

Ευαγγελία-Γεωργία Κωστάκη

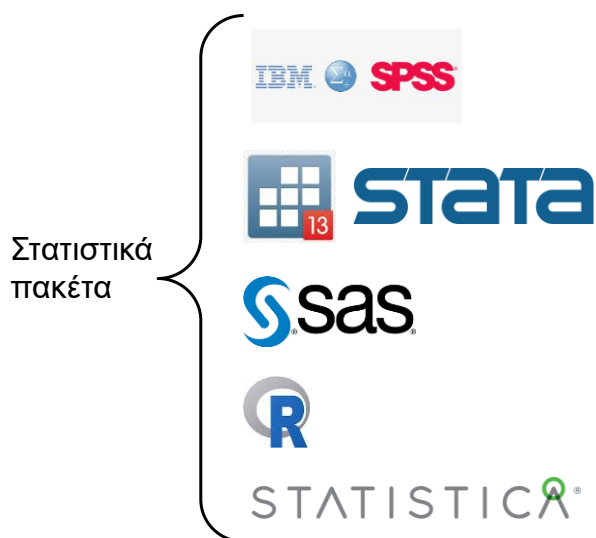
Βιοστατιστικός-Επιδημιολόγος PhD,
Μέλος ΕΔΙΠ
Εργαστήριο Υγιεινής,
Επιδημιολογίας και Ιατρικής
Στατιστικής
Ιατρική Σχολή, Εθνικό και
Καποδιστριακό Πανεπιστήμιο
Αθηνών

ekostakh@med.uoa.gr

**Χρήση στατιστικών πακέτων και εισαγωγή στο SPSS
Επαναληπτικό Ι**

1

Ο Ρόλος των Η/Υ στη Στατιστική



2

Εισαγωγικά στο SPSS



- Ολοκληρωμένο πρόγραμμα διαχείρισης πληροφοριών (δεδομένων) και εφαρμογής στατιστικών αναλύσεων
- Ευρύ φάσμα στατιστικών αναλύσεων
- Εύκολος χειρισμός δεδομένων και γραφικών, συμβατότητα με άλλα πακέτα
- Εύκολος χειρισμός εντολών (menu bar)
- Διαφορετικά παράθυρα για ευκολότερο χειρισμό

3

Βήματα για τη στατιστική ανάλυση στο SPSS

- Εισαγωγή δεδομένων στον Data Editor
- Επιλογή κατάλληλης στατιστικής μεθόδου
- Επιλογή των σχετικών μεταβλητών για την ανάλυση
- Λήψη αποτελεσμάτων στον SPSS Viewer

4

Συλλογή δεδομένων [προκαταρκτικό βήμα] (I)

Μέσω Ερωτηματολογίου

ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ		
Ηλικία μητέρας	[Εισάγετε τον αριθμό]
Βάρος μητέρας (κ.γ.)	
Ύψος μητέρας (εκ.)	
Προγεστερόνη 1η μετρ. (nmol/L)	
Προγεστερόνη 2η μετρ. (nmol/L)	
Φύλο μωρού	<input type="checkbox"/> Αγόρι <input type="checkbox"/> Κορίτσι	Επιλέξτε το αντίστοιχο τετράγωνο
Βάρος μωρού (γρ.)	
...		

5

Συλλογή δεδομένων [προκαταρκτικό βήμα] (II)

Μέσω Ερωτηματολογίου

ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ		
Ηλικία μητέρας	..32.....	[Εισάγετε τον αριθμό]
Βάρος μητέρας (κ.γ.)	..59.....	
Ύψος μητέρας (εκ.)	170.....	
Προγεστερόνη 1η μετρ. (nmol/L)	187.....	
Προγεστερόνη 2η μετρ. (nmol/L)	190.....	
Φύλο μωρού	<input checked="" type="checkbox"/> 1 <input checked="" type="checkbox"/> 2	Αγόρι Κορίτσι Επιλέξτε το αντίστοιχο τετράγωνο
Βάρος μωρού (γρ.)	3700...	
...		

6

Συλλογή δεδομένων [προκαταρκτικό βήμα] (III)

Μέσω Ερωτηματολογίου

ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ			
Ηλικία μητέρας	[age]	.32....	[Εισάγετε τον αριθμό]
Βάρος μητέρας (κγ.)	[weight]	.59....	
Ύψος μητέρας (εκ.)	[height]	.170....	
Προγεστερόνη 1η μετρ. (nmol/L)		.187....	
Προγεστερόνη 2η μετρ. (nmol/L)		.190....	
Φύλο μωρού	[gender]	<input type="checkbox"/> 1 Αγόρι <input type="checkbox"/> 2 Κορίτσι	Επιλέξτε το αντίστοιχο τετράγωνο
Βάρος μωρού (γρ.)	[bweight]	.3700...	
...			

Μεταφορά δεδομένων από ερωτηματολόγιο στον Η/Υ με **κατάλληλη κωδικοποίηση** (ερώτηση → μεταβλητή)

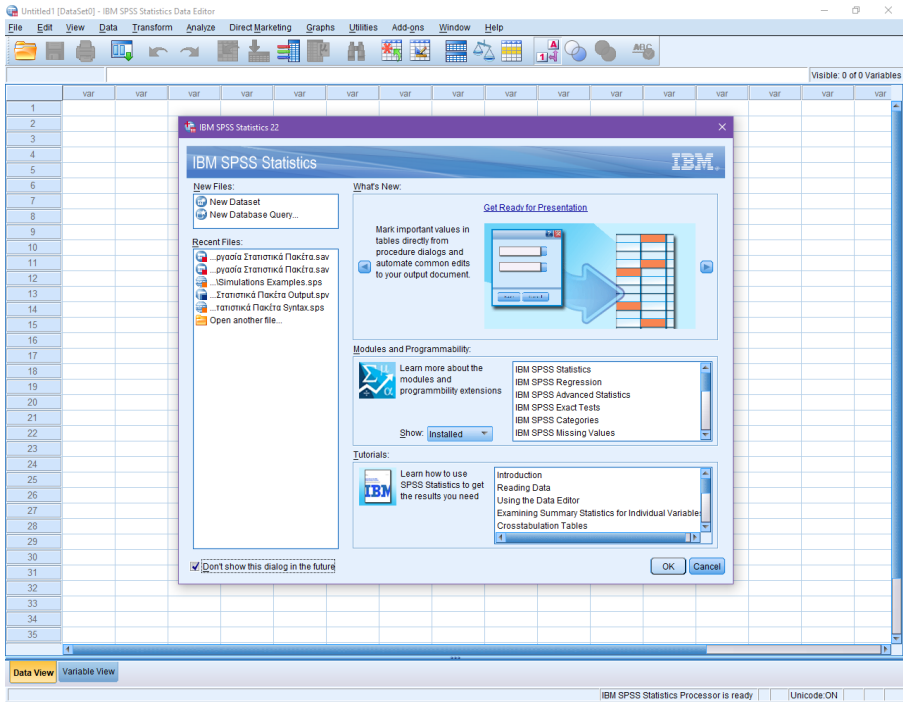
7

1ο Βήμα: Εισαγωγή δεδομένων

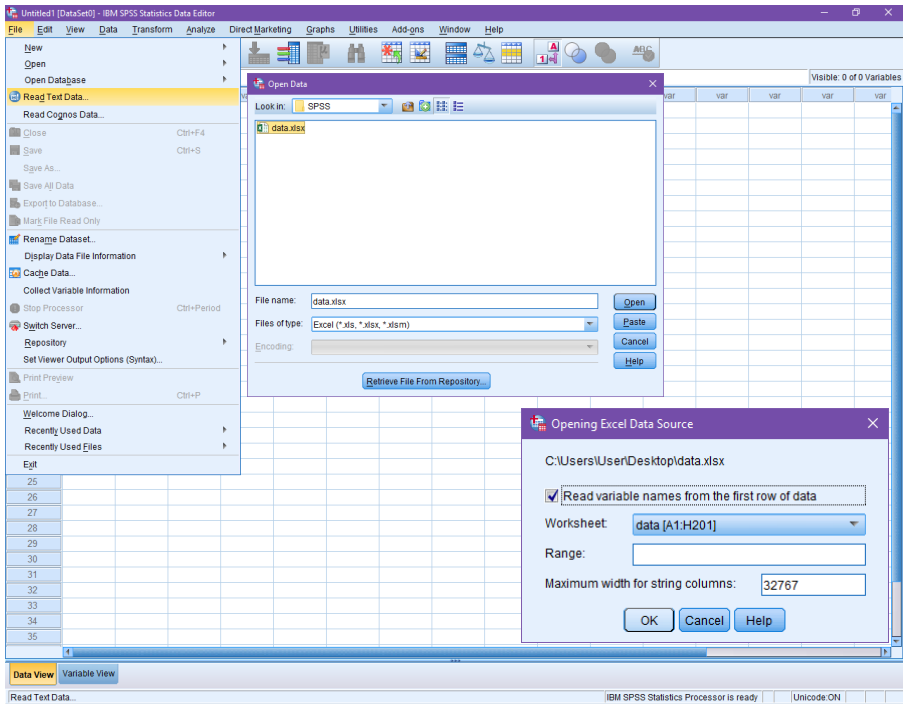
Εισαγωγή δεδομένων με πολλαπλούς τρόπους:

- ✓ Πληκτρολογώντας
(μέσα στα κελιά) με ENTER ή με τα βελάκια
- ✓ Αντιγραφή/επικόλληση (copy/paste)
(να δείχνει το σημείο που θέλουμε να επικολλήσουμε τα δεδομένα, συνήθως κελί "var1")
- ✓ Άνοιγμα προϋπάρχοντος αρχείου (.sav)
File ⇒ Open ⇒ Data
(Μπορεί να αποτελεί το 1^ο βήμα)

8



9



10

	id	age	CHD_group	male_sex	smoking	bmi	dyslipidemia	hdl	var	var	var
1	1	29	1	1	1	28.2	1	41.0			
2	2	35	1	1	1	36.4	1	53.0			
3	3	33	1	1	1	30.0	1	31.0			
4	4	35	1	0	1	20.7	1	45.0			
5	5	34	1	1	1	23.0	1	38.0			
6	6	25	1	0	1	25.4	1	45.0			
7	7	33	1	1	1	28.6	0	58.0			
8	8	36	1	1	1	30.0	1	51.0			
9	9	32	1	1	1	33.3	1	38.0			
10	10	30	1	1	1	28.2	1	27.0			
11	11	35	1	1	1	22.1	0	41.0			
12	12	34	1	1	1	37.0	0	29.0			
13	13	33	1	1	1	30.1	0	44.0			
14	14	30	1	1	0	39.7	0	45.0			
15	15	34	1	1	1	32.7	1	27.0			
16	16	35	1	1	1	24.7	1	37.0			
17	17	35	1	1	1	33.0	1	32.0			
18	18	33	1	1	0	35.8	1	31.0			
19	19	33	1	1	1	35.3	1	35.0			
20	20	35	1	1	1	27.7	0	32.0			
21	21	33	1	1	1	25.9	1	44.0			
22	22	35	1	1	1	21.8	1	43.0			
23	23	33	1	1	1	31.8	1	44.0			
24	24	29	1	0	1	26.4	0	68.0			
25	25	28	1	1	1	21.7	0	38.0			
26	26	30	1	1	1	34.7	1	29.0			
27	27	32	1	0	1	19.7	0	43.0			
28	28	32	1	0	1	29.7	1	35.0			
29	29	33	1	1	1	23.7	0	43.0			
30	30	33	1	1	1	25.6	0	66.0			
31	31	36	1	1	1	30.4	1	18.0			
32	32	34	1	1	1	29.0	0	36.0			
33	33	27	1	1	1	25.6	0	36.0			
34	34	26	1	1	1	25.4	1	50.0			
35	35	35	1	1	1	26.7	1	35.0			

11

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	id	Numeric	12	0		None	None	12	Right	Scale	Input
2	age	Numeric	12	0		None	None	12	Right	Scale	Input
3	CHD_group	Numeric	12	0		None	None	12	Right	Nominal	Input
4	male_sex	Numeric	12	0		None	None	12	Right	Nominal	Input
5	smoking	Numeric	12	0		None	None	12	Right	Nominal	Input
6	bmi	Numeric	12	1		None	None	12	Right	Scale	Input
7	dyslipidemia	Numeric	12	0		None	None	12	Right	Nominal	Input
8	hdl	Numeric	12	1		None	None	12	Right	Scale	Input
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											

2^ο Βήμα: Προσδιορισμός μεταβλητών

12

Είδη μεταβλητών

- **Ποιοτικές** (qualitative)
 - Π.χ. φύλο, χαρακτηρισμός πτυχίου
 - Διακρίνονται σε κατηγορικές (categorical) και διατάξιμες (ordinal)
- **Ποσοτικές** (quantitative)
 - Π.χ. ύψος, θερμοκρασία, ηλικία
 - Διακρίνονται σε συνεχείς (continues) και διακριτές (discrete)

13

Παράδειγμα

- Ποιες από τις παρακάτω μεταβλητές είναι ποιοτικές και ποιες ποσοτικές;
- Από τις ποσοτικές, ποιες είναι διακριτές και ποιες συνεχείς;
 - α) Βάρος
 - β) Αριθμός τροχαίων ατυχημάτων
 - γ) Φυλή
 - δ) Αριθμός παιδιών στην οικογένεια
 - ε) Οικογενειακή κατάσταση
 - στ) Στάθμη λίμνης του Μαραθώνα
 - ζ) Νούμερο γυναικείων παπουτσιών

14

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	id	Numeric	12	0		None	None	12	Right	Scale	Input
2	age	Numeric	12	0		None	None	12	Right	Scale	Input
3	CHD_group	Numeric	12	0		None	None	12	Right	Nominal	Input
4	male_sex	Numeric	12	0		None	None	12	Right	Nominal	Input
5	smoking	Numeric	12	0		None	None	12	Right	Nominal	Input
6	bmi	Numeric	12	1		None	None	12	Right	Scale	Input
7	dyslipidemia	Numeric	12	0		None	None	12	Right	Nominal	Input
8	hdl	Numeric	12	1		None	None	12	Right	Scale	Input
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											

15

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	id	Numeric	12	0	code	None	None	12	Right	Scale	Input
2	age	Numeric	12	0	age	None	None	12	Right	Scale	Input
3	CHD_group	Numeric	12	0	coronary heart ...	None	None	12	Right	Nominal	Input
4	male_sex	Numeric	12	0	gender	None	None	12	Right	Nominal	Input
5	smoking	Numeric	12	0	smoking	None	None	12	Right	Nominal	Input
6	bmi	Numeric	12	1	bmi	None	None	12	Right	Scale	Input
7	dyslipidemia	Numeric	12	0	dyslipidemia	None	None	12	Right	Nominal	Input
8	hdl	Numeric	12	0	hdl	None	None	12	Right	Scale	Input

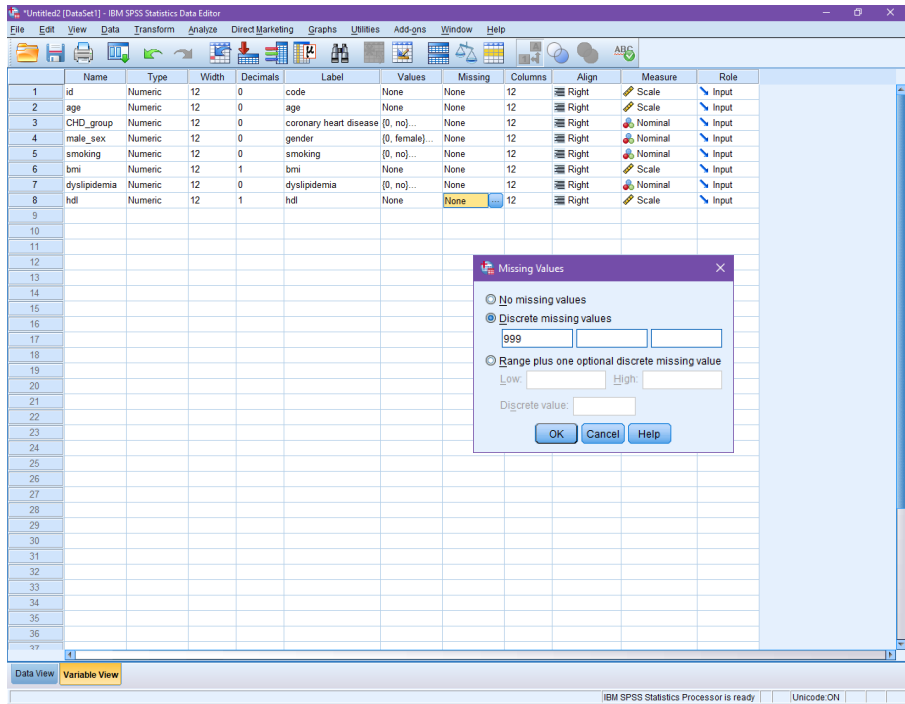
Value Labels

Value:

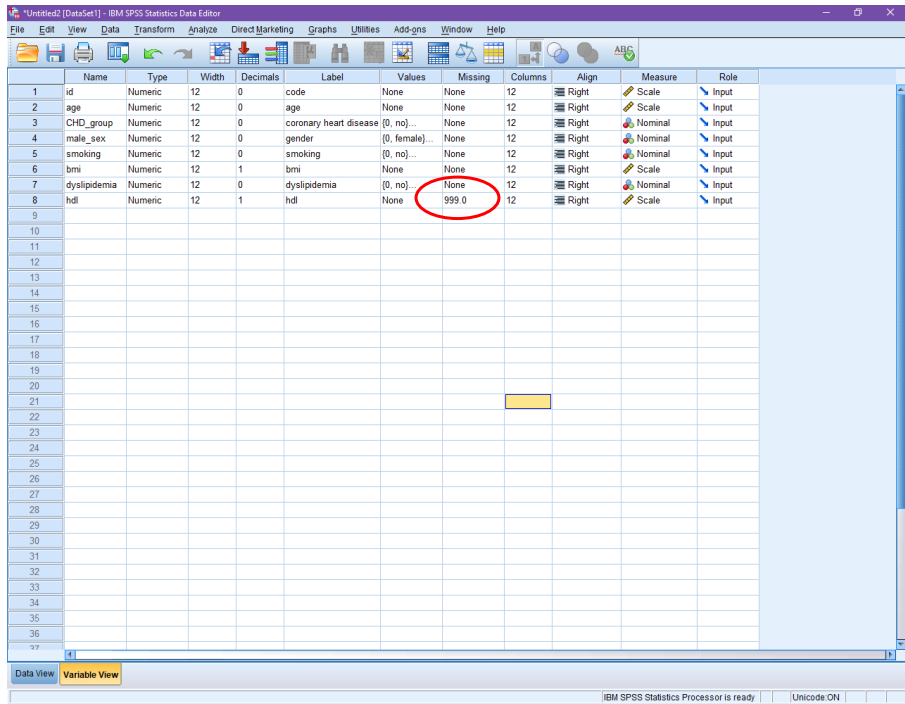
Label:

- 0 = "No"

16



17



18

	id	age	CHD_group	male_sex	smoking	bmi	dyslipidemia	hdl	var	var	var
1	1	29	yes	male	yes	28.2	yes	41.0			
2	2	35	yes	male	yes	36.4	yes	53.0			
3	3	33	yes	male	yes	30.0	yes	31.0			
4	4	35	yes	female	yes	20.7	yes	45.0			
5	5	34	yes	male	yes	23.0	yes	38.0			
6	6	25	yes	female	yes	25.4	yes	45.0			
7	7	33	yes	male	yes	28.6	no	58.0			
8	8	36	yes	male	yes	30.0	yes	51.0			
9	9	32	yes	male	yes	33.3	yes	38.0			
10	10	30	yes	male	yes	28.2	yes	27.0			
11	11	35	yes	male	yes	22.1	no	41.0			
12	12	34	yes	male	yes	37.0	no	29.0			
13	13	33	yes	male	yes	30.1	no	44.0			
14	14	30	yes	male	no	39.7	no	45.0			
15	15	34	yes	male	yes	32.7	yes	27.0			
16	16	35	yes	male	yes	24.7	yes	37.0			
17	17	35	yes	male	yes	33.0	yes	32.0			
18	18	33	yes	male	no	35.8	yes	31.0			
19	19	33	yes	male	yes	35.3	yes	35.0			
20	20	35	yes	male	yes	27.7	no	32.0			
21	21	33	yes	male	yes	25.9	yes	44.0			
22	22	35	yes	male	yes	21.8	yes	43.0			
23	23	33	yes	male	yes	31.8	yes	44.0			
24	24	29	yes	female	yes	26.4	no	68.0			
25	25	28	yes	male	yes	21.7	no	38.0			
26	26	30	yes	male	yes	34.7	yes	28.0			
27	27	32	yes	female	yes	19.7	no	43.0			
28	28	32	yes	female	yes	29.7	yes	35.0			
29	29	33	yes	male	yes	23.7	no	43.0			
30	30	33	yes	male	yes	25.6	no	66.0			
31	31	36	yes	male	yes	30.4	yes	18.0			
32	32	34	yes	male	yes	29.0	no	36.0			
33	33	27	yes	male	yes	25.6	no	36.0			
34	34	26	yes	male	yes	25.4	yes	50.0			
35	35	35	yes	male	yes	26.7	yes	35.0			

19

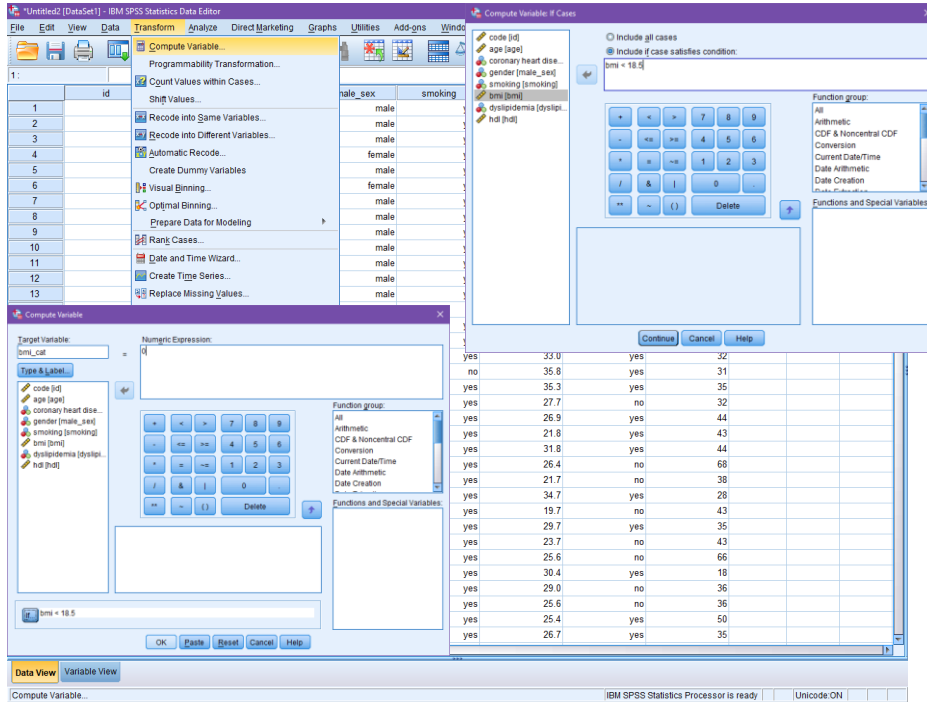
3^ο Βήμα: Χειρισμός των δεδομένων

Από το menu bar:

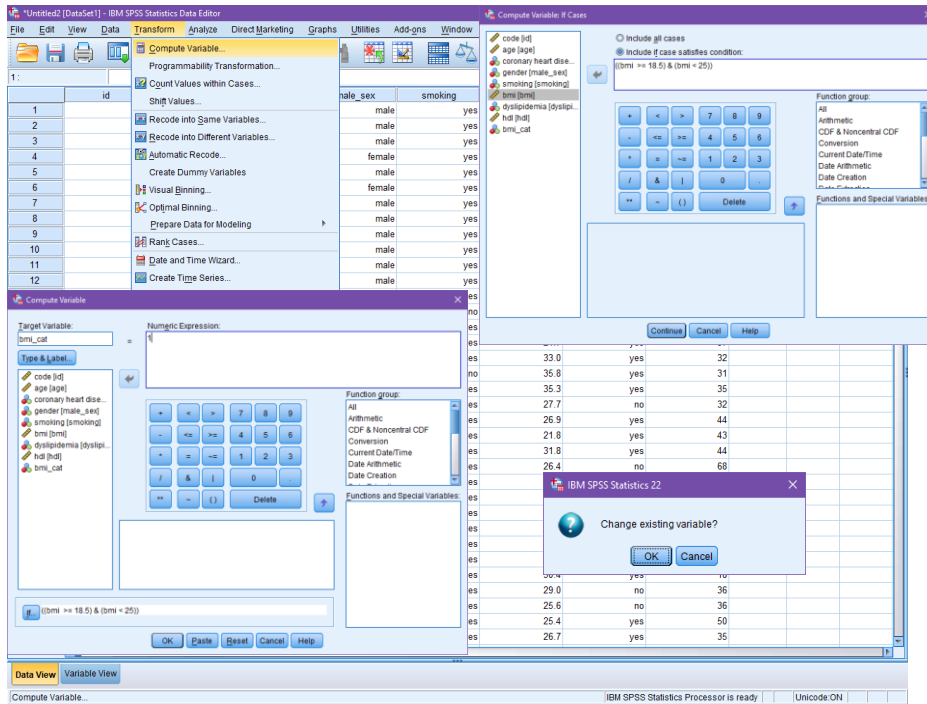
- ✓ Transform
 - Compute Variable
 - Recode into Same Variables
 - Recode into Different Variables

- ✓ Data
 - Sort Cases
 - Split File
 - Select Cases

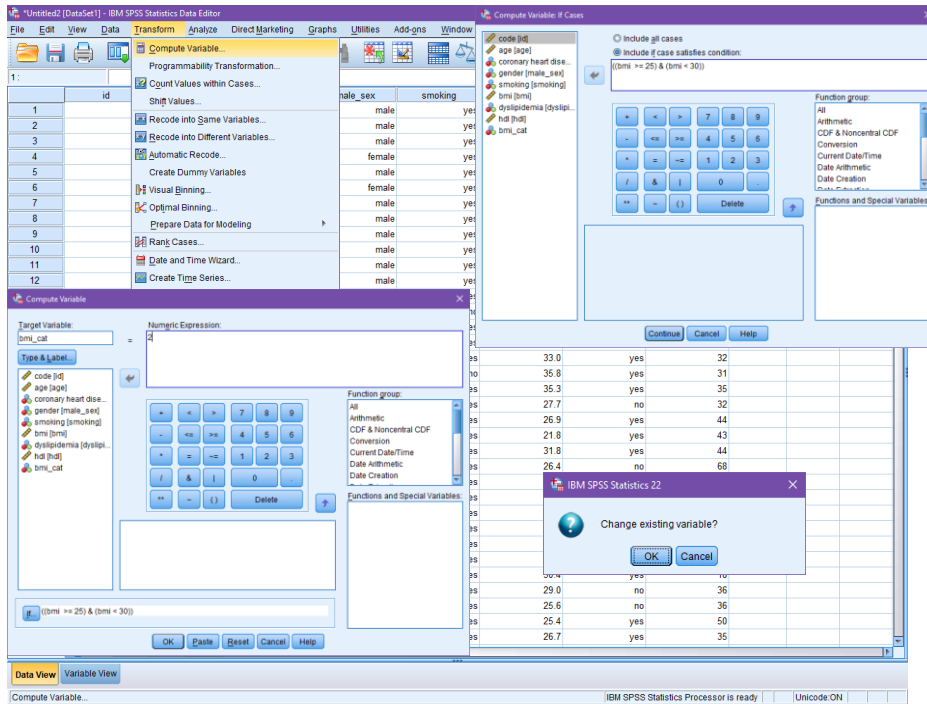
20



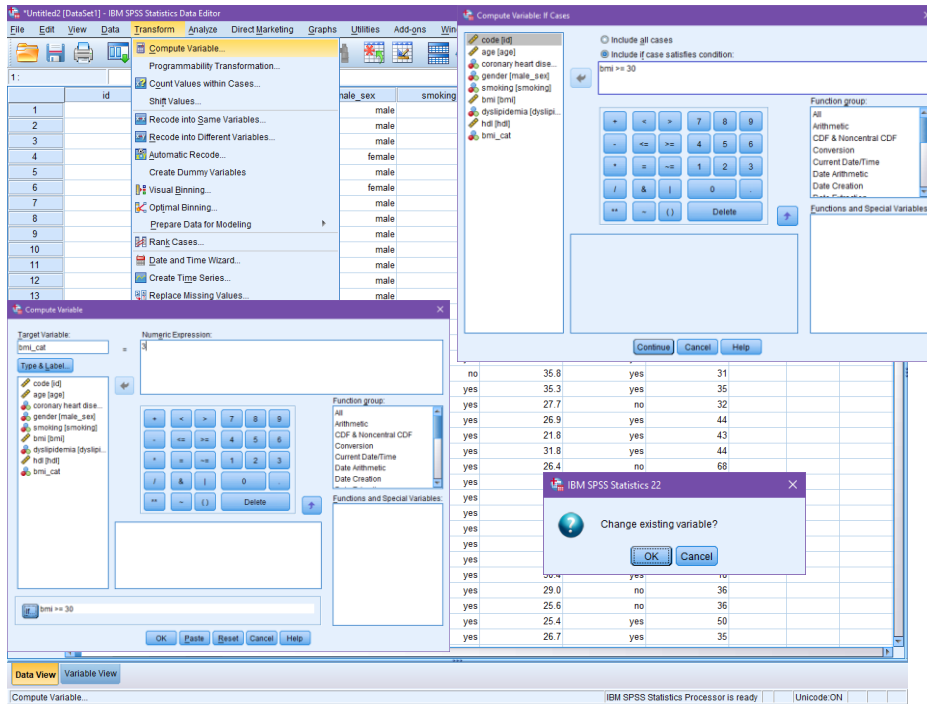
21



22



23



24

Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor

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

Visible: 9 of 9 Variables

	id	age	CHD_group	male_sex	smoking	bmi	dyslipidemia	hdl	bmi_cat	var	var
1	1	29	yes	male	yes	28.2	yes	41.0	2.00		
2	2	35	yes	male	yes	36.4	yes	53.0	3.00		
3	3	33	yes	male	yes	30.0	yes	31.0	3.00		
4	4	35	yes	female	yes	20.7	yes	45.0	1.00		
5	5	34	yes	male	yes	23.0	yes	38.0	1.00		
6	6	25	yes	female	yes	25.4	yes	45.0	2.00		
7	7	33	yes	male	yes	28.6	no	58.0	2.00		
8	8	36	yes	male	yes	30.0	yes	51.0	3.00		
9	9	32	yes	male	yes	33.3	yes	38.0	3.00		
10	10	30	yes	male	yes	28.2	yes	27.0	2.00		
11	11	35	yes	male	yes	22.1	no	41.0	1.00		
12	12	34	yes	male	yes	37.0	no	29.0	3.00		
13	13	33	yes	male	yes	30.1	no	44.0	3.00		
14	14	30	yes	male	no	39.7	no	45.0	3.00		
15	15	34	yes	male	yes	32.7	yes	27.0	3.00		
16	16	35	yes	male	yes	24.7	yes	37.0	1.00		
17	17	35	yes	male	yes	33.0	yes	32.0	3.00		
18	18	33	yes	male	no	35.8	yes	31.0	3.00		
19	19	33	yes	male	yes	35.3	yes	35.0	3.00		
20	20	35	yes	male	yes	27.7	no	32.0	2.00		
21	21	33	yes	male	yes	25.9	yes	44.0	2.00		
22	22	35	yes	male	yes	21.8	yes	43.0	1.00		
23	23	33	yes	male	yes	31.8	yes	44.0	3.00		
24	24	29	yes	female	yes	26.4	no	68.0	2.00		
25	25	28	yes	male	yes	21.7	no	38.0	1.00		
26	26	30	yes	male	yes	34.7	yes	28.0	3.00		
27	27	32	yes	female	yes	19.7	no	43.0	1.00		
28	28	32	yes	female	yes	29.7	yes	35.0	2.00		
29	29	33	yes	male	yes	23.7	no	43.0	1.00		
30	30	33	yes	male	yes	25.6	no	66.0	2.00		
31	31	36	yes	male	yes	30.4	yes	18.0	3.00		
32	32	34	yes	male	yes	29.0	no	36.0	2.00		
33	33	27	yes	male	yes	25.6	no	36.0	2.00		
34	34	26	yes	male	yes	25.4	yes	50.0	2.00		
35	35	35	yes	male	yes	26.7	yes	35.0	2.00		

Data View Variable View

IBM SPSS Statistics Processor is ready | Unicode ON

25

Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor

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

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	id	Numeric	12	0	code	None	None	12	Right	Scale	Input
2	age	Numeric	12	0	age	None	None	12	Right	Scale	Input
3	CHD_group	Numeric	12	0	coronary heart ...	{0, No}...	None	12	Right	Nominal	Input
4	male_sex	Numeric	12	0	gender	{0, female}...	None	12	Right	Nominal	Input
5	smoking	Numeric	12	0	smoking	{0, no}...	None	12	Right	Nominal	Input
6	bmi	Numeric	12	1	bmi	None	None	12	Right	Scale	Input
7	dyslipidemia	Numeric	12	0	dyslipidemia	{0, no}...	None	12	Right	Nominal	Input
8	hdl	Numeric	12	0	hdl	None	None	12	Right	Scale	Input
9	bmi_cat	Numeric	8	0	bmi categorical	None	None	10	Right	Nominal	Input

Value Labels

Value: 3

Label: Obesity

0 = "Underweight"
1 = "Normal weight"
2 = "Pre-obesity"

OK Cancel Help

Data View Variable View

IBM SPSS Statistics Processor is ready | Unicode ON

26

	id	age	CHD_group	male_sex	smoking	bmi	dyslipidemia	hdi	bmi_cat	var	var
1	1	29	yes	male	yes	28.2	yes	41.0	Pre-obesity		
2	2	35	yes	male	yes	36.4	yes	53.0	Obesity		
3	3	33	yes	male	yes	30.0	yes	31.0	Obesity		
4	4	35	yes	female	yes	20.7	yes	45.0	Normal weight		
5	5	34	yes	male	yes	23.0	yes	38.0	Normal weight		
6	6	25	yes	female	yes	25.4	yes	45.0	Pre-obesity		
7	7	33	yes	male	yes	28.6	no	58.0	Pre-obesity		
8	8	36	yes	male	yes	30.0	yes	51.0	Obesity		
9	9	32	yes	male	yes	33.3	yes	38.0	Obesity		
10	10	30	yes	male	yes	28.2	yes	27.0	Pre-obesity		
11	11	35	yes	male	yes	22.1	no	41.0	Normal weight		
12	12	34	yes	male	yes	37.0	no	29.0	Obesity		
13	13	33	yes	male	yes	30.1	no	44.0	Obesity		
14	14	30	yes	male	no	39.7	no	45.0	Obesity		
15	15	34	yes	male	yes	32.7	yes	27.0	Obesity		
16	16	35	yes	male	yes	24.7	yes	37.0	Normal weight		
17	17	35	yes	male	yes	33.0	yes	32.0	Obesity		
18	18	33	yes	male	no	35.8	yes	31.0	Obesity		
19	19	33	yes	male	yes	35.3	yes	35.0	Obesity		
20	20	35	yes	male	yes	27.7	no	32.0	Pre-obesity		
21	21	33	yes	male	yes	25.9	yes	44.0	Pre-obesity		
22	22	35	yes	male	yes	21.8	yes	43.0	Normal weight		
23	23	33	yes	male	yes	31.8	yes	44.0	Obesity		
24	24	29	yes	female	yes	26.4	no	68.0	Pre-obesity		
25	25	28	yes	male	yes	21.7	no	38.0	Normal weight		
26	26	30	yes	male	yes	34.7	yes	28.0	Obesity		
27	27	32	yes	female	yes	19.7	no	43.0	Normal weight		
28	28	32	yes	female	yes	29.7	yes	35.0	Pre-obesity		
29	29	33	yes	male	yes	23.7	no	43.0	Normal weight		
30	30	33	yes	male	yes	25.6	no	66.0	Pre-obesity		
31	31	36	yes	male	yes	30.4	yes	18.0	Obesity		
32	32	34	yes	male	yes	29.0	no	36.0	Pre-obesity		
33	33	27	yes	male	yes	25.6	no	36.0	Pre-obesity		
34	34	26	yes	male	yes	25.4	yes	50.0	Pre-obesity		
35	35	35	yes	male	yes	26.7	yes	35.0	Pre-obesity		

27

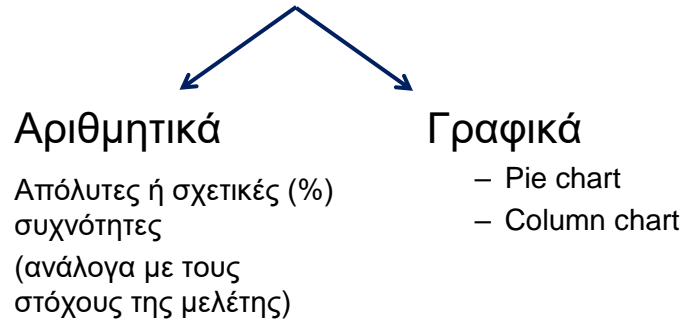
Κλάδοι στατιστικής

- Η **περιγραφική στατιστική** ανάλυση (descriptive statistics) ασχολείται με τη συνοπτική και περιγραφική παρουσίαση των δεδομένων μίας στατιστικής έρευνας.
- Η ανάλυση και εξαγωγή συμπερασμάτων ονομάζεται **επαγωγική στατιστική** ή **στατιστική συμπερασματολογία** (statistical inference, inferential statistics).

28

Περιγραφική στατιστική (I)

Μέθοδοι περιγραφής **ποιοτικών** δεδομένων



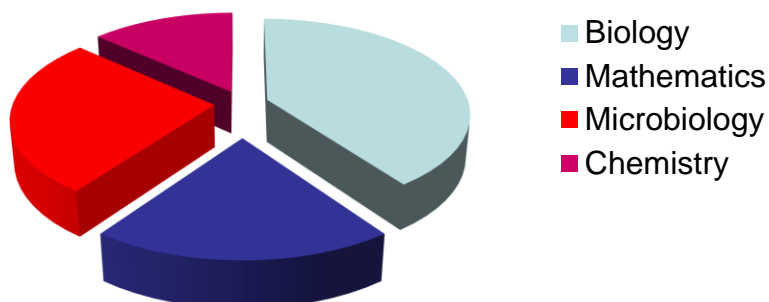
29

Παράδειγμα

All drug users (<i>n</i> = 544)	
Gender, <i>n</i> (%)	
Male	498 (91.5)
Female	46 (8.5)
Age (years)	
mean(SD)	36 (9)
Nationality, <i>n</i> (%)	
Greek	489 (90.1)
Other	54 (9.9)
Education, <i>n</i> (%)	
Never attended school	16 (2.9)
1–6 years	192 (35.4)
7–12 years	276 (50.8)
> 12 years	59 (10.9)
Employment status, <i>n</i> (%)	
Full-time work	336 (62.2)
Temporary work	136 (25.2)
Unemployed	54 (10.0)
Other	14 (2.6)

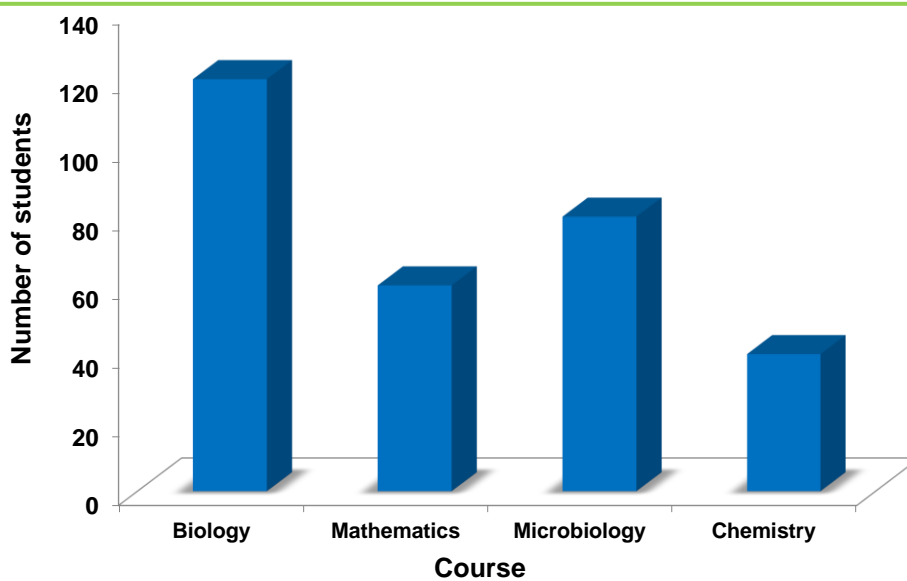
30

Γραφική παρουσίαση ποιοτικών δεδομένων - Διάγραμμα πίτας



31

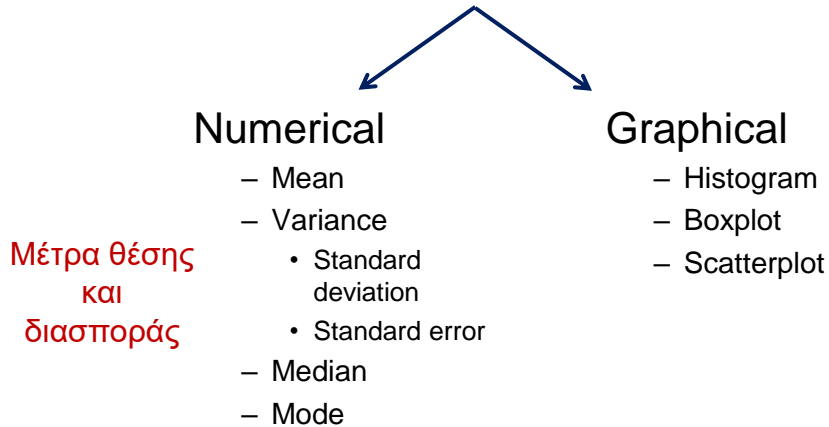
Ραβδόγραμμα



32

Περιγραφική στατιστική (II)

Μέθοδοι περιγραφής **ποσοτικών** δεδομένων



33

Μέτρα θέσης

- Τα πιο συνηθισμένα μέτρα που χρησιμοποιούνται για την περιγραφή της θέσης ενός συνόλου δεδομένων πάνω στον οριζόντιο άξονα x , είναι:
- **Αριθμητικός μέσος** ή **μέση τιμή** (*arithmetic mean* ή *average*)
- **Διάμεση τιμή** (*median*)
- **Κορυφή** ή **επικρατούσα τιμή** (*mode*)

34

$$3, 7, 10, 8, 31, 10, 2$$

$$\text{Mean (avg)} = \frac{3 + 7 + 10 + 8 + 31 + 10 + 2}{7} = \frac{71}{7}$$

↓
10.14

$$\text{Median} = 2, 3, 7, 8, 10, 10, 31$$

↓
8

↑
middle

$$\text{Mode} = 3, 7, 10, 8, 31, 10, 2$$

↓
10

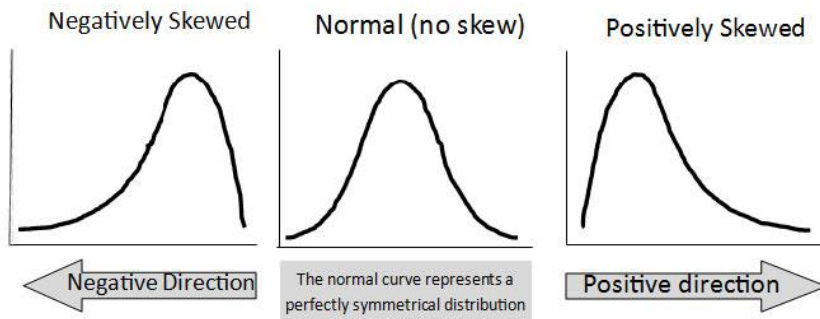
35

Επιλογή μέτρων θέσης

- Εξαρτάται από το σχήμα της κατανομής:
 - Εάν μία κατανομή είναι συμμετρική, μπορούμε να χρησιμοποιήσουμε τη μέση τιμή.
 - Εάν είναι ασύμμετρη προς τα αριστερά ή προς τα δεξιά επιλέγουμε τη διάμεση τιμή.

36

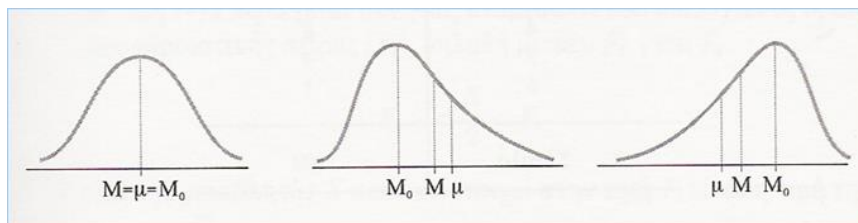
Παραδείγματα κατανομών (I)



37

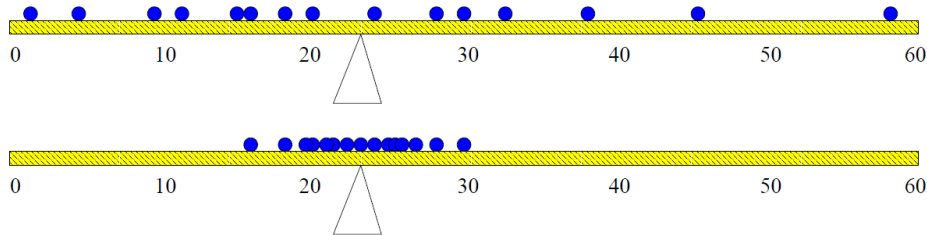
Παραδείγματα κατανομών (II)

M = διάμεση τιμή
 μ = μέση τιμή
 M_o = επικρατούσα τιμή



38

Μέτρα διασποράς



39

Διακύμανση (s^2)

- Για να υπολογίσουμε τη διασπορά (**διακύμανση, variance**) των παρατηρήσεων t_1, t_2, \dots, t_n μίας μεταβλητής x , θα αφαιρέσουμε τη μέση τιμή της x από κάθε παρατήρηση και θα βρούμε τον αριθμητικό μέσο των διαφορών αυτών:

$$s^2 = \frac{1}{v} \sum_{i=1}^v (t_i - \bar{x})^2$$

- Για πίνακα συχνοτήτων: $s^2 = \frac{1}{v} \sum_{i=1}^k (x_i - \bar{x})^2 v_i$

όπου x_1, x_2, \dots, x_k οι τιμές της μεταβλητής (ή τα κέντρα των κλάσεων) με αντίστοιχες συχνότητες v_1, v_2, \dots, v_k

40

Τυπική Απόκλιση (s) (I)

- Η διακύμανση δεν εκφράζεται με τις μονάδες με τις οποίες εκφράζονται οι παρατηρήσεις.
- Αν εκτιμήσουμε την τετραγωνική ρίζα της διακύμανσης, θα έχουμε ένα μέτρο διασποράς που θα εκφράζεται με την ίδια μονάδα μέτρησης του χαρακτηριστικού.
- Η ποσότητα αυτή λέγεται **τυπική απόκλιση (*standard deviation*)**, συμβολίζεται με **s** και είναι ένα μέτρο της διασποράς των δεδομένων γύρω από τη μέση τιμή.

$$s = \sqrt{s^2} \qquad s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

41

Τυπική Απόκλιση (s) (II)

Sample

Data (X_i): 10 12 14 15 17 18 18 24

n = 8 Mean = \bar{X} = 16

$$s = \sqrt{\frac{(10 - \bar{X})^2 + (12 - \bar{X})^2 + (14 - \bar{X})^2 + \dots + (24 - \bar{X})^2}{n - 1}}$$

$$= \sqrt{\frac{(10 - 16)^2 + (12 - 16)^2 + (14 - 16)^2 + \dots + (24 - 16)^2}{7}}$$

$$= \sqrt{\frac{130}{7}} = \boxed{4.308} \rightarrow \text{A measure of the "average" scatter around the mean}$$

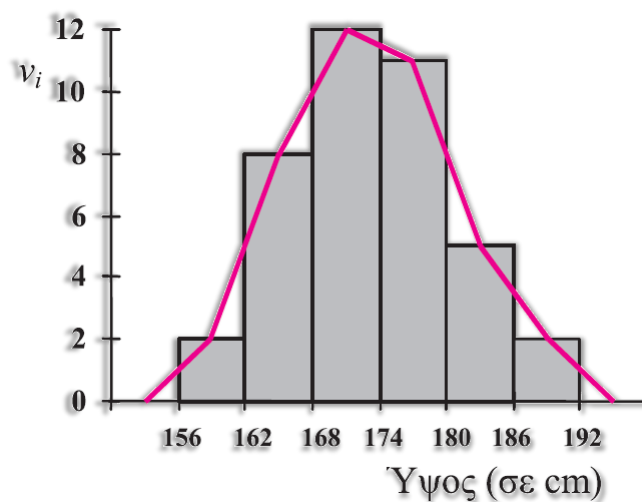
42

Γραφική παρουσίαση ποσοτικών δεδομένων

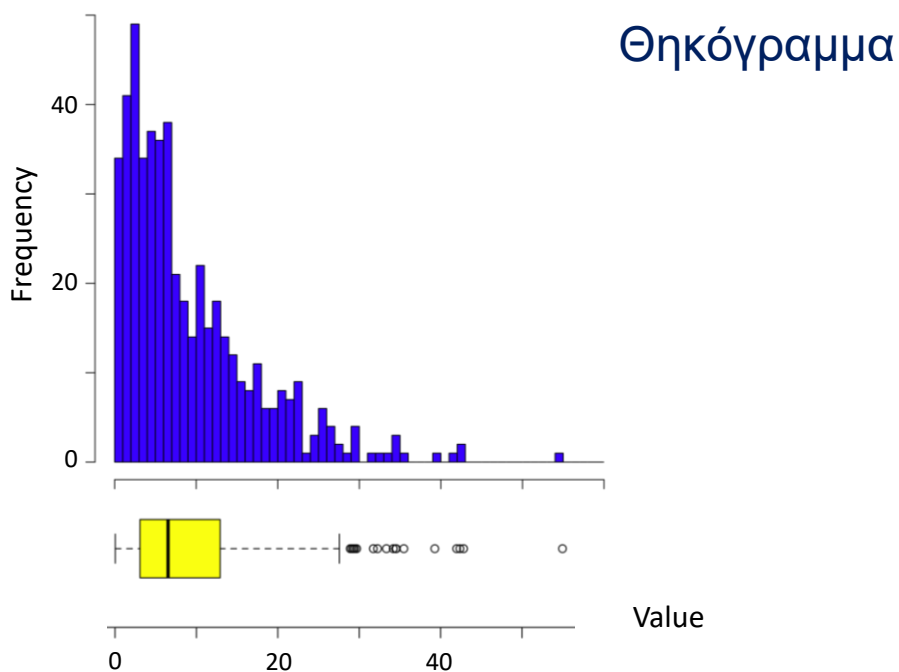
- Για τα **συνεχή ποσοτικά δεδομένα** η γραφική απεικόνιση μίας κατανομής συχνοτήτων συνεχών δεδομένων μπορεί να γίνει με:
 - Ιστόγραμμα συχνοτήτων
 - Καμπύλη συχνοτήτων
 - Θηκόγραμμα

43

Ιστόγραμμα συχνοτήτων



44



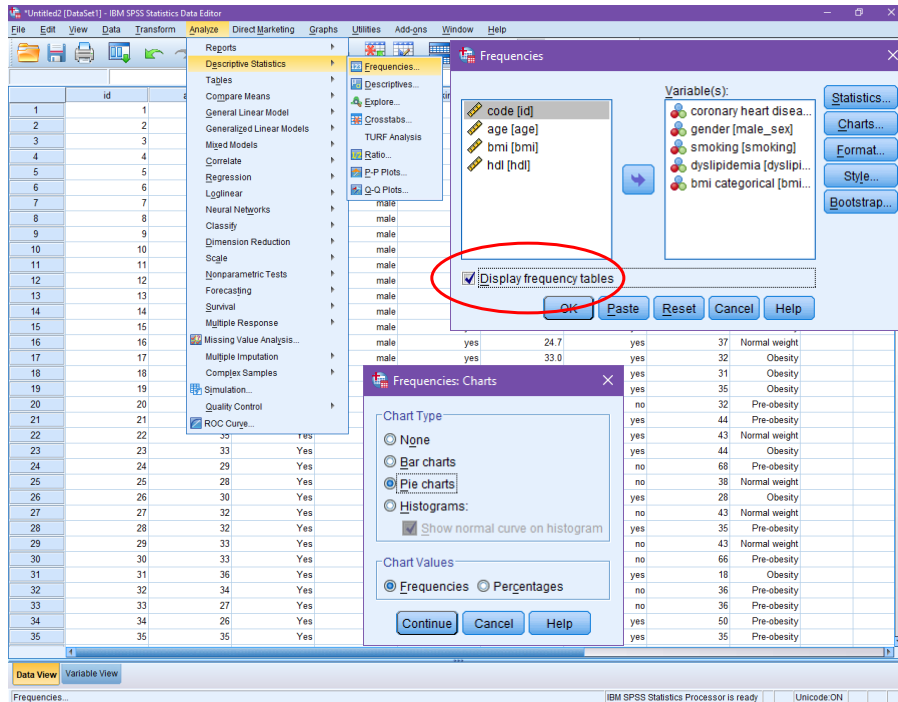
45

4^ο Βήμα: Στατιστικές αναλύσεις (Analyze)

Παράθεση δεδομένων (περιγραφική στατιστική):

- Μέσω του μενού Analyze ⇨ Descriptive Statistics ⇨
 - ✓ Descriptives, αντλούμε βασικές πληροφορίες για τη μέγιστη, την ελάχιστη, τη μέση τιμή και τη τυπική απόκλιση των μεταβλητών
 - ✓ Frequencies για συχνότητες ή περιγραφή κατανομής

46



47

Frequency Table

gender

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid female	26	13.0	13.0	13.0
male	174	87.0	87.0	100.0
Total	200	100.0	100.0	

smoking

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid no	50	25.0	25.0	25.0
yes	150	75.0	75.0	100.0
Total	200	100.0	100.0	

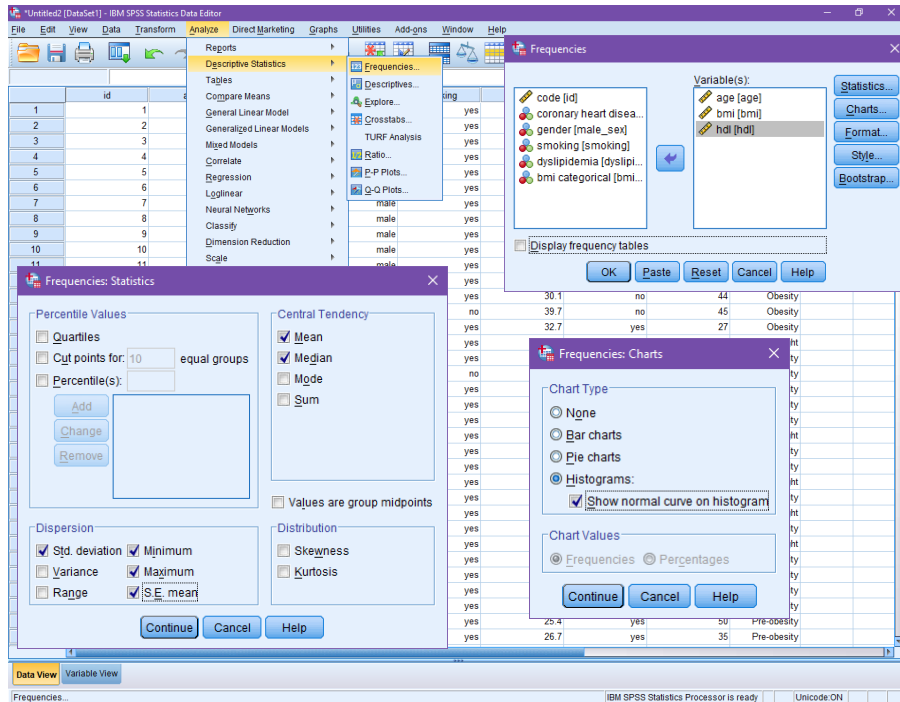
dyslipidemia

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid no	134	67.0	67.0	67.0
yes	66	33.0	33.0	100.0
Total	200	100.0	100.0	

bmi categorical

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Normal weight	85	42.5	42.5	42.5
Pre-obesity	61	30.5	30.5	73.0
Obesity	54	27.0	27.0	100.0
Total	200	100.0	100.0	

48



49

Frequencies

Statistics

		age	bmi	hdi
N	Valid	200	200	198
	Missing	0	0	2
Mean		32.15	27.174	45.121
Std. Error of Mean		.229	.3361	1.0358
Median		33.00	26.475	43.000
Std. Deviation		3.239	4.7531	14.5754
Minimum		23	19.6	18.0
Maximum		36	50.5	98.0

50

Δημιουργία / μεταβολή σχημάτων (Graph)

- Bar chart
 - Pie Chart
 - Histogram
 - ...
- Τα σχήματα εμφανίζονται στον Output Viewer όπου και μπορούν να τροποποιηθούν μέσω του Chart Editor (παράθυρο που ανοίγει με διπλό κλικ πάνω στο γράφημα)

51

Στατιστική συμπερασματολογία

Έλεγχος υποθέσεων:

- Θα πρέπει να σχετίζεται με μία επιστημονική υπόθεση.
- Διατυπώνονται **σε ζεύγη**:
 - Μηδενική υπόθεση (H_0):
Η υπόθεση που θέλουμε να απορρίψουμε
 - Εναλλακτική Υπόθεση (H_A)
Η υπόθεση που επιθυμούμε να ελέγξουμε την ισχύ της (η ενδιαφέρουσα υπόθεση)
- Οι στατιστικές δοκιμασίες επιχειρούν (μαθηματικά) να απορρίψουν τη μηδενική υπόθεση.

52

Οι τιμές p εκτιμούν σημαντικότητα

Η τιμή p μίας στατιστικής δοκιμασίας είναι η πιθανότητα να παρατηρήσουμε τις τιμές του δείγματός μας υποθέτοντας ότι η H_0 είναι αληθής.

- Αν p είναι πολύ μικρή, **δεν** είναι πολύ πιθανό η H_0 να είναι ορθή.
- Ποιες τιμές p είναι αποδεκτές;
 - Εξαρτάται από το υπό μελέτη ερώτημα
 - 0.05 αποτελεί το **σύνηθες όριο**
 - Αν $p < 0,05$, τότε η πιθανότητα να παρατηρήσουμε το δείγμα μας εάν η μηδενική υπόθεση ήταν αληθής είναι μικρότερη από 5%

53

Στατιστικές μέθοδοι

Μεταβλητή	Μεταβλητή	
	Ποσοτική	Ποιοτική
Ποιοτική	• t-test	• Chi-square test
Ποσοτική		• t-test

Σημείωση: Ποσοτικές μεταβλητές που ακολουθούν κανονική κατανομή

54

Παραδείγματα

- Υπάρχει διαφορά της αρτηριακής πίεσης μεταξύ των δύο φύλων;
 - Αρτηριακή πίεση: **ποσοτική** μεταβλητή
 - Φύλο: **ποιοτική** μεταβλητή
 - t-test

- Υπάρχει διαφορά της καπνιστικής συνήθειας μεταξύ των δύο φύλων;
 - Κάπνισμα: **ποιοτική** μεταβλητή
 - Φύλο: **ποιοτική** μεταβλητή
 - Chi-square test

55

The screenshot displays the IBM SPSS Statistics interface. The main window shows a data table with columns 'id' and 'smoking'. The 'Crosstabs' dialog box is open, with 'code [id]' selected as the Row(s) variable and 'smoking [smoking]' as the Column(s) variable. The 'Crosstabs: Cell Display' dialog box is also open, showing 'Observed' and 'Expected' counts selected, and 'Compare column proportions' checked. The 'Crosstabs: Statistics' dialog box is open, showing 'Chi-square' selected. The 'Crosstabs: Cell Display' dialog box also shows 'Row' and 'Column' percentages selected, and 'Total' percentages selected. The 'Crosstabs: Statistics' dialog box also shows 'Nominal' and 'Ordinal' statistics selected.

56

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
coronary heart disease * smoking	200	100.0%	0	0.0%	200	100.0%

Crosstabs

coronary heart disease * smoking Crosstabulation

			smoking		Total
			no	yes	
coronary heart disease	no	Count	46	54	100
		Expected Count	25.0	75.0	100.0
		% within coronary heart disease	46.0%	54.0%	100.0%
		% within smoking	92.0%	36.0%	50.0%
yes	yes	Count	4	96	100
		Expected Count	25.0	75.0	100.0
		% within coronary heart disease	4.0%	96.0%	100.0%
		% within smoking	8.0%	64.0%	50.0%
Total		Count	50	150	200
		Expected Count	50.0	150.0	200.0
		% within coronary heart disease	25.0%	75.0%	100.0%
		% within smoking	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	47.040 ^a	1	.000		
Continuity Correction ^b	44.827	1	.000		
Likelihood Ratio	53.356	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	46.805	1	.000		
N of Valid Cases	200				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 25.00.

57

Είδη t-test

Έλεγχος ενός δείγματος:

- Συγκρίνουμε τη μέση τιμή ενός δείγματος με μία συγκεκριμένη τιμή

Πολλαπλά δείγματα:

Εξαρτημένα δείγματα (με αντιστοιχία κατά ζεύγη):

- Συγκρίνουμε τη μέση τιμή δύο κατανομών στις οποίες οι παρατηρήσεις π.χ. αφορούν τον ίδιο πληθυσμό

Ανεξάρτητα (μη σχετιζόμενα) δείγματα:

- Συγκρίνουμε τη μέση τιμή δύο ανεξάρτητων κατανομών

58

t-test

- Παραδοχές: Οι παρατηρήσεις ακολουθούν κανονική κατανομή.

59

The screenshot shows the IBM SPSS Statistics Data Editor interface. The main window displays a dataset with the following variables: code [id], age [age], gender [male_sex], smoking [smoking], dyslipidemia [dyslipi...], hdl [hdl], and bmi_cat. The 'Explore' dialog box is open, showing the dependent list as 'bmi (bmi)' and the factor list as 'coronary heart disea...'. The 'Explore: Plots' dialog box is also open, showing options for boxplots and descriptive statistics.

Explore Dialog Box:

- Dependent List: bmi (bmi)
- Factor List: coronary heart disea...
- Display: Both (Statistics, Plots)

Explore: Plots Dialog Box:

- Boxplots:
 - Factor levels together
 - Dependents together
 - None
- Descriptive:
 - Stem-and-leaf
 - Histogram
- Normality plots with tests:
 - Spread vs Level with Levene Test:
 - None
 - Power estimation
 - Transformed Power: Natural log
 - Untransformed

60

Explore

coronary heart disease

Case Processing Summary						
coronary heart disease	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
bmi no	100	100.0%	0	0.0%	100	100.0%
yes	100	100.0%	0	0.0%	100	100.0%

coronary heart disease			Statistic	Std. Error
bmi no	Mean		25.582	.3797
	95% Confidence Interval for Mean	Lower Bound	24.898	
		Upper Bound	26.335	
	5% Trimmed Mean		25.273	
	Median		23.000	
	Variance		14.417	
	Std. Deviation		3.7970	
	Minimum		20.7	
	Maximum		37.9	
	Range		17.2	
	Interquartile Range		5.5	
	Skewness		1.219	.241
	Kurtosis		.427	.478
	yes	Mean		28.766
95% Confidence Interval for Mean		Lower Bound	27.766	
		Upper Bound	29.775	
5% Trimmed Mean			28.568	
Median			28.600	
Variance			25.876	
Std. Deviation			5.0868	
Minimum			19.6	
Maximum			50.5	
Range			30.9	
Interquartile Range			6.8	
Skewness			.838	.241
Kurtosis			2.276	.478

Tests of Normality

coronary heart disease	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
bmi no	.302	100	.000	.780	100	.000
yes	.060	100	.200	.960	100	.004

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

61

62

T-Test

coronary heart disease	N	Mean	Std. Deviation	Std. Error Mean
bmi no	100	25.582	3.7970	.3797
bmi yes	100	28.766	5.0868	.5087

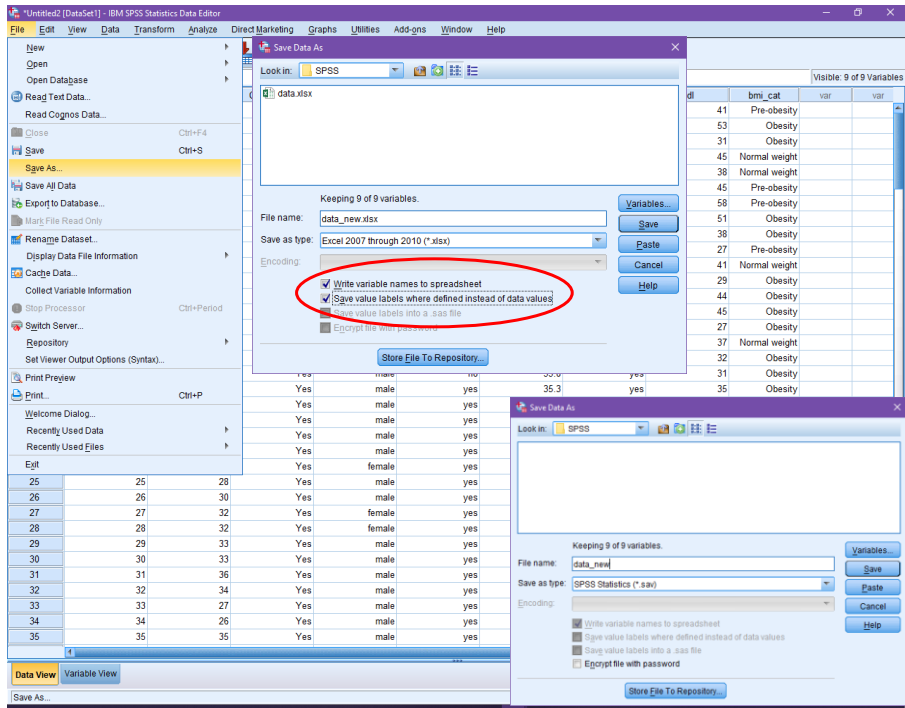
		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
										Lower	Upper
bmi	Equal variances assumed	3.429	.066	-5.016	198	.000	-3.1839	.6348	-4.4357	-1.9321	
	Equal variances not assumed			-5.016	183.185	.000	-3.1839	.6348	-4.4363	-1.9315	

63

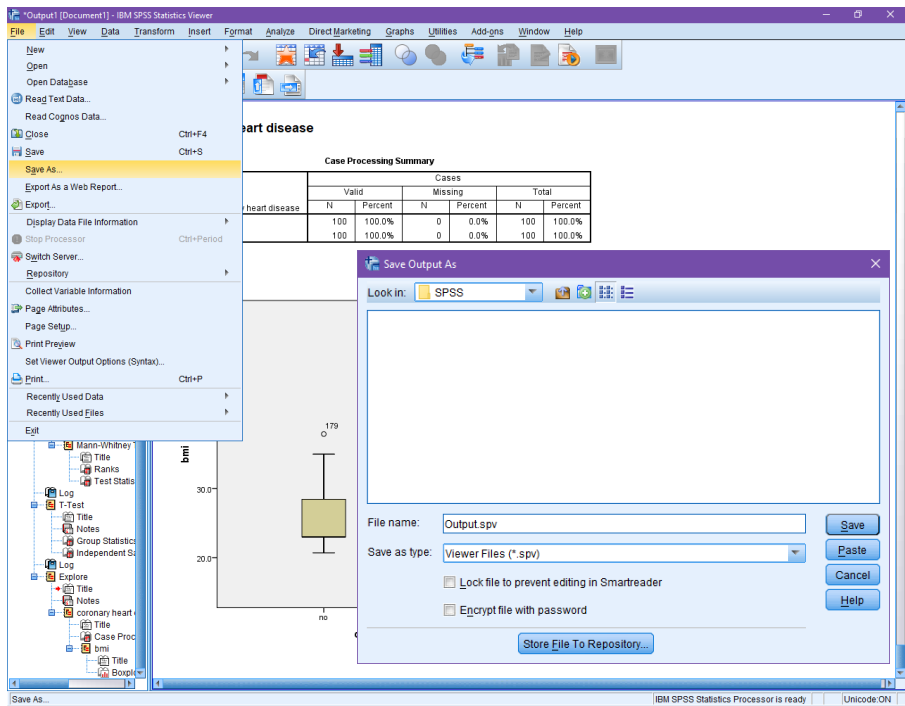
5^ο Βήμα: Αποθήκευση / εκτύπωση / έξοδος

- Αποθήκευση μέσω του μενού File ⇒ Save ή Save As, και σώζουμε το αρχείο με κατάληξη **.sav**
Προσοχή με το Save!
- Εκτύπωση μέσω του μενού File ⇒ Print (αφορά κυρίως τον Output Viewer από όπου μπορούμε να εκτυπώσουμε επιλεκτικά)
- Έξοδος μέσω του μενού File ⇒ Exit ή πατώντας το «x» στο παράθυρο πάνω δεξιά

64



65



66