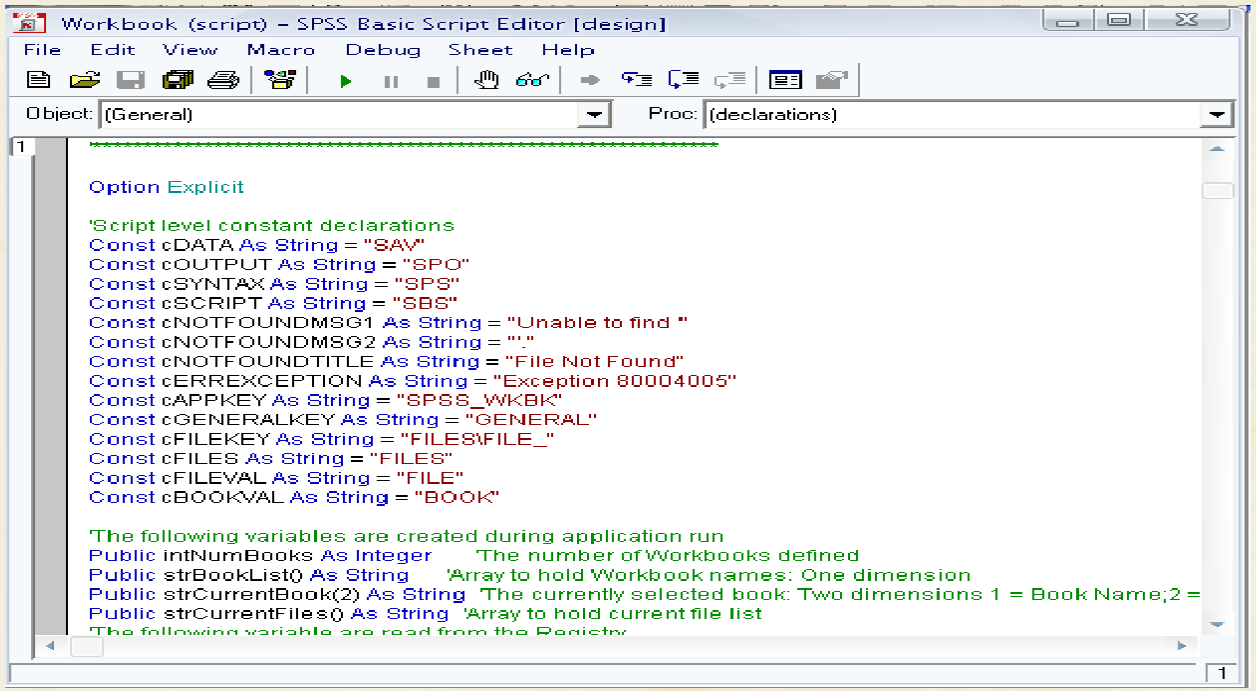
  Text editor for syntax composition. Extension of the saved file will be “sps.”



The Four Windows: Script Window

- Script Window

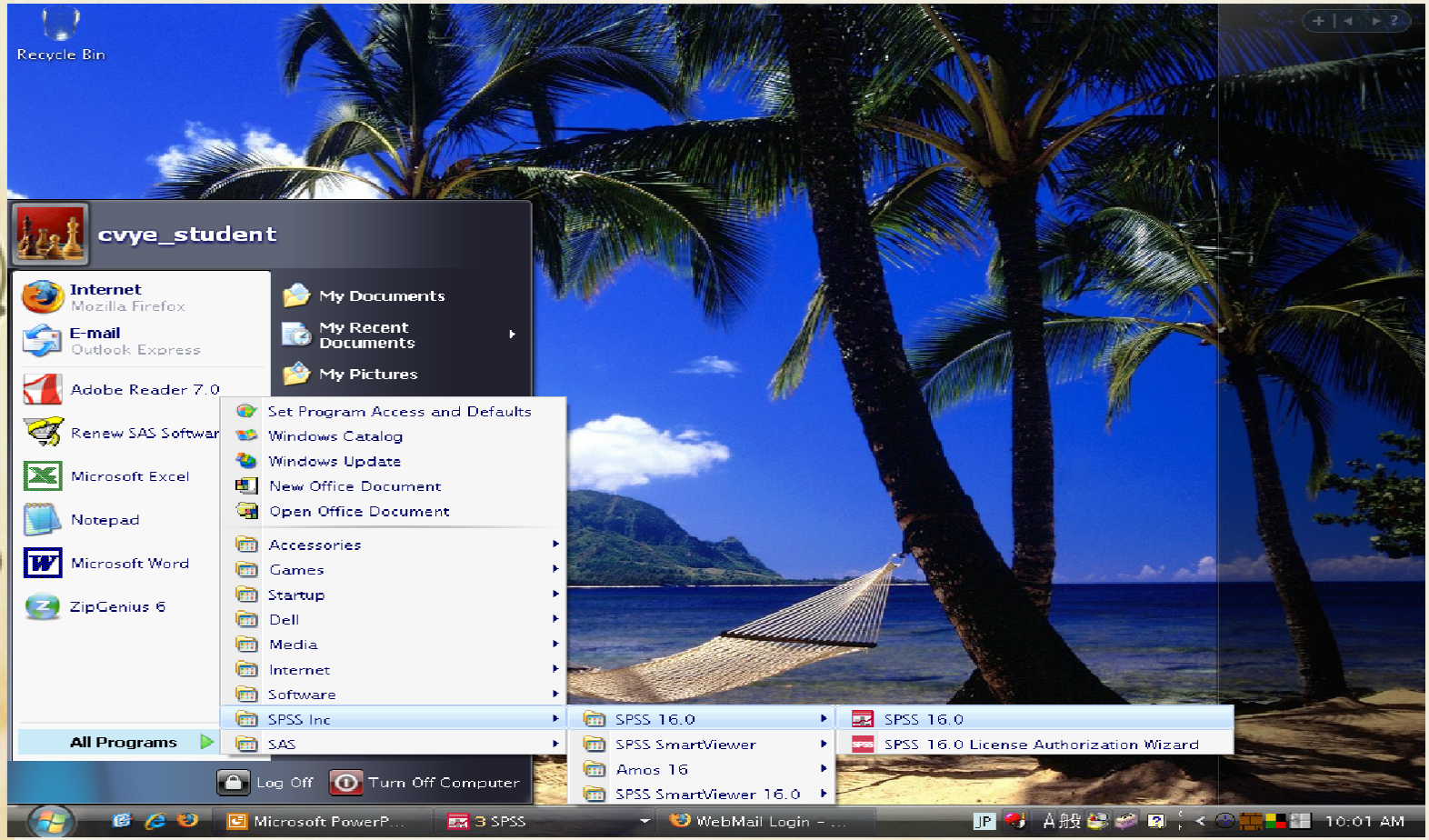
  Provides the opportunity to write full-blown programs, in a BASIC-like language. Text editor for syntax composition. Extension of the saved file will be “sbs.”




The basics of managing
data files

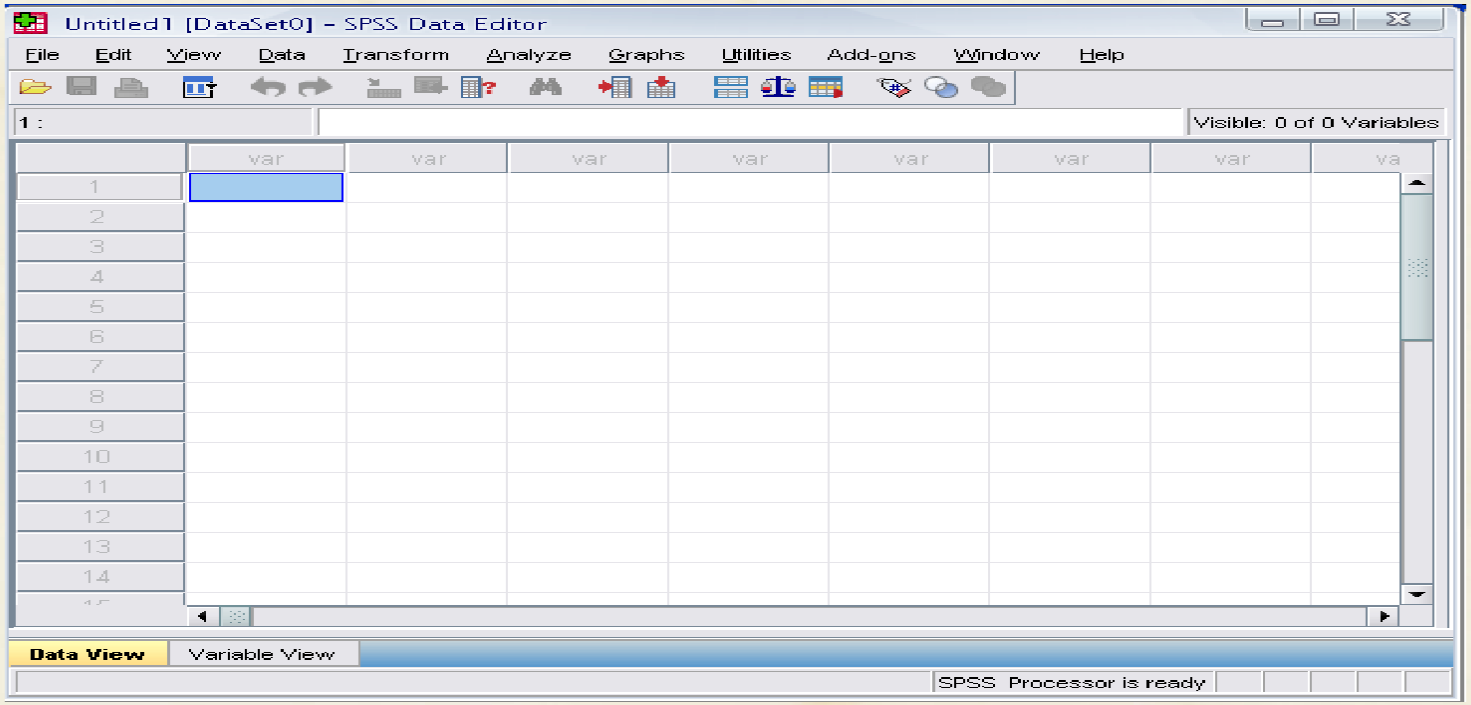
Opening SPSS

- Start → All Programs → SPSS Inc→ SPSS 16.0 → SPSS 16.0



Opening SPSS

- The default window will have the data editor
- There are two sheets in the window:
  1. Data view
  2. Variable view

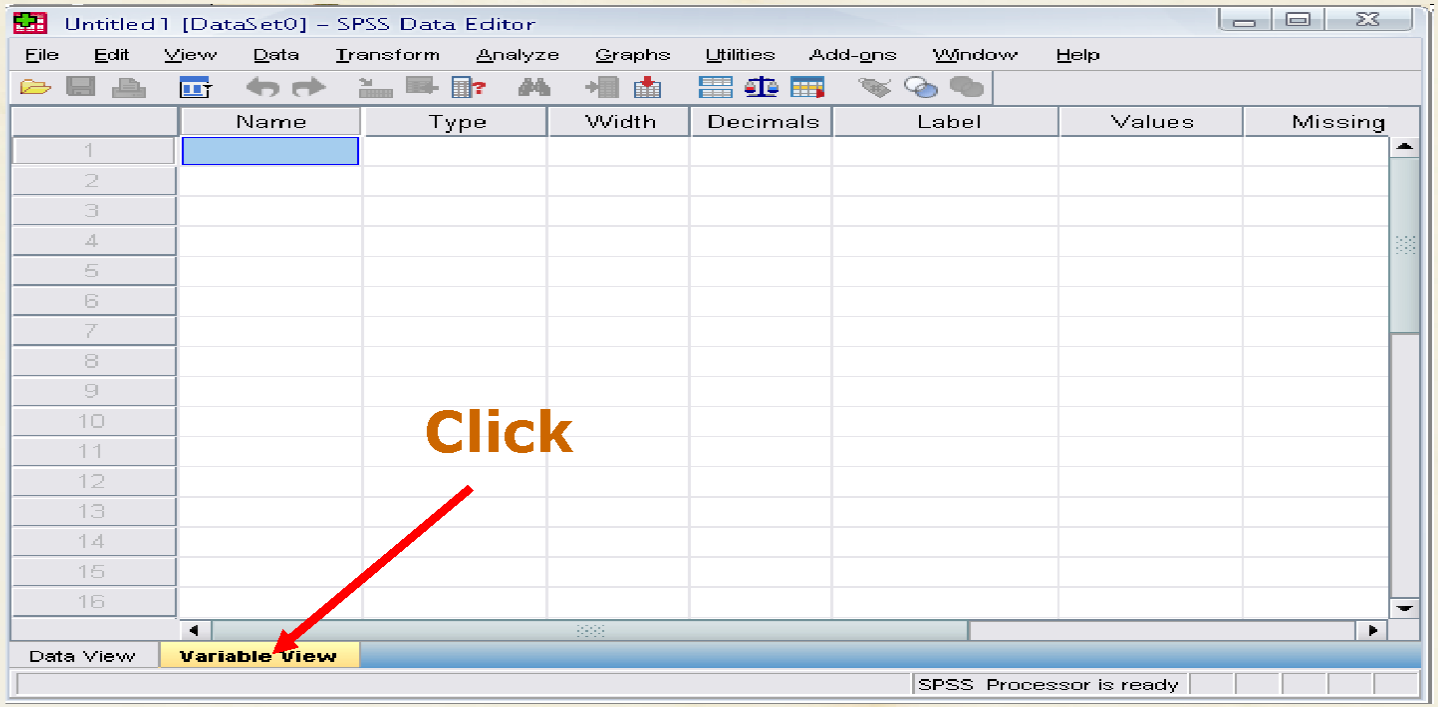


Data View window

- The Data View window

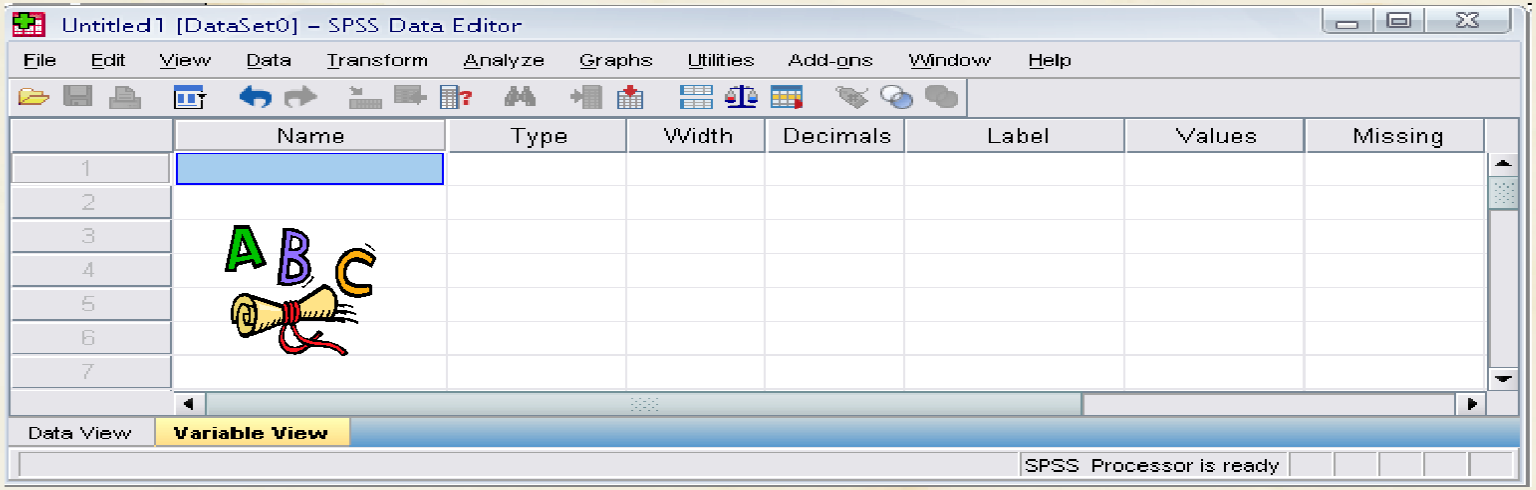
  This sheet is visible when you first open the Data Editor and this sheet contains the data

- Click on the tab labeled Variable View



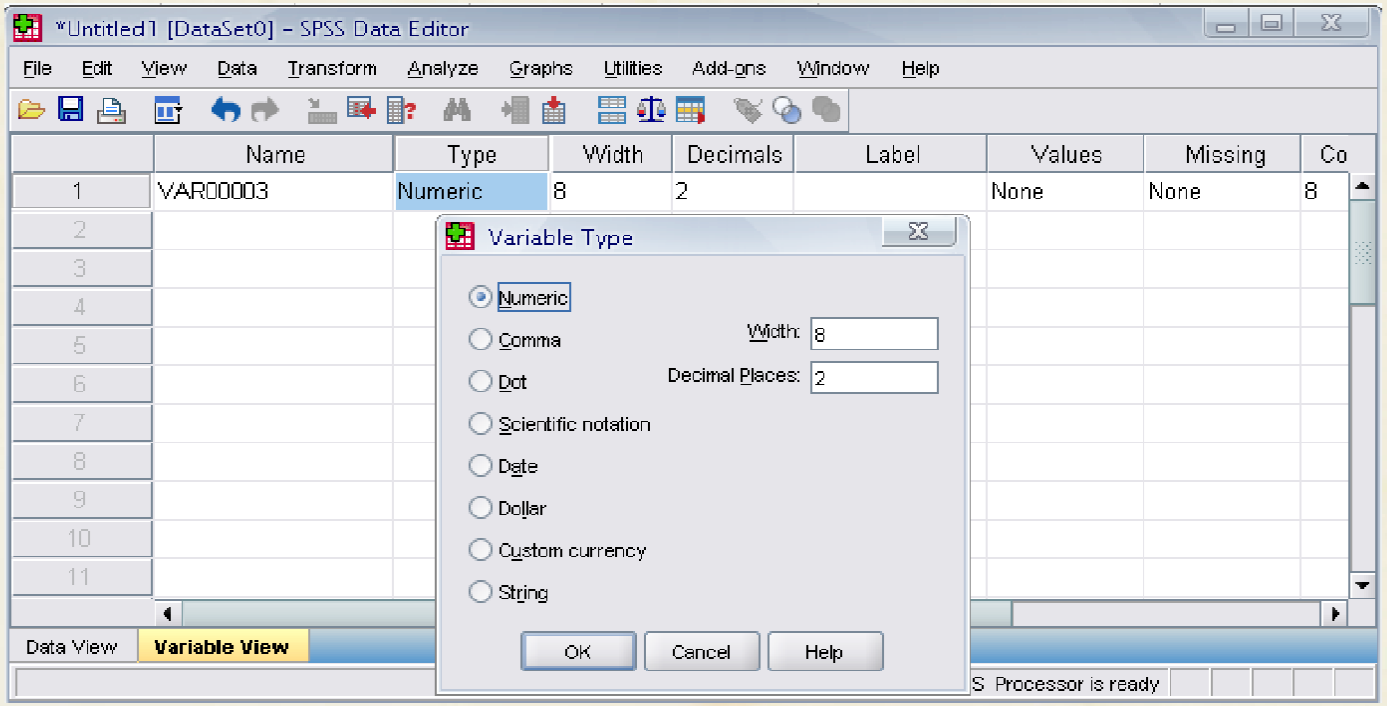
Variable View window

- This sheet contains information about the data set that is stored with the dataset
- Name
  - The first character of the variable name must be alphabetic
  - Variable names must be unique, and have to be less than 64 characters.
  - Spaces are NOT allowed.



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.

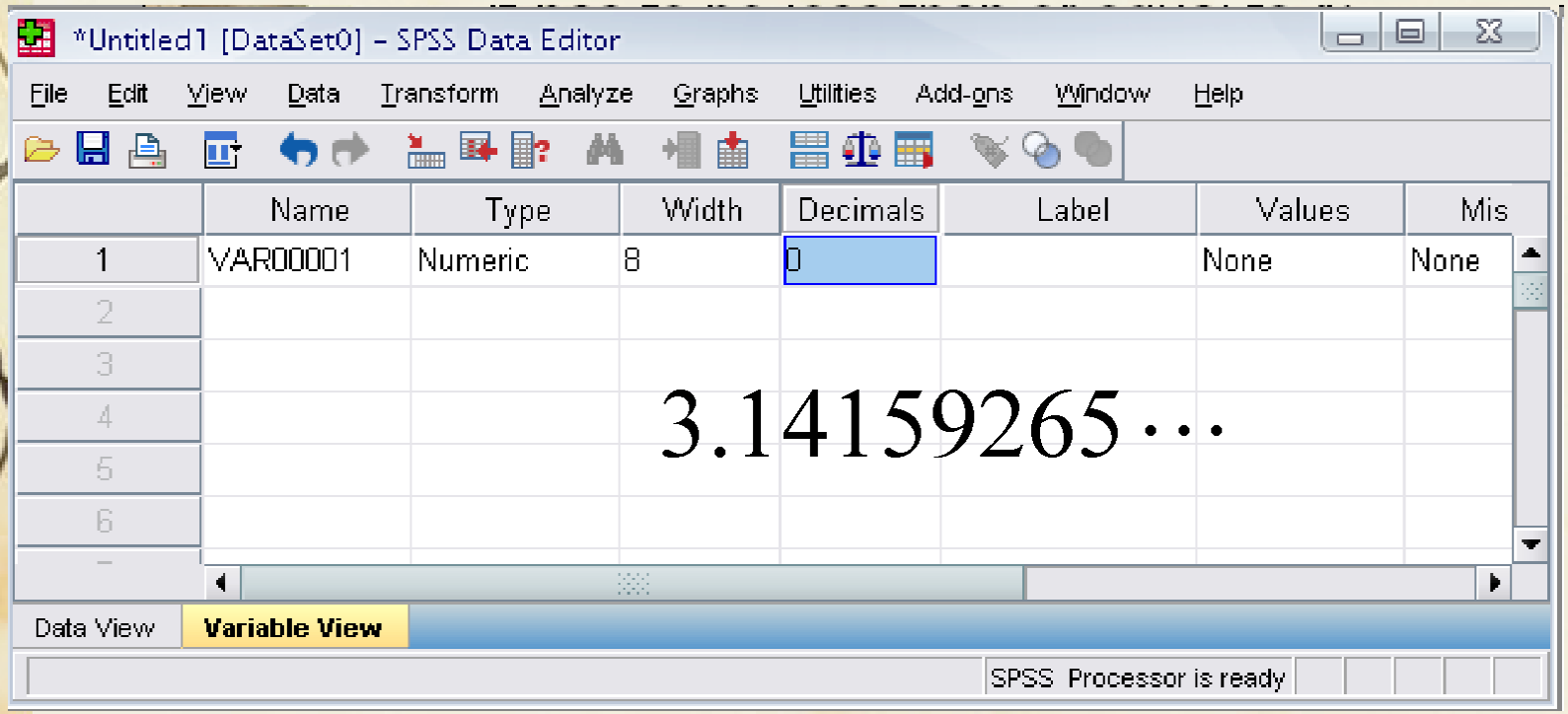


Variable View window: Width

- Width
  - Width allows you to determine the number of characters SPSS will allow to be entered for the variable

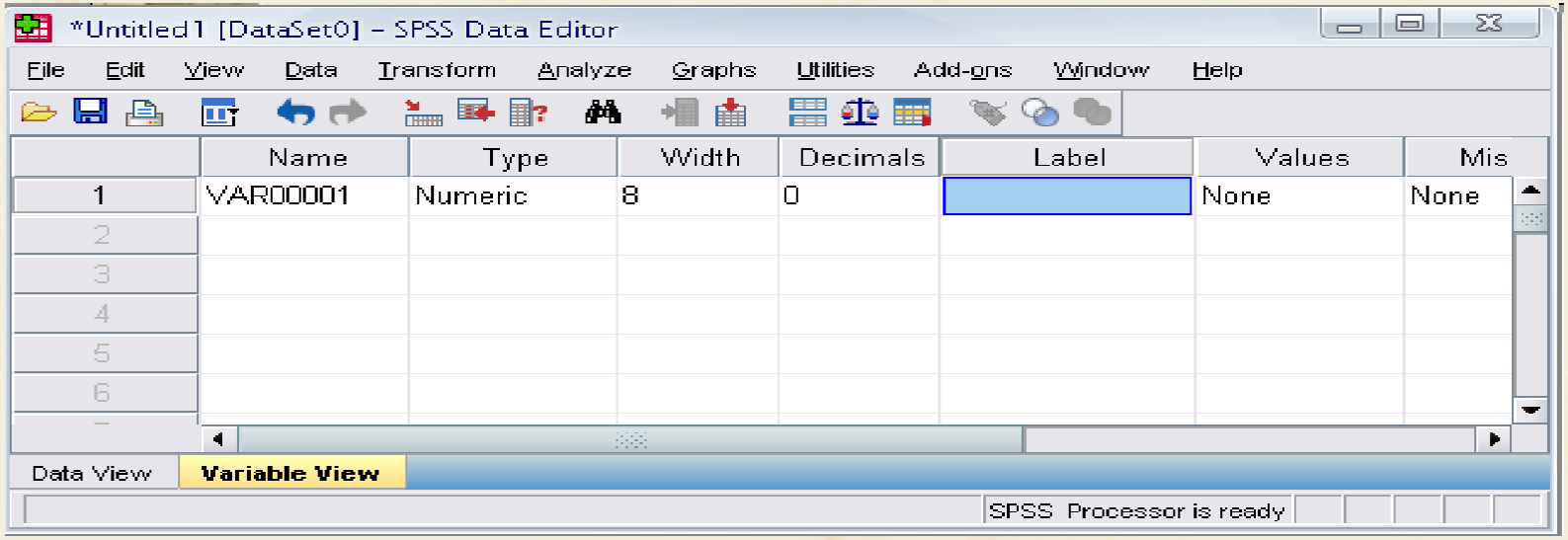
Variable View window: Decimals

- Decimals
  - Number of decimals
  - It has to be less than or equal to 16



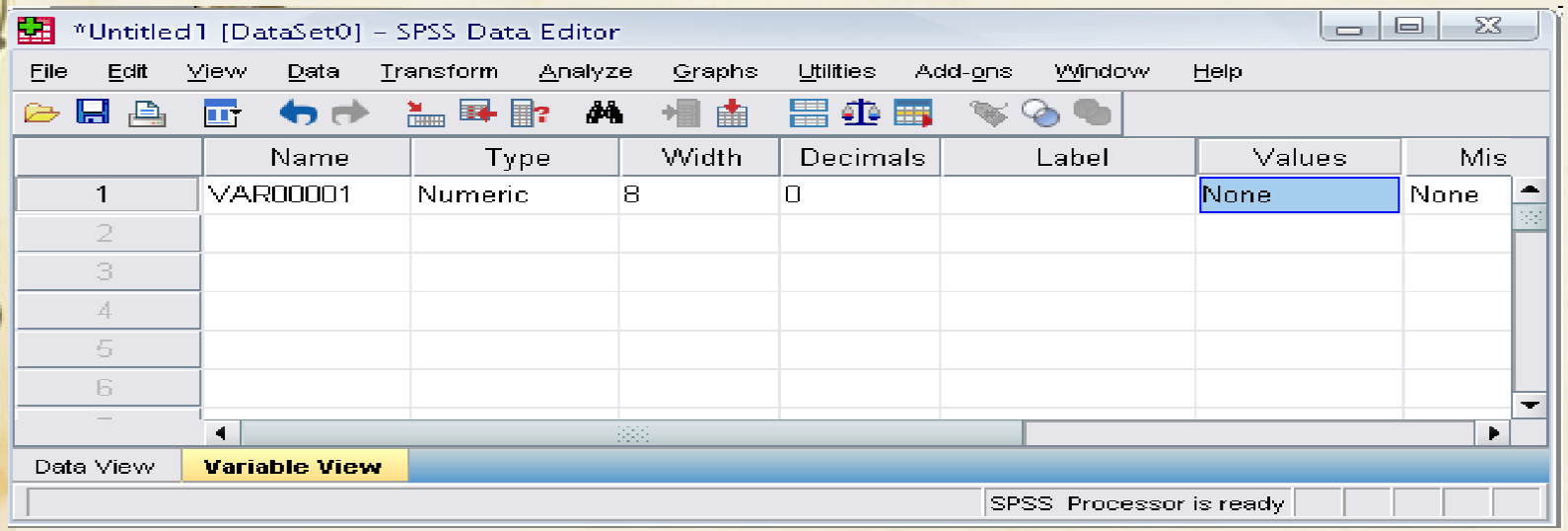
Variable View window: Label

- Label
  - You can specify the details of the variable
  - You can write characters with spaces up to 256 characters



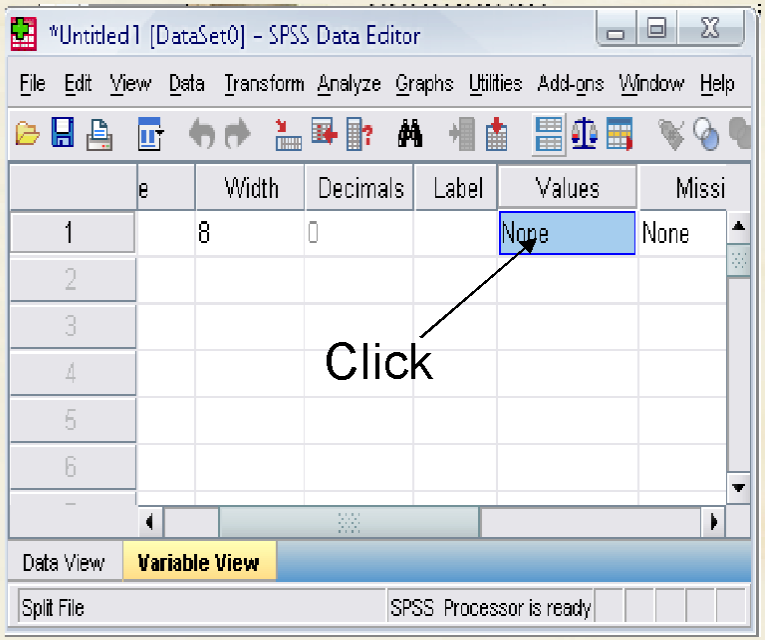
Variable View window: Values

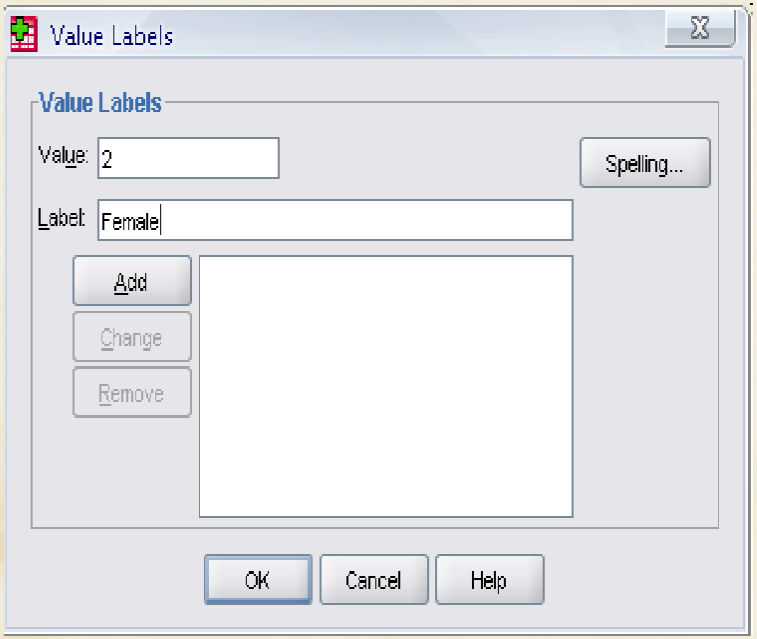
- Values
  - This is used and to suggest which numbers represent which categories when the variable represents a category



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.





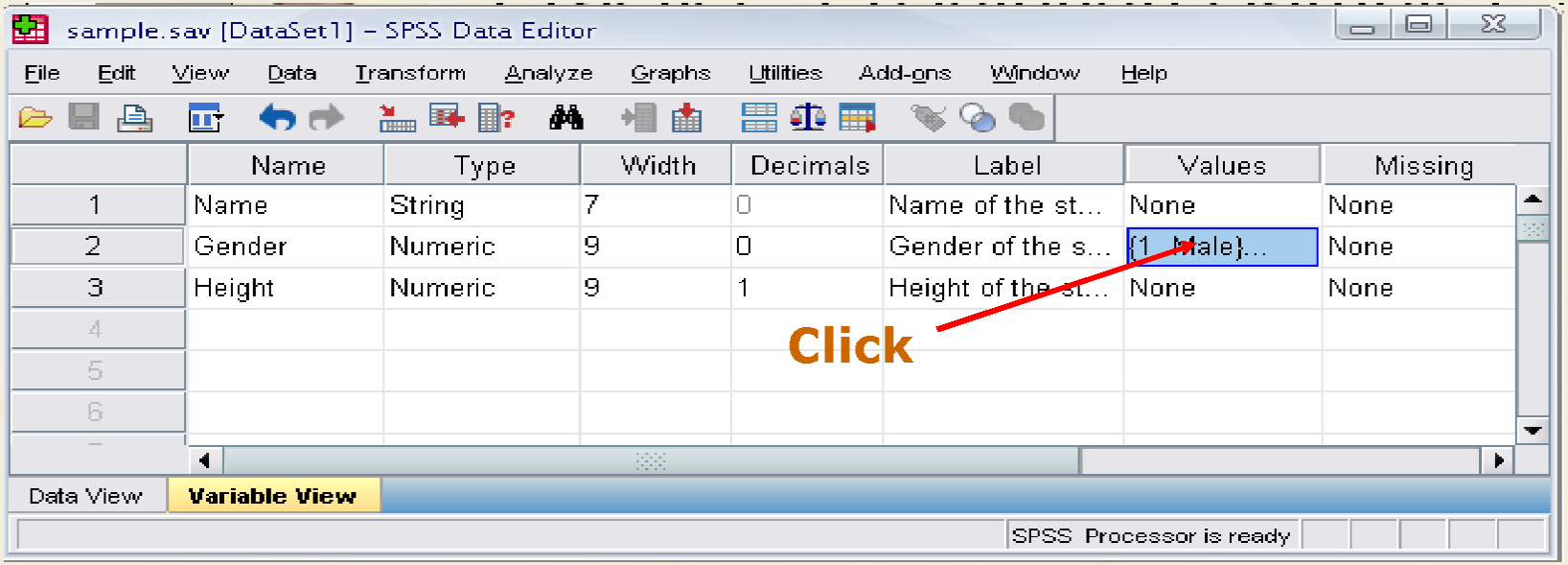
Practice 1

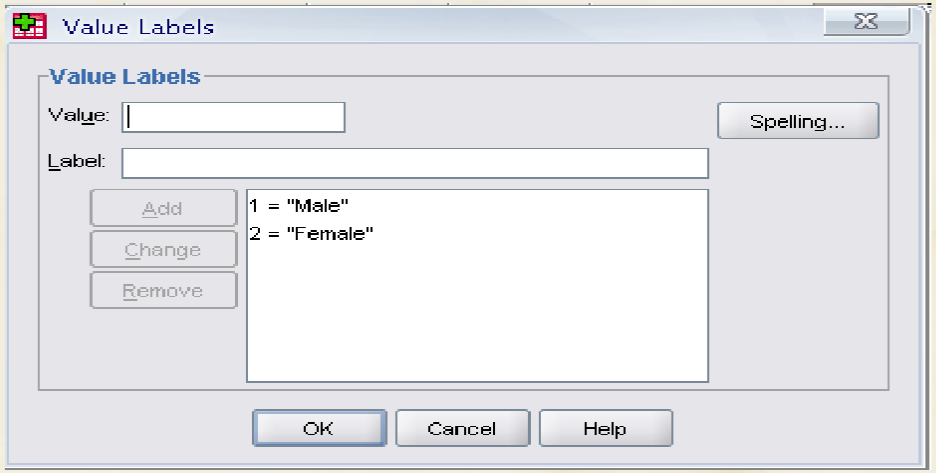
- How would you put the following information into SPSS?

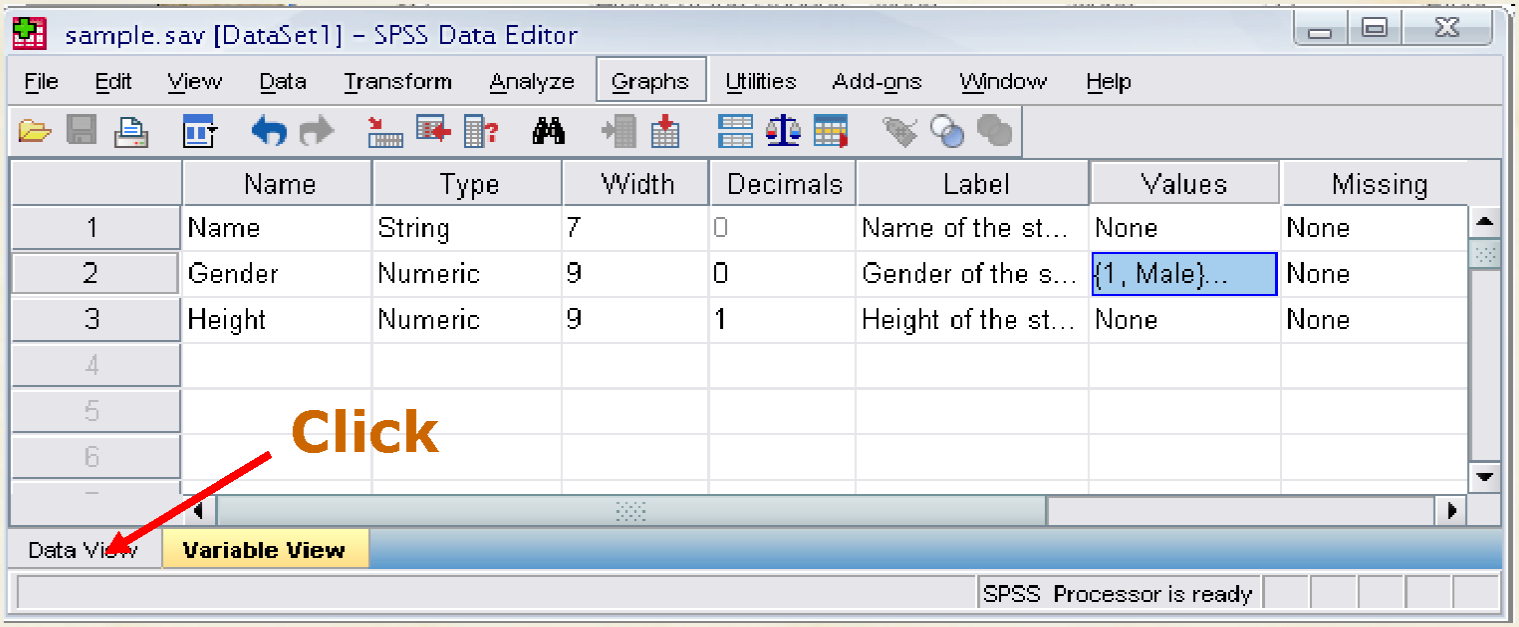
Name	Gender	Height
JAUNITA	2	5.4
SALLY	2	5.3
DONNA	2	5.6
SABRINA	2	5.7
JOHN	1	5.7
MARK	1	6
ERIC	1	6.4
BRUCE	1	5.9

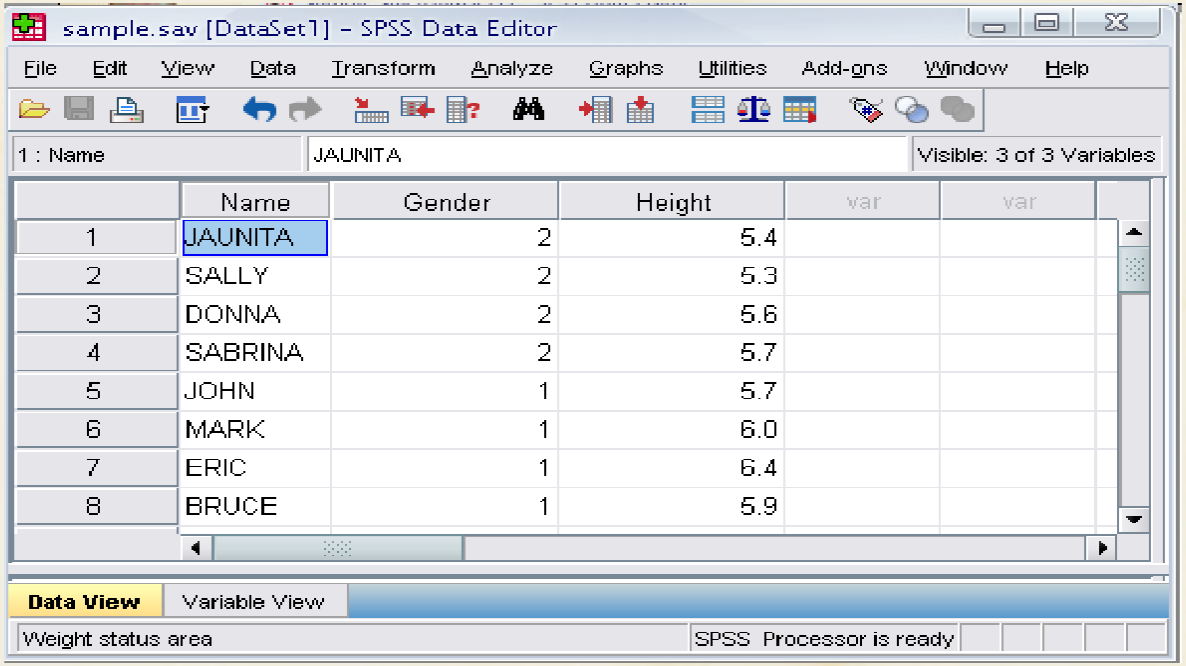
Value = 1 represents Male and Value = 2 represents Female

Practice 1 (Solution Sample)



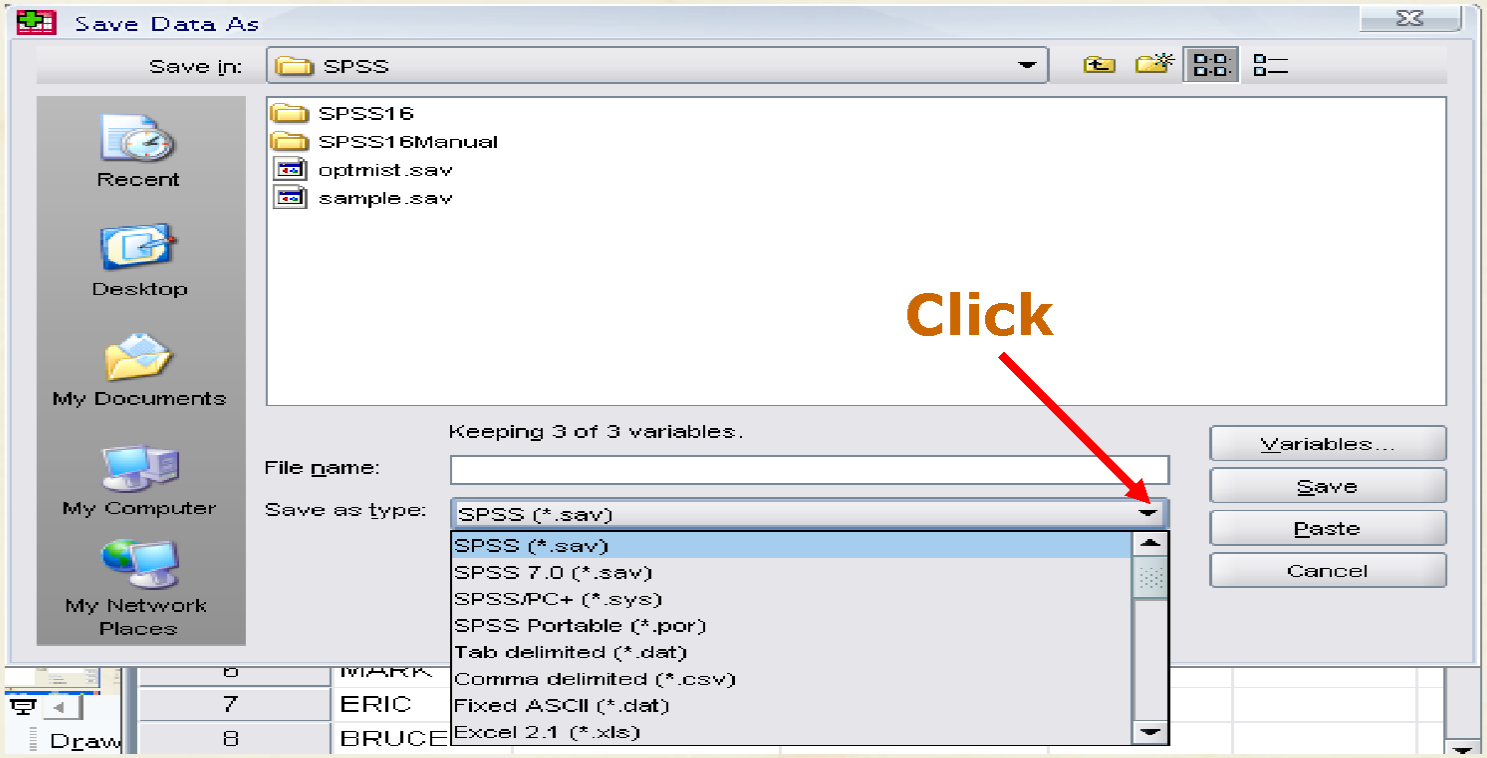



sample.sav [DataSet1] – SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Name Type Width Decimals Label Values Missing
1 Name String 7 0 Name of the st... None None
2 Gender Numeric 9 0 Gender of the s... {1, Male}... None
3 Height Numeric 9 1 Height of the st... None None
4
5
6
Click
Data View Variable View
SPSS Processor is ready


sample.sav [DataSet1] – SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
1 : Name
JAUNITA
Visible: 3 of 3 Variables
Name Gender Height var var
1 JAUNITA 2 5.4
2 SALLY 2 5.3
3 DONNA 2 5.6
4 SABRINA 2 5.7
5 JOHN 1 5.7
6 MARK 1 6.0
7 ERIC 1 6.4
8 BRUCE 1 5.9
Data View Variable View
Weight status area
SPSS Processor is ready

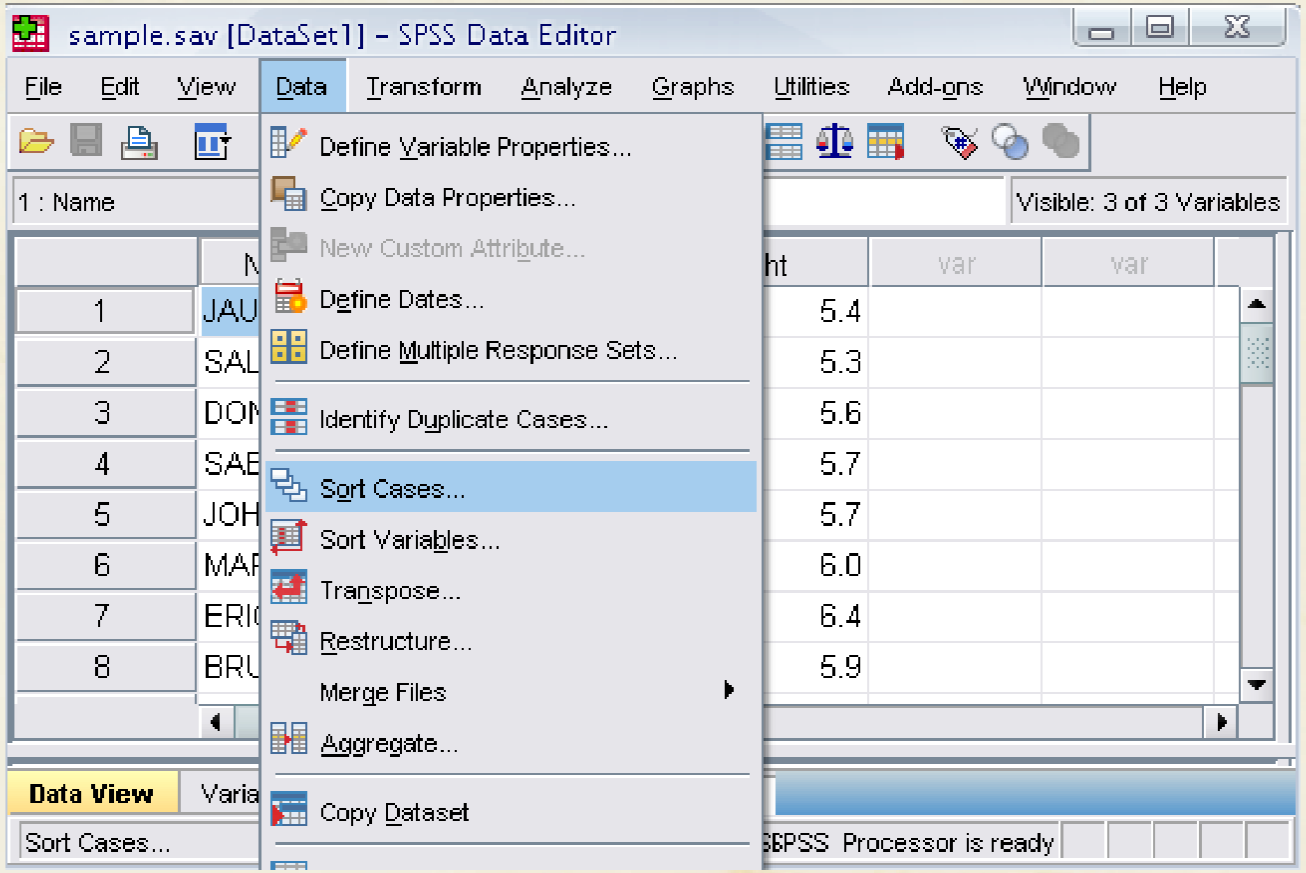
Saving the data

- 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.”



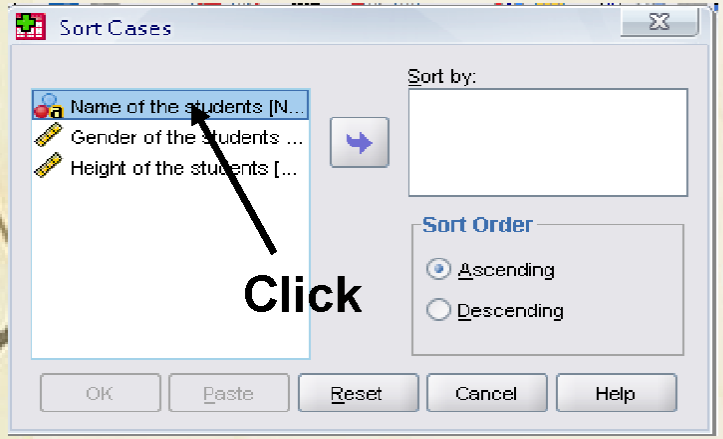
Sorting the data

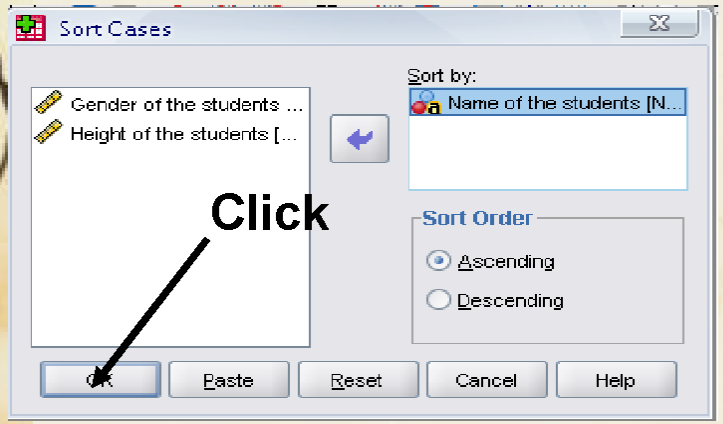
- Click ‘Data’ and then click Sort Cases

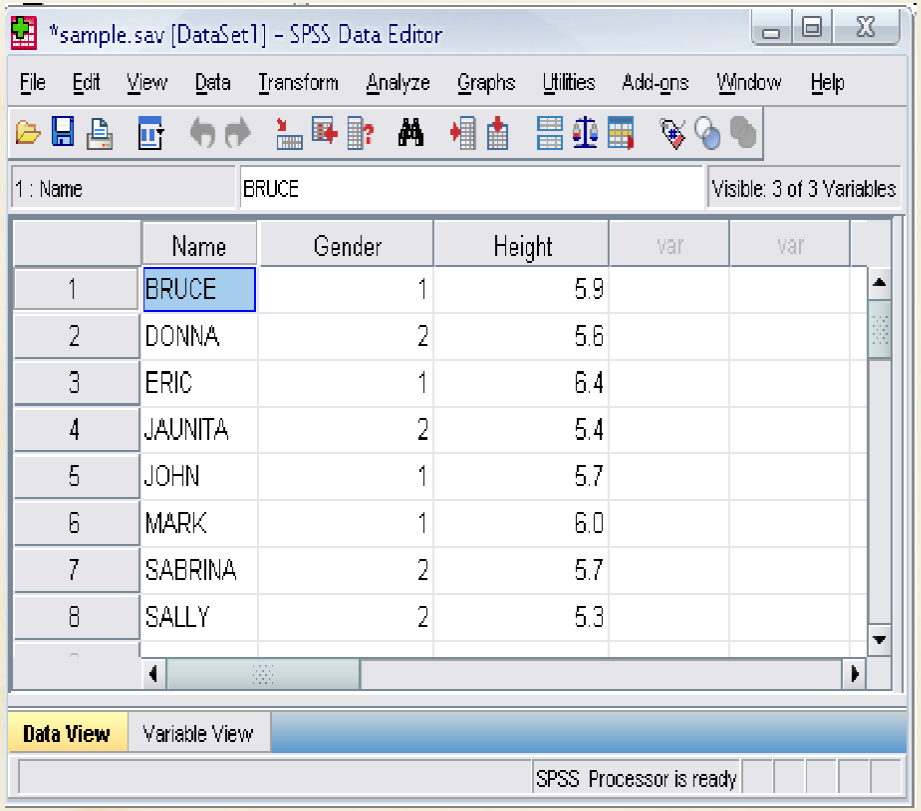


Sorting the data (cont'd)

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

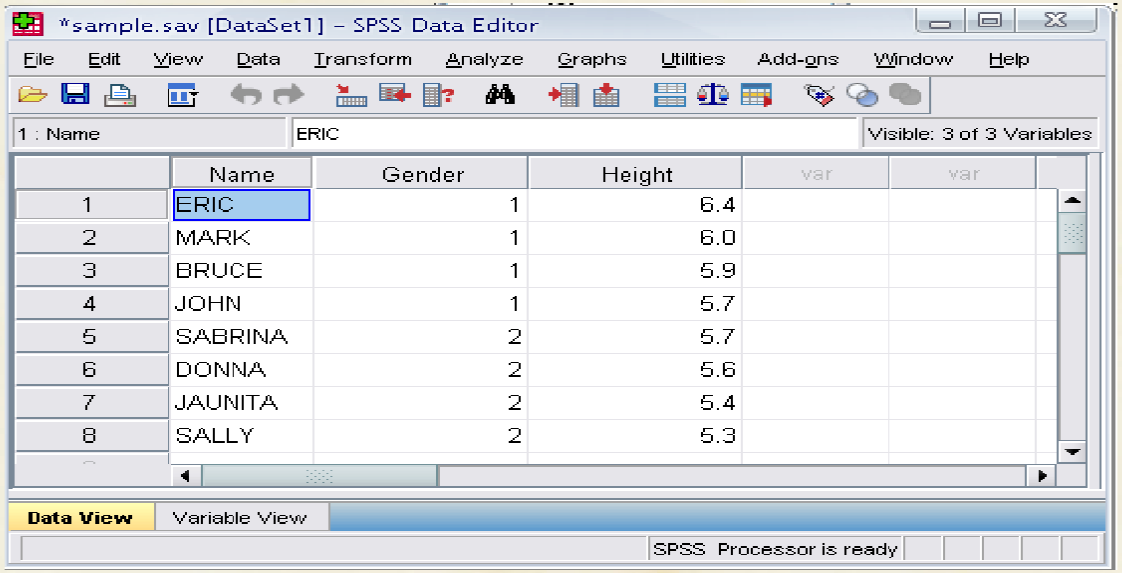






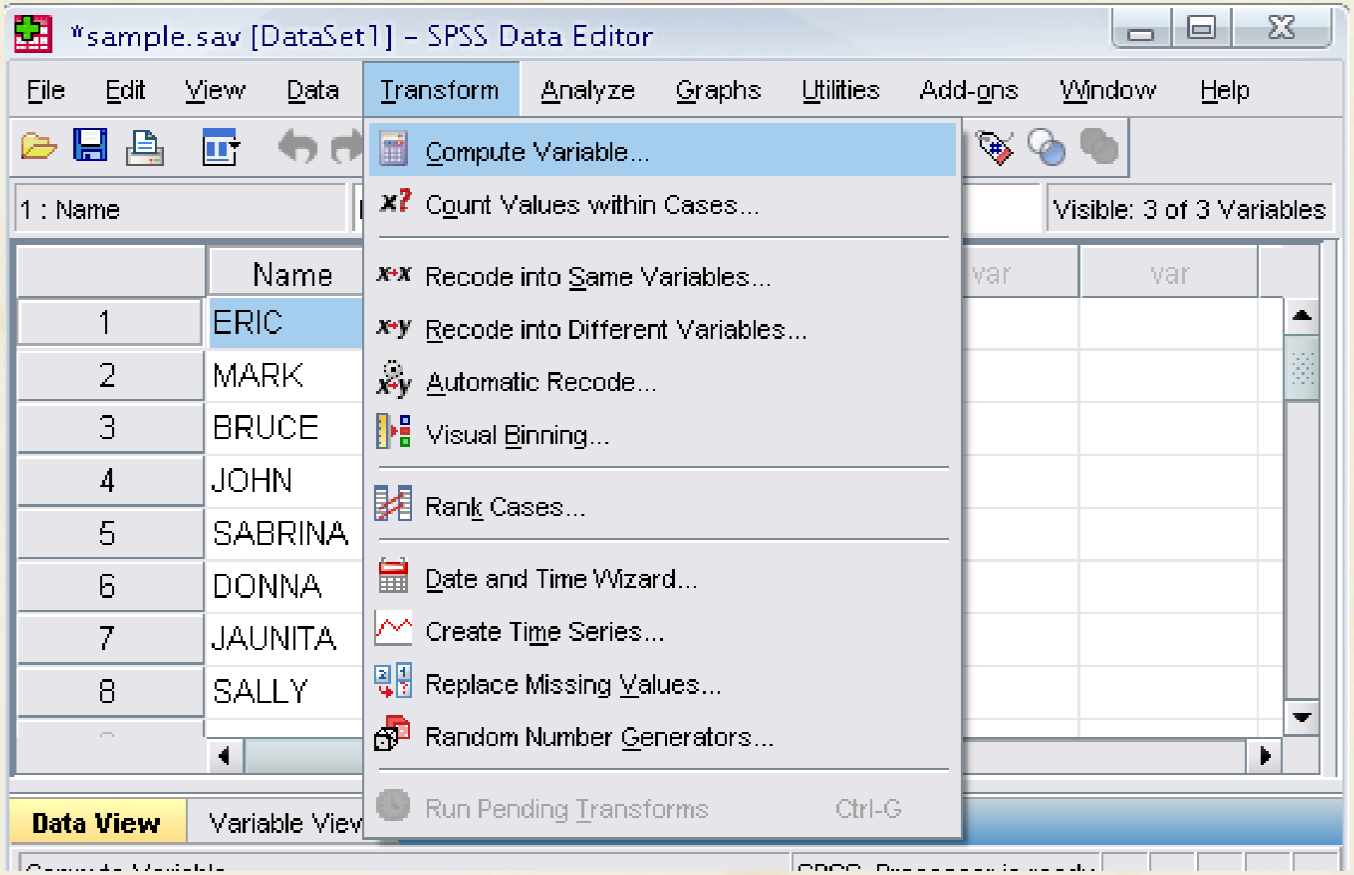
Practice 2

- How would you sort the data by the ‘Height’ of students in descending order?
- Answer
  - Click data, sort cases, double click ‘height of students,’ click ‘descending,’ and finally click ok.



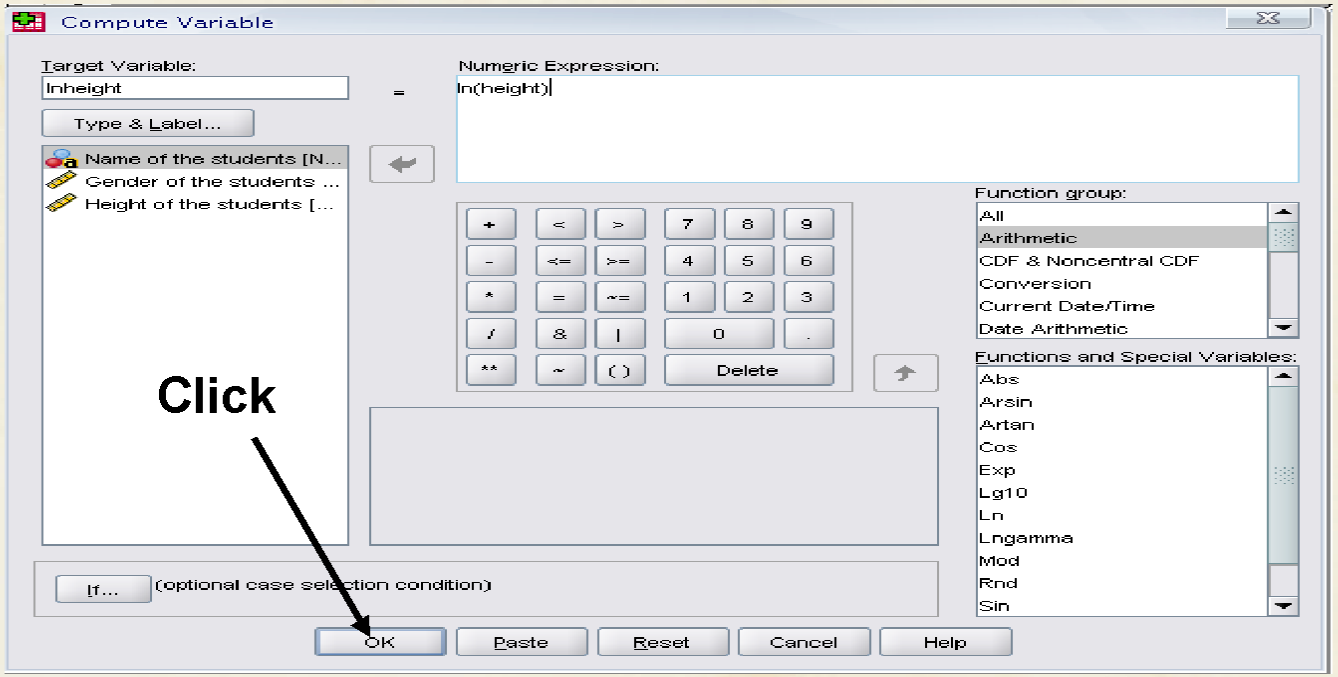
Transforming data

- Click ‘Transform’ and then click ‘Compute Variable…’



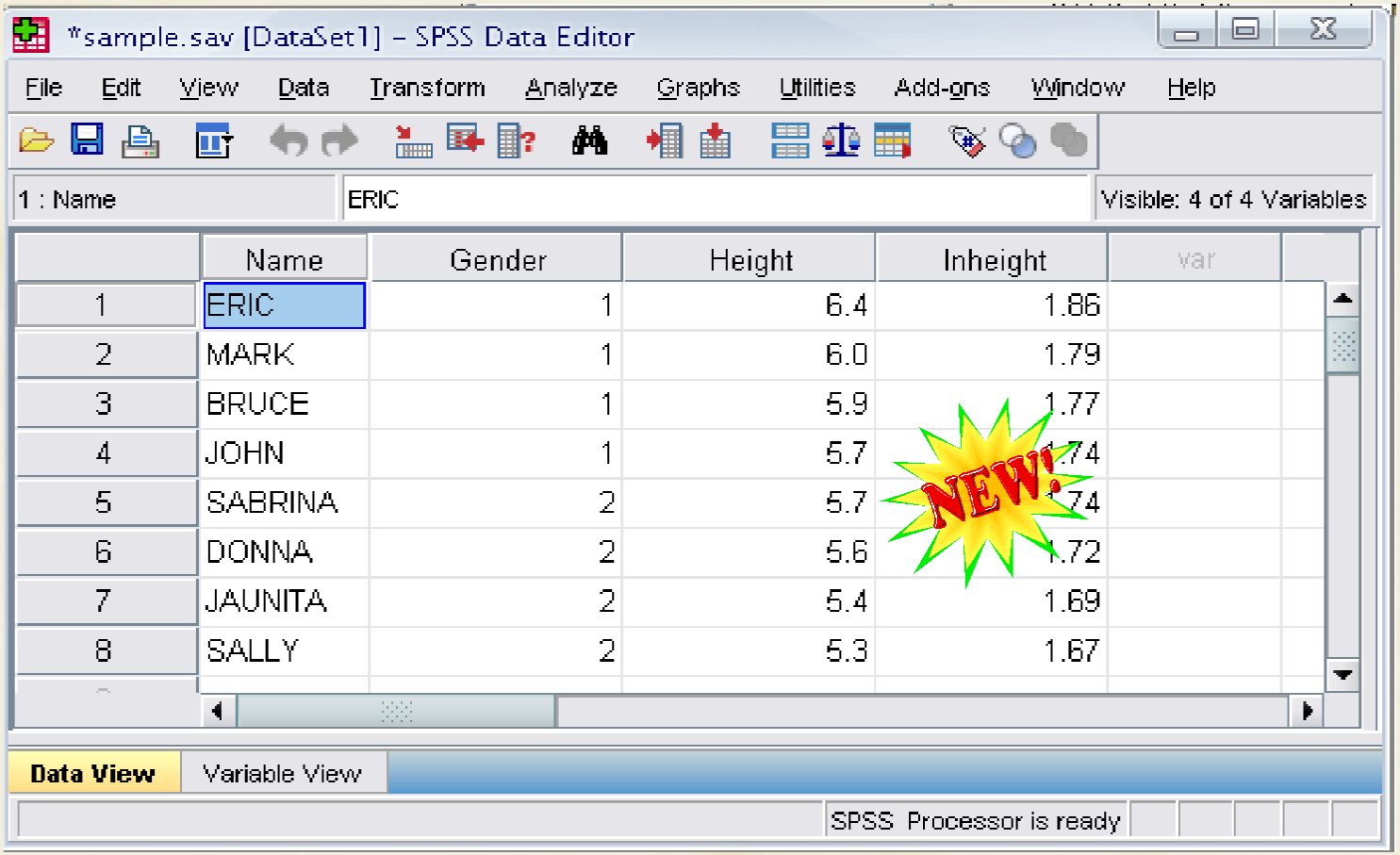
Transforming data (cont'd)

- Example: Adding a new variable named 'lnheight' which is the natural log of height
  - Type in lnheight in the 'Target Variable' box. Then type in 'ln(height)' in the 'Numeric Expression' box. Click OK



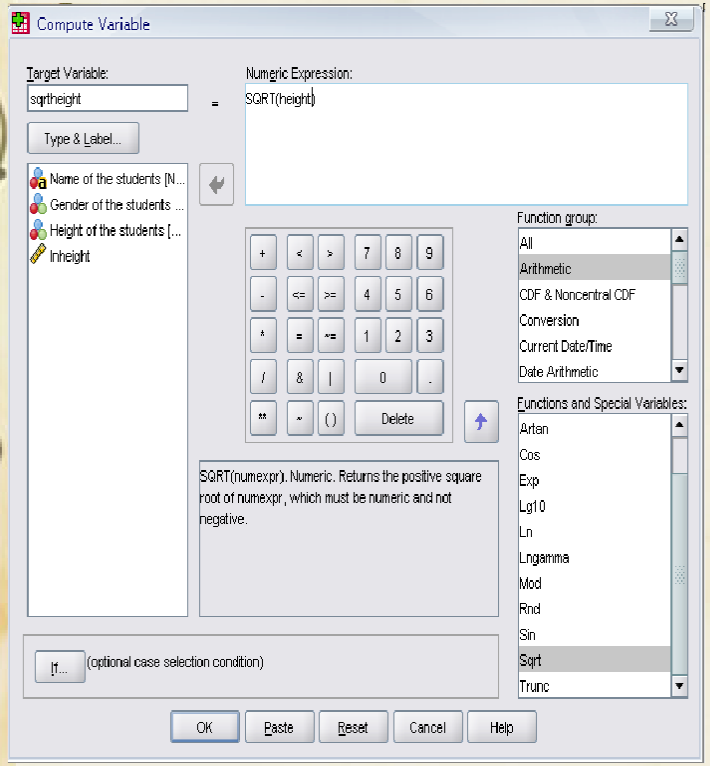
Transforming data (cont'd)

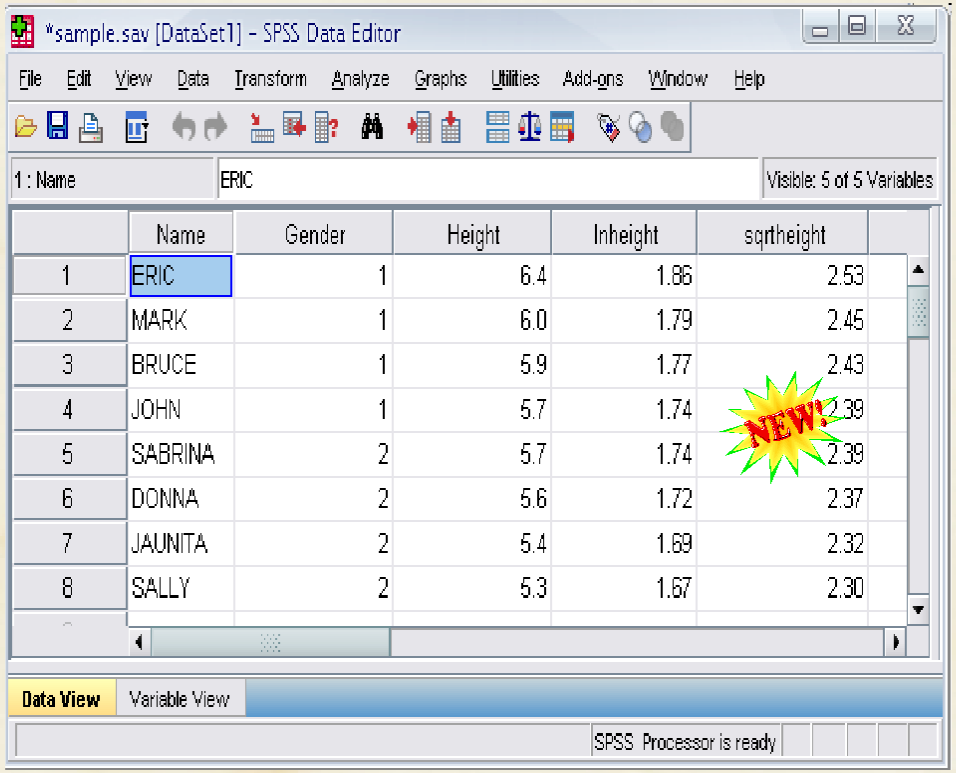
- A new variable 'Inheight' is added to the table



Practice 3

- Create a new variable named “sqrtheight” which is the square root of height.
- Answer





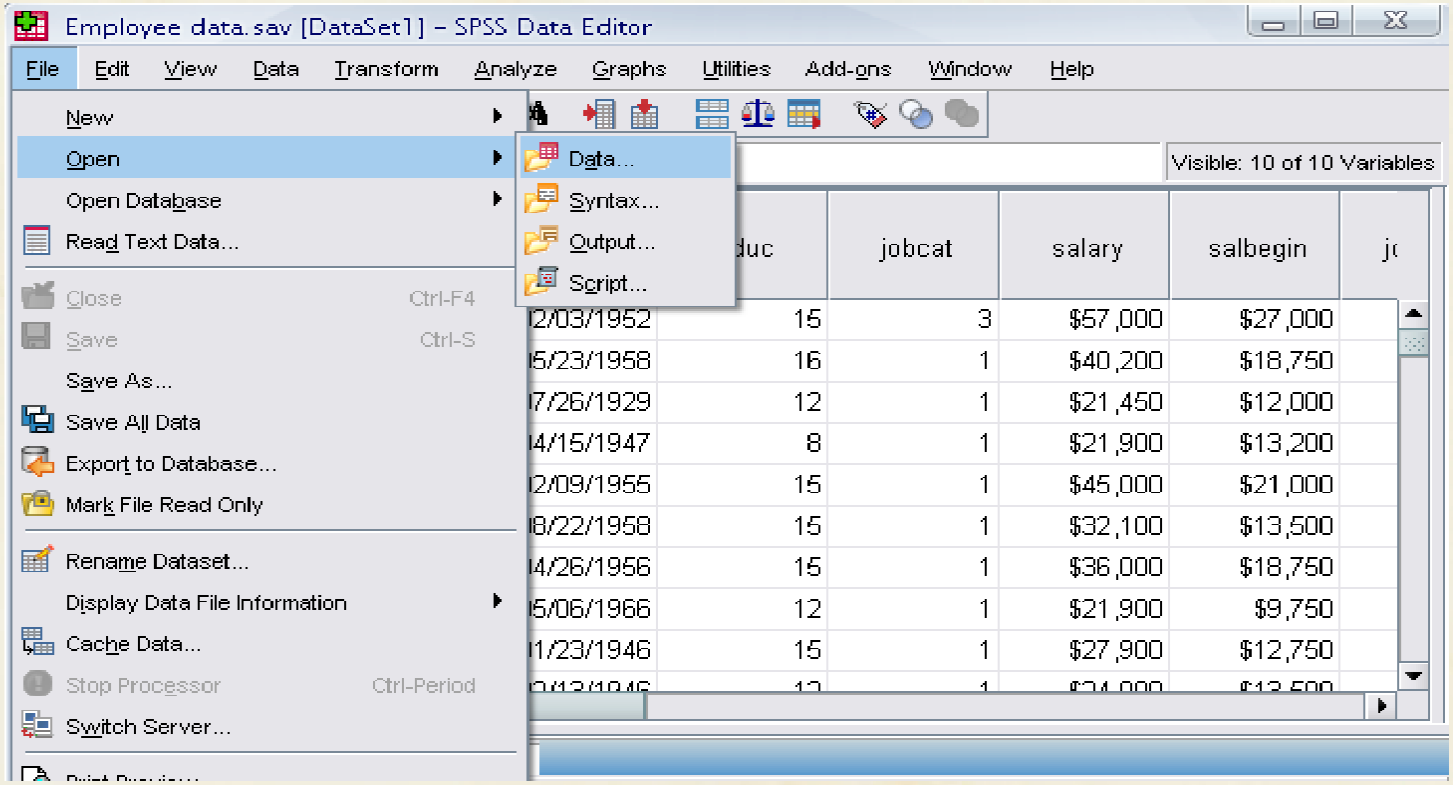
The basic analysis

The basic analysis of SPSS that will be introduced in this class

- Frequencies
  - This analysis produces frequency tables showing frequency counts and percentages of the values of individual variables.
- Descriptives
  - This analysis shows the maximum, minimum, mean, and standard deviation of the variables
- Linear regression analysis
  - Linear Regression estimates the coefficients of the linear equation

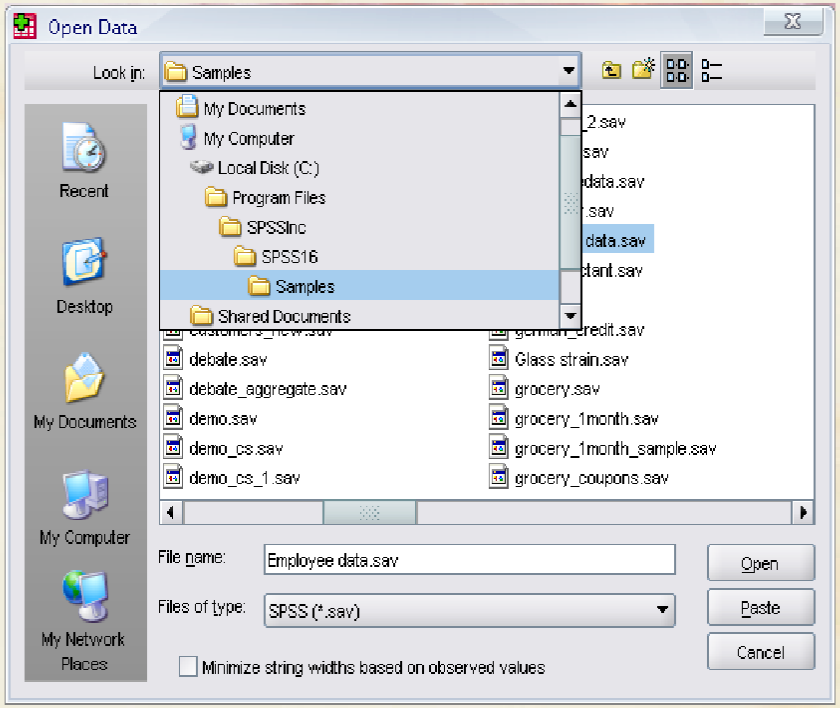
Opening the sample data

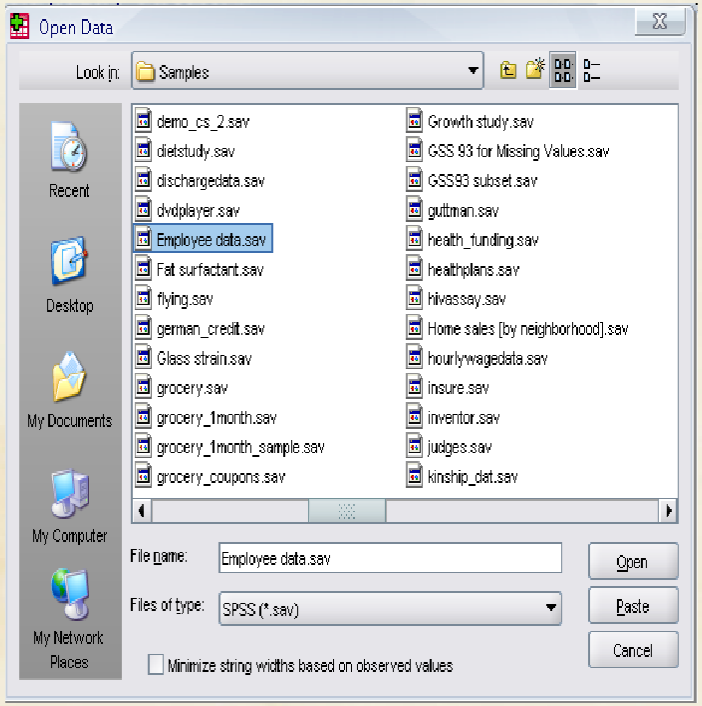
- Open ‘Employee data.sav’ from the SPSS
  - Go to “File,” “Open,” and Click Data



Opening the sample data

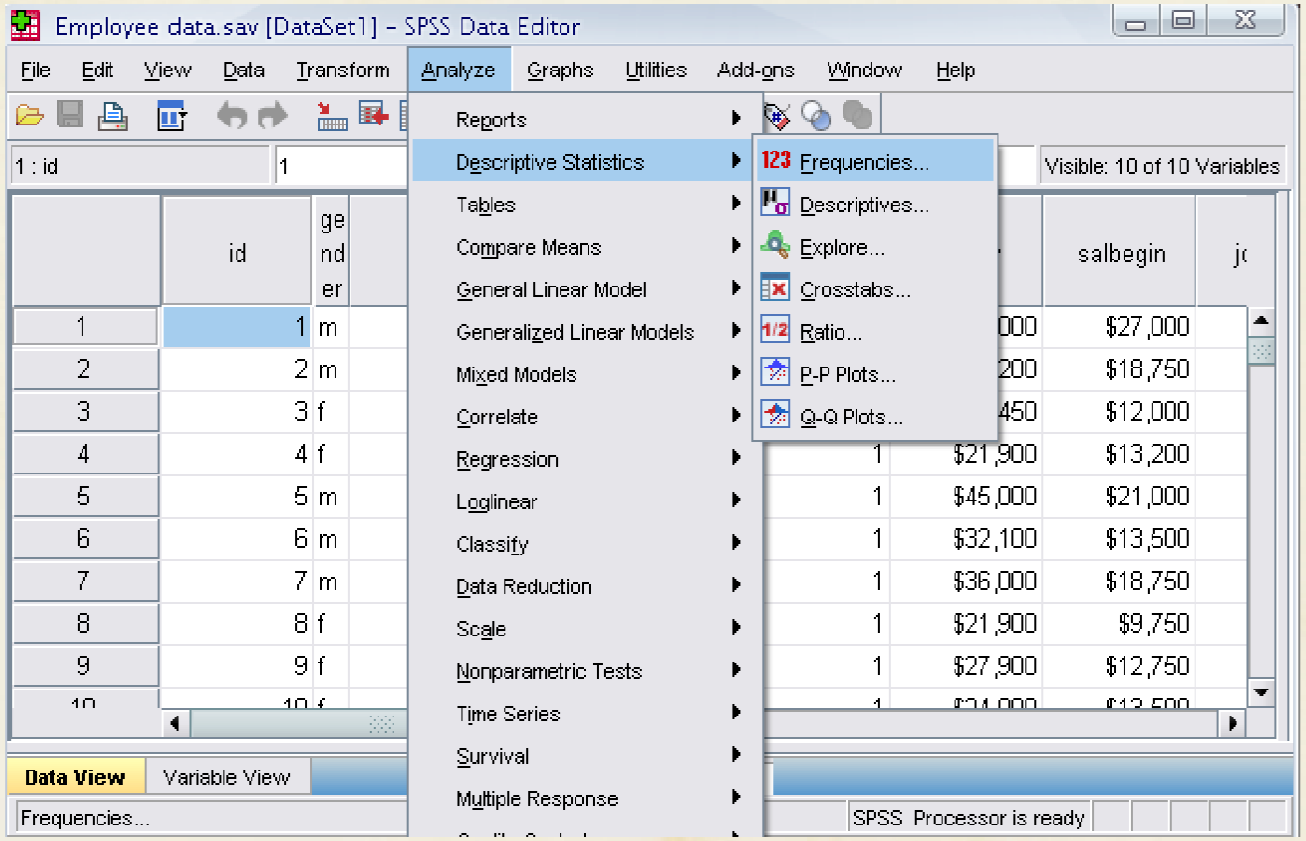
- Go to Program Files," "SPSSInc," "SPSS16," and "Samples" folder.
- Open "Employee Data.sav" file





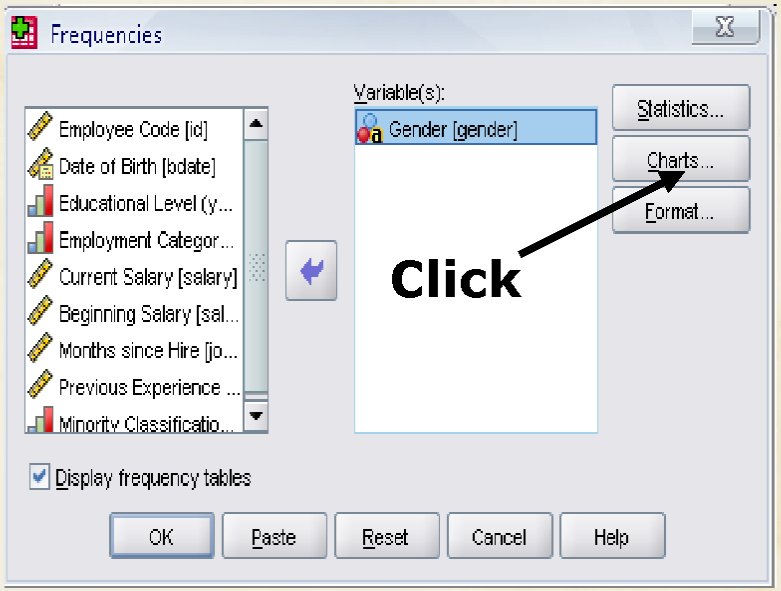
Frequencies

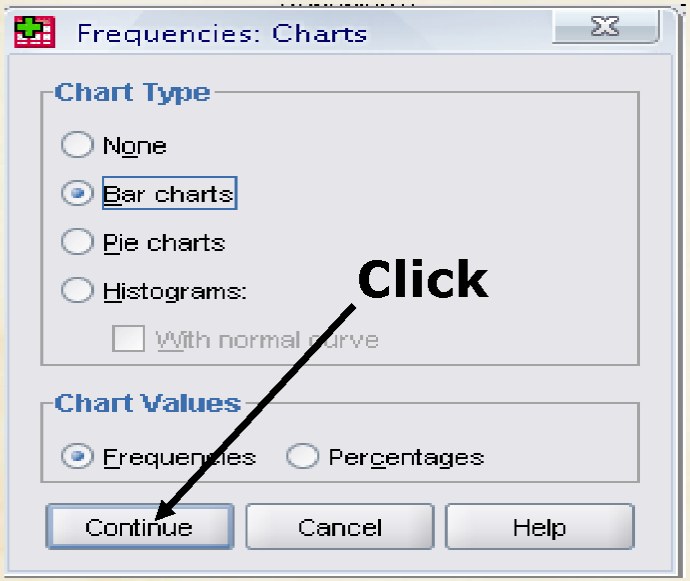
- Click ‘Analyze,’ ‘Descriptive statistics,’ then click ‘Frequencies’



Frequencies

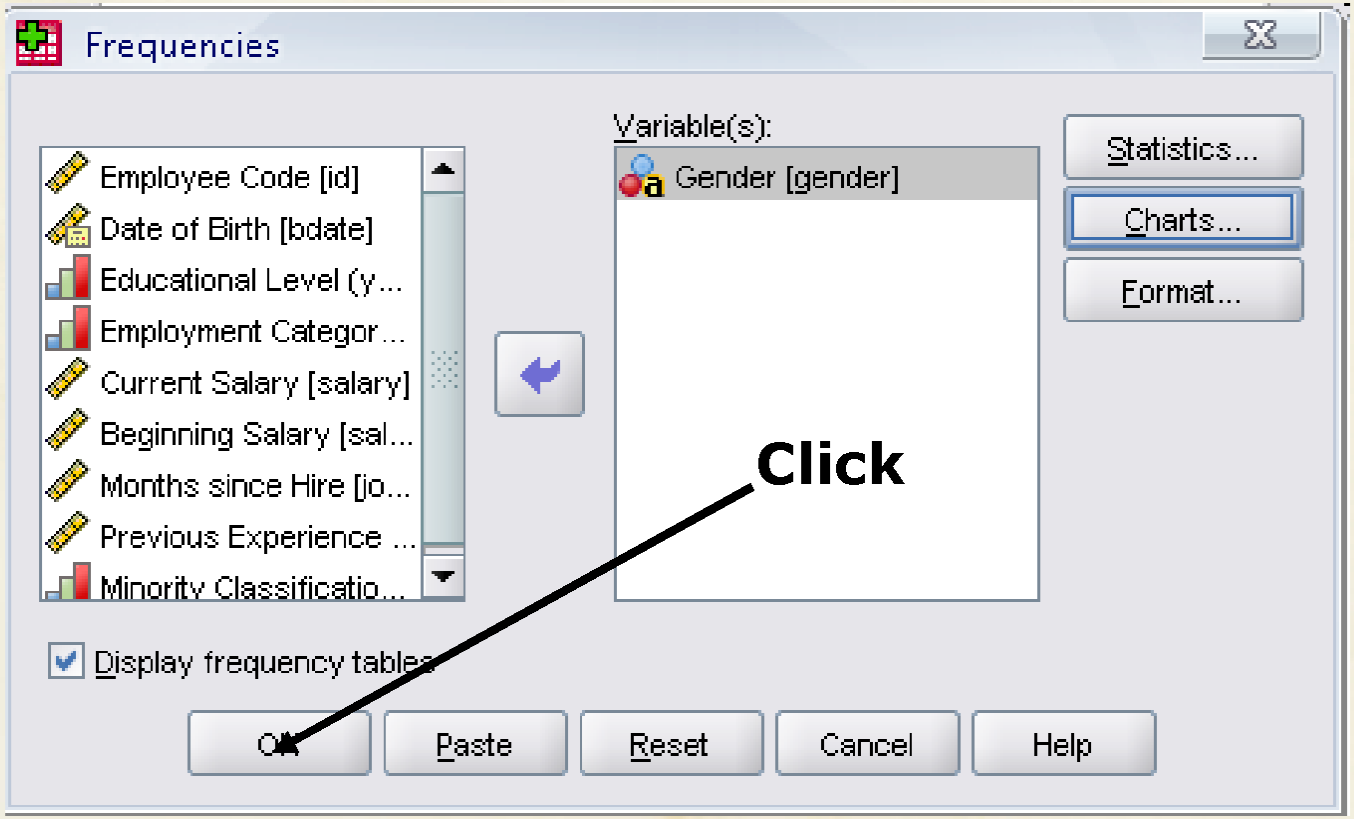
- Click gender and put it into the variable box.
- Click ‘Charts.’
- Then click ‘Bar charts’ and click ‘Continue.’





Frequencies

- Finally Click OK in the Frequencies box.



] - SPSS Viewer

Transform Insert Format Analyze Graphs Utilities Add-ons Window Help

Frequencies

[DataSet1] C:\Program Files\SPSSInc\SPSS16\Samples\Employee

Statistics

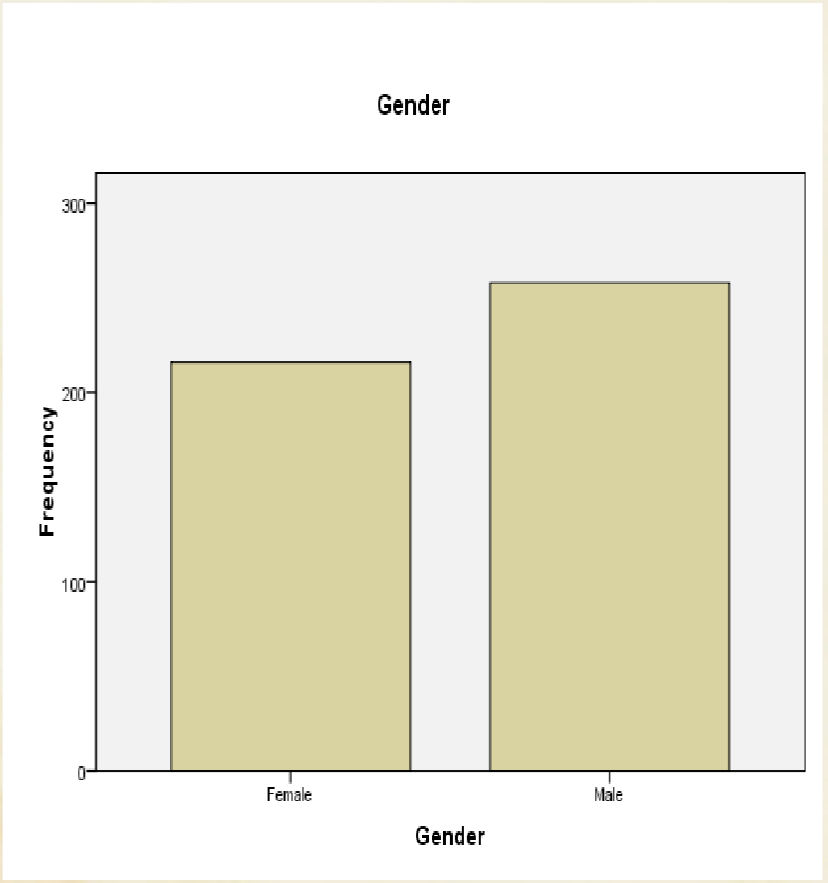
Gender

N	Valid	474
	Missing	0

Gender

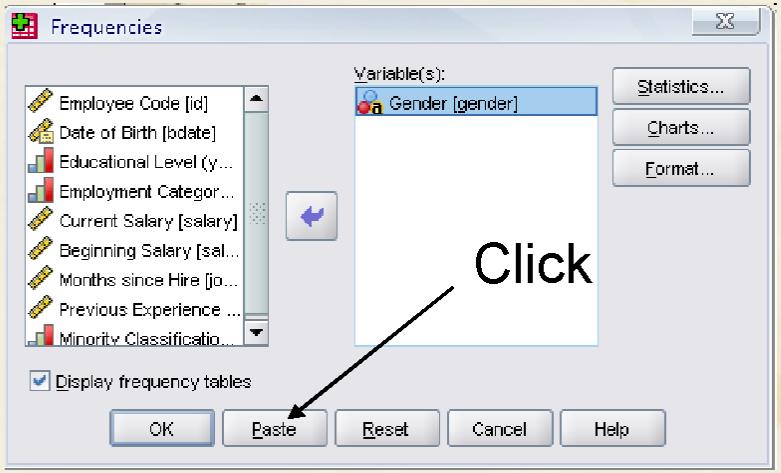
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	216	45.6	45.6	45.6
	Male	258	54.4	54.4	100.0
	Total	474	100.0	100.0	

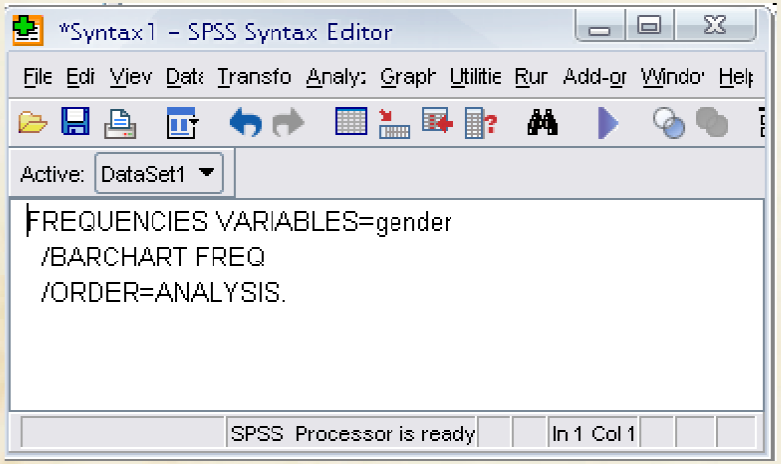
SPSS Processor is ready



Using the Syntax editor

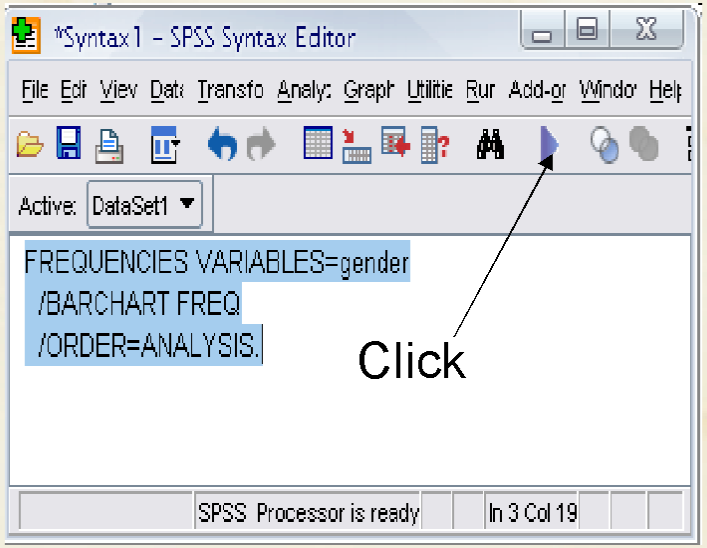
- Click ‘Analyze,’ ‘Descriptive statistics,’ then click ‘Frequencies.’
- Put ‘Gender’ in the Variable(s) box.
- Then click ‘Charts,’ ‘Bar charts,’ and click ‘Continue.’
- Click ‘Paste.’

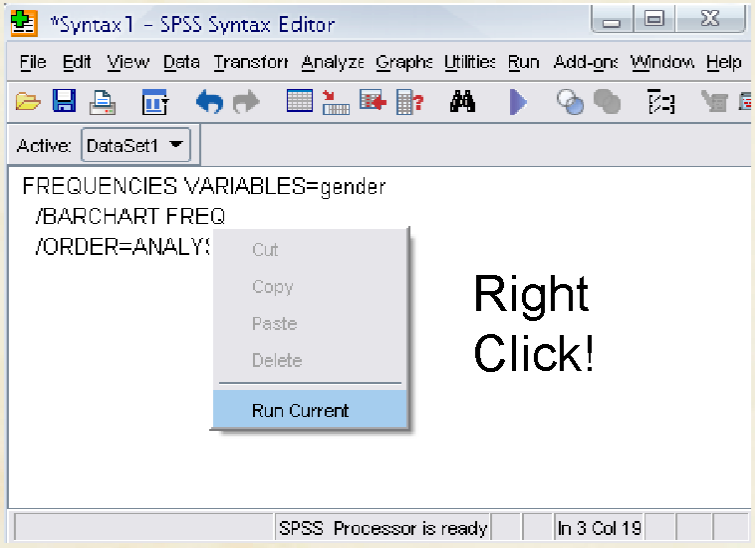




Using the Syntax editor

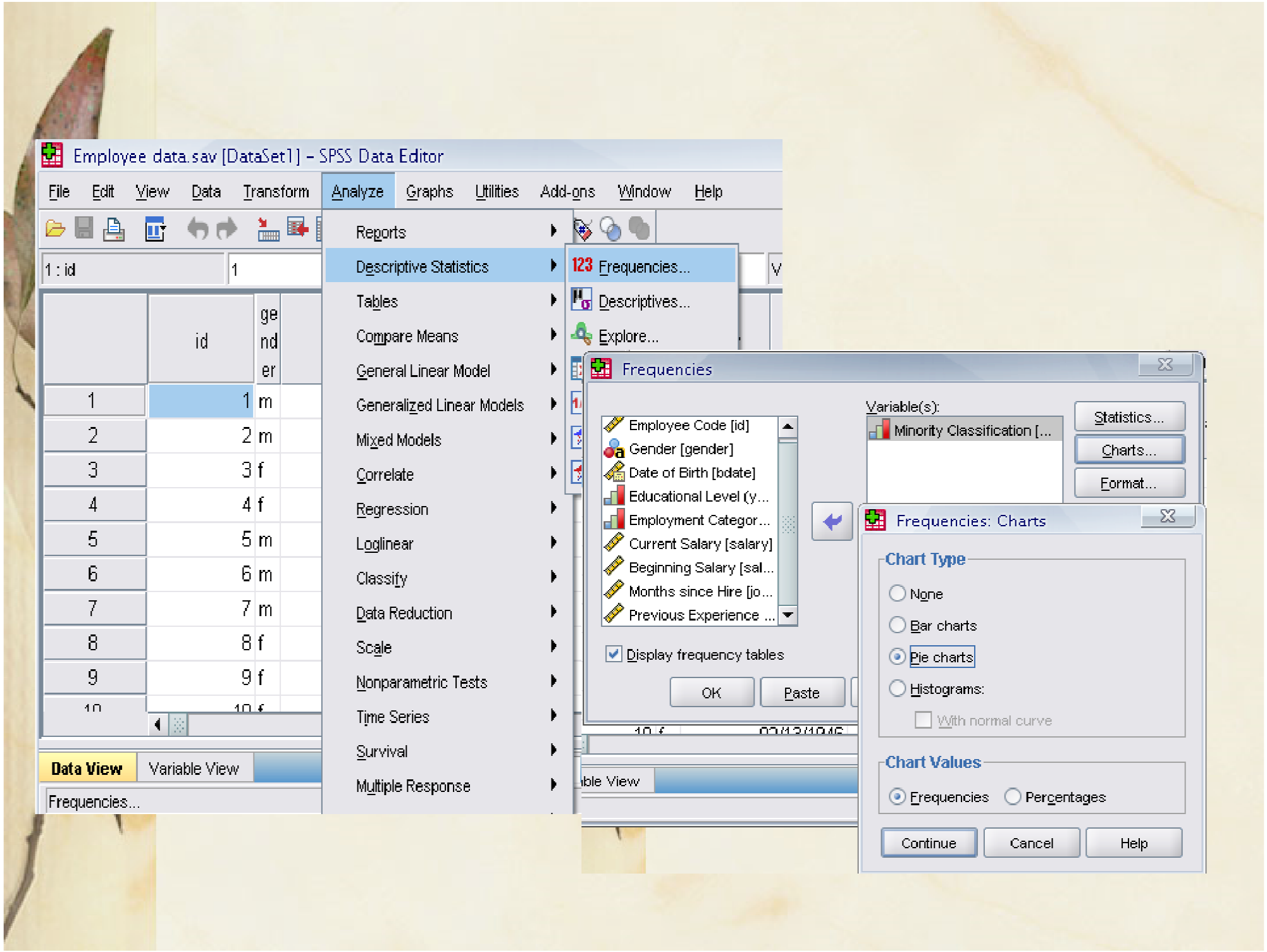
- Highlight the commands in the Syntax editor and then click the run icon.
- You can do the same thing by right clicking the highlighted area and then by clicking ‘Run Current’



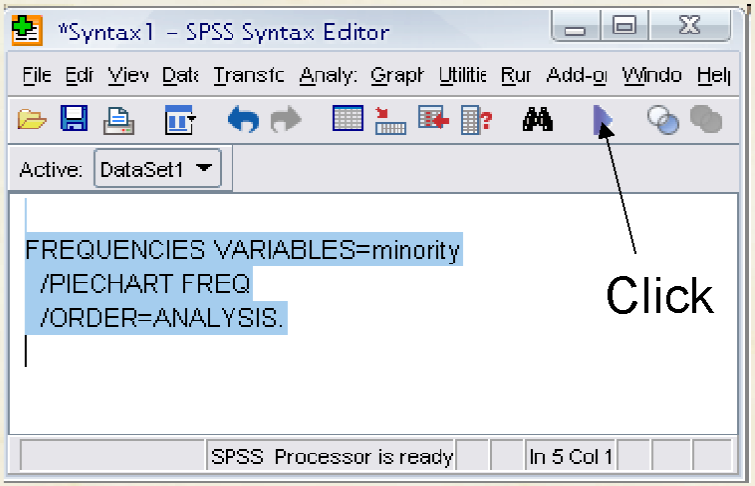


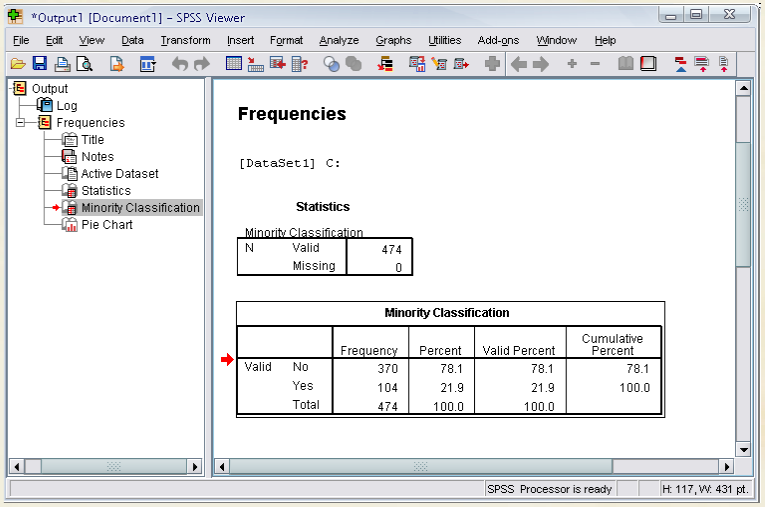
Practice 4

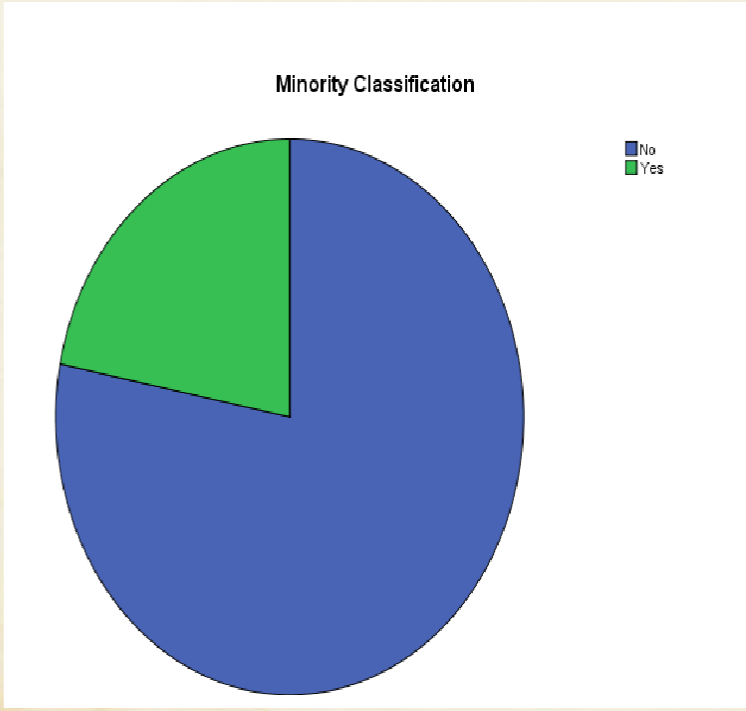
- Do a frequency analysis on the variable “minority”
- Create pie charts for it
- Do the same analysis using the syntax editor


Employee data.sav [DataSet1] – SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
1 : id
1
id
gender
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Frequencies...
Descriptives...
Explore...
Data View
Variable View
Frequencies...
Frequencies
Employee Code [id]
Gender [gender]
Date of Birth [bdate]
Educational Level (y...
Employment Categor...
Current Salary [salary]
Beginning Salary [sal...
Months since Hire [jo...
Previous Experience ...
Variable(s):
Minority Classification [...
Statistics...
Charts...
Format...
Display frequency tables
OK
Paste
Frequencies: Charts
Chart Type
None
Bar charts
Pie charts
Histograms:
With normal curve
Chart Values
Frequencies
Percentages
Continue
Cancel
Help

Answer

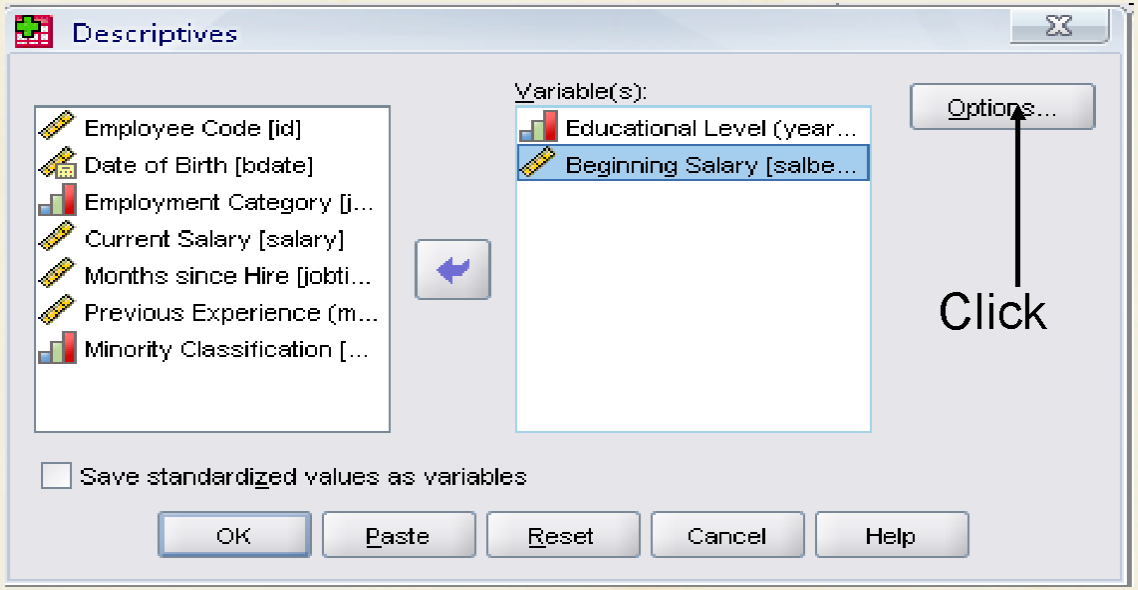

*Syntax1 – SPSS Syntax Editor
File Edit View Data Transform Analyze Graphs Utilities Run Add-ons Window Help
Active: DataSet1
FREQUENCIES VARIABLES=minority
/PIECHART FREQ
/ORDER=ANALYSIS.
Click
SPSS Processor is ready
In 5 Col 1


*Output1 [Document1] – SPSS Viewer
File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help
Output
Log
Frequencies
Title
Notes
Active Dataset
Statistics
Minority Classification
Pie Chart
Frequencies
[DataSet1] C:
Statistics
Minority Classification
N Valid 474
Missing 0
Minority Classification
Frequency Percent Valid Percent Cumulative Percent
Valid No 370 78.1 78.1 78.1
Yes 104 21.9 21.9 100.0
Total 474 100.0 100.0
SPSS Processor is ready
H: 117, W: 431 pt.


Minority Classification
No
Yes

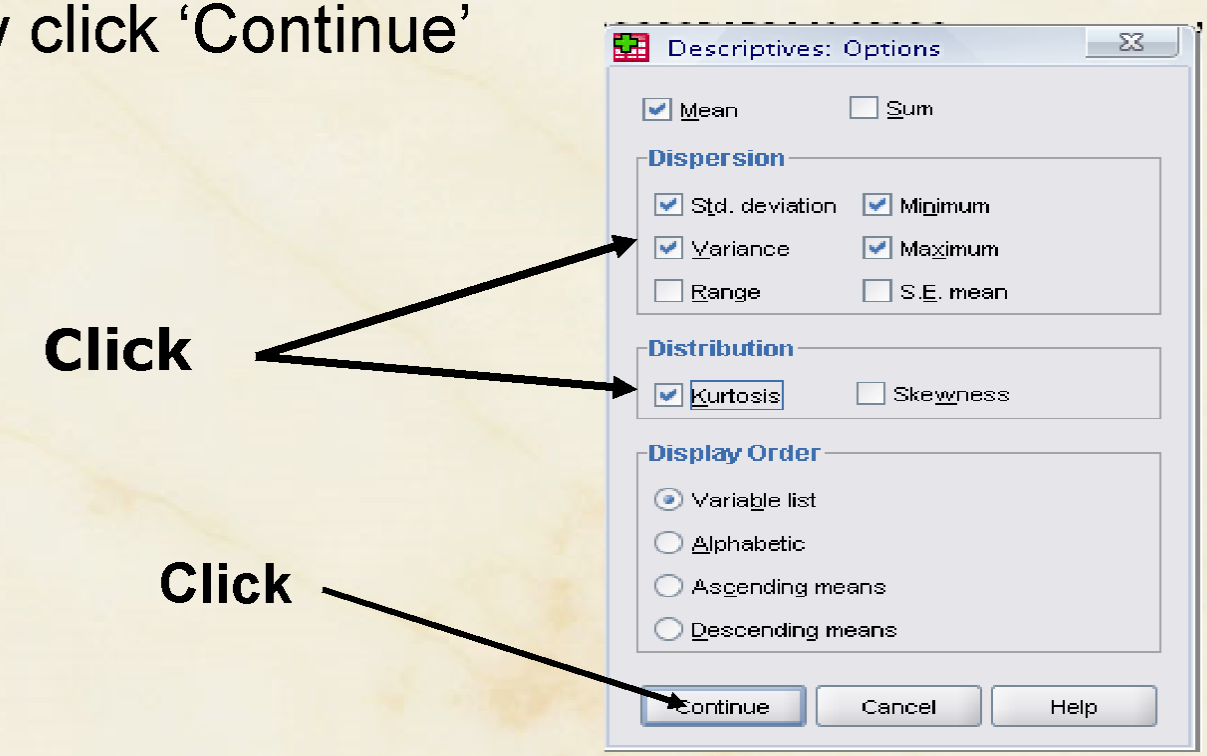
Descriptives

- Click ‘Analyze,’ ‘Descriptive statistics,’ then click ‘Descriptives…’
- Click ‘Educational level’ and ‘Beginning Salary,’ and put it into the variable box.
- Click Options



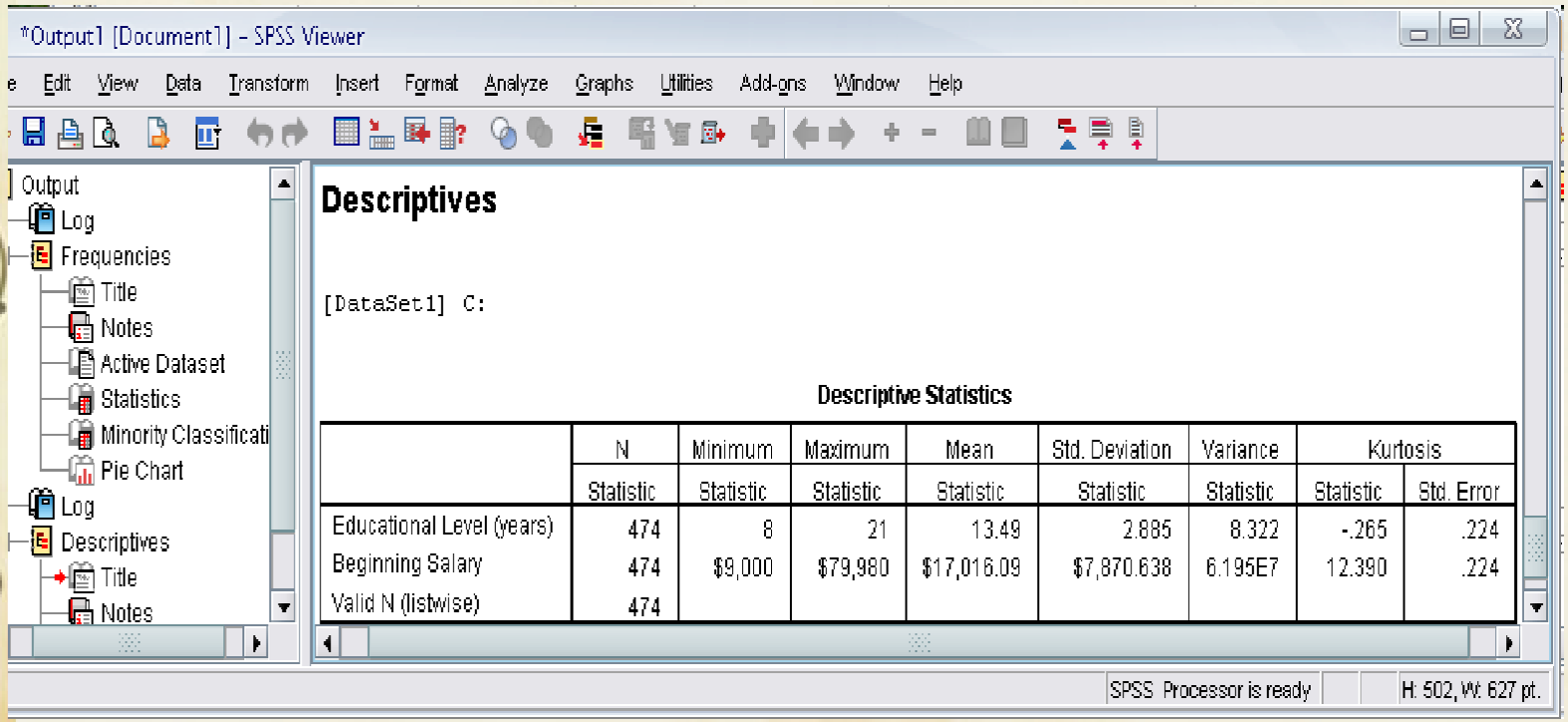
Descriptives

- The options allows you to analyze other descriptive statistics besides the mean and Std.
- Click ‘variance’ and ‘kurtosis’
- Finally click ‘Continue’



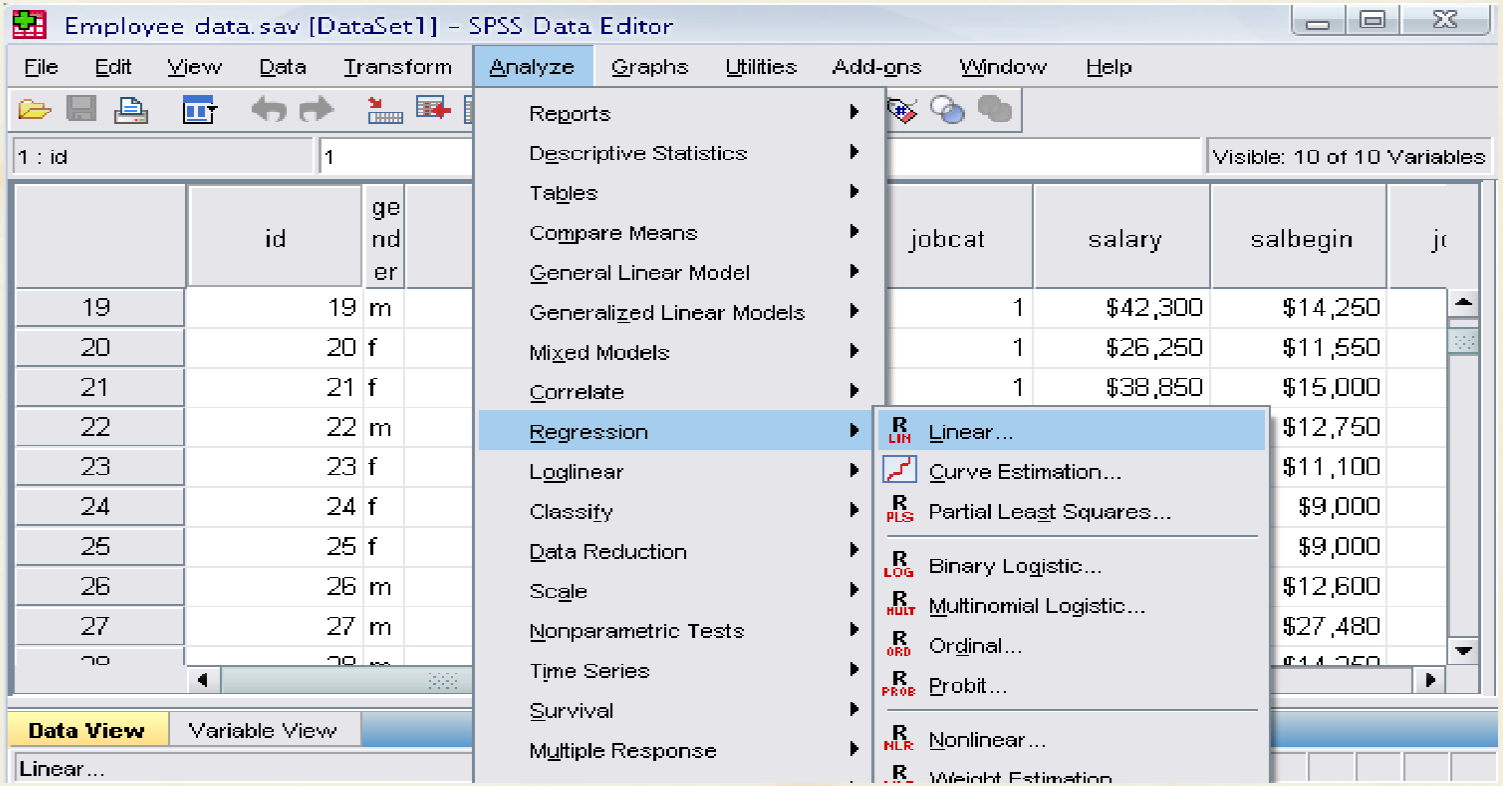
Descriptives

- Finally Click OK in the Descriptives box. You will be able to see the result of the analysis.



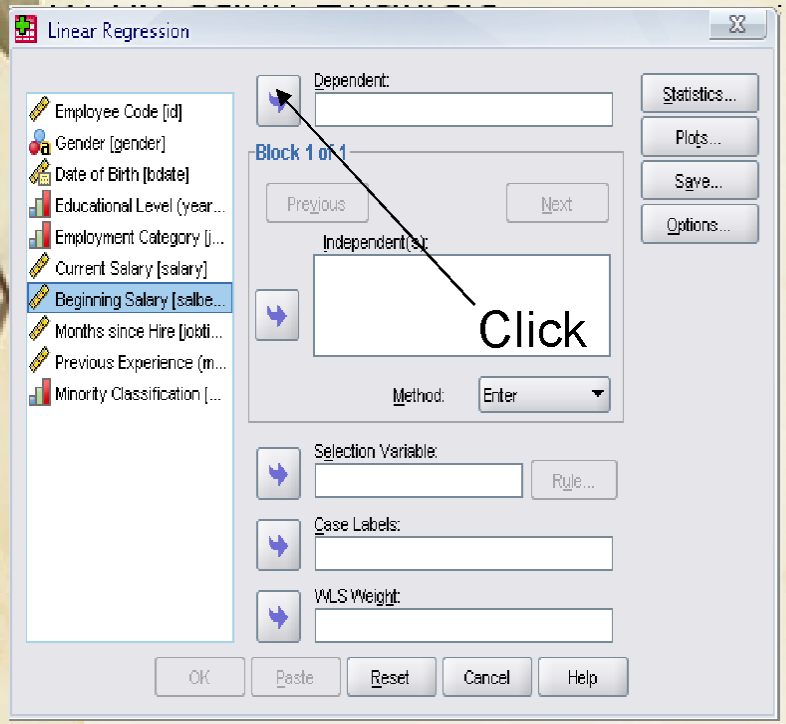
Regression Analysis

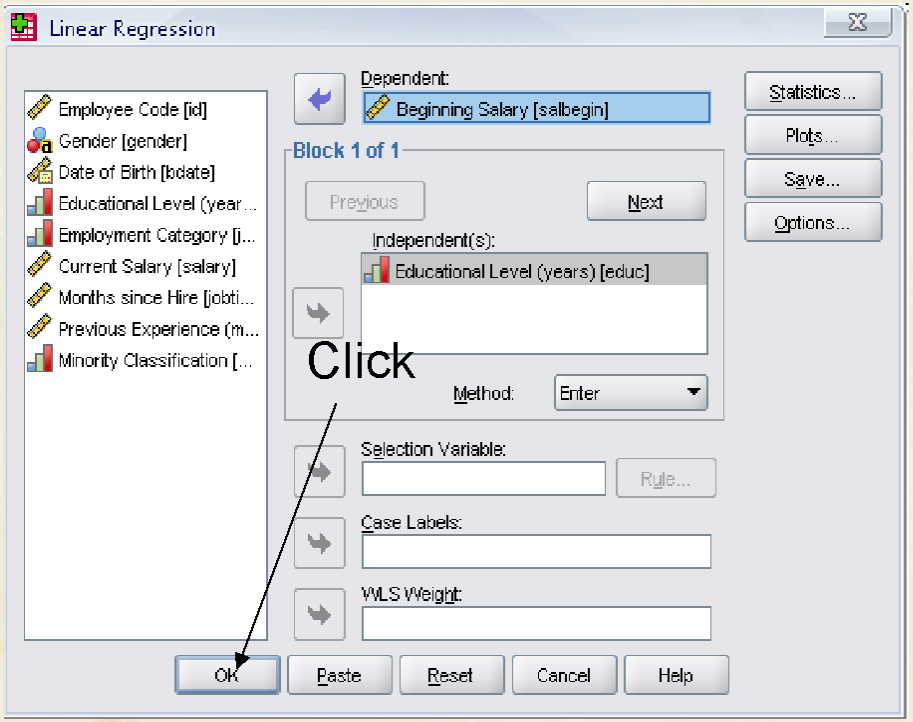
- Click ‘Analyze,’ ‘Regression,’ then click ‘Linear’ from the main menu.



Regression Analysis

- For example let's analyze the model $\text{salbegin} = \beta_0 + \beta_1 \text{edu} + \varepsilon$
- Put 'Beginning Salary' as Dependent and 'Educational Level' as Independent.





Regression Analysis

- Clicking OK gives the result

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.633[a]	.401	.400	$6,098.259

a. Predictors: (Constant), Educational Level (years)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.175E10	1	1.175E10	315.897	.000[a]
	Residual	1.755E10	472	3.719E7		
	Total	2.930E10	473			

a. Predictors: (Constant), Educational Level (years)

b. Dependent Variable: Beginning Salary

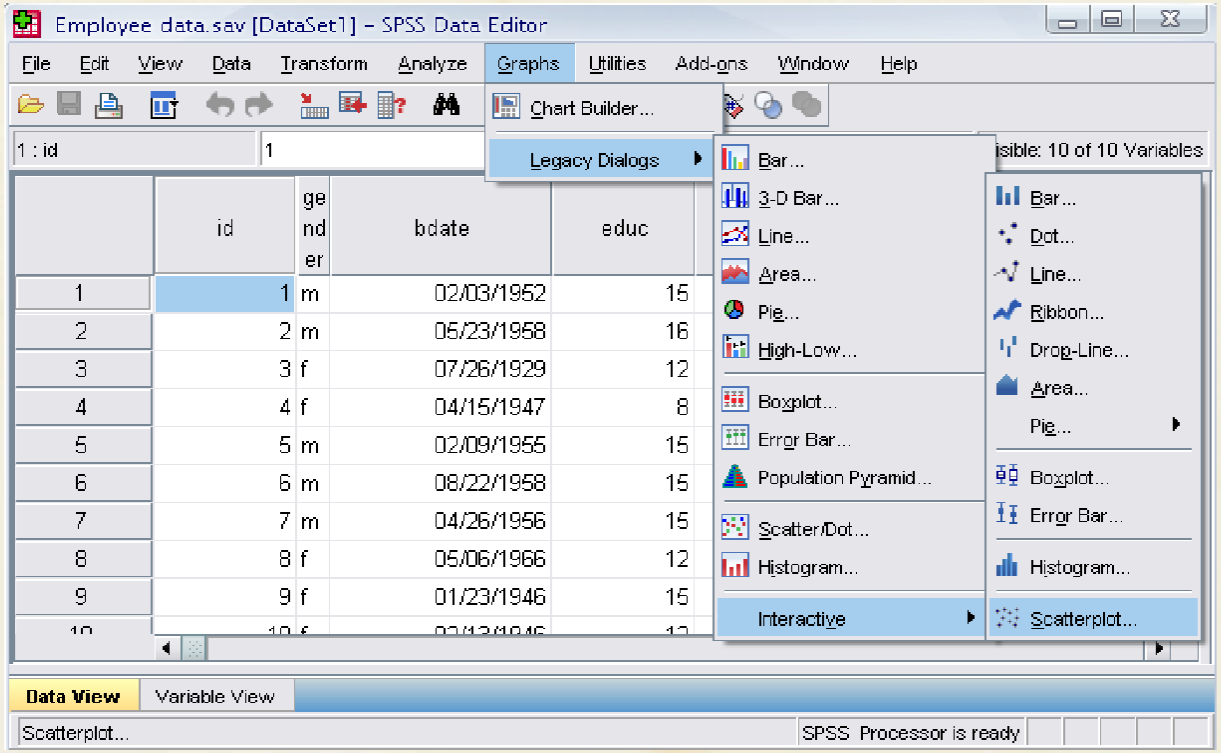
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-6290.967	1340.920		-4.692	.000
	Educational Level (years)	1727.528	97.197	.633	17.773	.000

a. Dependent Variable: Beginning Salary

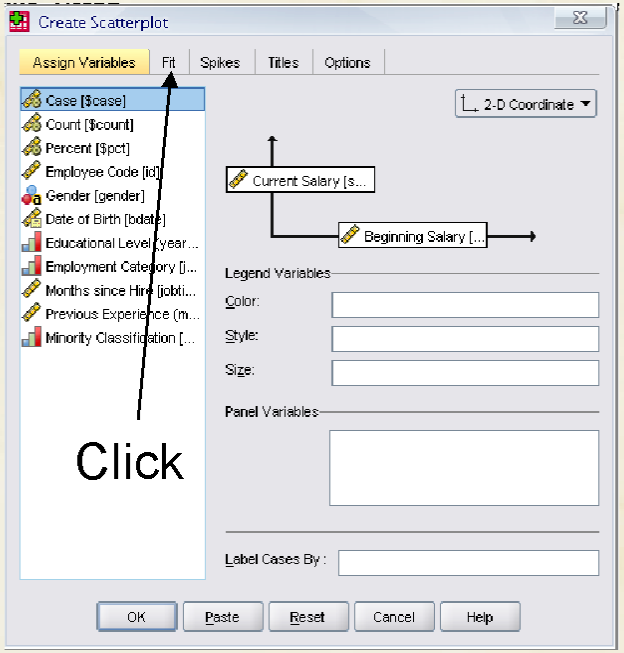
Plotting the regression line

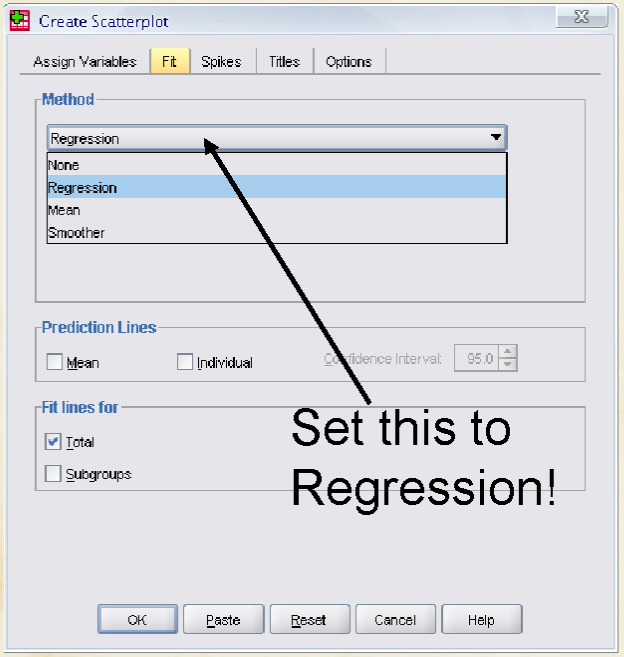
- Click ‘Graphs,’ ‘Legacy Dialogs,’ ‘Interactive,’ and ‘Scatterplot’ from the main menu.

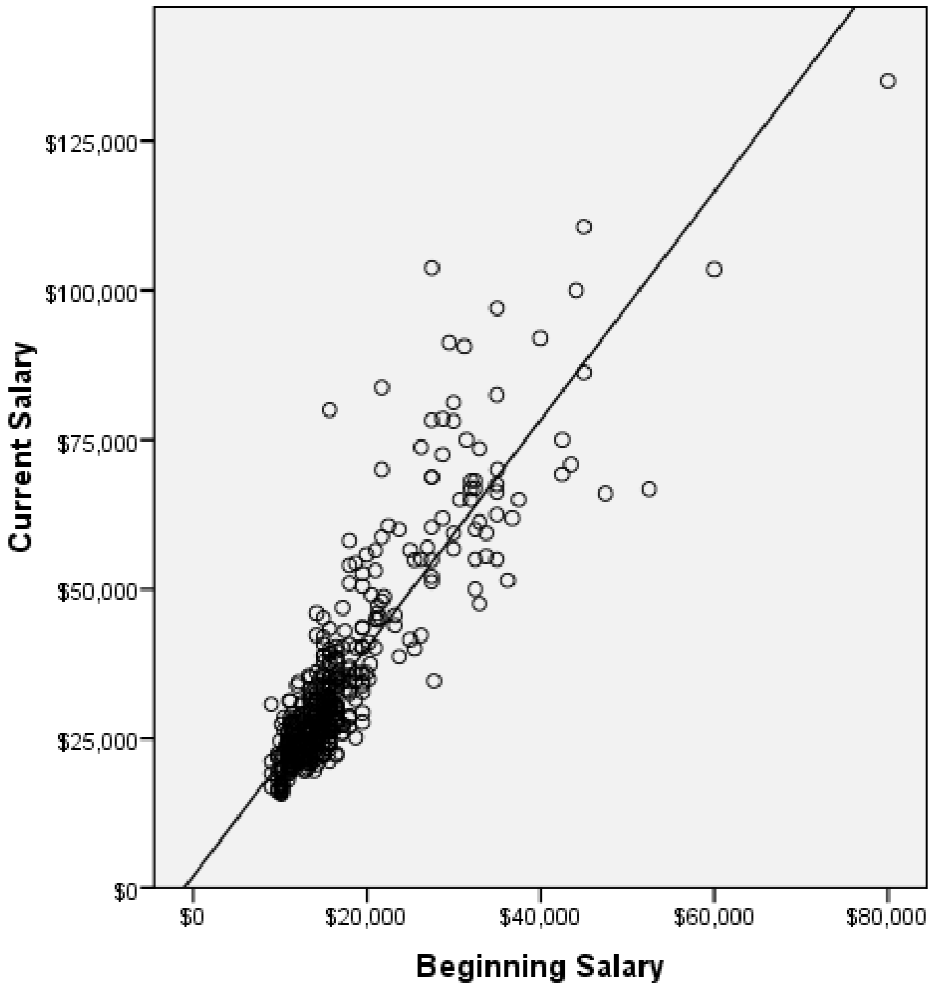


Plotting the regression line

- Drag ‘Current Salary’ into the vertical axis box and ‘Beginning Salary’ in the horizontal axis box.
- Click ‘Fit’ bar. Make sure the Method is regression in the Fit box. Then click ‘OK’.



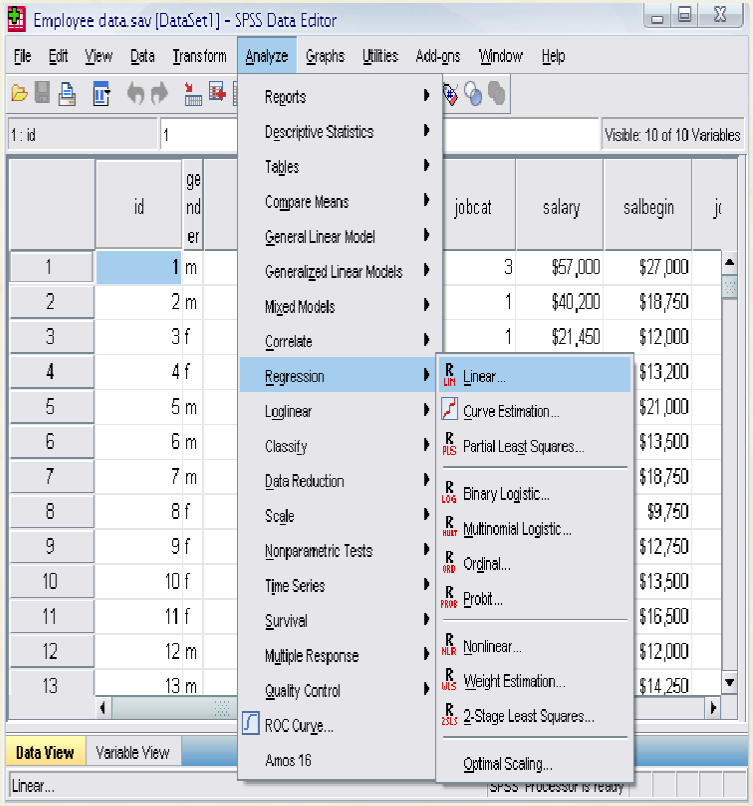


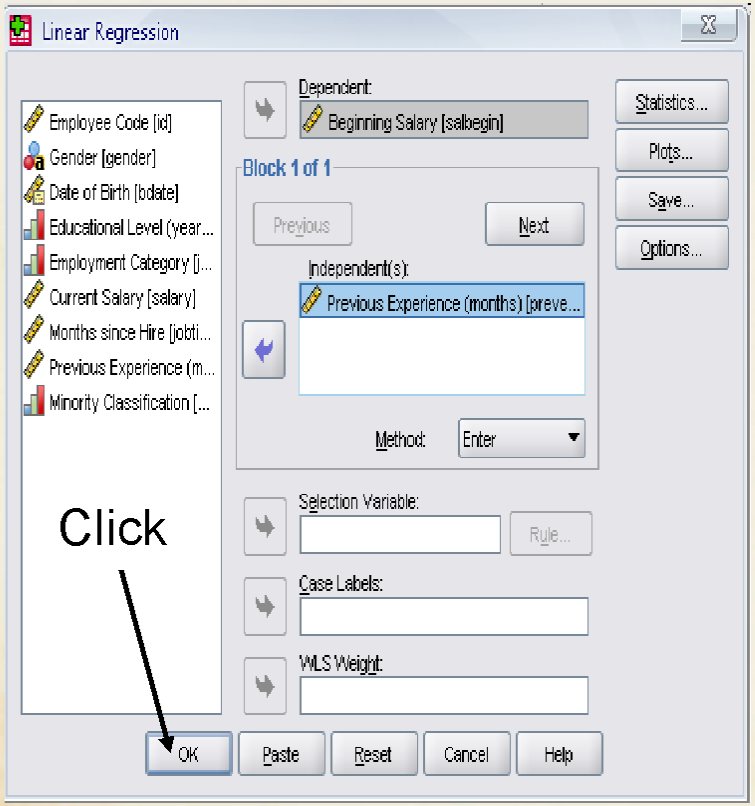

Linear Regression
$125,000
$100,000
$75,000
$50,000
$25,000
$0
Current Salary
$0
$20,000
$40,000
$60,000
$80,000
Beginning Salary
R Sq Linear = 0.775

Practice 5

- Find out whether or not the previous experience of workers has any affect on their beginning salary?
  - Take the variable “salbegin,” and “prevexp” as dependent and independent variables respectively.
- Plot the regression line for the above analysis using the “scatter plot” menu.

Answer


Employee data.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Quality Control
ROC Curve...
Amos 16
Linear...
Curve Estimation...
Partial Least Squares...
Binary Logistic...
Multinomial Logistic...
Ordinal...
Probit...
Nonlinear...
Weight Estimation...
2-Stage Least Squares...
Optimal Scaling...
Data View
Variable View


Linear Regression
Employee Code [id]
Gender [gender]
Date of Birth [bdate]
Educational Level (year...
Employment Category [j...
Current Salary [salary]
Months since Hire [jobti...
Previous Experience (m...
Minority Classification [...
Dependent:
Beginning Salary [salbegin]
Block 1 of 1
Previous
Next
Independent(s):
Previous Experience (months) [preve...
Method:
Enter
Selection Variable:
Rule...
Case Labels:
WLS Weight:
Statistics...
Plots...
Save...
Options...
Click
OK
Paste
Reset
Cancel
Help

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.045[a]	.002	.000	$7,870.942

a. Predictors: (Constant), Previous Experience (months)

b. Dependent Variable: Beginning Salary

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5.969E7	1	5.969E7	.964	.327[a]
	Residual	2.924E10	472	6.195E7		
	Total	2.930E10	473			

a. Predictors: (Constant), Previous Experience (months)

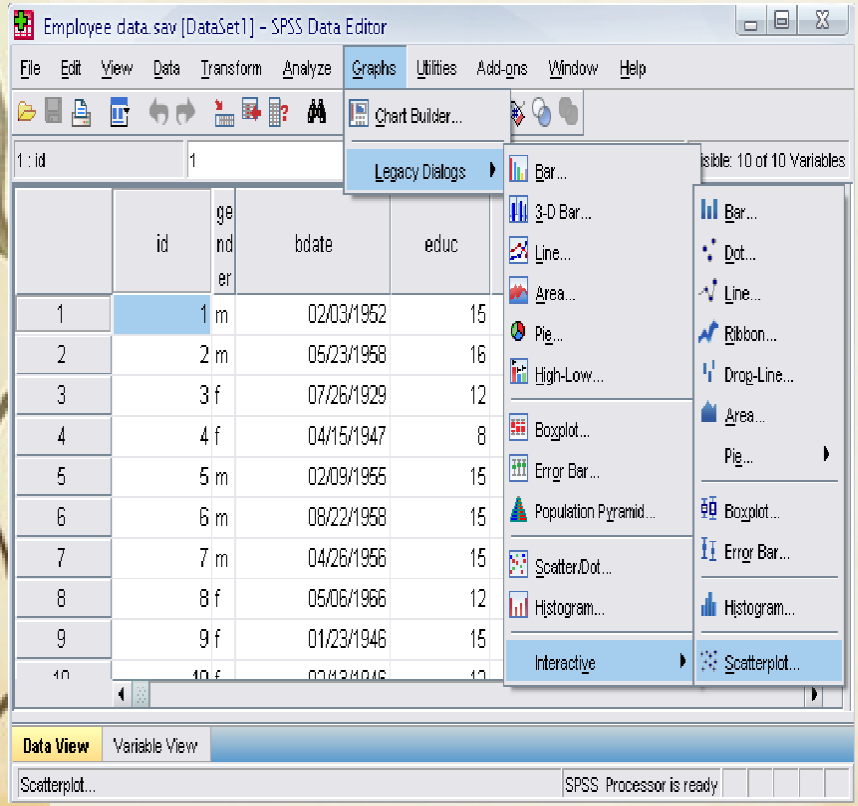
b. Dependent Variable: Beginning Salary

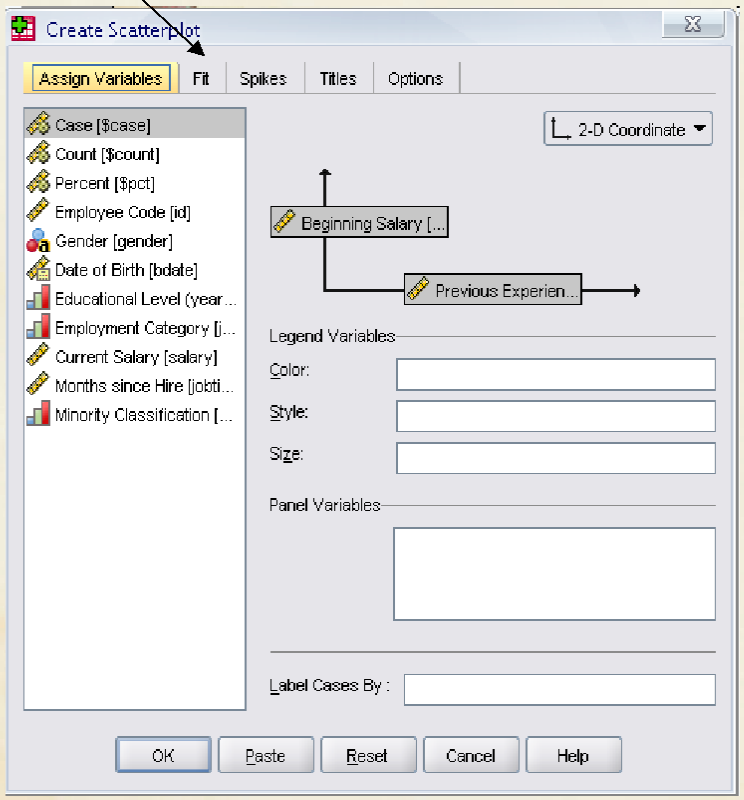
Coefficients[a]

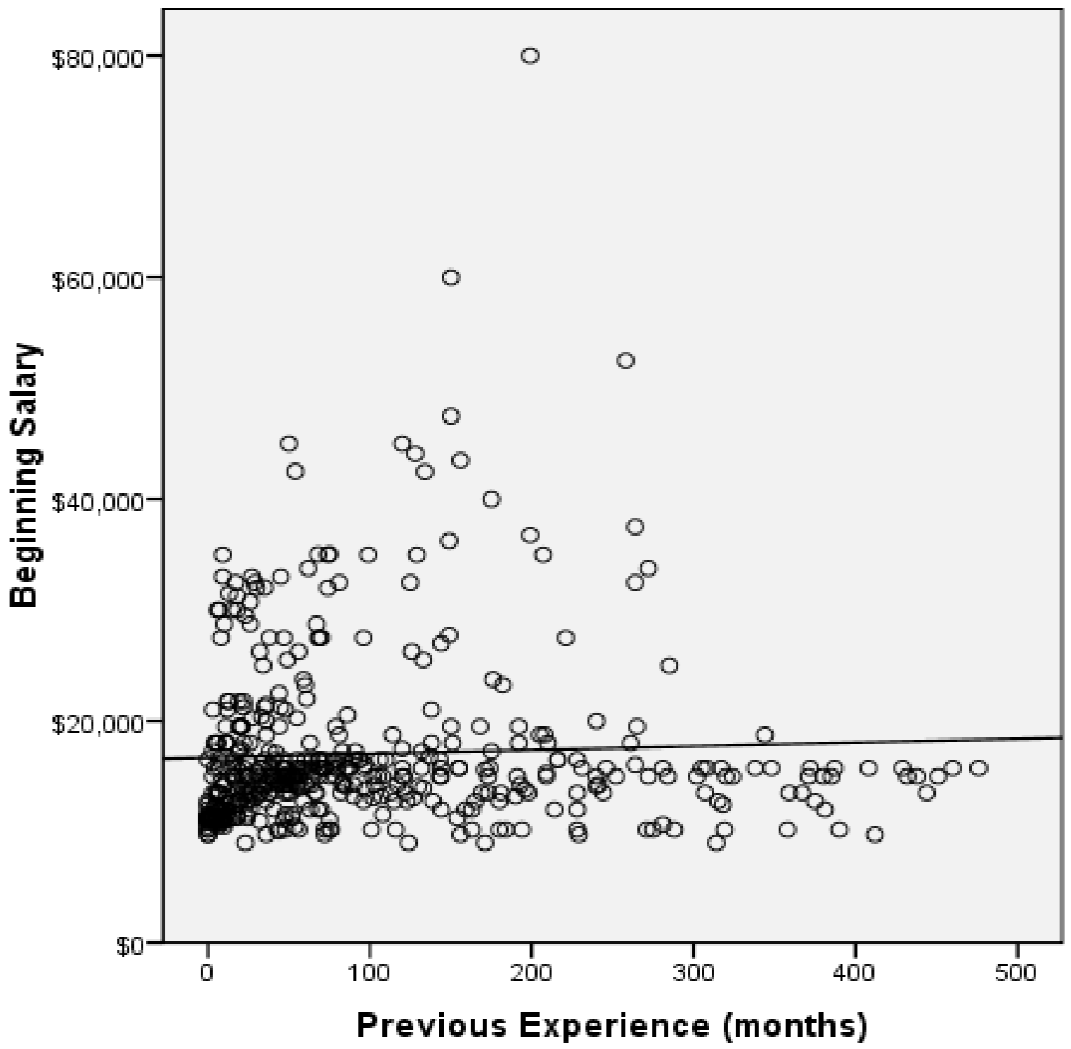
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	16690.478	490.646		34.017	.000
	Previous Experience (months)	3.397	3.460	.045	.982	.327

a. Dependent Variable: Beginning Salary

Click on the “fit” tab to make sure the method is regression






Linear Regression
$80,000
$60,000
$40,000
$20,000
$0
Beginning Salary
0
100
200
300
400
500
Previous Experience (months)
R Sq Linear = 0.002