



Southeast Asian Ministers of Education Organization
Regional Training Center

QUANTITATIVE DATA ANALYSIS WITH SPSS

Tin T. Dang, PhD

datati@gmail.com | VietCALL.org

1



Objectives

- Design meaningful and measurable variables that align with research questions and reflect theoretical constructs
- Classify, recode, and manage variables to prepare them for accurate and efficient analysis
- Construct clear and organized codebooks to standardize data coding and support replicable research practices

2

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

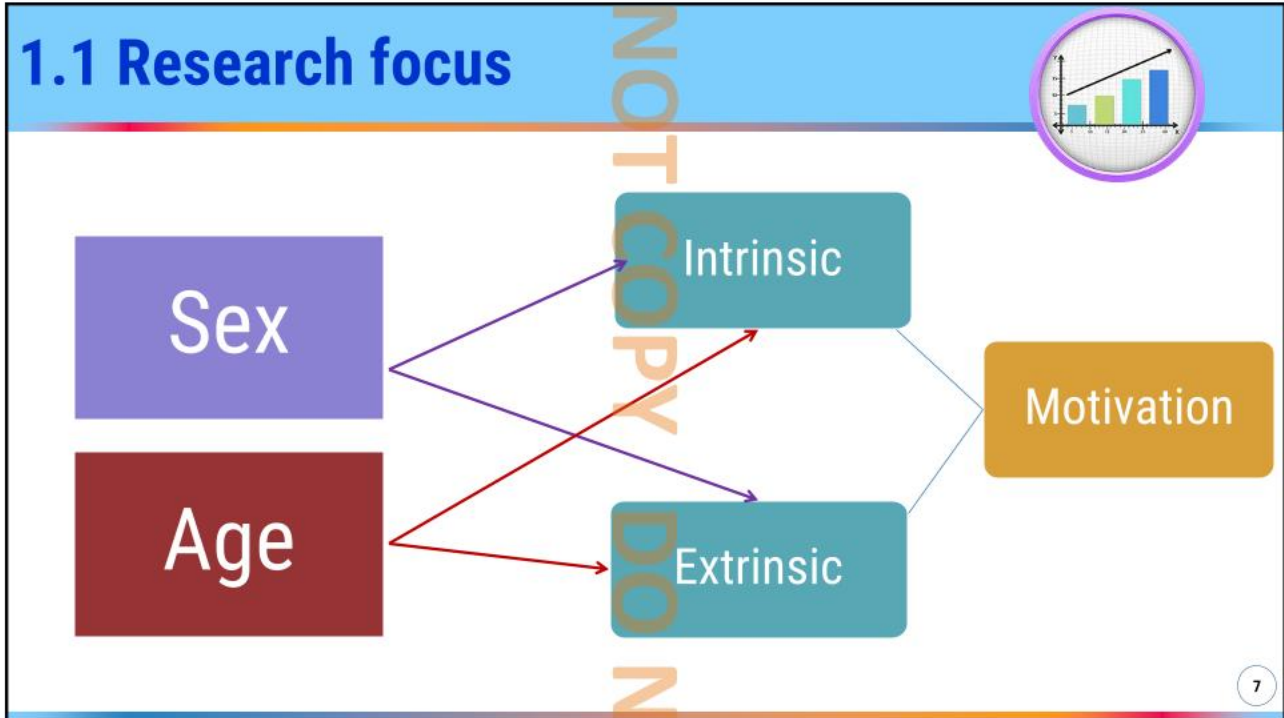
- 1.1 Research focus
- 1.2 Instrument drafting
- 1.3 Validity of the instrument
- 1.4 Data collection

5

1.1 Research focus

```
graph LR; Learning[Learning] --- LK[Learning for knowledge]; Learning --- LR[Learning for rewards]; LK --- Intrinsic[Intrinsic]; LR --- Extrinsic[Extrinsic]; Intrinsic --- Motivation[Motivation]; Extrinsic --- Motivation;
```

6



1.1 Research focus

Research 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

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

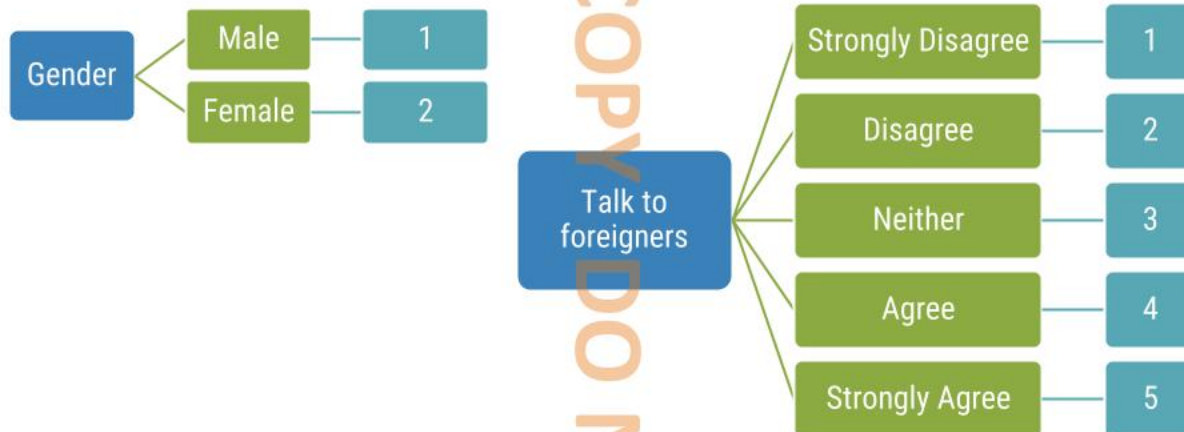
Align	Measure	Role
Right	Nominal	Input
	Scale	
	Ordinal	
	Nominal	

16

2.2 Coding variables



Assigning values to levels of variables



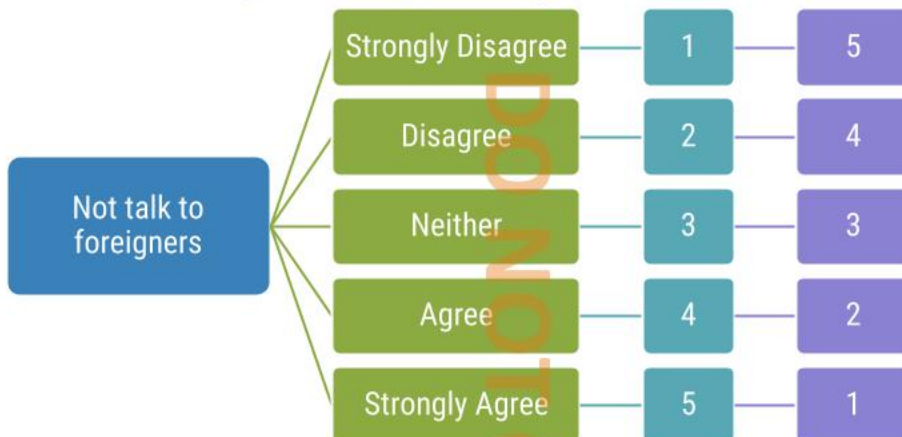
17

17

2.2 Coding variables



Reverse coding for negatively phrased items in scales



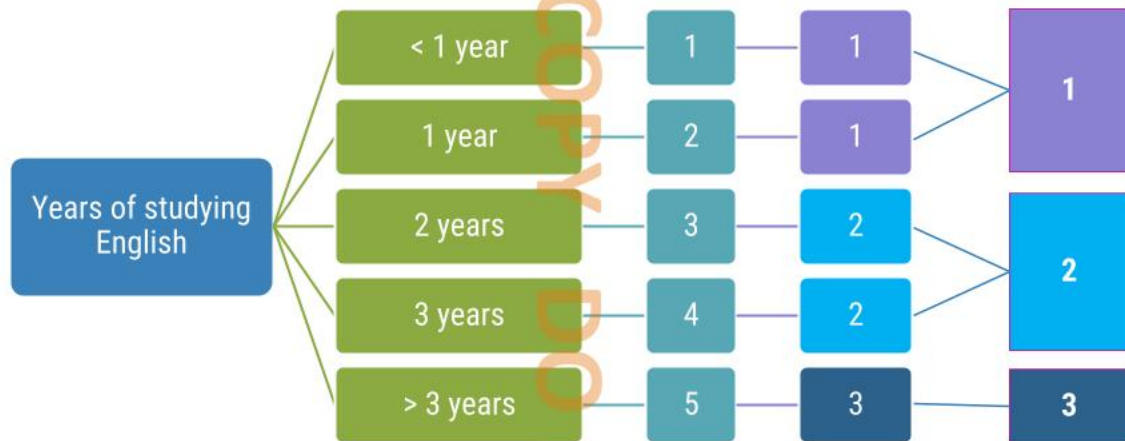
18

18

2.2 Coding variables



■ Combining categories



19

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

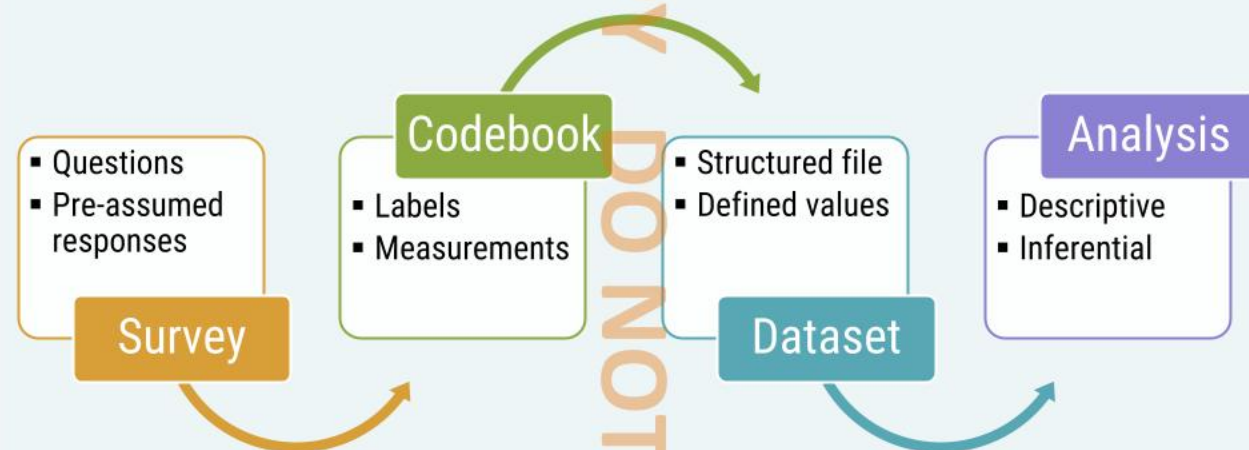


- A structured document that describes the variables, values, and coding rules in a dataset
- Acting as a reference guide for researchers and collaborators
- Essential for transparency, replication, and data sharing.

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

Value Labels

Value Labels

Value:

Label:

- 1 = "LS1"
- 2 = "LS2"
- 3 = "LS3"
- 4 = "LS4"
- 5 = "LS5"

27

4. Data Screening



- 4.1** Entering data
- 4.2** Checking missing values
- 4.3** Normal distribution
- 4.4** Reliability test

28

4.1 Entering data

The screenshot shows the SPSS Variable View window. The table below represents the data structure shown in the interface:

	Name	Type	Wi...	De...	Label	Values	Missing	Columns
1	Sex	Nume...	8	0	Sex	{1, Male}...	None	8
2	Age	Nume...	8	0	Age	None	None	8
3								
4								
5								
6								
7								

At the bottom, the 'Variable View' tab is highlighted with a red box and a red arrow points to it from the 'Data View' tab.

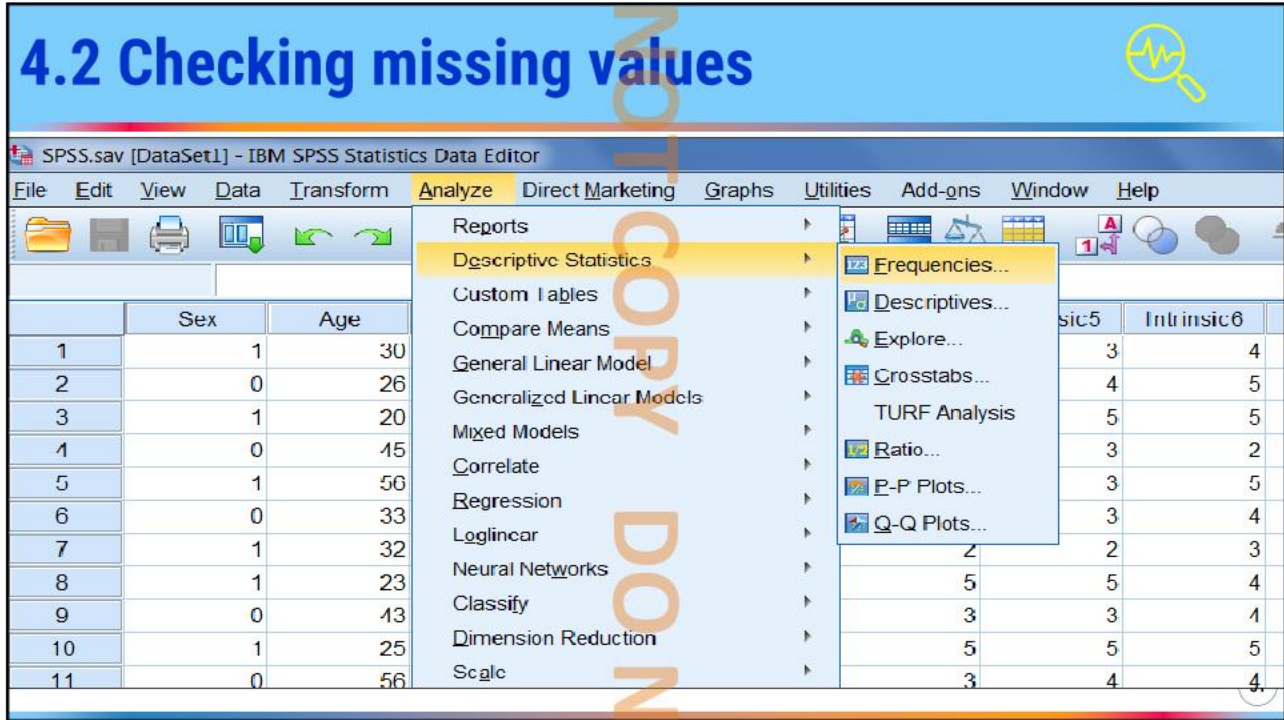
29

4.1 Entering data

The screenshot shows the SPSS Data View window. The grid is currently empty, with the header row containing ten 'var' labels. The status bar at the bottom indicates 'Visible: 0 of 0 Variables'. The 'Data View' tab is highlighted with a red box, and a red arrow points to it from the 'Variable View' tab.

30

4.2 Checking missing values

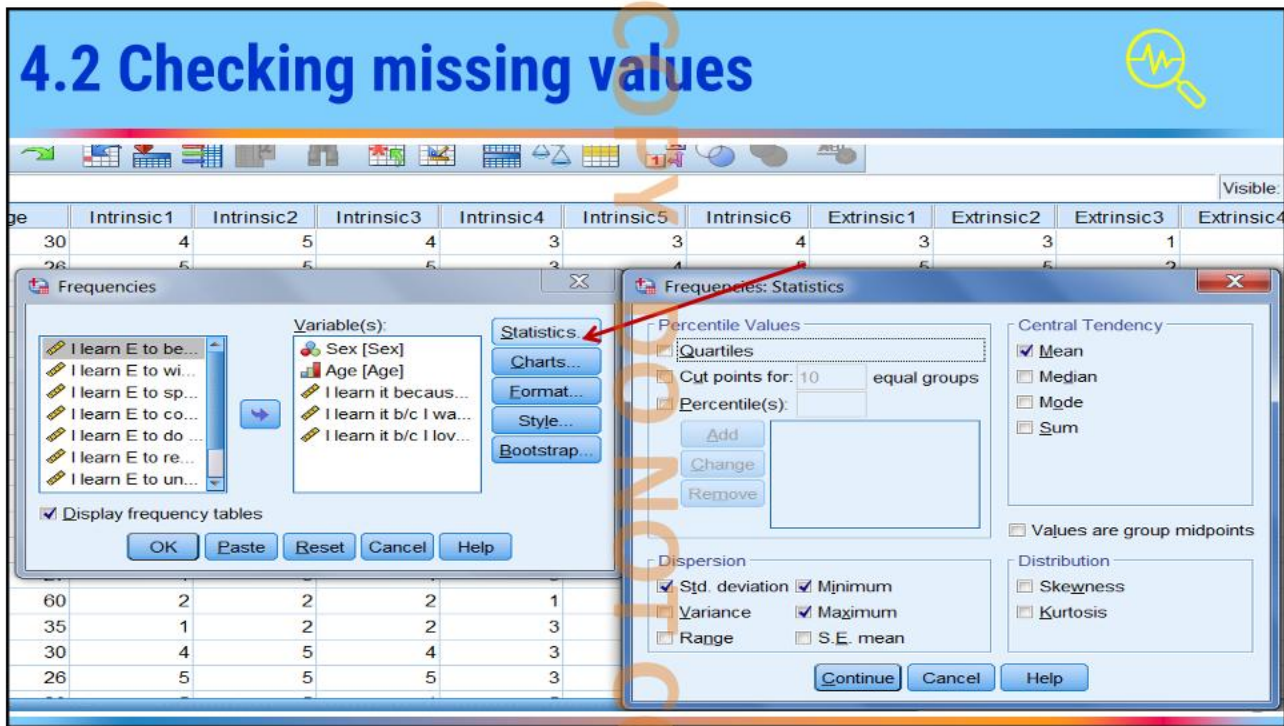


The screenshot shows the SPSS Statistics Data Editor interface. The 'Analyze' menu is open, and the path 'Analyze > Descriptive Statistics > Frequencies...' is highlighted. The background shows a data table with columns 'Sex' and 'Age'.

	Sex	Age
1	1	30
2	0	26
3	1	20
4	0	45
5	1	50
6	0	33
7	1	32
8	1	23
9	0	43
10	1	25
11	0	56

31

4.2 Checking missing values



The screenshot shows the 'Frequencies' dialog box in SPSS. The 'Statistics' sub-dialog is open, showing various statistical options. A red arrow points to the 'Statistics' button in the main dialog. The background shows a data table with columns 'Intrinsic1' through 'Intrinsic6' and 'Extrinsic1' through 'Extrinsic6'.

	Intrinsic1	Intrinsic2	Intrinsic3	Intrinsic4	Intrinsic5	Intrinsic6	Extrinsic1	Extrinsic2	Extrinsic3	Extrinsic4
30	4	5	4	3	3	4	3	3	1	
26	5	5	5	3	4	3	5	5	2	
60	2	2	2	1						
35	1	2	2	3						
30	4	5	4	3						
26	5	5	5	3						

Frequencies Statistics

- Percentile Values
 - Quartiles
 - Cut points for: 10 equal groups
 - Percentile(s):
 - Buttons: Add, Change, Remove
- Dispersion
 - Std. deviation
 - Variance
 - Range
 - Minimum
 - Maximum
 - S.E. mean
- Central Tendency
 - Mean
 - Median
 - Mode
 - Sum
- Values are group midpoints
- Distribution
 - Skewness
 - Kurtosis

32



4.2 Checking missing values

Frequencies

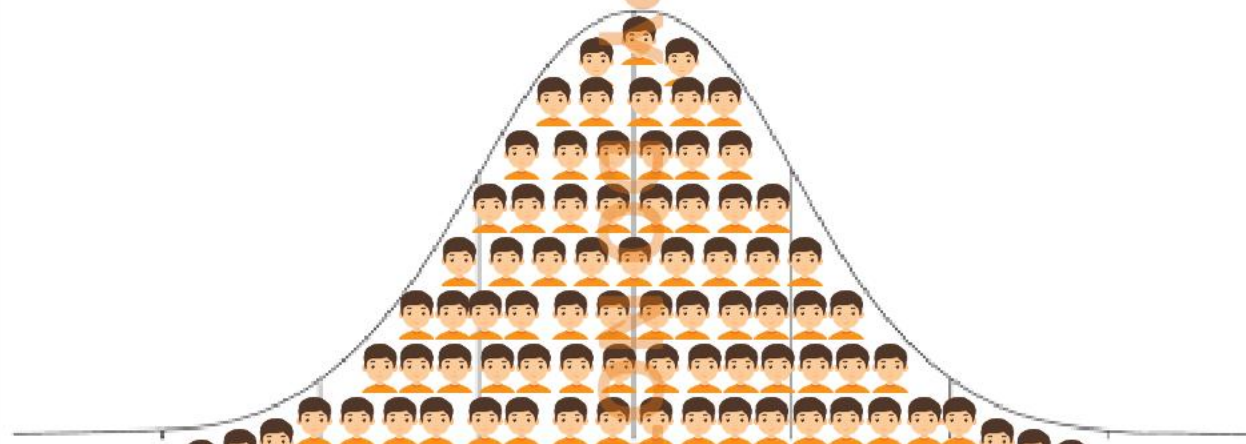
Statistics

		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

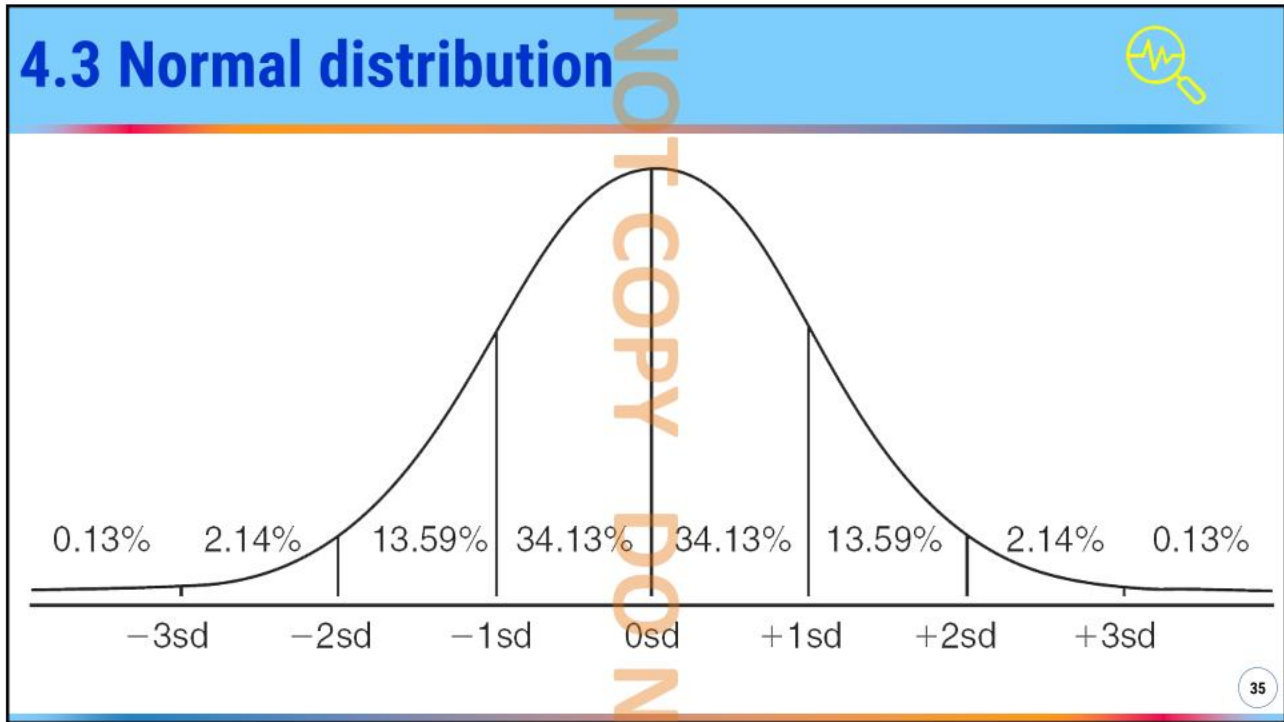


4.3 Normal distribution

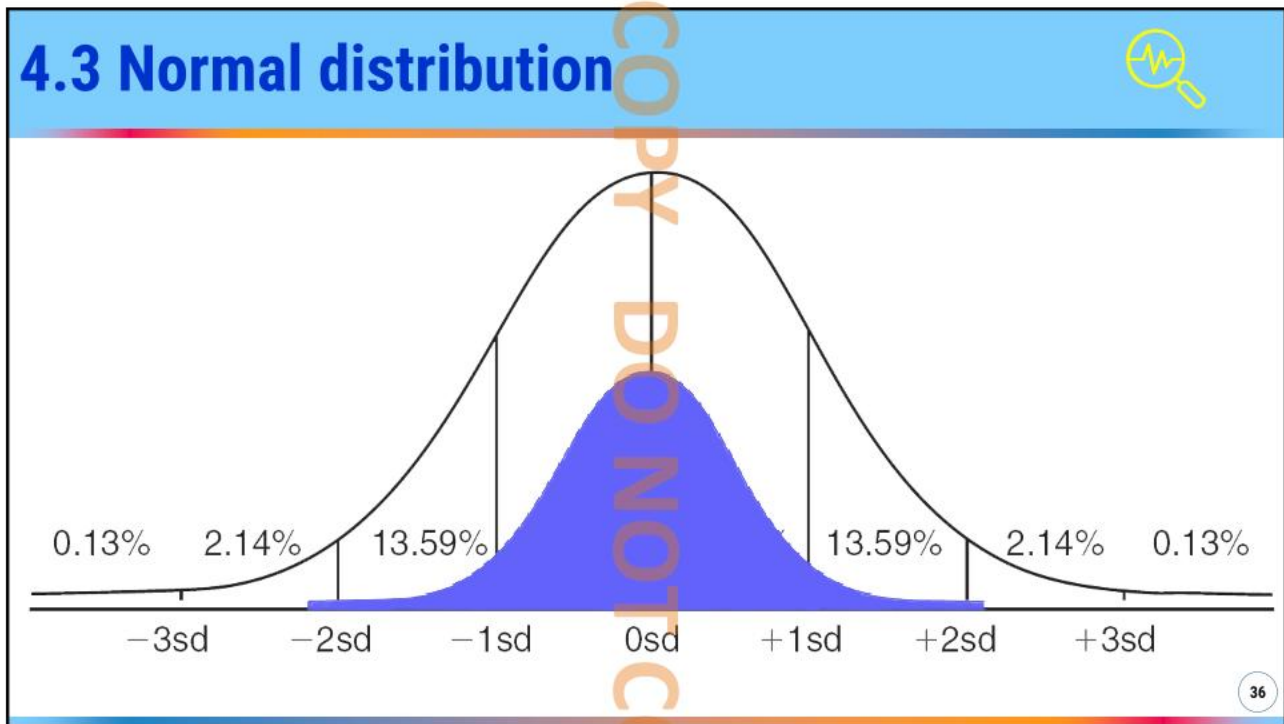


34

34



35



36

4.3 Normal distribution

The screenshot shows the SPSS Statistics Viewer interface. The 'Analyze' menu is open, and 'Explore...' is selected. The main window displays a 'Frequencies' table for the variable 'Sex'.

Sex	Frequency	Percentage
1	5	10.0%
2	14	28.0%
Total	19	100.0%

The 'Frequencies' table in the main window shows the following data:

5	14
Total	50

The 'Analyze' menu options include: Reports, Descriptive Statistics, Custom Tables, Compare Means, General Linear Model, Generalized Linear Models, Mixed Models, Correlate, Regression, Loglinear, Neural Networks, Classify, Dimension Reduction, Scale, Frequencies..., Descriptives..., Explore..., Crosstabs..., TURF Analysis, Ratio, P-P Plots..., and Q-Q Plots...

37

4.3 Normal distribution

The screenshot shows the 'Explore' dialog box and the 'Explore: Plots' sub-dialog box. The 'Explore' dialog has 'I like it [intrinsic1]' as the dependent variable and 'Sex' as the factor. The 'Explore: Plots' dialog has 'Normality plots with tests' checked.

The 'Explore' dialog box shows the following settings:

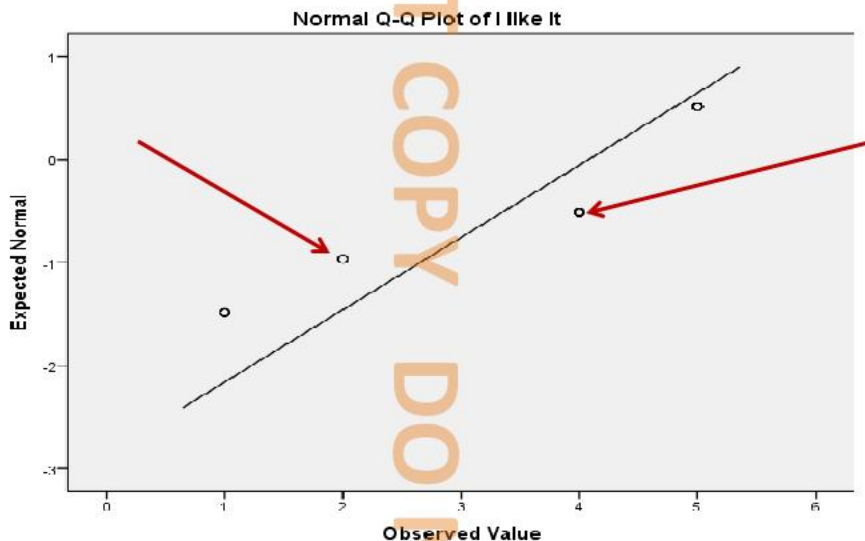
- Dependent List: I like it [intrinsic1]
- Factor List: Sex [Sex]
- Display: Both

The 'Explore: Plots' dialog box shows the following settings:

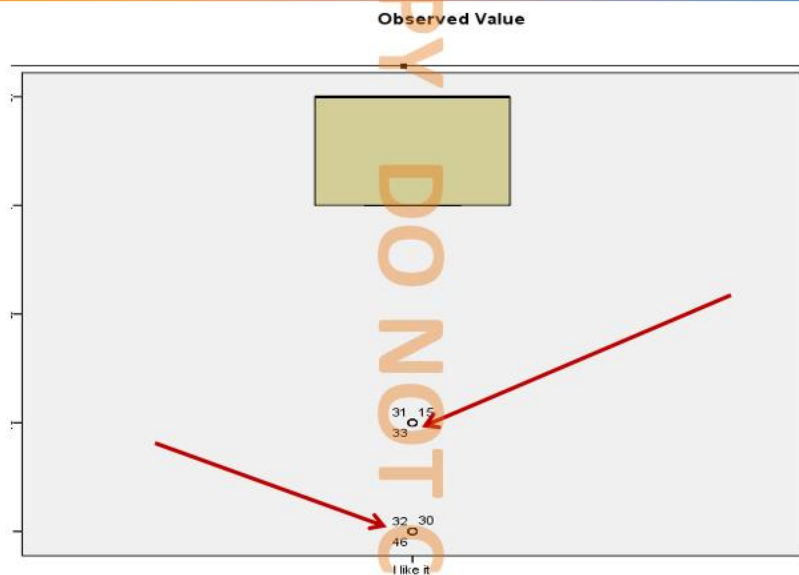
- Boxplots: Factor levels together
- Normality plots with tests:
- Spread vs Level with Levene Test: None

38

4.3 Normal distribution



4.3 Normal distribution



4.4 Reliability test

The screenshot shows the SPSS Data Editor interface. The 'Analyze' menu is open, and 'Reliability Analysis...' is highlighted. The background data table is partially visible.

	Sex	Age	Intrinsic4	Intrinsic5	Intrinsic6	Extrinsic1	Extrinsic2	Extrinsic3
1	1	30	3	3	4	3	3	3
2	0	26	3	4	5	5	5	5
3	1	20	5	5	5	4	4	4
4	0	45	2	3	2	2	1	1
5	1	56	4	3	5	5	5	5
6	0	33	2	3	4	5	4	4
7	1	32	2	2	3	1	1	1
8	1	23	5	5	4	5	4	4
9	0	43	3	3	4	3	2	2
10	1	25	5	5	5	4	5	5
11	0	56				2	3	3
12	1	21				1	2	2
13	0	32				2	4	4
14	1	27				4	2	2
15	1	60				2	4	4
16	1	35				3	5	5
17	1	30				3	3	3

41

4.4 Reliability test

The screenshot shows the 'Reliability Analysis' dialog box with the 'Statistics...' sub-dialog box open. The 'Reliability Analysis: Statistics' dialog is also open, showing options for Descriptives, Summaries, ANOVA Table, and Model.

Reliability Analysis Dialog:

- Model: Alpha
- Scale label: (empty)
- Items:
 - I like it [Intrinsic1]
 - I want to improve my ski...
 - I love foreign languages ...
 - I want to become a goo...
 - I want to widen my kno...
 - I want to speak E fluent!

Reliability Analysis: Statistics Dialog:

- Descriptives for:
 - Item
 - Scale
 - Scale if item deleted
- Inter-Item:
 - Correlations
 - Covariances
- Summaries:
 - Means
 - Variances
 - Covariances
 - Correlations
- ANOVA Table:
 - None
 - F test
 - Friedman chi-square
 - Cochran chi-square
- Hotelling's T-square:
- Tukey's test of additivity:
- Intraclass correlation coefficient:
- Model: Two-Way Mixed
- Type: Consistency
- Confidence interval: 95 %
- Test value: 0

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

6. Simple Inferential Statistics



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

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



The screenshot shows the SPSS Data Editor interface. The 'Analyze' menu is open, and 'Independent-Samples T Test' is highlighted. The data table below shows variables 'Scx' and 'Age'.

	Scx	Age
1	1	30
2	0	26
3	1	20
4	0	45
5	1	56
6	0	33
7	1	32
8	1	23
9	0	43
10	1	25
11	0	56
12	1	21
13	0	22

46

6.1 Independent-samples T-test



47

6.1 Independent-samples T-test



T-Test

Group Statistics				
	Sex	N	Mean	Std. Deviation
Intrinsic	0	18	3.8056	.46880
	1	32	3.5417	1.07597

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Intrinsic	Equal variances assumed	15.067	.000	.986	48	.329	-.26309	.26769	-.27434	.00212
	Equal variances not assumed			1.200	45.918	.236	-.26309	.21997	-.17091	.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 interpretation]

[Conclusion if necessary]

49

49

6.2 One-way ANOVA

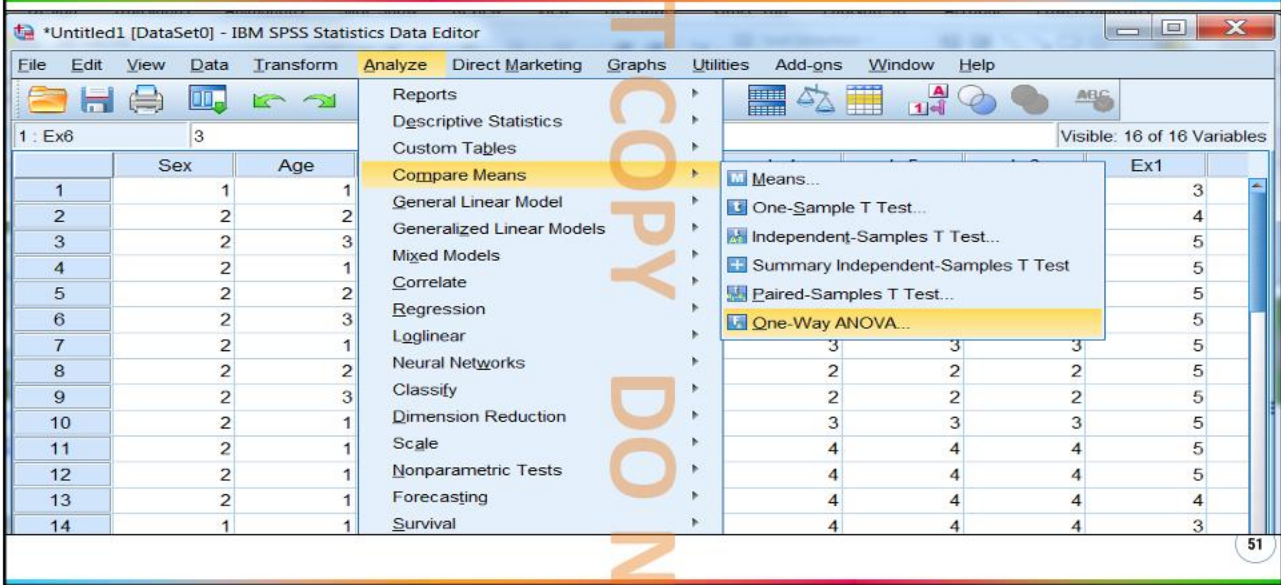


- Comparing means of a variable from more than 2 different groups.
- In this case, the one group' mean of intrinsic motivation level is compared to other groups' means of intrinsic motivation level.
- Refer to research question 2.

50

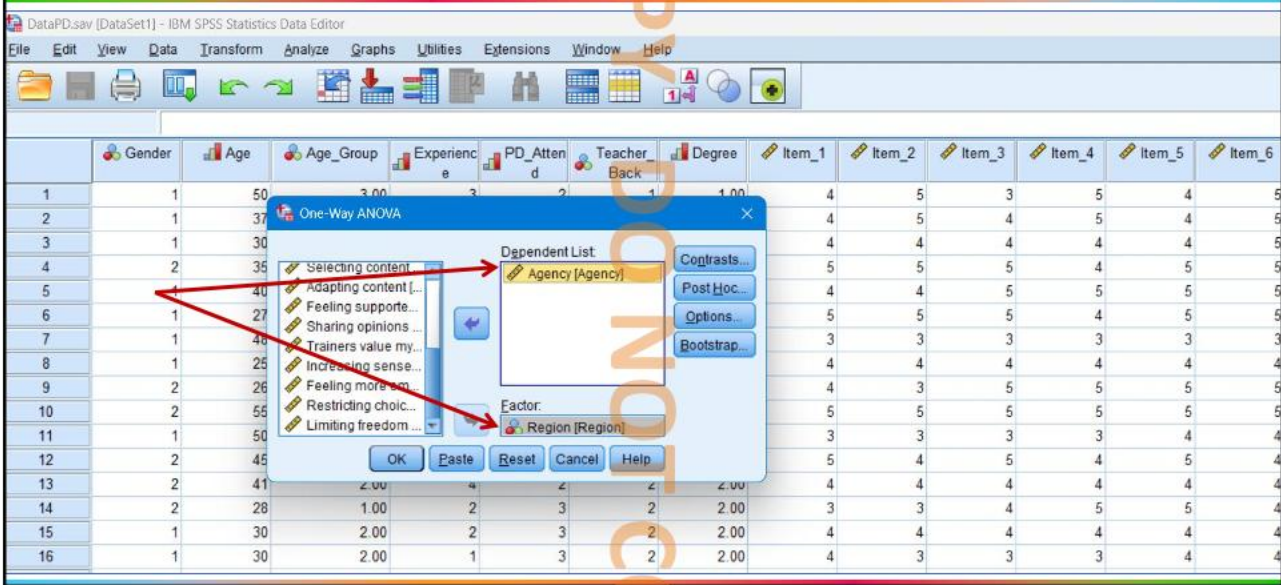
50

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	0.372
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



The screenshot shows the SPSS Statistics Data Editor window for 'QuantitativePractice.sav'. The 'Analyze' menu is open, and 'Crosstabs...' is highlighted. The main window displays a data table with columns 'Gender' and 'FavTeacher'.

	Gender	FavTeacher
1	1.00	1.00
2	1.00	2.00
3	1.00	2.00
4	1.00	1.00
5	1.00	2.00
6	1.00	1.00
7	1.00	1.00
8	1.00	1.00
9	1.00	2.00
10	2.00	1.00

54

6.3 Chi-Square

The screenshot shows the SPSS Crosstabs dialog box with 'Proficiency Level [Prof]' in the Row(s) list and 'Gender' in the Column(s) list. The Statistics sub-dialog box is open, showing 'Chi-square' checked under the 'Nominal' section. Other options like 'Contingency coefficient', 'Phi and Cramer's V', and 'Lambda' are unchecked. The background shows a data table with variables like 'FavTeacher', 'Prof', 'Use', 'Goals', 'Plan', 'Monitor', 'Resources', 'Strategies', 'Performance', 'Reflection', and 'LA'.

55

6.3 Chi-Square

	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

7. Correlation and regression

- 7.1 Correlations
- 7.2 Regression
- 7.3 Factor Analysis

57



Thank you!

datati@gmail.com
VietCALL.org

58