Introduction to Data Transformation

Ismail SA

Data Transformations

Two most commonly used data transformations in SPSS include:

1. COMPUTE
   Create new variable based on existing variable/s
2. RECODE
   Can be used to:
   a. Recategorize values
   b. Create categories based on metric (interval/ratio) variables

Compute

Compute

Create a new variable based on existing variables

No	New variable	Existing variable	# of item
1.	Attitude	A1 – A7	7
2.	QWL	Q1 – Q9	9

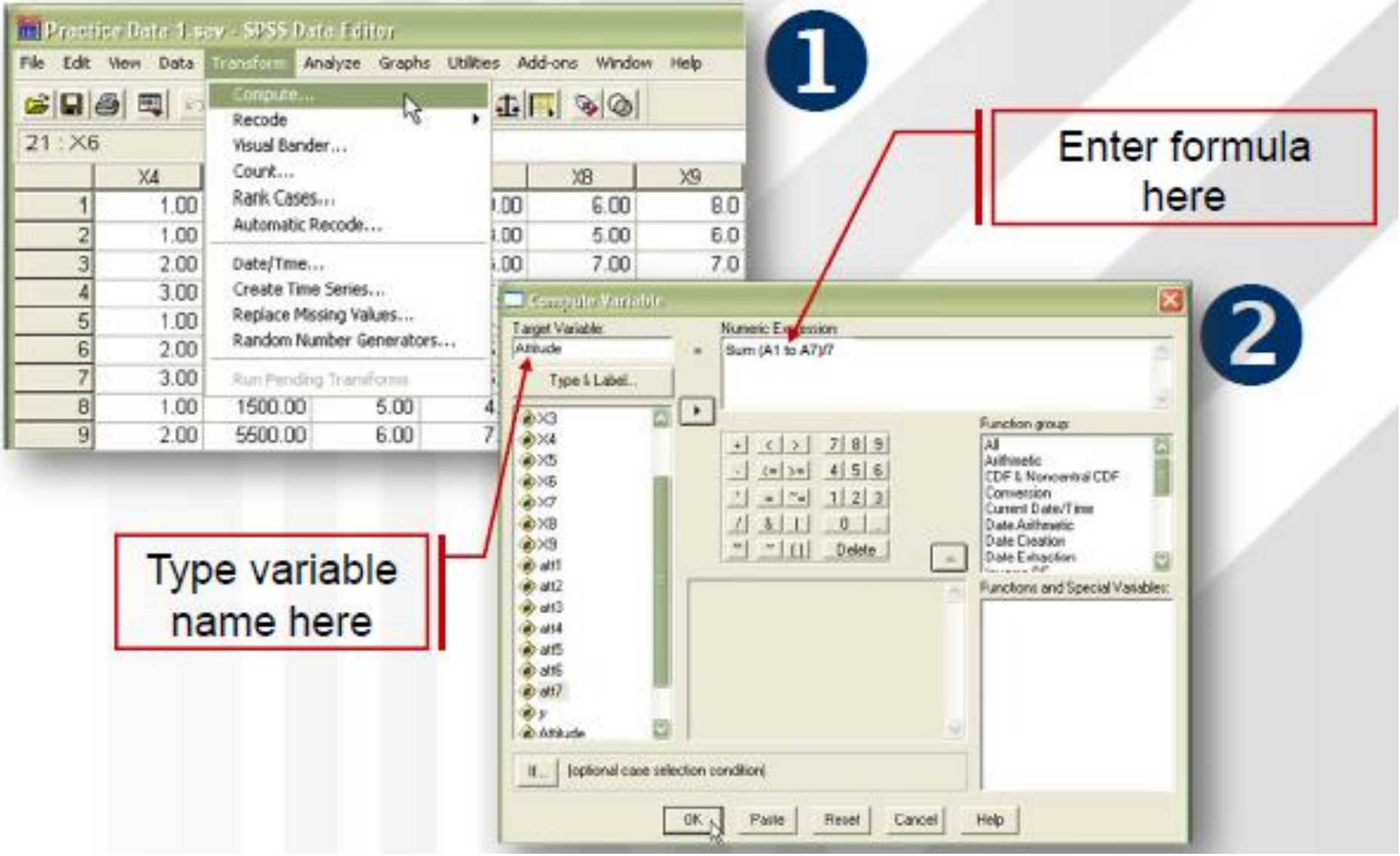
Compute an_income = X5 * 12

Compute Attitude = Mean (A1 to A7)

Variable name

Formula

COMPUTE Procedures:



Recode

Recode

Categorize scores into categories

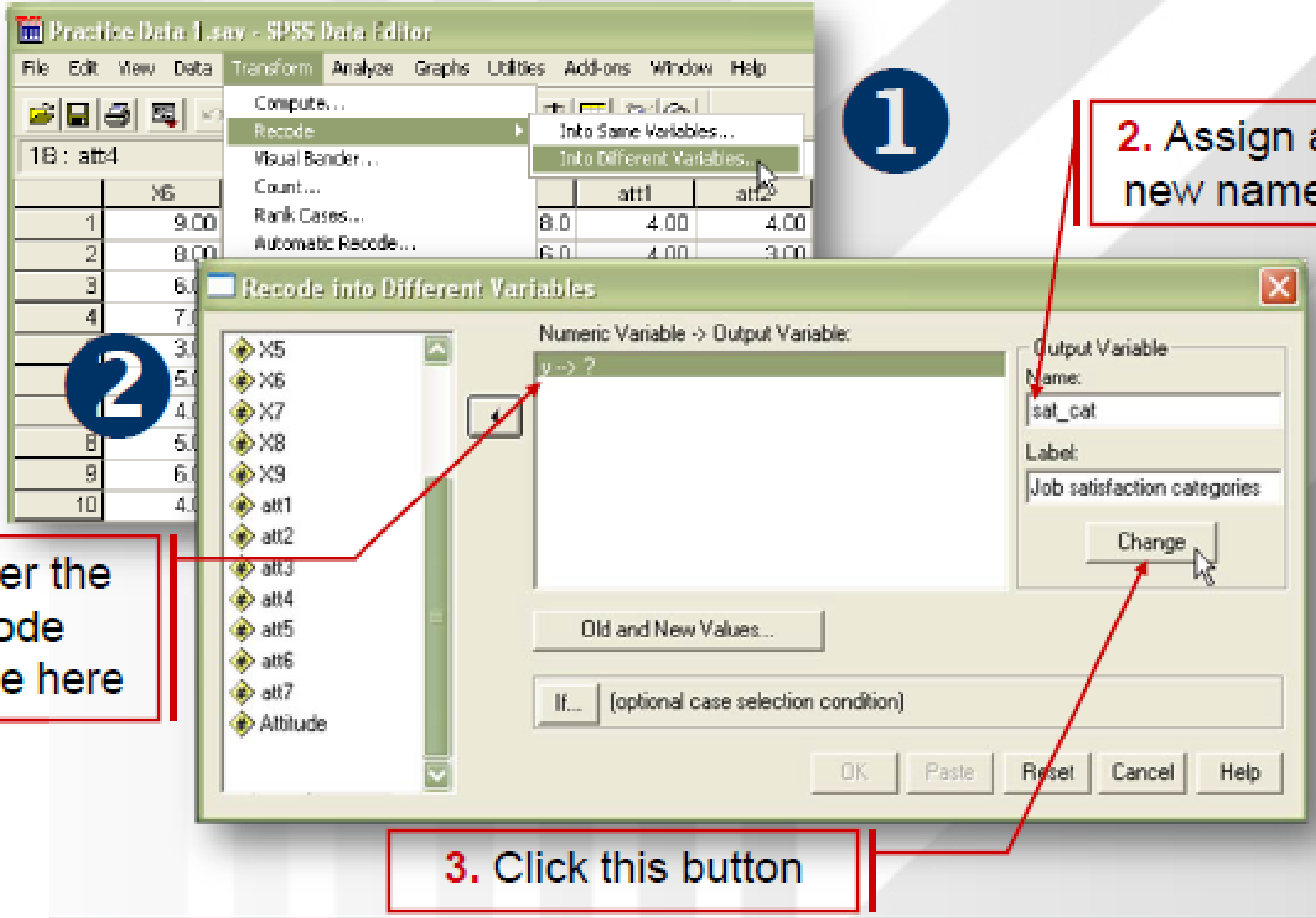
Ex. 1: Recode Y into Sat_cat

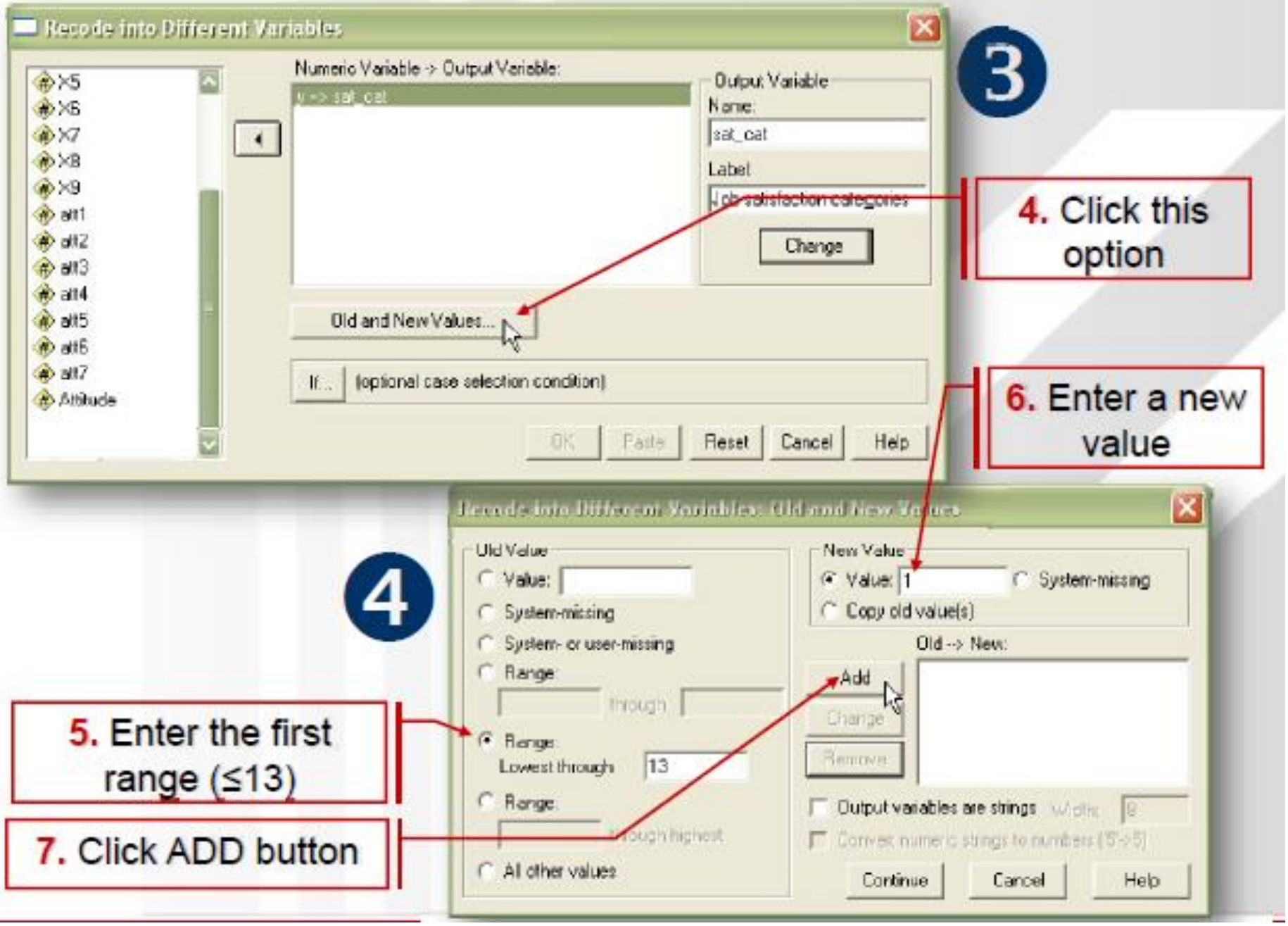
Category	Level	Range
1	Low	≤13
2	Moderate	14 – 16
3	High	>16

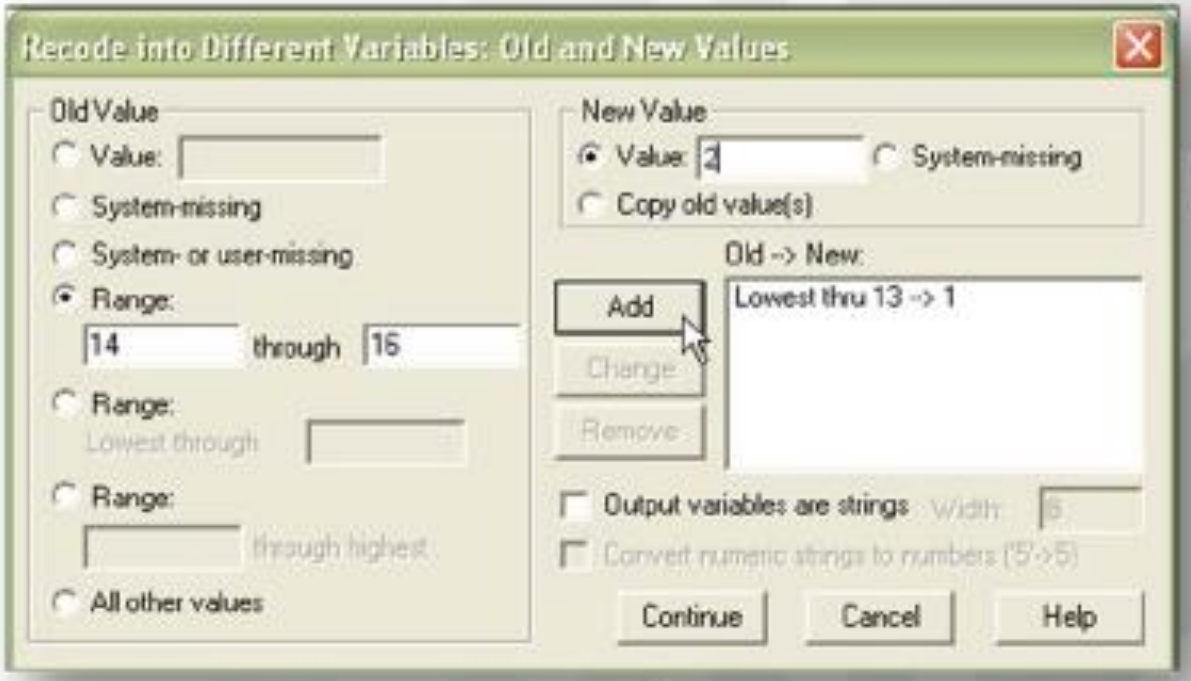
Ex. 2: Recode Attitude into Attitude_cat

Category	Level	Range
1	Low	1.00 – 2.33
2	Moderate	2.34 – 3.66
3	High	3.67 – 5.00

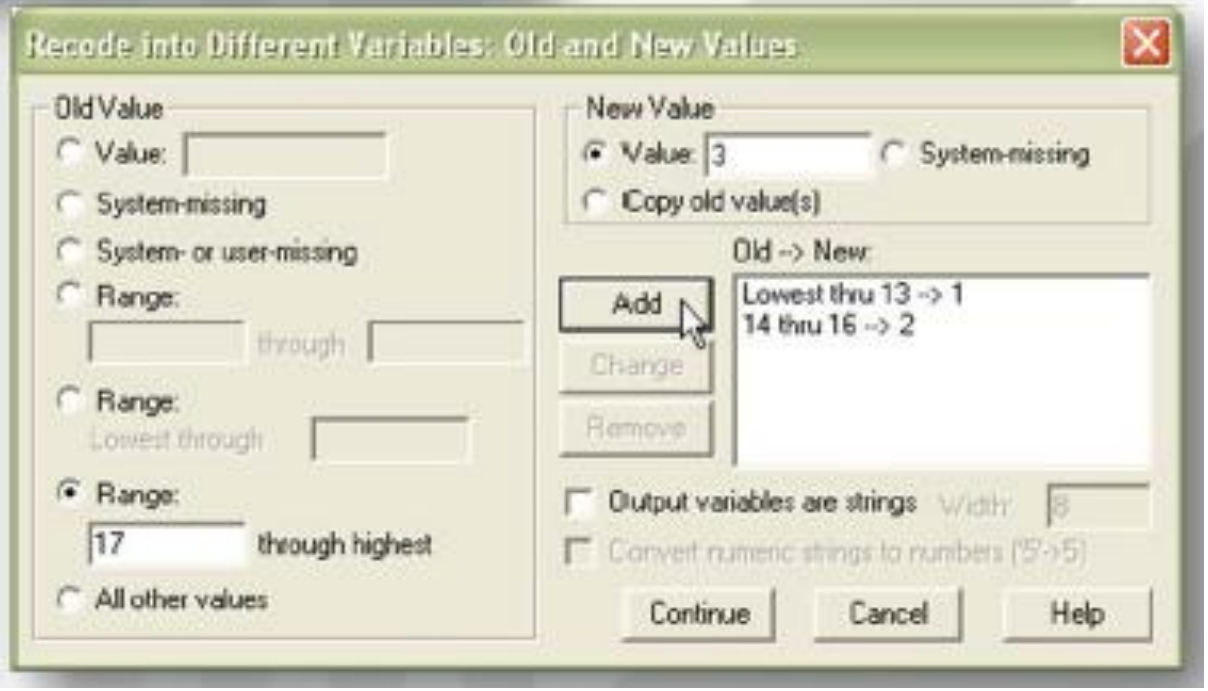
RECODE Procedures:



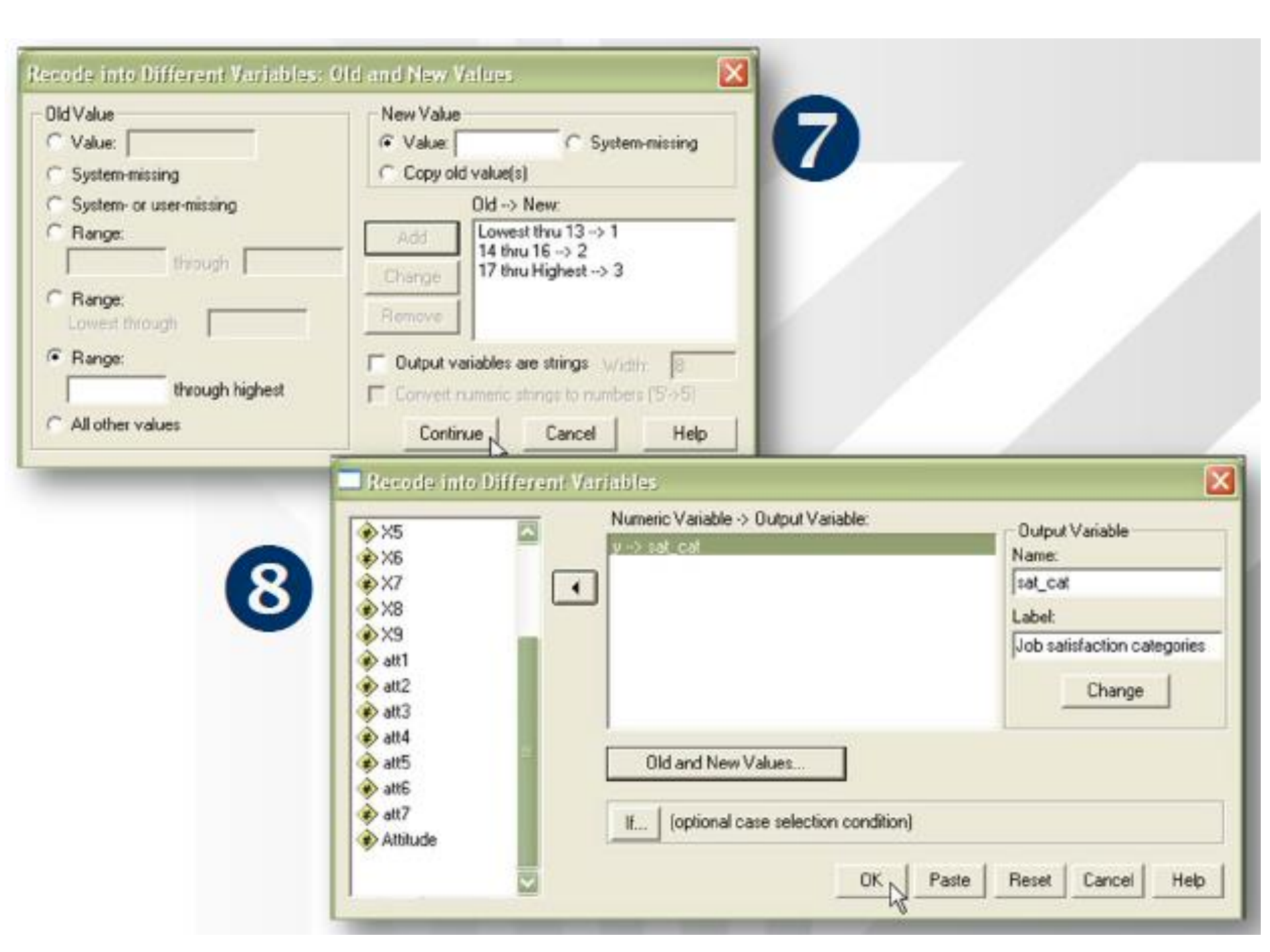

Recode into Different Variables
X5
X6
X7
X8
X9
alt1
alt2
alt3
alt4
alt5
alt6
alt7
Attitude
Numeric Variable -> Output Variable:
Output Variable
Name:
sat_cat
Label
Change
Old and New Values...
If...
(optional case selection condition)
OK
Paste
Reset
Cancel
Help
3
4. Click this option
6. Enter a new value
4
Old Value
Value:
System-missing
System- or user-missing
Range
through
Range:
Lowest through
13
Range:
through highest
All other values
New Value
Value:
1
System-missing
Copy old value(s)
Old --> New:
Add
Change
Remove
Output variables are strings
Width:
Continue
Cancel
Help
5. Enter the first range (≤13)
7. Click ADD button


Recode into Different Variables: Old and New Values
Old Value
Value:
System-missing
System- or user-missing
Range:
14
through
16
Range:
Lowest through
Range:
through highest
All other values
New Value
Value: 2
System-missing
Copy old value(s)
Old --> New:
Add
Change
Remove
Lowest thru 13 --> 1
Output variables are strings
Width:
Convert numeric strings to numbers ('5'->5)
Continue
Cancel
Help

5


Recode into Different Variables: Old and New Values
Old Value
Value:
System-missing
System- or user-missing
Range:
through
Range:
Lowest through
Range:
17
through highest
All other values
New Value
Value: 3
System-missing
Copy old value(s)
Old --> New:
Add
Change
Remove
Lowest thru 13 --> 1
14 thru 16 --> 2
Output variables are strings
Width:
Convert numeric strings to numbers ('5'->5)
Continue
Cancel
Help

6


Recode into Different Variables: Old and New Values
Old Value
Value:
System-missing
System- or user-missing
Range:
through
Range:
Lowest through
Range:
through highest
All other values
New Value
Value:
System-missing
Copy old value(s)
Old --> New:
Add
Change
Remove
Lowest thru 13 --> 1
14 thru 16 --> 2
17 thru Highest --> 3
Output variables are strings
Width:
Convert numeric strings to numbers ('5'->5)
Continue
Cancel
Help
7
Recode into Different Variables
X5
X6
X7
X8
X9
att1
att2
att3
att4
att5
att6
att7
Attitude
Numeric Variable -> Output Variable:
v -> sat_cat
Output Variable
Name:
sat_cat
Label:
Job satisfaction categories
Change
Old and New Values...
If...
(optional case selection condition)
OK
Paste
Reset
Cancel
Help
8

Exercise

Data Set 3:

The above data set comprises the following variables:

Variables	Item
Support from peer	S1 – S9
Work environment	W1 – W11
Motivation	M1 – M12
Job performance (Y)	J1 – J13

Question

1. Calculate the mean cumulative scores for each of the variables

   Assign the new variables as:
   - Support
   - Work
   - Motive
   - Perform

2. Categorize the above mean scores into three categories below:

1	Low	1.00 – 2.33
2	Moderate	2.34 – 3.66
3	High	3.67 – 5.00

Assign the new variables as:

- Support_cat
- Work_cat
- Motive_cat
- Perform_cat

3. Present the results in the following tables:

Table 1: Distribution of Peer Support and Work Environment Scores

Variable	Freq	%	Mean	SD
Peer support			___	___
Low (1.00 – 2.33)	___	___		
Moderate (2.34 – 3.66)	___	___		
High (3.67 – 5.00)	___	___		
Work environment			___	___
Low (1.00 – 2.33)	___	___		
Moderate (2.34 – 3.66)	___	___		
High (3.67 – 5.00)	___	___		

Table 2: Distribution of Motivation and Job Performance Scores

Variable	Freq	%	Mean	SD
Motivation			———	———
Low (1.00 – 2.33)	———	———		
Moderate (2.34 – 3.66)	———	———		
High (3.67 – 5.00)	———	———		
Job performance			———	———
Low (1.00 – 2.33)	———	———		
Moderate (2.34 – 3.66)	———	———		
High (3.67 – 5.00)	———	———		