ΣΤΑΤΙΣΤΙΚΗ (ΜΥ0202)

Μ.Ν. Ντυκέν,
Πανεπιστήμιο Θεσσαλίας
Τ.Μ.Χ.Π.Π.Α.

Βόλος, 2018-2019

ΔΙΑΛΕΞΗ 06

Περιεχόμενο της Διάλεξης
Διερευνητική ανάλυση: μέτρα θέσης για Ομαδοποιημένα Δεδομένα

1.β. Διερευνητική Ανάλυση

Ομαδοποιημένα Δεδομένα / Δεδομένα κατά τάξεις

1β. ΜΕΤΡΑ ΘΕΣΗΣ – ΟΜΑΔΟΠΟΙΗΜΕΝΑ ΔΕΔΟΜΕΝΑ

2.1. Μέσος όρος – αριθμητικός μέσος: μ (πληθυσμός) $\bar{X}$ (δείγμα)

Τα δεδομένα είναι ομαδοποιημένα σε **k** τάξεις ή **k** ομάδες.
Αρκετοί οργανισμοί και στατιστικές υπηρεσίες προσφέρουν δευτερογενή δεδομένα υπό αυτή τη μορφή.

Έχουμε τα ακόλουθα:

- $\Rightarrow$ **n** άτομα
- $\Rightarrow$ **k** τάξεις / ομάδες
- $\Rightarrow$ Κάθε τάξη περιλαμβάνει $\mathbf{n_i}$ άτομα, $\mathbf{i = 1,\ldots, k}$ και $f_i = \frac{n_i}{n}$

$$\boxed{\sum_{i=1}^{k} n_i = n \ \& \ \sum_{i}^{k} f_i = 1}$$

Ο μέσος όρος υπολογίζεται ως έξης:

$$\boxed{\overline{X} = \frac{1}{n}\sum_{i=1}^{k} n_i X_i = \sum_{i=1}^{k} f_i X_i}$$

όπου $\mathbf{X_i}$ = **κέντρο της τάξης i, όταν η ομαδοποίηση αφορά τάξεις μεγέθους**

1β. ΜΕΤΡΑ ΘΈΣΗΣ – ΟΜΑΔΟΠΟΙΗΜΕΝΑ ΔΕΔΟΜΕΝΑ

2 εναλλακτικές μορφές ομαδοποιημένων δεδομένων

Μηνιαίο Εισόδημα Νοικοκυριών

i	Εισόδημα (σε 1000 €)	Αριθμός νοικοκυριών
1	[0-2)	15
2	[2-4)	10
3	[4-6)	5
4	[6-8)	2
	Σύνολο	32

Δεδομένα <u>κατά τάξεις εισοδήματος</u>

➔**Xi = κέντρο της κάθε τάξης**

Μέγεθος Νοικοκυριού (μελή ανά νοικοκυριό)

Μέγεθος Νοικοκυριού	Αριθμός νοικοκυριών
1	10
2	18
3	22
4	20
5	14
6	6
Σύνολο	90

Δεδομένα <u>ομαδοποιημένα με βάση τον αριθμό μελών</u>

➔**Xi = 1, 2, …., 6**

1β. ΜΕΤΡΑ ΘΈΣΗΣ – ΟΜΑΔΟΠΟΙΗΜΕΝΑ ΔΕΔΟΜΕΝΑ

i	Εισόδημα (σε 1000 €)	Αριθμός νοικοκυριών	κέντρο της τάξης (Xi)	ni Xi
1	[0-2)	15	1	15
2	[2-4)	10	3	30
3	[4-6)	5	5	25
4	[6-8)	2	7	14
	Σύνολο	32		84

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} n_i X_i = \frac{1}{32} \times 84 = 2{,}6$$

Κατά μέσο όρο, το μέσο εισόδημα ανέρχεται σε 2.600€

Μέγεθος Νοικοκυριού (Xi)	Αριθμός νοικοκυριών (ni)	ni Xi	fi	fi Xi
1	10	10	0,111	0,111
2	18	36	0,200	0,400
3	22	66	0,244	0,733
4	20	80	0,222	0,889
5	14	70	0,156	0,778
6	6	36	0,067	0,400
Σύνολο	90	298	1,000	3,311

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} n_i X_i = \frac{1}{90} \times 298 = 3{,}3$$

$$\overline{X} = \sum_{i=1}^{k} f_i X_i = 3{,}3$$

Κατά μέσο όρο, το μέγεθος των νοικοκυριών ξεπερνά τα 3 άτομα ανά νοικοκυριό

1β. ΜΕΤΡΑ ΘΈΣΗΣ – ΟΜΑΔΟΠΟΙΗΜΕΝΑ ΔΕΔΟΜΕΝΑ

2.2. Διάμεσος: M_d ή Q_2

Εφόσον έχουμε **ομαδοποιημένα** δεδομένα, πρέπει καταρχήν να βρούμε την **τάξη που περιλαμβάνει τη διάμεσο**: αυτό απαιτεί τον ***υπολογισμό των αθροιστικών συχνοτήτων***.

Έστω **m** = αριθμός της τάξης που περιλαμβάνει τη διάμεσο όπου $1 \leq m \leq k$

Η διάμεσος Q_2 υπολογίζεται ως έξης:

$$Q_2 = L_m + \frac{w}{n_m}(0,5.n - N_{m-1})$$

ή

$$Q_2 = L_m + \frac{w}{f_m}(0,5 - F_{m-1})$$

L_m = αριστερό άκρο της **τάξης αναφοράς m** (τάξη μέσα στην οποία βρίσκεται ο δείκτης Q_p)

n = πλήθος ατόμων, μέγεθος δείγματος

w = πλάτος του διαστήματος της τάξης m

N_{m-1} = **απόλυτη** ***αθροιστική*** συχνότητα του διαστήματος που προηγείται του διαστήματος αναφοράς

n_m = **απόλυτη** ***απλή*** συχνότητα του διαστήματος αναφοράς

F_{m-1} = **σχετική** ***αθροιστική*** συχνότητα του διαστήματος που προηγείται του διαστήματος αναφοράς

f_m =**σχετική** ***απλή*** συχνότητα του διαστήματος αναφοράς Ε

1β. ΜΕΤΡΑ ΘΈΣΗΣ – ΟΜΑΔΟΠΟΙΗΜΕΝΑ ΔΕΔΟΜΕΝΑ

2.3. Τεταρτημόρια – Δεκατημόρια: Q_p - D_p

Τα Τεταρτημόρια $\mathbf{Q_p}$ δεν είναι τίποτα άλλο από $\mathbf{Q_1}$ (25%), $\mathbf{Q_2 = M_d}$ (50%) και $\mathbf{Q_3}$ (75%).

Ακολουθούν πάντα την διάταξη: $X_{min} \leq Q_1 \leq Q_2 \leq Q_3 \leq X_{max}$

Το 2ο Τεταρτημόριο (Q_2) συμπίπτει με τη Διάμεσο (M_d).

Τα Δεκατημόρια $\mathbf{D_p}$ χωρίζουν τα διατεταγμένα δεδομένα (αύξουσα σειρά) σε δέκατα (ανά 10%).

Υπάρχουν 9 δεκατημόρια: $\mathbf{D_1}$(10%), $\mathbf{D_2}$ (20%), …, $\mathbf{D_8}$ (80%), $\mathbf{D_9}$ (90%).

Ακολουθούν πάντα την διάταξη: $X_{min} \leq D_1 \leq D_2 \leq \ldots \leq D_5 \leq \ldots \leq D_9 \leq X_{max}$

Το 5ο Δεκατημόριο (D_5) συμπίπτει με τη Διάμεσο ($M_d = Q_2$).

1β. ΜΕΤΡΑ ΘΈΣΗΣ – ΟΜΑΔΟΠΟΙΗΜΕΝΑ ΔΕΔΟΜΕΝΑ

2.3. Τεταρτημόρια – Δεκατημόρια: Q_p - D_p

Έστω **m** = αριθμός της τάξης που περιλαμβάνει το Τεταρτημόριο ή το Δεκατημόριο, ανάλογα με αυτό που αναζητούμε.

Το Τεταρτημόριο **Q_p** υπολογίζεται ως έξης:

$$Q_p \text{ ή } D_p = L_m + \frac{w}{n_m}(p.n - N_{m-1})$$

ή

$$Q_p \text{ ή } D_p = L_m + \frac{w}{f_m}(p - F_{m-1})$$

L_m = αριστερό άκρο της **τάξης αναφοράς m** (τάξη μέσα στην οποία βρίσκεται ο δείκτης Q_p)

n = πλήθος ατόμων, μέγεθος δείγματος

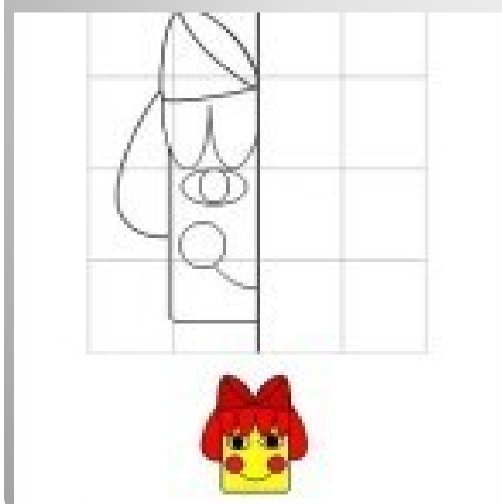
w = πλάτος διαστήματος της τάξης m

p = 0,25 (Q_1), 0,75 (Q_3), 0,1 ($_{D1}$), 0,2 (D_2),…, 0,9 (D_9)

N_{m-1} = **απόλυτη** ***αθροιστική*** συχνότητα του διαστήματος που προηγείται του διαστήματος αναφοράς

F_{m-1} = **σχετική** ***αθροιστική*** συχνότητα του διαστήματος που προηγείται του διαστήματος αναφοράς

ΥΠΟΛΟΓΙΣΜΟΣ ΜΕΤΡΩΝ ΘΕΣΗΣ & ΠΑΡΑΓΩΓΗ ΘΗΚΟΓΡΑΜΜΑΤΟΣ



ΜΗΚΟΣ ΑΚΤΩΝ 90 ΝΗΣΙΩΝ ΤΗΣ ΕΛΛΑΔΑΣ [01]

Ο πίνακας που ακολουθεί, μας δίνει την κατανομή των 90 μεγαλύτερων νησιών της Ελλάδας (εκτός Κρήτης) με βάση το μήκος των ακτών τους.

i	Μήκος ακτών (σε χλμ)	x_i (κέντρο τάξης)	Αριθμός νησιών n_i	N_i
1	[10 – 30)	20	19	19
2	[30 – 50)	40	27	46
3	[50 – 100)	75	18	64
4	[100 – 150)	125	16	80
5	[150 – 200)	175	4	84
6	[200 – 420)	310	6	**90**
ΣΥΝΟΛΟ			90	

1. *Να βρείτε (α) το μέσο μήκος ακτών και (β) τη διάμεσο.*
2. *Να σχεδιάσετε το Θηκόγραμμα.*
3. *Ποια τα συμπεράσματα σας;*

ΜΗΚΟΣ ΑΚΤΩΝ 90 ΝΗΣΙΩΝ ΤΗΣ ΕΛΛΑΔΑΣ [02]

(1.a) Το μέσο μήκος ακτών των 90 νησιών ανέρχεται σε : 7370 / 90 = ***81,9 χλμ***

i	Μήκος ακτών (σε χλμ)	X_i (κέντρο τάξης)	Αριθμός νησιών n_i	ni Xi	Ni
1	[10 – 30)	20	19	380	19
2	[30 – 50)	40	27	1080	46
3	[50 – 100)	75	18	1350	64
4	[100 – 150)	125	16	2000	80
5	[150 – 200)	175	4	700	84
6	[200 – 420)	310	6	1860	**90**
ΣΥΝΟΛΟ			**90**	**7370**	

Θέση της Q2:

$$m = 2$$

$$Q_2 = 30 + \frac{20}{27}(45 - 19)$$

$$Q_2 = 49,3$$

(1.b) Για να βρούμε τη διάμεσο, υπολογίζουμε το n/2 = 45 → η διάμεσος βρίσκεται στην 2η τάξη [30-50) δεδομένου ότι μέχρι την 1η τάξη, έχουμε 19 νησιά ενώ μέχρι και την 2η τάξη, έχουμε 46 νησιά > 45.

ΜΗΚΟΣ ΑΚΤΩΝ 90 ΝΗΣΙΩΝ ΤΗΣ ΕΛΛΑΔΑΣ [03]

i	Μήκος ακτών (σε χλμ)	X_i (κέντρο τάξης)	Αριθμός νησιών n_i	ni Xi	Ni
1	[10 – 30)	20	19	380	19
2	[30 – 50)	40	27	1080	46
3	[50 – 100)	75	18	1350	64
4	[100 – 150)	125	16	2000	80
5	[150 – 200)	175	4	700	84
6	[200 – 420)	310	6	1860	**90**
ΣΥΝΟΛΟ			**90**	**7370**	

<u>1ο Τεταρτημόριο</u>: Θέση του Q_1: $\mathbf{p \times n = 0,25 \times 90 = 22,5}$ ➔ Q_1 βρίσκεται μεταξύ των 22ου και 23ου νησιών, δηλαδή στην 2η τάξη (m=2):
L = 30, w = 20, n_2 = 27, N_1 = 19

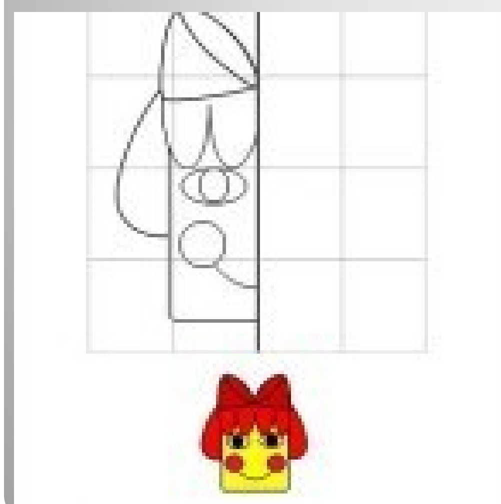
$$\boldsymbol{Q_1 = 30 + \frac{20}{27}(22,5 - 19) = 30 + 2,59 = 32,59}$$

<u>3ο Τεταρτημόριο</u>: Θέση του Q_3: $\mathbf{p \times n = 0,75 \times 90 = 67,5}$ ➔ Q_3 βρίσκεται μεταξύ των 67ου και 68ου νησιών, δηλαδή στην 4η τάξη (m=4):
L = 100, w = 50, n_4 = 16, N_3 = 64

$$\boldsymbol{Q_3 = 100 + \frac{50}{16}(67,5 - 64) = 100 + 10,94 = 110,94}$$

ΠΑΡΑΡΤΗΜΑ:

ΠΑΡΑΔΕΙΓΜΑΤΑ ΜΕ ΛΥΣΕΙΣ



ΠΑΡΑΔΕΙΓΜΑ 1

ΠΥΚΝΟΤΗΤΑ ΠΛΗΘΥΣΜΟΥ ΤΩΝ 13 ΠΕΡΙΦΕΡΕΙΩΝ ΤΗΣ ΕΛΛΑΔΑΣ [01]

Πυκνότητα πληθυσμού ανά Περιφέρεια, 2001-2011

$$\text{Πυκνότητα} = \frac{\text{Πληθυσμός}}{\text{Επιφάνεια}} = \frac{P}{S}$$

ΠΕΡΙΦΕΡΕΙΕΣ	Επιφάνεια (τ.χμ.)	Πληθυσμός		Πυκνότητα	
		2001	2011	2001	2011
	S	P01	P11	D01	D11
ΑΝΑΤ ΜΑΚΕΔΟΝΙΑ & ΘΡΑΚΗ	14.063,1	607.162	608.182	43,2	43,2
ΚΕΝΤΡΙΚΗ ΜΑΚΕΔΟΝΙΑ	18.885,0	1.876.558	1.882.108	99,4	99,7
ΔΥΤΙΚΗ ΜΑΚΕΔΟΝΙΑ	9.258,0	294.317	283.689	31,8	30,6
ΗΠΕΙΡΟΣ	9.084,1	336.392	336.856	37,0	37,1
ΘΕΣΣΑΛΙΑ	14.003,5	740.115	732.762	52,9	52,3
ΣΤΕΡΕΑ ΕΛΛΑΣ	15.435,0	558.144	547.390	36,2	35,5
ΙΟΝΙΟΙ ΝΗΣΟΙ	2.299,1	209.608	207.855	91,2	90,4
ΔΥΤΙΚΗ ΕΛΛΑΔΑ	11.067,0	721.541	679.796	65,2	61,4
ΠΕΛΟΠΟΝΝΗΣΟΣ	15.475,0	597.622	577.903	38,6	37,3
ΑΤΤΙΚΗ	3.805,6	3.894.573	3.828.434	1023,4	1006,0
ΒΟΡΕΙΟ ΑΙΓΑΙΟ	3.823,0	205.235	199.231	53,7	52,1
ΝΟΤΙΟΥ ΑΙΓΑΙΟΥ	5.286,0	298.462	309.015	56,5	58,5
ΚΡΗΤΗ	8.335,3	594.368	623.065	71,3	74,8
ΕΛΛΑΔΑ	**130.819,7**	**10.934.097**	**10.816.286**	**83,6**	**82,7**

$$\frac{608.162}{14063{,}1} = 43{,}2$$

$$\frac{3.828.434}{3805{,}6} = 1006{,}0$$

Η πυκνότητα πληθυσμού δίνει τον ***αριθμό μονίμων κατοίκων ανά km²***. Ο συγκεκριμένος δείκτης αντανακλά το βαθμό συγκέντρωσης του πληθυσμού και αποτελεί καλή ένδειξη του βαθμού αστικοποίησης των περιοχών.

Πηγή: Ελ.Στατ, Απόγραφες πληθυσμού

ΠΥΚΝΟΤΗΤΑ ΠΛΗΘΥΣΜΟΥ ΤΩΝ 13 ΠΕΡΙΦΕΡΕΙΩΝ ΤΗΣ ΕΛΛΑΔΑΣ [02]

- Εξετάζοντας τον πίνακα, παρατηρούμε ότι, έχουμε δεδομένα για την μεταβλητή «πυκνότητα» για δυο διαφορετικά έτη. Κατά συνέπεια μπορούμε να εξετάζουμε την μεταβλητή σε δύο διαφορετικές περιόδους (πρακτικά σημαίνει ότι, έχουμε 2 μεταβλητές).
- Από την ανάλυση του πίνακα, είναι φανερό ότι, η αναζήτηση της επικρατούσας τιμής δεν έχει νόημα: για κάθε έτος ξεχωριστά, όλες οι Περιφέρειες έχουν διαφορετικές τιμές μεταξύ τους.
- Η ανάλυση των μέτρων κεντρικής τάσης περιλαμβάνει επομένως:

  (i) τον υπολογισμό της μέσης τιμής για τις 13 περιφέρειες

  (ii) τον υπολογισμό της διάμεσου

  (iii) τον υπολογισμό των Τεταρτημόριων

  (iv) την παραγωγή του θηκογράμματος

Για τους υπολογισμούς, θα πρέπει τα δεδομένα να είναι ταξινομημένα με αύξουσα σειρά.

ΠΥΚΝΟΤΗΤΑ ΠΛΗΘΥΣΜΟΥ ΤΩΝ 13 ΠΕΡΙΦΕΡΕΙΩΝ ΤΗΣ ΕΛΛΑΔΑΣ [03]

Εφόσον τα δεδομένα είναι ατομικά (n=13), ο μέσος όρος δίνεται από τον ακόλουθο τύπο:

$$\boxed{\mu = \frac{\sum_{i=1}^{n} X_i}{n}}$$

i	ΠΕΡΙΦΕΡΕΙΕΣ	Πυκνότητα	
		2001	2011
1	ΑΝΑΤ ΜΑΚΕΔΟΝΙΑ & ΘΡΑΚΗ	43,2	43,2
2	ΚΕΝΤΡΙΚΗ ΜΑΚΕΔΟΝΙΑ	99,4	99,7
3	ΔΥΤΙΚΗ ΜΑΚΕΔΟΝΙΑ	31,8	30,6
4	ΗΠΕΙΡΟΣ	37,0	37,1
5	ΘΕΣΣΑΛΙΑ	52,9	52,3
6	ΣΤΕΡΕΑ ΕΛΛΑΣ	36,2	35,5
7	ΙΟΝΙΟΙ ΝΗΣΟΙ	91,2	90,4
8	ΔΥΤΙΚΗ ΕΛΛΑΔΑ	65,2	61,4
9	ΠΕΛΟΠΟΝΝΗΣΟΣ	38,6	37,3
10	ΑΤΤΙΚΗ	1023,4	1006,0
11	ΒΟΡΕΙΟ ΑΙΓΑΙΟ	53,7	52,1
12	ΝΟΤΙΟΥ ΑΙΓΑΙΟΥ	56,5	58,5
13	ΚΡΗΤΗ	71,3	74,8
	ΣΥΝΟΛΟ	**1700,4**	**1678,9**

Από τον πίνακα, έχουμε:

2001: $\sum_{i=1}^{13} X_i = 1700{,}4$

2002: $\sum_{i=1}^{13} X_i = 1678{,}9$

2001: $\mu = \frac{1700{,}4}{13} = 130{,}8$

2011: $\mu = \frac{1678{,}9}{13} = 129{,}1$

Ο αριθμητικός μέσος του δείκτη είναι πολύ διαφορετικός από την πραγματική πυκνότητα της χώρας!

Η πραγματική πυκνότητα σε επίπεδο Ελλάδας = 83,6 (2001) & 82,7 (2011)

ΠΥΚΝΟΤΗΤΑ ΠΛΗΘΥΣΜΟΥ ΤΩΝ 13 ΠΕΡΙΦΕΡΕΙΩΝ ΤΗΣ ΕΛΛΑΔΑΣ [03]

Υπολογισμός της διάμεσου: *n*=*13 (μόνος)*

i	ΠΕΡΙΦΕΡΕΙΕΣ	Πυκνότητα 2001
1	ΔΥΤΙΚΗ ΜΑΚΕΔΟΝΙΑ	31,8
2	ΣΤΕΡΕΑ ΕΛΛΑΣ	36,2
3	ΗΠΕΙΡΟΣ	37,0
4	ΠΕΛΟΠΟΝΝΗΣΟΣ	38,6
5	ΑΝΑΤ ΜΑΚΕΔΟΝΙΑ & ΘΡΑΚΗ	43,2
6	ΘΕΣΣΑΛΙΑ	52,9
7	**ΒΟΡΕΙΟ ΑΙΓΑΙΟ**	**53,7**
8	ΝΟΤΙΟΥ ΑΙΓΑΙΟΥ	56,5
9	ΔΥΤΙΚΗ ΕΛΛΑΔΑ	65,2
10	ΚΡΗΤΗ	71,3
11	ΙΟΝΙΟΙ ΝΗΣΟΙ	91,2
12	ΚΕΝΤΡΙΚΗ ΜΑΚΕΔΟΝΙΑ	99,4
13	ΑΤΤΙΚΗ	1023,4
	ΕΛΛΑΔΑ	**83,6**

i	ΠΕΡΙΦΕΡΕΙΕΣ	Πυκνότητα 2011
1	ΔΥΤΙΚΗ ΜΑΚΕΔΟΝΙΑ	30,6
2	ΣΤΕΡΕΑ ΕΛΛΑΣ	35,5
3	ΗΠΕΙΡΟΣ	37,1
4	ΠΕΛΟΠΟΝΝΗΣΟΣ	37,3
5	ΑΝΑΤ ΜΑΚΕΔΟΝΙΑ & ΘΡΑΚΗ	43,2
6	ΒΟΡΕΙΟ ΑΙΓΑΙΟ	52,1
7	**ΘΕΣΣΑΛΙΑ**	**52,3**
8	ΝΟΤΙΟΥ ΑΙΓΑΙΟΥ	58,5
9	ΔΥΤΙΚΗ ΕΛΛΑΔΑ	61,4
10	ΚΡΗΤΗ	74,8
11	ΙΟΝΙΟΙ ΝΗΣΟΙ	90,4
12	ΚΕΝΤΡΙΚΗ ΜΑΚΕΔΟΝΙΑ	99,7
13	ΑΤΤΙΚΗ	1006,0
	ΕΛΛΑΔΑ	**82,7**

Προσοχή: η κατάταξη άλλαξε μεταξύ Θεσσαλίας και Β. Αιγαίου

Τυπολόγιο:
n=13 ➔ η θέση δίνεται από (n+1)/2 = 7 ➔ **Md = 53,7 (2001) & Md = 52,3 (2011)**

ΠΥΚΝΟΤΗΤΑ ΠΛΗΘΥΣΜΟΥ ΤΩΝ 13 ΠΕΡΙΦΕΡΕΙΩΝ ΤΗΣ ΕΛΛΑΔΑΣ [03]

Υπολογισμός των Τεταρτημόριων: *n=13 (μόνος),*

i	ΠΕΡΙΦΕΡΕΙΕΣ	Πυκνότητα 2001	i	ΠΕΡΙΦΕΡΕΙΕΣ	Πυκνότητα 2011
1	ΔΥΤΙΚΗ ΜΑΚΕΔΟΝΙΑ	31,8	1	ΔΥΤΙΚΗ ΜΑΚΕΔΟΝΙΑ	30,6
2	ΣΤΕΡΕΑ ΕΛΛΑΣ	36,2	2	ΣΤΕΡΕΑ ΕΛΛΑΣ	35,5
3	ΗΠΕΙΡΟΣ	37,0	3	ΗΠΕΙΡΟΣ	37,1
4	ΠΕΛΟΠΟΝΝΗΣΟΣ	38,6	4	ΠΕΛΟΠΟΝΝΗΣΟΣ	37,3
5	ΑΝΑΤ ΜΑΚΕΔΟΝΙΑ & ΘΡΑΚΗ	43,2	5	ΑΝΑΤ ΜΑΚΕΔΟΝΙΑ & ΘΡΑΚΗ	43,2
6	ΘΕΣΣΑΛΙΑ	52,9	6	ΒΟΡΕΙΟ ΑΙΓΑΙΟ	52,1
7	ΒΟΡΕΙΟ ΑΙΓΑΙΟ	53,7	7	ΘΕΣΣΑΛΙΑ	52,3
8	ΝΟΤΙΟΥ ΑΙΓΑΙΟΥ	56,5	8	ΝΟΤΙΟΥ ΑΙΓΑΙΟΥ	58,5
9	ΔΥΤΙΚΗ ΕΛΛΑΔΑ	65,2	9	ΔΥΤΙΚΗ ΕΛΛΑΔΑ	61,4
10	ΚΡΗΤΗ	71,3	10	ΚΡΗΤΗ	74,8
11	ΙΟΝΙΟΙ ΝΗΣΟΙ	91,2	11	ΙΟΝΙΟΙ ΝΗΣΟΙ	90,4
12	ΚΕΝΤΡΙΚΗ ΜΑΚΕΔΟΝΙΑ	99,4	12	ΚΕΝΤΡΙΚΗ ΜΑΚΕΔΟΝΙΑ	99,7
13	ΑΤΤΙΚΗ	1023,4	13	ΑΤΤΙΚΗ	1006,0
	ΕΛΛΑΔΑ	**83,6**		**ΕΛΛΑΔΑ**	**82,7**

n=13

➔ $n' = (n+1)/2$

➔ $n' = 7$

➔ n' μόνος

Βλέπε 1η στήλη του πίνακα (Τυπολόγιο)

Θέση για Q_1: $(n'+1)/2 = 4$

Θέση για Q_3: $(3n'+1)/2 = 11$

Q_1: (θέση i = 4) ➔ $Q_1 = X_4 = 38{,}6$ (2001) & $Q_1 = X_4 = 37{,}3$ (2011)

Q_3: (θέση i = 11) ➔ $Q_3 = X_{11} = 91{,}2$ (2001) & $Q_3 = X_{11} = 90{,}7$ (2011)

ΠΥΚΝΟΤΗΤΑ ΠΛΗΘΥΣΜΟΥ ΤΩΝ 13 ΠΕΡΙΦΕΡΕΙΩΝ ΤΗΣ ΕΛΛΑΔΑΣ [03]

Θηκογράμματα

Μέτρα		2001	2011
Min	Ελάχιστη τιμή	31,8	30,6
Max	Μέγιστη τιμή	1023,4	1006,0
M_d	Διάμεσος	53,7	52,3
Q_1	1ο Τεταρτημόριο	38,6	37,3
Q_3	3ο Τεταρτημόριο	91,2	90,7
D_F	Ενδοτεταρτημοριακό διάστημα (Q3-Q1)	52,6	53,4
W1	Κάτω Εσωτερικό φράχτη	-40,3	-42,8
W3	Άνω Εσωτερικό φράχτη	170,1	170,8
WW1	Κάτω Εξωτερικό φράχτη	…	…
WW3	Άνω Εξωτερικό φράχτη	249,0	250,9

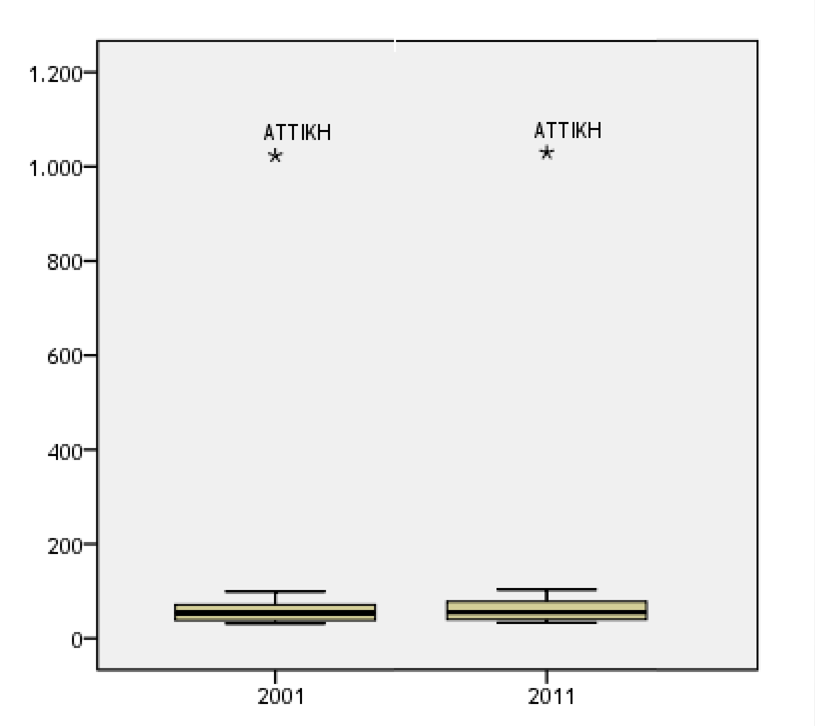
➔ W1 = 0

Η πυκνότητα δεν μπορεί να έχει αρνητικές τιμές

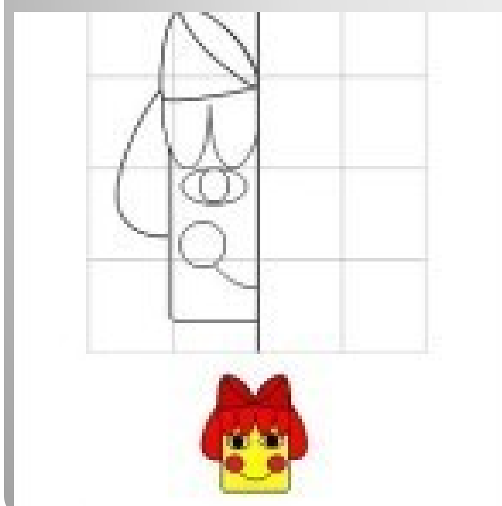
W3 < Max ➔ Πρέπει να υπολογίσουμε το WW3

ΠΥΚΝΟΤΗΤΑ ΠΛΗΘΥΣΜΟΥ ΤΩΝ 13 ΠΕΡΙΦΕΡΕΙΩΝ ΤΗΣ ΕΛΛΑΔΑΣ [03]

Θηκογράμματα



Η Αττική αποτελεί «άτυπη» περίπτωση, ακραία τιμή σε σχέση με τις άλλες 12 περιφέρειες, δεδομένου ότι, η πυκνότητα της Αττικής είναι μεγαλύτερη όχι μόνο από τον εσωτερικό φράχτη άλλα και από τον εξωτερικό.



ΠΑΡΑΔΕΙΓΜΑ

2

Διατακτική μεταβλητή με Likert scale

- Για την εκτίμηση της ποιότητας ενός προϊόντος, ρωτήσαμε σε 300 καταναλωτές να αξιολογήσουν την ποιότητα του προϊόντος σε μια κλίμακα από 1 έως 7, χρησιμοποιώντας την γνωστή Κλίμακα Αθροιστικής Βαθμολόγησης (**Likert scale**).

Ποιότητα προϊόντος	X_i	Αριθμός καταναλωτών (n_i)
Απαράδεκτη	1	12
Πολύ κακή	2	27
Κακή	3	48
Μέτρια	4	66
Καλή	5	84
Πολύ καλή	6	45
Τέλεια	7	18
Σύνολο		300

- Η μεταβλητή που εξετάζουμε είναι <u>ποιοτική</u> (Ποιότητα προϊόντος) όμως κωδικοποιήσαμε τις 7 κατηγορίες έτσι ώστε να υπάρχει μια <u>λογική ιεράρχηση</u> ως προς την ποιότητα. Πρόκειται για **διατακτική μεταβλητή** με 7 κατηγορίες.

Διατακτική μεταβλητή
με Likert scale

Εφόσον έχουμε **αρκετές κατηγορίες (τιμές) οι οποίες αντανακλούν ένα βαθμό προτιμήσεων,** μπορεί να γίνει **διευρυμένη στατιστική ανάλυση** η οποία περιλαμβάνει:

i. τον υπολογισμό των συχνοτήτων, βασική προϋπόθεση για την περαιτέρω ανάλυση,

ii. την αναζήτηση της επικρατούσας τιμής (κατηγορίας),

iii. τον υπολογισμό του μέσου όρου (μέση ικανοποίηση ως προς την ποιότητα σε μια κλίμακα από 1 έως 7)

iv. τον υπολογισμό της διάμεσου (Md = Q2),

v. Τον υπολογισμό του πρώτου τεταρτημόριου (Q1) και του τρίτου τεταρτημόριου (Q3).

vi. την γραφική αναπαράσταση: το Θηκόγραμμα (Box-Plot) για την αναζήτηση «**ακραίων τιμών**».

(i) ***Υπολογισμός σχετικών συχνοτήτων***
(ii) ***Επικρατούσα τιμή***

Ποιότητα προϊόντος	Xi	Αριθμός καταναλωτών: Απόλυτες συχνότητες		Σχετικές συχνότητες	
		n_i	N_i	f_i	F_i
Απαράδεκτη	1	12	12	4,0	4,0
Πολύ κακή	2	27	39	9,0	13,0
Κακή	3	48	87	16,0	29,0
Μέτρια	4	66	153	22,0	51,0
Καλή	5	84	237	28,0	79,0
Πολύ καλή	6	45	282	15,0	94,0
Τέλεια	7	18	**300**	6,0	**100,0**
Σύνολο		**300**		**100,0**	

Η <u>επικρατούσα τιμή</u> αντιστοιχεί στην κατηγορία με την μέγιστη συχνότητα, δηλαδή στον χαρακτηρισμό: **Καλή ποιότητα**.
Δεν προσφέρει σημαντική πληροφορία, ειδικά όταν αντιστοιχεί σε περιορισμένο %.

Διατακτική μεταβλητή με Likert scale

(iii) Υπολογισμός του μέσου όρου *(με ομαδοποιημένα δεδομένα)*

Ποιότητα προϊόντος	Xi	Αριθμός καταναλωτών: Απόλυτες συχνότητες	Σχετικές συχνότητες	Υπολογισμός	
		n_i	f_i	$n_i X_i$	$f_i X_i$
Απαράδεκτη	1	12	0,04	12	0,04
Πολύ κακή	2	27	0,09	54	0,18
Κακή	3	48	0,16	144	0,48
Μέτρια	4	66	0,22	264	0,88
Καλή	5	84	0,28	420	1,40
Πολύ καλή	6	45	0,15	270	0,90
Τέλεια	7	18	0,06	126	0,42
Σύνολο		**300**	**1,0**	1290	**4,30**

$$\overline{X} = \frac{\sum_{i=1}^{k} n_i X_i}{n} \text{ ή } \overline{X} = \sum_{i=1}^{k} f_i X_i$$

$$\sum_{i=1}^{7} n_i X_i = 1290$$

$$\Rightarrow$$

$$\overline{X} = 1290/300 = 4,3$$

k = 7
n = 300

Η μέση αξιολόγηση της ποιότητας του προϊόντος είναι λίγο μεγαλύτερη από το μέτριο επίπεδο.

(iv) Αναζήτηση της Διάμεσου

Ποιότητα προϊόντος	Xi	Αριθμός καταναλωτών: Απόλυτες συχνότητες		Σχετικές συχνότητες	
		n_i	N_i	f_i	F_i
Απαράδεκτη	1	12	12	4,0	4,0
Πολύ κακή	2	27	39	9,0	13,0
Κακή	3	48	87	16,0	29,0
Μέτρια	4	66	153	22,0	51,0
Καλή	5	84	237	28,0	79,0
Πολύ καλή	6	45	282	15,0	94,0
Τέλεια	7	18	**300**	6,0	**100,0**
Σύνολο		**300**		**100,0**	

29% δηλώσαν έως και Κακή (3) ενώ 51% δηλώσαν έως και Μέτρια (4)

n = 300 ➔ **η διάμεσος** βρίσκεται μεταξύ n/2 και (n+2)/1 : 150 και 151
Με βάση τα N_i, η διάμεσος βρίσκεται στην κατηγορία «Μέτρια» ➔ **$M_d = Q_2 = 4$**

(v) ***Αναζήτηση Τεταρτημόριων*** Q_1 ***και*** Q_2

- Η θέση τους εξαρτάται από το μέγεθος του δείγματος **n.**
- Το 1ο τεταρτημόριο Q_1 είναι εκείνη η τιμή της μεταβλητής για την οποία το **25%** των ατόμων δηλώνουν τιμή μικρότερη από το Q1 (επομένως το 75% των παρατηρήσεων βρίσκεται πάνω από Q1).
- Με παρόμοιο τρόπο, ορίζεται το 3ο τεταρτημόριο Q3. Πρόκειται για τη τιμή της μεταβλητής για την οποία το **75%** των ατόμων δηλώνουν τιμή μικρότερη από το Q3 (επομένως το 25% των παρατηρήσεων βρίσκεται πάνω από το Q3).

Διατακτική μεταβλητή με Likert scale

(v) Αναζήτηση Τεταρτημόριων Q_1 και Q_2

Ποιότητα προϊόντος	Xi	Αριθμός καταναλωτών: Απόλυτες συχνότητες		Σχετικές συχνότητες	
		n_i	N_i	f_i	F_i
Απαράδεκτη	1	12	12	4,0	4,0
Πολύ κακή	2	27	39	9,0	13,0
Κακή	3	48	87	16,0	29,0
Μέτρια	4	66	153	22,0	51,0
Καλή	5	84	237	28,0	79,0
Πολύ καλή	6	45	282	15,0	94,0
Τέλεια	7	18	**300**	6,0	**100,0**
Σύνολο		**300**		**100,0**	

n = 300 ➔
n′ = n/2= 150
n′ ζυγός

4η στήλη στον πίνακα

$$Q_1=\frac{1}{2}\left(X_{\frac{n'}{2}}+X_{\frac{n'}{2}+1}\right)=\frac{1}{2}\left(X_{\frac{150}{2}}+X_{\frac{150}{2}+1}\right)=\frac{1}{2}(X_{75}+X_{76})=3 \quad Q_3=\frac{1}{2}\left(X_{\frac{3n'}{2}}+X_{\frac{3n'}{2}+1}\right)=\frac{1}{2}(X_{225}+X_{226})=5$$

➔ $Q_1 = 3$ & $Q_3 = 5$

(vi) Δημιουργία Θηκόγραμμα – BOX-PLOT

Αποτελείται από:

(α) ένα ορθογώνιο κουτί η βάση του οποίου αντιστοιχεί στο 1ο τεταρτημόριο και η κορυφή στο 3ο τεταρτημόριο: το ύψος του κουτιού = d_F.

(β) μέσα στο κουτί, σημειώνεται μια κάθετος που αντιστοιχεί στη διάμεσο.

(γ) υπολογίζουμε τους εσωτερικούς και εξωτερικούς φράχτες.

Κάτω και Άνω Εσωτερικοί φράχτες:

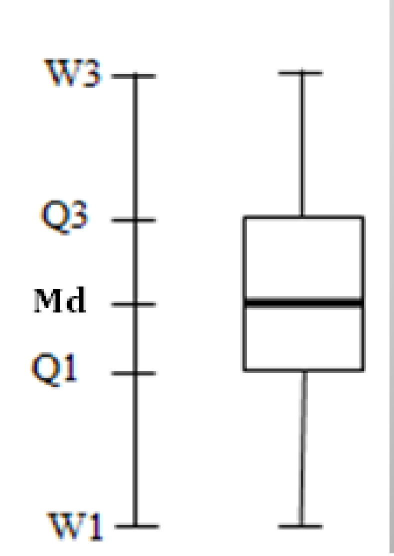
Κάτω: $W_1 = Q_1 - 1.5 \times (Q_3\text{-}Q_1) = Q_1 - 1.5 \times d_F$

Άνω : $W_3 = Q_3 + 1.5 \times (Q_3\text{-}Q_1) = Q_1 + 1.5 \times d_F$

Κάτω και Άνω Εξωτερικοί φράχτες:

Κάτω: $WW_1 = Q_1 - 3 \times (Q_3\text{-}Q_1) = Q_1 - 3 \times d_F$

Άνω : $WW_3 = Q_3 + 3 \times (Q_3\text{-}Q_1) = Q_1 + 3 \times d_F$



(vi) Δημιουργία Θηκόγραμμα : Ακραίες τιμές

Με βάση τις τιμές που βρήκαμε για τη διάμεσο και τα τεταρτημόρια, έχουμε:

$Q_1 = 3$, $Q_2 = 4$, $Q_3 = 5 \Rightarrow Q_3 - Q_1 = d_F = 2$

$W_1 = Q_1 - 1{,}5\,.[Q_3 - Q_1] = Q_1 - 1{,}5.d_F = 3 - 1{,}5 \times 2 = 0$

$W_3 = Q_3 + 1{,}5\,.[Q_3 - Q_1] = Q_3 + 1{,}5.d_F = 5 + 1{,}5 \times 2 = 8$

Δεδομένου ότι:

- Η ελάχιστη τιμή στις προτιμήσεις των καταναλωτών = 1 ➔ Min > W_1 (=0)
  ➔ δεν υπάρχει ακραία τιμή
- Η μέγιστη τιμή στις προτιμήσεις των καταναλωτών = 7 ➔ Max < W_3 (=8)
  ➔ δεν υπάρχει ακραία τιμή

Έχουμε μια **ομαλή κατανομή των προτιμήσεων**. Δεν μπορούμε να θεωρήσουμε ότι, ορισμένοι καταναλωτές έχουν «ακραία» γνώμη για την ποιότητα του προϊόντος.