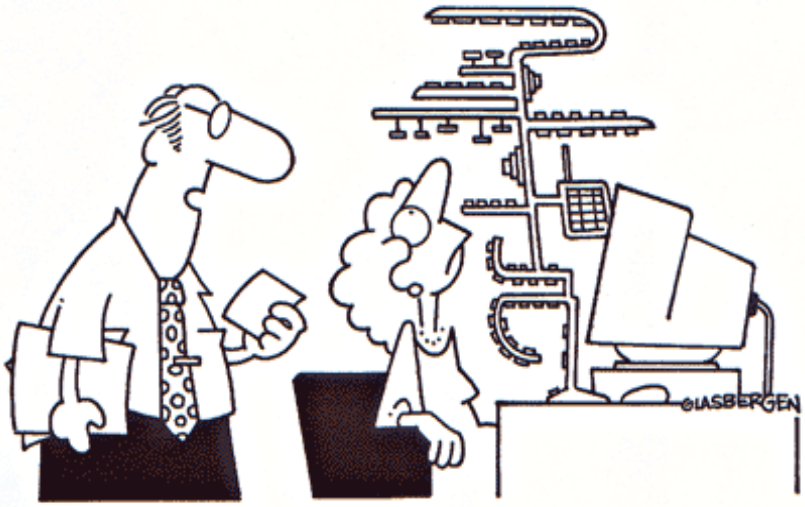
INTRODUCTORY SPSS

Dr Feroz Mahomed Swalaha
fswalaha@dut.ac.za
x2689

1

- Statistics (the systematic collection and display of numerical data) is the most abused area of numeracy.
- "97% of statistics are made up on the spot"
- Unless appropriate statistical techniques are used, the results are meaningless:
  - Garbage In = Garbage Out!
- The most important aspect of statistics is:
- Choosing the right statistical test !

2



"It's the new keyboard for the statistics lab. Once you learn how to use it, it will make computation of the standard deviation easier."

3

PROBLEM SOLVING METHODS

- What is a problem?
- A problem is a question that motivates you to search for a solution.
- What is problem solving?
- Finding a solution to a problem by developing an understanding of the problem through the creation and/or manipulation of processes and concepts.
  - Understand and explore the problem;
  - Find a strategy;
  - Use the strategy to solve the problem;
  - Look back and reflect on the solution.

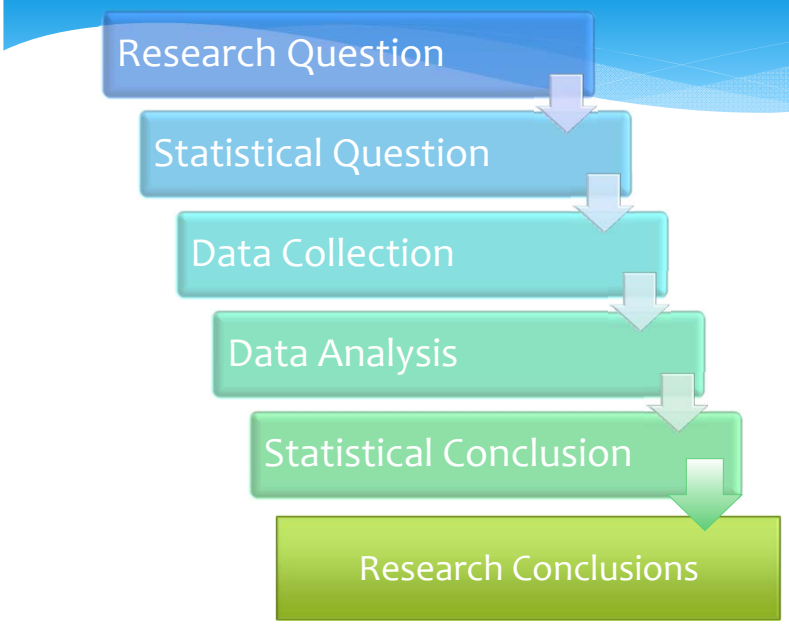
4

PROBLEM SOLVING STRATEGIES

Problem solving strategies:

- Split problems into parts.
- Analyze the given values
- Draw (this includes drawing pictures and diagrams)
- Make a List (this includes making a table)
- Think (this means using skills you already know)
- Think about the statistical methods that are used to solve the problem.
- Analyze the efficiency of the result.

5



6

Statistics

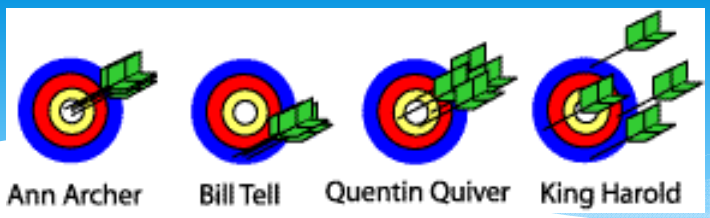
* using the scientific methods of
  - collecting
  - processing
  - reducing
  - presenting
  - analysing
  - interpreting data
  - making inferences
* drawing conclusions from
* numerical data

7

Data measurement instrument

Measurement error can be subdivided into:
VALIDITY, meaning how close the test comes to measuring the variable we are interested in (sensitivity & specificity),
and
RELIABILITY, meaning how consistent the test is when used by different observers or over different periods of time.

8



* Ann Archer: Good precision (grouping) and accuracy (bull's-eye!)
* Bill Tell: Good precision, poor accuracy
* Quentin Quiver: Reasonable accuracy, poor precision
* King Harold: Poor accuracy and precision

9

WHAT TYPE OF VARIABLE?

* Four types of variable
* Nominal (Categorical)
* Ordinal
* Interval
* Ratio
* (latter two grouped as "scale" in SPSS)

10

Nominal

- Mutually exclusive unordered categories to classify
- Examples
  - Genotypes
  - Motor vehicle makes
  - City names
  - Gender
- Computing average would be meaningless
- Easy way to remember – "named"

11

Ordinal

- Order matters but not difference in values
- Examples
  - Pain on a scale of 1 to 10
  - 7>5>3 but difference between each is not not the same
  - Movie ratings from ★ to ★★★★★
  - Level of bacterial contamination 1-5
- Types of mathematical analysis that can be conducted on ordinal level data are limited.

12

Interval

Difference between two values is meaningful

Examples

Difference between 100°C and 90°C same as between 90°C and 80°C

pH

- -log [H]

13

Ratio

* Same properties as Interval variable
* Has a clear definition of 0.0
* Examples
  * Height
  * Weight
  * Temp in K
* Many statistical analyses can be carried out on this category of variable

14

OK to compute....	Nominal	Ordinal	Interval	Ratio
frequency distribution.	Yes	Yes	Yes	Yes
median and percentiles.	No	Yes	Yes	Yes
add or subtract.	No	No	Yes	Yes
mean, standard deviation, standard error of the mean.	No	No	Yes	Yes
ratio, or coefficient of variation.	No	No	No	Yes

15

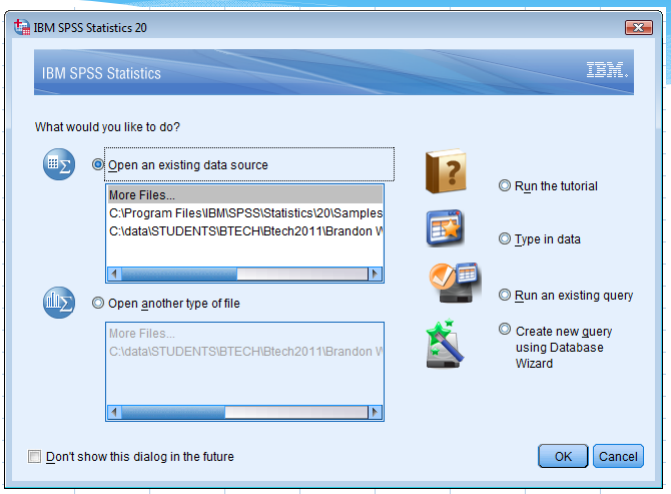
Some concerns

* **Colour in design - undefined**
* **Colour in physics – ratio**
* **Taste**
* **Perceptions**

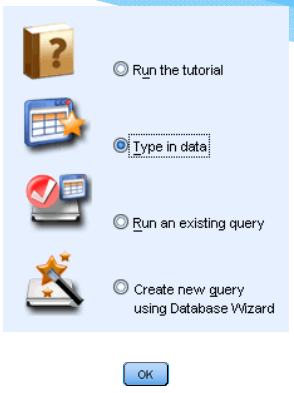
* **“If you torture data enough, it will confess” (anon.)**

16

OPEN SPSS

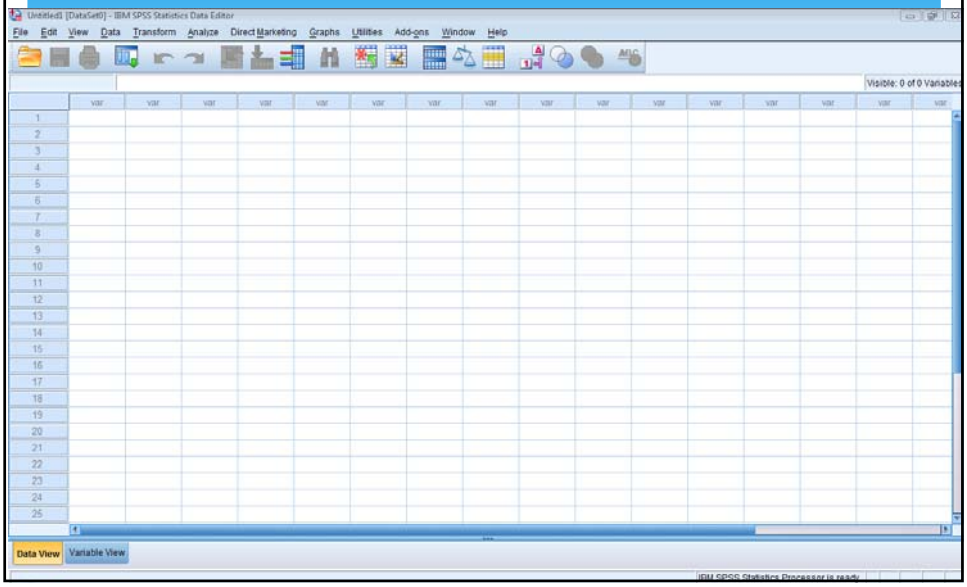


17



18

Data View

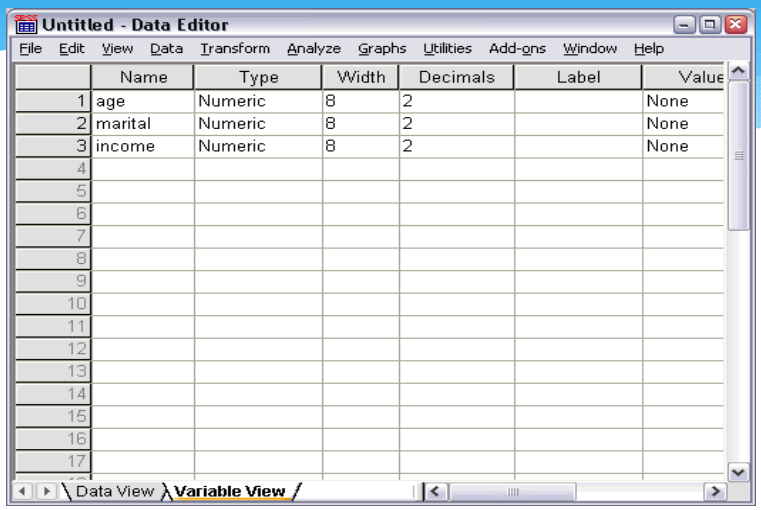


ENTERING DATA IN NEW SHEET

- In Data View, columns represent variables, and rows represent cases (observations).
- In Variable View, each row is a variable, and each column is an attribute that is associated with that variable
- Click the Variable View tab at the bottom of the Data Editor window.
- You need to define the variables that will be used. In this case, only three variables are needed:
- *age*, *marital status*, and *income*
- Click the Data View tab to continue entering the data.

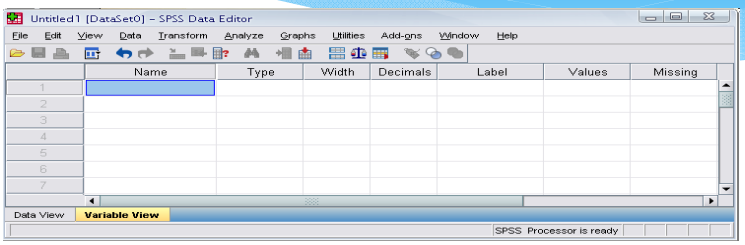
20

Variable View



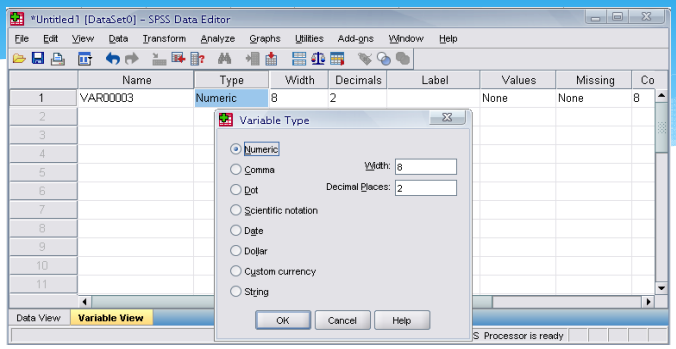
* **Information about the data stored in the dataset**

21



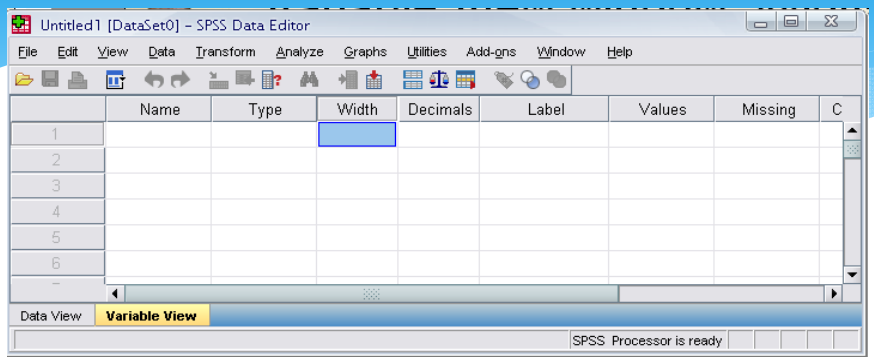
* **Name**
  * **The first character of the variable name must be alphabetic**
  * **Variable names must be unique, and have to be less than 64 characters.**
  * **Spaces are NOT allowed.**

22



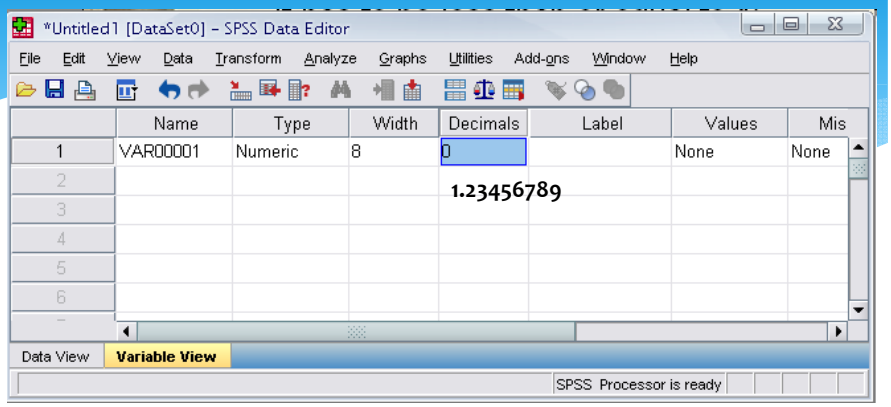
- **Type**
  - **Click on the 'type' box. The two basic types of variables that you will use are numeric and string. This column enables you to specify the type of variable.**

23



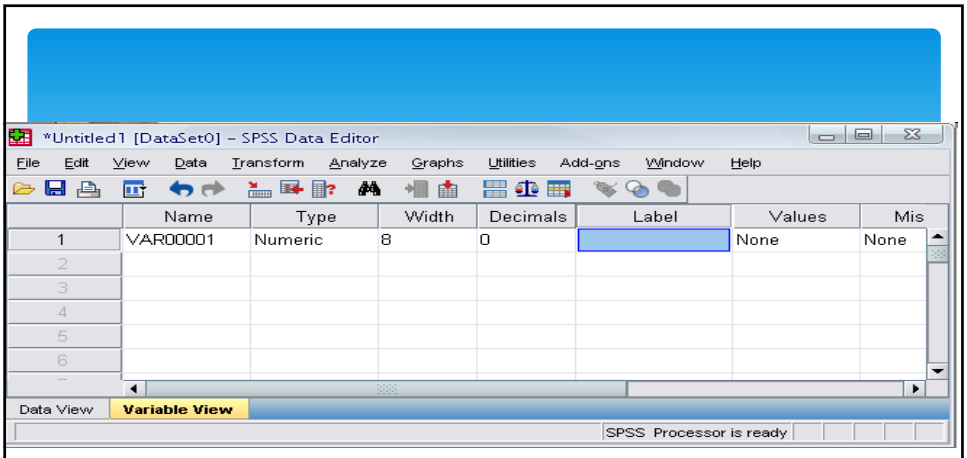
- **Width**
  - **Width allows you to determine the number of characters SPSS will allow to be entered for the variable**

24



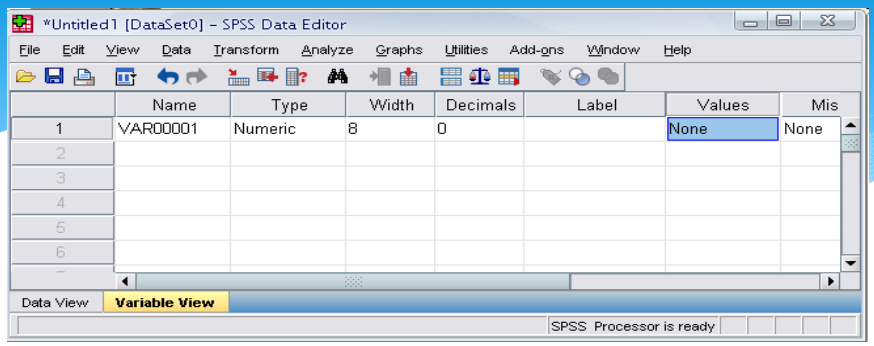
- **Decimals**
  - **Number of decimals**
  - **It has to be less than or equal to 16**

25



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

26



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

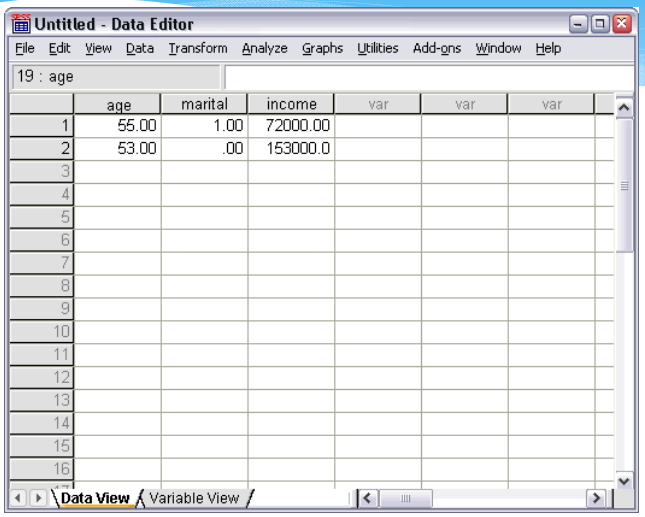
27

DATA COLLECTED

NAME	Age	Marital status	Income	Sex
Bob	55	Married	72000	Male
Jane	53	Single	153000	Female

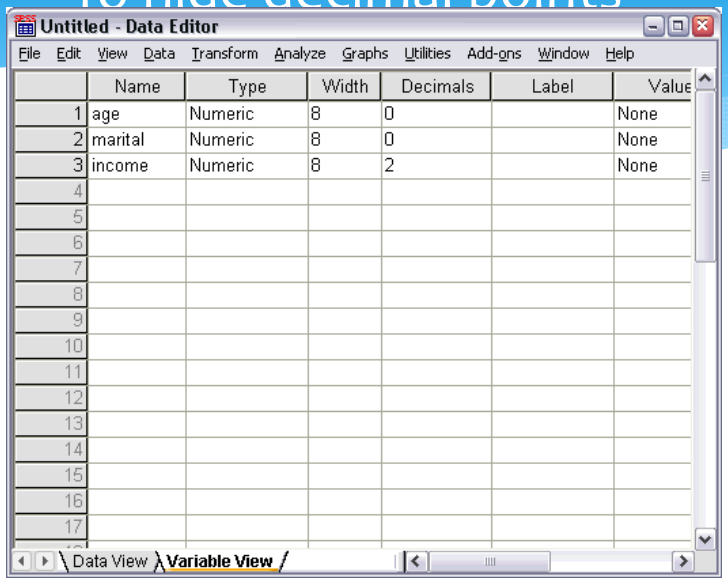
28

Data view



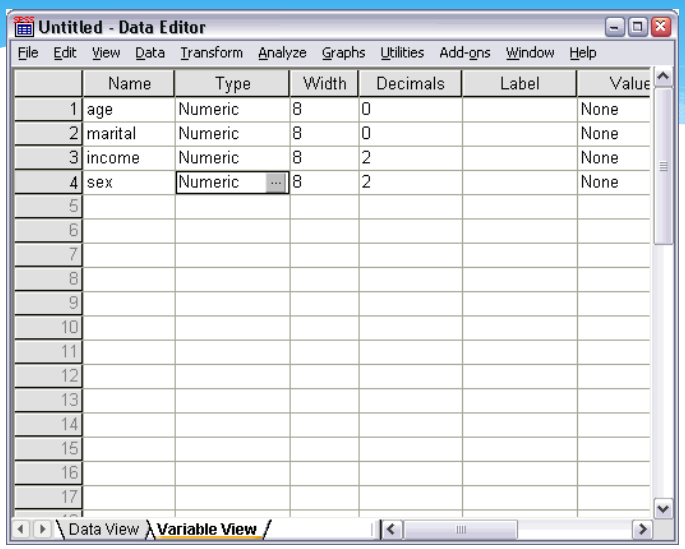
29

To hide decimal points

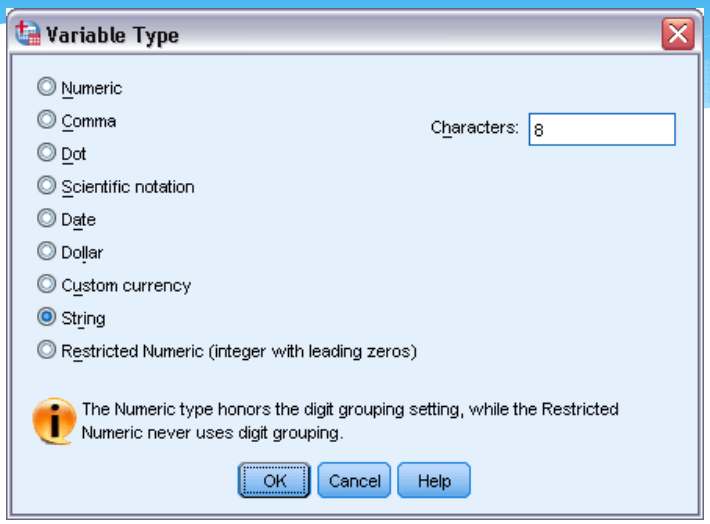


30

Entering String Data

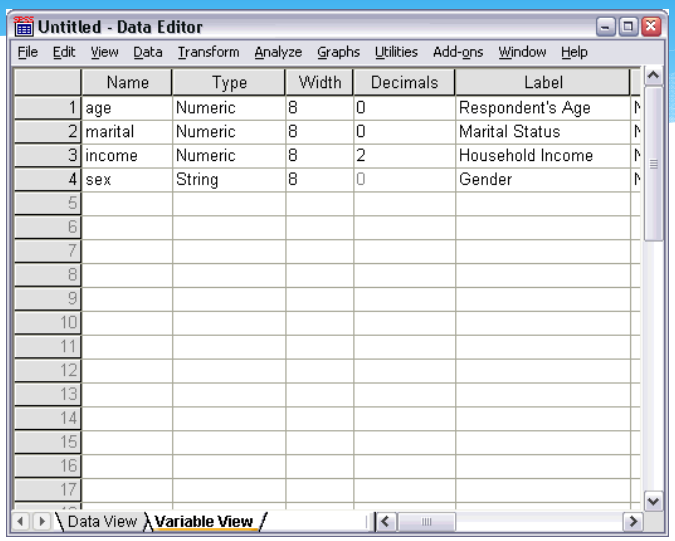


31



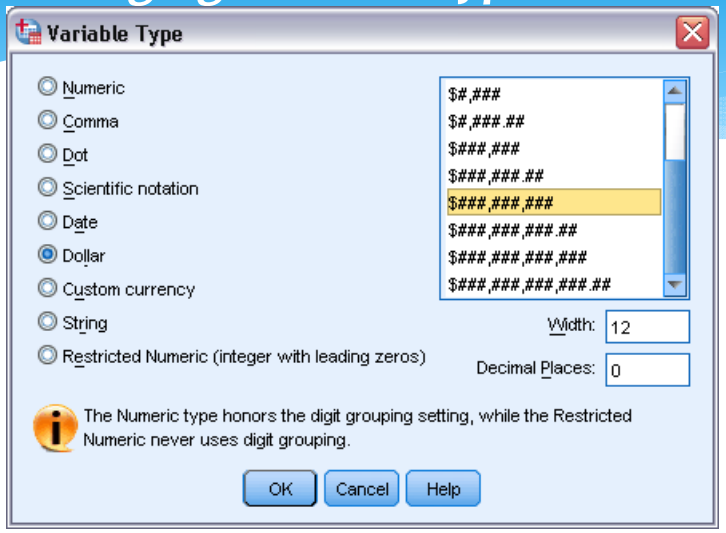
32

Adding Variable Labels



33

Changing Variable Type and Format

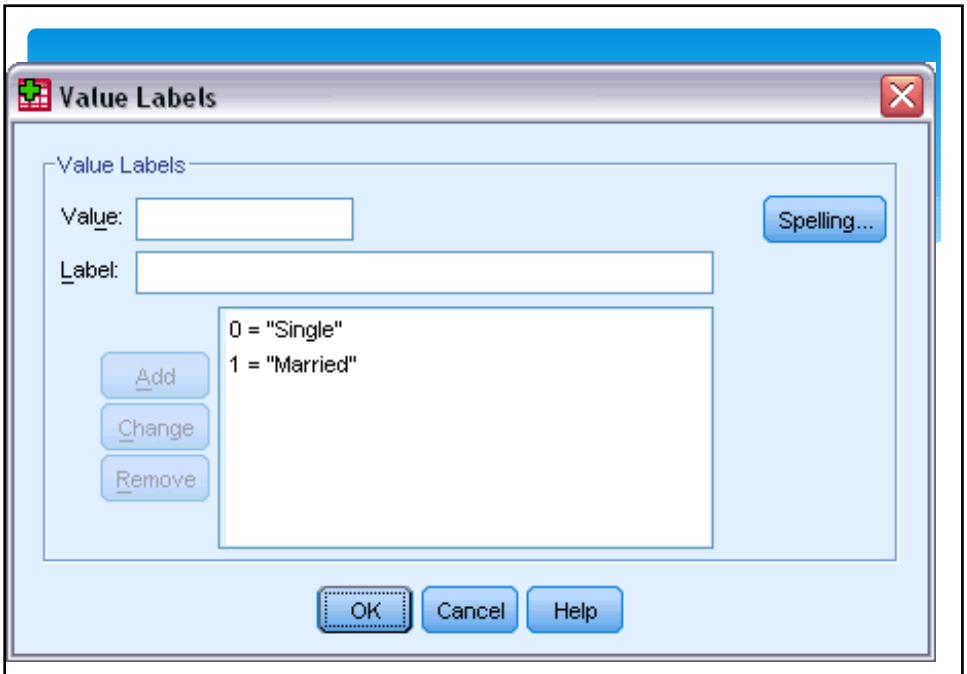


34

Adding Value Labels for Numeric Variables

- Value labels provide a method for mapping your variable values to a string label. In this example,
- there are two acceptable values for the *marital* variable. A value of 0 means that the subject is single,
- and a value of 1 means that he or she is married

35

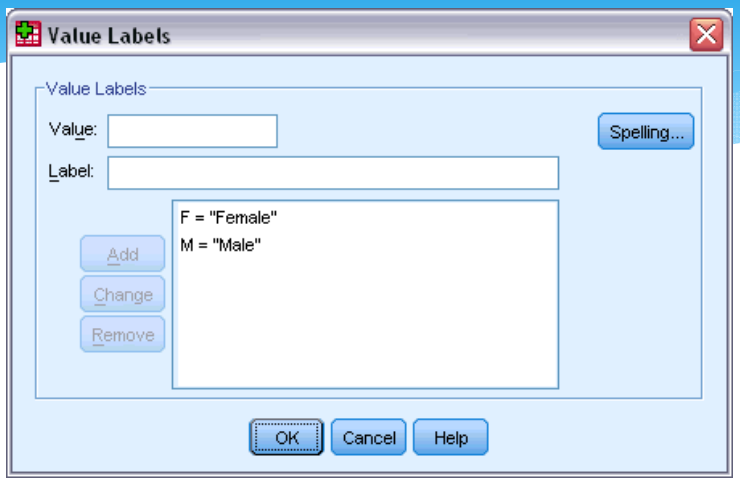


36

Adding Value Labels for String Variables

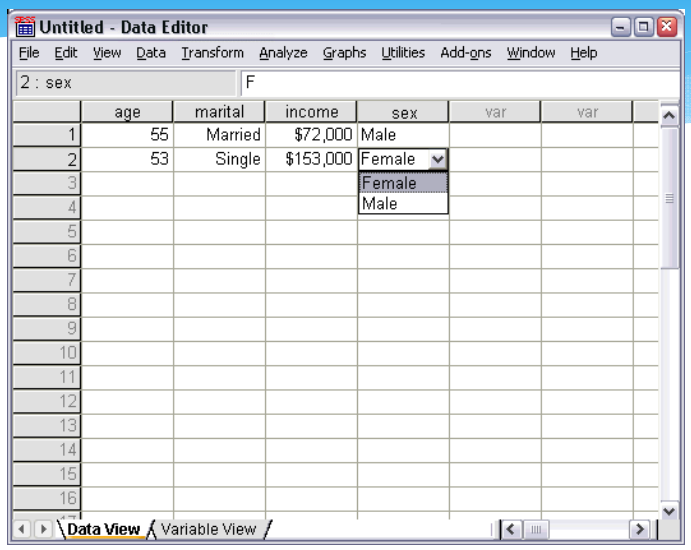
- String variables may require value labels as well. For example, your data may use single letters,
- *M* or *F*, to identify the sex of the subject. Value labels can be used to specify that *M* stands
- for *Male* and *F* stands for *Female*.

37



38

Using Value Labels for Data Entry

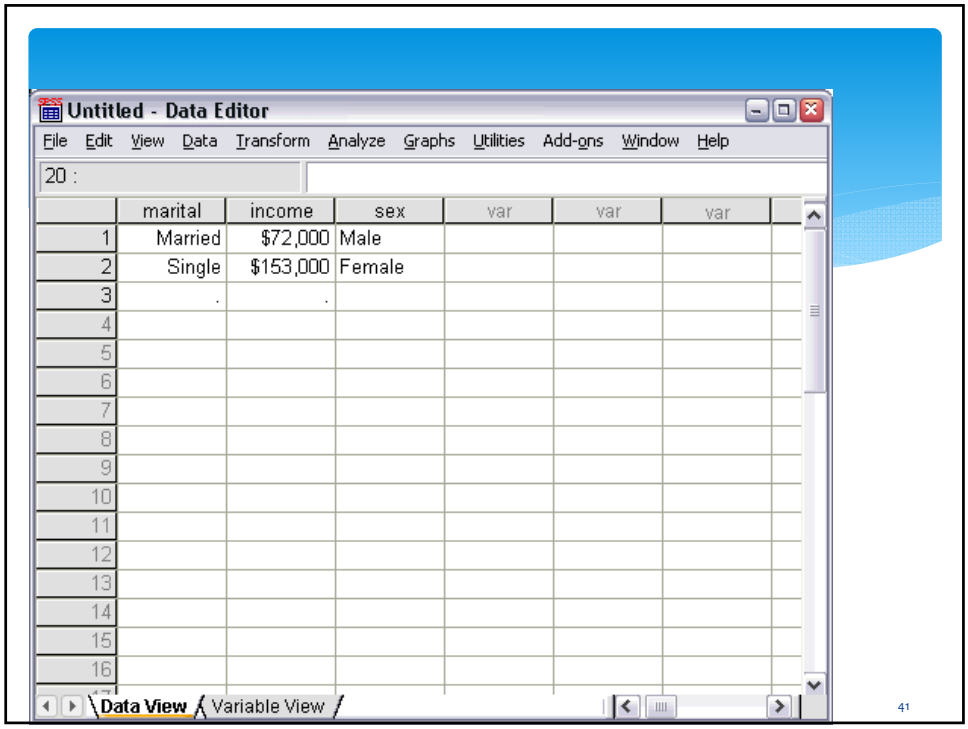


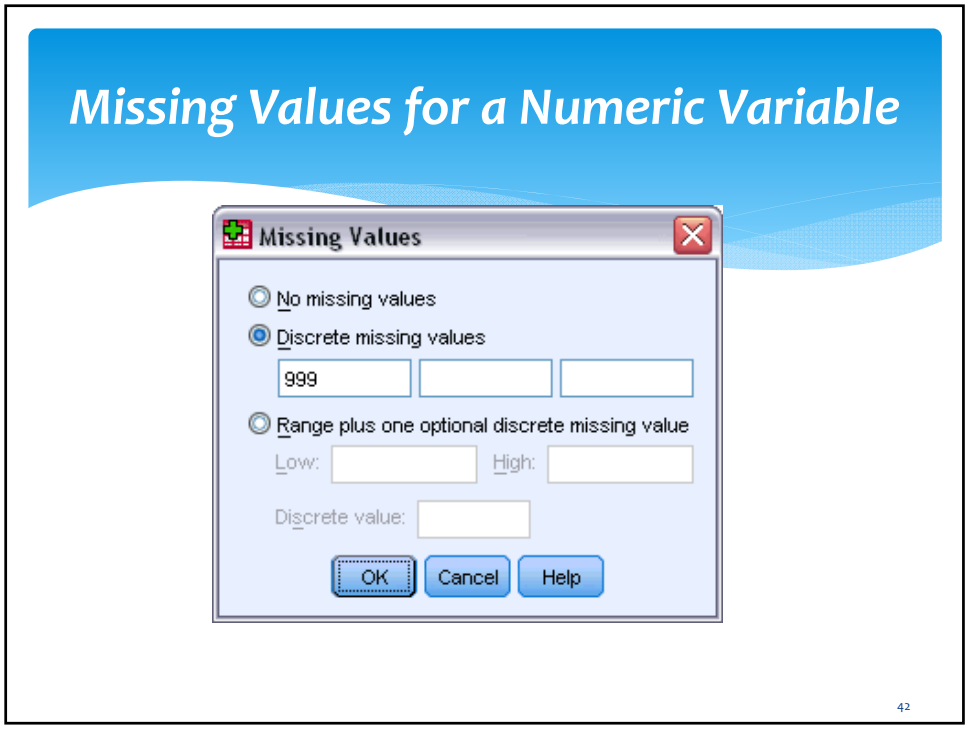
39

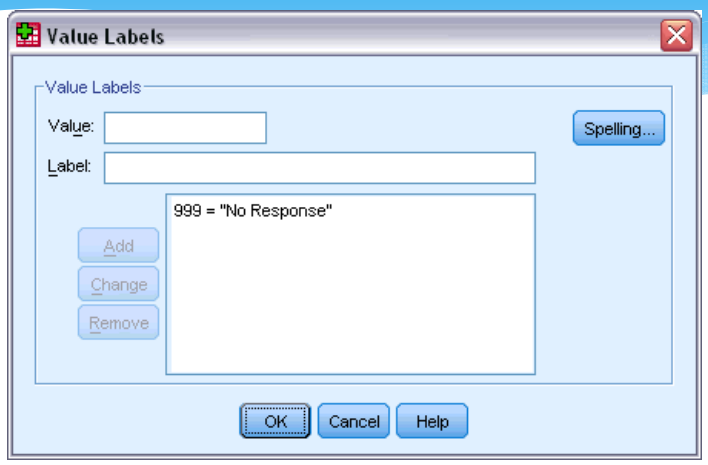
Handling Missing Data

- Missing or invalid data are generally too common to ignore.
- Survey respondents may refuse to answer certain questions, may not know the answer, or may answer in an unexpected format.
- If you don't filter or identify these data, your analysis may not provide accurate results.
- For numeric data, empty data fields or fields containing invalid entries are converted to system-missing, which is identifiable by a single period
- The reason a value is missing may be important to your analysis.
- For example, you may find it useful to distinguish between those respondents who refused to answer a question
- and those respondents who didn't answer a question because it was not applicable

40


Untitled - Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
20 :
marital income sex var var var
1 Married $72,000 Male
2 Single $153,000 Female
3 . .
Data View Variable View
41


Missing Values for a Numeric Variable
Missing Values
No missing values
Discrete missing values
999
Range plus one optional discrete missing value
Low:
High:
Discrete value:
OK
Cancel
Help
42

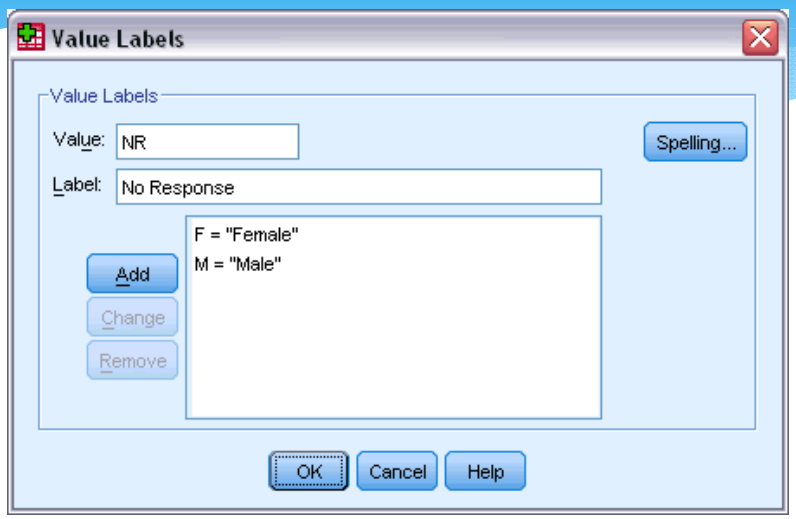


43

Missing Values for a String Variable

- Missing values for string variables are handled similarly to the missing values for numeric variables.
- However, unlike numeric variables, empty fields in string variables are not designated as system-missing.
- Rather, they are interpreted as an empty string.

44

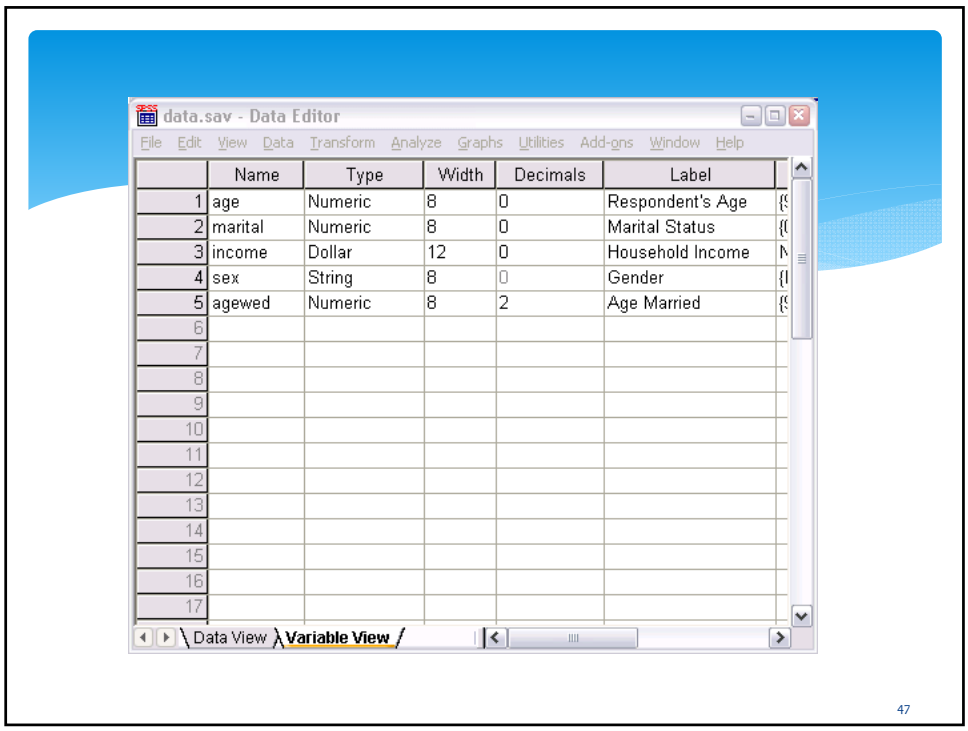


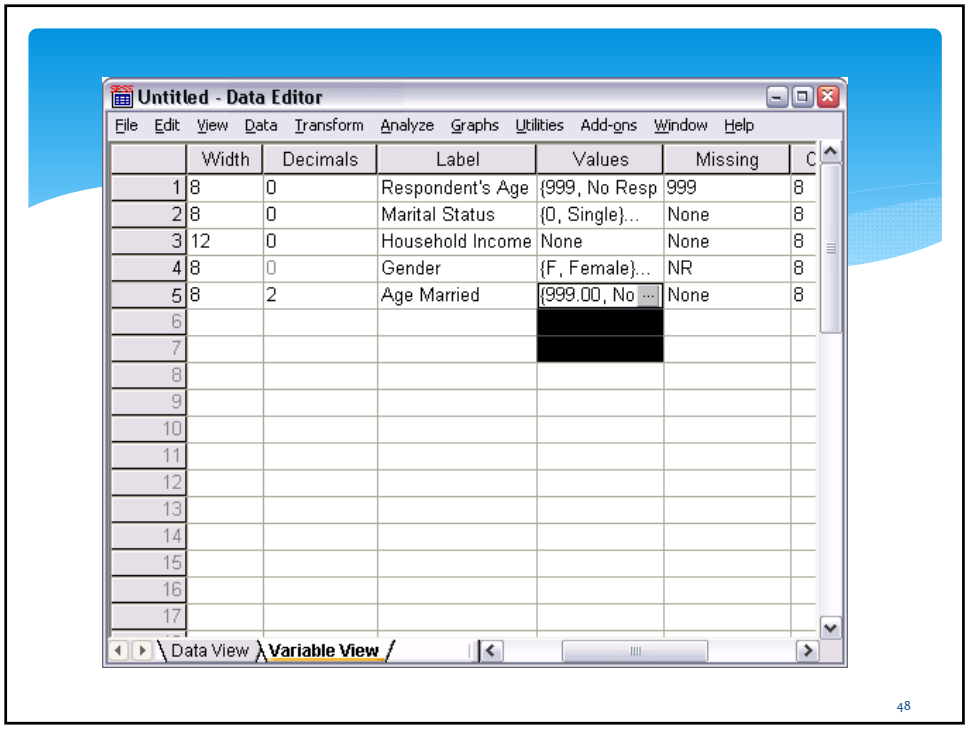
45

Copying and Pasting Variable Attributes

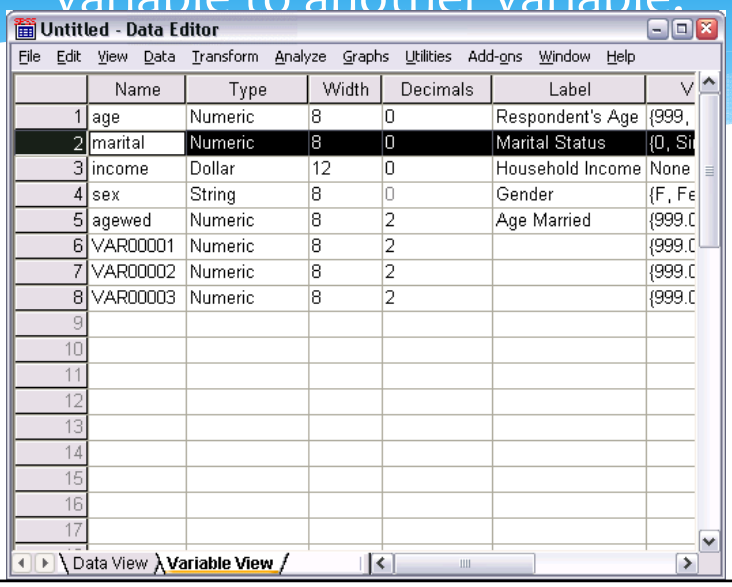
* After you've defined variable attributes for a variable, you can copy these attributes and apply them to other variables.

46

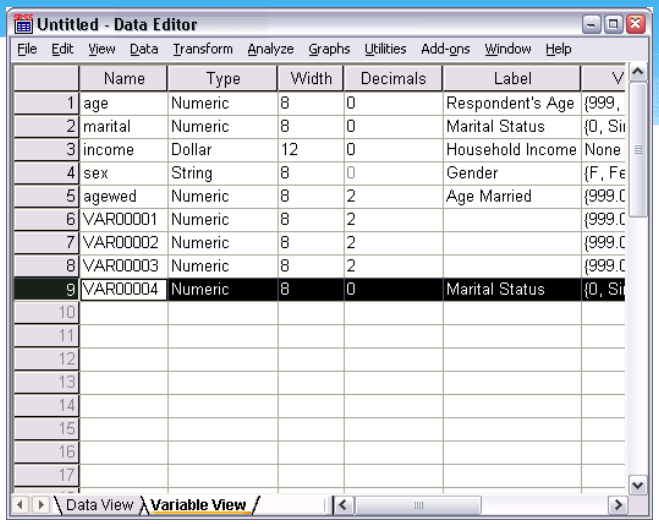

data.sav - Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Name Type Width Decimals Label
1 age Numeric 8 0 Respondent's Age
2 marital Numeric 8 0 Marital Status
3 income Dollar 12 0 Household Income
4 sex String 8 0 Gender
5 agewed Numeric 8 2 Age Married
Data View Variable View
47


Untitled - Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Width Decimals Label Values Missing
1 8 0 Respondent's Age {999, No Resp 999 8
2 8 0 Marital Status {0, Single}... None 8
3 12 0 Household Income None None 8
4 8 0 Gender {F, Female}... NR 8
5 8 2 Age Married {999.00, No None 8
Data View Variable View
48

To copy all attributes from one variable to another variable:



49



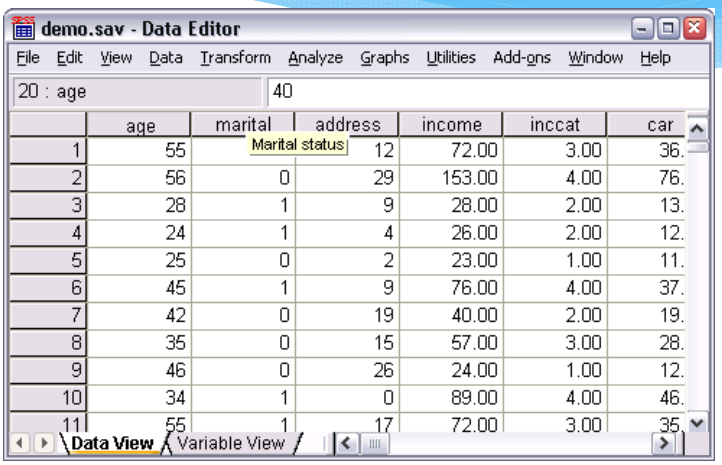
50

Opening a Data File

- From the menus choose:
- File > Open > Data...
- Alternatively, you can use the Open File button on the toolbar.
- *Open File toolbar button*
- A dialog box for opening files is displayed.
- We will be opening demo.sav

51

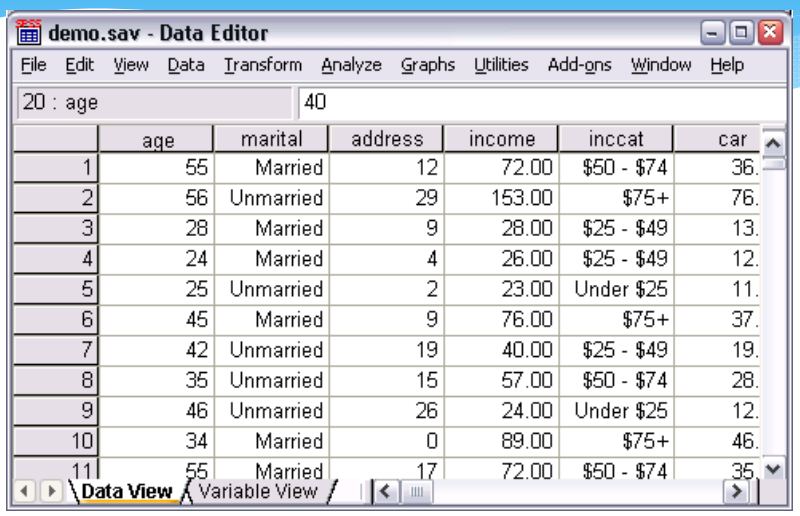
c:\program files\ibm\spss\statistics\20\samples\english\



	age	marital	address	income	inccat	car
1	55		12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35.

52

Value labels displayed



	age	marital	address	income	inccat	car
1	55	Married	12	72.00	$50 - $74	36.
2	56	Unmarried	29	153.00	$75+	76.
3	28	Married	9	28.00	$25 - $49	13.
4	24	Married	4	26.00	$25 - $49	12.
5	25	Unmarried	2	23.00	Under $25	11.
6	45	Married	9	76.00	$75+	37.
7	42	Unmarried	19	40.00	$25 - $49	19.
8	35	Unmarried	15	57.00	$50 - $74	28.
9	46	Unmarried	26	24.00	Under $25	12.
10	34	Married	0	89.00	$75+	46.
11	55	Married	17	72.00	$50 - $74	35.

53

Changing Value Labels

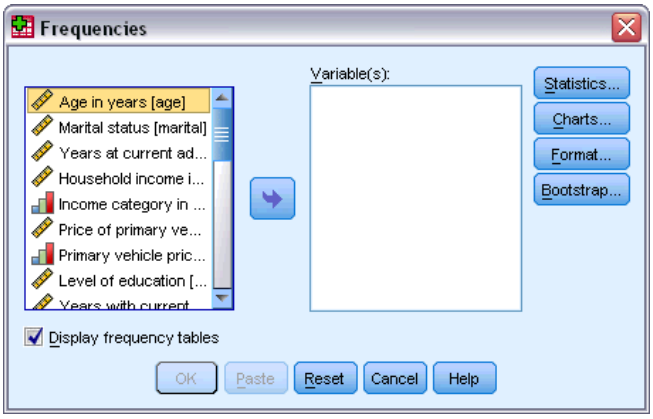
- By default, the actual data values are displayed.
- To display labels:
- From the menus choose:
- View > Value Labels
- Alternatively, use Value Lab[illegible] on toolbar.

- Descriptive value labels will be displayed

54

Create Simple Frequency Table

- From the menus choose:
- Analyze > Descriptive Statistics > Frequencies...



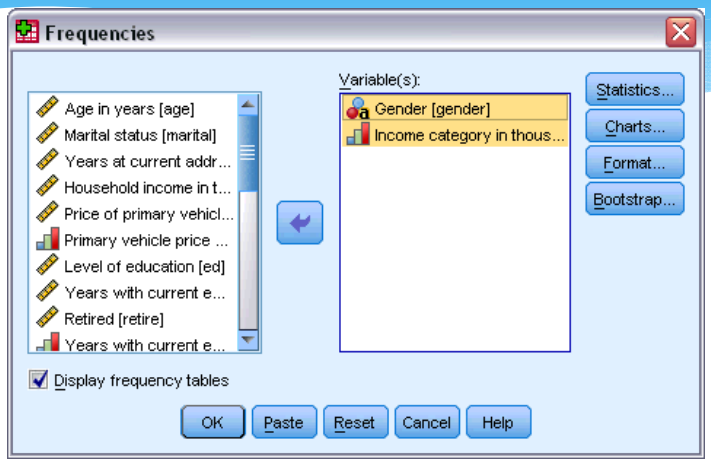
55

What the symbols mean

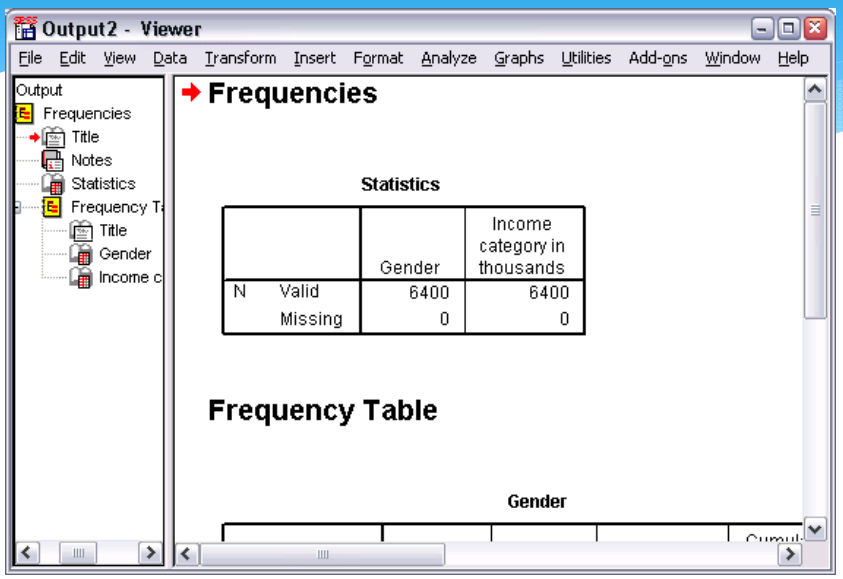
	Numeric	String	Date	Time
Scale (Continuous)		n/a		
Ordinal				
Nominal				

56

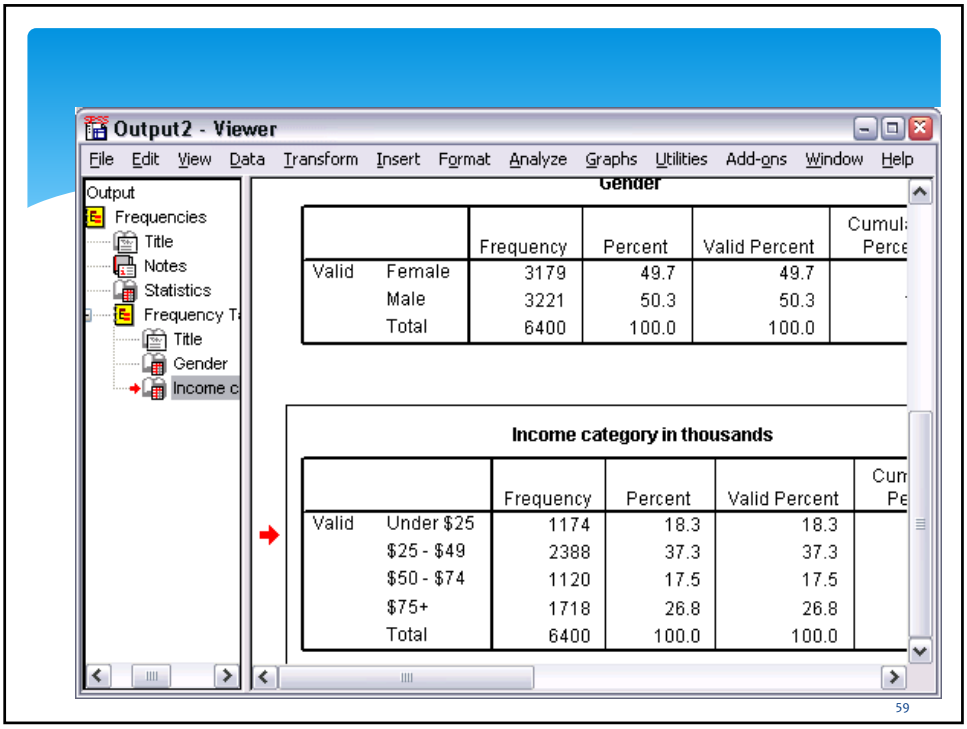
gender and income category frequencies



57



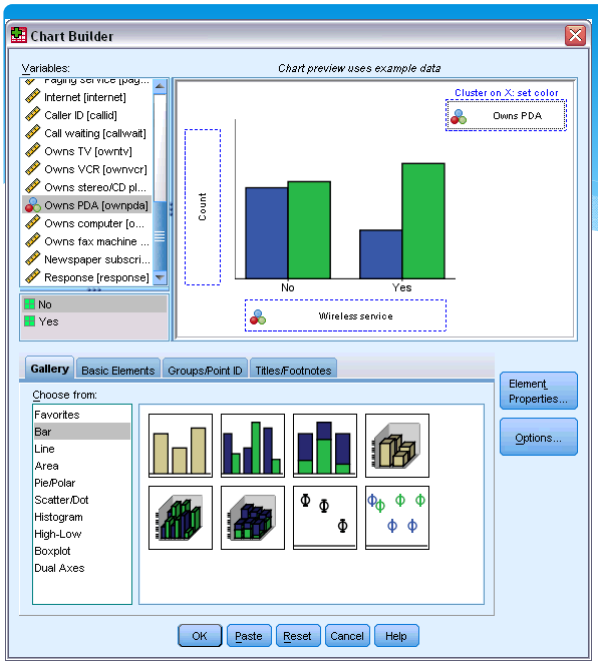
58



Creating Graphs

- From the menus choose:
- Graphs > Chart Builder...
- Click the Gallery tab (if it is not selected).
- Click Bar (if it is not selected).
- Drag the Clustered Bar icon onto the canvas, which is the large area above the Gallery.

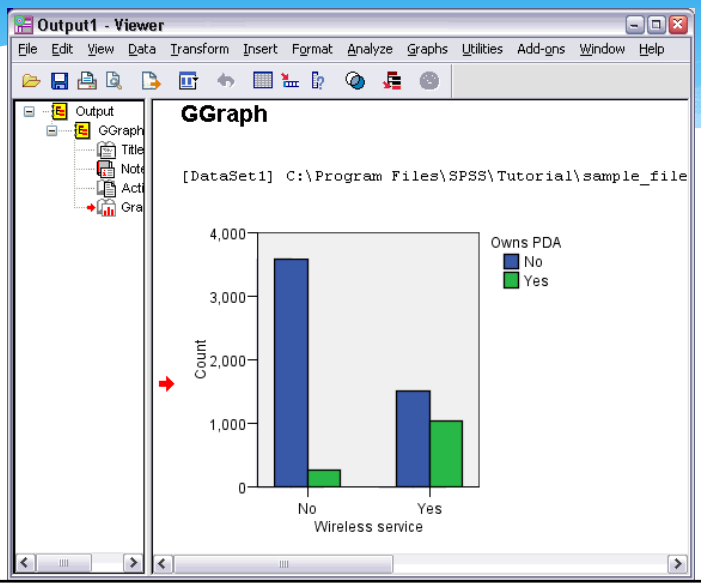
60



- Scroll down the Variables list, right-click *Wireless service [wireless]*, and then choose Nominal
- as its measurement level.
- Drag the *Wireless service [wireless]* variable to the *x* axis.
- Right-click *Owns PDA [ownpda]* and choose Nominal as its measurement level.
- Drag the *Owns PDA [ownpda]* variable to the cluster drop zone in the upper right corner of
- the canvas.
- Click OK to create the chart.

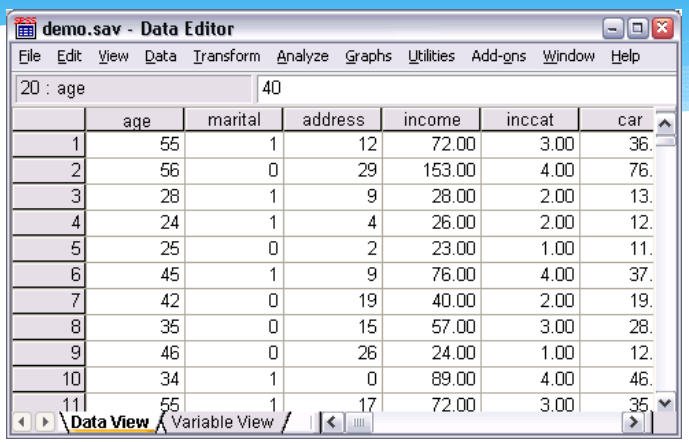
61

Bar Chart in viewer window



62

Reading and Importing Data

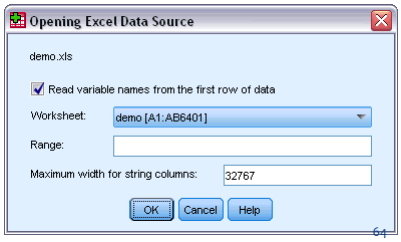


	age	marital	address	income	inccat	car
1	55	1	12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35.

63

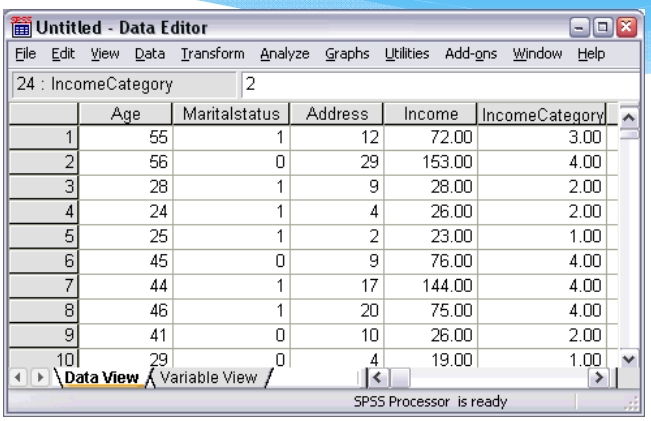
Importing data from Excel

- Open *demo.xls*.
- The Opening Excel Data Source dialog box is displayed, allowing you to specify whether variable names are to be included in the spreadsheet, as well as the cells that you want to import.



64

- Make sure first row is column heading
- It will be changed to variable name in SPSS



	Age	Maritalstatus	Address	Income	IncomeCategory
1	55	1	12	72.00	3.00
2	56	0	29	153.00	4.00
3	28	1	9	28.00	2.00
4	24	1	4	26.00	2.00
5	25	1	2	23.00	1.00
6	45	0	9	76.00	4.00
7	44	1	17	144.00	4.00
8	46	1	20	75.00	4.00
9	41	0	10	26.00	2.00
10	29	0	4	19.00	1.00

SPSS can also import from databases and text files – see notes for specific instructions

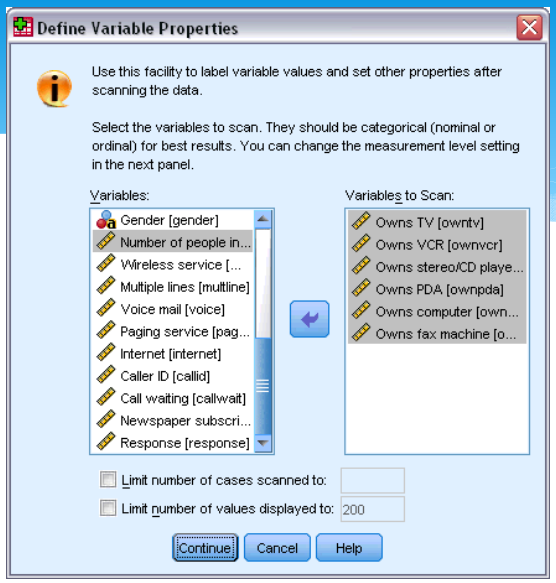
Defining Variable Properties for Categorical Variables

- For categorical (nominal, ordinal) data, you can use Define Variable Properties to define value labels and other variable properties.
- The Define Variable Properties process:
- Scans the actual data values and lists all unique data values for each selected variable.
- Identifies unlabeled values and provides an "auto-label" feature.
- Provides the ability to copy defined value labels from another variable to the selected variable or from the selected variable to additional variables.

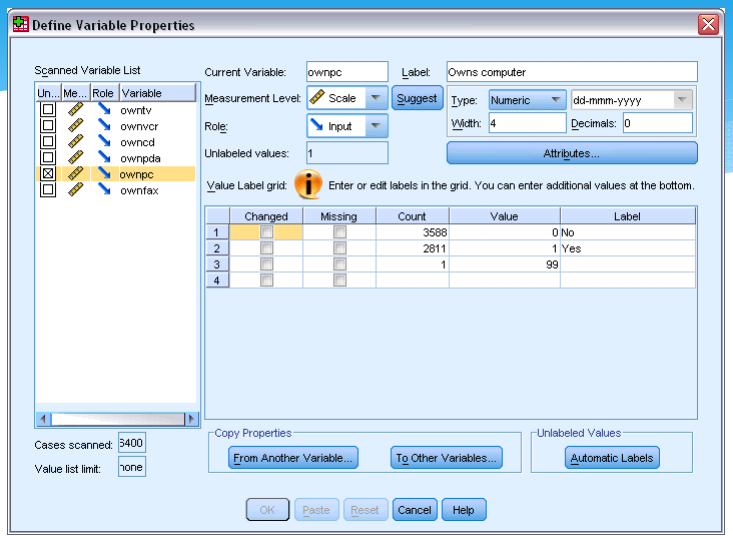
66

- This example uses the data file *demo.sav*.
- This data file already has defined value labels, so we will enter a value for which there is no defined value label.
- In Data View of the Data Editor, click the first data cell for the variable *ownpc* (you may have to
- scroll to the right), and then enter 99.
- From the menus choose:
- Data > Define Variable Properties..

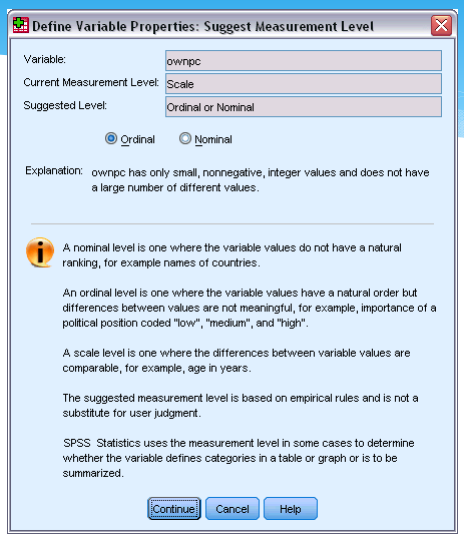
67



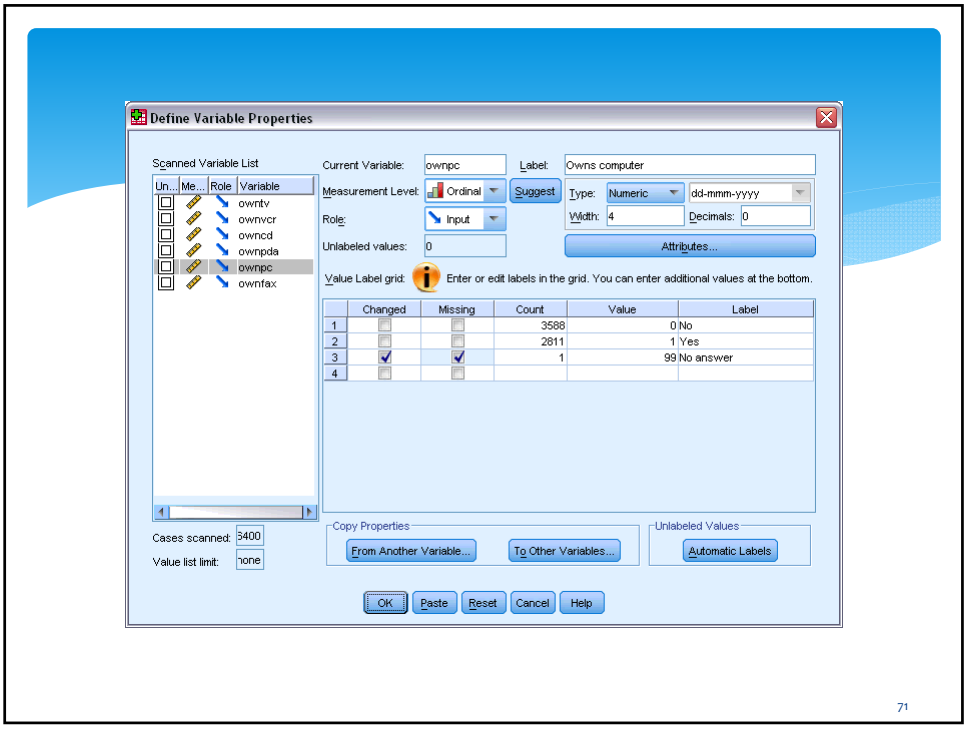
68

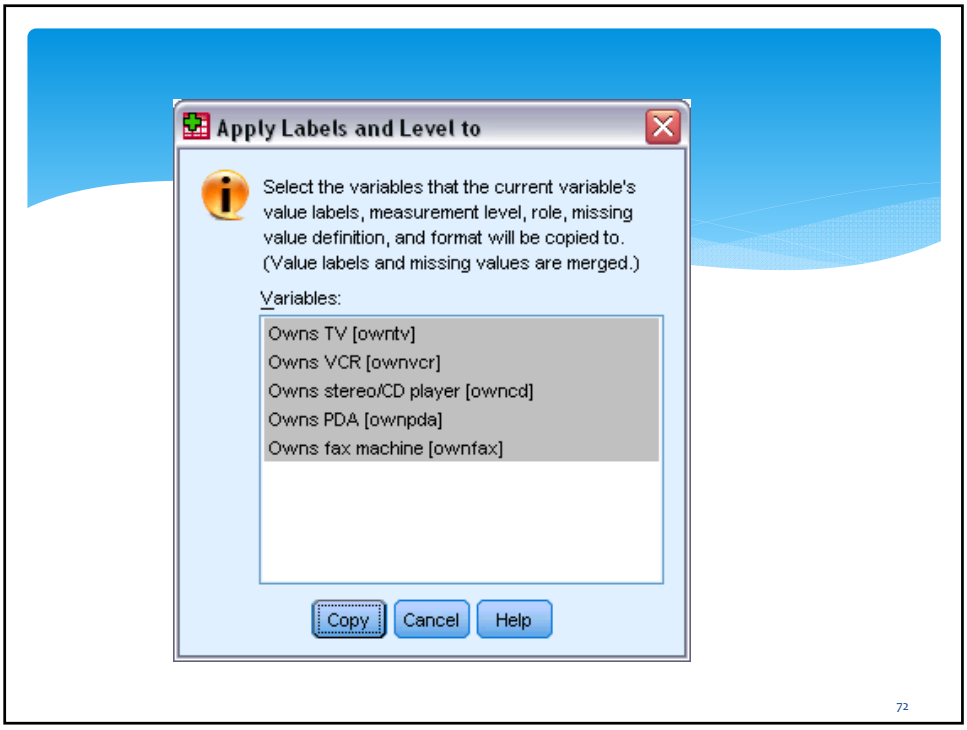

Define Variable Properties
Scanned Variable List
Un... Me... Role Variable
owntv
ownvcr
owncd
ownpda
ownpc
ownfax
Current Variable: ownpc
Label: Owns computer
Measurement Level: Scale
Suggest
Role: Input
Unlabeled values: 1
Type: Numeric
dd-mmm-yyyy
Width: 4
Decimals: 0
Attributes...
Value Label grid: Enter or edit labels in the grid. You can enter additional values at the bottom.
Changed Missing Count Value Label
1 3588 0 No
2 2811 1 Yes
3 1 99
4
Cases scanned: 3400
Value list limit: none
Copy Properties
From Another Variable...
To Other Variables...
Unlabeled Values
Automatic Labels
OK Paste Reset Cancel Help

69


Define Variable Properties: Suggest Measurement Level
Variable: ownpc
Current Measurement Level: Scale
Suggested Level: Ordinal or Nominal
Ordinal
Nominal
Explanation: ownpc has only small, nonnegative, integer values and does not have a large number of different values.
A nominal level is one where the variable values do not have a natural ranking, for example names of countries.
An ordinal level is one where the variable values have a natural order but differences between values are not meaningful, for example, importance of a political position coded "low", "medium", and "high".
A scale level is one where the differences between variable values are comparable, for example, age in years.
The suggested measurement level is based on empirical rules and is not a substitute for user judgment.
SPSS Statistics uses the measurement level in some cases to determine whether the variable defines categories in a table or graph or is to be summarized.
Continue Cancel Help

70


Define Variable Properties
Scanned Variable List
Un... Me... Role Variable
owntv
ownvcr
owncd
ownpda
ownpc
ownfax
Current Variable: ownpc
Label: Owns computer
Measurement Level: Ordinal
Suggest
Role: Input
Unlabeled values: 0
Type: Numeric
dd-mmm-yyyy
Width: 4
Decimals: 0
Attributes...
Value Label grid: Enter or edit labels in the grid. You can enter additional values at the bottom.
Changed Missing Count Value Label
1 3588 0 No
2 2811 1 Yes
3 1 99 No answer
4
Copy Properties
From Another Variable...
To Other Variables...
Unlabeled Values
Automatic Labels
Cases scanned: 3400
Value list limit: none
OK Paste Reset Cancel Help
71


Apply Labels and Level to
Select the variables that the current variable's value labels, measurement level, role, missing value definition, and format will be copied to. (Value labels and missing values are merged.)
Variables:
Owns TV [owntv]
Owns VCR [ownvcr]
Owns stereo/CD player [owncd]
Owns PDA [ownpda]
Owns fax machine [ownfax]
Copy Cancel Help
72

End of First Workshop