

25/04/2020

Πιθανοθεωρία, – Στατιστική & Συνδυαστική Ανάλυση

7^η Διάλεξη

(Περιγραφική Στατιστική)

Πληθυσμός ή στατιστικός πληθυσμός

- Ονομάζεται η κατανομή των τιμών μιας τ.μ., δηλαδή η κατανομή των τιμών που παίρνει ένα χαρακτηριστικό μιας ομάδας υποκειμένων
- Κάθε υποκείμενο επί του οποίου μετράται/παρατηρείται η τιμή ενός κοινού χαρακτηριστικού ονομάζεται **απλό στοιχείο ή δειγματοληπτική / πειραματική μονάδα**

Τυχαίο δείγμα

- Τυχαίο δείγμα μεγέθους n από έναν πληθυσμό ονομάζουμε n τ.μ. X_1, X_2, \dots, X_n που παίρνουν τιμές από τον πληθυσμό αυτό (και έχουν επομένως την ίδια κατανομή)

Πραγματοποίηση τυχαίου δείγματος

- Οι συγκεκριμένες τιμές x_1, x_2, \dots, x_n που είναι διαθέσιμες για επεξεργασία μετά τη λήψη του δείγματος αποτελούν μια πραγματοποίηση των τ.μ. X_1, X_2, \dots, X_n και ονομάζονται **δεδομένα ή παρατηρήσεις**

Πίνακας κατανομής συχνοτήτων (1)

ο Ποσοτικές μεταβλητές

- Στην 1^η στήλη του Πίνακα καταγράφονται σε αύξουσα σειρά οι διαφορετικές τιμές y_1, y_2, \dots, y_k ($k \leq n$) από τις x_1, x_2, \dots, x_n που εμφανίστηκαν στο δείγμα
- Στις επόμενες στήλες για κάθε μεταβλητή y_i , $i = 1, 2, \dots, k$ καταγράφεται η εξής πληροφορία:

Πίνακας κατανομής συχνοτήτων (2)

ο Ποσοτικές μεταβλητές

1. Η **συχνότητα** της, n_i (πόσες φορές εμφανίστηκε)
2. Η **σχετική συχνότητα** της, $f_i = \frac{n_i}{n}$
3. Η **αθροιστική συχνότητα** της, N_i (το άθροισμα των συχνοτήτων των τιμών που είναι $\leq y_i$)
4. Η **αθροιστική σχετική συχνότητα** της, F_i (το άθροισμα των σχετικών συχνοτήτων των τιμών που είναι $\leq y_i$)

Πίνακας κατανομής συχνοτήτων (3)

y_i	n_i	f_i	N_i	F_i
5	67	0.5826	67	0.5826
6	26	0.2261	93	0.8087
7	12	0.1043	105	0.9130
8	5	0.0435	110	0.9565
9	3	0.0261	113	0.9826
10	2	0.0174	115	1.0000
Σύνολο	115	1.0000		

Πίνακας κατανομής συχνοτήτων (4)

ο Ποσοτικές μεταβλητές

- Αν έχει γίνει ομαδοποίηση των τιμών, στην 1^η στήλη αντί των διαφορετικών τιμών καταγράφονται οι διαφορετικές κλάσεις τιμών
- Στις επόμενες στήλες καταγράφονται η συχνότητα, η σχετική συχνότητα, η αθροιστική συχνότητα και η αθροιστική σχετική συχνότητα κάθε κλάσης τιμών
- Επιλογή κλάσεων: $k = 1 + 3.32 \cdot \log_{10} n$
- Εύρος κλάσεων: $r = \frac{R}{k} = \frac{x_{max} - x_{min}}{k}$

Πίνακας κατανομής συχνοτήτων (5)

Χρόνος επανεμφάνισης	n_i	f_i	N_i	F_i
[0.0, 1.5)	21	0.42	21	0.42
[1.5, 3.0)	13	0.26	34	0.68
[3.0, 4.5)	6	0.12	40	0.80
[4.5, 6.0)	4	0.08	44	0.88
[6.0, 7.5)	3	0.06	47	0.94
[7.5, 9.0)	2	0.04	49	0.98
[9.0, 10.5)	1	0.02	50	1.00
Σύνολα	50	1.00		

Πίνακας κατανομής συχνοτήτων (6)

ο Ποιοτικές μεταβλητές

- Στις ποιοτικές μεταβλητές **κατηγορίας** δεν έχει νόημα η διάταξη των διαφορετικών τιμών y_1, y_2, \dots, y_k και επομένως δεν έχουν νόημα ούτε οι αθροιστικές ούτε οι αθροιστικές σχετικές συχνότητες αλλά μόνο οι συχνότητες και οι σχετικές συχνότητες
- Στις ποιοτικές μεταβλητές **διάταξης** η διάταξη των διαφορετικών τιμών y_1, y_2, \dots, y_k έχει νόημα και επομένως έχουν νόημα τόσο οι συχνότητες και οι σχετικές συχνότητες όσο και οι αθροιστικές και οι αθροιστικές σχετικές συχνότητες

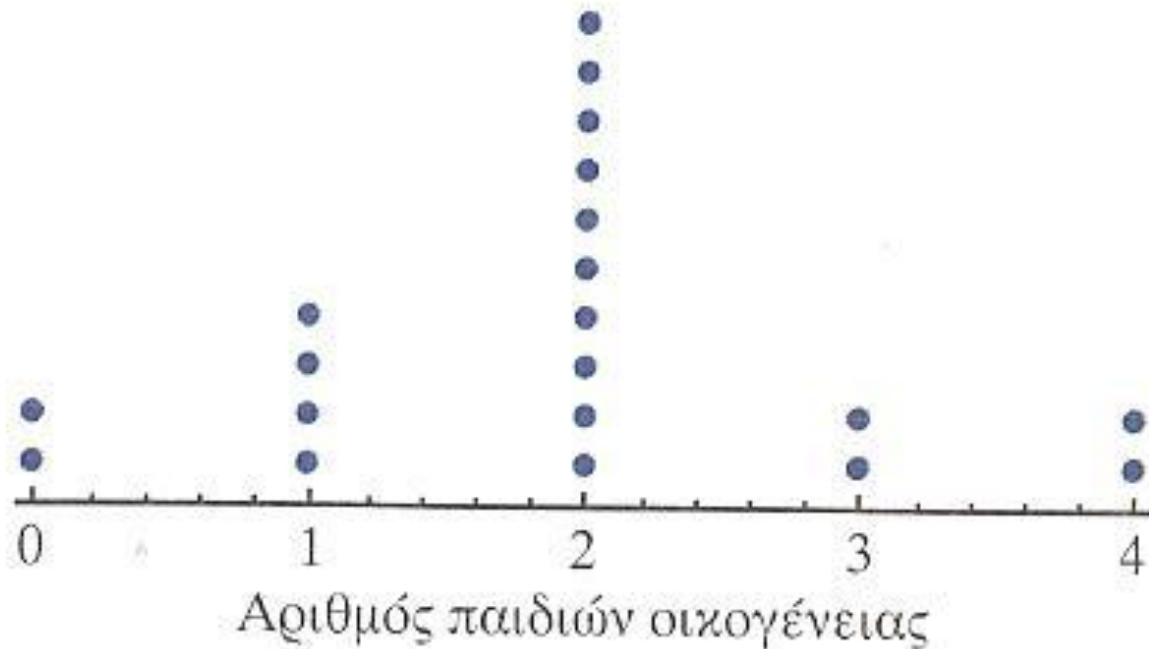
Γραφική παρουσίαση κατανομής συχνοτήτων (1)

ο Ποσοτικές μεταβλητές

1. Σημειόγραμμα
2. Ραβδόγραμμα συχνοτήτων και σχετικών συχνοτήτων
3. Κυκλικό διάγραμμα συχνοτήτων και σχετικών συχνοτήτων
4. Διάγραμμα συχνοτήτων και σχετικών συχνοτήτων
5. Ιστόγραμμα συχνοτήτων / σχετικών συχνοτήτων / αθροιστικών συχνοτήτων / αθροιστικών σχετικών συχνοτήτων
6. Πολύγωνο συχνοτήτων / σχετικών συχνοτήτων / αθροιστικών συχνοτήτων / αθροιστικών σχετικών συχνοτήτων
7. Φυλλογράφημα
8. Θηκόγραμμα

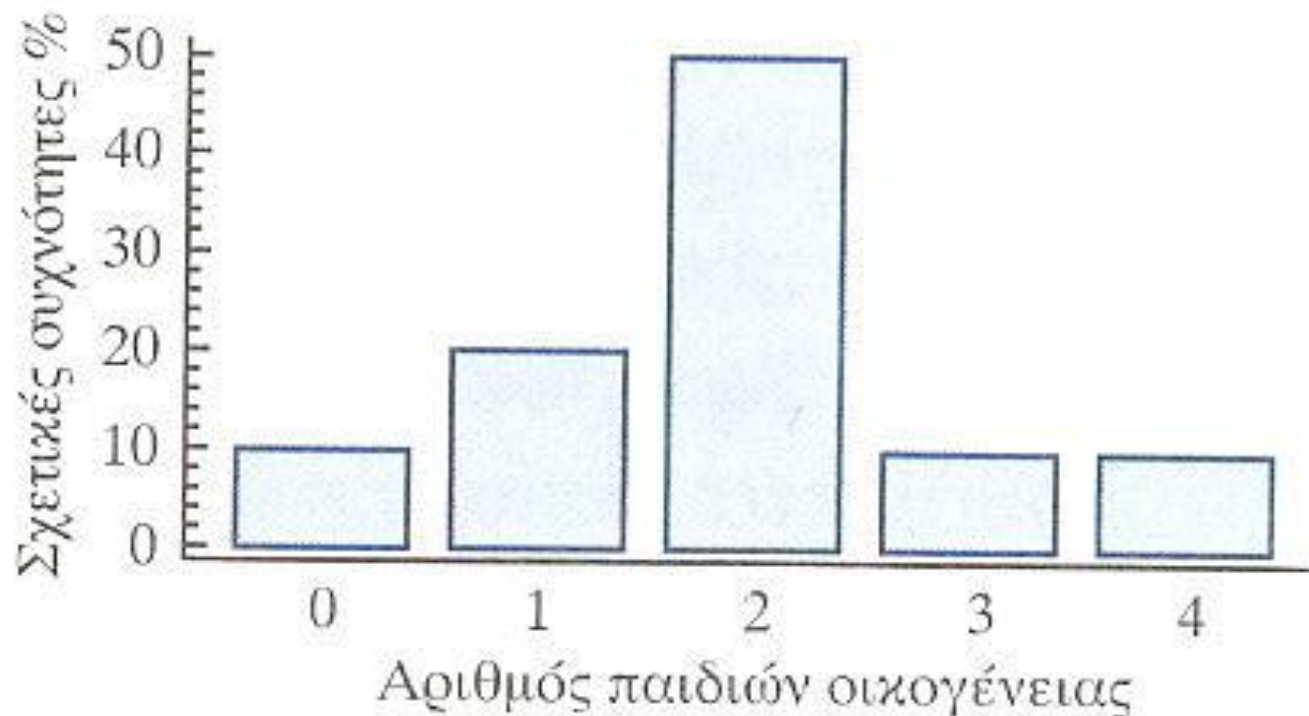
Γραφική παρουσίαση κατανομής
συχνοτήτων (1.1)

- Ποσοτικές μεταβλητές – Σημειόγραμμα
(dot diagram)



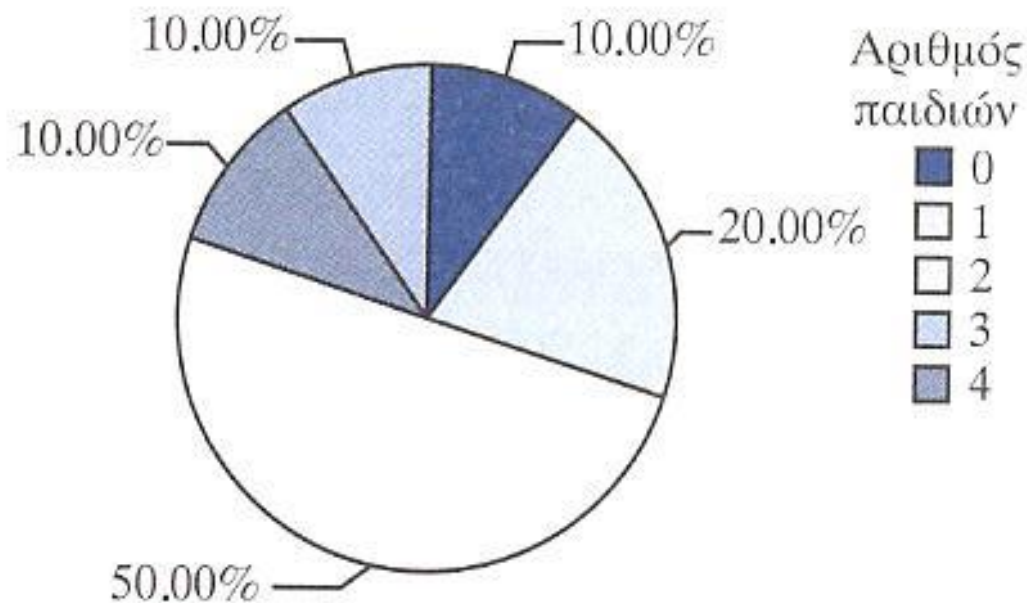
Γραφική παρουσίαση κατανομής
συχνοτήτων (1.2)

● Ποσοτικές μεταβλητές – Ραβδόγραμμα



Γραφική παρουσίαση κατανομής συχνοτήτων (1.3)

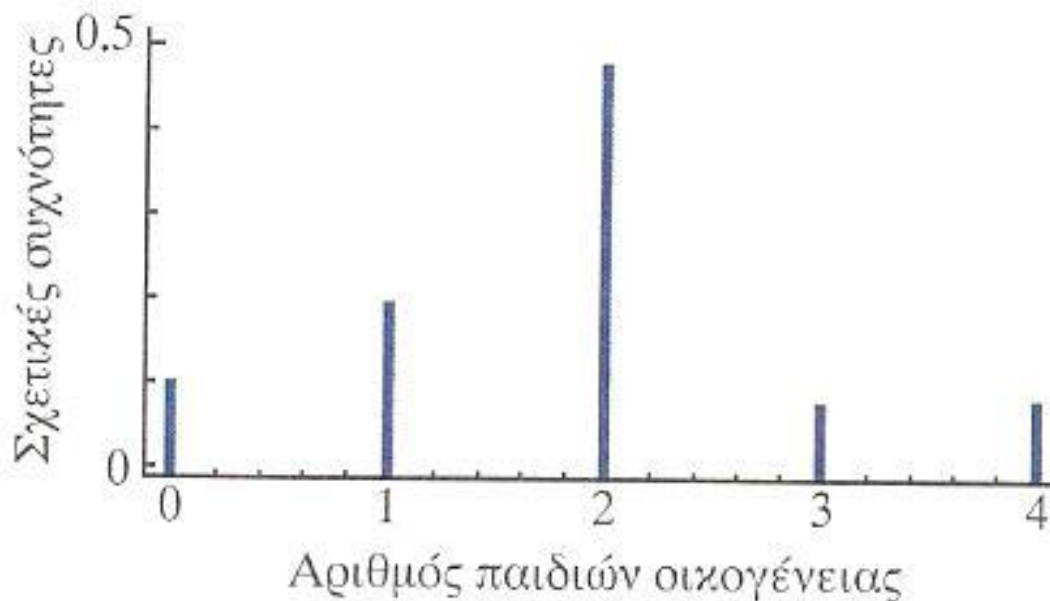
● Ποσοτικές μεταβλητές – Κυκλικό διάγραμμα (pie chart)



γ. Κυκλικό διάγραμμα.

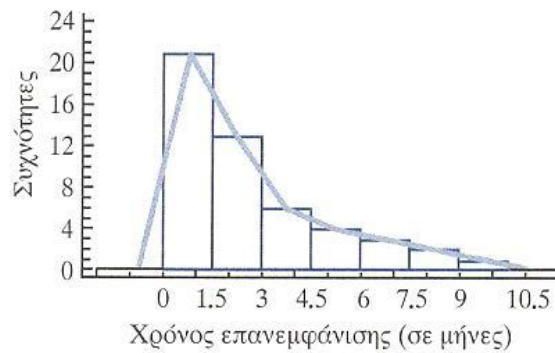
Γραφική παρουσίαση κατανομής συχνοτήτων (1.4)

- Ποσοτικές μεταβλητές – Διάγραμμα γραμμής (line diagram) σχετικών συχνοτήτων

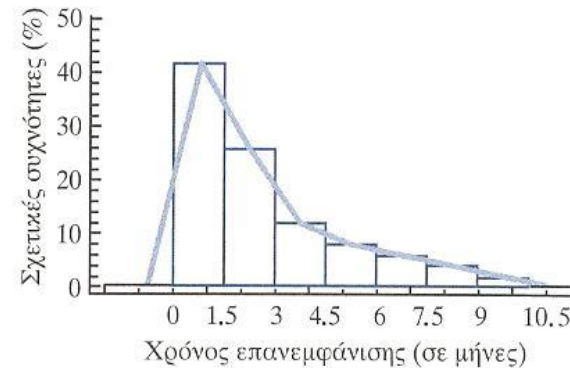


Γραφική παρουσίαση κατανομής συχνοτήτων (1.5)

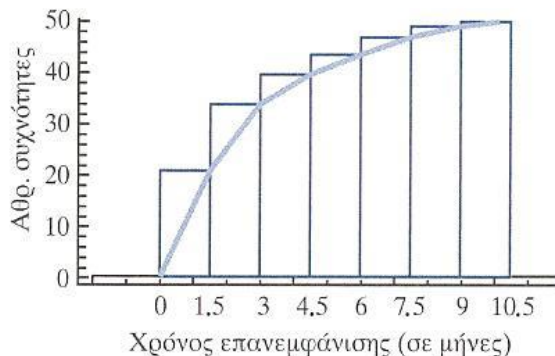
- Ποσοτικές μεταβλητές – Ιστόγραμμα (histogram) και Πολύγωνο (polygon)



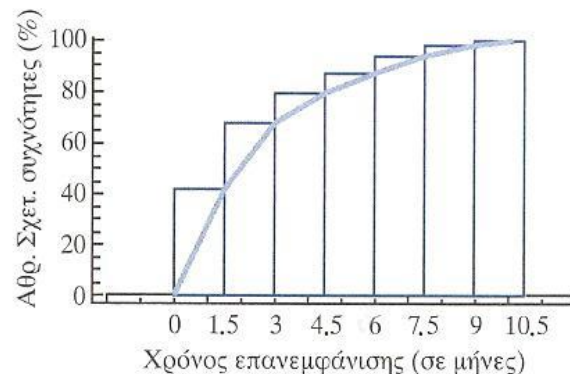
α. Ιστόγραμμα και πολύγωνο συχνοτήτων.



β. Ιστόγραμμα και πολύγωνο σχετικών συχνοτήτων.



γ. Ιστόγραμμα και πολύγωνο



δ. Ιστόγραμμα και πολύγωνο

Γραφική παρουσίαση κατανομής συχνοτήτων (1.6.1)

● Ποσοτικές μεταβλητές – Φυλλογράφημα (steam-leaf plot)

2.1	1.7	0.8	0.8	4.1	8.7	1.4	2.9	1.9	2.7
4.4	2.2	5.5	7.0	1.8	0.2	1.0	0.9	4.0	0.7
2.0	6.5	0.7	4.3	0.2	5.6	2.4	1.4	1.3	1.2
0.5	3.9	7.4	3.3	8.8	0.3	2.0	5.7	0.8	2.6
9.9	1.6	2.8	1.0	0.6	1.3	0.8	5.9	0.9	0.4

Γραφική παρουσίαση κατανομής συχνοτήτων (1.6.2)

ο Ποσοτικές μεταβλητές – Φυλλογράφημα (stem-leaf plot)

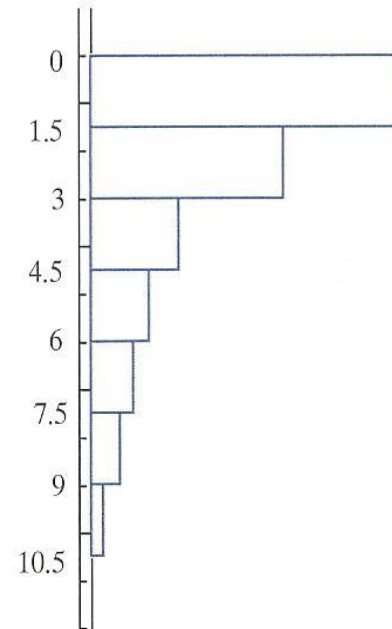
0		22345677888899
1		00233446789
2		001246789
3		39
4		0134
5		5679
6		5
7		04
8		7
		HI 9.9

Ακραία
Τιμή



(Leaf unit = 0.1)

α. Φυλλογράφημα.



β. Ιστόγραμμα συχνοτήτων.

Γραφική παρουσίαση κατανομής συχνοτήτων (2)

○ Ποιοτικές μεταβλητές

1. Ραβδόγραμμα συχνοτήτων και σχετικών συχνοτήτων
2. Κυκλικό διάγραμμα συχνοτήτων και σχετικών συχνοτήτων

Αριθμητικά περιγραφικά μέτρα (1)

ο Ποσοτικές μεταβλητές

■ Μέτρα Θέσης

1. Δειγματικός μέσος, \bar{x}

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \sum_{i=1}^k n_i y_i = \sum_{i=1}^k f_i y_i$$

➤ Σταθμικός μέσος: $\bar{x}_w = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$

2. Κορυφή του δείγματος (επικρατούσα τιμή), M_0

Η τιμή με τη μεγαλύτερη συχνότητα

Αριθμητικά περιγραφικά μέτρα (2)

ο Ποσοτικές μεταβλητές

- Μέτρα Θέσης

3. Διάμεσος του δείγματος, δ ή Q_2

Το πολύ 50 % των τιμών του δείγματος είναι μικρότερες από τη διάμεσο και επίσης το πολύ των 50 % των τιμών του δείγματος είναι μεγαλύτερες από τη διάμεσο

Σε αύξουσα διάταξη των x_1, x_2, \dots, x_n , τη θέση της διαμέσου δίνει ο αριθμός $0.5(n+1)$ εφόσον είναι ακέραιος, ενώ αν δεν είναι ακέραιος, τότε η διάμεσος είναι ίση με το ημιάθροισμα των δύο τιμών που οι θέσεις τους είναι οι πλησιέστερες στον αριθμό $0.5(n+1)$

Αριθμητικά περιγραφικά μέτρα (3)

ο Ποσοτικές μεταβλητές

▪ Μέτρα Θέσης

4. p -ποσοστιαία σημεία του δείγματος, x_p , $0 < p < 1$
Το πολύ $100p$ % των τιμών του δείγματος είναι μικρότερες από το p -ποσοστιαίο σημείο και επίσης το πολύ $100(1 - p)$ % των τιμών του δείγματος είναι μεγαλύτερες από το p -ποσοστιαίο σημείο
Σε αύξουσα διάταξη των x_1, x_2, \dots, x_n , τη θέση του p -ποσοστιαίου σημείου δίνει ο αριθμός $p(n+1)$ εφόσον είναι ακέραιος, ενώ αν δεν είναι ακέραιος, τότε το p -ποσοστιαίο σημείο εκτιμάται με παρεμβολή μεταξύ των δύο τιμών που οι θέσεις τους είναι οι πλησιέστερες στον $p(n+1)$

Αριθμητικά περιγραφικά μέτρα (4)

○ Ποσοτικές μεταβλητές

▪ Μέτρα Θέσης

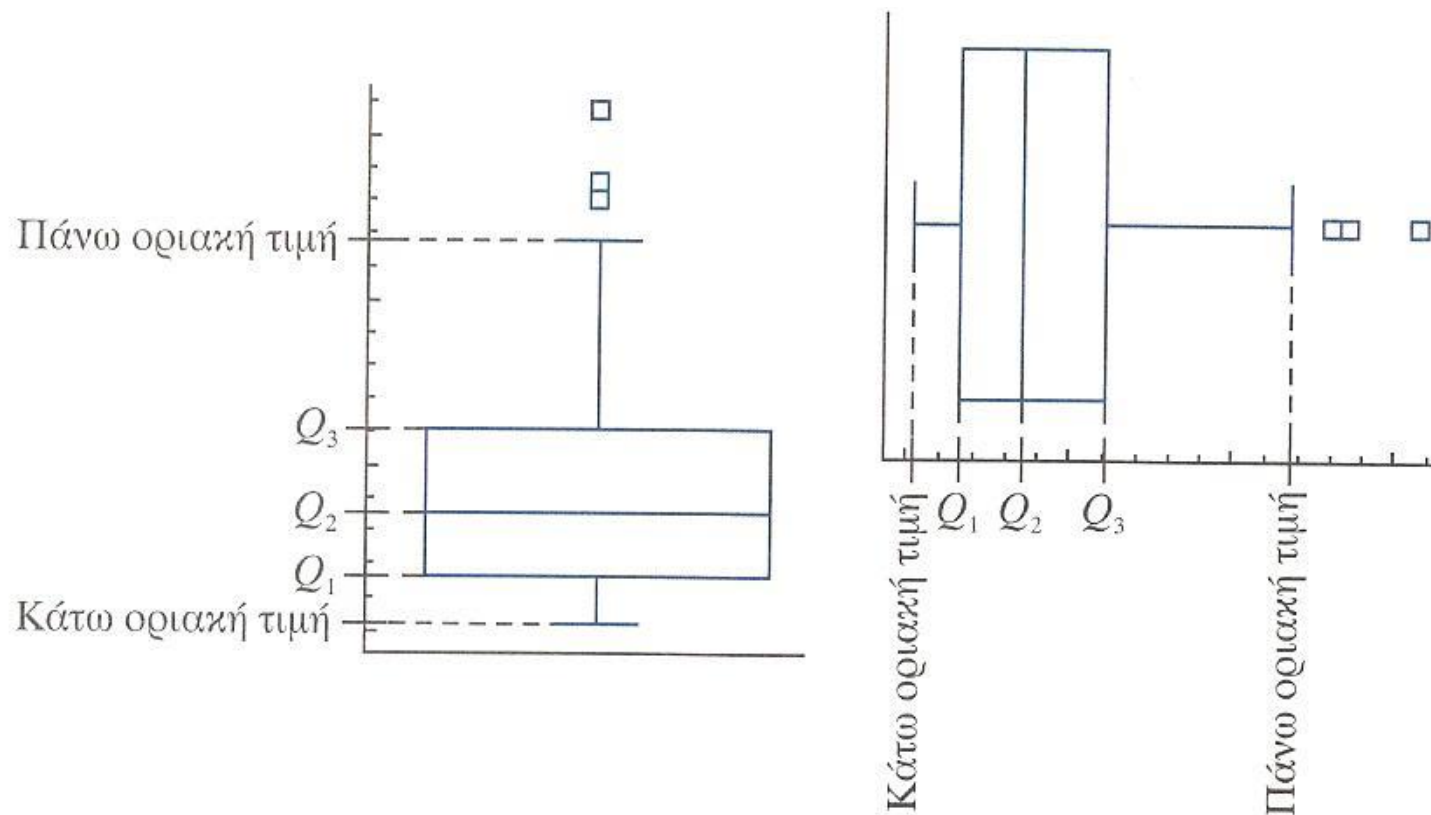
5. Τεταρτημόρια, Q_1 , Q_2 , Q_3

$$Q_1 = X_{0.25}$$

$$Q_2 = X_{0.5} = \delta$$

$$Q_3 = X_{0.75}$$

Θηκόγραμμα (box plot) (1)



Θηκόγραμμα (box plot) (2)

ο Πάνω οριακή τιμή

- Μέγιστη τιμή του δείγματος

ή

- Η μεγαλύτερη τιμή του δείγματος που είναι μικρότερη ή ίση από το ανώτερο εσωτερικό φράγμα $Q_3 + 1.5(Q_3 - Q_1)$

ή

- Η μεγαλύτερη τιμή του δείγματος που είναι μικρότερη ή ίση από το ανώτερο εξωτερικό φράγμα $Q_3 + 3(Q_3 - Q_1)$

Θηκόγραμμα (box plot) (3)

- **Κάτω οριακή τιμή**

- Ελάχιστη τιμή του δείγματος

ή

- Η μικρότερη τιμή του δείγματος που είναι μεγαλύτερη ή ίση από το κατώτερο εσωτερικό φράγμα $Q_1 - 1.5(Q_3 - Q_1)$

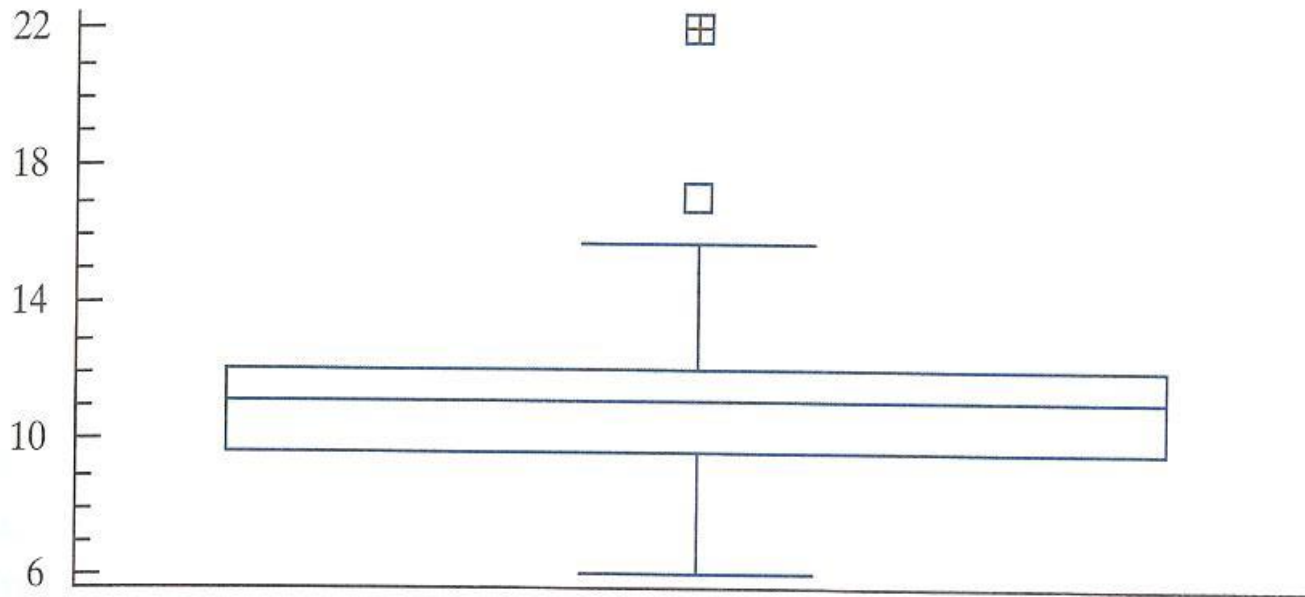
ή

- Η μικρότερη τιμή του δείγματος που είναι μεγαλύτερη ή ίση από το κατώτερο εξωτερικό φράγμα $Q_1 - 3(Q_3 - Q_1)$

Θηκόγραμμα (box plot) (4)

- Σε περίπτωση που ως οριακές τιμές δε χρησιμοποιούνται η μέγιστη και η ελάχιστη τιμή του δείγματος στο θηκόγραμμα σημειώνονται, όταν φυσικά υπάρχουν, οι **ακραίες** τιμές και οι **εξαιρετικά ακραίες** τιμές

Θηκόγραμμα (box plot) (5)



Ομαδοποίηση των τιμών του δείγματος σε k κλάσεις (1)

● Υπολογισμός κορυφής

$$M_0 = L_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} c_i$$

όπου

L_i : το κάτω άκρο της επικρατούσας κλάσης

c_i : το πλάτος της επικρατούσας κλάσης

$$\Delta_1 = n_i - n_{i-1}$$

$$\Delta_2 = n_i - n_{i+1}$$

n_i : η συχνότητα της επικρατούσας κλάσης

Ομαδοποίηση των τιμών του δείγματος σε k κλάσεις (2)

● Υπολογισμός του δειγματικού μέσου

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i y_i = \sum_{i=1}^k f_i y_i$$

όπου τα $y_i, i = 1, 2, \dots, k$ είναι οι κεντρικές τιμές των κλάσεων

Ομαδοποίηση των τιμών του δείγματος σε k κλάσεις (3)

● Υπολογισμός της διαμέσου

$$\delta = L_i + \frac{0.5n - N_{i-1}}{n_i} c_i$$

όπου

L_i : το κάτω άκρο της μεσαίας κλάσης

c_i : το πλάτος της μεσαίας κλάσης

n_i : η συχνότητα της μεσαίας κλάσης

N_{i-1} : η αθροιστική συχνότητα της προηγούμενης κλάσης από τη μεσαία

Ομαδοποίηση των τιμών του δείγματος σε k κλάσεις (4)

● Υπολογισμός p -ποσοσטיαίων σημείων

$$x_p = L_i + \frac{pn - N_{i-1}}{n_i} c_i$$

όπου

L_i : το κάτω άκρο της κλάσης στην οποία βρίσκεται το x_p

c_i : το πλάτος της

n_i : η συχνότητά της

N_{i-1} : η αθροιστική συχνότητα της προηγούμενης κλάσης

Παράδειγμα 1^ο

- Συμπληρώστε τον παρακάτω Πίνακα και υπολογίστε το δειγματικό μέσο:

y_i	n_i	f_i	N_i	F_i	$n_i y_i$	$n_i y_i^2$
0	2	0.1	2	0.1	0	0
1	4	0.2	6	0.3	4	4
2	10	0.5	16	0.8	20	40
3	2	0.1	18	0.9	6	18
4	2	0.1	20	1.0	8	32
Σύνολο	20	1.0			38	94

Παράδειγμα 2^ο

- Ένας οδηγός φορτηγού διανομής τροφίμων αγόρασε σε μια ημέρα πετρέλαιο από τρία διαφορετικά πρατήρια
- Από το 1^ο αγόρασε 6 λίτρα προς 0.75 € το λίτρο, από το 2^ο 12 λίτρα προς 0.84 € το λίτρο και από το 3^ο 5 λίτρα προς 0.76 € το λίτρο
- Υπολογίστε το μέσο ποσό που πλήρωσε ανά λίτρο

Παράδειγμα 3^ο

1. Έστω οι παρατηρήσεις 5, 2, 6, 9, 11
Βρείτε τη διάμεσο

2. Έστω οι παρατηρήσεις 2, 5, 6, 27, 11, 9
Βρείτε τη διάμεσο

Παράδειγμα 4^ο

● Έστω ο παρακάτω Πίνακας:

Εισόδημα	n_i	f_i	N_i	F_i
[900, 1100)	1	0.05	1	0.05
[1100, 1300)	4	0.20	5	0.25
[1300, 1500)	6	0.30	11	0.55
[1500, 1700)	4	0.20	15	0.75
[1700, 1900)	3	0.15	18	0.90
[1900, 2100)	2	0.10	20	1.00

● Να υπολογίσετε τη διάμεσο και το 0.95-ποσοστιαίο σημείο της κατανομής του δείγματος

Παράδειγμα 5^ο

- Να υπολογίσετε τα τεταρτημόρια της κατανομής του δείγματος:

15, 11, 11, 11, 22, 9, 11, 7, 11, 12, 12,
16, 8, 11, 15, 9, 10, 14, 9, 10, 11, 10, 6,
17, 11, 10, 8, 11

Παράδειγμα 5^ο

y_i	n_i	f_i	N_i	F_i
6	1	0,0357	1	0,0357
7	1	0,0357	2	0,0714
8	2	0,0714	4	0,1428
9	3	0,1071	7	0,2500
10	4	0,1428	11	0,3928
11	9	0,3214	20	0,7143
12	2	0,0714	22	0,7857
14	1	0,0357	23	0,8214
15	2	0,0714	25	0,8928
16	1	0,0357	26	0,9286
17	1	0,0357	27	0,9643
22	1	0,0357	28	1,000
Σύνολα	28	1,000		