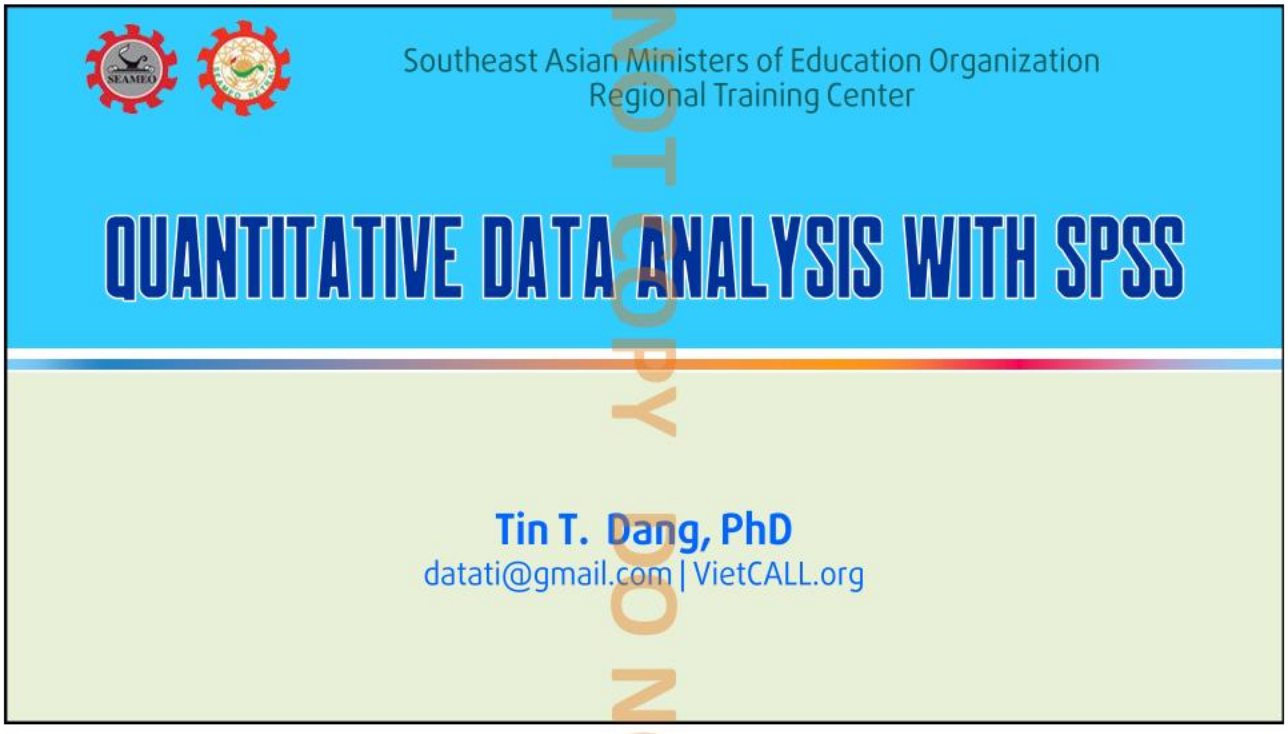
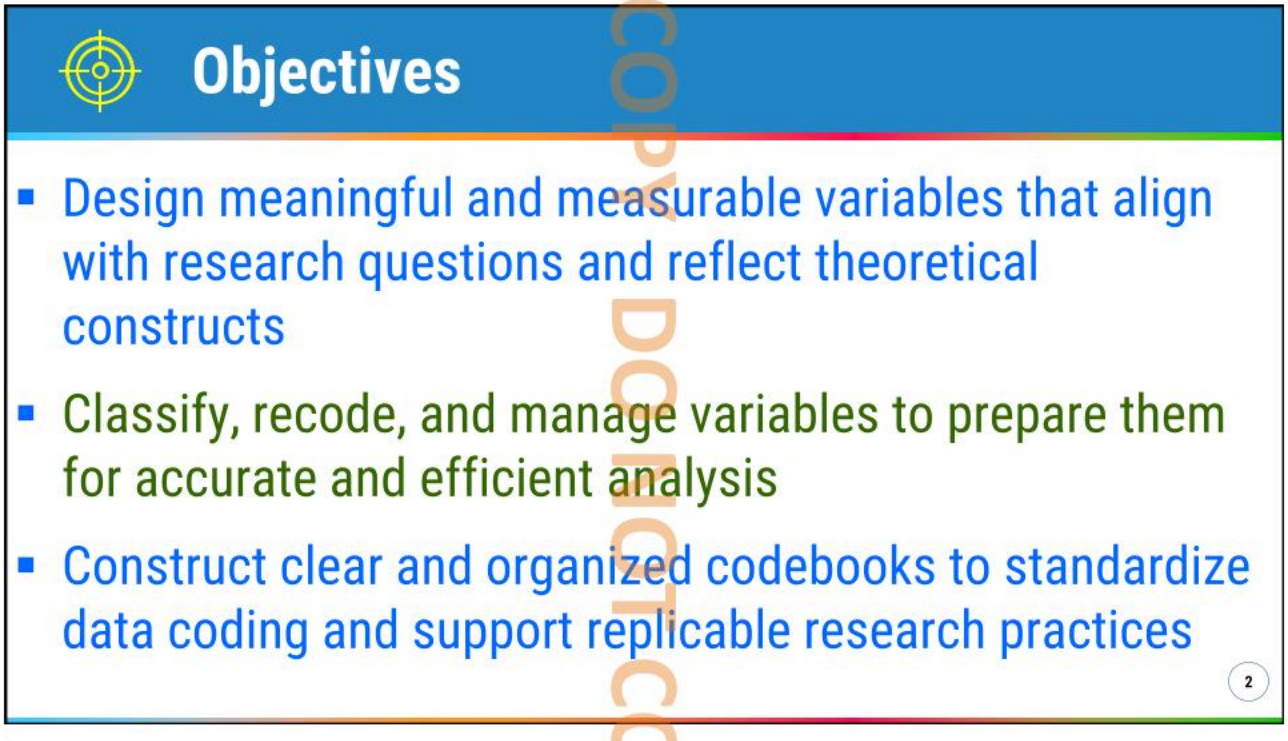


1



2

Objectives

- Clean datasets and apply descriptive statistics to ensure data quality and summarize patterns and distributions
- Apply inferential statistical methods to test hypotheses and draw generalizable conclusions from sample data
- Conduct correlation and regression analyses to explore and interpret relationships between key variables.

3

3

Outline

1. Variable design
2. Variable handling
3. Codebook building
4. Cleaning data
5. Descriptive statistics
6. Simple inferential statistics
7. Correlation and regression

4

4

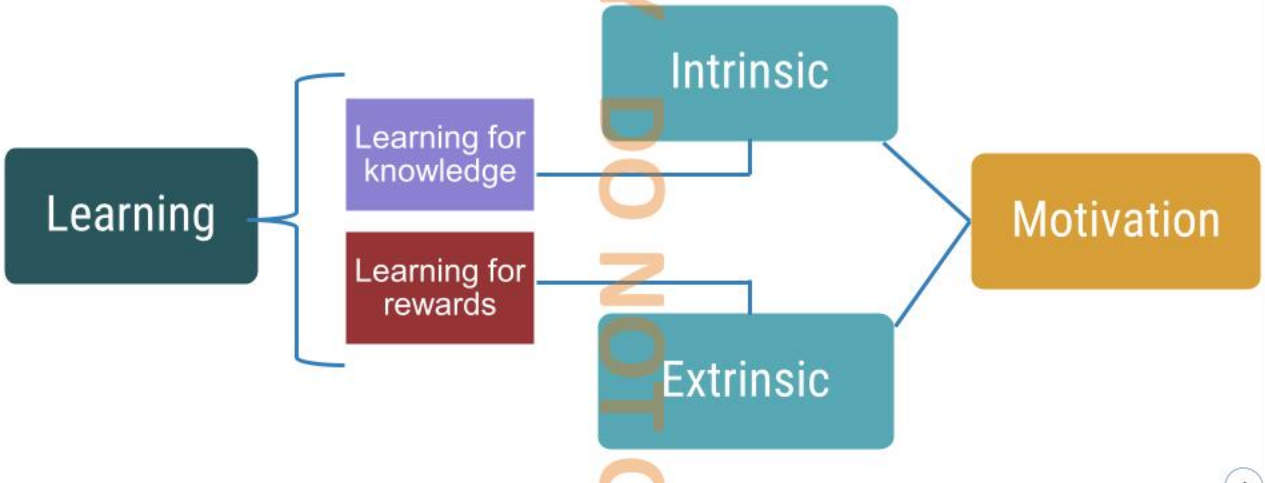
1. Variable Design

- 1.1 Research focus
- 1.2 Instrument drafting
- 1.3 Validity of the instrument
- 1.4 Data collection

5

5

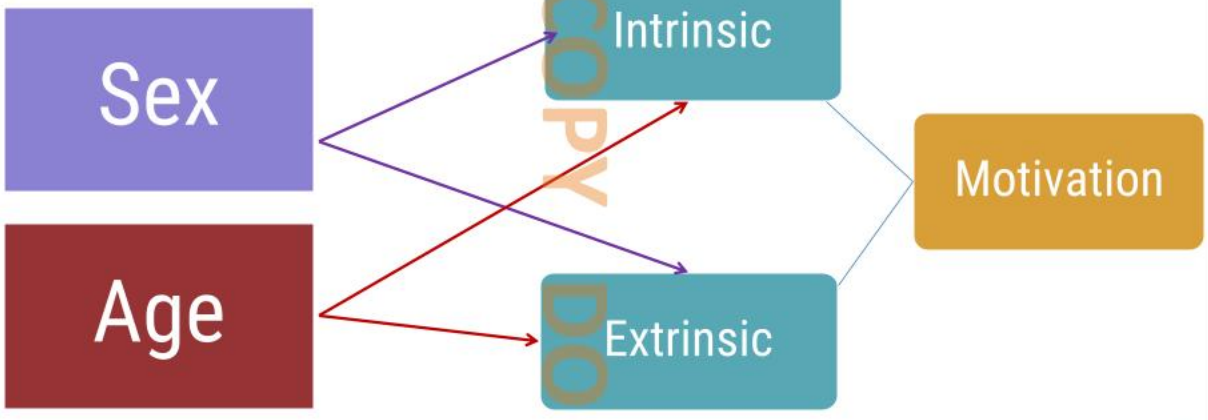
1.1 Research focus



6

6

1.1 Research focus



7

7

1.1 Research focus

Reseach questions

1. What is the impact of EFL students' sex on their level of learning motivation?
2. To what extent is student's level of motivation correlated with their age level?

8

8

1.2 Instrument drafting

- What is your gender?

  Male Female

- What is your age level?

  10-15 16-18 19-22

9

9

1.2 Instrument drafting

- Indicate your agreement with the following statements

INTRINSIC MOTIVATION	SD	D	N	A	SA
1. I learn English to know about US culture.					
2. I learn English as I want to improve myself.					
3.					
4.					
5.					
6.					

10

10

1.2 Instrument drafting

- Indicate your agreement with the following statements

EXTRINSIC MOTIVATION	SD	D	N	A	SA
1. I learn English to get a good job.					
2. I learn English to make my parents happy.					
3.					
4.					
5.					
6.					

11

11

1.3 Validity

- Do the items in the instrument really mean what you want to measure?
- Discuss with your partners and experts to choose the best items possible. Improve your items if necessary
- Finalizing a list of at least 6 items for each type of motivation.

12

12

1.4 Data collection

- Distribute your questionnaire to the participants
- Ask your participants to complete the questionnaire
- Collect responses from the participants
- Ensure that you can get responses from around 50 participants.

13

13

2. Variable Handling

2.1	Types of variables
2.2	Coding variables
2.3	Managing variables
2.4	General advice

14

14

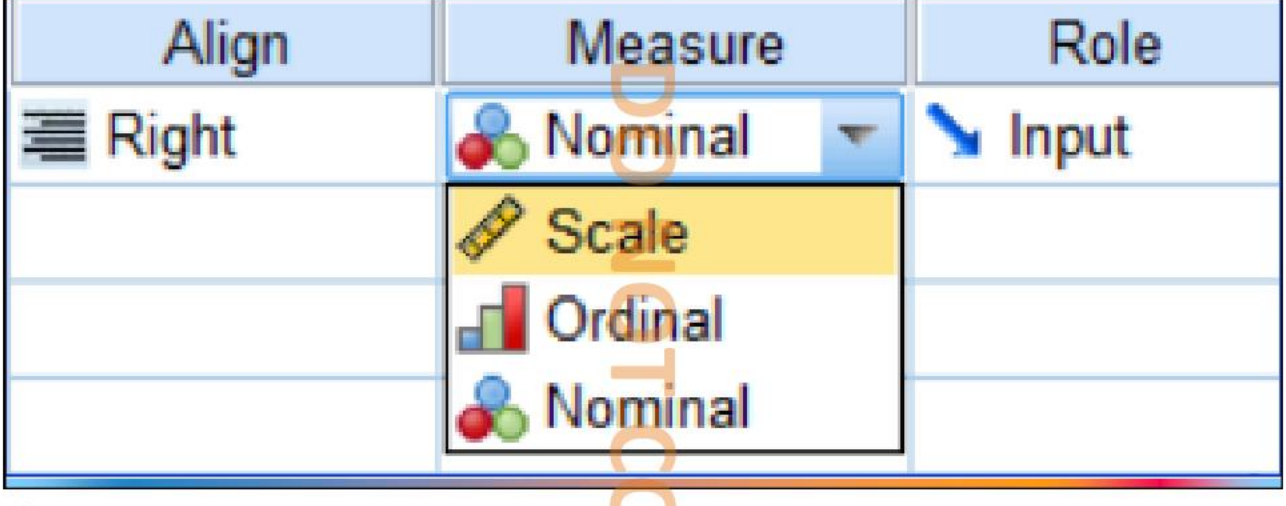
2.1 Types of variables

Nominal	Categorical variables with no order or ranking	Gender, region
Ordinal	Categorical variables with a meaningful order, but intervals are not equal or known	Edu level, class rank
Interval	Numerical variables with equal intervals between values but no true zero point	Temperature, year period
Ratio	Numerical variables with equal intervals and a true zero point	Age, income
Scale	A general term for interval or ratio	Likert scale, screen time

15

2.1 Types of variables

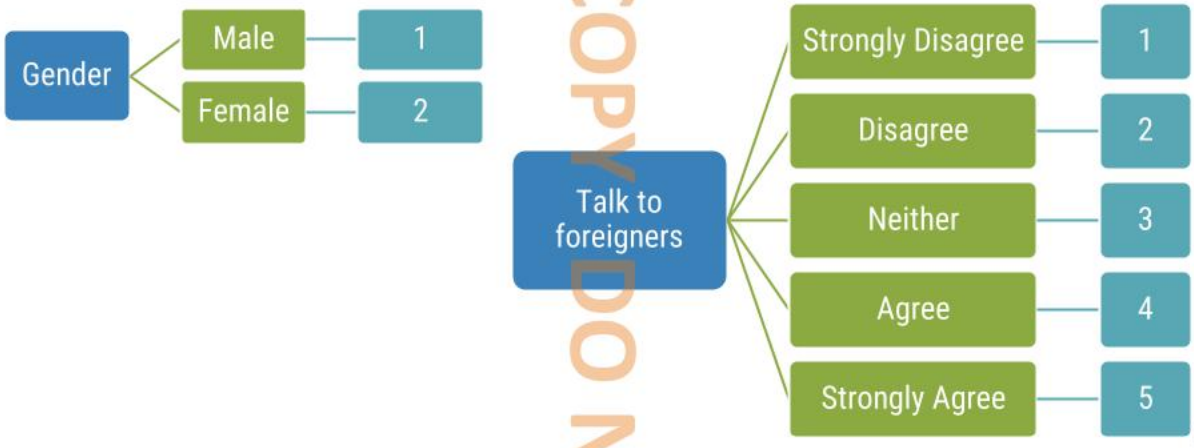
On SPSS



16

2.2 Coding variables

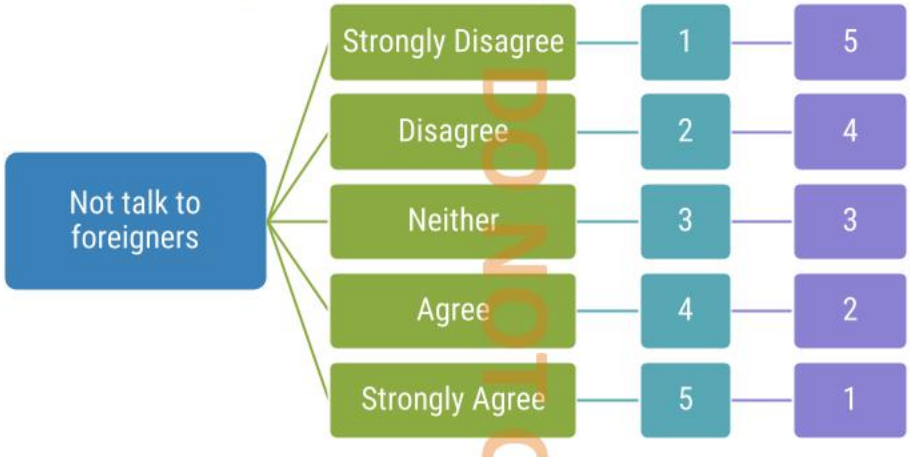
- Assigning values to levels of variables



17

2.2 Coding variables

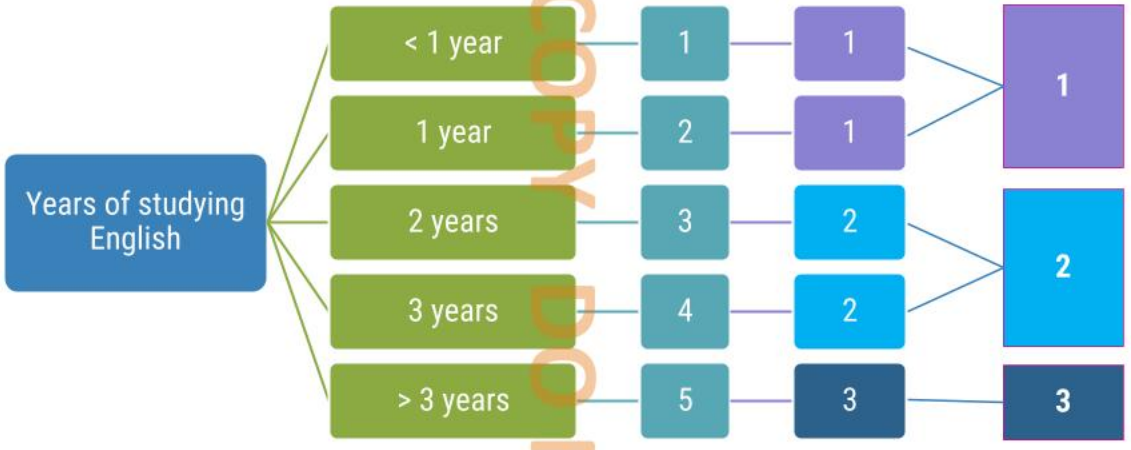
- Reverse coding for negatively phrased items in scales



18

2.2 Coding variables

- Combining categories



19

2.3 Managing variables

- Using clear, meaningful, and consistent variable names
- Checking for missing values and duplicates
- Documenting all changes for transparency and reproducibility
- Keeping in mind the statistical tests associated with the variable types.

20

20

2.4 General advice

- Classifying variables correctly to choose appropriate analyses
- Recoding when needed to make data usable and consistent
- Managing variables methodically to ensure analytical accuracy.

21

21

3. Codebook Building

3.1 Concept of a codebook

3.2 Purpose of a codebook

3.3 Elements of a codebook

3.4 Example of a codebook

22

22

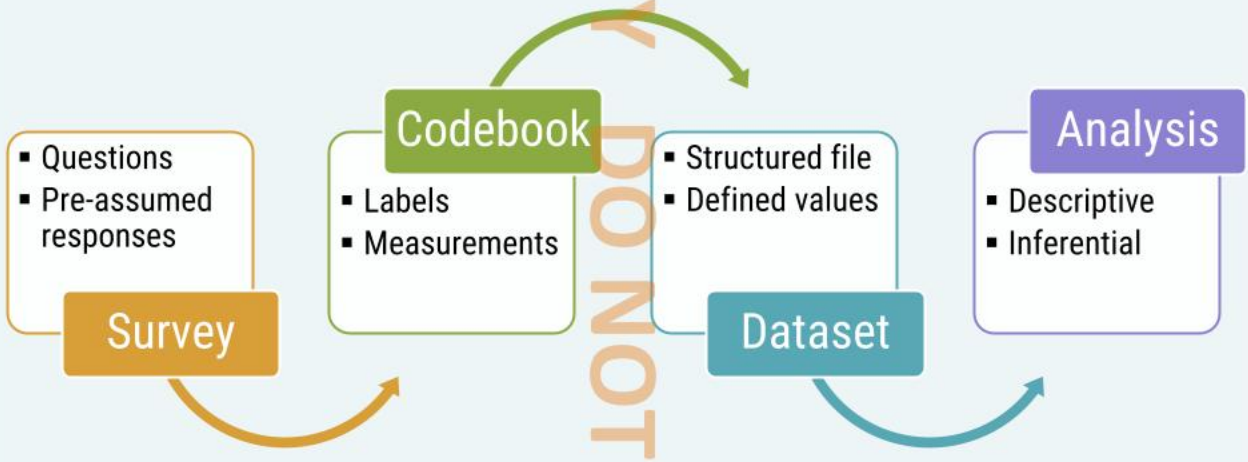
3.1 Concept of a codebook

- A structured document that describes the variables, values, and coding rules in a dataset
- Acting as a reference guide for researchers and collaborators
- Essential for transparency, replication, and data sharing.

23

23

3.2 Purpose of a codebook



24

24

3.2 Purpose of a codebook

- Clarifying what each variable measures
- Standardizing coding across team members and time
- Helping troubleshoots, cleaning, and analyzing data efficiently
- Required for publication or data archiving in most reputable journals.

25

25

3.3 Elements of a codebook

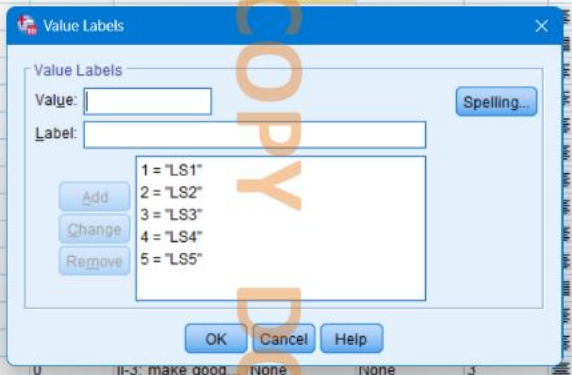
Variable Name	Gender, age, planning
Variable Label	Gender, age, making learning plans
Value Labels	0 for female and 1 for male; 1 for strongly disagree and 5 for strongly agree
Variable Type	Nominal, ordinal, scale
Notes/Comments	Transforming data, history, problem when collecting data

26

26

3.4 Example of a codebook

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	ClassName	String	3	0	Class Name	None	None	7	Left	Nominal	Input
2	Class	Numeric	3	0		{1, LS1}...	None	4	Center	Nominal	Input
3	Group	Numeric	3						Center	Nominal	Input
4	Groups2	Numeric	8						Right	Scale	Input
5	ID	String	10						Left	Nominal	Input
6	Name	String	23						Left	Nominal	Input
7	ScorePre	Numeric	11						Center	Scale	Input
8	ScorePost	Numeric	11						Center	Scale	Input
9	Gender	Numeric	11						Center	Nominal	Input
10	YOB	Numeric	11						Center	Nominal	Input
11	Age	Numeric	8						Center	Scale	Input
12	IT	Numeric	11						Center	Ordinal	Input
13	TimeY	Numeric	11						Right	Nominal	Input
14	II1	Numeric	11						Center	Ordinal	Input
15	II2	Numeric	11						Center	Ordinal	Input
16	II3	Numeric	11	0	II-3: make good...	None	None	3	Center	Ordinal	Input
17	II4	Numeric	11	0	II-4: look for op...	None	None	3	Center	Ordinal	Input
18	II5	Numeric	11	0	II-5: try to find a...	None	None	3	Center	Ordinal	Input
19	II6	Numeric	11	0	II-6: will ask the...	None	None	3	Center	Ordinal	Input



27

27

4. Data Screening

4.1 Entering data

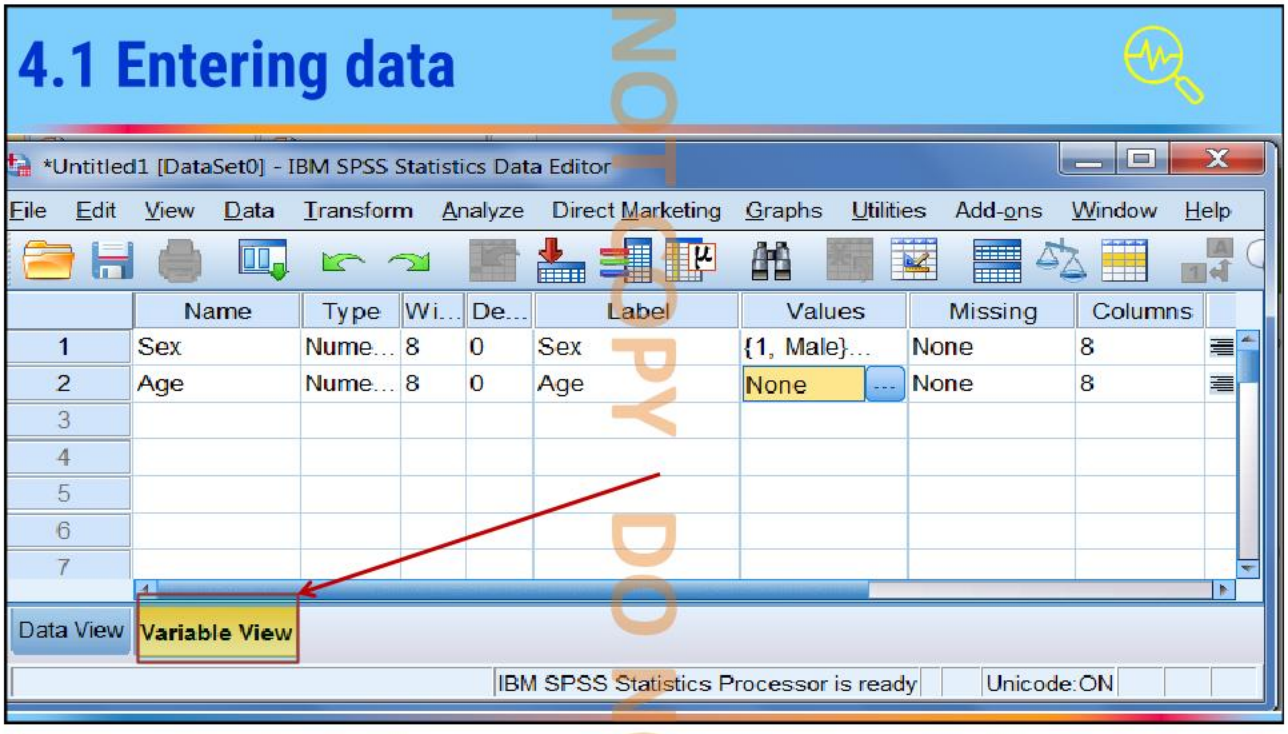
4.2 Checking missing values

4.3 Normal distribution

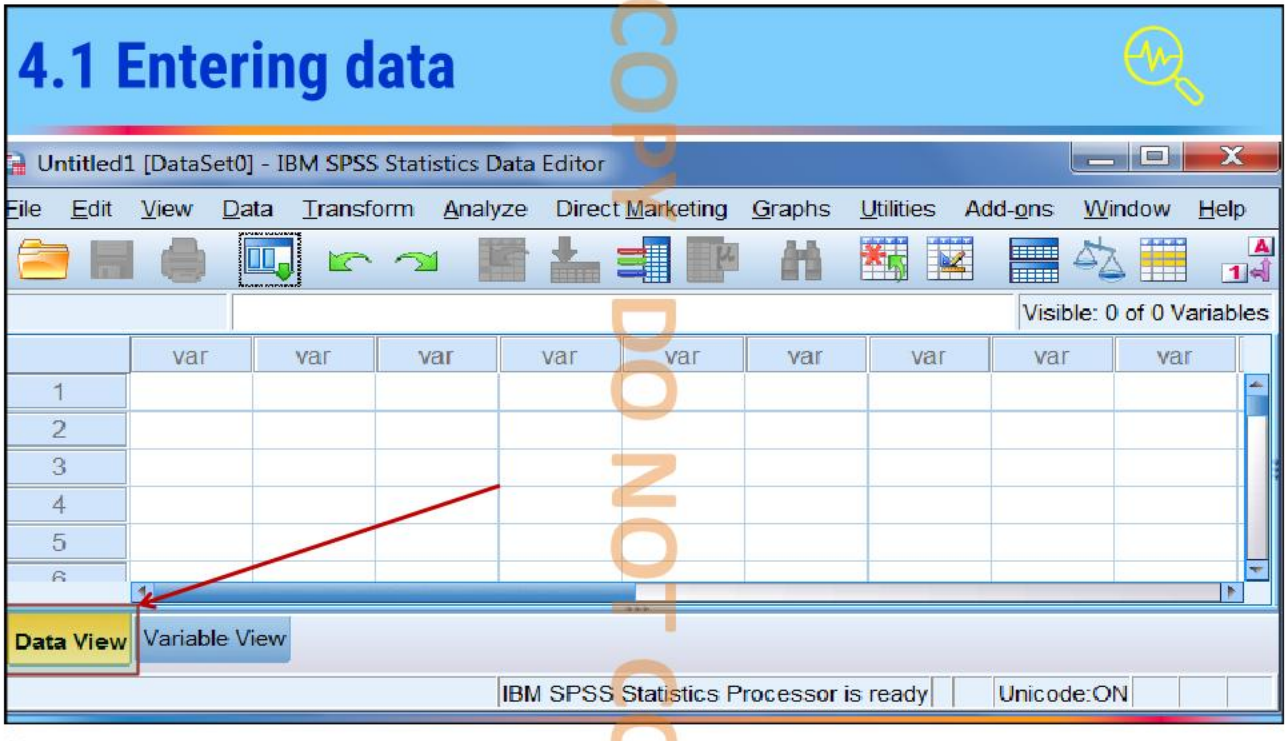
4.4 Reliability test

28

28

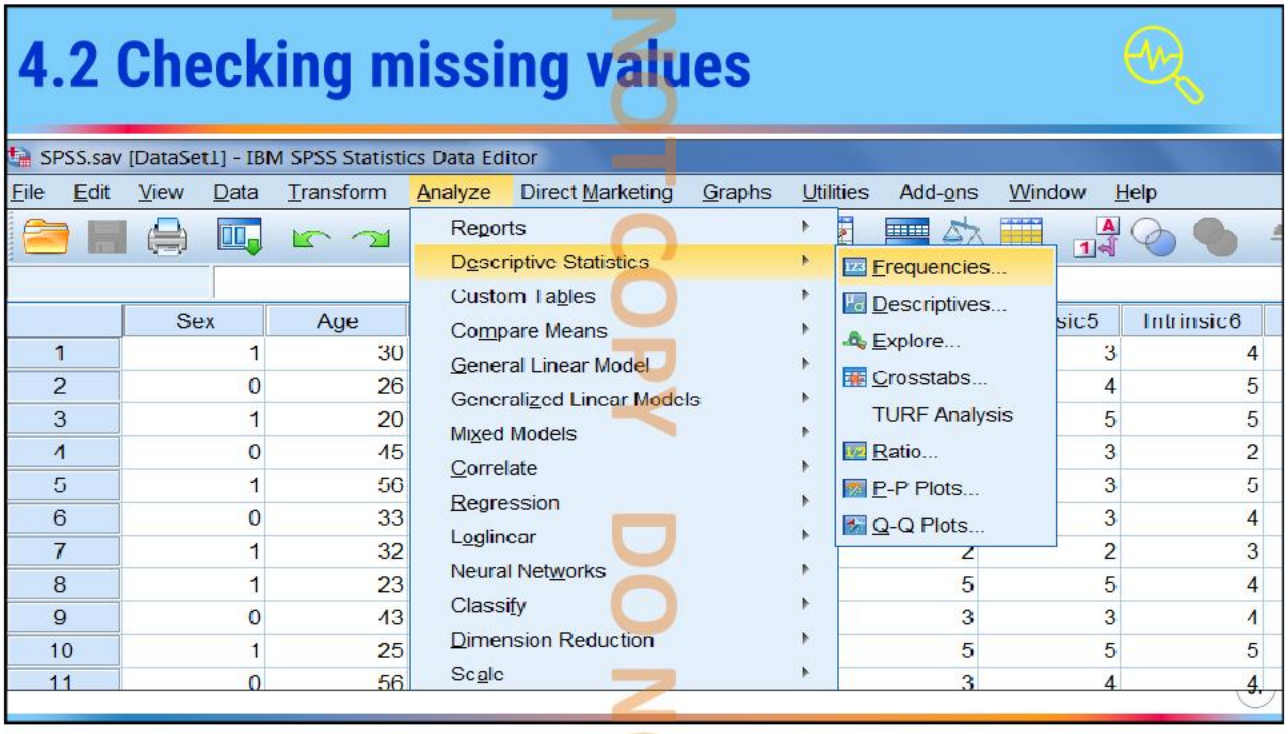


29



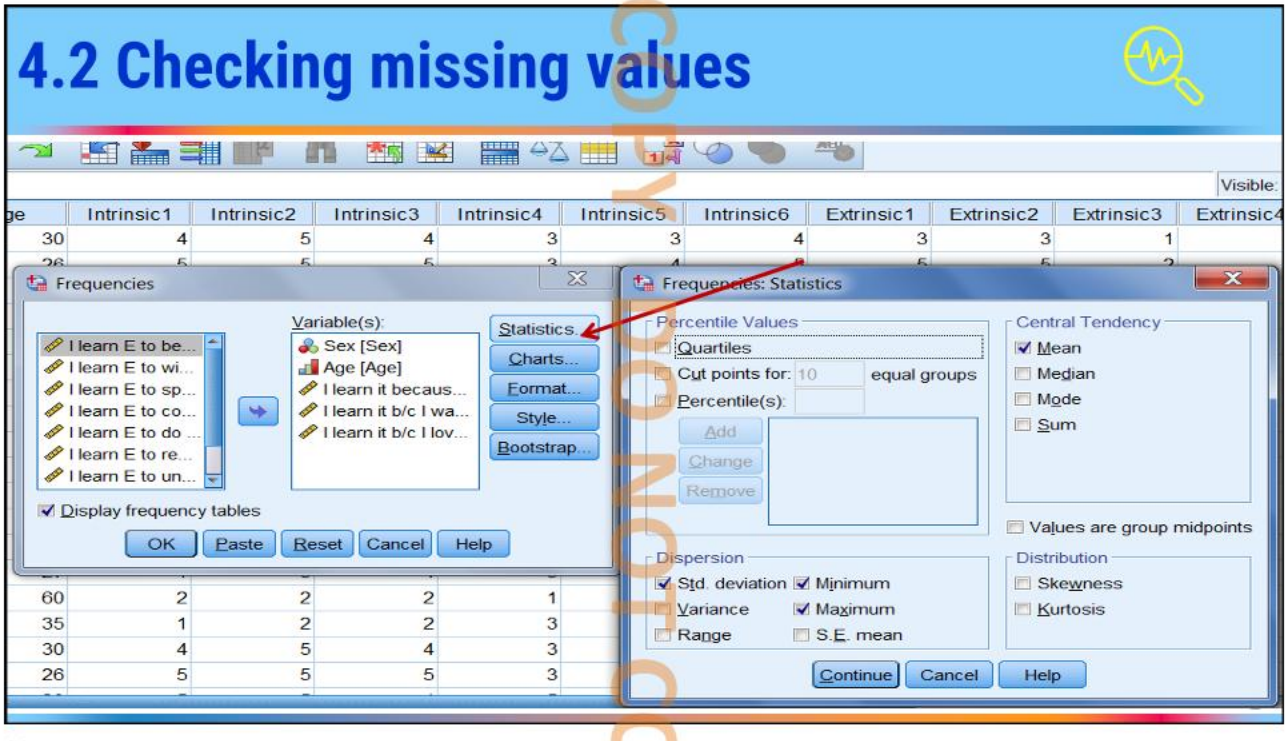
30

4.2 Checking missing values



31

4.2 Checking missing values



32

4.2 Checking missing values

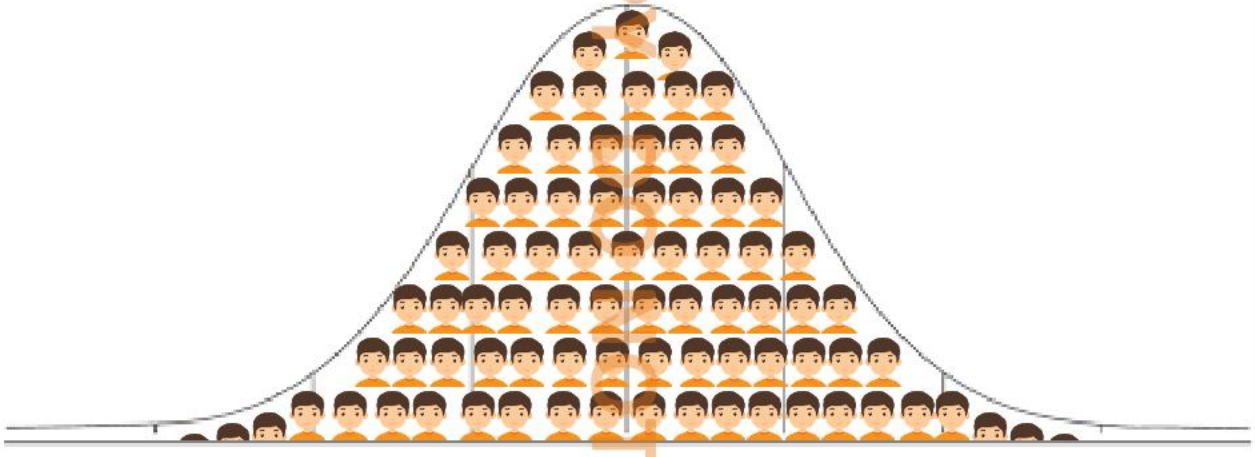
Frequencies

Statistics

		Sex	Age	I like it	I want to improve my skills	I love foreign languages
N	Valid	50	50	50	50	50
	Missing	0	0	0	0	0
Mean		.64	34.94	4.08	4.26	3.76
Std. Deviation		.485	12.566	1.426	1.084	1.061
Minimum		0	20	1	2	2
Maximum		1	60	5	5	5

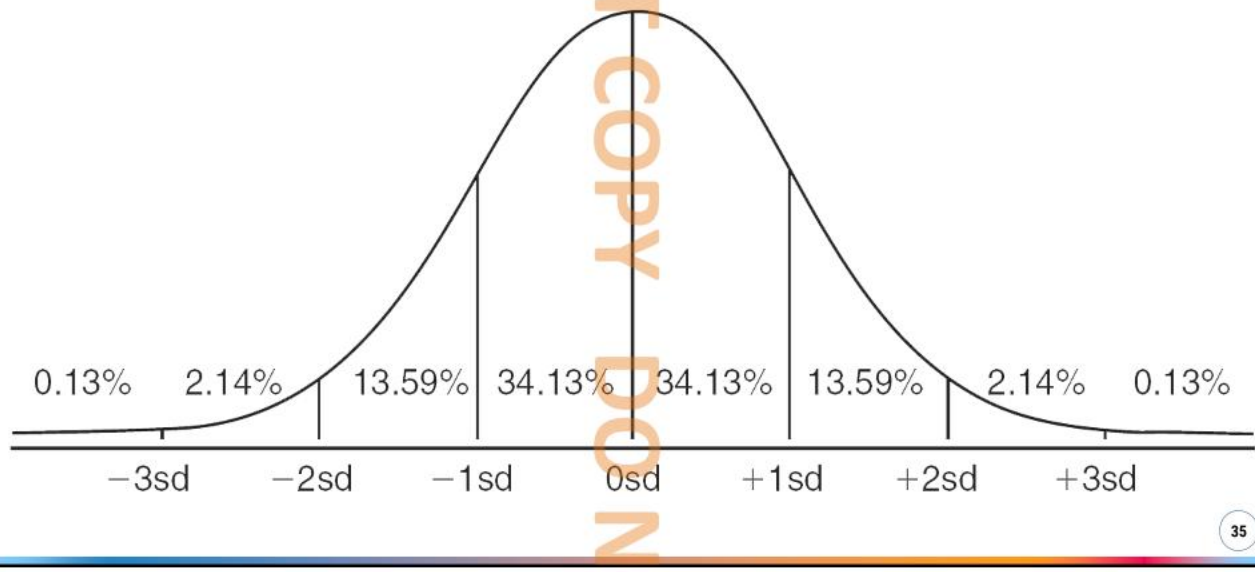
33

4.3 Normal distribution



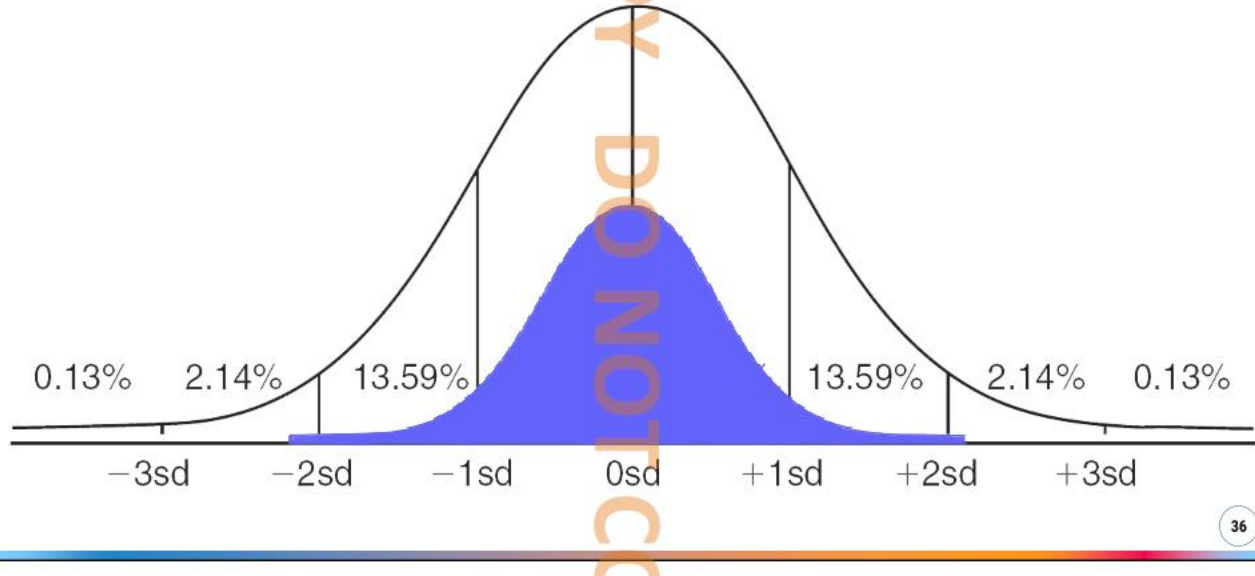
34

4.3 Normal distribution

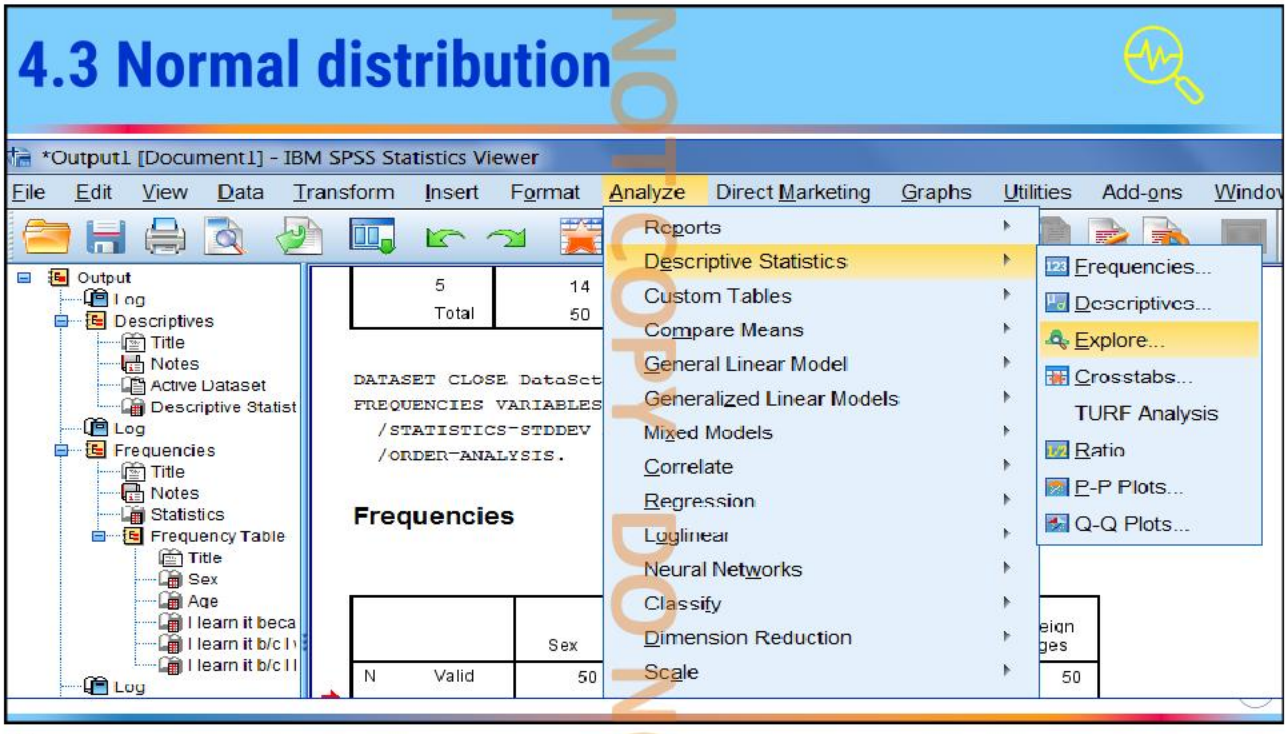


35

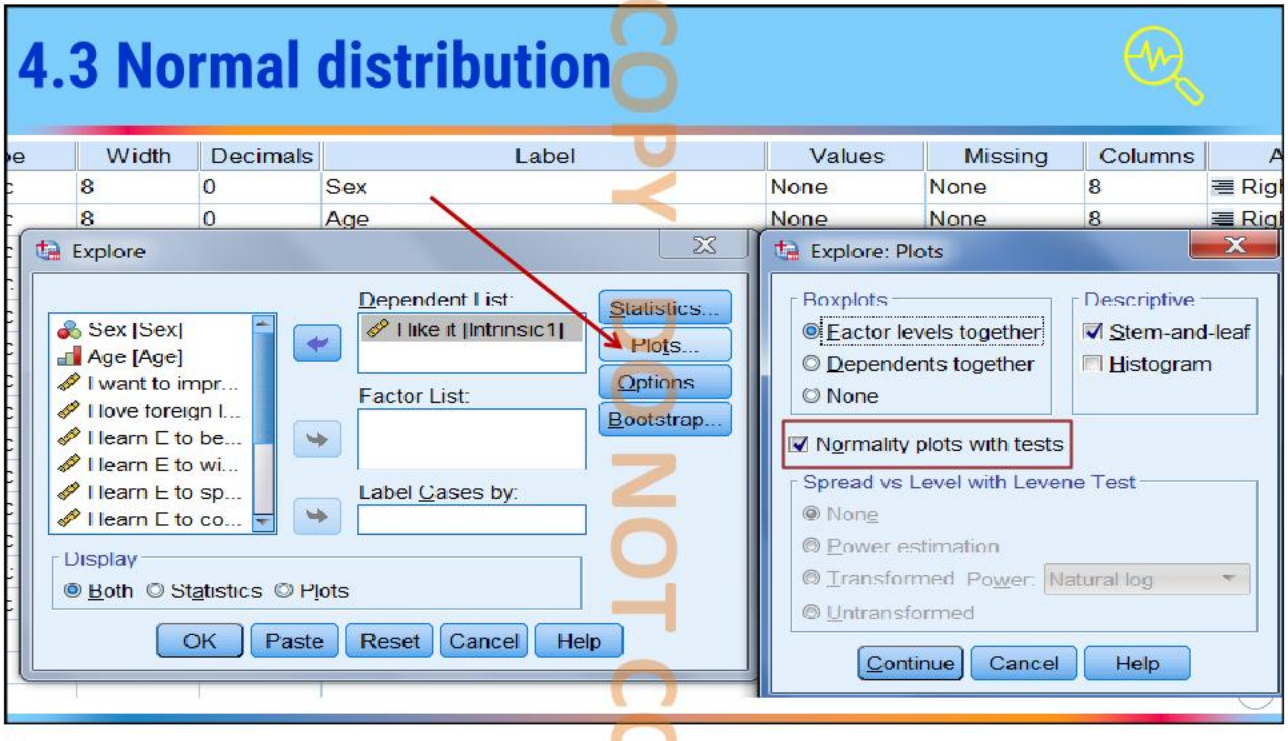
4.3 Normal distribution



36

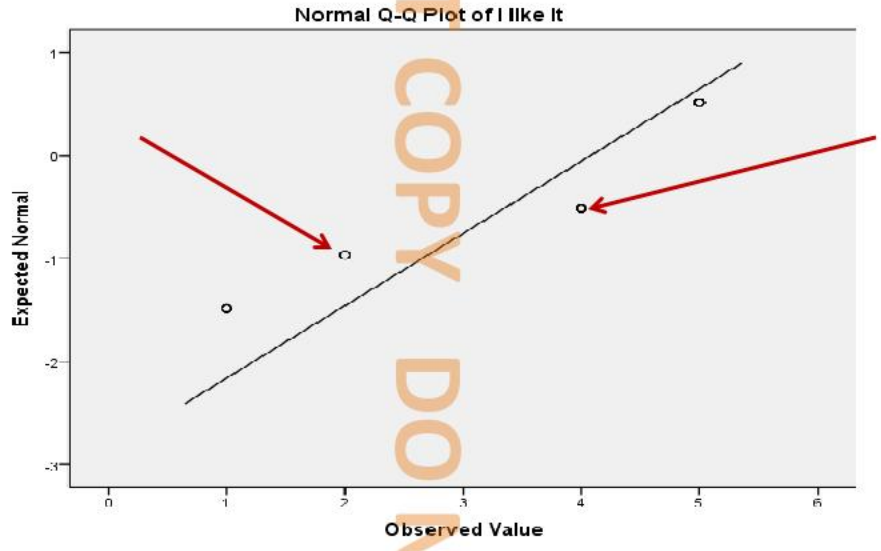


37



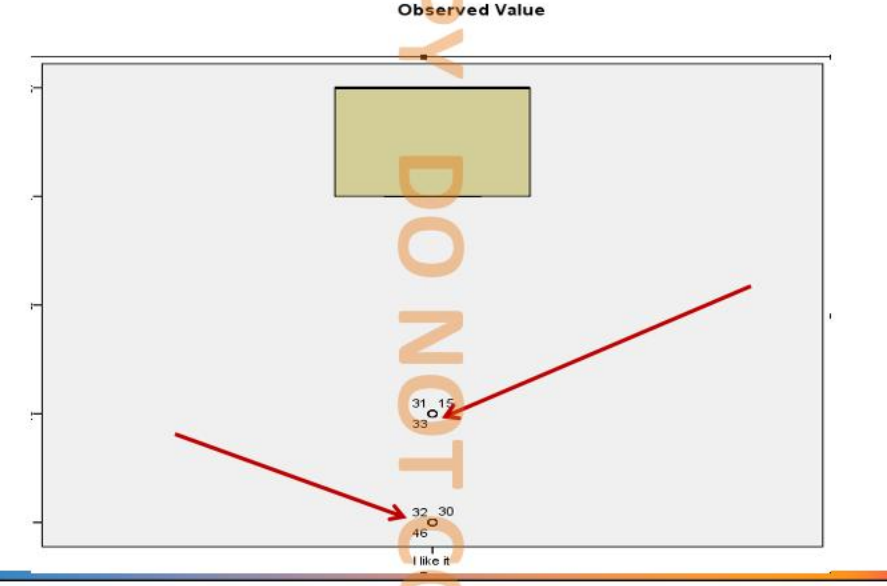
38

4.3 Normal distribution



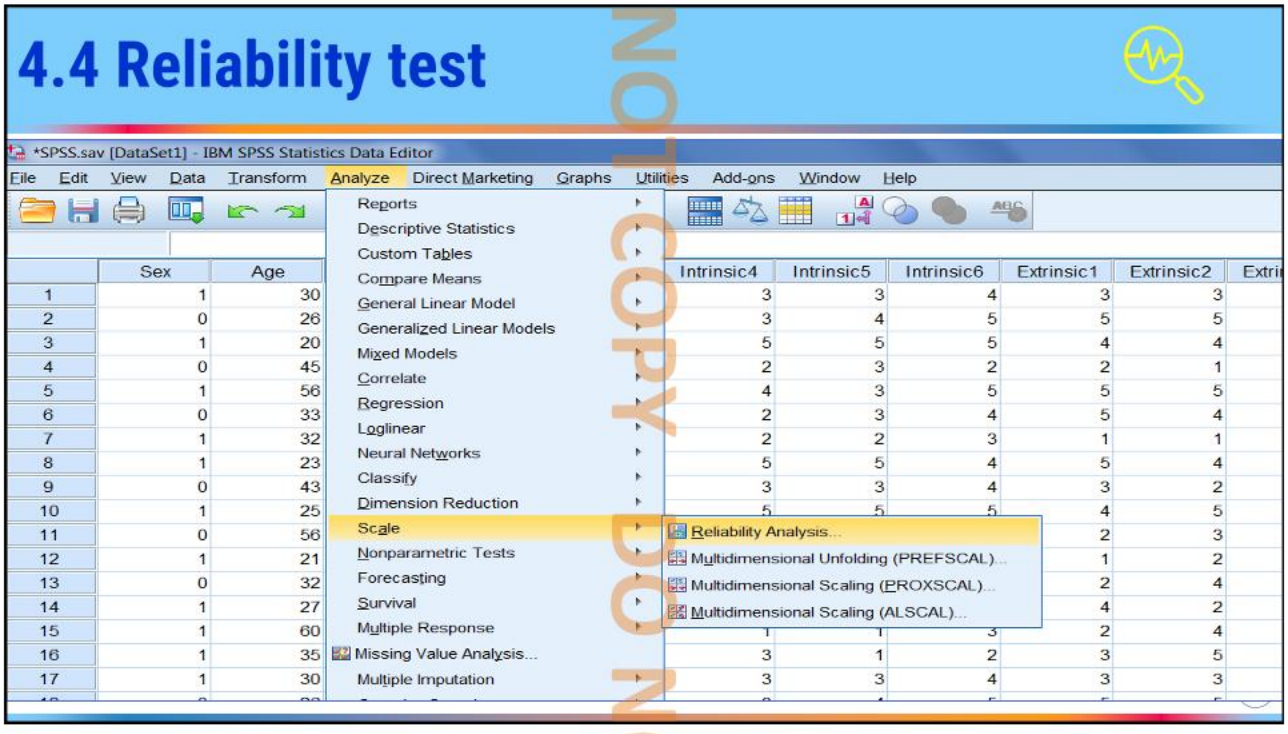
39

4.3 Normal distribution



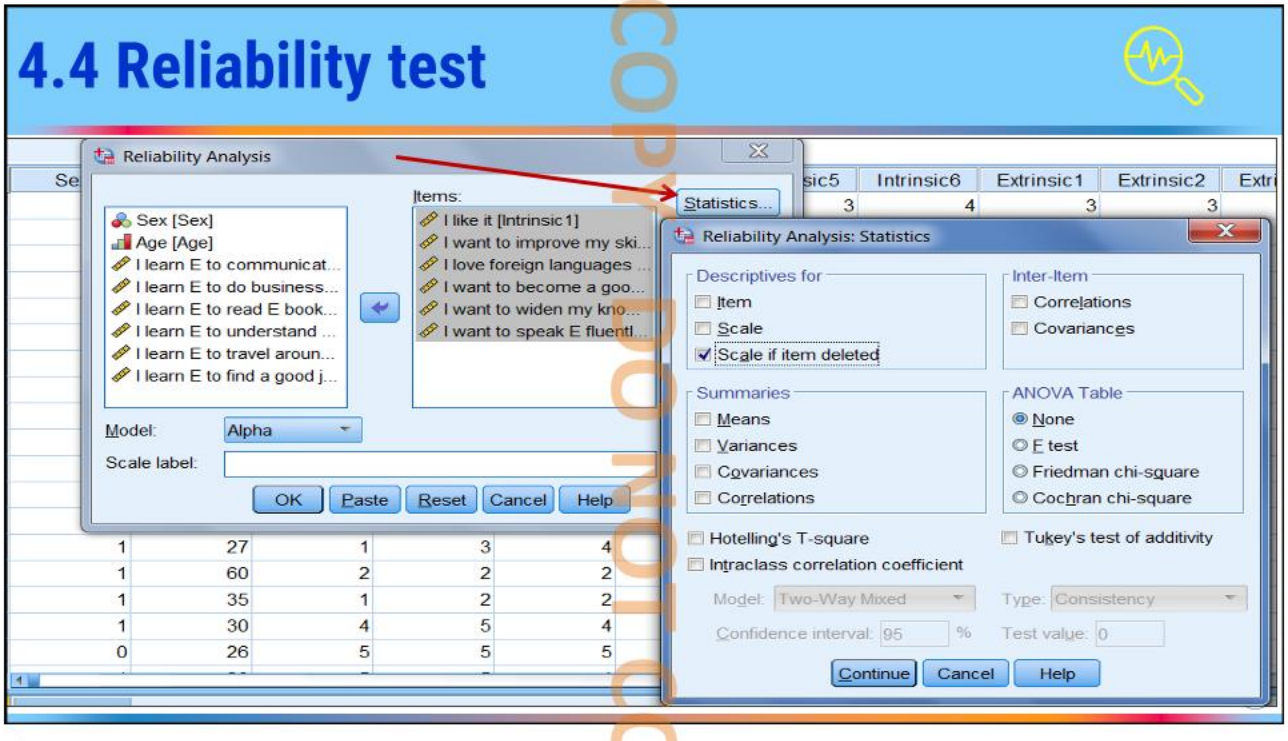
40

4.4 Reliability test



41

4.4 Reliability test



42

4.4 Reliability test

Reliability Statistics

Cronbach's Alpha	N of Items
.816	6

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
I like it	17.74	23.543	.298	.853
I want to improve my skills	17.56	22.211	.618	.781
I love foreign languages	18.06	21.935	.669	.772
I want to become a good E speaker	18.68	21.732	.522	.799
I want to widen my knowledge	18.86	17.715	.794	.731
I want to speak E fluently	18.20	21.184	.673	.768

43

43

6. Simple Inferential Statistics

- 6.1 Independent-samples T-tests
- 6.2 One-way ANOVA
- 6.3 Chi-square

44

44

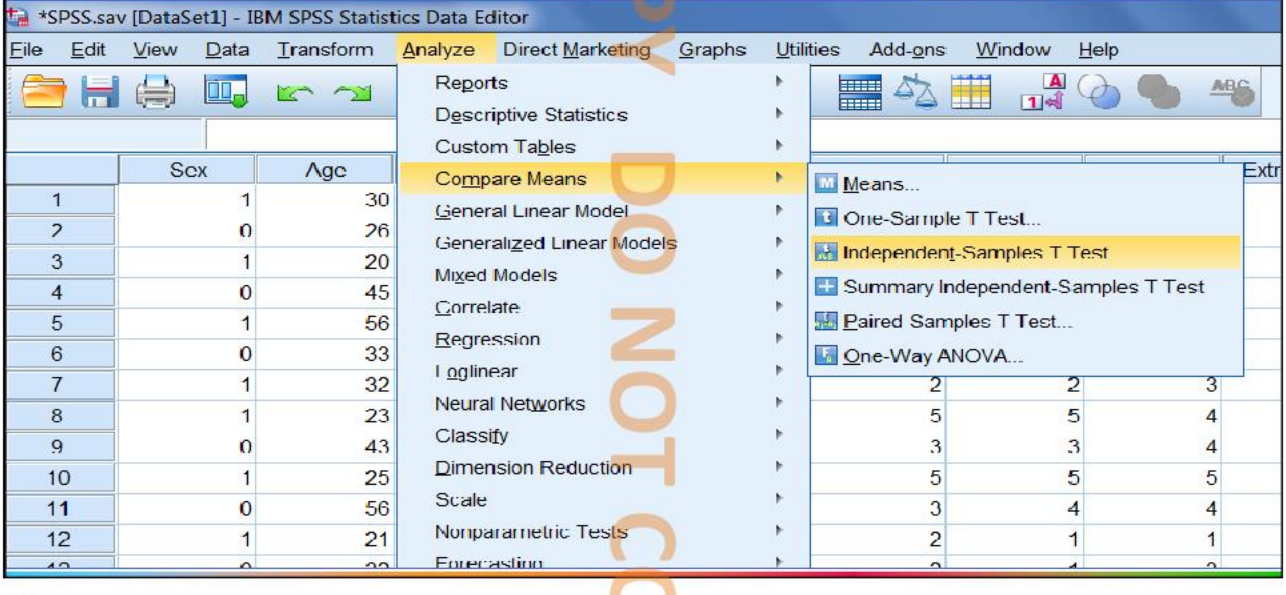
6.1 Independent-samples T-test

- Comparing means of a variable from 2 different groups.
- In this case, the males' mean of intrinsic motivation level is compared to the females' mean of intrinsic motivation level.
- Refer to research question 1.

45

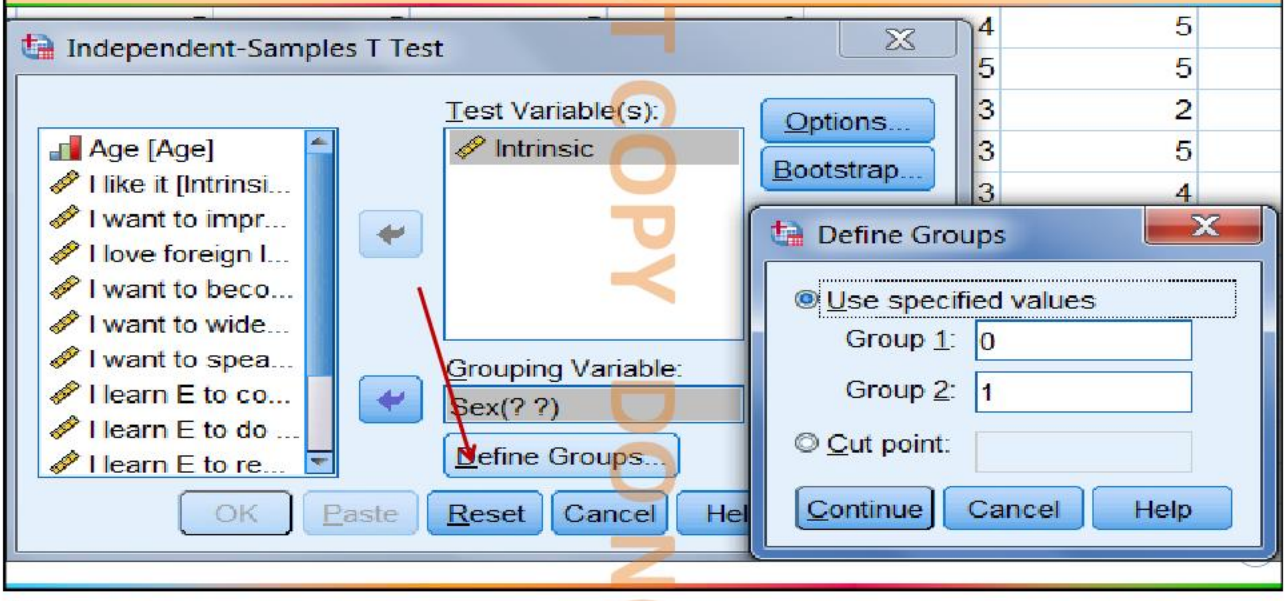
45

6.1 Independent-samples T-test



46

6.1 Independent-samples T-test



47

6.1 Independent-samples T-test

T-Test

Group Statistics

	Sex	N	Mean	Std. Deviation	Std. Error Mean
Intrinsic	0	18	3.8056	.46880	.11050
	1	32	3.5417	1.07597	.19021

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					
								95% Confidence Interval of the Difference	
		F	Sig	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Intrinsic	Equal variances assumed	15.067	.000	.986	48	.329	.26389	.26769	-.27434	.80212
	Equal variances not assumed			1.200	45.918	.236	.26389	.21997	-.17891	.70669

Conclusion: No statistically significant difference between these two means. In other words, male students' level of intrinsic motivation is not different from that of their female counterparts.

48

6.1 Independent-samples T-test

Reporting

[Purpose] [Name of statistical test]

[Test result] [Result intepretation]

[Conclusion if necessary]

49

49

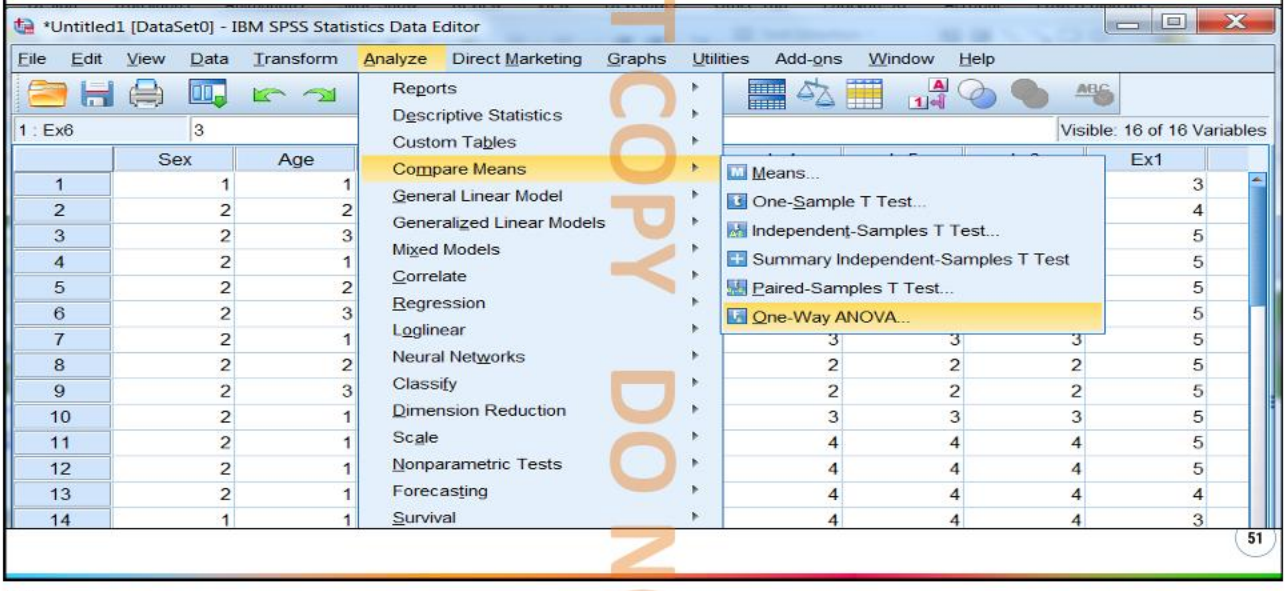
6.2 One-way ANOVA

- Comparing means of a variable from more than 2 different groups.
- In this case, the one group' mean of intrinsic motivation level is compared to other groups' means of intrinsic motivation level.
- Refer to research question 2.

50

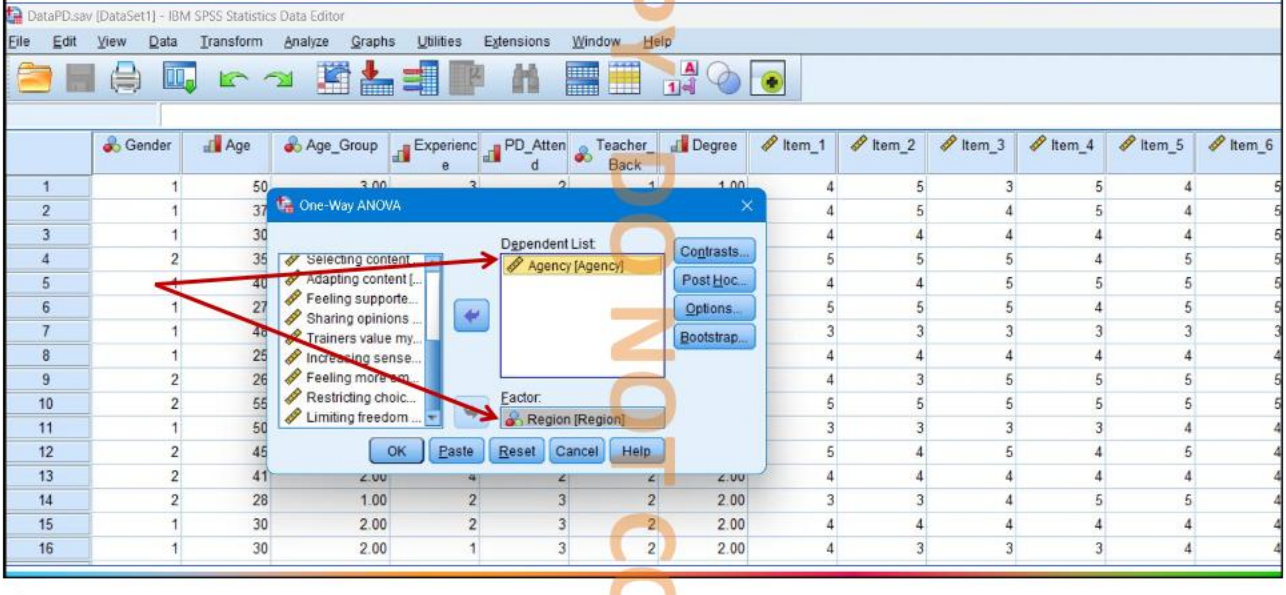
50

6.2 One-way ANOVA



51

6.2 One-way ANOVA



52

6.2 One-way ANOVA

ANOVA

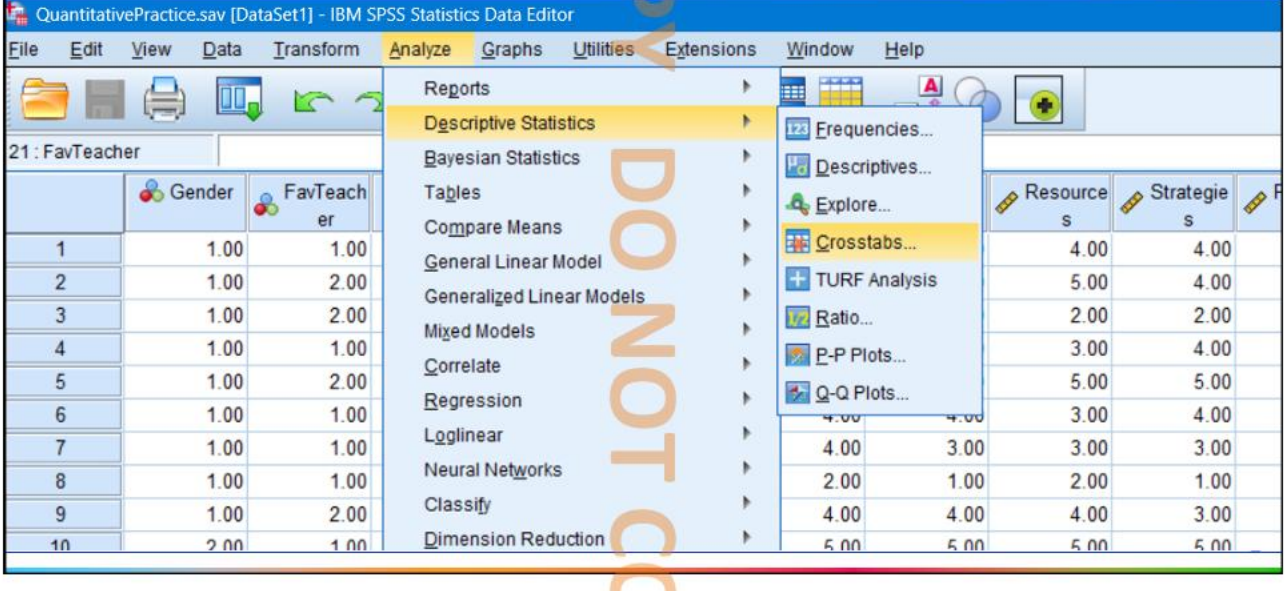
Agency

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.538	2	.269	1.037	.372
Within Groups	5.451	21	.260		
Total	5.990	23			

Conclusion: No statistically significant difference between these three means. In other words, teachers' level of agency is not different across three regions.

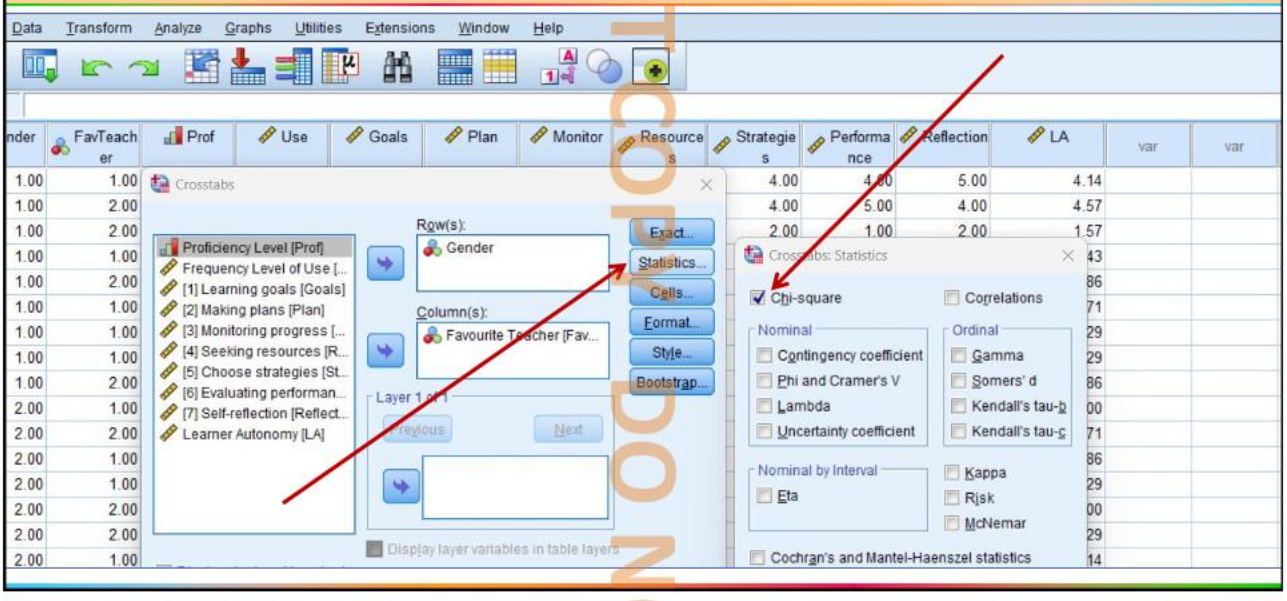
53

6.3 Chi-Square



54

6.3 Chi-Square



55

6.3 Chi-Square

Chi-Square Tests

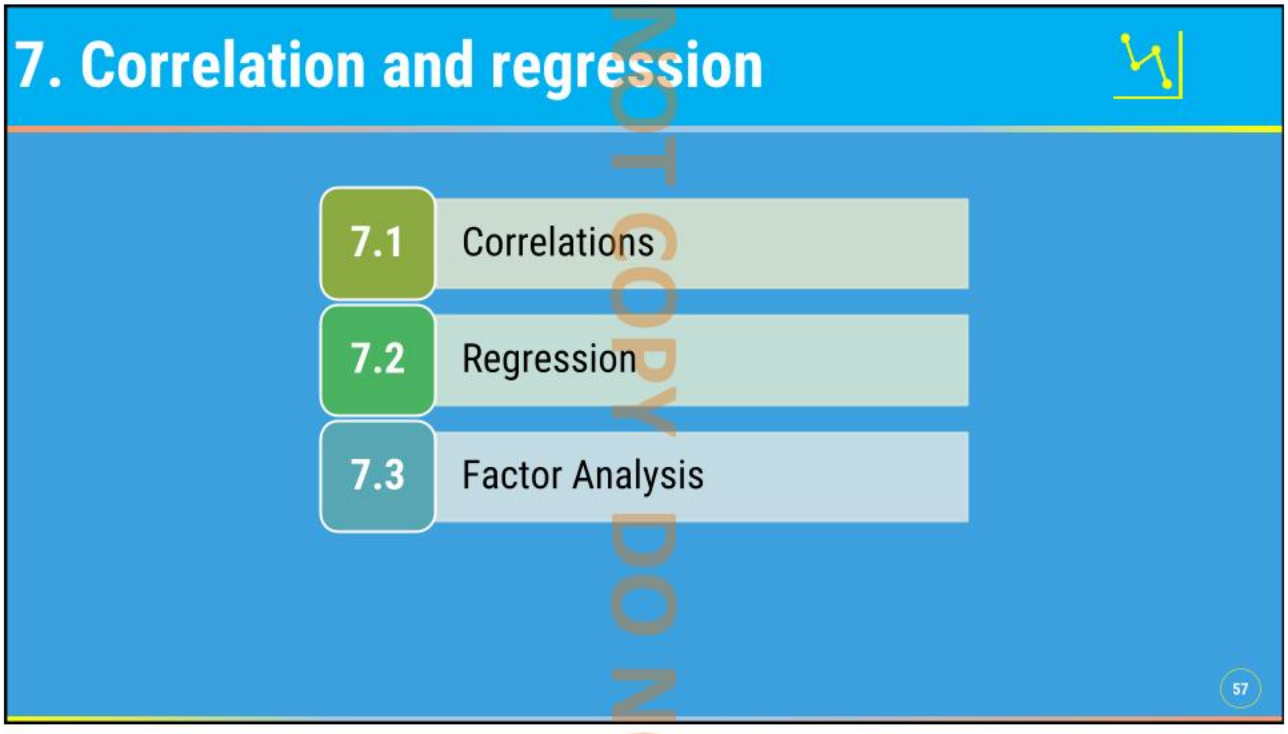
	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.202[a]	1	.653		
Continuity Correction[b]	.000	1	1.000		
Likelihood Ratio	.202	1	.653		
Fisher's Exact Test				1.000	.500
Linear-by-Linear Association	.192	1	.661		
N of Valid Cases	20				

a. 2 cells (50.0%) have expected count less than 5. The minimum expected count is 4.50.

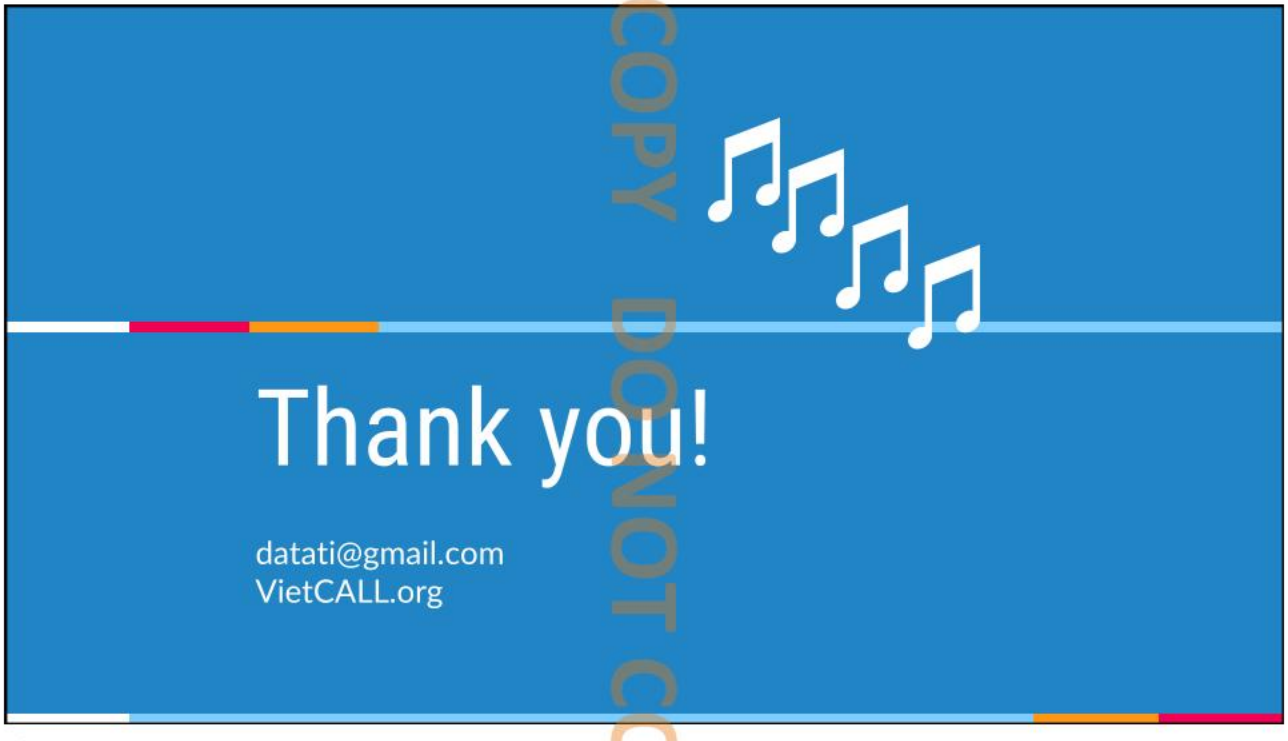
b. Computed only for a 2x2 table

56

56



57



58