

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:

1

Enter formula here

Type variable name here

The screenshot shows the SPSS Data Editor with a data table containing 9 rows and 6 columns. The columns are labeled X4, X8, and X9. The data values are as follows:

	X4	X8	X9
1	1.00	6.00	8.0
2	1.00	5.00	6.0
3	2.00	7.00	7.0
4	3.00		
5	1.00		
6	2.00		
7	3.00		
8	1.00	1500.00	5.00
9	2.00	5500.00	6.00

The 'Transform' menu is open, and the 'Compute...' option is selected. A red arrow points from the 'Type variable name here' box to the 'Target Variable:' field in the 'Compute Variable' dialog box. Another red arrow points from the 'Enter formula here' box to the 'Numeric Expression:' field in the same dialog box. The dialog box shows the formula 'Sum (A1 to A7)/7'.

2

OK Paste Reset Cancel Help

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:

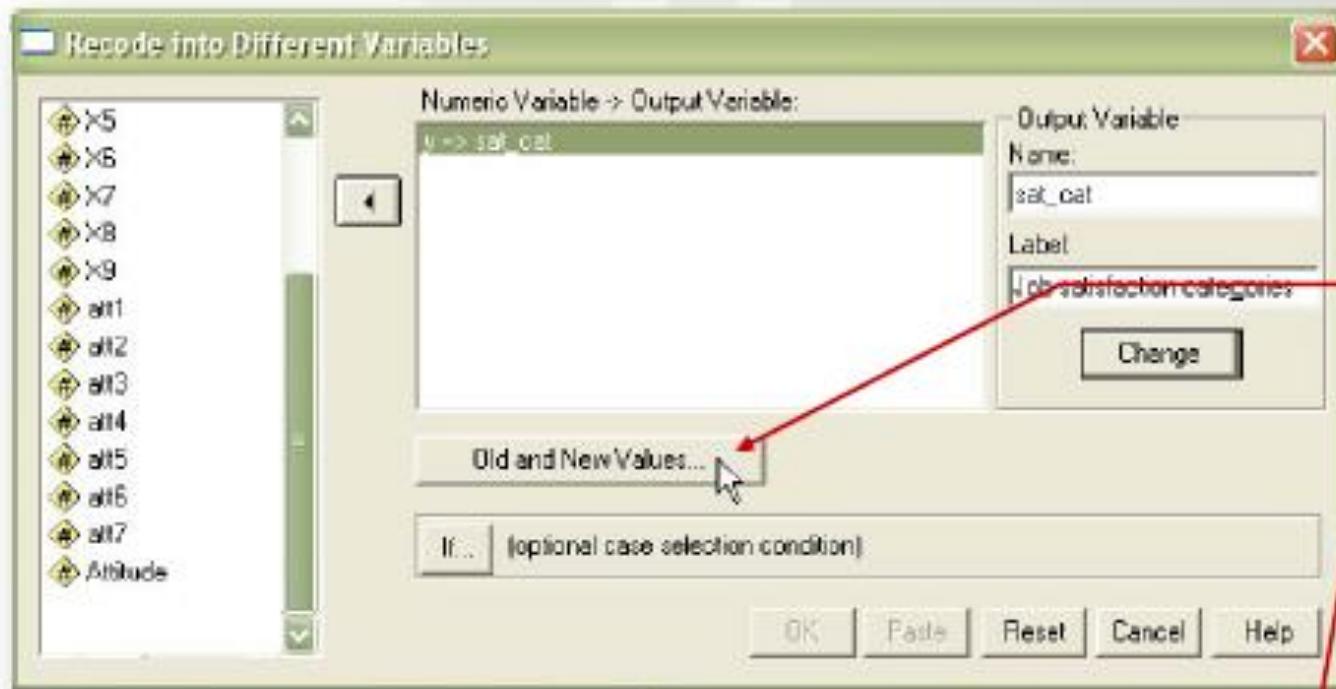
1

2. Assign a new name

1. Enter the recode variable here

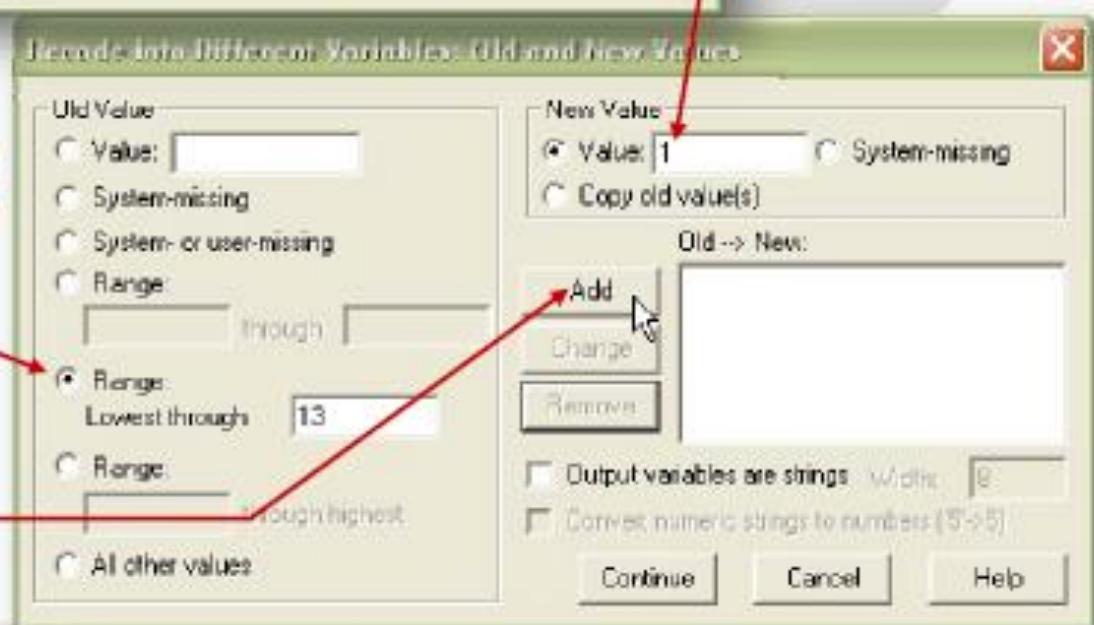
3. Click this button

The screenshot shows the SPSS Data Editor interface. The 'Transform' menu is open, with 'Recode' selected. A sub-menu shows 'Recode into Same Variables...' and 'Recode into Different Variables...', with the latter being the active choice. The main dialog box is titled 'Recode into Different Variables'. On the left, a list of variables includes X5, X6, X7, X8, X9, att1, att2, att3, att4, att5, att6, att7, and Altitude. In the center, under 'Numeric Variable > Output Variable:', there is a dropdown menu currently set to 'y > ?'. To the right, the 'Output Variable' section has 'Name:' set to 'sat_cat' and 'Label:' set to 'Job satisfaction categories'. Below these are 'Old and New Values...', 'If... (optional case selection condition)', and buttons for 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'. A large blue circle labeled '2' points to the 'sat_cat' entry field. A red arrow points from the text '1. Enter the recode variable here' to the 'sat_cat' field. Another red arrow points from the text '3. Click this button' to the 'Change' button.



3

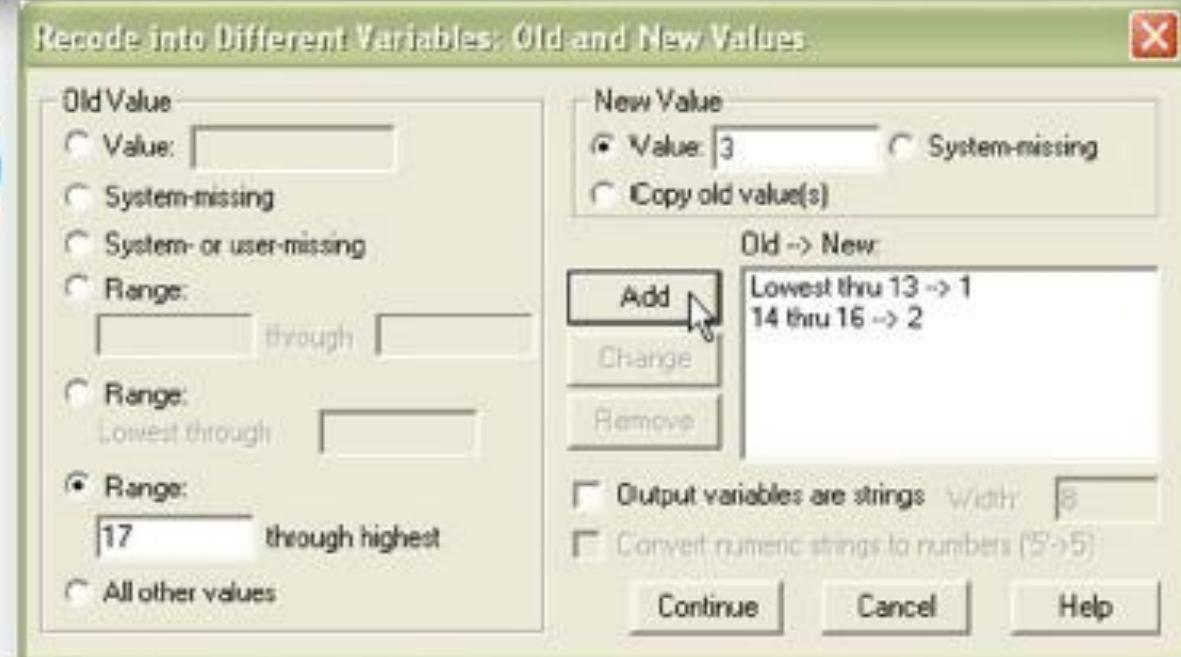
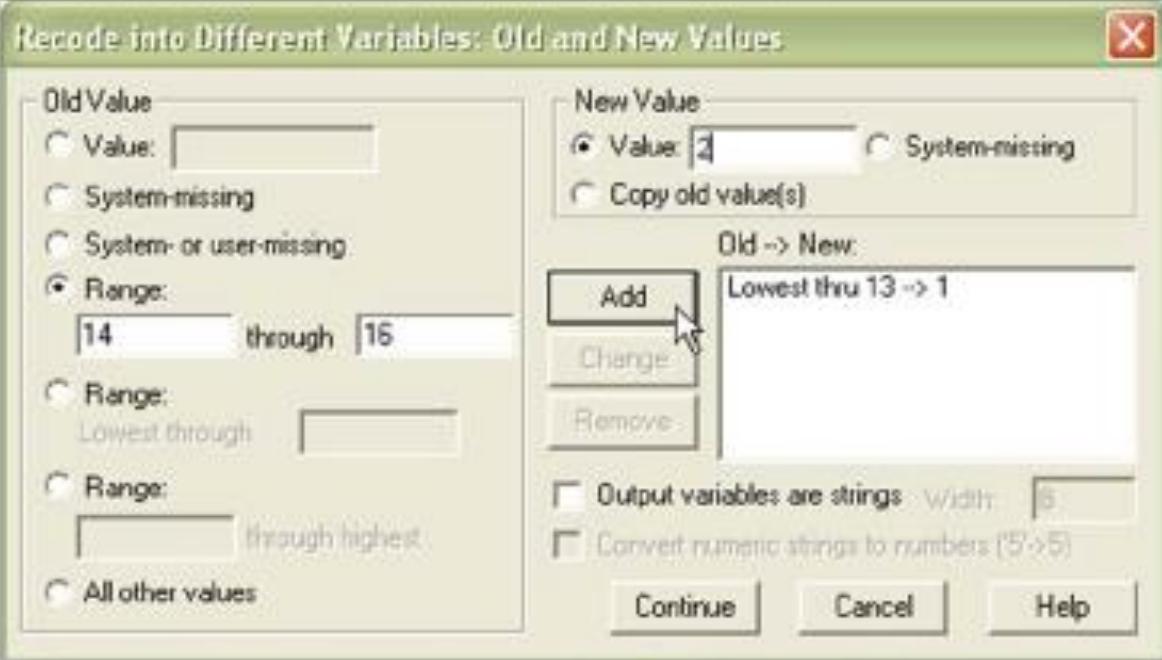
4. Click this option



4

5. Enter the first range (≤ 13)

7. Click ADD button



7

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:

Lowest thru 13 → 1
14 thru 16 → 2
17 thru Highest → 3

Add Change Remove

Output variables are strings: width:
 Convert numeric strings to numbers (5→5)

Continue Cancel Help

8

Recode into Different Variables

Numeric Variable → Output Variable:
v → sat_cat

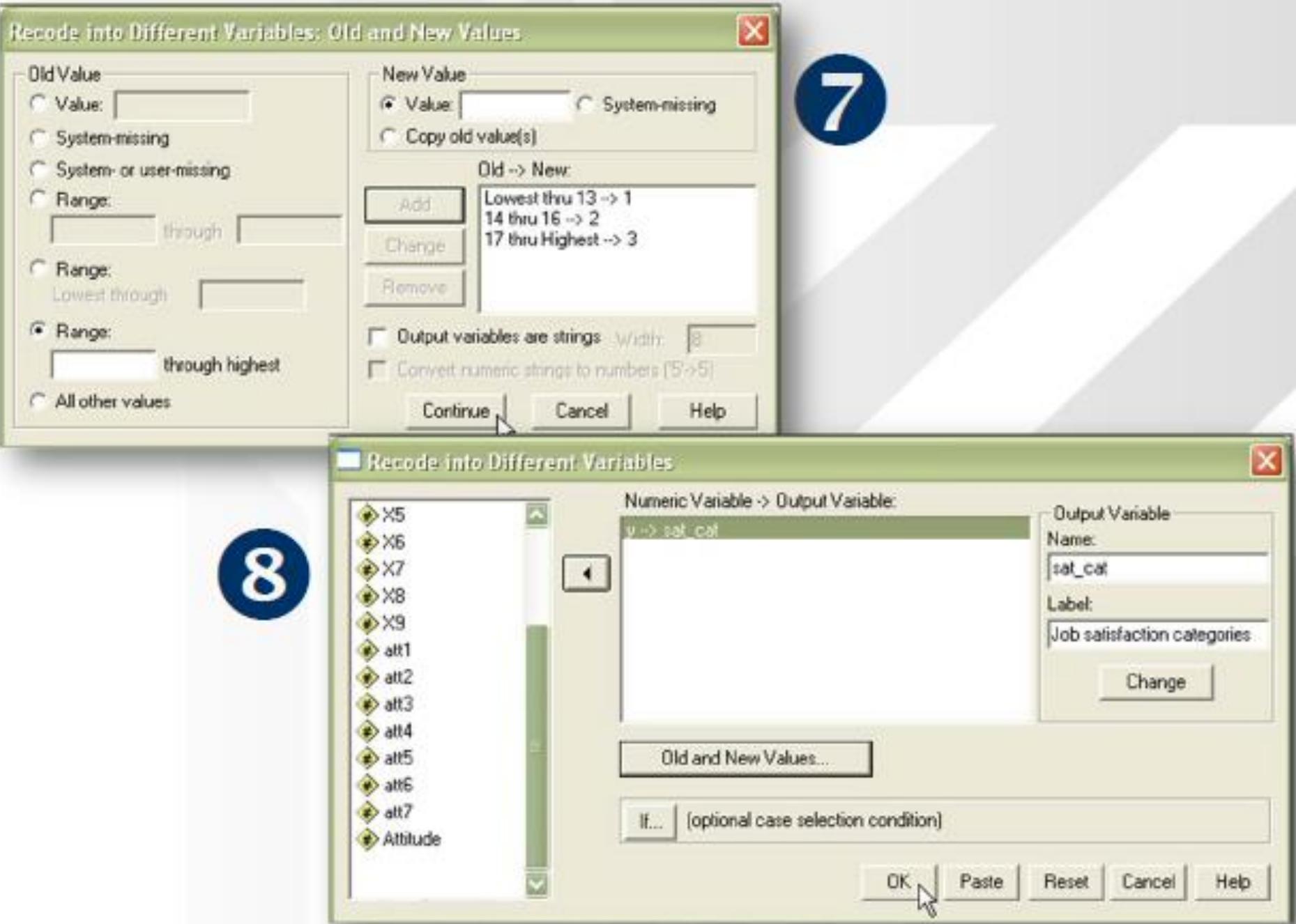
Output Variable

Name: sat_cat
Label: Job satisfaction categories
Change

Old and New Values...

If... (optional case selection condition)

OK Paste Reset Cancel Help



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)	—	—	—	—