

# SPSS FOR BEGINNERS

---

THAMARAT JANGSIRIWATTANA

## 2 WHAT IS IN THIS WORKSHOP

---

- SPSS interface: data view and variable view
- How to enter data in SPSS
- How to import external data into SPSS
- How to transform variables
- How to sort and select cases
- How to get descriptive statistics

# 3 SPSS INTERFACE

---

- Data view
  - The place to enter data
  - Columns: variables
  - Rows: records
- Variable view
  - The place to enter variables
  - List of all variables
  - Characteristics of all variables

# 4 ENTER DATA IN SPSS 19.0

The screenshot displays the IBM SPSS Statistics Data Editor interface. The window title is "Untitled2 [DataSet3] - IBM SPSS Statistics Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, and Help. The toolbar contains various icons for file operations and data manipulation. The main area is a grid with 20 columns and 36 rows. The first column is labeled "var" and contains row numbers from 1 to 36. The other columns are also labeled "var". Three red callout boxes with arrows point to specific parts of the grid: "Columns: variables" points to the first column, "Rows: cases" points to the first row, and "Under Data View" points to the "Data View" button at the bottom left. The status bar at the bottom right indicates "IBM SPSS Statistics Processor is ready".

Columns: variables

Rows: cases

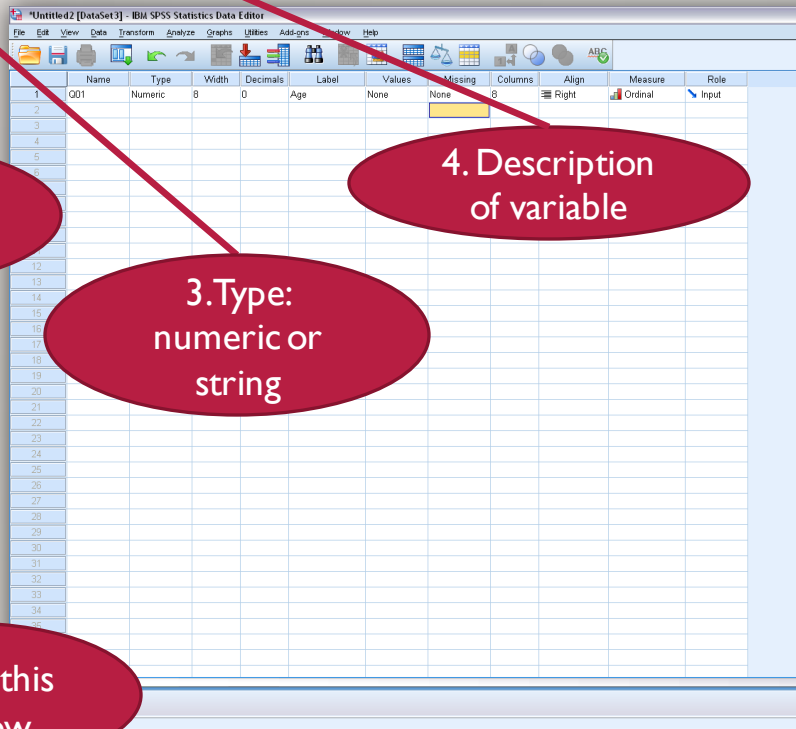
Under Data View

Data View Variable View

Visible: 0 of 0 Variables

IBM SPSS Statistics Processor is ready

# 5 ENTER VARIABLES



1. Click Variable View
2. Type variable name under Name column (e.g. Q01).  
**NOTE:** Variable name can be 64 bytes long, and the first character must be a letter or one of the characters @, #, or \$.
3. Type: Numeric, string, etc.
4. Label: description of variables.

2. Type variable name

3. Type: numeric or string

4. Description of variable

1. Click this Window

# 6 ENTER VARIABLES

Based on your code book!

The image shows a sequence of steps in SPSS. In the background, the 'Data Editor' window displays a data table with columns: Name, Type, Width, Decimals, Label, and Values. The first row contains: 001, Numeric, 8, 0, Age, None, Nor. A red arrow points from the 'Label' column to the 'Value Labels' dialog box.

The first 'Value Labels' dialog box is empty, with 'Value:' and 'Label:' fields. A 'Spelling...' button is visible.

The second 'Value Labels' dialog box shows the 'Value:' field containing '3' and the 'Label:' field containing '14 years'. Below these fields is a list of defined labels: '1 = "12 years old or young"' and '2 = "13 years"', with the second item highlighted. Buttons for 'Add', 'Change', and 'Remove' are on the left. 'OK', 'Cancel', and 'Help' buttons are at the bottom.

# 7 ENTER CASES

The screenshot shows the IBM SPSS Statistics Data Editor interface. The main window displays a data grid with two columns: 'Code' and 'Q01'. The data is as follows:

	Code	Q01	var	var	var	var	var	var	var	var	var	var	var	var	var	var	var	var	var
1	1001	1																	
2	1002	2																	
3	1003	2																	
4	1004	3																	
5	1005	2																	
6																			
7																			
8																			
9																			
10																			
11																			
12																			
13																			
14																			
15																			
16																			
17																			
18																			
19																			
20																			
21																			
22																			
23																			
24																			
25																			
26																			
27																			
28																			
29																			
30																			
31																			
32																			
33																			
34																			
35																			
36																			
37																			

The interface includes a menu bar (File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, Help) and a toolbar with various icons. The status bar at the bottom indicates 'Data View' and 'Variable View' tabs, with 'Data View' selected. A red oval points to the 'Data View' tab with the text 'Under Data View'. A red box contains a list of four points explaining the variables. A red arrow points from the red box to the 'Q01' column header.

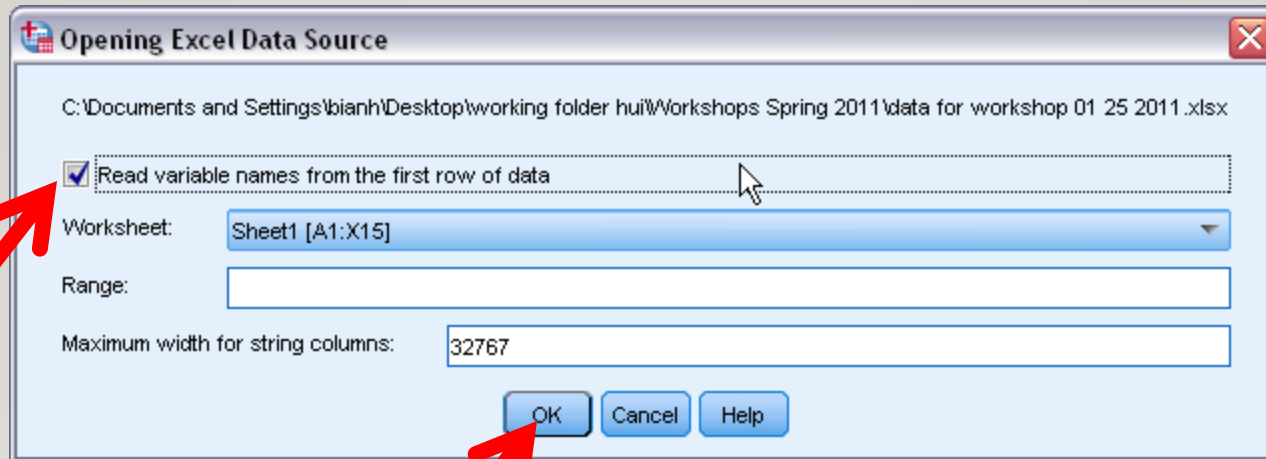
1. Two variables in the data set.
2. They are: Code and Q01.
3. Code is an ID variable, used to identify individual case (NOT people's real IDs).
4. Q01 is about participants' ages: 1 = 12 years or younger, 2 = 13 years, 3 = 14 years...

Under Data View

IBM SPSS Statistics Processor is ready

## 8 IMPORT DATA FROM EXCEL

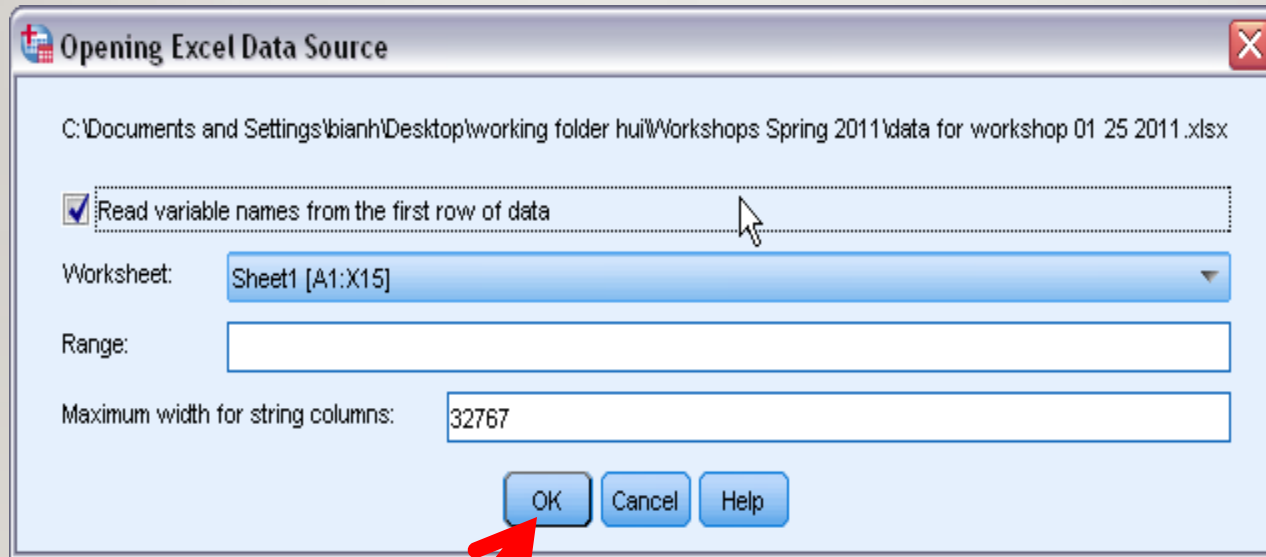
- Select File → Open → Data →
- Choose **Excel** as file type
- Select the file you want to import
- Then click Open






# 9 OPEN EXCEL FILES IN SPSS

---



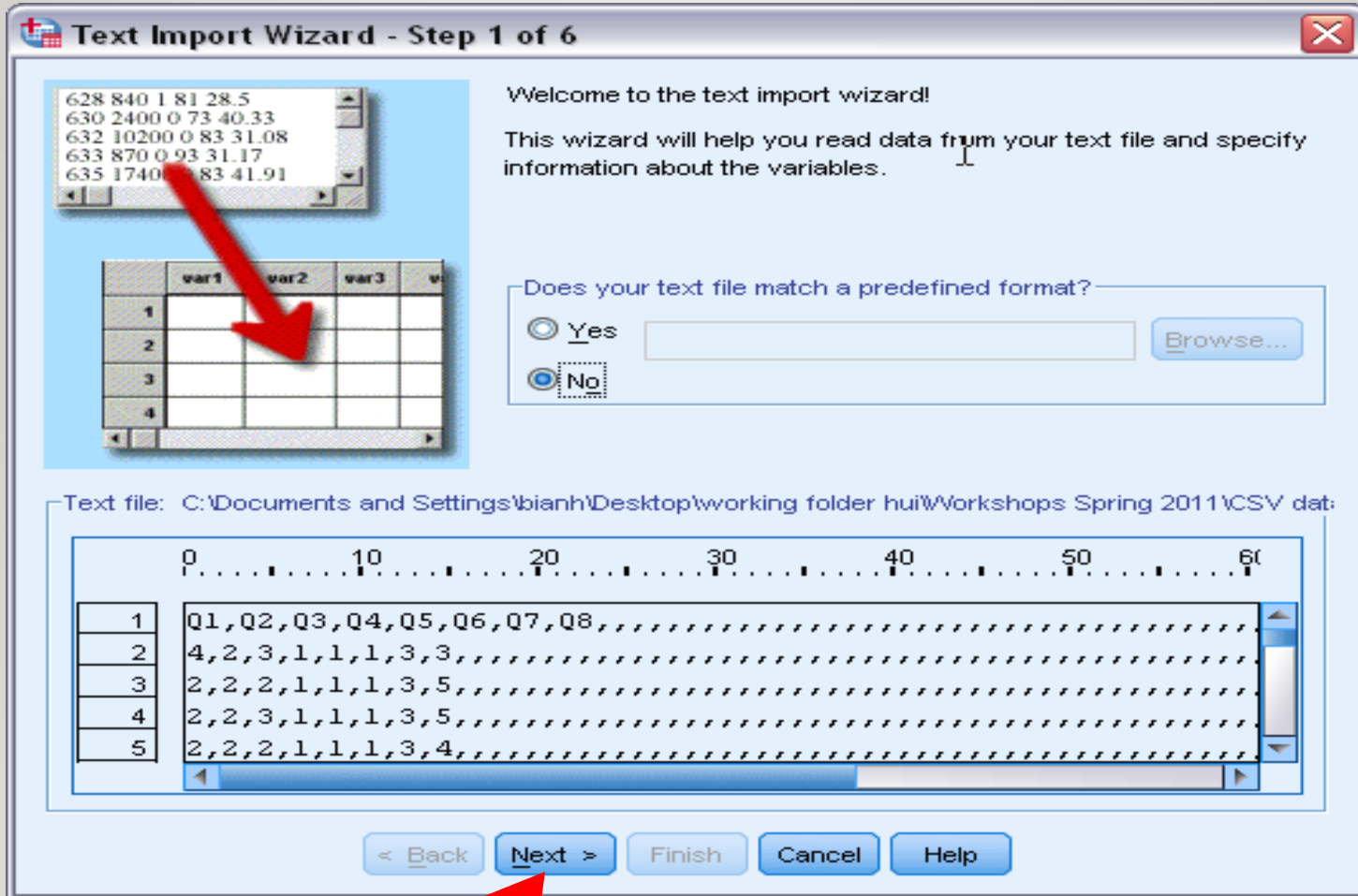
# 10 IMPORT DATA FROM CVS FILE

---

- CVS is a comma-separated values file.
- If you use Qualtrics to collect data (online survey), you will get a ~~CVS data~~ file. 
- Select File                      Open                      Data
- Choose **All files** as file type
- Select the file you want to import
- Then click Open

11

## CONTINUE



# 12 CONTINUE

**Text Import Wizard - Step 2 of 6**

How are your variables arranged?

Delimited - Variables are delimited by a specific character (i.e., comma, tab).  
 Fixed width - Variables are aligned in fixed width columns.

Are variable names included at the top of your file?

Yes  
 No

Text file: C:\Documents and Settings\bianh\Desktop\working folder hui\Workshops Spring 2011\CSV data

1	4,2,3,1,1,1,3,3,,,,,,
2	2,2,2,1,1,1,3,5,,,,,,
3	2,2,3,1,1,1,3,5,,,,,,
4	2,2,2,1,1,1,3,4,,,,,,
5	1,2,1,1,1,1,1,5,,,,,,

< Back   **Next >**   Finish   Cancel   Help

# 13 CONTINUE

The first case of data begins on which line number? 2

How are your cases represented?

Each line represents a case

A specific number of variables represents a case: 80

How many cases do you want to import?

All of the cases

The first 1000 cases.

A random percentage of the cases (approximate): 10 %

Data preview

1	4,2,3,1,1,1,3,3
2	2,2,2,1,1,1,3,5
3	2,2,3,1,1,1,3,5

< Back Next > Finish Cancel Help



# 14 CONTINUE



Text Import Wizard - Delimited Step 4 of 6

Which delimiters appear between variables?

Tab  Space  
 Comma  Semicolon  
 Other:

What is the text qualifier?

None  
 Single quote  
 Double quote  
 Other:

Data preview

Q1	Q2	Q3	Q4	Q5	Q6	Q7
1	2	1	2	1	1	1
1	2	1	1	1	1	1
1	2	1	1	1	1	1
2	2	1	1	1	1	1
4	1	2	1	1		3
1	2	1	1	1		1
3	2	5	1	1		3
2	2	2	1	1		1

< Back Next > Finish Cancel Help



# 15 CONTINUE

Text Import Wizard - Step 5 of 6

Specifications for variable(s) selected in the data preview

Variable name:  Original Name: Q1

Data format:

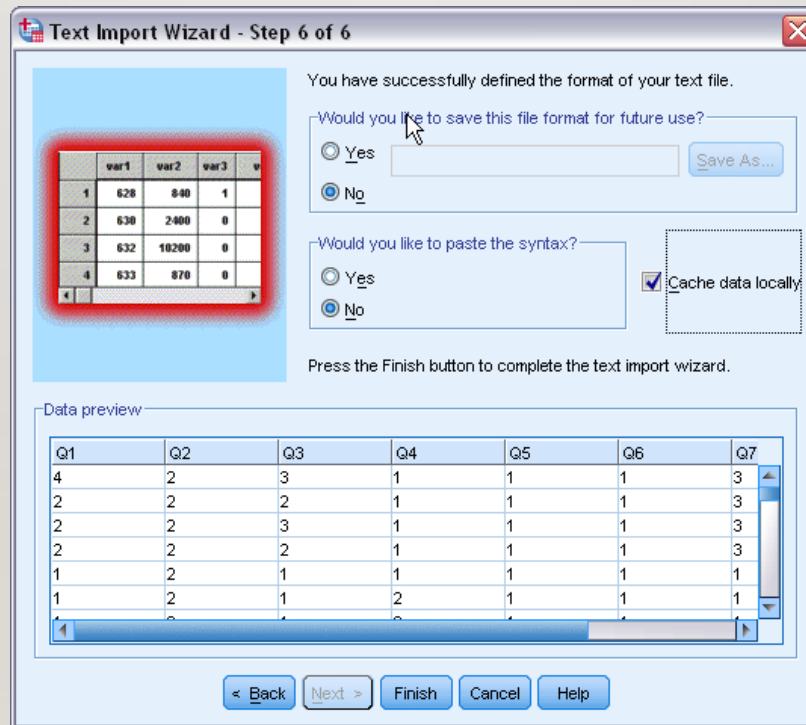
Data preview

Q1	Q2	Q3	Q4	Q5	Q6	Q7
4	2	3	1	1	1	3
2	2	2	1	1	1	3
2	2	3	1	1	1	3
2	2	2	1	1	1	3
1	2	1	1	1	1	1

< Back Next > Finish Cancel Help



# 16 CONTINUE





# 17 CONTINUE

13:04 | 1 | Visible: 0 of 0 Variables

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	var	var	var	var	var	var	var	var	var	var
1			3		1	1	1	3	3									
2	Female		2	1	1	1	1	3	5									
3	Female		3	1	1	1	1	3	5									
4	Male		2	1	1	1	1	3	4									
5	Male		1	1	1	1	1	1	5									
6	Female		1	2	1	1	1	1	1									
7	Male		1	2	1	1	1	1	5									
8	Female		1	1	1	1	1	1	3									
9	Female		1	1	1	1	1	1	5									
10	Female		1	1	1	1	1	1	1									
11	Female		2	1	1	1	1	3	5									
12	Male		1	1	1	1	1	1	3									
13	Male		5	1	1	1	1	3	5									
14	Male		2	1	1	1	1	1	5									
15	Male		1	2	1	1	1	1	1									
16	Female		1	1	3	1	1	1	4									
17	Male		1	1	1	1	1	3	3									
18	Female		1	1	1	1	1	1	3									
19	Female		1	1	1	1	1	1	3									
20	Male		2	1	1	1	1	3	1									
21	Female		1	1	1	1	1	3	5									
22	Female		1	1	1	1	1	1	4									
23	Female		2	1	1	1	1	3	5									
24	Male		1	1	1	1	1	3	1									
25	Male		1	1	1	1	1	1	3									
26	Female		1	2	1	1	1	1	3									
27	Male		1	3	1	1	1	1	5									
28	Female		2	1	3	1	1	3	5									
29	Male		2	1	2	1	1	3	5									
30	Male		4	1	1	1	1	3	1									
31	Male		1	1	1	1	1	1	5									
32	Female		1	1	1	1	1	1	3									
33	Male		1	1	1	1	1	1	4									
34	Female		1	3	1	1	1	1	2									
35	Male		1	1	1	1	1	1	5									
36	Male		2	1	1	1	1	3	2									
37	Male		3	1	1	1	1	3	2									

Data View | Variable View | IBM SPSS Statistics Processor is ready

Save this file as SPSS data

18

## CLEAN DATA AFTER IMPORT DATA FILES

---

- Key in values and labels for each variable
- Run frequency for each variable
- Check outputs to see if you have variables with wrong values.
- Check missing values and physical surveys if you use paper surveys, and make sure they are real missing.
- Sometimes, you need to recode string variables into numeric variables

# 19 CONTINUE

Q1 How old are you

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 12 years old or younger	19	.1	.1	.1
	2 13 years old	17	.1	.1	.2
	3 14 years old	1851	11.3	11.3	11.5
	4 15 years old	4045	24.7	24.7	36.3
	5 16 years old	4234	25.8	25.9	62.2
	6 17 years old	3963	24.2	24.2	86.4
	7 18 years old or older	2215	13.5	13.6	100.0
	8	1	.0	.0	100.0
	9	1	.0	.0	100.0
	Total	16344	99.6	100.0	
Missing	System	66	.4		
Total		16410	100.0		

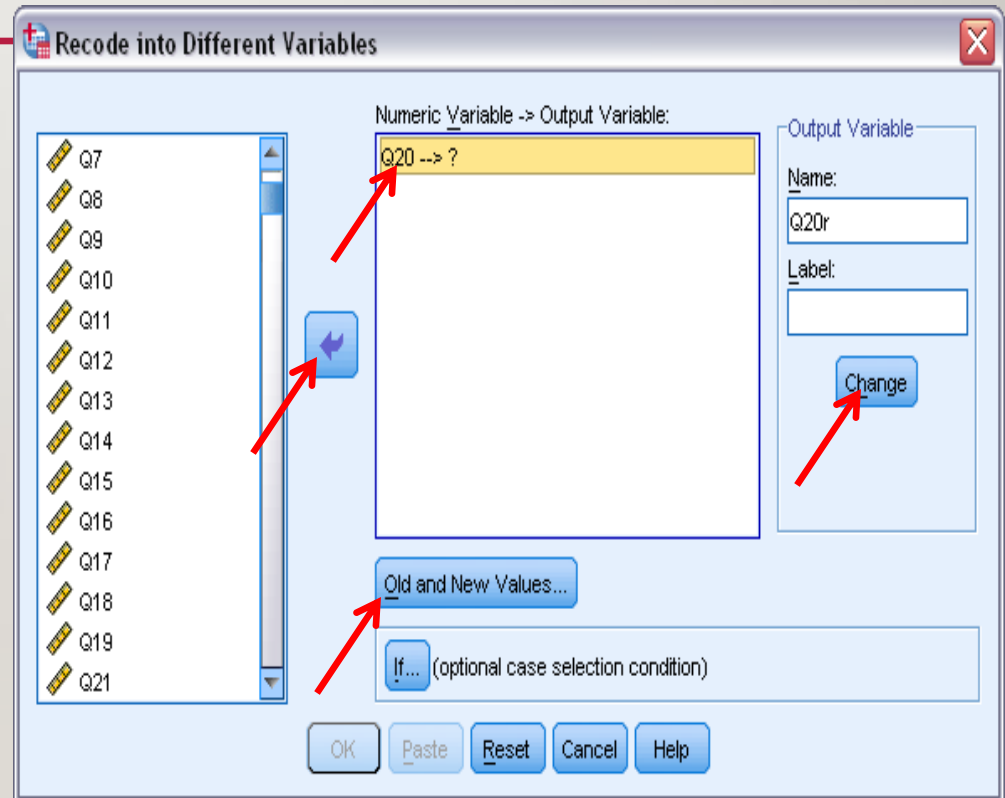


Wrong entries

## 20 VARIABLE TRANSFORMATION

- Recode variables

1. Select Transform → Recode into Different Variables
2. Select variable that you want to transform (e.g. Q20): we want 1= Yes and 0 = No
3. Click Arrow button to put your variable into the right window
4. Under Output Variable: type name for new variable and label, then click Change
5. Click Old and New Values



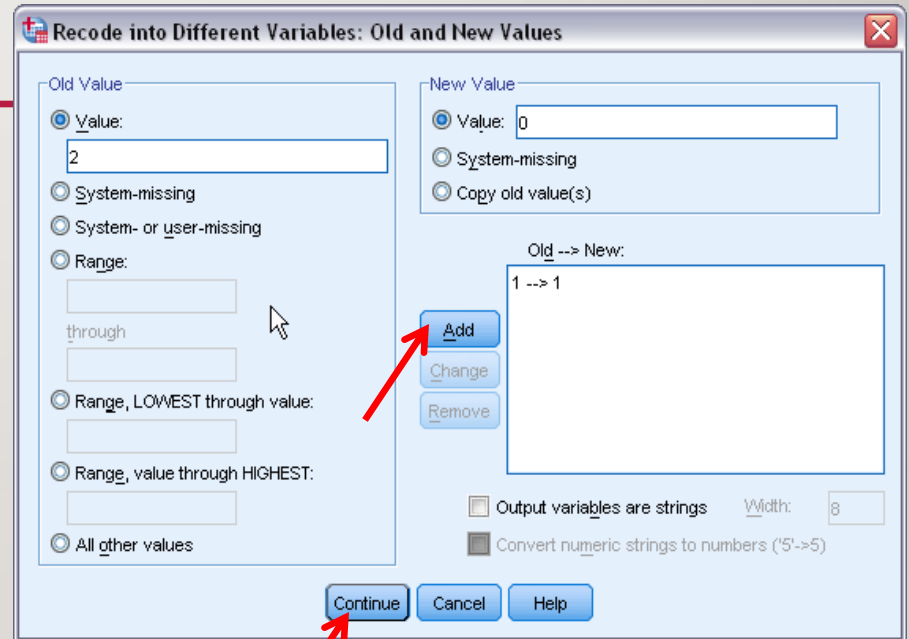
21

# CONTINUE

6. Type 1 under Old Value and 1 under New Value, click Add. Then type 2 under Old Value, and 0 under New Value, click Add.

7. Click Continue after finish all the changes.

8. Click Ok



# VARIABLE TRANSFORMATION

---

- Compute variable (use YRBSS 2009 data)
  - Example 1. Create a new variable: drug\_use (During the past 30 days, any use of cigarettes, alcohol, and marijuana is defined as use, else as non-use). There are two categories for the new variable (use vs. non-use). Coding: 1 = Use and 0 = Non-use
    1. Use Q30, Q41, and Q47 from 2009 YRBSS survey
    2. Non-users means those who answered **0** days/times to all three ~~questions~~ questions.
    3. Go to Transform            Compute Variable

## 23 CONTINUE

4. Type “drug\_use” under

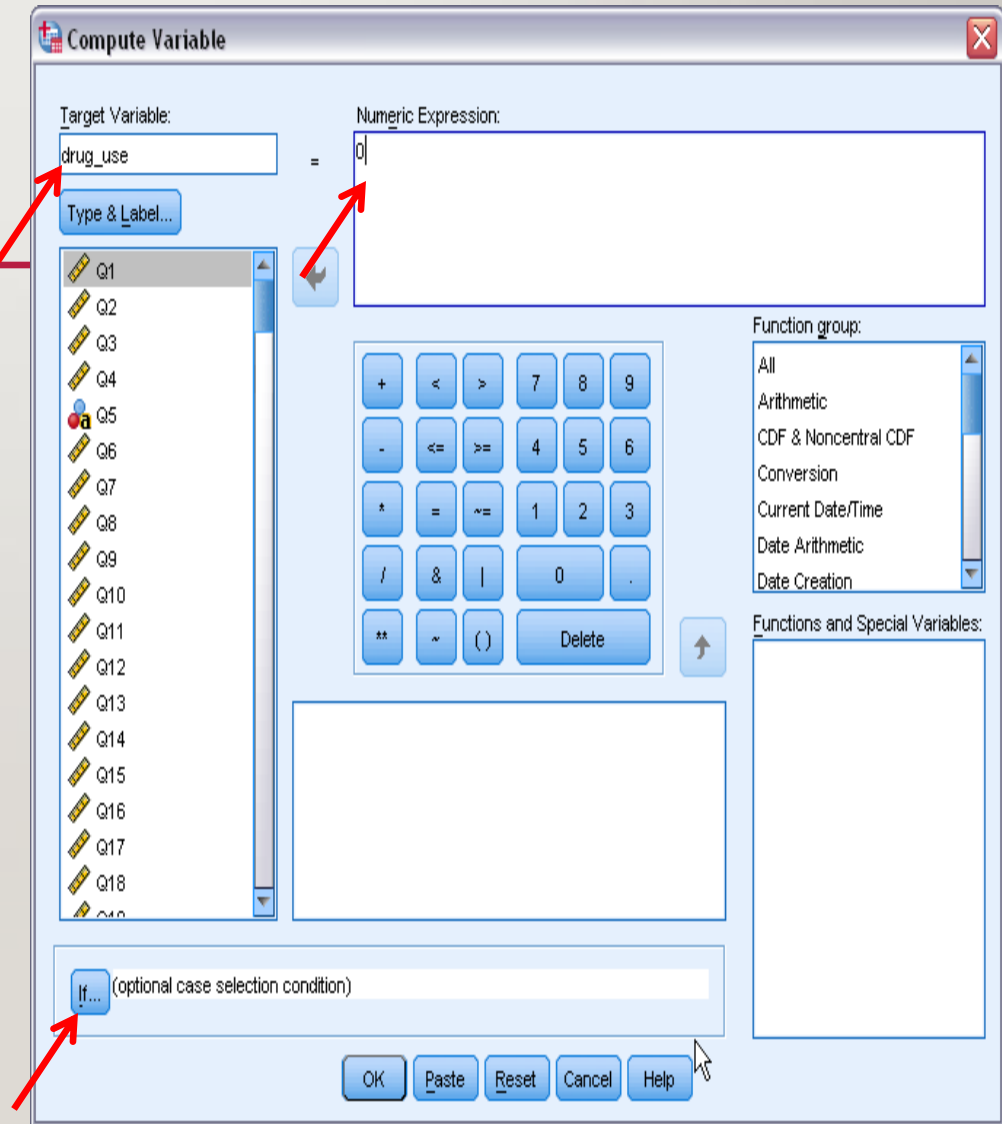
Target Variable

5. Type “0” under Numeric

Expression. 0 means

Non-use

6. Click If button.

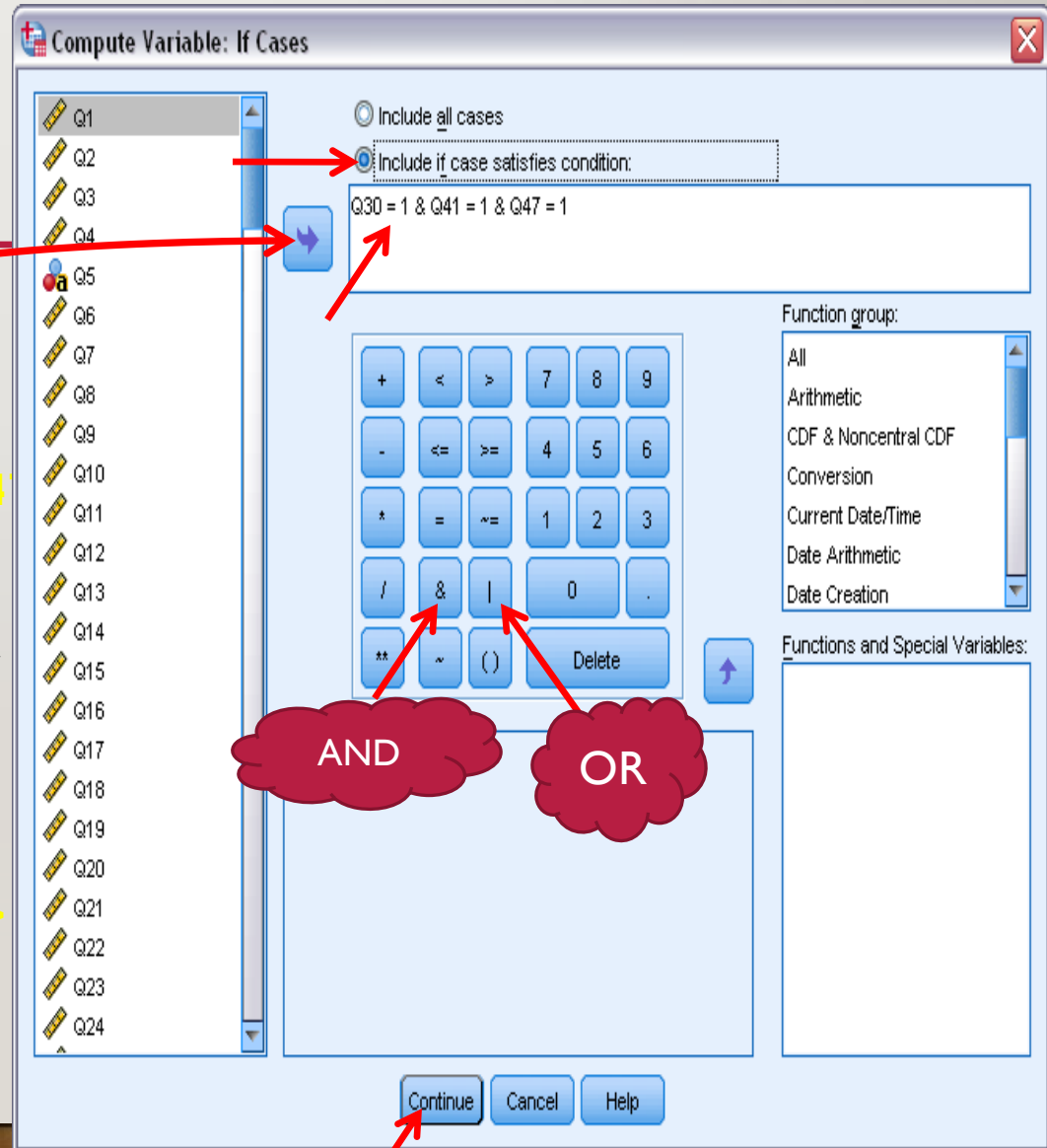


24

# CONTINUE

7. With help of that  
Arrow button, type  
**Q30 = 1 & Q41 = 1 & Q47**  
then click Continue

8. Do the same thing for  
Use, but the numeric  
expression is different:  
**Q30 > 1 | Q41 > 1 | Q47 >**





25

# CONTINU

9. Click OK

10. After click OK,

a small window asks

if you want to

change existing

variable because

drug\_use was already

created when you

first define non-use.

11. Click ok.

The screenshot shows the 'Compute Variable' dialog box in IBM SPSS Statistics 19. The 'Target Variable' is 'drug\_use' and the 'Numeric Expression' is '1'. A list of variables (Q1-Q18) is visible on the left. A calculator keypad is in the center. A 'Function group' list is on the right. At the bottom, there is an 'If...' button and a text field containing 'Q30 > 1 | Q41 > 1 | Q47 > 1'. A smaller dialog box titled 'IBM SPSS Statistics 19' with a question mark icon asks 'Change existing variable?' with 'OK' and 'Cancel' buttons. Red arrows point from the text instructions to the 'OK' button in the warning dialog, the 'If...' button, and the 'Numeric Expression' field.

## 26 CONTINUE

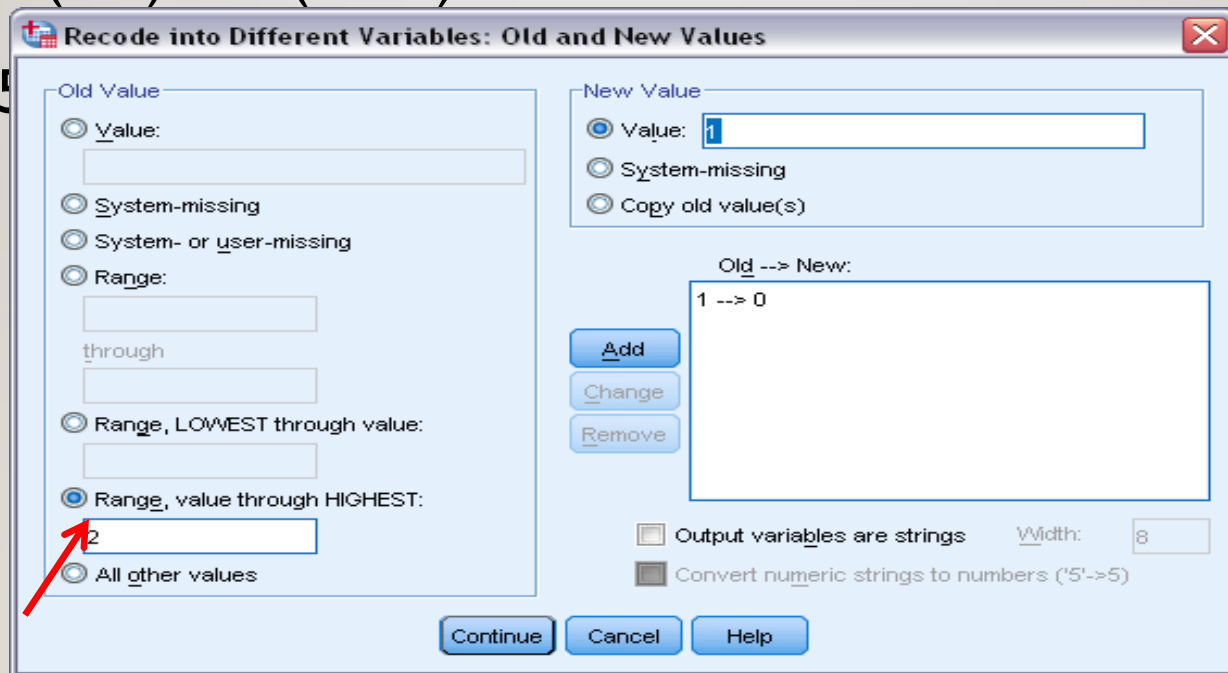
---

- Compute variables
  - Example 2. Create a new variable drug\_N that assesses total number of drugs that adolescents used during the last 30 days.
    1. Use Q30 (cigarettes), 41 (alcohol), 47 (marijuana), and 50 (cocaine). The number of drugs used should be between 0 and 4.
    2. First, recode all four variables into two categories: 0 = non-use (0 days), 1 = use (at least 1 day/time)
    3. Four variables have 6 or 7 categories

27

# CONTINUE

4. Recode four variables: 1 (old) = 0 (new), 2-6/7 (old) = 1 (New).  $\rightarrow$



The image shows the 'Recode into Different Variables: Old and New Values' dialog box in SPSS. The 'Old Value' section has 'Range, value through HIGHEST:' selected, with the value '2' entered in the text box below it. A red arrow points to this '2'. The 'New Value' section has 'Value:' selected, with '1' entered in the text box. The 'Old --> New:' list contains '1 --> 0'. At the bottom, there are 'Continue', 'Cancel', and 'Help' buttons.

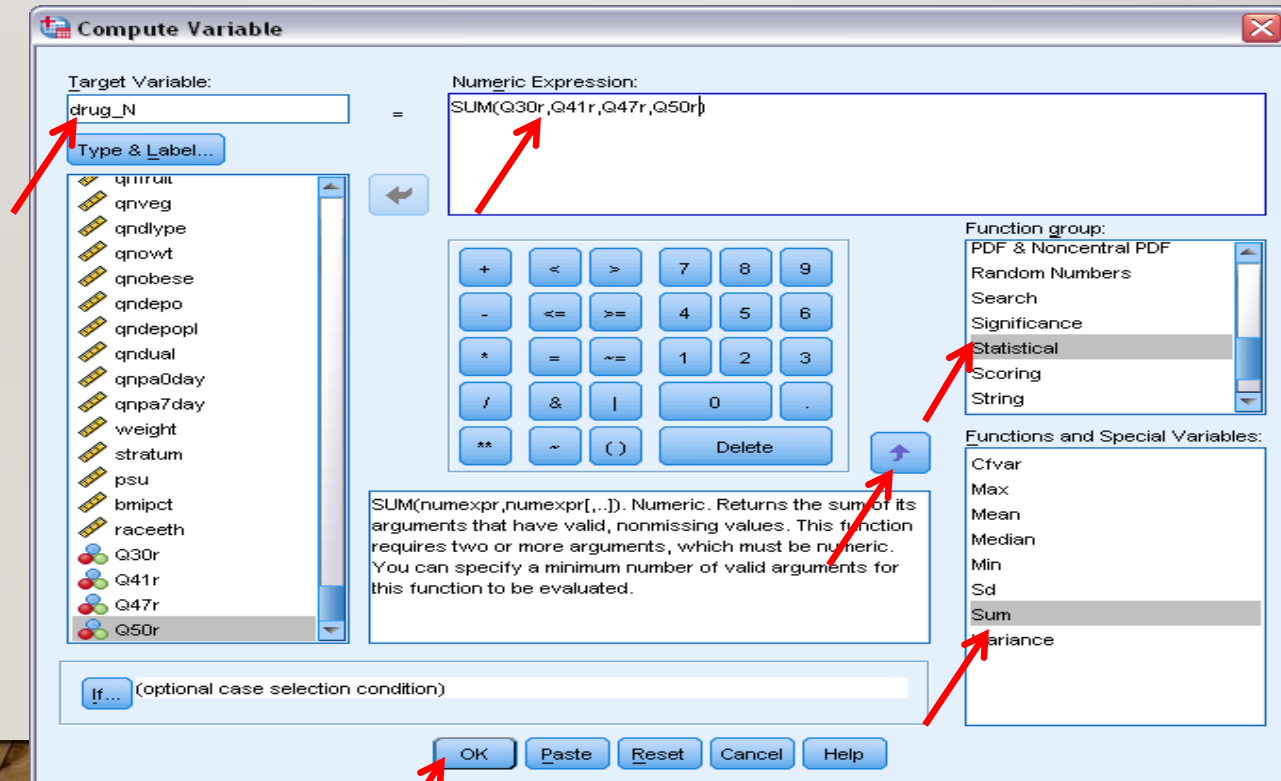
le

# 28 CONTINUE

6. Type drug\_N under Target Variable

7. Numeric Expression: SUM (Q30r,Q41r,Q47r,Q50r)

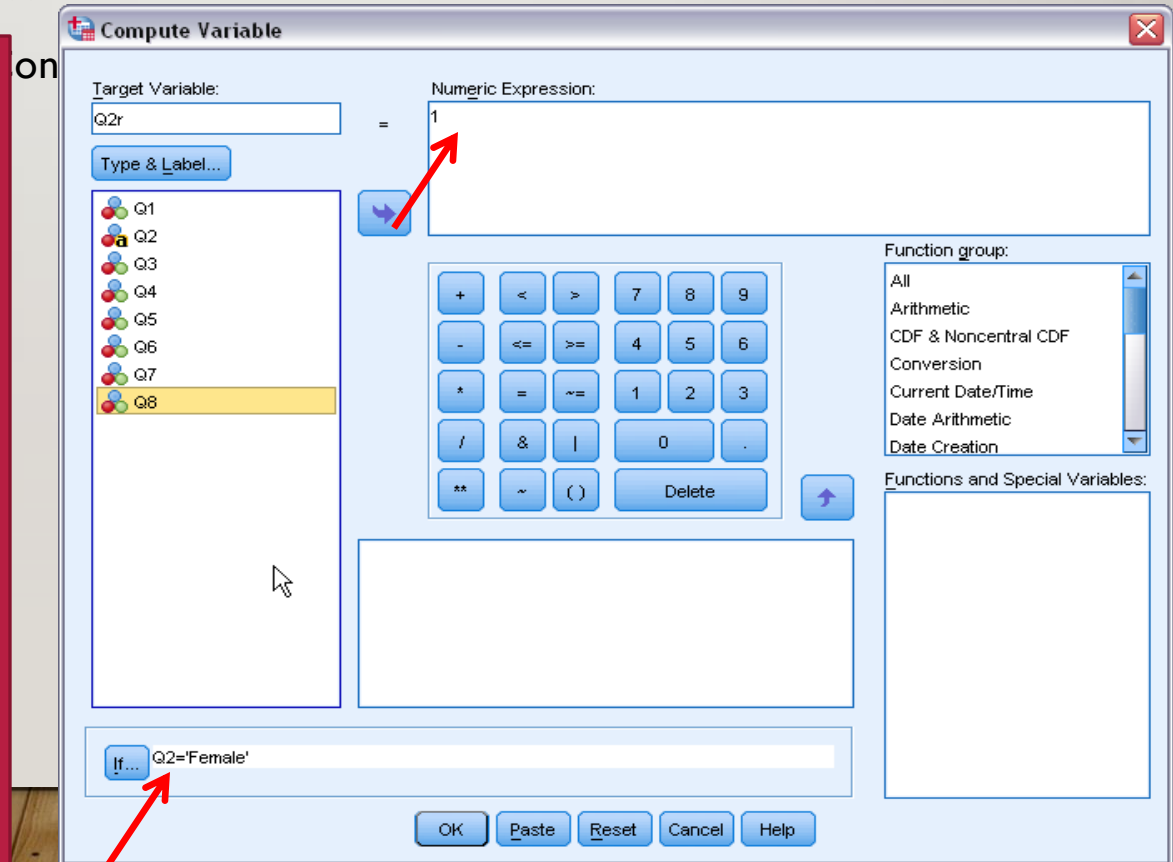
8. Click OK



# 29 CONTINUE

- Compute variables

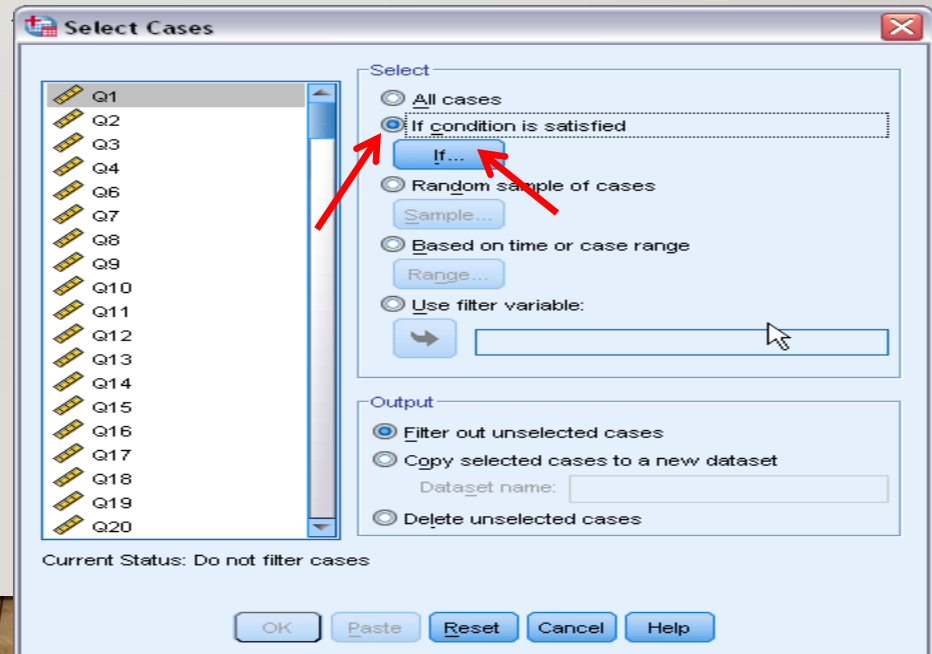
1. Enter 1 at Numeric Expression.
2. Click If button and type Q2 = 'Female'
3. Then click Ok.
4. Enter 2 at Numeric Expression.
5. Click If button and type Q2 = 'Male'
6. Then click Ok





# SORT AND SELECT CASES

- Select cases
  - Example I. Select Females for analysis.
  - 1. Go to Data → Select Cases
  - 2. Under Select: check
  - 3. Click If button



## CONTINU

4. Q2 (gender) = 1

1 means Female

5. Click Continue

6. Click Ok

Unselected cases :  
Q2 = 2

\*yrbs09.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

420: Q1 3 Q2 1 Q3 1

410 3 1 1

411 3 1 1

412 3 1 2

413 3 1 2

414 3 1 1

415 3 1 1

416 3 2 1

417 3 2 1

418 3 2 1

419 3 2 1

420 3 2 1

421 3 2 1

422 3 1 1

423 3 1 1

424 3 1 1

425 3 2 1

426 3 2 1

427 3 2 1

428 3 2 1

429 3 2 1

430 3 1 1

431 3 1 1

432 3 1 1

433 3 1 1

434 3 1 1

435 3 1 1

436 3 1 1

437 3 1 1

438 3 1 1

439 3 1 1

440 3 2 1

441 3 2 1

442 3 2 1

443 3 2 1

444 3 1 1

445 3 1 1

Visible: 208 of 209 Variables

Select Cases: If

Q1 Q2 Q3

Q2 = 1

Function group:

All

Arithmetic

CDF & Noncentral CDF

Conversion

Current Date/Time

Date Arithmetic

Date Creation

Functions and Special Variables:

Continue Cancel Help

IBM SPSS Statistics Processor is ready Filter On Weight On



# 33 SORT AND SELECT CASES

The screenshot displays the IBM SPSS Statistics Data Editor interface. The main window shows a list of variables with the following columns: Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, Measure, and Role. The variable 'filter\_\$' is highlighted in yellow, and a red arrow points to it from the left. Another red arrow points to the 'Variable View' tab at the bottom left of the window.

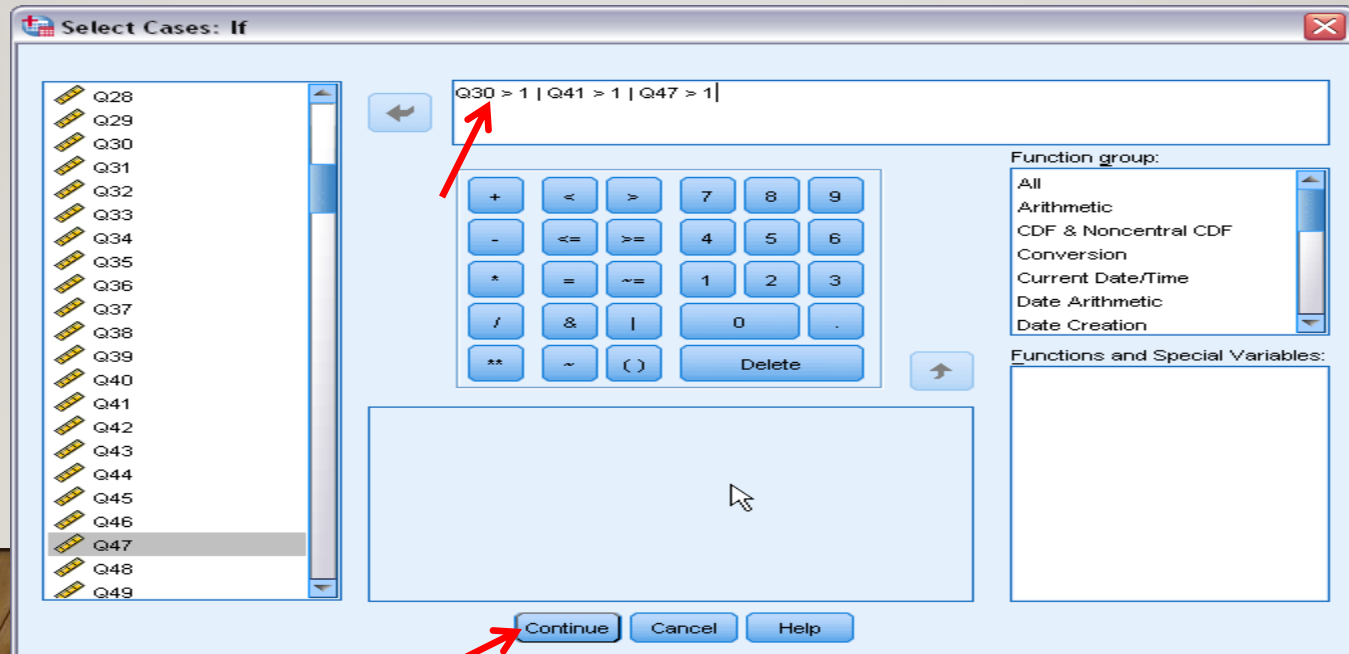
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
188	QN97	Numeric	1	0	Get 8+ hour...	None	-179769313...	6	Right	Scale	Input
189	QN98	Numeric	1	0	Grades mo...	None	-179769313...	6	Right	Scale	Input
190	site	String	3	0	Site Code	None		6	Left	Nominal	Input
191	qnfrcig	Numeric	1	0	Smoked on...	None	-179769313...	9	Right	Scale	Input
192	qnanytob	Numeric	1	0	Used any t...	None	-179769313...	10	Right	Scale	Input
193	qnfveg	Numeric	1	0	Ate 5+ fruit...	None	-179769313...	8	Right	Scale	Input
194	qnfrcig	Numeric	1	0	Ate 2+ fruit...	None	-179769313...	9	Right	Scale	Input
195	qmwveg	Numeric	1	0	Ate 3+ vege...	None	-179769313...	7	Right	Scale	Input
196	qndlype	Numeric	1	0	Attended P...	None	-179769313...	9	Right	Scale	Input
197	qnowt	Numeric	1	0	Overweight	None	-179769313...	7	Right	Scale	Input
198	qnobese	Numeric	1	0	Obese	None	-179769313...	9	Right	Scale	Input
199	qndepo	Numeric	1	0	Of current s...	None	-179769313...	8	Right	Scale	Input
200	qndepopl	Numeric	1	0	Of current s...	None	-179769313...	10	Right	Scale	Input
201	qndual	Numeric	1	0	Of current s...	None	-179769313...	8	Right	Scale	Input
202	qnpa0day	Numeric	1	0	Active for 6...	None	-179769313...	10	Right	Scale	Input
203	qnpa7day	Numeric	1	0	Active for 6...	None	-179769313...	10	Right	Scale	Input
204	weight	Numeric	10	0	Weight	None	-179769313...	12	Right	Scale	Input
205	stratum	Numeric	3	0	Stratum	None	-179769313...	9	Right	Scale	Input
206	psu	Numeric	7	0	Primary Sa...	None	-179769313...	9	Right	Scale	Input
207	bmipct	Numeric	5	0	Body Mass...	None	-179769313...	8	Right	Scale	Input
208	raceeth	Numeric	5	0	Race/Ethni...	{1, Am Indi...	-179769313...	9	Right	Scale	Input
209	filter_\$	Numeric	1	0	Q2 = 1 (FIL...	{0, Not Sele...	None	10	Right	Nominal	Input
210											
211											
212											
213											
214											
215											
216											
217											
218											
219											
220											
221											
222											
223											
224											
225											

# SORT AND SELECT CASES

- Select cases

→ Example 2. Select cases who used any of cigarettes, alcohol, and marijuana during the last 30 days.

1. Data Select Cases



# 35 BASIC STATISTICAL ANALYSIS

- Descriptive statistics
  - Purposes:
    1. Find wrong entries
    2. Have basic knowledge about the sample and targeted variables in a study
    3. Summarize data

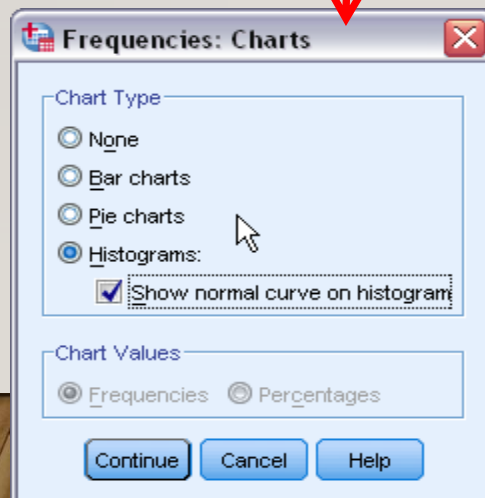
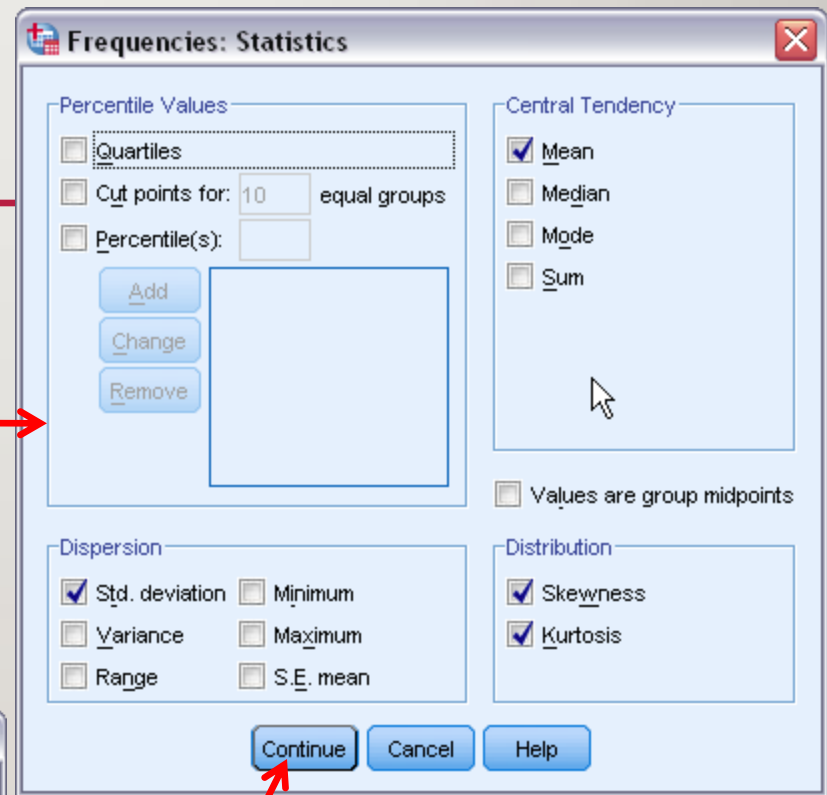
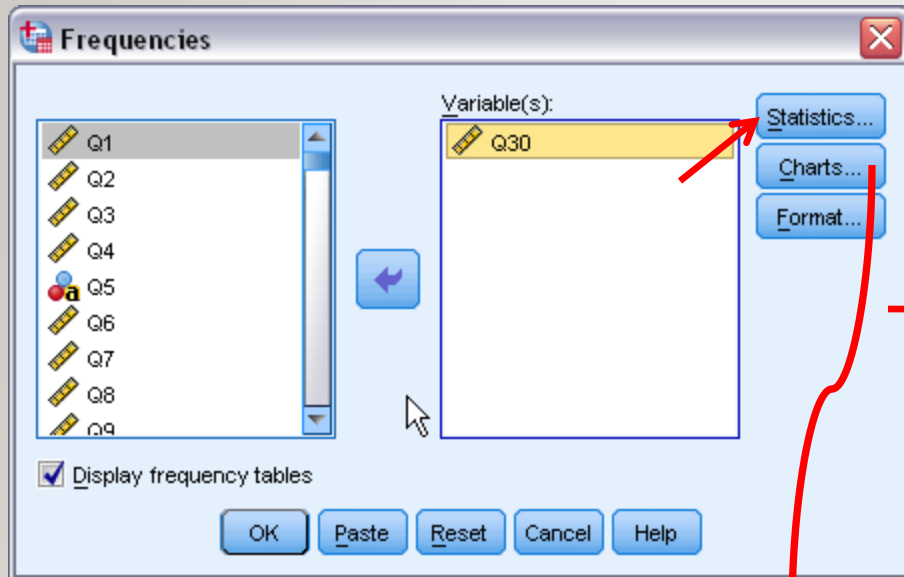
Analyze

Descriptive statistics

Frequency



# 36 CONTINUE



# 37 FREQUENCY TABLE

The screenshot displays the IBM SPSS Statistics Viewer interface. The main window shows the output of a frequency analysis for the variable 'Q30 How many days smoked 30 days'. The output is divided into several sections:

- Summary Table:** A table showing the distribution of responses for 'Q30 How many days smoked 30 days'. The categories and their corresponding counts, percentages, and cumulative percentages are as follows:

Category	Count	Percent	Cumulative Percent
4 to 9 days	282	1.7	1.8
5 to 19 days	360	2.2	2.3
6 to 29 days	310	1.9	2.0
7 All 30 days	845	5.1	5.3
Total	15846	96.6	100.0
Missing System	564	3.4	
Total	16410	100.0	

- Statistics:** A table providing summary statistics for the variable 'Q30 How many days smoked 30 days':

Statistic	Value
N Valid	15846
N Missing	564
Mean	1.627
Std. Deviation	1.6227
Skewness	2.449
Std. Error of Skewness	.019
Kurtosis	4.630
Std. Error of Kurtosis	.039

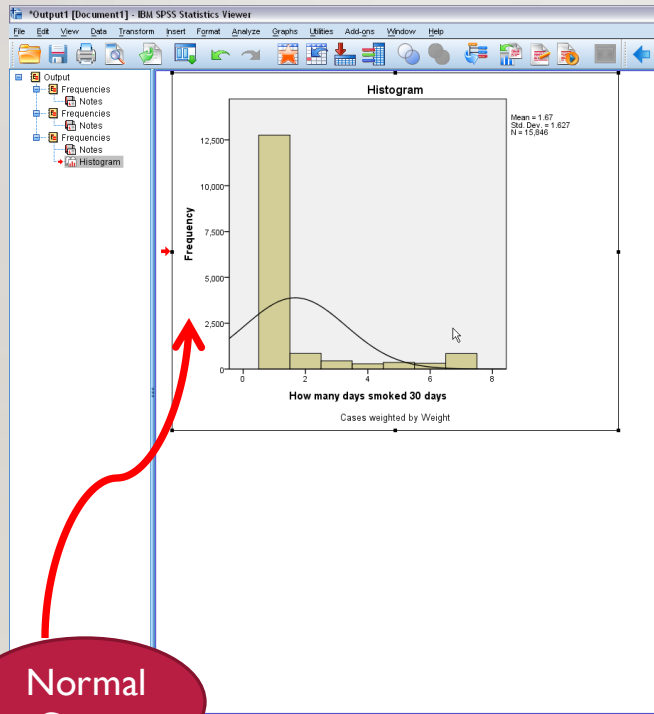
- Statistics:** A table providing summary statistics for the variable 'Q30 How many days smoked 30 days' (repeated):

Statistic	Value
N Valid	15846
N Missing	564
Mean	1.627
Std. Deviation	1.6227
Skewness	2.449
Std. Error of Skewness	.019
Kurtosis	4.630
Std. Error of Kurtosis	.039

- Statistics:** A table providing summary statistics for the variable 'Q30 How many days smoked 30 days' (repeated):

Statistic	Value
N Valid	15846
N Missing	564
Mean	1.627
Std. Deviation	1.6227
Skewness	2.449
Std. Error of Skewness	.019
Kurtosis	4.630
Std. Error of Kurtosis	.039

Two red arrows point to the 'Statistics' section and the 'Q30 How many days smoked 30 days' table.



Normal  
Curve

1. *Skewness*: a measure of the asymmetry of a distribution. The normal distribution is symmetric and has a skewness value of zero.

*Positive skewness*: a long right tail.  
*Negative skewness*: a long left tail.

Departure from symmetry : a skewness value more than twice its standard error.

2. *Kurtosis*: A measure of the extent to which observations cluster around a central point. For a normal distribution, the value of the kurtosis statistic is zero. *Leptokurtic* data values are more peaked, whereas *platykurtic* data values are flatter and more dispersed along the X axis.

39

---

**Q & A**

