Exploring SPSS for Data Analysis

دانشگاه علوم پزشکی ساوه

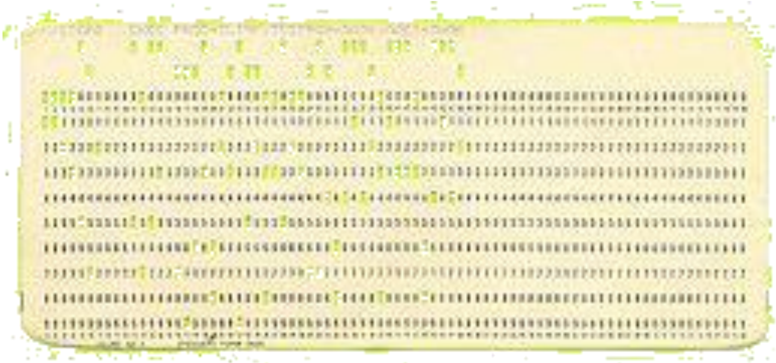
Introduction

- SPSS is a software used for statistical analysis
- First released in 1968 and was developed by Norman H Nie, Dale H. Bent and C. Hadial Hull
- Since its release, SPSS was under SPSS Inc.
- However in July 28, 2009 SPSS was acquired by IBM for US$1.2 billion
- Versions 17 and 18 were known as PASW (Predictive and Analytical Software)
- Version 19 was renamed as SPSS Statistics

Next ▷

SPSS Versions

– The earlier versions of SPSS ran on mainframe computers
– SPSS/PC+ was first introduced in 1984
– SPSS 6 for Windows was introduced in mid 1990's
– SPSS 15 - November 2006
– SPSS 16 - April 2008
– PASW Statistics 17 – December 2008
– PASW Statistics 18 – August 2009
– SPSS Statistics 19 – 2010
– SPSS Statistics 20 – 2011
– SPSS Statistics 21 – 2012
– SPSS Statistics 22 – 2013
– SPSS Statistics 23 – 2015



IBM 80-column punched card

Card punch machine

Next ▷

Learning Objectives

Participants to be able to:

❶ Understand three SPSS windows

❷ 7 Steps in Data Preparation

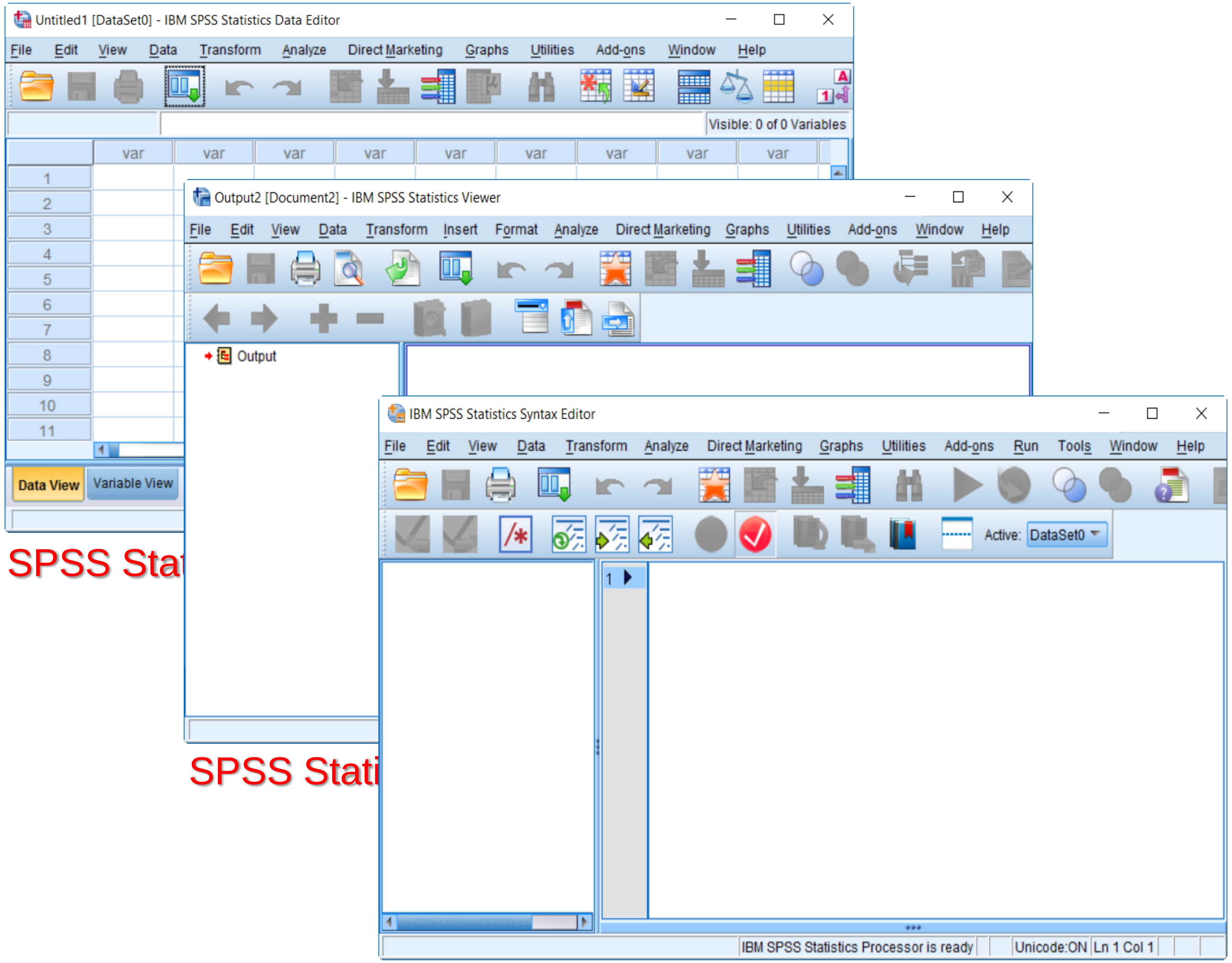
❸ Define variable and enter data into Data Editor

❹ Perform data editing and transformation

❺ Run selected statistical procedures

❻ Use SPSS coaches

end



SPSS Stat

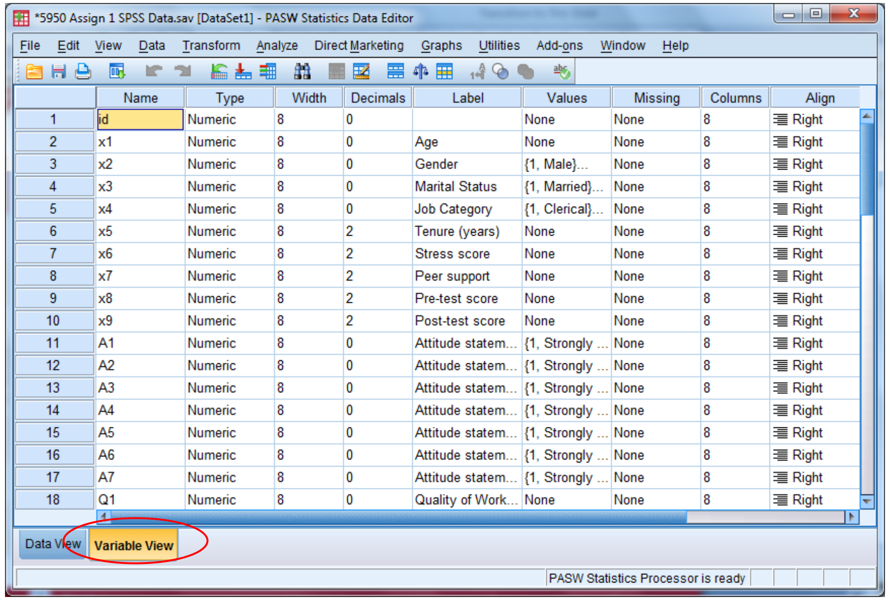
SPSS Stati

SPSS Statistics SYNTAX EDITOR

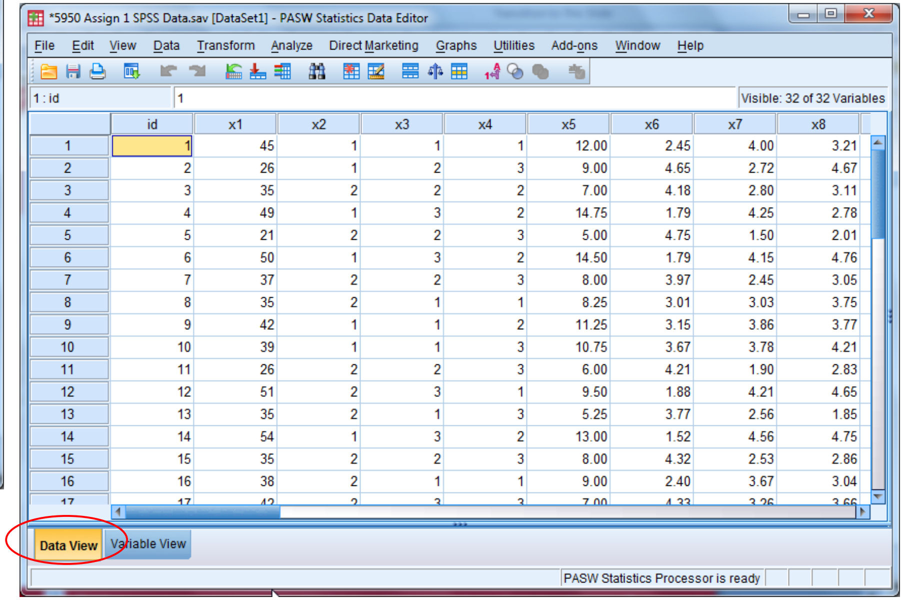
3 SPSS Windows

1 SPSS Statistics Data Editor

– Include two (2) tabs – 1) Variable view and 2) Data view
– Two (2) tasks – 1) Define variables and 2) Enter data



Variable view: Define variables



Data view: Enter data

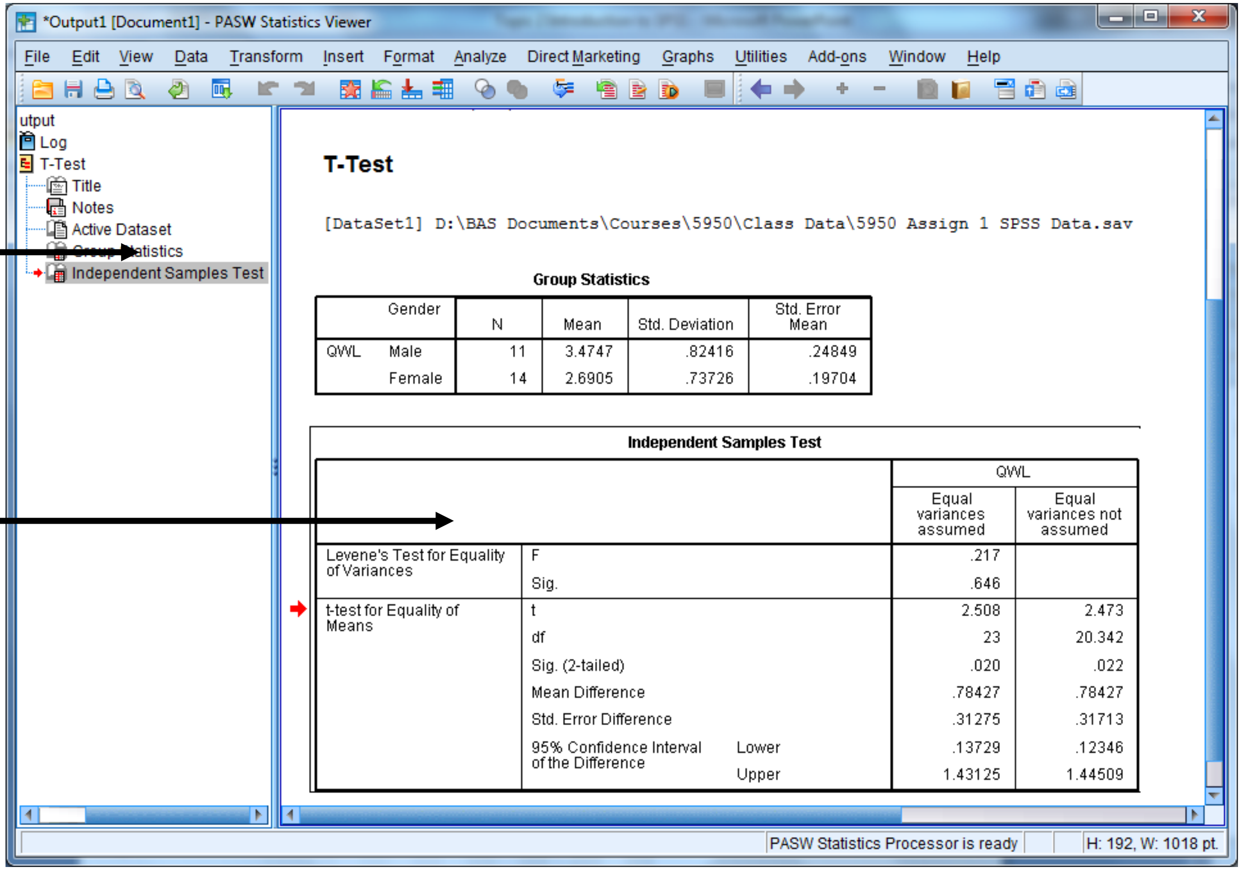
Next ▷

2 SPSS Statistics Viewer

- Display results of data analyses
- Display two (2) panels/panes

Navigator to help easy browsing

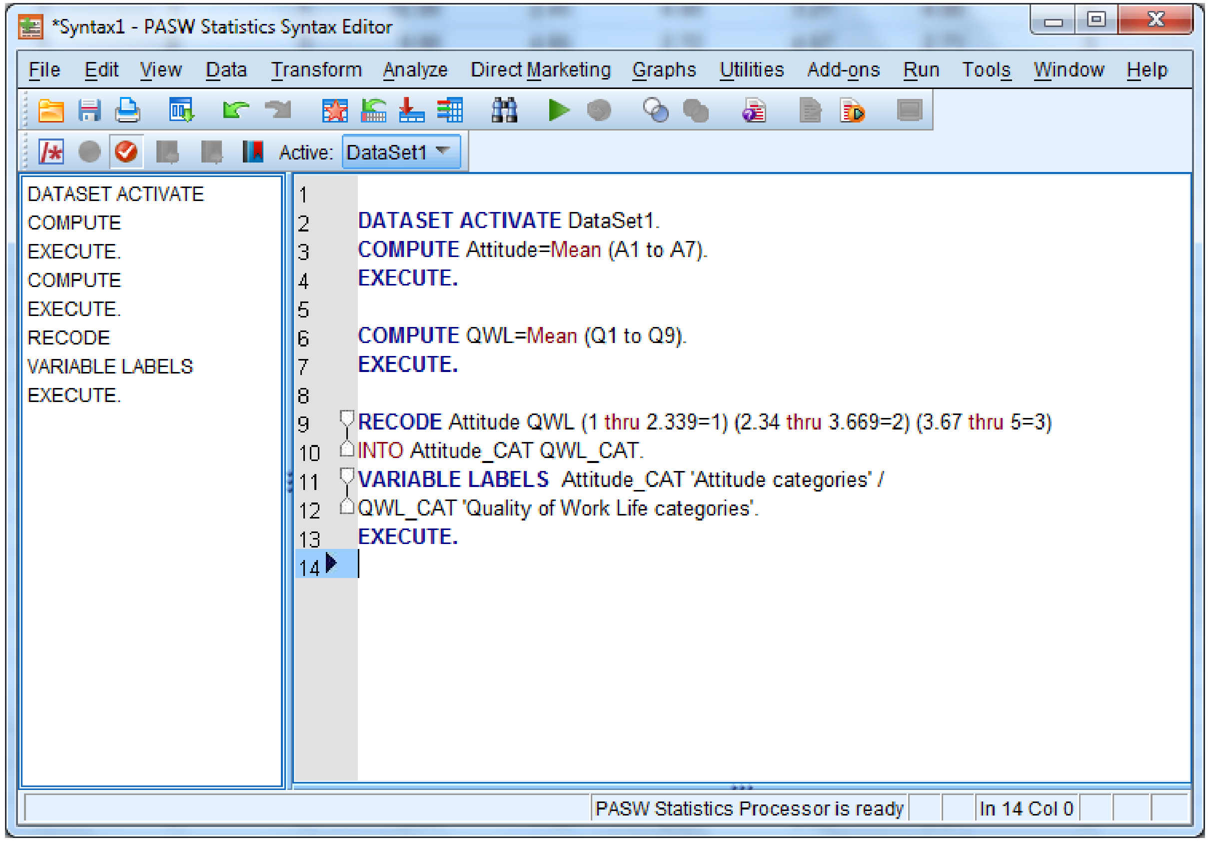
Results of data analyses



Next ▷

3 SPSS Statistics Syntax Editor

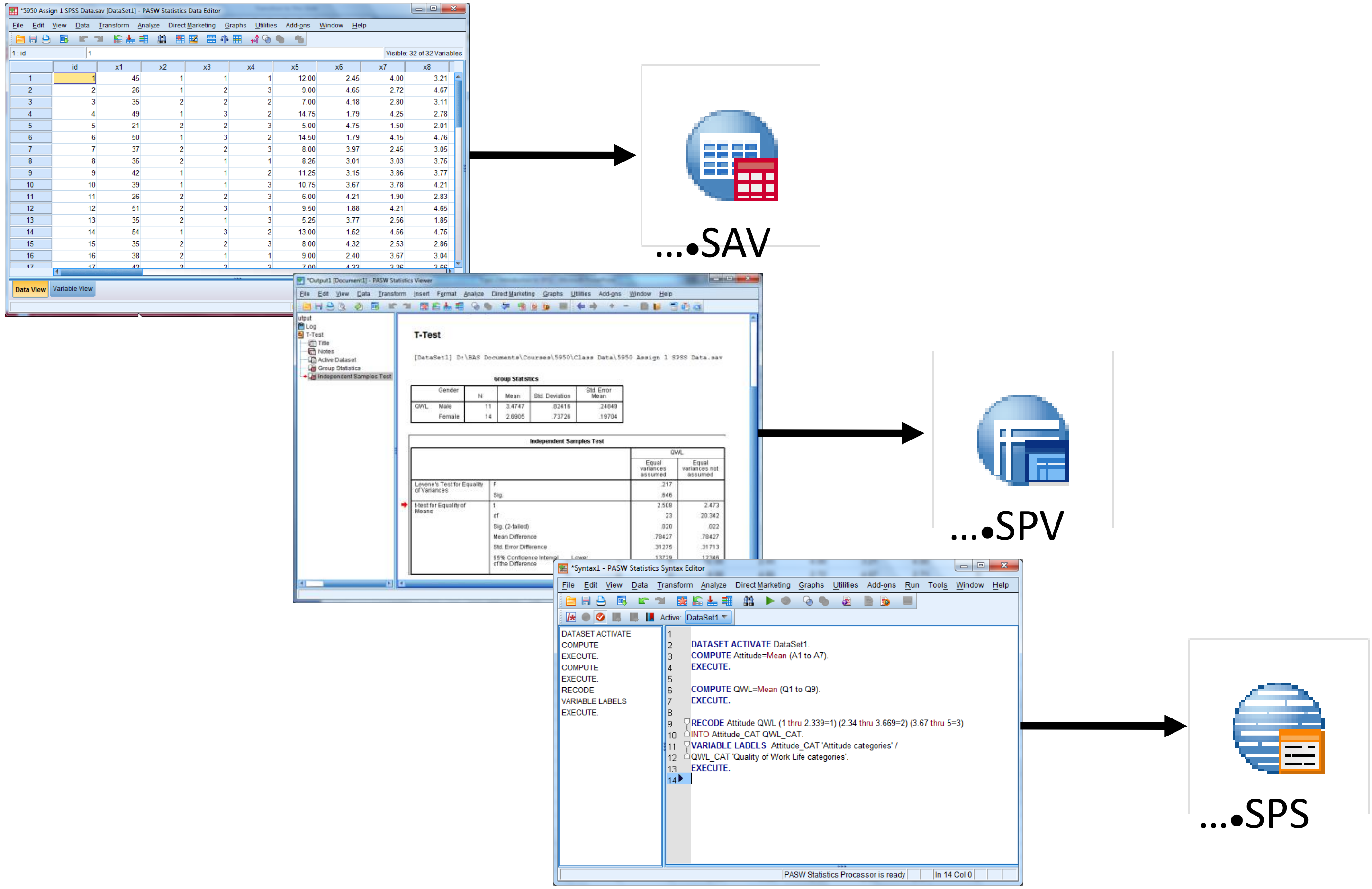
- Write, display, retrieve, run and save syntax/commands
- Use for two (2) purposes:
  1. For future reference
  2. Automate data analysis



Purpose of Syntax Editor:
1. For future reference
2. Automate data analysis

Next ▷

SPSS File Format



◁ Back

7 Steps in Data Preparation

Steps in Data Preparation

1. Define variables
2. Enter data
3. Run frequencies
4. Edit data
5. Test reliability
6. Transform data:
   - Compute
   - Recode
7. Run Exploratory Data Analysis (EDA)

Define Variables 1

Enter Data 2

Run Frequency 3

Data Editing 4

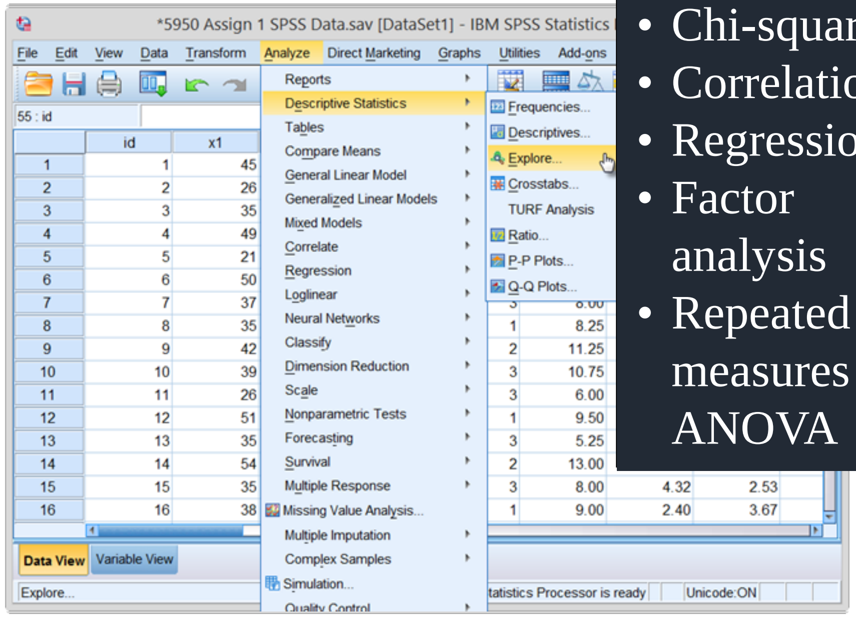
Reliability Test 5

Transfor-mations 6

EDA Normality 7

Data Analyses

- Frequency
- t-test
- ANOVA
- Chi-square
- Correlation
- Regression
- Factor analysis
- Repeated measures ANOVA



◁ Back

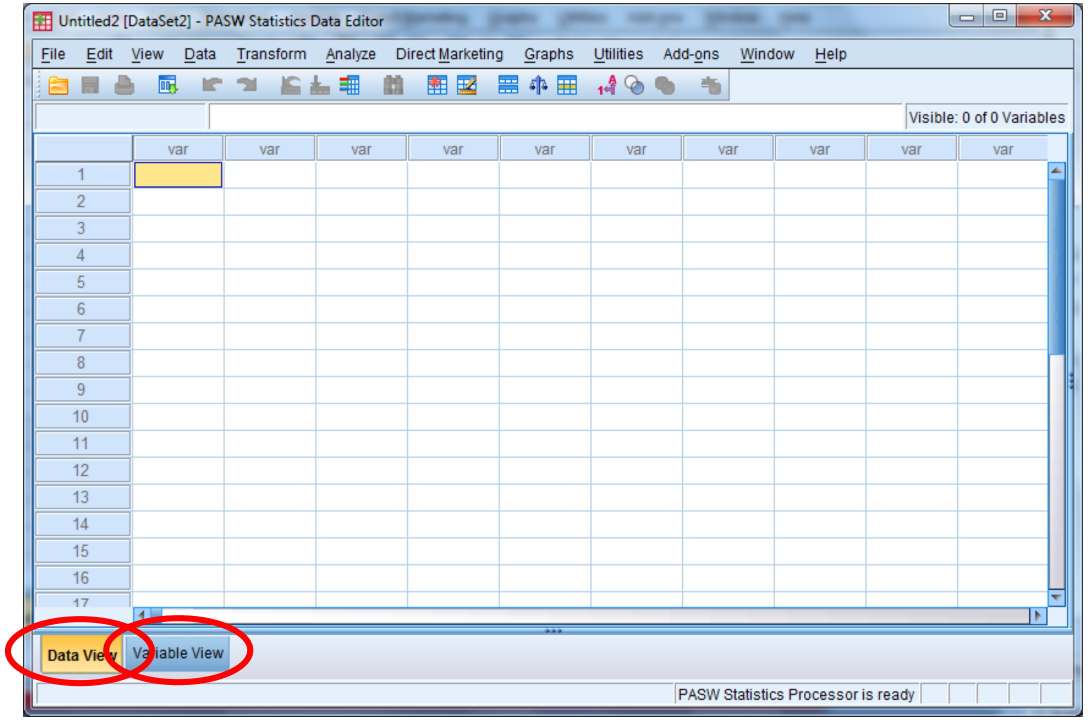
Define Variables & Enter Data

SPSS Statistics Data Editor

In the Data Editor, you can:

❶ Define variables
in Variable View window

❷ Enter data
in Data View window



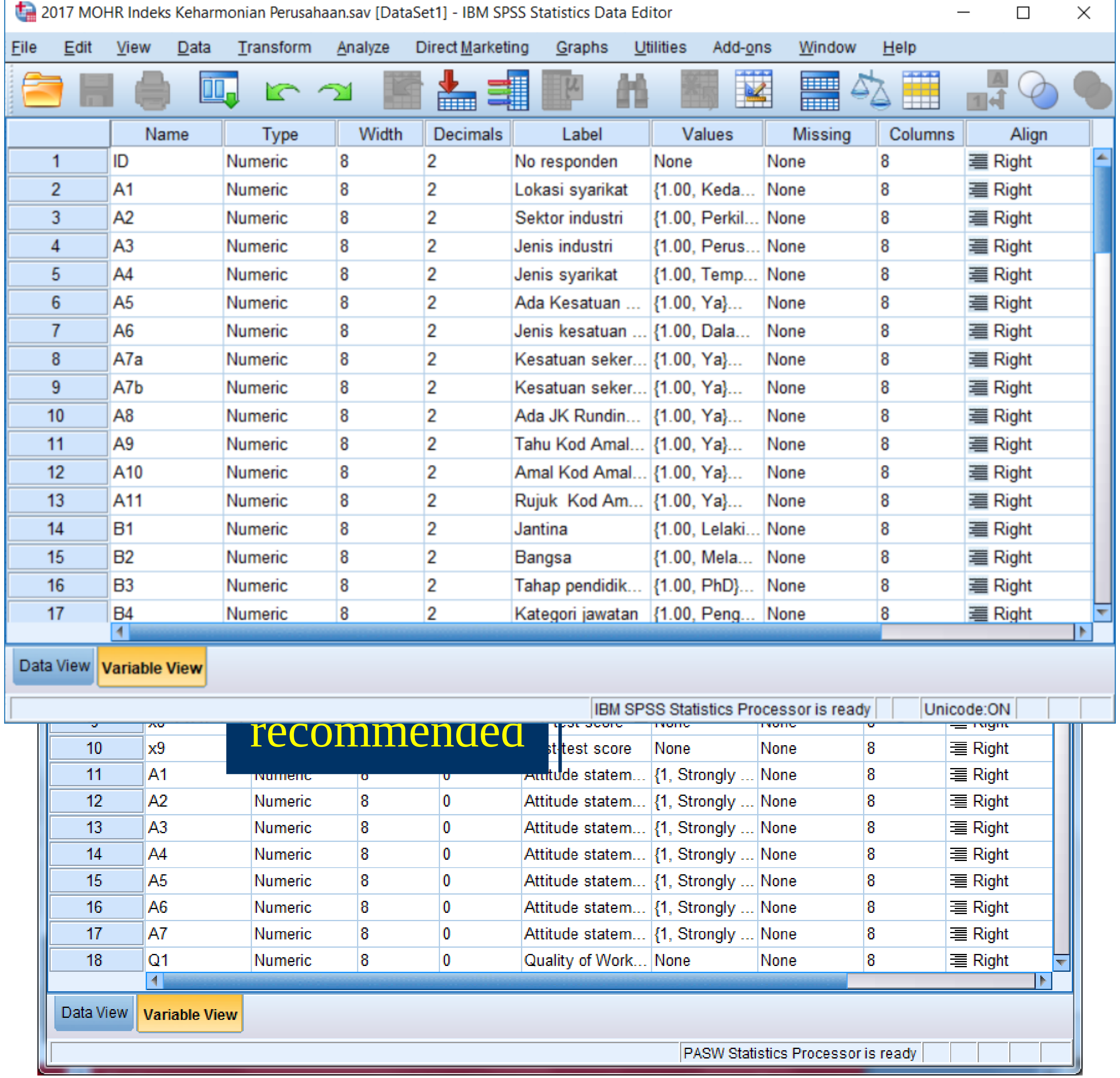
Next ▷

1 Define Variables

Click the Variable View tab

Define:

- Name
- Label
- Values



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align
1	ID	Numeric	8	2	No responden	None	None	8	Right
2	A1	Numeric	8	2	Lokasi syarikat	{1.00, Keda...	None	8	Right
3	A2	Numeric	8	2	Sektor industri	{1.00, Perkil...	None	8	Right
4	A3	Numeric	8	2	Jenis industri	{1.00, Perus...	None	8	Right
5	A4	Numeric	8	2	Jenis syarikat	{1.00, Temp...	None	8	Right
6	A5	Numeric	8	2	Ada Kesatuan ...	{1.00, Ya}...	None	8	Right
7	A6	Numeric	8	2	Jenis kesatuan ...	{1.00, Dala...	None	8	Right
8	A7a	Numeric	8	2	Kesatuan seker...	{1.00, Ya}...	None	8	Right
9	A7b	Numeric	8	2	Kesatuan seker...	{1.00, Ya}...	None	8	Right
10	A8	Numeric	8	2	Ada JK Rundin...	{1.00, Ya}...	None	8	Right
11	A9	Numeric	8	2	Tahu Kod Amal...	{1.00, Ya}...	None	8	Right
12	A10	Numeric	8	2	Amal Kod Amal...	{1.00, Ya}...	None	8	Right
13	A11	Numeric	8	2	Rujuk Kod Am...	{1.00, Ya}...	None	8	Right
14	B1	Numeric	8	2	Jantina	{1.00, Lelaki...	None	8	Right
15	B2	Numeric	8	2	Bangsa	{1.00, Mela...	None	8	Right
16	B3	Numeric	8	2	Tahap pendidik...	{1.00, PhD}...	None	8	Right
17	B4	Numeric	8	2	Kategori jawatan	{1.00, Peng...	None	8	Right

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align
10	x9				st test score	None	None	8	Right
11	A1	Numeric	8	0	Attitude statem...	{1, Strongly ...	None	8	Right
12	A2	Numeric	8	0	Attitude statem...	{1, Strongly ...	None	8	Right
13	A3	Numeric	8	0	Attitude statem...	{1, Strongly ...	None	8	Right
14	A4	Numeric	8	0	Attitude statem...	{1, Strongly ...	None	8	Right
15	A5	Numeric	8	0	Attitude statem...	{1, Strongly ...	None	8	Right
16	A6	Numeric	8	0	Attitude statem...	{1, Strongly ...	None	8	Right
17	A7	Numeric	8	0	Attitude statem...	{1, Strongly ...	None	8	Right
18	Q1	Numeric	8	0	Quality of Work...	None	None	8	Right

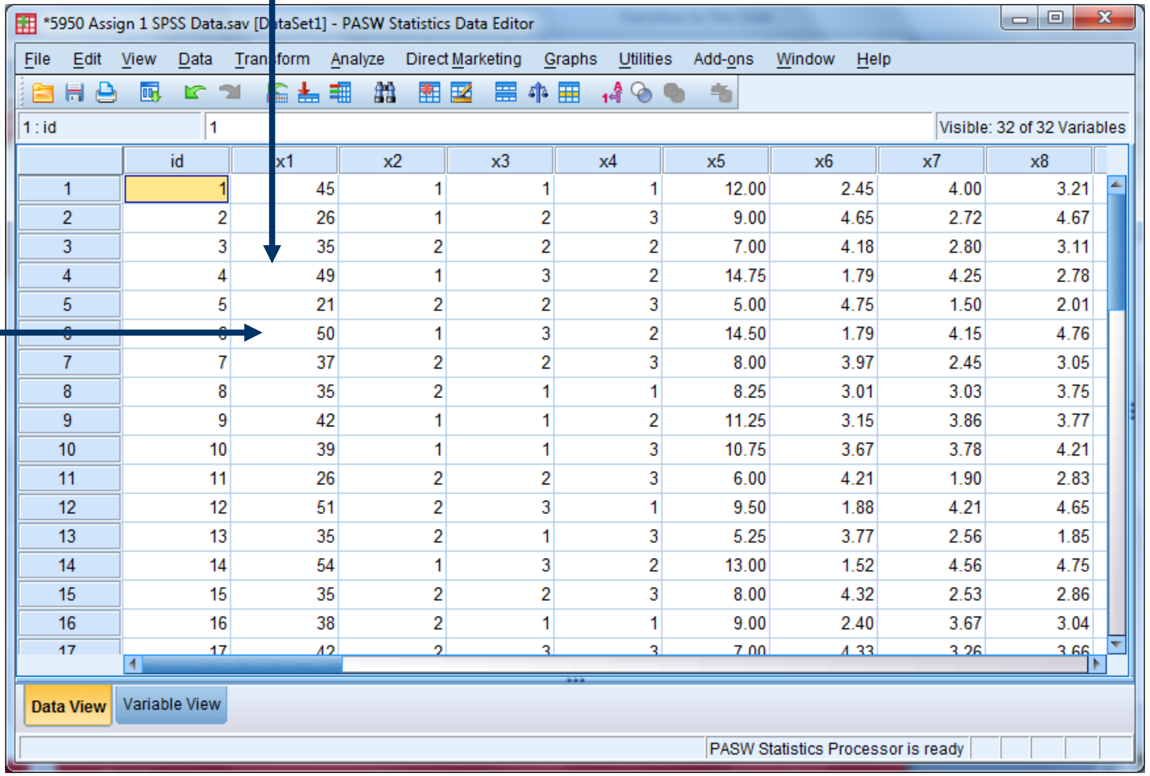
Next ▷

2 Enter Data

Enter data in Data View windows

One column refers to one variable

One row refers to one case or observation



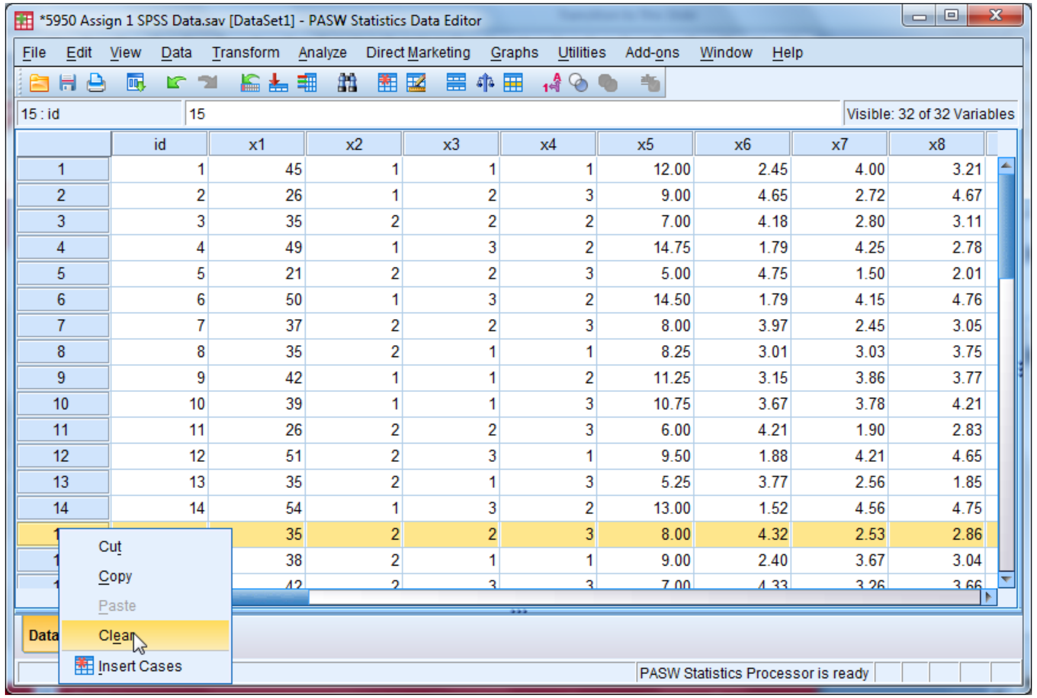
	id	x1	x2	x3	x4	x5	x6	x7	x8
1	1	45	1	1	1	12.00	2.45	4.00	3.21
2	2	26	1	2	3	9.00	4.65	2.72	4.67
3	3	35	2	2	2	7.00	4.18	2.80	3.11
4	4	49	1	3	2	14.75	1.79	4.25	2.78
5	5	21	2	2	3	5.00	4.75	1.50	2.01
6	6	50	1	3	2	14.50	1.79	4.15	4.76
7	7	37	2	2	3	8.00	3.97	2.45	3.05
8	8	35	2	1	1	8.25	3.01	3.03	3.75
9	9	42	1	1	2	11.25	3.15	3.86	3.77
10	10	39	1	1	3	10.75	3.67	3.78	4.21
11	11	26	2	2	3	6.00	4.21	1.90	2.83
12	12	51	2	3	1	9.50	1.88	4.21	4.65
13	13	35	2	1	3	5.25	3.77	2.56	1.85
14	14	54	1	3	2	13.00	1.52	4.56	4.75
15	15	35	2	2	3	8.00	4.32	2.53	2.86
16	16	38	2	1	1	9.00	2.40	3.67	3.04
17	17	42	2	3	3	7.00	4.33	3.26	3.66

◁ Back

Data Editing & Transformations

Data Editing

- Change data value
- Cut, copy and paste data value
- Add or delete case
- Add or delete variable
- Change sequence of variables



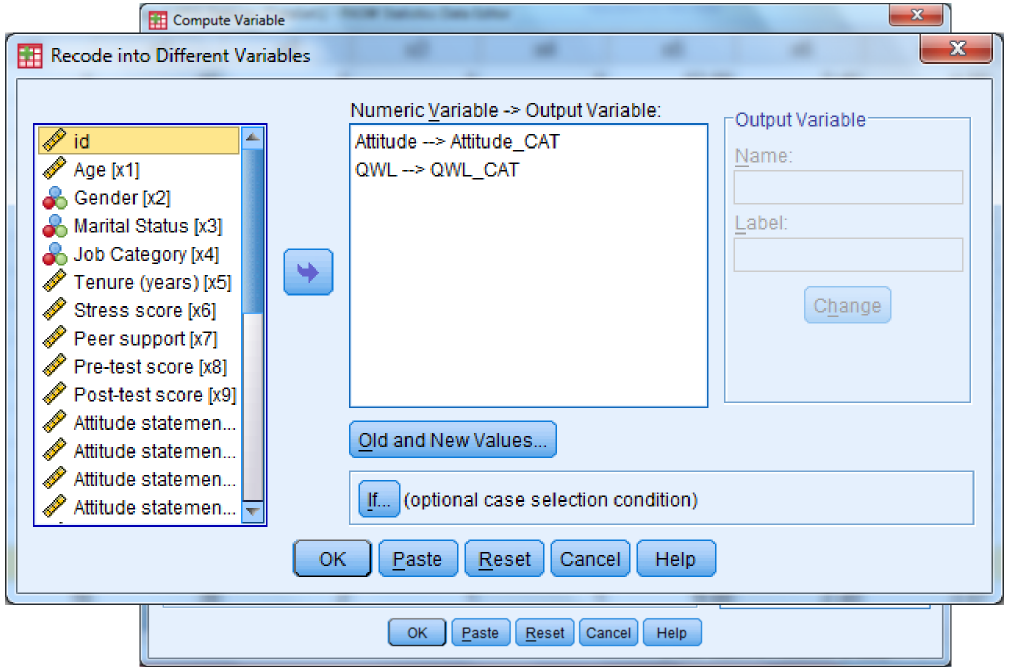
Next ▷

Data Transformations

Two commonly used data transformations:

① Compute
– create new variable based on existing variable/s

② Recode
used to:
- Create categories from continuous variable
- Change values
  Example: 1 → 2
  2 → 1



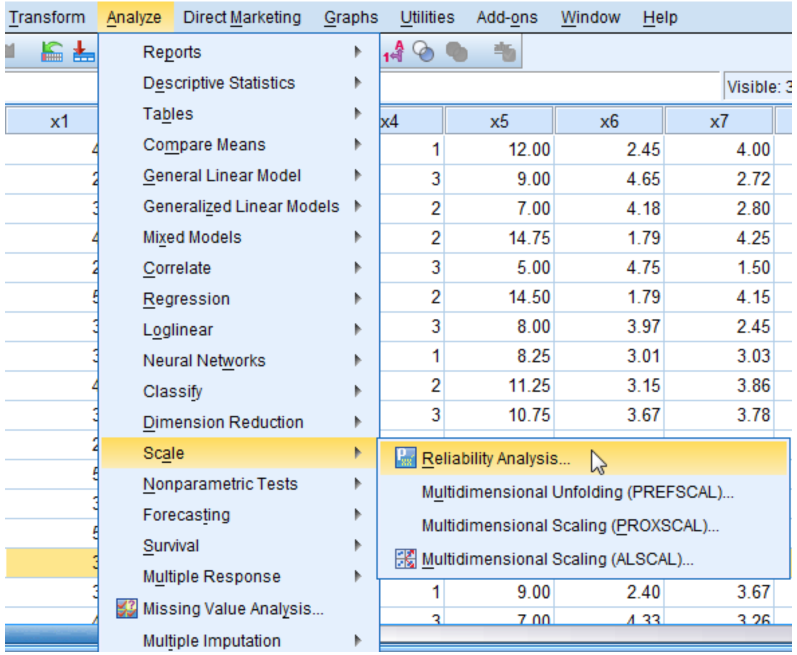
◁ Back

Statistical Procedures

Statistical Procedures

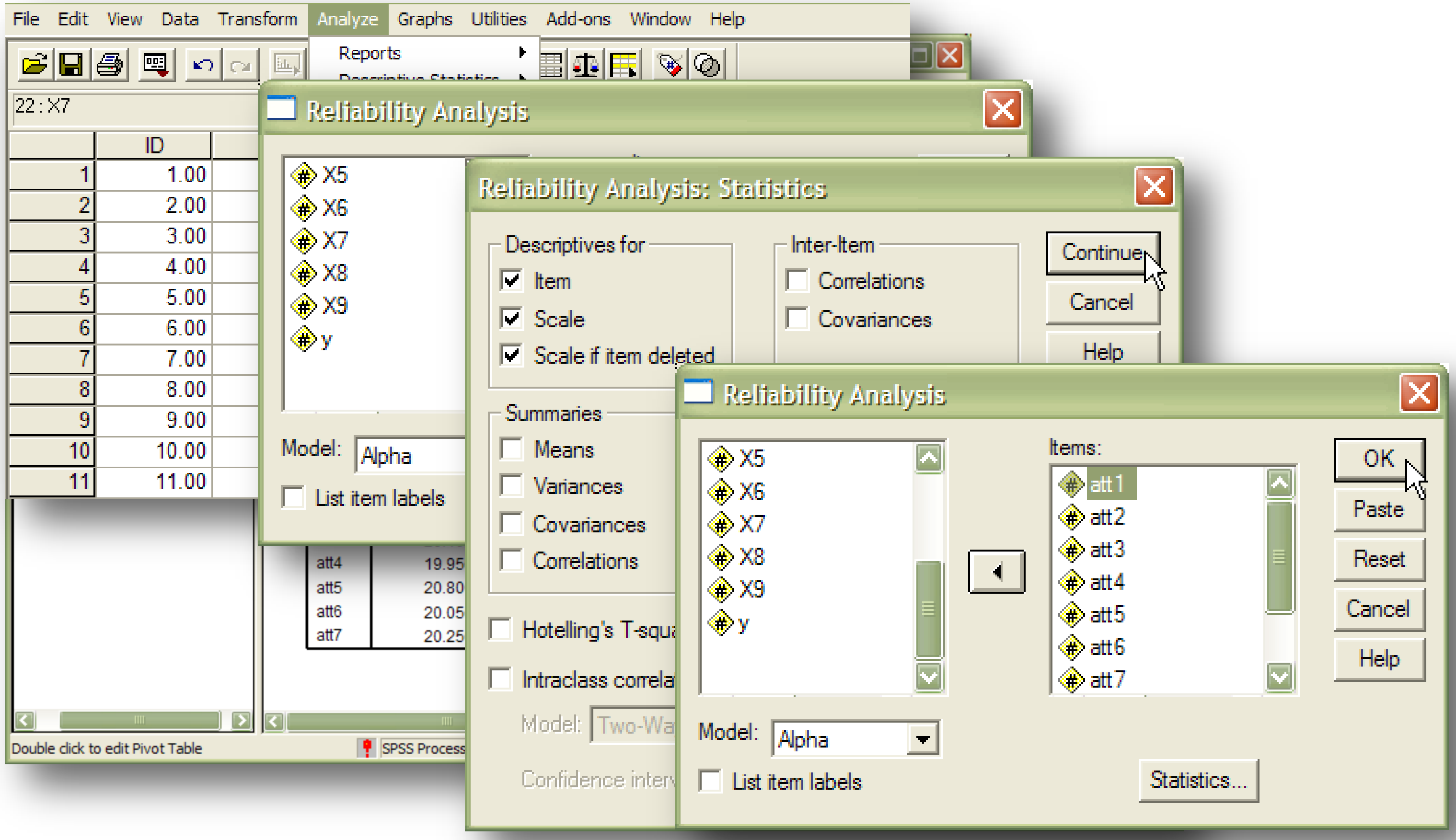
Statistical Procedures:

- Reliability test
- Exploratory data analysis
- Descriptive Statistics
  – Frequencies, crosstabs
- Compare group means
  – t-test and ANOVA
- Relationship between variables
  – Chi-square, Spearman rho, Pearson correlation, and regression analysis



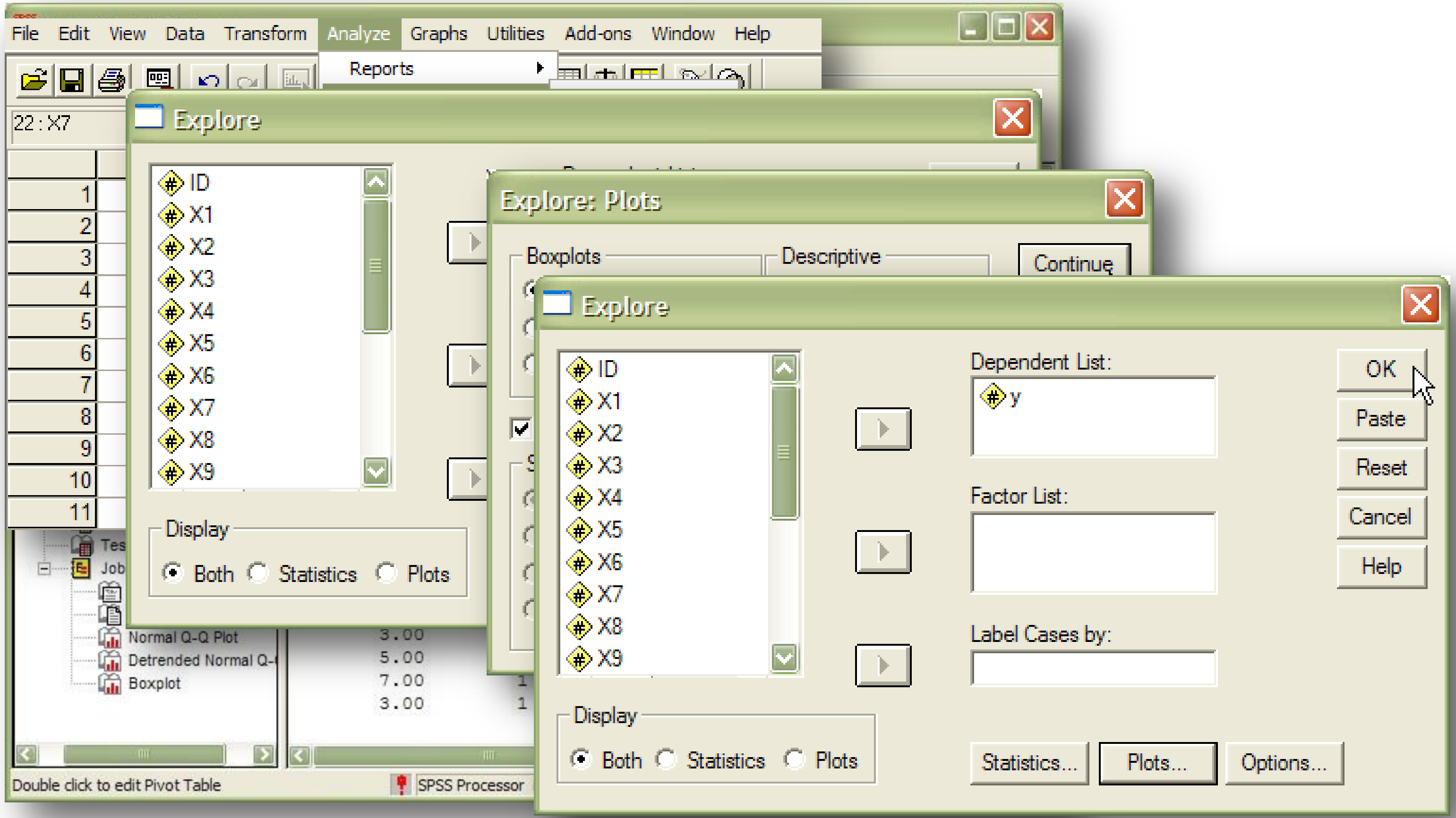
◁ Back

Reliability Test:



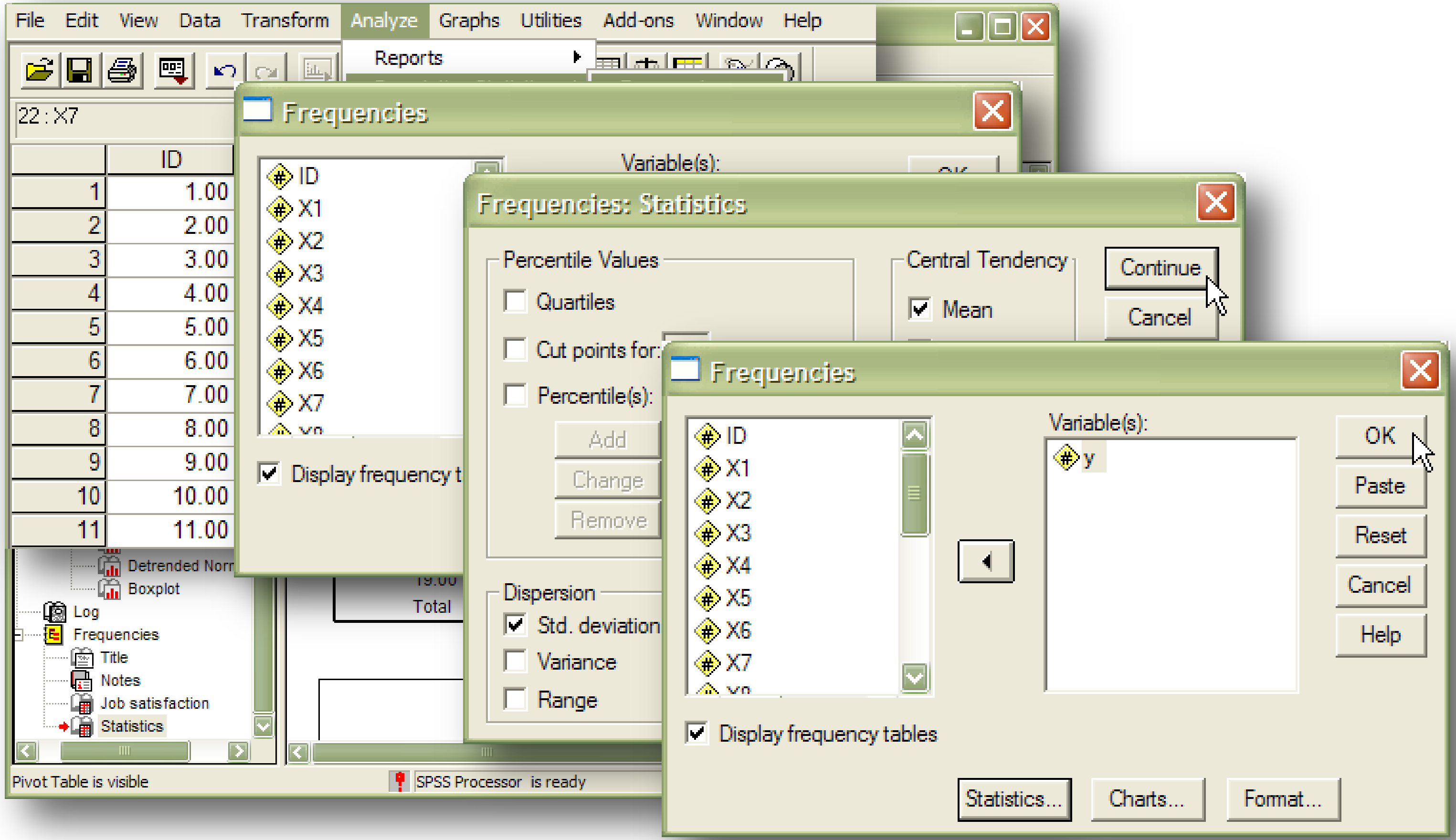
◁ Back

Exploratory Data Analysis:



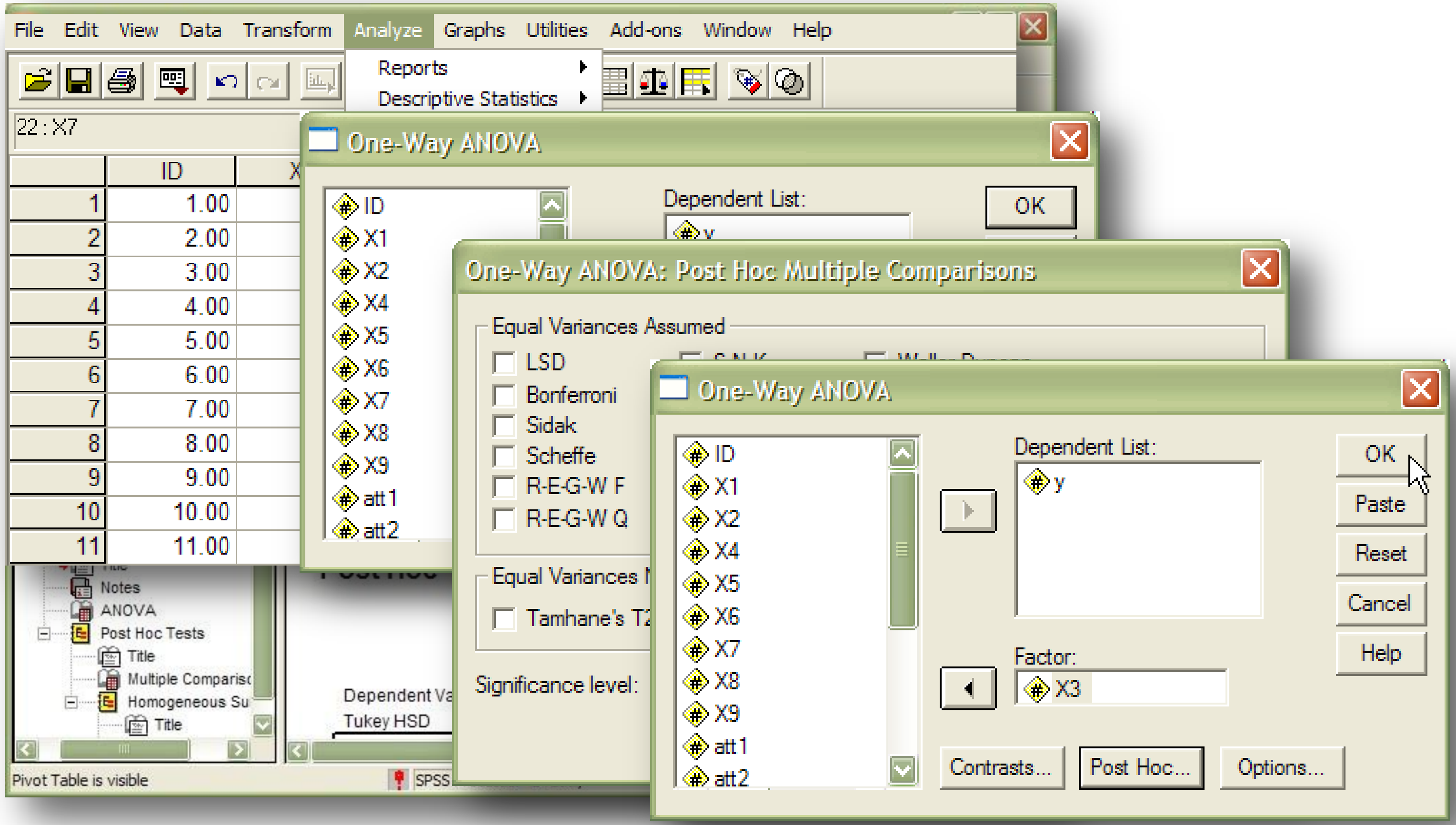
◁ Back

Frequencies:



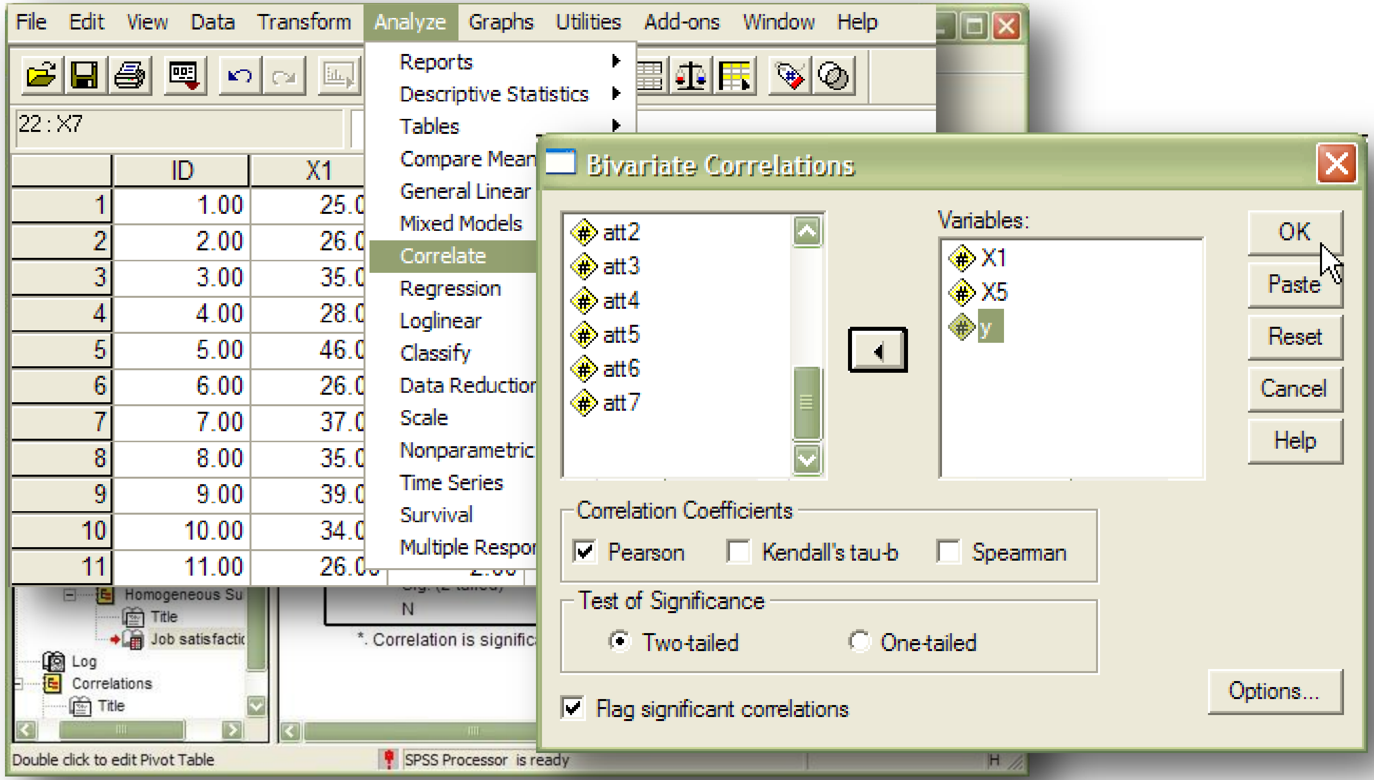
◁ Back

ANOVA:


File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Reports
Descriptive Statistics
22 : X7
ID
One-Way ANOVA
Dependent List:
OK
One-Way ANOVA: Post Hoc Multiple Comparisons
Equal Variances Assumed
LSD
Bonferroni
Sidak
Scheffe
R-E-G-W F
R-E-G-W Q
Tamhane's T2
Significance level:
Notes
ANOVA
Post Hoc Tests
Title
Multiple Comparis
Homogeneous Su
Dependent Va
Tukey HSD
Pivot Table is visible
SPSS
ID
X1
X2
X4
X5
X6
X7
X8
X9
att1
att2
y
Factor:
X3
Paste
Reset
Cancel
Help
Contrasts...
Post Hoc...
Options...

◁ Back

Correlation Analyses:



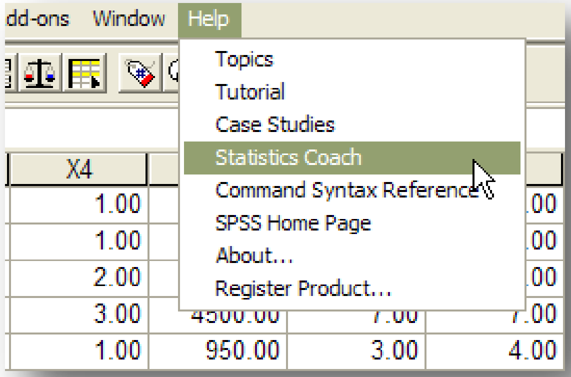
◁ Back

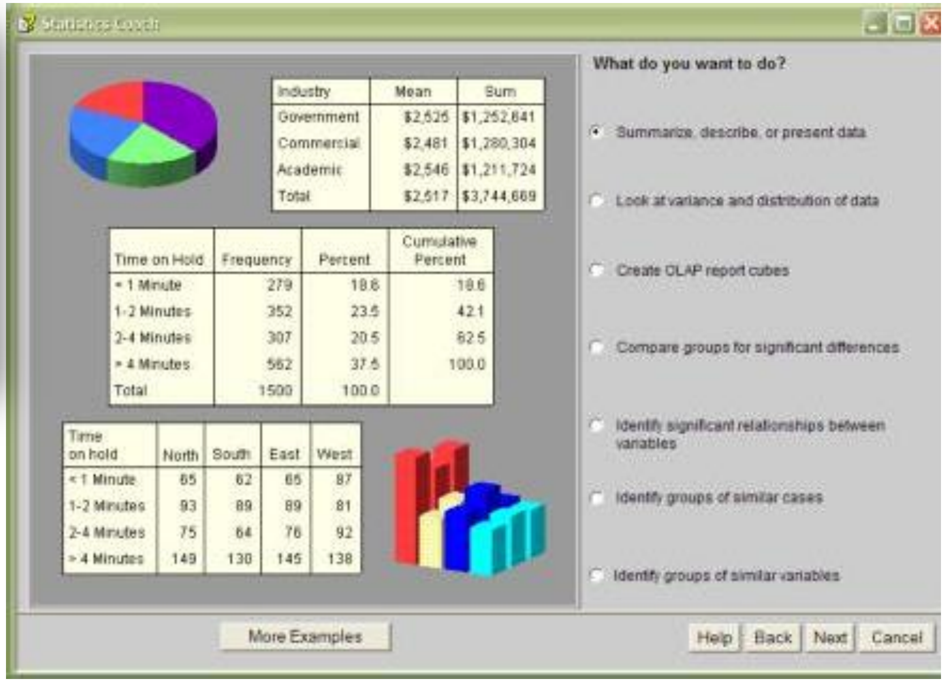
SPSS Coaches

Guide in selecting the most appropriate statistical analysis

1 Statistic Coach

Guide in selecting the most appropriate statistical analysis



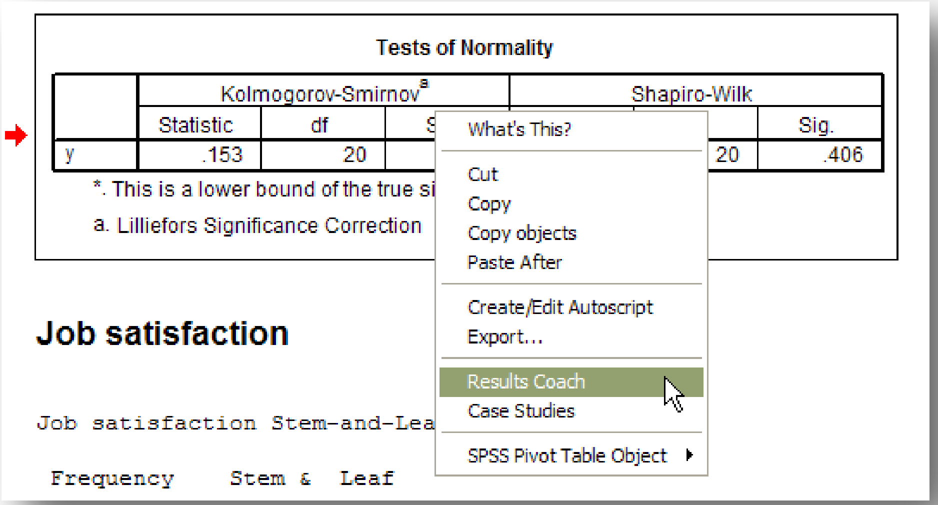


Next ▷

2 Case Studies

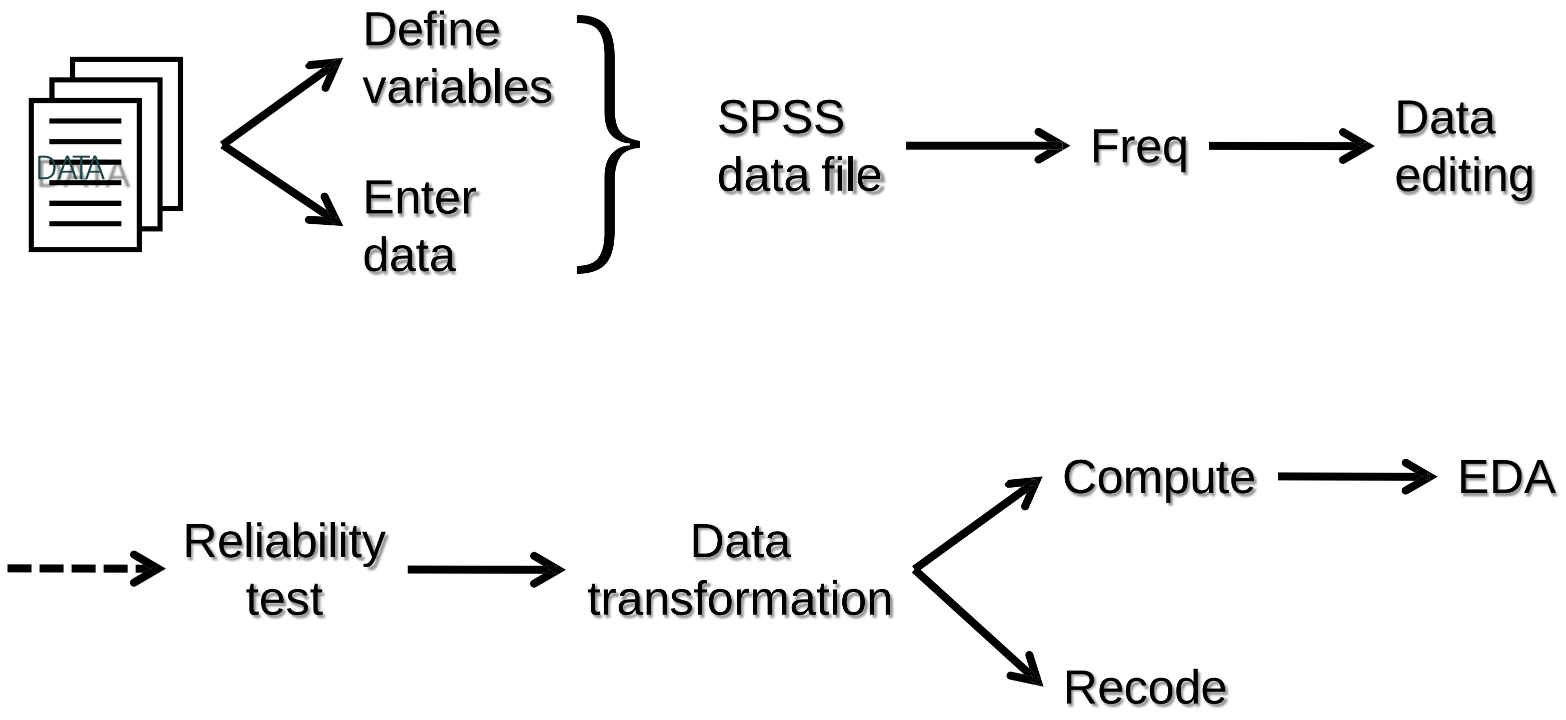
Help to interpret statistical results

Help to interpret statistical results



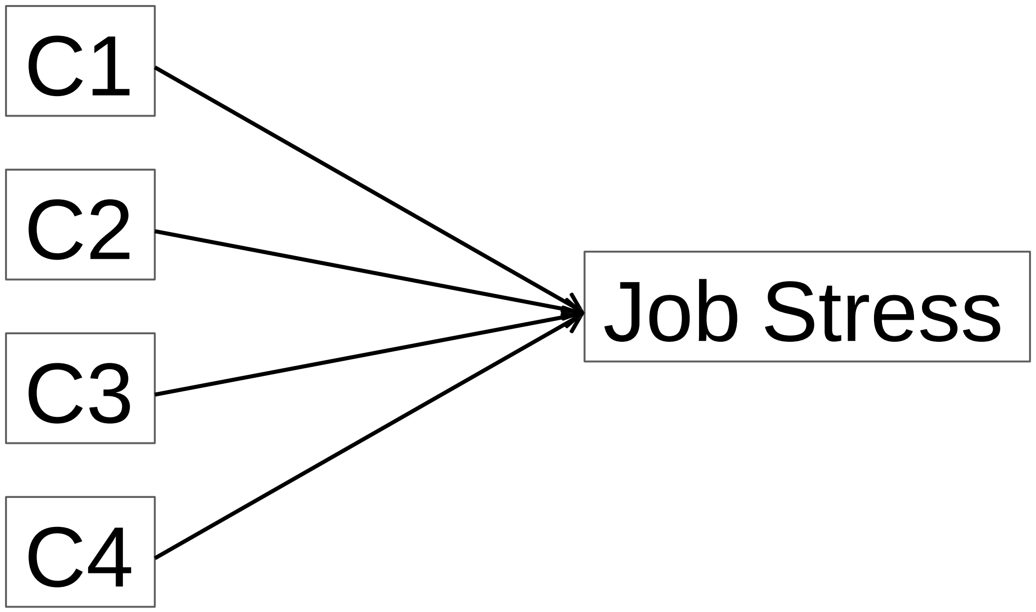
◁ Back

Basic Steps



DATA
Define variables
Enter data
SPSS data file
Freq
Data editing
Reliability test
Data transformation
Compute
EDA
Recode

Why Compute?



Why Recode?

Stress

		Frequency	Percent
Valid	1.50	1	3.8
	1.75	1	3.8
	2.00	2	7.7
	2.25	3	11.5
	2.75	1	3.8
	3.00	7	26.9
	3.25	2	7.7
	3.50	2	7.7
	3.75	2	7.7
	4.00	3	11.5
	4.25	1	3.8
	4.50	1	3.8
	Total	26	100.0

Table 3: Frequency Distribution of Job

Variables	Freq	Percent
Job Stress		
Low (1 – 2.33)	_____	_____
Moderate (2.34 – 3.66)	_____	_____
High (3.67 – 5)	_____	_____

Reliability Analysis

Reliabillity

- Reliability relates to the quality of measurement.
- In its everyday sense, reliability is the "consistency" or "repeatability" of the study measures
- The extent to which a measure or instrument will yield the same score when administered in different times, locations, or populations

Types of Reliability

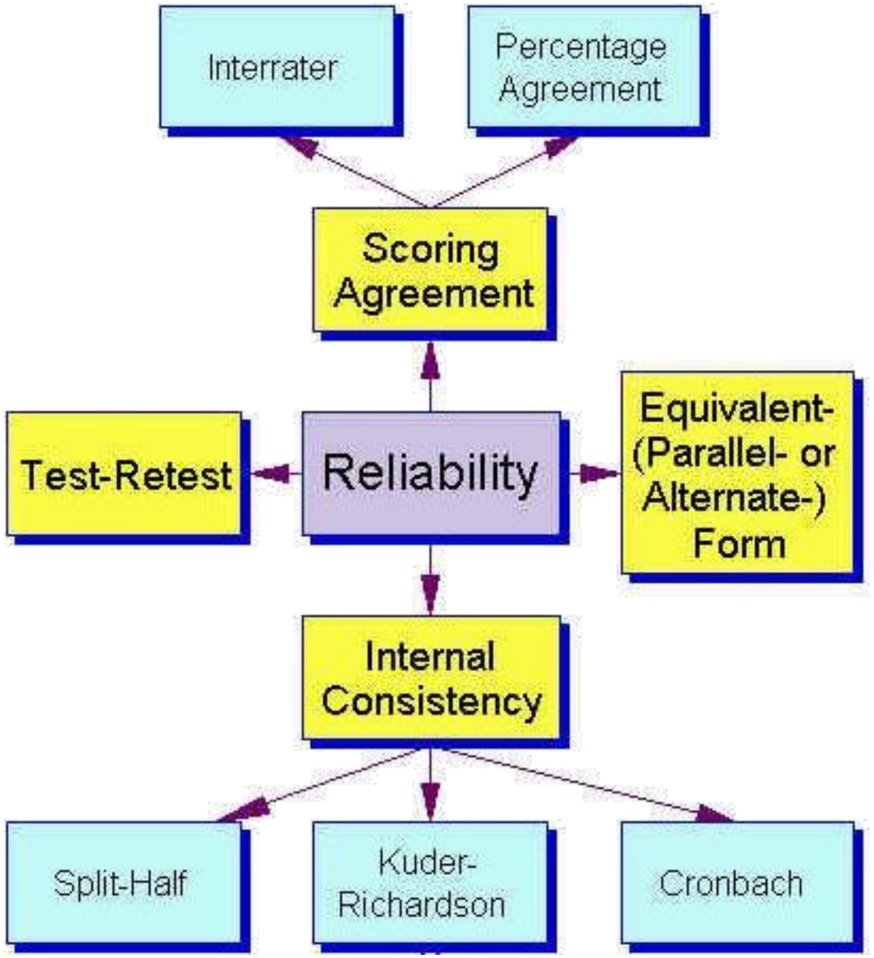
There are four general types of reliability estimates:

1 Inter-Rater Reliability

2 Test-Retest Reliability

3 Parallel-Forms Reliability

4 Internal Consistency Reliability



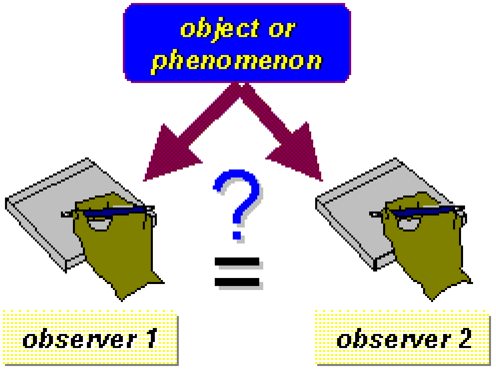
Types of Reliability

NEXT ▷

1 Inter-Rater Reliability

– Determine whether two observers are consistent in their observations

– Inter-rater reliability should be established prior the actual data collection



– For categorical data, use Kappa (a measure of agreement between the raters)

– For continuous data, correlation coefficient is used a measure of reliability

NEXT ▷

Kappa

- Cohen's kappa coefficient is a measure of inter-rater agreement for categorical items
- Kappa ranges between -1 to 1
- Large numbers means better reliability
- Values near 0 suggest that agreement is attributable to chance
- Values < 0 signify that agreement is even less than that which could be attributed to chance
- Most statisticians prefer for Kappa values to be at least .60 and most often > .70 for a good level of agreement

Interpretation

κ	Indicator
< .20	Poor agreement
.20 - .40	Fair agreement
.40 - .60	Moderate agreement
.60 - .80	Good agreement
.80 – 1.00	Very good agreement

[Altman DG (1991). Practical Statistics for Medical Research. London England: Chapman and Hall]

NEXT ▷

Example 1: Kappa

Two raters were requested to rate 50 research projects on a scale of excellent, good and poor. Data are as below:

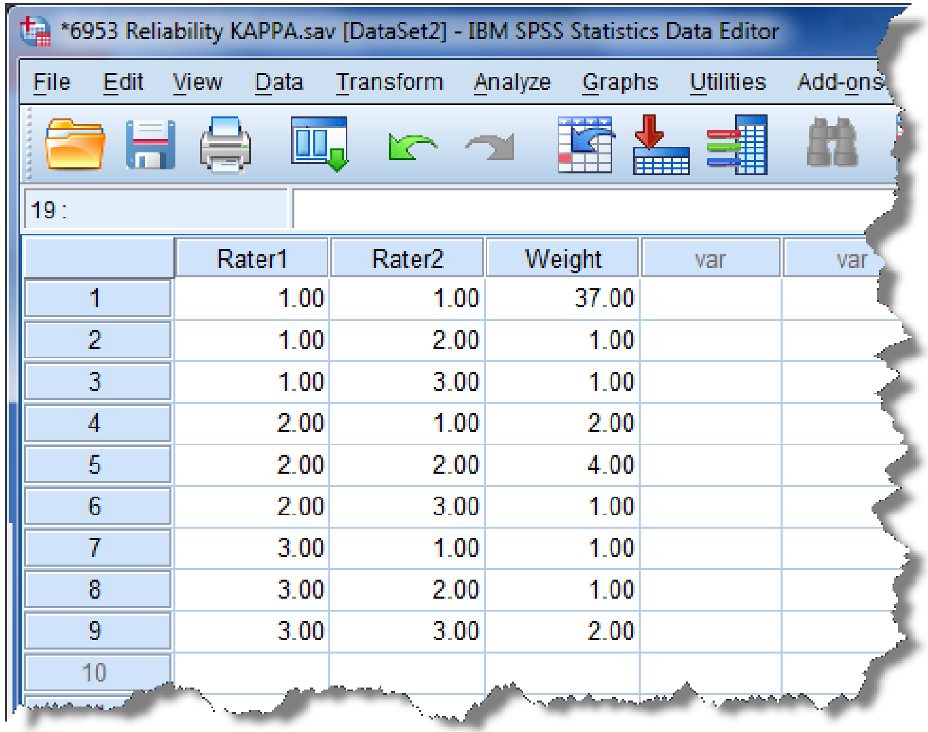
Rater A	Rater B		
	Excellent	Good	Poor
Excellent	37	1	1
Good	2	4	1
Poor	1	1	2

Data: 6953 Reliability KAPPA

NEXT ▷

SPSS Data Editor and Procedures

- Data | Weight Cases – by Count
- Analyze | Descriptive Statistics | Crosstabs Statistics - Kappa



	Rater1	Rater2	Weight	var	var
1	1.00	1.00	37.00		
2	1.00	2.00	1.00		
3	1.00	3.00	1.00		
4	2.00	1.00	2.00		
5	2.00	2.00	4.00		
6	2.00	3.00	1.00		
7	3.00	1.00	1.00		
8	3.00	2.00	1.00		
9	3.00	3.00	2.00		
10					

NEXT ▷

SPSS Results

- Report Kappa, sig. and CI
- 95% CI = Estimate ± 1.96 (SE)
  = .603 ± 1.96 (.124)
  = .360, .846

Symmetric Measures

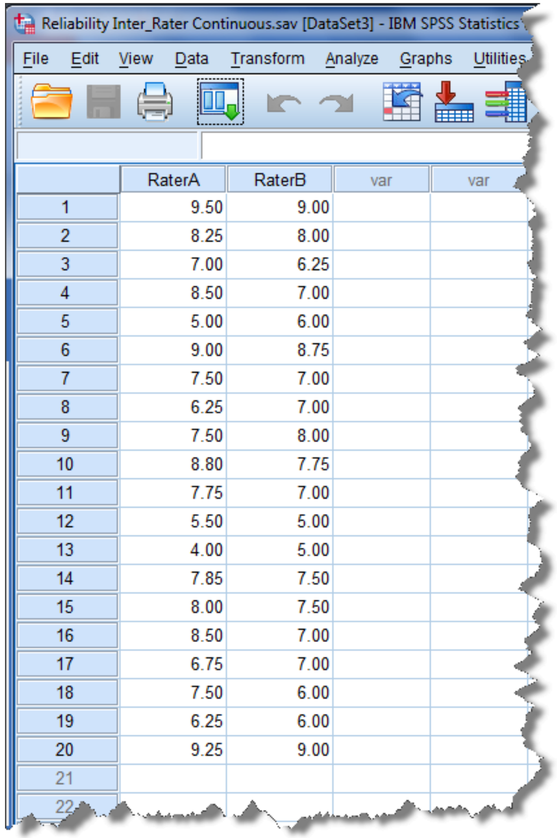
		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Measure of Agreement	Kappa	.603	.124	5.503	.000
N of Valid Cases		50			

a. Not assuming the null hypothesis.
b. Using the asymptotic standard error assuming the null hypothesis.

NEXT ▷

Example 2: Correlation

Two raters were assigned to assess 20 essays written by students. The scores assigned range between 1 (poor) and 10 (excellent)



	RaterA	RaterB	var	var
1	9.50	9.00		
2	8.25	8.00		
3	7.00	6.25		
4	8.50	7.00		
5	5.00	6.00		
6	9.00	8.75		
7	7.50	7.00		
8	6.25	7.00		
9	7.50	8.00		
10	8.80	7.75		
11	7.75	7.00		
12	5.50	5.00		
13	4.00	5.00		
14	7.85	7.50		
15	8.00	7.50		
16	8.50	7.00		
17	6.75	7.00		
18	7.50	6.00		
19	6.25	6.00		
20	9.25	9.00		
21				
22				

Data set: Reliability Inter-Rater Continuous

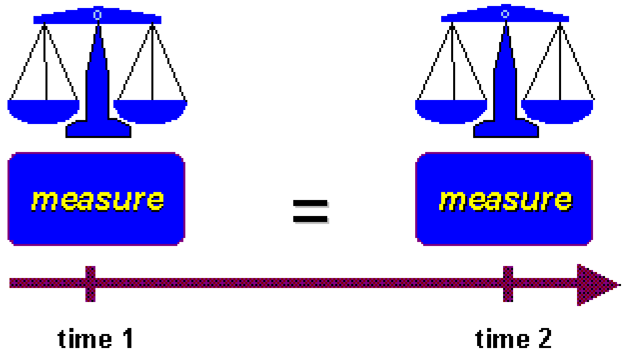
Correlations

		RaterA	RaterB
RaterA	Pearson Correlation	1	.857**
	Sig. (2-tailed)		.000
	N	20	20
RaterB	Pearson Correlation	.857**	1
	Sig. (2-tailed)	.000	
	N	20	20

**. Correlation is significant at the 0.01 level (2-tailed).

2 Test-Retest Reliability

- Administer the same test to the same sample on two different occasions
- Assumes no substantial change in the instrument
- Use correlation to measure estimate of reliability
- This estimate of reliability is affected by time elapses



NEXT ▷

Example 3: Test-Retest

A selected group comprises 15 students was given a test and after a time lapse, the same test was administered to the group.

Variables:

Test1 Scores on Test 1

Test2 Scores on Test 2

Reliability Test_Retest.sav [DataSet4] - IBM SPSS Statistics

File Edit View Data Transform Analyze Graph

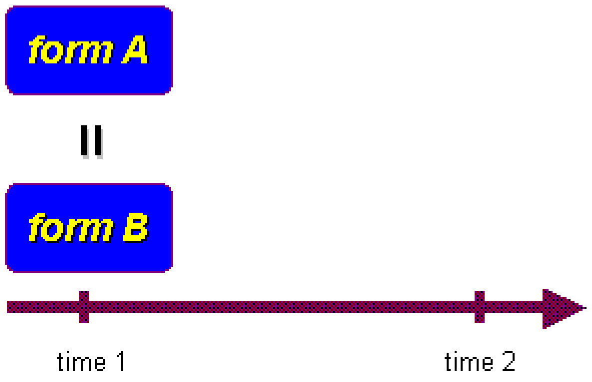
9 :

	Test1	Test2	var
1	16.00	17.00	
2	18.00	18.00	
3	16.00	17.00	
4	17.00	16.00	
5	14.00	15.00	
6	19.00	19.00	
7	12.00	14.00	
8	15.00	16.00	
9	16.00	15.00	
10	18.00	17.00	
11	16.00	17.00	
12	17.00	17.00	
13	19.00	20.00	
14	18.00	19.00	
15	16.00	17.00	
16			

Data set: Reliability Test-Retest

3 Parallel-Forms Reliability

- Prepare two parallel forms to measure a construct
- Administer the instruments to the same group of respondent



- This parallel-forms approach is similar to the split-half reliability
- The major different is the parallel forms can be used independent of each other

NEXT ▷

Example 4: Parallel-Forms

Two set of instruments (forms) were developed to measure perception of patients towards medical services received. These two instruments were administered to a group of patients.

The instruments:

Form A:

ItemA_1 to ItemA_6

Form B:

ItemB_1 to ItemB_6

	FormA_1	FormA_2	FormA_3	FormA_4	FormA_5	FormA_6	FormB_1	FormB_2	FormB_3	FormB_4	FormB_5	FormB_6	FormA
1	3.00	4.00	4.00	3.00	4.00	4.00	4.00	3.00	5.00	3.00	4.00	3.00	3.67
2	4.00	4.00	4.00	5.00	4.00	4.00	4.00	3.00	4.00	5.00	5.00	4.00	4.17
3	5.00	4.00	4.00	4.00	4.00	4.00	5.00	4.00	4.00	5.00	4.00	4.00	4.17
4	4.00	4.00	4.00	5.00	4.00	4.00	4.00	3.00	4.00	5.00	4.00	4.00	4.17
5	5.00	5.00	5.00	5.00	5.00	5.00	5.00	4.00	5.00	4.00	5.00	4.00	5.00
6	4.00	4.00	4.00	5.00	4.00	4.00	4.00	3.00	4.00	5.00	4.00	3.00	4.17
7	3.00	3.00	4.00	3.00	3.00	3.00	3.00	3.00	4.00	4.00	3.00	3.00	3.17
8	4.00	4.00	5.00	4.00	5.00	4.00	4.00	4.00	3.00	5.00	5.00	4.00	4.33
9	5.00	4.00	4.00	4.00	4.00	4.00	5.00	5.00	3.00	5.00	3.00	4.00	4.17
10	4.00	5.00	4.00	5.00	5.00	5.00	4.00	4.00	5.00	5.00	5.00	4.00	4.67
11	4.00	5.00	4.00	5.00	3.00	5.00	4.00	4.00	3.00	5.00	4.00	4.00	4.33
12	4.00	4.00	4.00	4.00	4.00	4.00	3.00	5.00	3.00	4.00	5.00	5.00	4.00
13	5.00	4.00	4.00	4.00	4.00	4.00	5.00	4.00	4.00	3.00	4.00	4.00	4.17
14	4.00	5.00	4.00	4.00	5.00	4.00	4.00	3.00	5.00	4.00	4.00	5.00	4.33
15	5.00	3.00	4.00	5.00	3.00	4.00	5.00	5.00	4.00	3.00	4.00	5.00	4.00
16	4.00	4.00	5.00	4.00	5.00	4.00	4.00	4.00	5.00	5.00	4.00	3.00	4.33
17	3.00	3.00	5.00	4.00	5.00	3.00	4.00	3.00	3.00	4.00	4.00	4.00	3.83
18	4.00	4.00	5.00	4.00	5.00	4.00	5.00	3.00	5.00	4.00	5.00	4.00	4.33
19	5.00	4.00	5.00	5.00	4.00	5.00	5.00	5.00	4.00	4.00	5.00	4.00	4.67
20	4.00	4.00	4.00	4.00	5.00	4.00	4.00	4.00	4.00	4.00	5.00	4.00	4.17

Data set: Reliability Parallel Form

4 Internal Consistency Reliability

- Use single measurement instrument to a group of respondents on one occasion
- Estimate the degree of consistency among the items that make up the instrument/scale
- Two estimates of internal consistency:
  1. Cronbach's alpha
  2. Split-half reliability

NEXT ▷

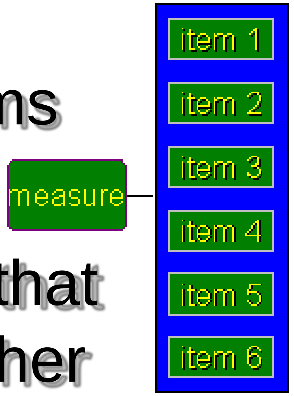
Cronbach's Alpha

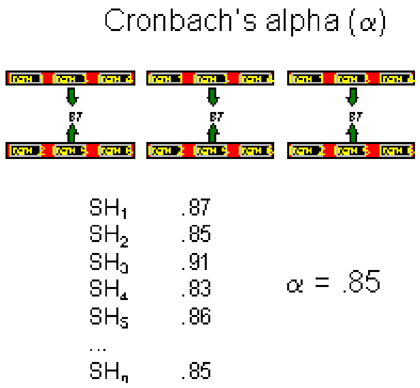
– Cronbach's Alpha coefficient is the most common estimate of internal consistency

– It is mathematically equivalent to the average of all possible split-half estimates

– Kuder-Richardson 20 (KR20) is similar to Cronbach's alpha for dichotomous items

– In social science, the widely-accepted cut-off is that alpha should be .70 or higher





NEXT ▷

Interpretation

Alpha	Indicator
.9 – 1.0	Very good
.8 – .9	Good
.7 – .8	Acceptable
.6 – .7	Questionable
.4 – .6	Weak
< .4	Unacceptable

(George and Mallery, 2001)

NEXT ▷

Example 5: Cronbach Alpha

An instrument was used to measure emotional control. The instrument comprises 10 items using a 5-point Likert like scale (0 to 4)

The list of items:

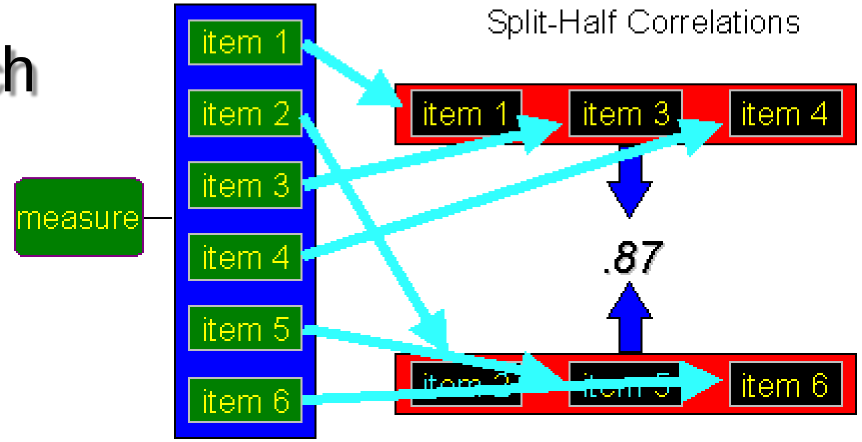
Item1 to Item10

Item4 and Item6 are negative statements. Recode these two items into Item4R and Item6R

Data set: Reliability Cronbach

Split-Half Reliability

- Randomly divide all the items into two sets
- Administer the entire instruments to a sample
- Total scores will be calculated for each set
- A reliability coefficient will be generated which is just the correlation between the two total scores



NEXT ▷

- Important to carefully choose items to include in each half so that the two halves are as equivalent as possible
- Different item splits may produce dramatically different results
- The best split of items is the one that produces equivalent halves

NEXT ▷

Example 6: Split Half

An instrument was used to measure emotional control. The instrument comprises 10 items using a 5-point Likert like scale (0 to 4)

The list of items:

Half 1 – item1, item3, item5, item8, and item10
Half 2 – item2, item4R, item6R, item7, and item9

Data set: Reliability Split Half

END

Data Transformations

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

Next ▷

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

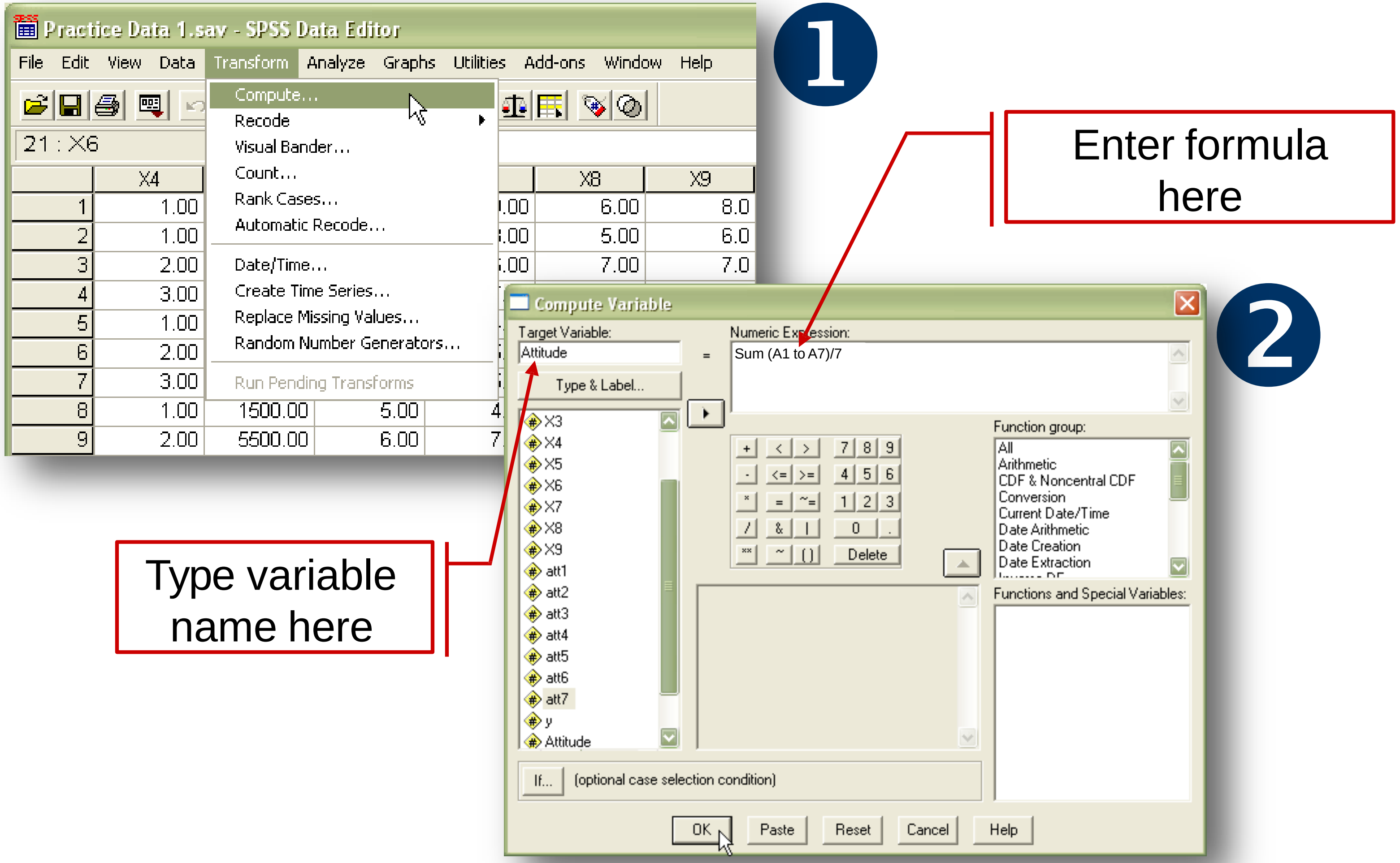
Compute an_income = X5 * 12

Compute Attitude = Mean (A1 to A7)

Variable name

Formula

COMPUTE Procedures:



Recode

Recode

Categorize scores into categories

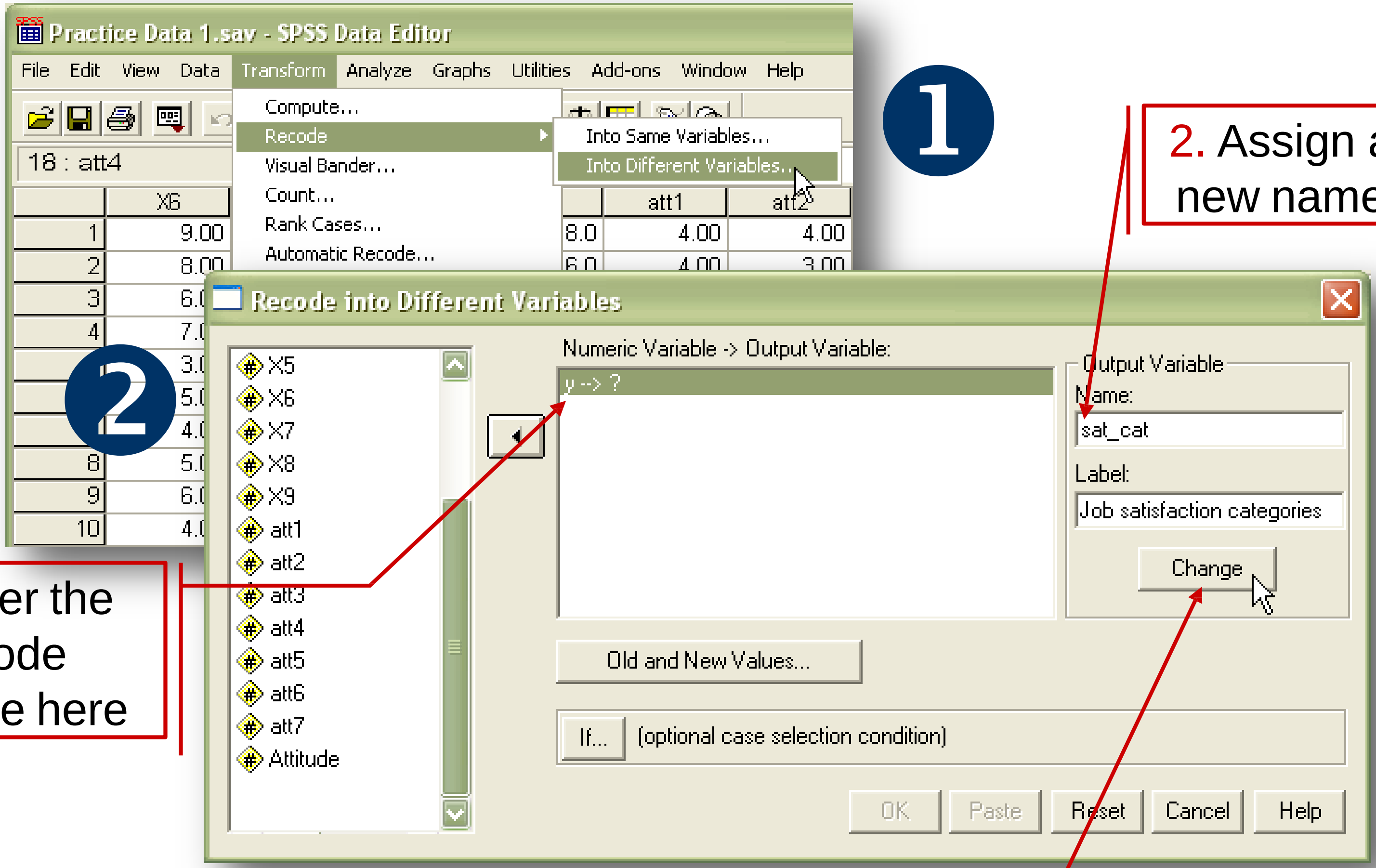
Ex. 1: Recode Y into Sat_cat

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

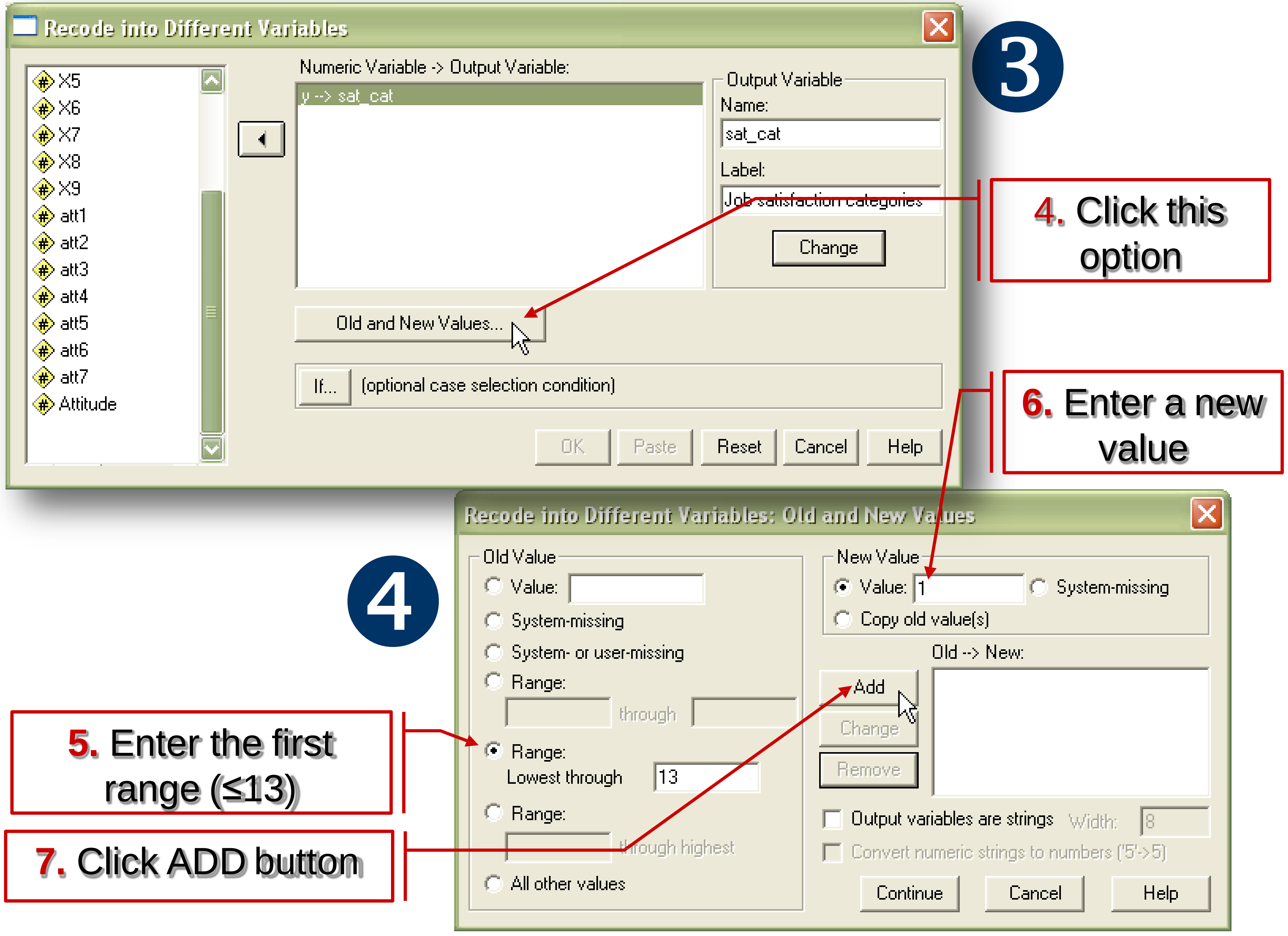
RECODE Procedures:



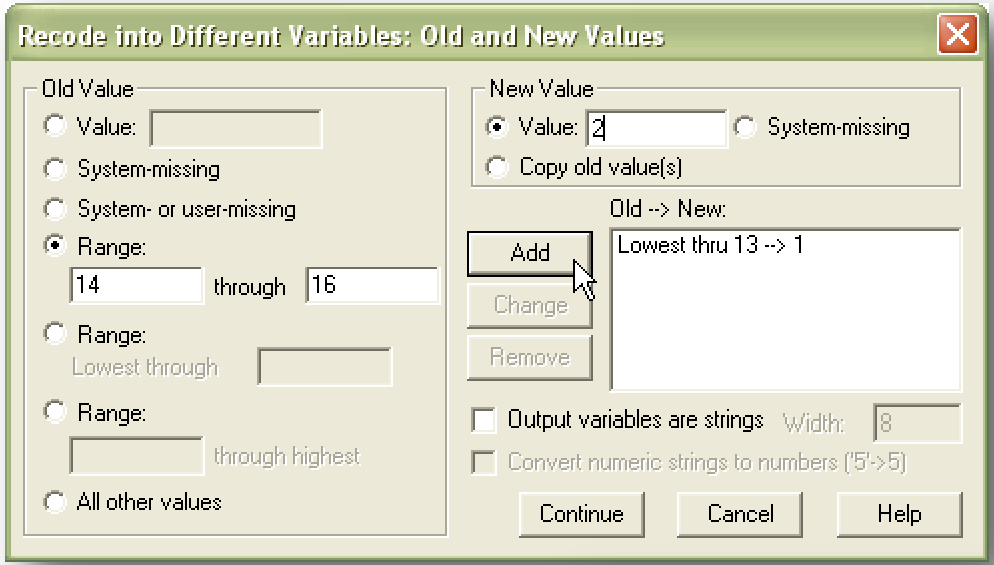
1. Enter the recode variable here

2. Assign a new name

3. Click this button

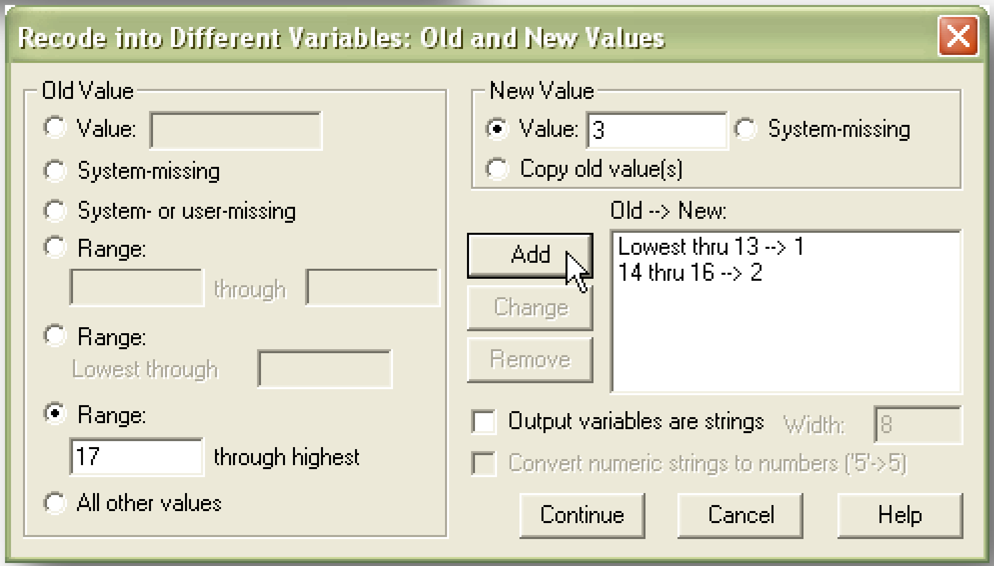


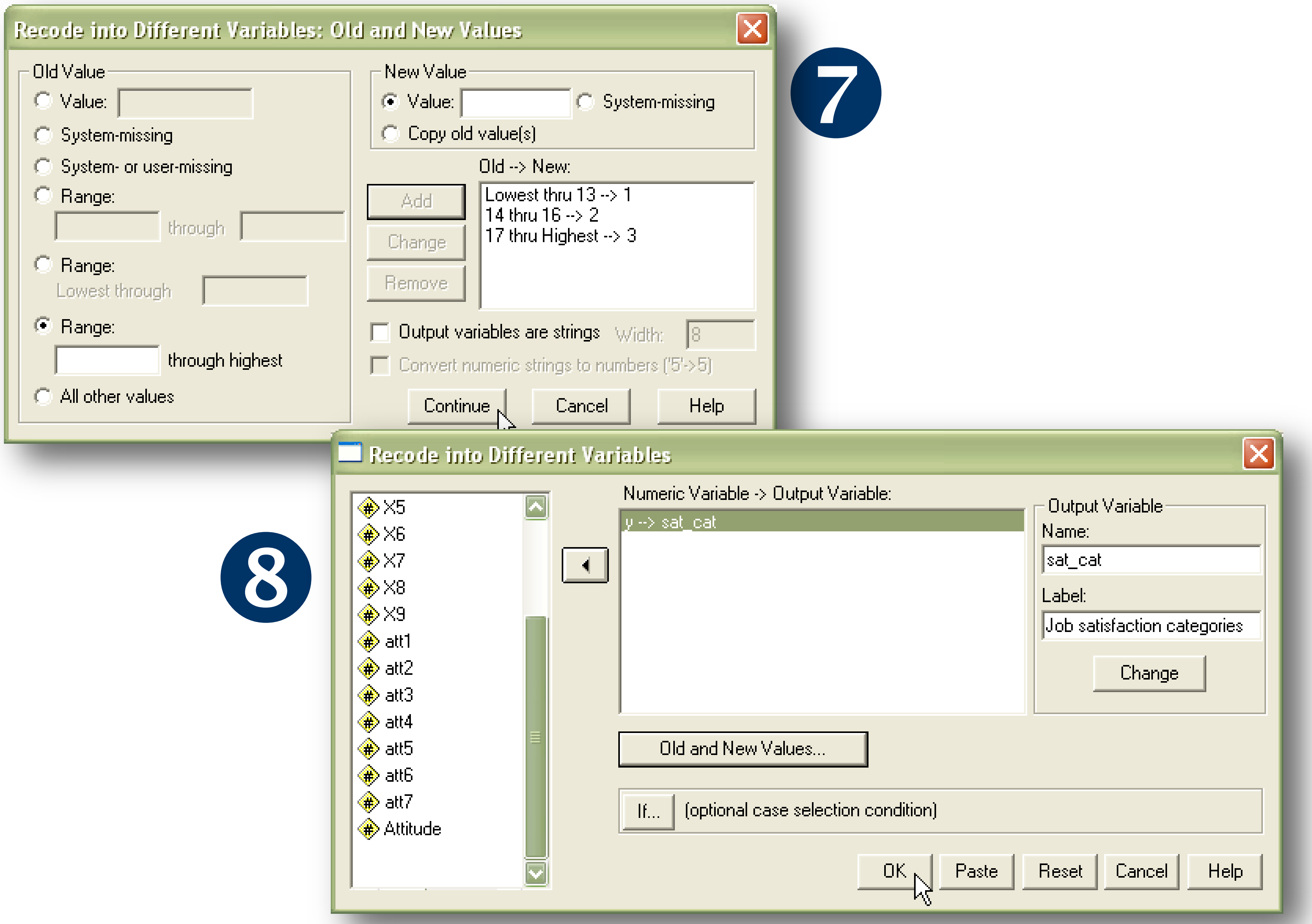
Recode into Different Variables
X5
X6
X7
X8
X9
att1
att2
att3
att4
att5
att6
att7
Attitude
Numeric Variable -> Output Variable:
y --> sat_cat
Output Variable
Name:
sat_cat
Label:
Job satisfaction categories
Change
Old and New Values...
If...
(optional case selection condition)
OK
Paste
Reset
Cancel
Help
3
4. Click this option
6. Enter a new value
Recode into Different Variables: Old and New Values
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:
8
Convert numeric strings to numbers ('5'->5)
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: 8
Convert numeric strings to numbers ('5'->5)
Continue
Cancel
Help

5

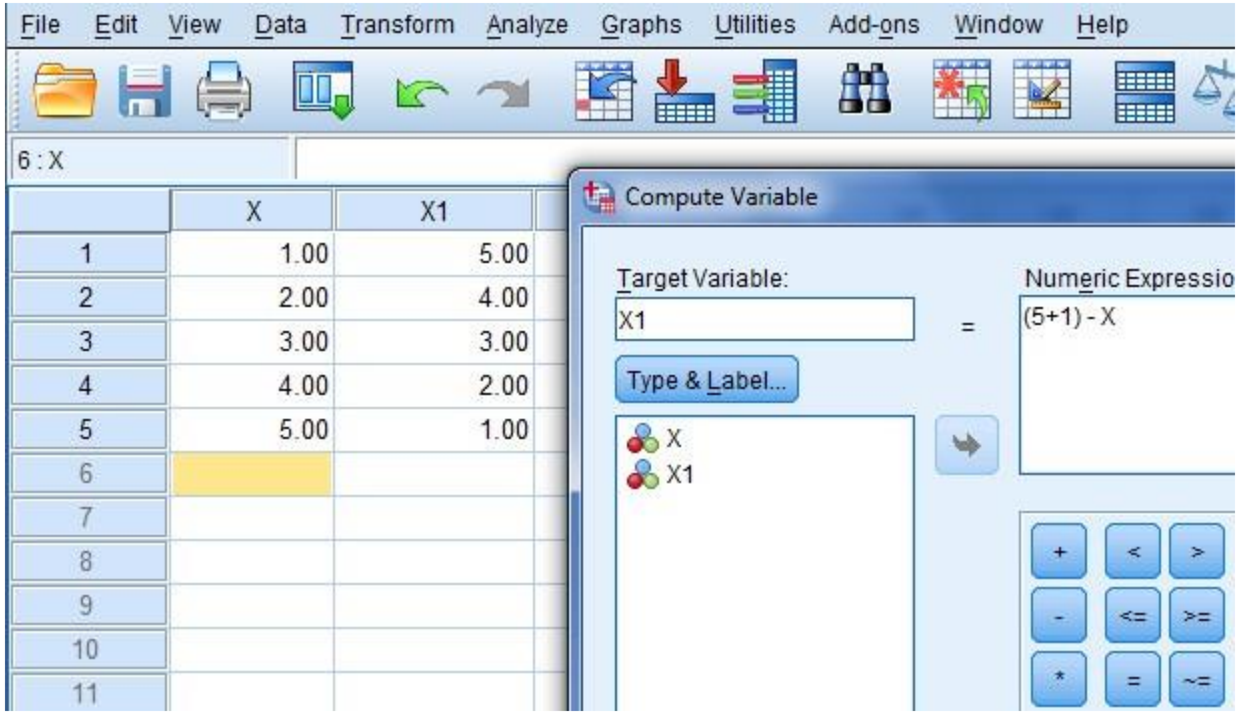
6


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: 8
Convert numeric strings to numbers ('5'->5)
Continue
Cancel
Help


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:
Add
Change
Remove
Lowest thru 13 --> 1
14 thru 16 --> 2
17 thru Highest --> 3
Output variables are strings
Width:
8
Convert numeric strings to numbers ('5'->5)
Continue
Cancel
Help
8
Recode into Different Variables
X5
X6
X7
X8
X9
att1
att2
att3
att4
att5
att6
att7
Attitude
Numeric Variable -> Output Variable:
y --> 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

Recode

Reverse coding



Application Exercises

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)	____	____		

end

Test of Normality

Normality Test

- One of the major assumption for parametric statistics is data in the population must be normally distributed
- How to check whether your data meet the above assumption?
- Use Exploratory Data Analysis (EDA) in SPSS
- SPSS provides two statistics:
  1. Kolmogorov-Smirnov
  2. Shapiro-Wilk

- You data meet the assumption of normality
  - If the sig-value > alpha (.05)
- In addition, SPSS also produces Normality Plots:
  - Normal Q-Q Plot
  - Detrended Normal Q-Q Plot
- You data can be considered to be normally distributed
  - If majority of the points in the Detrended Normal Q-Q plot are within -.3 and +.3
- Data can be considered normal if skewness is between -1 and +1. However values between ±2 are in many cases acceptable (George, D and Mallery, P,2005) and Pallant (2001)*

Skew is the tilt (or lack of it) in a distribution. The more common type is right skew, where the tail points to the right. Less common is left skew, where the tail is points left. A common rule-of-thumb test for normality is to run descriptive statistics to get skewness and kurtosis, then divide these by the standard errors. Skew should be within the +2 to -2 range when the data are normally distributed. Some authors use +1 to -1 as a more stringent criterion when normality is critical.

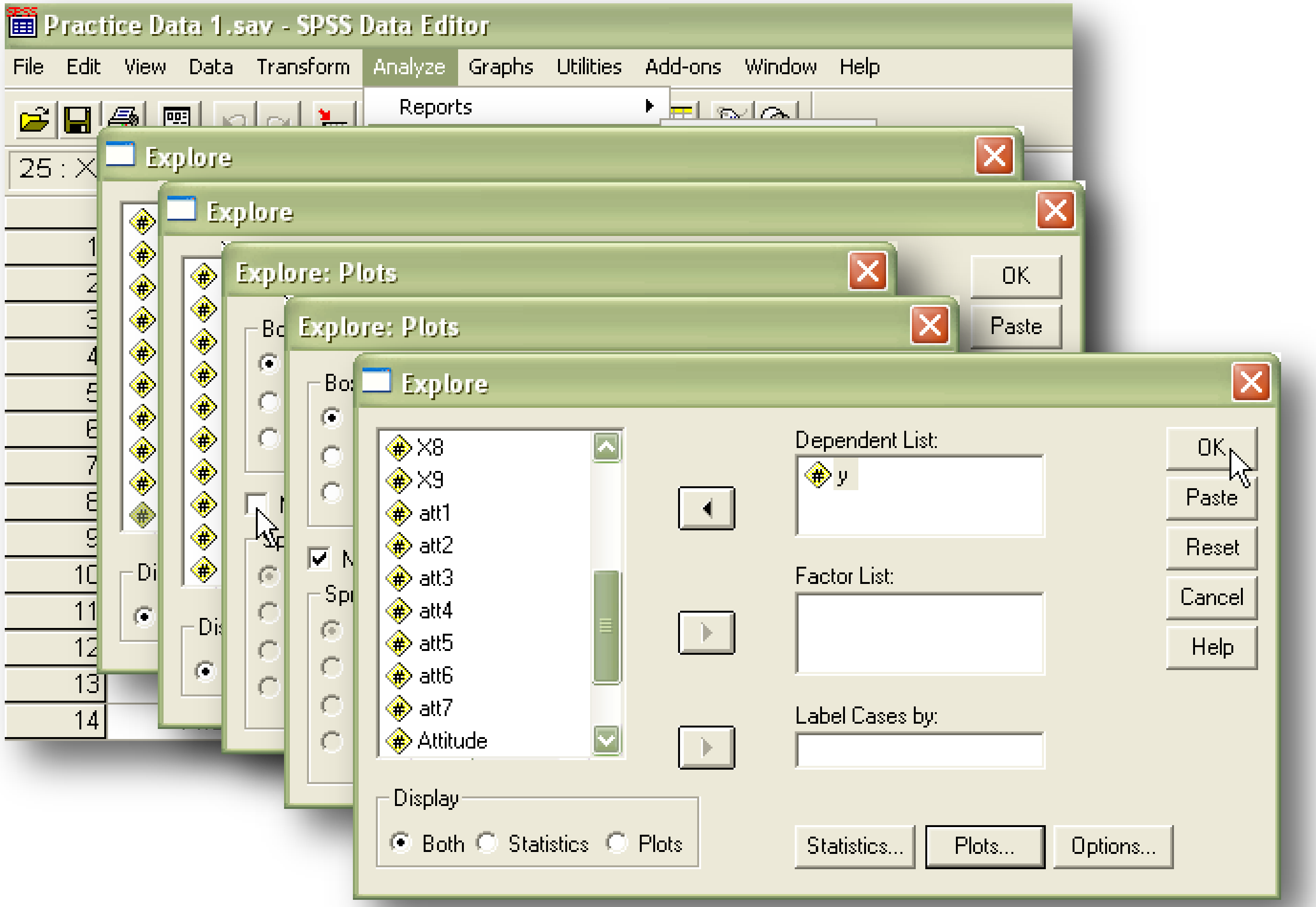
http://faculty.chass.ncsu.edu/garson/PA765/assumpt.htm#normal

Skewness. The question arises in statistical analysis of deciding how skewed a distribution can be before it is considered a problem. One way of determining if the degree of skewness is "significantly skewed" is to compare the numerical value for "Skewness" with twice the "Standard Error of Skewness" and include the range **from minus twice** the Std. Error of Skewness **to plus twice** the Std. Error of Skewness. If the value for Skewness falls within this range, the skewness is considered not seriously violated.
For example, from the above, twice the Std. Error of Skewness is 2 X .183 = .366. We now look at the range from �0.366 to + .366 and check whether the value for Skewness falls within this range. If it does we can consider the distribution to be approximately normal. If it doesn�t (as here), we conclude that the distribution is significantly non-normal and in this case is significantly positvely skewed.

http://www.une.edu.au/WebStat/unit_materials/c4_descriptive_statistics/determine_skew_kurt.html

Kurtosis..
The same numerical process can be used to check if the kurtosis is significantly non normal. A normal distribution will have Kurtosis value of zero. So again we construct a range of "normality" by multiplying the Std. Error of Kurtosis by 2 and going from minus that value to plus that value. Here 2 X .363 = .726 and we consider the range from �0.726 to + 0.726 and check if the value for Kurtosis falls within this range. Here it doesn�t (12.778), so this distribution is also significantly non normal in terms of Kurtosis (leptokurtic).

http://www.une.edu.au/WebStat/unit_materials/c4_descriptive_statistics/determine_skew_kurt.html



Practice Data 1.sav - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Reports
Explore
Explore
Explore: Plots
Explore: Plots
OK
Paste
Explore
X8
X9
att1
att2
att3
att4
att5
att6
att7
Attitude
Dependent List:
y
Factor List:
Label Cases by:
OK
Paste
Reset
Cancel
Help
Display
Both
Statistics
Plots
Statistics...
Plots...
Options...

Descriptives

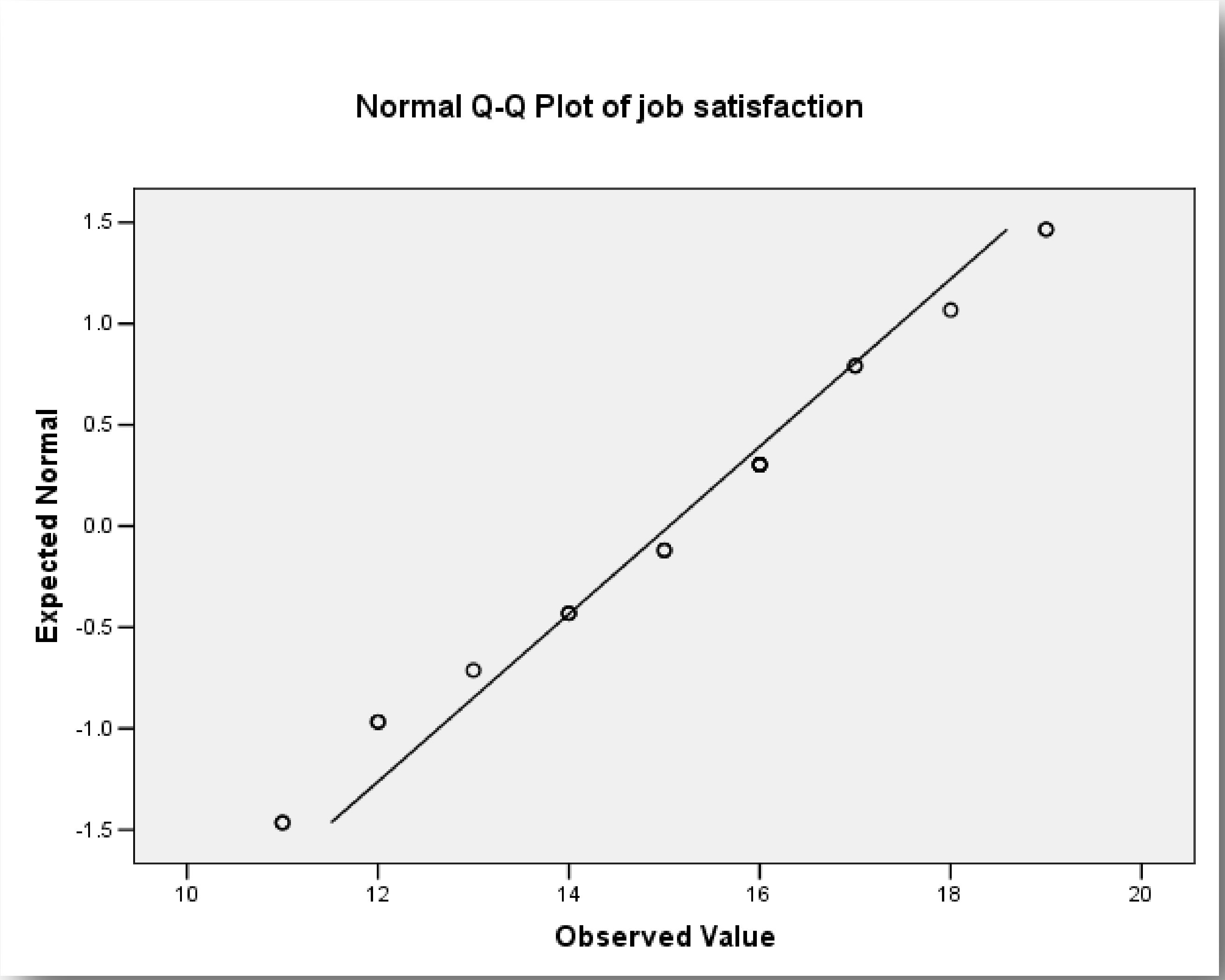
			Statistic	Std. Error
y	Mean		15.0500	.54035
	95% Confidence Interval for Mean	Lower Bound	13.9190	
		Upper Bound	16.1810	
	5% Trimmed Mean		15.0556	
	Median		15.5000	
	Variance		5.839	
	Std. Deviation		2.41650	
	Minimum		11.00	
	Maximum		19.00	
	Range		8.00	
	Interquartile Range		3.50	
	Skewness		-.139	.512
	Kurtosis		-.726	.992

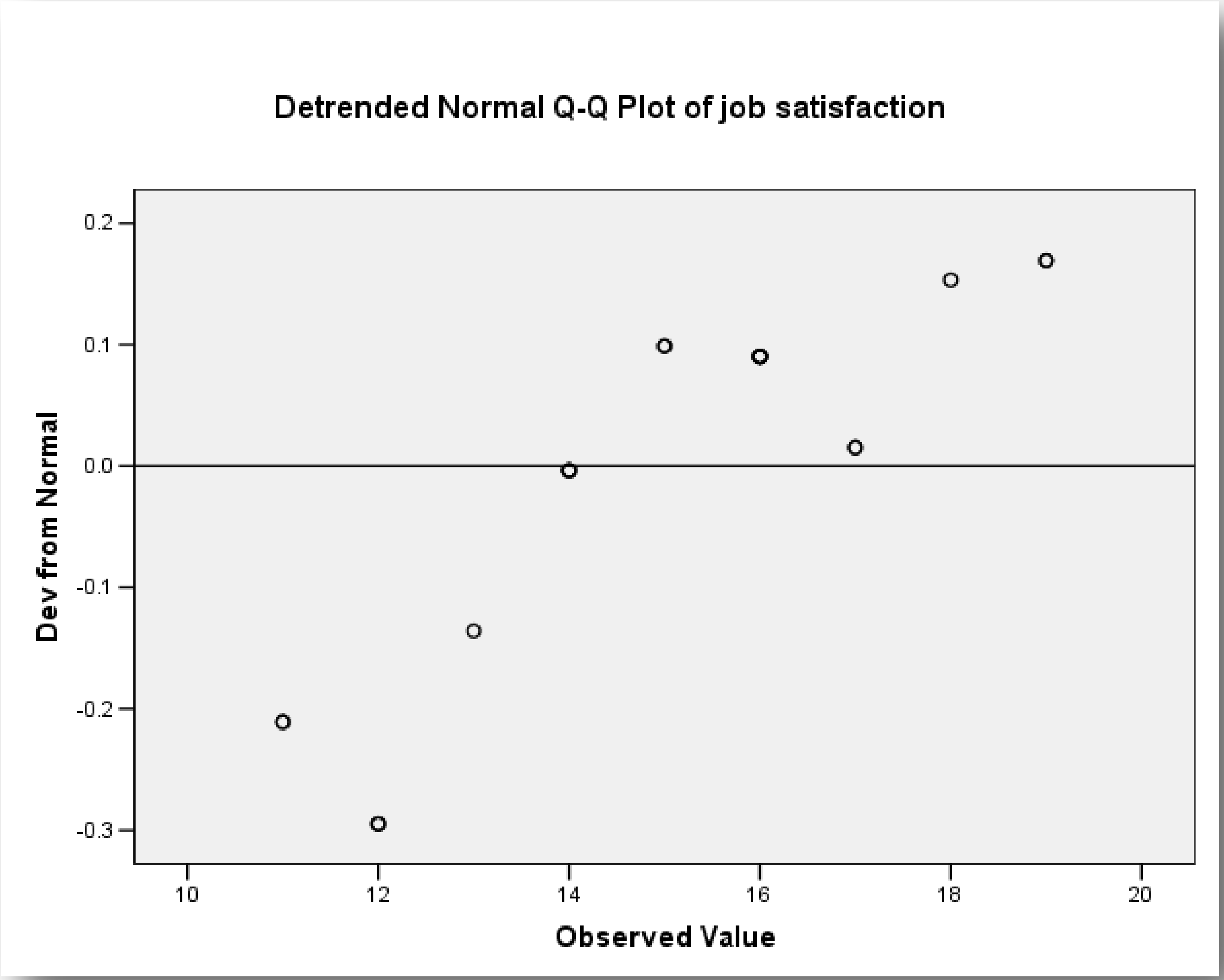
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
y	.153	20	.200*	.952	20	.406

*. This is a lower bound of the true signif icance.

a. Lillief ors Signif icance Correction


Normal Q-Q Plot of job satisfaction
Expected Normal
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
10
12
14
16
18
20
Observed Value


Detrended Normal Q-Q Plot of job satisfaction
0.2
0.1
0.0
-0.1
-0.2
-0.3
Dev from Normal
10
12
14
16
18
20
Observed Value

Application Exercises

Data Set 3:

The above data set comprises the following variables:

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

Question

Test the normality assumption of the following variables:

- Support
- Work
- Motive
- Perform

State your conclusion and justify your answer

Table 1: Normality Test of Study Instruments

Instrument	Kolmogorov	*p*
Support from Peers	____	___
Work environment	____	___
Motivation	____	___
Job Performance	____	___

Basic Statistics

Objectives

Participants to be able to:

1. Run frequency procedure
2. Extract relevant information to be presented in appropriate presentation mode
3. Prepare tables and charts

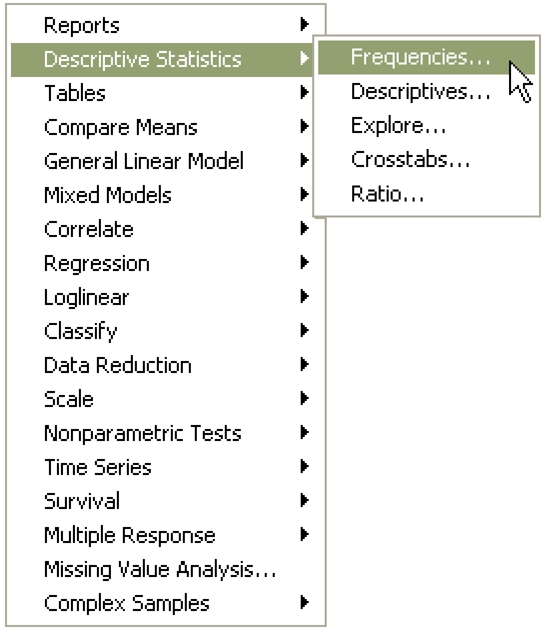
Frequencies

Frequency is an SPSS procedure to obtain:

– Frequency distribution

– Percentage distribution

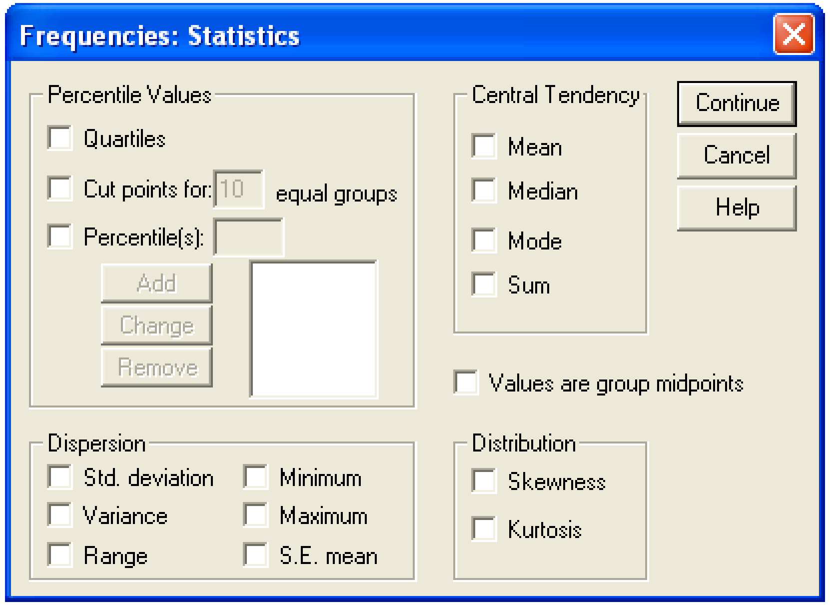
– Basic statistics



Basic Statistics

Statistics option in Frequency procedure provides the following statistics:

- Percentile Values
- Central Tendency
- Dispersion
- Distribution

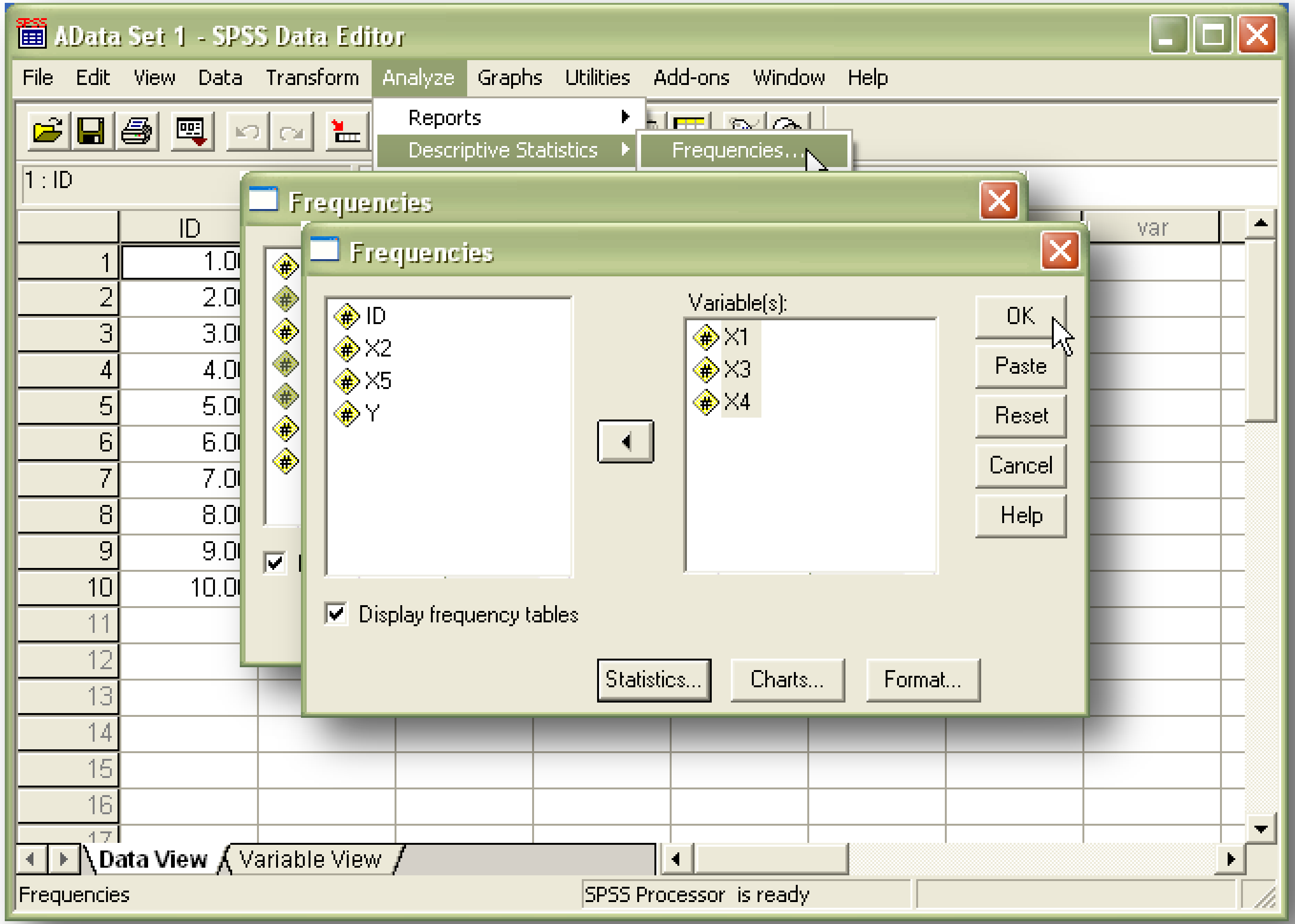


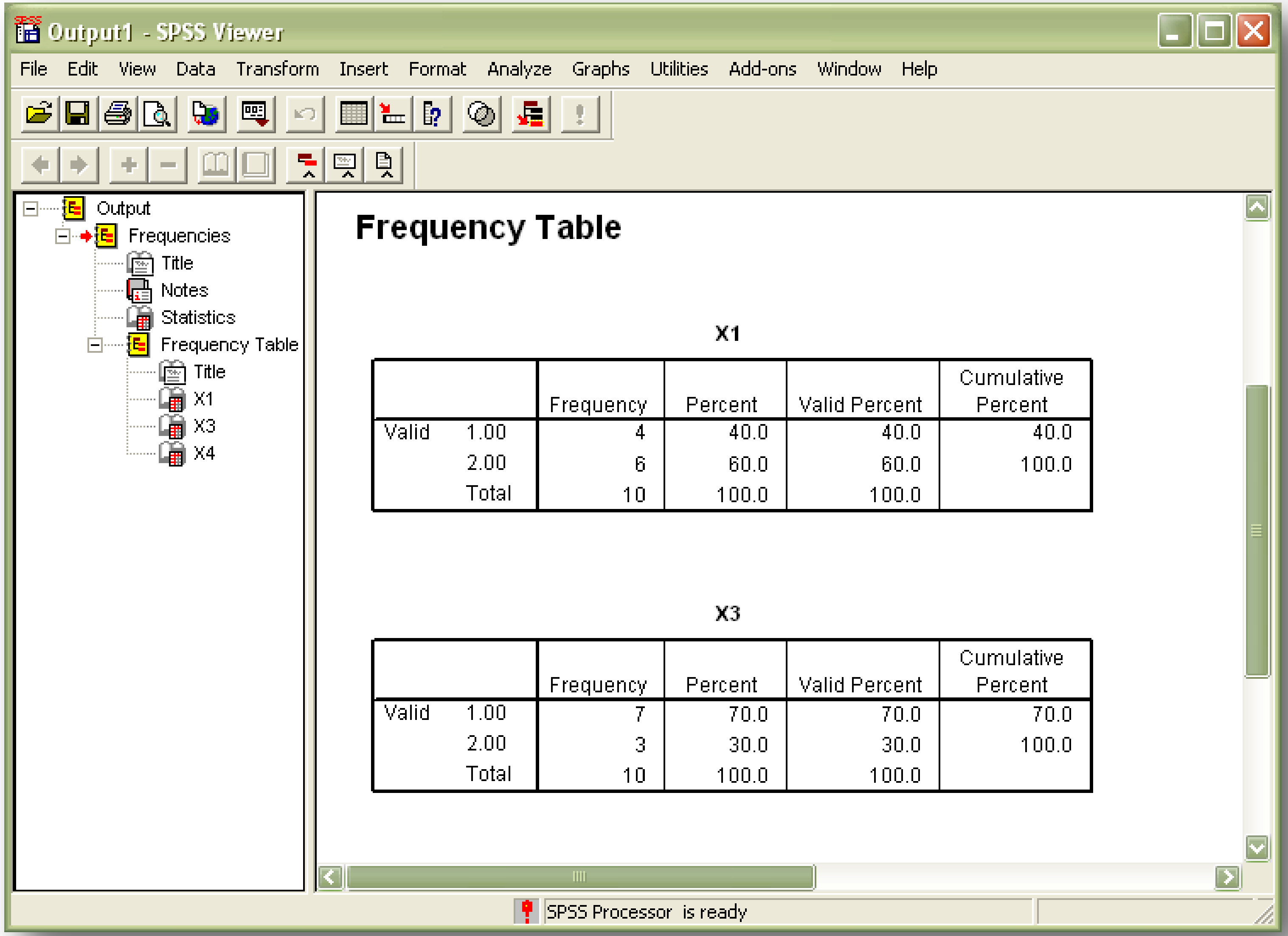
working Examples

In this example, you will be using the Practice Data. Run the Frequencies procedure, extract and present the results in the given tables.

1. Run Frequencies for Gender (X1), Marital status (X3) dan Job Category (X4)
2. Run Frequencies for Age (X2), Tenure (X5), Job Commitment (X6) and Job Performance (Y). Request for Mean and Std. deviation

Frequencies: X1, X3 and X4





Frequency Table

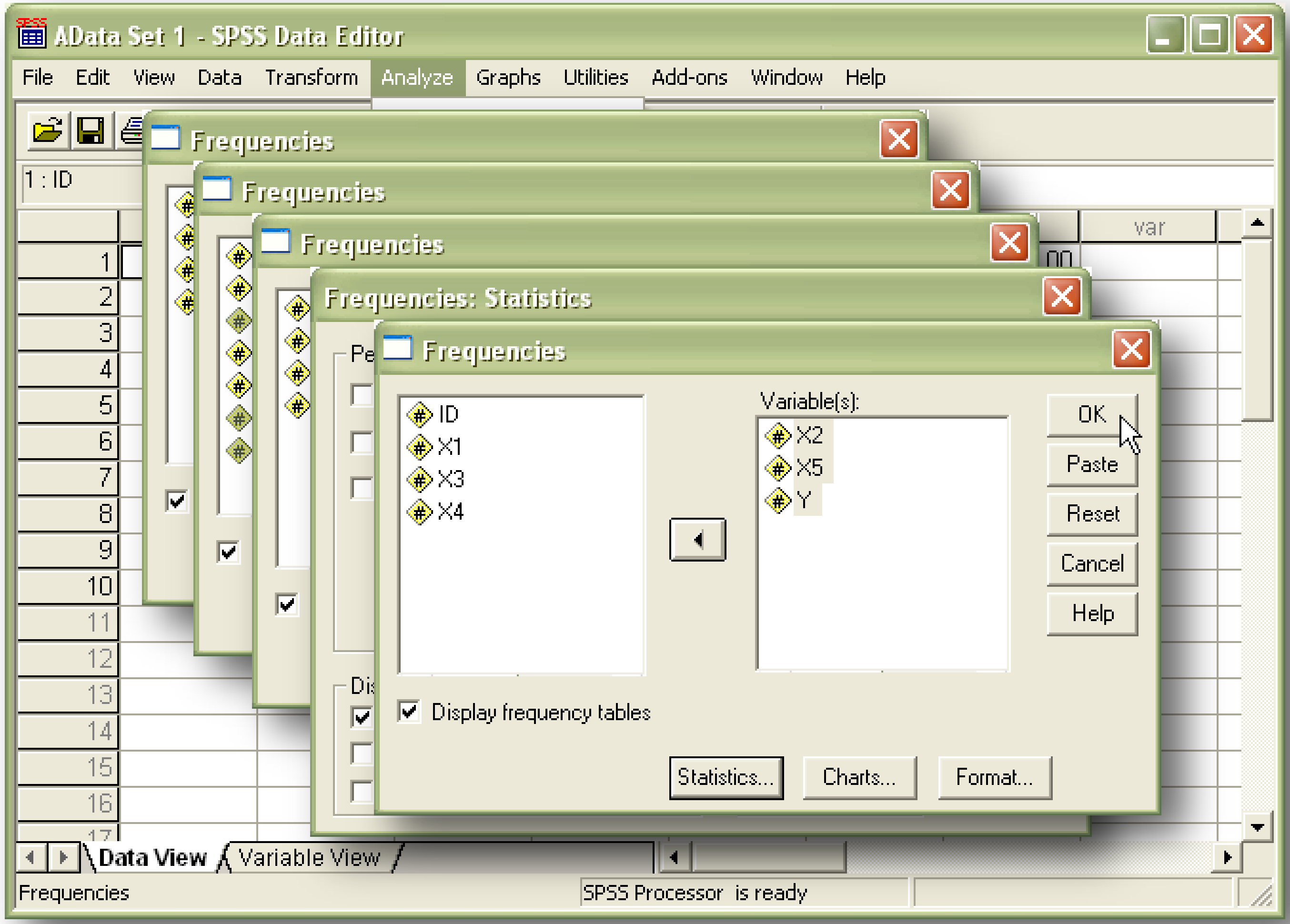
X1

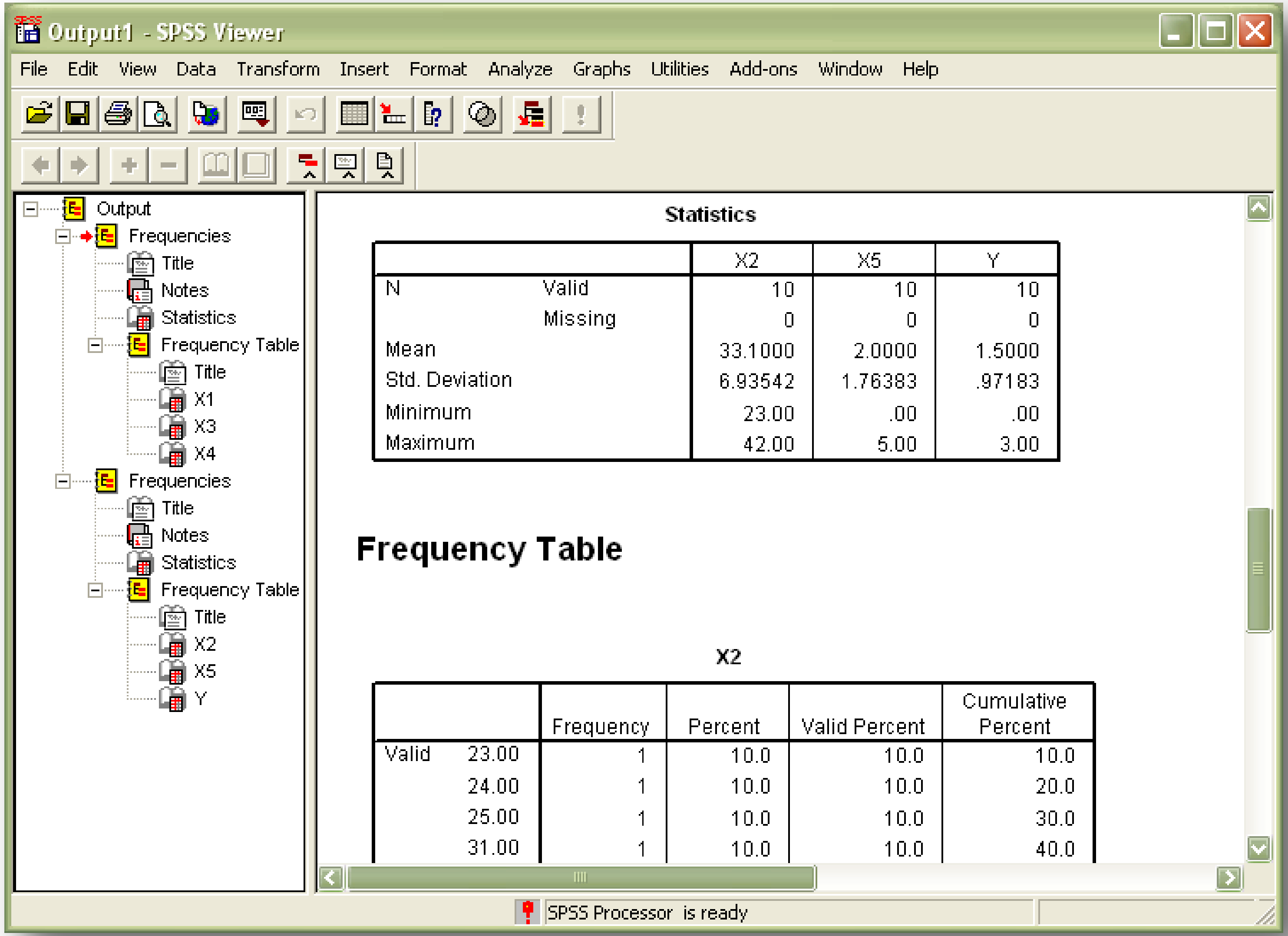
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	4	40.0	40.0	40.0
	2.00	6	60.0	60.0	100.0
	Total	10	100.0	100.0	

X3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	7	70.0	70.0	70.0
	2.00	3	30.0	30.0	100.0
	Total	10	100.0	100.0	

Frequencies: X2, X5 and Y




Output1 - SPSS Viewer
File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help
Output
Frequencies
Title
Notes
Statistics
Frequency Table
Title
X1
X3
X4
Frequencies
Title
Notes
Statistics
Frequency Table
Title
X2
X5
Y
Statistics
X2 X5 Y
N Valid 10 10 10
Missing 0 0 0
Mean 33.1000 2.0000 1.5000
Std. Deviation 6.93542 1.76383 .97183
Minimum 23.00 .00 .00
Maximum 42.00 5.00 3.00
Frequency Table
X2
Frequency Percent Valid Percent Cumulative Percent
Valid 23.00 1 10.0 10.0 10.0
24.00 1 10.0 10.0 20.0
25.00 1 10.0 10.0 30.0
31.00 1 10.0 10.0 40.0
SPSS Processor is ready

Answer:

Table 1: Gender, Marital Status and Job Categories

Variables	Freq	%
Gender		
Male		
Female		
Marital Status		
Married		
Widowed		
Bachelor		
Job Categories		
Support		
Clerical		
Administrator		

Table 2: Age, and Tenure

Variable	Freq	%	Mean	SD
Age (years)			___	___
< 30	___	___		
30 – 40	___	___		
> 40	___	___		
Tenure (years)			___	___
1 – 3	___	___		
4 – 6	___	___		
> 6	___	___		

Table 3: Job Commitment and Performance

Variable	Freq	%	Mean	SD
Job Commitment			___	___
Low (1 – 3)	___	___		
Moderate (4 – 6)	___	___		
High (7 -- 9)	___	___		
Job Performance			___	___
Low (6 – 13)	___	___		
Moderate (14 – 22)	___	___		
High (23 – 30)	___	___		

Application Exercises

Use the QWL Data and run Frequencies for the following variables and present the results in the given tables.

1. Frequencies for Marital status (X3) Job category (X4)
2. Run Frequencies for Peer Support (X7), Attitude, and Quality of Work Life (QWL). Request for Mean, Std. deviation, Minimum and Maximum

Table 4: Marital status and Job Categories

Variables	Freq	%
Marital Status		
Married	___	___
Divorced	___	___
Bachelor	___	___
Job Categories		
Support	___	___
Administrator	___	___
Management	___	___

Table 5: Peer Support, Attitude and QWL

Variable	Freq	%	Mean	SD
Peer Support			___	___
Low (≤ 3)	___	___		
Moderate (4-6)	___	___		
High (> 6)	___	___		
Attitude			___	___
Low (1 - 2.33)	___	___		
Moderate (2.34 - 3.66)	___	___		
High (3.67 - 5)	___	___		
Quality of Work Life			___	___
Low (1 - 2.33)	___	___		
Moderate (2.34 - 3.66)	___	___		
High (3.67 - 5)	___	___		

Hypothesis Testing

Objectives

Participants to be able to:

❶ Define hypothesis

❷ Name two (2) types of hypotheses

❸ List five (5) steps in hypothesis testing

❹ Define criteria in making decision
  - Manual calculation
  - SPSS

❺ Describe two (2) types of errors

Research Concerns/Objectives

Major Research Concerns

❶ **Describe Phenomenon**

- Frequency/Percent
- MCT
- MD

❷ **Comparison between Groups**

- T Test
- ANOVA
- Mann-Whitney
- Kruskal Wallis

❸ **Relationship between Variables**

- Chi square
- Spearman rank correlation
- Pearson PM correlation
- Regression Analysis

Comparing Group Difference

Are you interested to prove:

→ No difference $\mu_1 = \mu_2$

→ There is a difference $\mu_1 \neq \mu_2$

?

Hypothesis:

→ H_O Null hypothesis

→ H_A Alternative hypothesis

Hypothesis Testing

- Hypothesis refers to educated guess or assumption to be tested
- Hypothesis is formulated following the review of related literature, prior to the execution of the study
- Setting up and testing hypotheses is an essential part of statistical inference
- Types of hypotheses:
  1. Research hypothesis - H_A
  2. Null hypothesis - H_O

Characteristics

- A major characteristic of a good research hypothesis is that it is consistent with previous research
- A good hypothesis is a tentative, reasonable explanation for the occurrence of certain behaviors, phenomena, or events
- A good hypothesis states as clearly and concisely as possible the expected relationship or difference between two variables
- A well-stated and defined hypothesis must be testable

Research Hypothesis – H_A

- Also known as alternative hypothesis
- A statement of what a statistical hypothesis test is set up to establish
- In an experiment, the alternative hypothesis might be that the new teaching method has a different effect, on average, compared to that of the current method
- Or the alternative hypothesis might also be that the new method is better, on average, than the current method

Null Hypothesis – H_0

- Also known as hypothesis of NO DIFFERENCE or NO RELATIONSHIP

  Example: H_O: $\mu_1 = \mu_2$

  H_O: $\rho = 0$

- Used to facilitate testing of the research hypothesis

  ~~H_O: $\mu_1 = \mu_2$~~

  H_A: $\mu_1 \neq \mu_2$

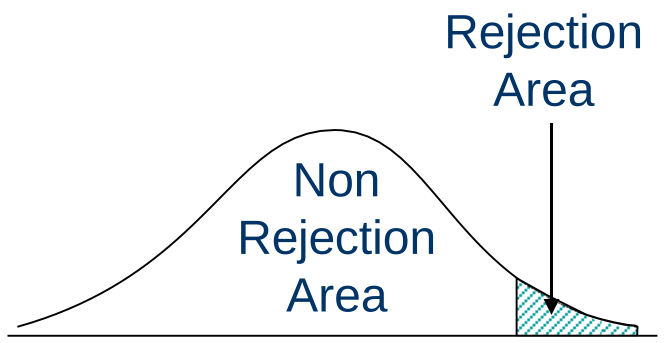
The logic:
It is difficult to prove something to be TRUE but is much easier to prove something to be NOT TRUE

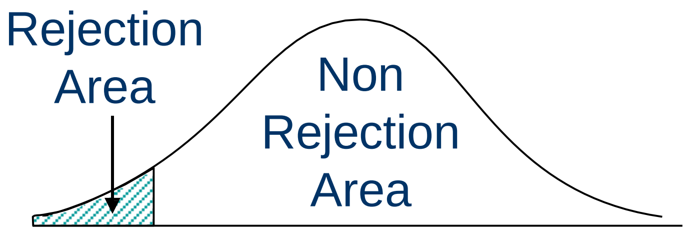
Types of Hypothesis

❶ One-tailed (directional)

H_A: $\rho > 0$

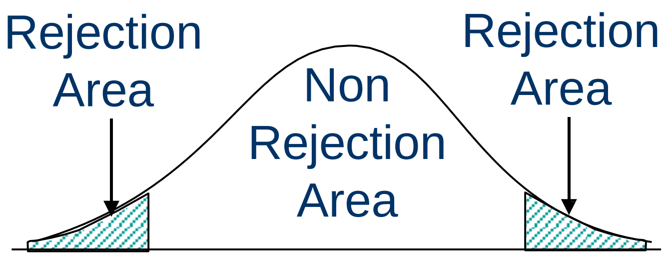
H_A: $\rho < 0$





❷ Two-tailed (non directional)

H_A: $\rho \neq 0$



Steps in Hypothesis Testing

1. State the null and alternative hypotheses
2. Set the confident/alpha level
3. Report test statistic and sig. values
4. Make decision
5. Conclusion

State Hypothesis

Comparison bet. groups:

$H_O: \mu_1 = \mu_2$

$H_A: \mu_1 \neq \mu_2$ ← Two-tailed

$\mu_1 > \mu_2$ ← One-tailed (More than)

$\mu_1 < \mu_2$ ← One-tailed (Less than)

Relationship bet. variables:

$H_O: \rho = 0$

$H_A: \rho \neq 0$ ← Two-tailed

$\rho > 0$ ← One-tailed (Positive)

$\rho < 0$ ← One-tailed (Negative)

Set Confidence Level

Generally, in social science studies, alpha is set at .05

◁ Back

Report Test Stat and sig.

From the SPSS output, report:

- Value of the test statistic
- Sig-value (*p*)

1. Degrees of freedom – *df*
2. Confidence level (α)
   By convention, in social science
   α = .05

◁ Back

Make Decision

Decision Criteria

SPSS

Reject H_O: sig-value < α
Fail to reject H_O: sig-value ≥ α

Criteria	*Decision*
Sig-$F < \alpha$	Reject H_O
Sig-$F \geq \alpha$	Fail to reject H_O

◁ Back

Conclusion

Relate to the hypothesis

If Reject H_O: Significant Difference/ Relationship

If Fail to Reject H_O: No Significant Difference/ Relationship

Types of Errors

Hypothesis

Decision	H_O True	H_O False
Reject H_O	Type I Error	Correct Decision
Fail to Reject H_O	Correct Decision	Type II Error

Type I Error

- Type I error occurs when the null hypothesis is rejected when it is in fact true; that is, H_O is wrongly rejected
- A type I error is often considered to be more serious, and therefore more important to avoid, than a type II error
- The probability of a type I error can be precisely computed as:

P(type I error) = significance level = alpha (α)

Type II Error

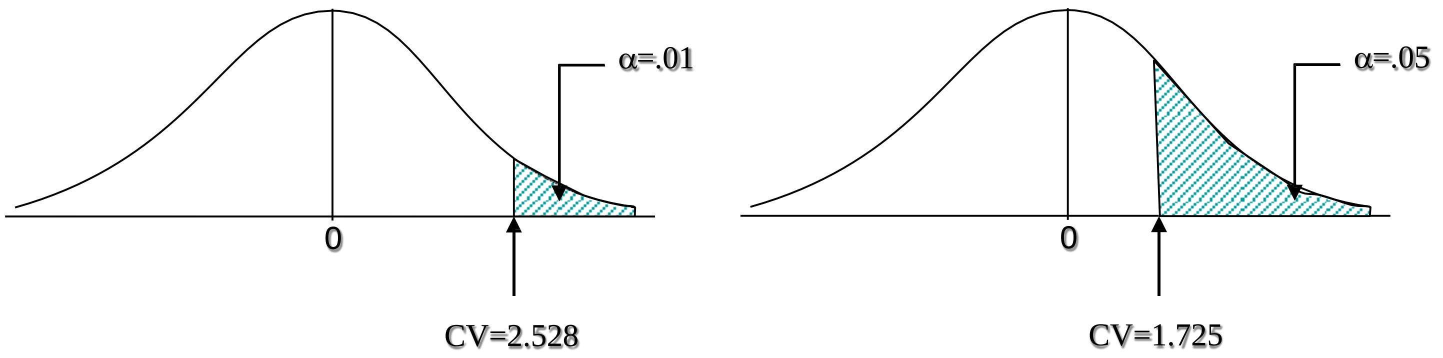
- A type II error occurs when the null hypothesis is not rejected when it is in fact false
- A type II error is frequently due to sample sizes being too small.
- The probability of a type II error is symbolized as beta

P(type II error) = beta (but is generally unknown).

1 General Observation

Reference:
Statistic: *t*
df: 20

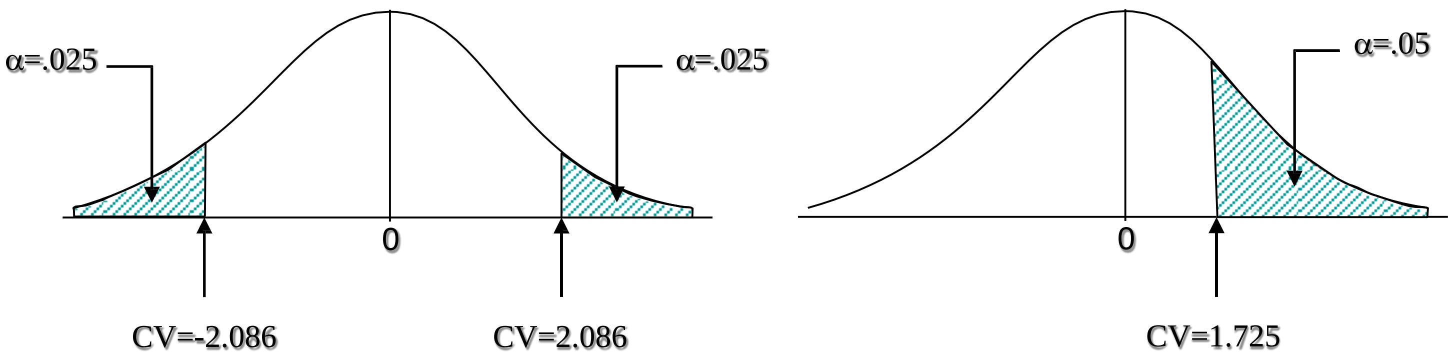
- If you reject H_O at .01, you will SURELY reject the H_O at .05
- If you reject H_O at .05, you may OR may not reject the H_O at .01



2 General Observation

Reference:
Statistic: *t*
$\alpha = .05$
df = 20

- If you reject an H_o at a two-tailed test, you will SURELY reject the H_o at a one-tailed test
- If you reject an H_o at a one-tailed test, you may OR may not reject the H_o at a two-tailed test

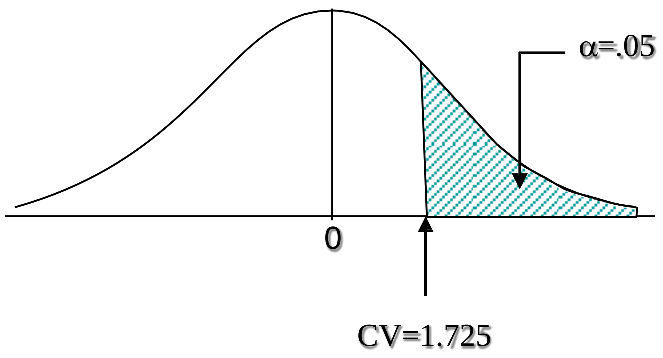


3 General Observation

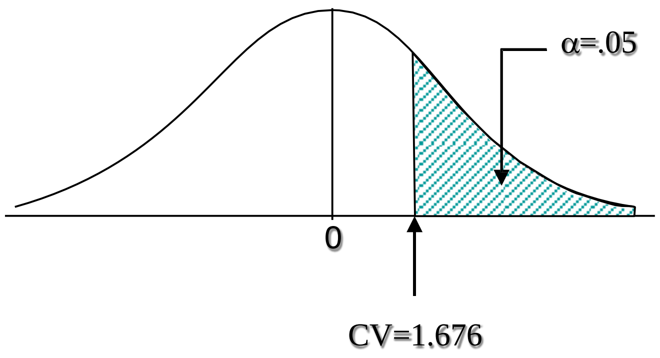
Reference:
Statistic: *t*
$\alpha = .05$

- If you reject an H_O at n=20, you will SURELY reject the H_O at a higher n (50)
- If you reject an H_O at n=50, you may OR may not reject the H_O at a smaller n (20)

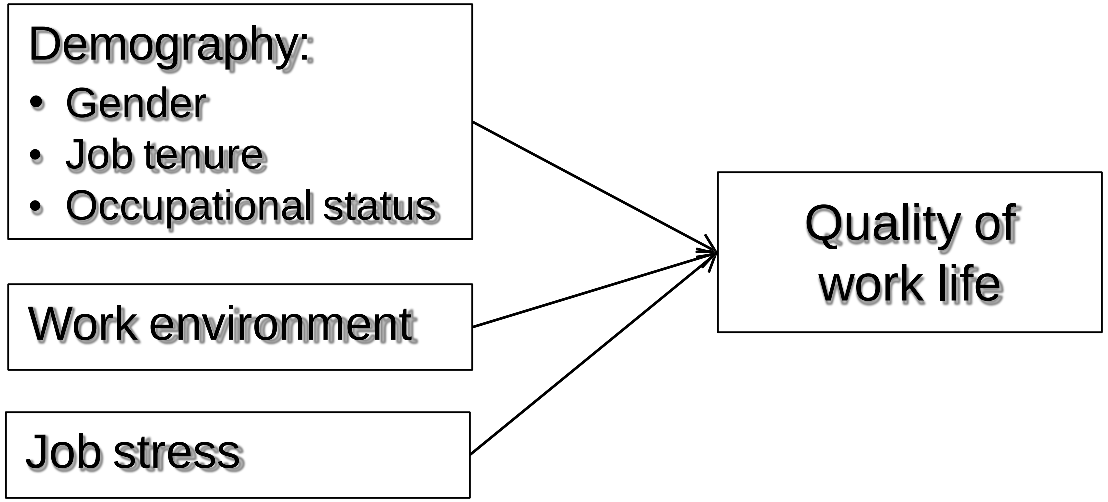
n = 20



n = 50



Hypothesis Testing



Objectives:

1. Compare differences in Quality of work life by gender
2. Determine relationship between job stress and quality of work life

What would be the possible hypotheses?

1. Quality of work life is different by gender
2. Job stress correlates with quality of work life

What are these hypotheses?

These are your research or alternative hypotheses, H_A

The hypothesis can be written as:

H_A: $\mu_m \neq \mu_f$

H_A: $\rho \neq 0$

However, in any hypothesis test, you need to have the null hypothesis.

H_O: $\mu_m = \mu_f$

H_O: $\rho = 0$

BUT - Why do you need the null hypothesis, H_O?

ANSWER:

Difficult to prove something to be true BUT much easier to prove something to be not true

~~H_O: $\mu_m = \mu_f$~~

H_A: $\mu_m \neq \mu_f$

~~H_O: $\rho = 0$~~

H_A: $\rho \neq 0$

One or Two-tailed Hypotheses

Also known as directional and non-directional hypotheses

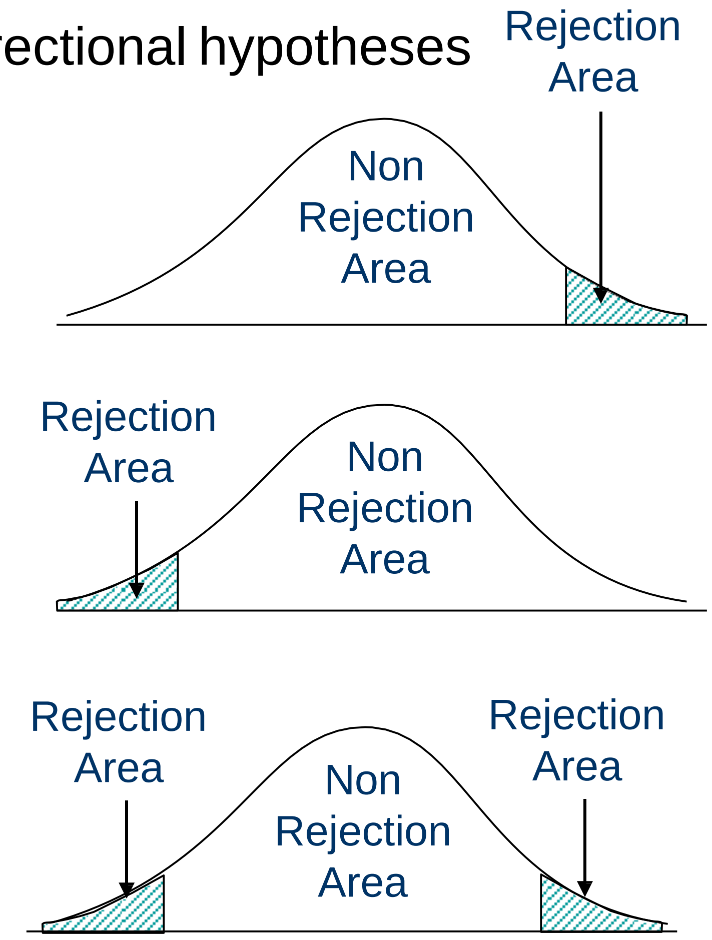
One-tailed = directional

Two-tailed = non-directional

For the above hypotheses:

H_O:	$\mu_m = \mu_f$	H_O:	$\rho = 0$
H_A:	$\mu_m \neq \mu_f$	H_A:	$\rho \neq 0$
	$\mu_m > \mu_f$		$\rho > 0$
	$\mu_m < \mu_f$		$\rho < 0$

Which one is MORE POWERFUL?
One-tailed OR two-tailed?



Steps in
Hypothesis Testing

1. State the null and alternative hypotheses
2. Set the confident/alpha level
3. Run statistical analysis and report: test statistic and sig. (*p*)value
4. Make decision
5. Conclusion

State Hypothesis

Comparison bet. groups:

$H_O: \mu_1 = \mu_2$

$H_A: \mu_1 \neq \mu_2$ ← Two-tailed

$\mu_1 > \mu_2$ ← One-tailed (More than)

$\mu_1 < \mu_2$ ← One-tailed (Less than)

Relationship bet. variables:

$H_O: \rho = 0$

$H_A: \rho \neq 0$ ← Two-tailed

$\rho > 0$ ← One-tailed (Positive)

$\rho < 0$ ← One-tailed (Negative)

◁ Back

Set Confidence Level

Generally, in social science research, alpha is set at .05

α

◁ Back

Report Test Stat and sig.

From the SPSS output, report:

- Value of the test statistic
- Sig-value (p)

◁ Back

Make Decision

Decision Criteria

SPSS

Reject H_O: sig-value < α

Fail to reject H_O: sig-value ≥ α

Criteria	*Decision*
Sig- < α	Reject H_O
	Fail to reject H_O

◁ Back

Conclusion

Relate to the hypothesis

If Reject H_O: Significant Difference/ Relationship

If Fail to Reject H_O: No Significant Difference/ Relationship

◁ Back

T-Test Statistics

Objectives

Participants to be able to:

1. Understand when to apply t-test
2. Differentiate between three types of t-test
3. Run each *t*-test using SPSS
4. Interpret results of t-test analyses

Next ▷

Types of T-Test

❶ One Sample t-test

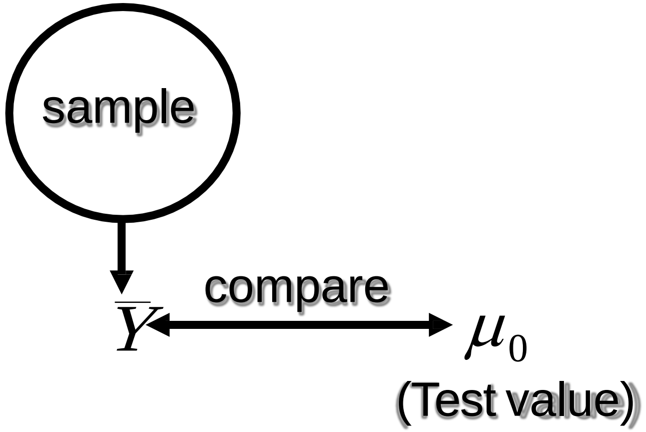
❷ Paired or Dependent Sample t-test

❸ Independent Sample t-test

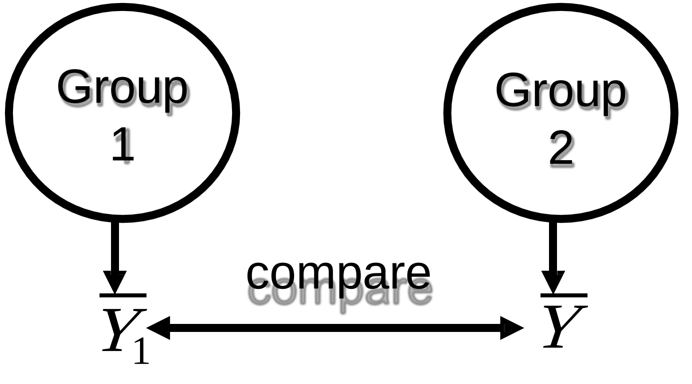
Next ▷

Differentiation

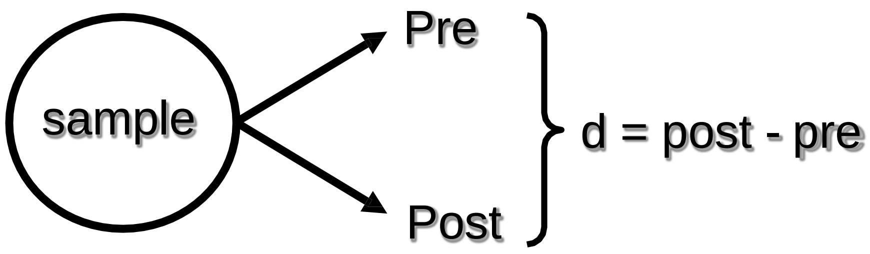
Three Types of T-Test



One-sample t-test



Independent sample t-test



Paired-sample t-test

Next ▶

Paired-Sample T-Test

Next

Introduction

- Dependent sample t-test is a bivariate, parametric and inferential statistics
- Employed in experimental research that involves repeated or dependent measures
- Example: Test effect of experimental treatment by comparing pre- and post-test scores

Purpose

- Compare differences between two (2) dependent mean scores
- Example: Test the effect of diet formula to loose weight by comparing pre- and post-test mean weight

$$\overline{Y}_{pre} \leftrightarrow \overline{Y}_{post}$$

Requirements

- Pre-test and Post-test scores are Interval or Ratio
- Ex: Compare mean weight between pre-test and post test
- Mean weight: Ratio

Assumptions

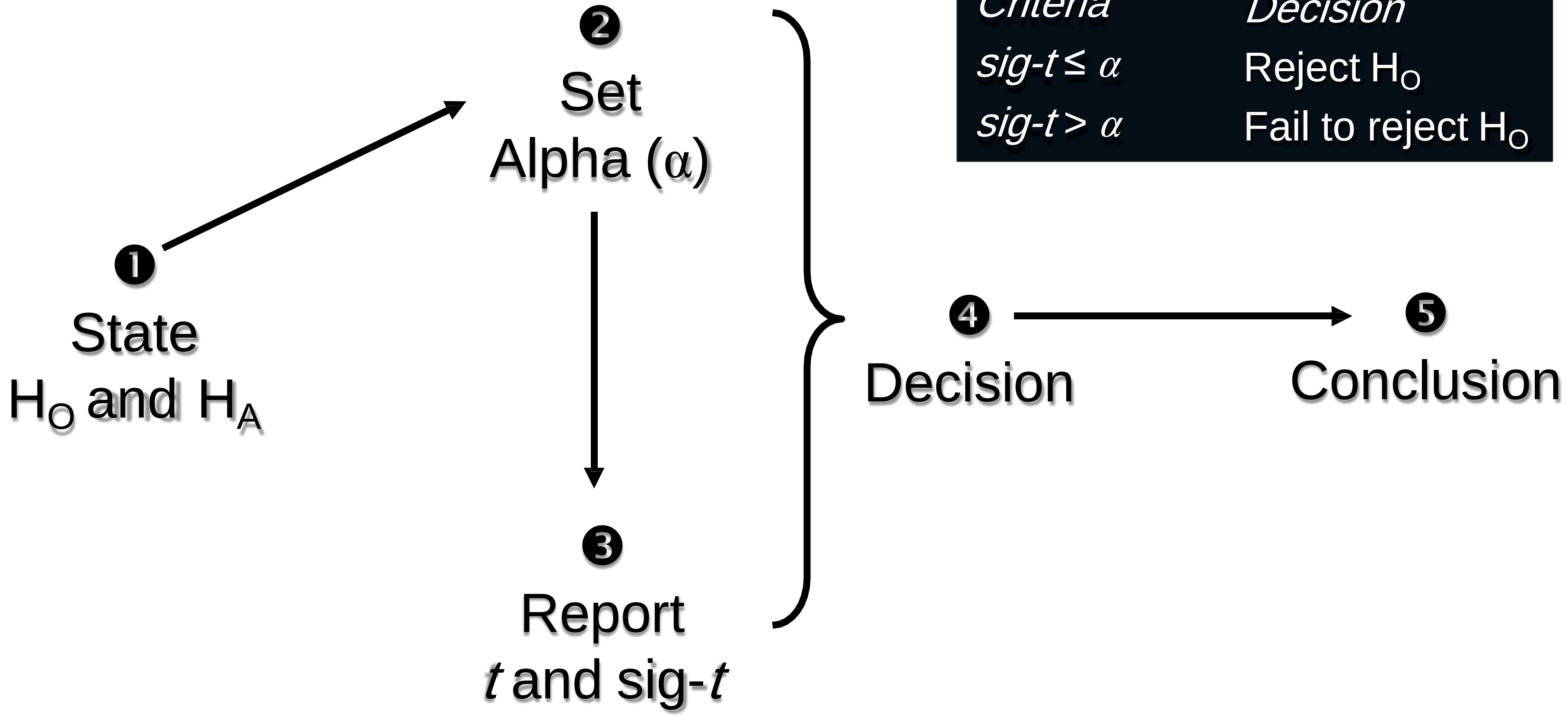
1. The pre-test and post-test scores are normally distributed
2. The cases represent random samples from the populations and the scores on the test variable are independent of each other

Next ▷

What to Expect?

Manual SPSS

Hypothesis Test



Criteria	*Decision*
sig-t ≤ *α*	Reject H_O
sig-t > *α*	Fail to reject H_O

Next ▶

5-Step Hypothesis Test

5-Steps Hypothesis Test

1. State H_O and H_A
2. Set Confidence Level (α)
3. Report *t* and sig-*t*
4. Decision
5. Conclusion

Next ▶

Step 1: State H_0 & H_A

H_O: $\mu_d = 0$

H_A: $\mu_d \neq 0$

$\mu_d > 0$

$\mu_d < 0$

Next ▶

Step 2: Set Confidence Level

Generally, confident level is set at .05

$$\alpha = .05$$

Next ▷

Step 3: Report *t* and sig-*t*

Simply report:

1. *t*
2. sig-*t*

Paired Samples Test

			Pair 1 post - pre
Paired Differences	Mean		.50000
	Std. Deviation		.70711
	Std. Error Mean		.22361
	95% Confidence Interval of the Difference	Lower	-.00583
		Upper	1.00583
t			2.236
df			9
Sig. (2-tailed)			.052

Next ▷

Step 4: Decision

- Only two (2) possible decisions.
- Reject or Fail to Reject H_O

Reject H_O: sig-$t \leq \alpha$

Fail to reject H_O: sig-$t > \alpha$

Criteria	*Decision*
sig-$t \leq \alpha$	Reject H_O
sig-$t > \alpha$	Fail to reject H_O

Next ▶

Step 5: Conclusion

Reject H_O

There is significant difference between pre-test and post-test scores

Fail to reject H_O

There is no significant difference between pre-test and post-test scores

Next ▷

Effect Size

- The magnitude of the difference
- Cohen (1988) proposed d as a measure of effect size

$$d = \frac{\bar{d}}{s_d}$$

Interpretation:

<.2	Trivial
.2	Small
.5	Medium
.8	Large

Note

- For *t*-test analysis, SPSS does not provide option for a one-tailed test.
- For the two-tailed test, simply use the given sig-*t* and compare against alpha (α) to make your decision
- For a one-tailed test, divide the sig-*t* (2-tailed) by two (2) and use this value to compare against alpha (α)

Next ▷

Example/Exercise

A training program was conducted to improve participants' knowledge on ICT. Data were collected from a selected sample both before and after the ICT training program.

1. Test the hypothesis that the training is effective to improve participants knowledge on ICT at $\alpha = .05$
2. Calculate and interpret the effect size (d)

Data set:

Pre	Post
12	13
14	15
13	13
11	12
12	13
10	11
15	16
13	13
9	8
14	14

One- OR Two-Tailed Test?

Data Paired ttest

Next ▷

1. Hypothesis: Significance of relationship

a. Hypotheses

H_O: $\mu_d = 0$

H_A: $\mu_d > 0$

b. Set confidence level

$\alpha = .05$

c. Report:

$t = 2.236$

sig-t = .026 (1-tailed)

Paired Samples Test

			Pair 1 post - pre
Paired Differences	Mean		.50000
	Std. Deviation		.70711
	Std. Error Mean		.22361
	95% Confidence Interval of the Difference	Lower	-.00583
		Upper	1.00583
t			2.236
df			9
Sig. (2-tailed)			.052

Next ▷

Criteria	*Decision*
sig-$t \leq \alpha$	Reject H_O
sig-$t > \alpha$	Fail to reject H_O

d. Decision

Since sig-t (.026) < α (.05)

$\therefore$ Reject H_O

e. Conclusion

There is significant increase in knowledge on ICT. Thus the training program is significantly effective to improve participants' knowledge on ICT at .05 level of significance

Table 1: Results of paired *t*-test between pre- and post test scores

Scores	*Mean*	*SD*	*t*	*p*
Post-test	12.8	2.20	2.236	.026
Pre-test	12.3	1.89		

2. Effect Size

$$d = \frac{\bar{d}}{s_d}$$

$$= \frac{.50}{.70711}$$

$$= .707$$

Medium effect size

Interpretation:

<.2	Trivial
.2	Small
.5	Medium
.8	Large

Paired Samples Test

			Pair 1 post - pre
Paired Differences	Mean		.50000
	Std. Deviation		.70711
	Std. Error Mean		.22361
	95% Confidence Interval of the Difference	Lower	-.00583
		Upper	1.00583
t			2.236
df			9
Sig. (2-tailed)			.052