

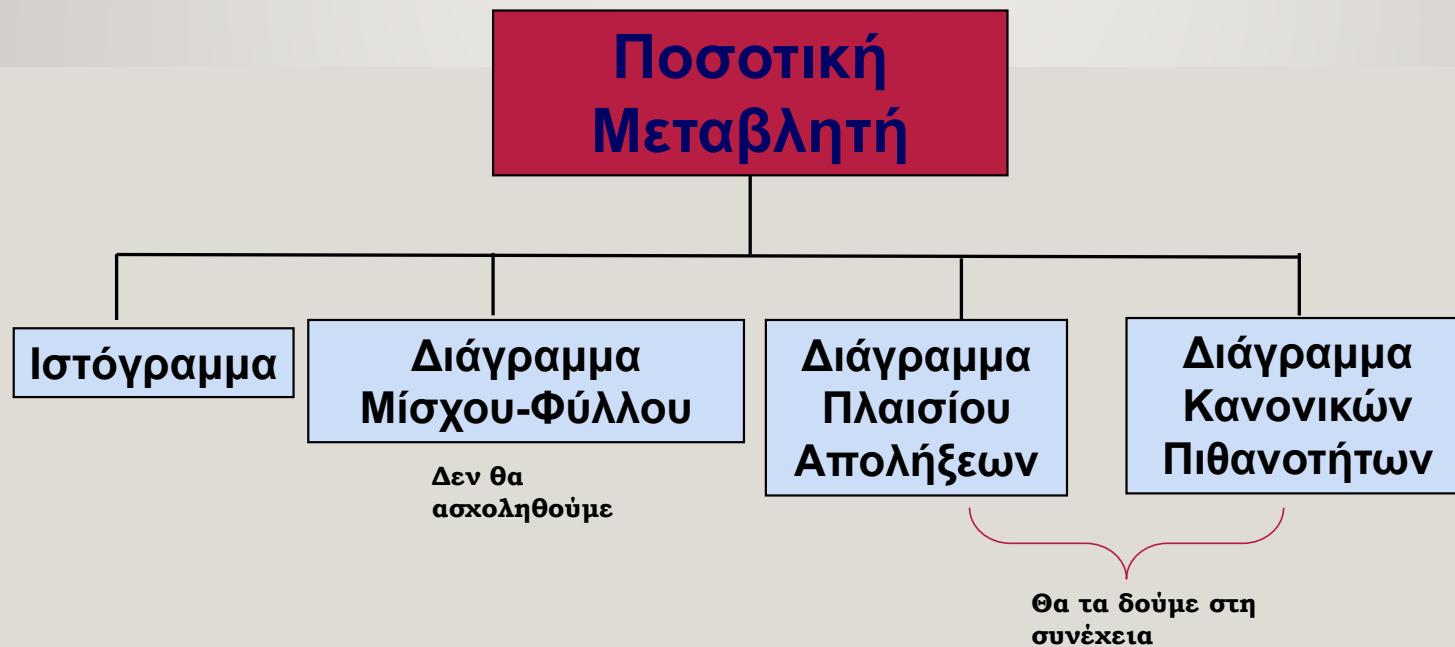
ΦΡΟΝΤΙΣΤΗΡΙΑ ΣΤΗ ΣΤΑΤΙΣΤΙΚΗ 1

3^ο ΦΡΟΝΤΙΣΤΗΡΙΟ



Δρ Ευγενία Τσομπανάκη

Απεικόνιση Ποσοτικών Μεταβλητών



Απεικόνιση Ποσοτικών Μεταβλητών

Ιστόγραμμα

Ημερήσιοι φοιτητές					
16	17	17	18	18	18
19	19	20	20	21	22
22	25	27	32	38	42

- Για να κατασκευαστεί το Ιστόγραμμα μίας Ποσοτικής μεταβλητής η Ποσοτική μεταβλητή μετατρέπεται σε **Διατάξιμη**

- Στον οριζόντιο άξονα (X) απεικονίζονται τα **διαστήματα τάξης** (*class boundaries*) ή τα **μέσα σημεία** (*class midpoints*).

Στην περίπτωση χρήσης Στατιστικού πακέτου η μετατροπή γίνεται «**αυθαίρετα**» από το πρόγραμμα αλλά ο χρήστης μπορεί να παρέμβει

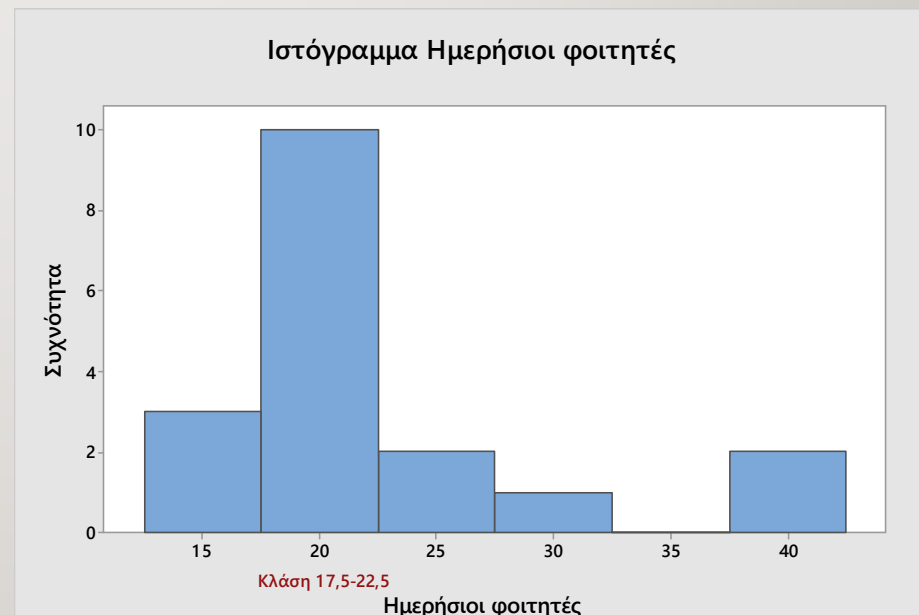
- Στον κάθετο άξονα (Y) απεικονίζεται η **συχνότητα** ή η **σχετική συχνότητα - ποσοστά**.

- Το ύψος των ράβδων απεικονίζει τη συχνότητα ή τη σχετική συχνότητα - ποσοστά.

Απεικόνιση Ποσοτικών Μεταβλητών

Ιστόγραμμα

Ημερήσιοι φοιτητές	Μέσο Τάξης	Συχνότητα
12,5 και κάτω του 17,5	15	3
17,5 και κάτω του 22,5	20	10
22,5 και κάτω του 27,5	25	2
27,5 και κάτω του 32,5	30	1
32,5 και κάτω του 37,5	35	0
37,5 και κάτω του 42,5	40	3
Σύνολο		18
Τάξεις εύρους 5		



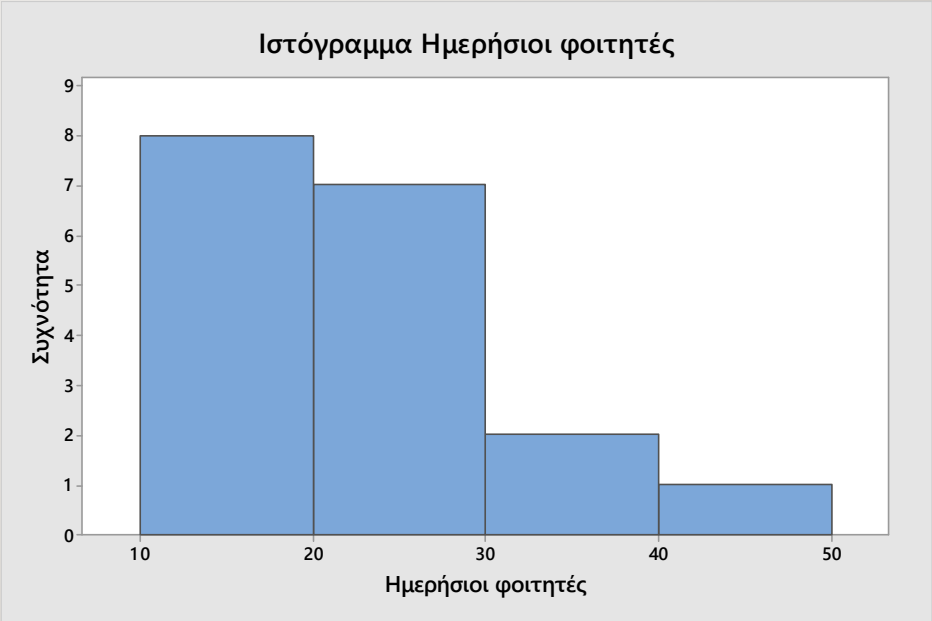
Απεικόνιση Ποσοτικών Μεταβλητών

Ιστόγραμμα

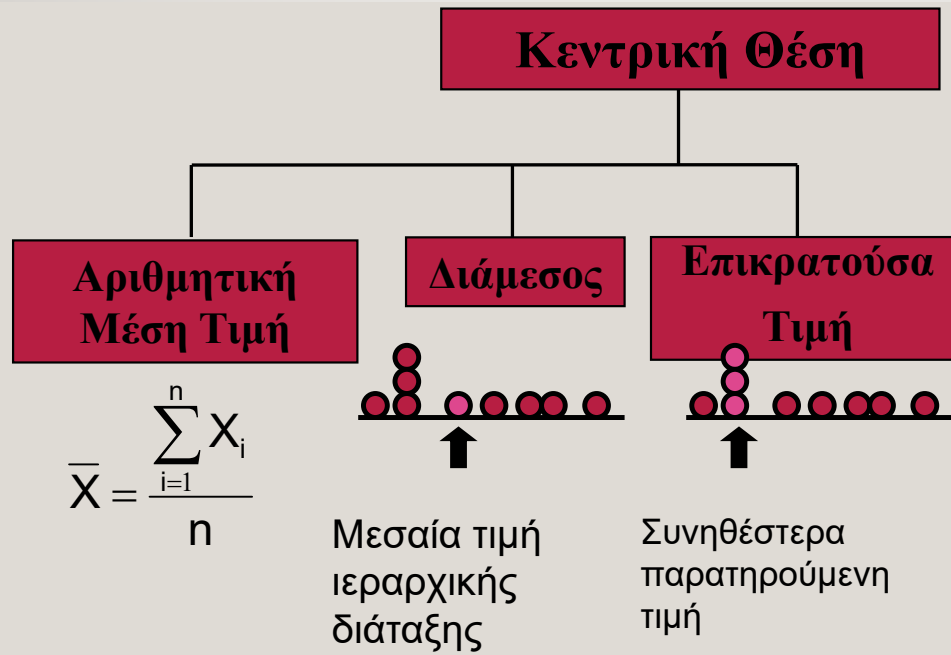
Ημερήσιοι φοιτητές					
16	17	17	18	18	18
19	19	20	20	21	22
22	25	27	32	38	42

Ημερήσιοι φοιτητές	Μέσο Τάξης	Συχνότητα
10 και κάτω του 20	15	8
20 και κάτω του 30	25	7
30 και κάτω του 40	35	2
40 και κάτω του 50	45	1
Σύνολο		18

Τάξεις εύρους 10

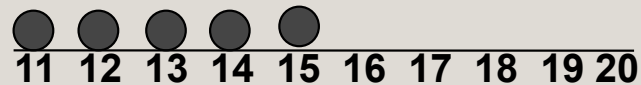


Μέτρα Κεντρικής Τάσης ή Θέσης: Συνοπτική Παρουσίαση



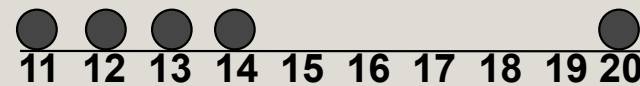
Μέτρα της Κεντρικής Τάσης: Η μέση τιμή

- Το πλέον κοινό μέτρο της κεντρικής τάσης
- Μέση τιμή = Το άθροισμα των τιμών διαιρούμενο με το πλήθος του δείγματος
- Επηρεάζεται από τις ακραίες τιμές (outliers)



Μέση τιμή = 13

$$\frac{11+12+13+14+15}{5} = \frac{65}{5} = 13$$



Μέση τιμή = 14

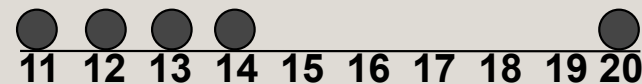
$$\frac{11+12+13+14+20}{5} = \frac{70}{5} = 14$$

Μέτρα της Κεντρικής Τάσης: Η Διάμεσος

- Η διάμεσος είναι η «μεσαία τιμή» μιας **ιεραρχημένης** σειράς δεδομένων (50% των τιμών είναι μεγαλύτερες της μεσαίας τιμής και 50% μικρότερες)



Διάμεσος = 13

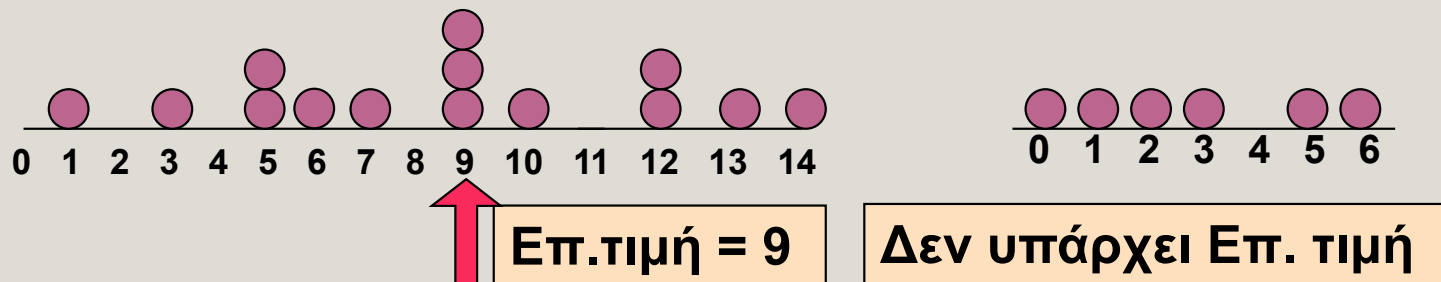


Διάμεσος = 13

- Δεν επηρεάζεται από τις ακραίες τιμές

Μέτρα της Κεντρικής Τάσης: Η Επικρατούσα Τιμή

- Είναι η τιμή που εμφανίζεται συχνότερα
- Δεν επηρεάζεται από τις ακραίες τιμές
- Χρησιμοποιείται τόσο για τις αριθμητικές μεταβλητές όσο και για τις κατηγορικές
- Μπορεί και να μην υπάρχει επικρατούσα τιμή
- Μπορεί να υπάρχουν περισσότερες επικρατούσες τιμές



1^ο Πρόβλημα

Ένας καθηγητής ρωτάει τους 20 μαθητές του να του πουν πόσες **ώρες** ασχολείται ο καθένας με τον αθλητισμό και καταγράφει τα στοιχεία στον παρακάτω πίνακα. Να **συνοψίσετε** τα δεδομένα με τη βοήθεια των μέτρων κεντρικής θέσης.

1	2	0	1	2	1	1	2	4	3	2	0	1	2	2	0	2	4	1	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Λύση

Μέση τιμή:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1+2+\dots+1}{20} = \frac{32}{20} = 1,6 \text{ ώρες}$$

1^ο Πρόβλημα

Ένας καθηγητής ρωτάει τους 20 μαθητές του να του πουν πόσες **ώρες** ασχολείται ο καθένας με τον αθλητισμό και καταγράφει τα στοιχεία στον παρακάτω πίνακα.

Ο τελευταίος μαθητής κατά λάθος έγραψε 11 αντί για 1, πως επηρεάζει αυτό **το λάθος** στην καταχώρηση τη μέση τιμή;

2	0	1	2	1	1	2	4	3	2	0	1	2	2	0	2	4	1	11	
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	--

Λύση

Μέση τιμή:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{X_1+X_2+.....+X_n}{n} = \frac{1+2+.....+11}{20} = \frac{42}{20} = 2,10 \text{ **ώρες**}$$

Η Μέση τιμή είναι ευαίσθητο μέτρο σε ακραίες τιμές. Δεν ενδείκνυται η χρήση της σε αυτή την περίπτωση



Σημαντική παρατήρηση

Για τον υπολογισμό της διαμέσου (δες αντίστοιχη ενότητα τεταρτημόρια) έχουν προταθεί διαφορετικοί τρόποι στη βιβλιογραφία. Εδώ αναφέρουμε δύο από αυτούς.

Η μέθοδος που **θα πρέπει να ακολουθήσετε** για τους υπολογισμούς και τη δημιουργία διαγράμματος box-plot αναφέρεται στις διαφάνειες ως **ΥΠΟΛΟΓΙΣΜΟΙ όπως το ΒΙΒΛΙΟ** ^[1]

Στη μέθοδο αυτή χρησιμοποιείται γραμμική παρεμβολή.

Σε άλλα εγχειρίδια ^[2] στρογγυλοποιούν τη θέση στον πλησιέστερο ακέραιο και μόνο αν η θέση είναι δεκαδικός αριθμός με δεκαδικό ,5 (πχ 1,5 ή 3,5 κ.ο.) θεωρούν ως τεταρτημόριο το μέσο όρο των παρατηρήσεων που αντιστοιχούν στις δύο γειτονικές τάξεις.

Θέση

$$2 * \frac{(n + 1)}{4}$$

Τιμή

$$Q_2 = X_{(A_Q)} + \Delta_Q [X_{(A_Q+1)} - X_{(A_Q)}]$$

$$A_Q : \text{το ακέραιο μέρος του } 2 * \frac{n+1}{4} \quad \Delta_Q : \text{το δεκαδικό μέρος του } 2 * \frac{n+1}{4}$$

^[1] Αγγελής, Β, Δημάκη, Αικ. ΣΤΑΤΙΣΤΙΚΗ ΤΟΜΟΣ Α. Εκδόσεις ΣΟΦΙΑ & Χαλκιάς Ι. Στατιστική Μέθοδοι Ανάλυσης για Επιχειρηματικές Αποφάσεις. 3^η Έκδοση, Rossili

^[2] Berenson, M. L., Levine, D. M., Szabat, K. A., & Stephan, D. F. (2020). *Βασικές Αρχές Στατιστικής για Επιχειρήσεις-Έννοιες και Εφαρμογές*. BROKEN HILL PUBLISHERS LTD.

0	0	0	1	1	1	1	1	1	1	1	2	2	2	2	2	2	3	4	4
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

ΥΠΟΛΟΓΙΣΜΟΙ όπως το
ΒΙΒΛΙΟ

2^ο Βήμα: Εύρεση θέσης Τεταρτημορίων & Τιμής

Θέση

$$2 * \frac{(n + 1)}{4}, i = 1,2,3$$

Τιμή

$$Q_2 = X_{(A_Q)} + \Delta_Q [X_{(A_Q+1)} - X_{(A_Q)}]$$

A_Q : το ακέραιο μέρος του $2 * \frac{n+1}{4}$ Δ_Q : το δεκαδικό μέρος του $2 * \frac{n+1}{4}$

$$2 * \frac{21}{4} = 10,5$$

A_Q : το ακέραιο μέρος του 10, 5 = 10

Δ_Q : το δεκαδικό μέρος του 10, 5 = 0, 5

$X_{(10)} = 1$

$X_{(10+1)} = 2$

$Q_2 = 1 + 0,5 * (2 - 1) = 1,5$ ώρες



Ο τελευταίος μαθητής όπως και προηγουμένως **κατά λάθος** έγραψε 11 αντί για 1, πως επηρεάζει αυτό το λάθος στην καταχώρηση τη διάμεσο;

Εύρεση **Διαμέσου** —————> 1^ο **Βήμα: Ταξινόμηση δεδομένων κατά αύξουσα σειρά**

0	0	0	1	1	1	1	1	1	1	1	2	2	2	2	2	2	3	4	11
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

ΥΠΟΛΟΓΙΣΜΟΙ όπως το ΒΙΒΛΙΟ

2^ο **Βήμα: Εύρεση θέσης Τεταρτημορίων & Τιμής**

Θέση

$$2 * \frac{(n + 1)}{4}, i = 1,2,3$$

$$2 * \frac{21}{4} = 10,5$$

A_Q : το ακέραιο μέρος του 10, 5 = 10

Δ_Q : το δεκαδικό μέρος του 10, 5 = 0, 5

Τιμή

$$Q_2 = X_{(A_Q)} + \Delta_Q [X_{(A_Q+1)} - X_{(A_Q)}]$$

A_Q : το ακέραιο μέρος του $2 * \frac{n+1}{4}$ Δ_Q : το δεκαδικό μέρος του $2 * \frac{n+1}{4}$

$$X_{(10)} = 1$$

$$X_{(10+1)} = 2$$

$$Q_2 = 1 + 0,5 * (2 - 1) = 1,5 \text{ ώρες}$$

Δεν επηρεάστηκε η τιμή της διαμέσου από την ύπαρξη της «ασυνήθιστης» τιμής

0	0	0	1	1	1	1	1	1	1	2	2	2	2	2	2	2	3	4
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

ΥΠΟΛΟΓΙΣΜΟΙ όπως το
ΒΙΒΛΙΟ

2^ο Βήμα: Εύρεση θέσης Τεταρτημορίων & Τιμής

Θέση

$$2 * \frac{(n + 1)}{4}, i = 1,2,3$$

Τιμή

$$Q_2 = X_{(A_Q)} + \Delta_Q [X_{(A_Q+1)} - X_{(A_Q)}]$$

A_Q : το ακέραιο μέρος του $2 * \frac{n+1}{4}$ Δ_Q : το δεκαδικό μέρος του $2 * \frac{n+1}{4}$

$$2 * \frac{20}{4} = 10$$

A_Q : το ακέραιο μέρος του 10 = 10

Δ_Q : το δεκαδικό μέρος του 10 = 0

$$X_{(10)} = 1$$

$$Q_2 = 1 + 0 * (2 - 1) = 1 \text{ ώρες}$$

$$X_{(10+1)} = 2$$

0	0	0	1	1	1	1	1	1	1	2	2	2	2	2	2	2	3	4	4
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

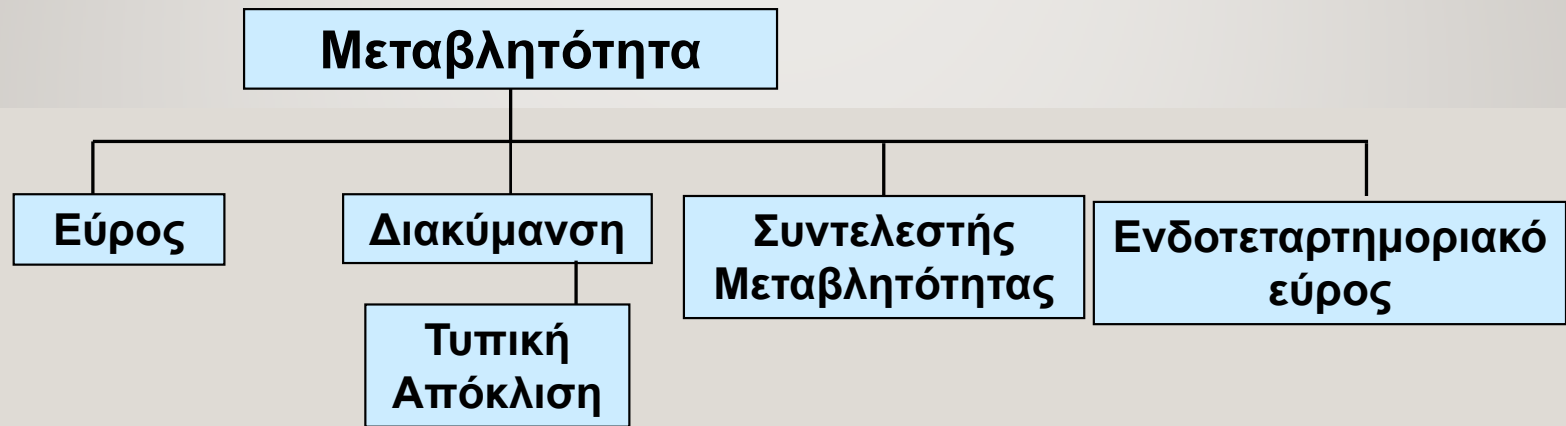
**Επικρατούσα
τιμή**

→ Τιμή που εμφανίζεται συχνότερα

Τιμή	Συχνότητα
0	3
1	7
2	7
3	1
4	2

Συνεπώς η τιμές **1** & **2** ώρες είναι οι επικρατούσες τιμή, καθώς παρατηρούνται τις περισσότερες φορές.

Μέτρα Μεταβλητότητας ή Διακύμανσης



- Τα μέτρα της διακύμανσης-μεταβλητότητας παρέχουν πληροφορίες για το **άπλωμα** (*spread*) ή την **μεταβλητότητα** (*variability or dispersion*) των τιμών.

Εύρος

1^ο Βήμα: **Ταξινόμηση** δεδομένων κατά αύξουσα σειρά

X	0	0	0	1	1	1	1	1	1	1	2	2	2	2	2	2	2	3	4	4
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

$$\text{Εύρος} = x_{\max} - x_{\min} = 4 - 1 = 3 \text{ ώρες}$$

Ο τελευταίος μαθητής όπως και προηγουμένως κατά λάθος έγραψε 11 αντί για 1, πως επηρεάζει αυτό το λάθος στην καταχώρηση το εύρος;

0	0	0	1	1	1	1	1	1	1	2	2	2	2	2	2	2	3	4	11
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

$$\text{Εύρος} = x_{\max} - x_{\min} = 11 - 1 = 10 \text{ ώρες}$$

Συνεπώς βλέπουμε ότι το εύρος **είναι** ευαίσθητο σε ακραίες τιμές



	X
1	1
2	2
3	0
4	1
5	2
6	1
7	1
8	2
9	4
10	3
11	2
12	0
13	1
14	2
15	2
16	0
17	2
18	4
19	1
20	1

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Δειγματική
διακύμανση

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{24,8}{19} = 1,31 \text{ ώρες}^2$$

$$\bar{X} = 1,6$$

	X	Xi-Xbar	(Xi-Xbar)^2
1	1	-0,60	0,36
2	2	0,40	0,16
3	0	-1,60	2,56
4	1	-0,60	0,36
5	2	0,40	0,16
6	1	-0,60	0,36
7	1	-0,60	0,36
8	2	0,40	0,16
9	4	2,40	5,76
10	3	1,40	1,96
11	2	0,40	0,16
12	0	-1,60	2,56
13	1	-0,60	0,36
14	2	0,40	0,16
15	2	0,40	0,16
16	0	-1,60	2,56
17	2	0,40	0,16
18	4	2,40	5,76
19	1	-0,60	0,36
20	1	-0,60	0,36
Sum	32	0,00	24,8
Average	1,6		1,31

Δειγματική
Τυπική
Απόκλιση



$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{24,8}{19}} = 1,142 \text{ ώρες}$$

Συντελεστής
Μεταβλητότητας



$$CV = \frac{S}{\bar{X}} = \frac{1,142}{1,6} = 0,714 \text{ Δεν έχει μονάδες}$$

	X	Xi-Xbar	(Xi-Xbar)^2
1	1	-0,60	0,36
2	2	0,40	0,16
3	0	-1,60	2,56
4	1	-0,60	0,36
5	2	0,40	0,16
6	1	-0,60	0,36
7	1	-0,60	0,36
8	2	0,40	0,16
9	4	2,40	5,76
10	3	1,40	1,96
11	2	0,40	0,16
12	0	-1,60	2,56
13	1	-0,60	0,36
14	2	0,40	0,16
15	2	0,40	0,16
16	0	-1,60	2,56
17	2	0,40	0,16
18	4	2,40	5,76
19	1	-0,60	0,36
20	1	-0,60	0,36
Sum	32	0,00	24,8
Average	1,6		1,31

Ένας καθηγητής ρωτάει τους 20 μαθητές του να του πουν πόσα **λεπτά** ασχολείται ο καθένας με τον αθλητισμό και καταγράφει τα στοιχεία στον παρακάτω πίνακα. Να **συνοψίσετε** τα δεδομένα.

Δειγματική μέση τιμή → 96 **λεπτά**

Δειγματική διακύμανση → 4698,947 **λεπτά²**

Δειγματική Τυπική Απόκλιση → 68,5 **λεπτά**

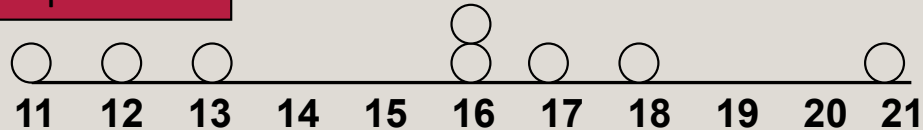
Συντελεστής Μεταβλητότητας → $CV = \frac{S}{\bar{X}} = \frac{68,5}{96} = 0,714$ **Δεν έχει μονάδες**

	X	Xi-Xbar	(Xi-Xbar)^2
1	60	-36,00	1.296,00
2	120	24,00	576,00
3	0	-96,00	9.216,00
4	60	-36,00	1.296,00
5	120	24,00	576,00
6	60	-36,00	1.296,00
7	60	-36,00	1.296,00
8	120	24,00	576,00
9	240	144,00	20.736,00
10	180	84,00	7.056,00
11	120	24,00	576,00
12	0	-96,00	9.216,00
13	60	-36,00	1.296,00
14	120	24,00	576,00
15	120	24,00	576,00
16	0	-96,00	9.216,00
17	120	24,00	576,00
18	240	144,00	20.736,00
19	60	-36,00	1.296,00
20	60	-36,00	1.296,00
Sum	1920	0,00	89280
Average	96		4.698,95

Μέτρα Μεταβλητότητας: Σύγκριση Τυπικών Αποκλίσεων

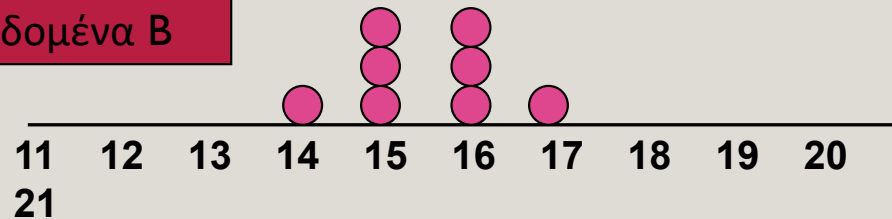
Μέση τιμή = 15.5

Δεδομένα Α



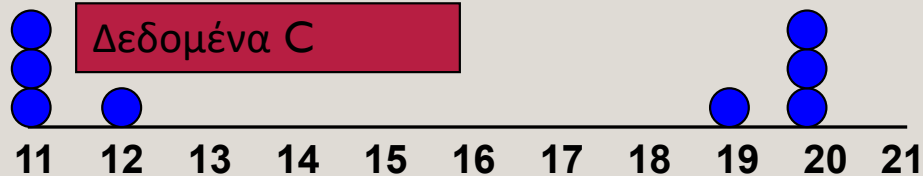
Μέση τιμή = 15.5
 $S = 3.338$

Δεδομένα Β



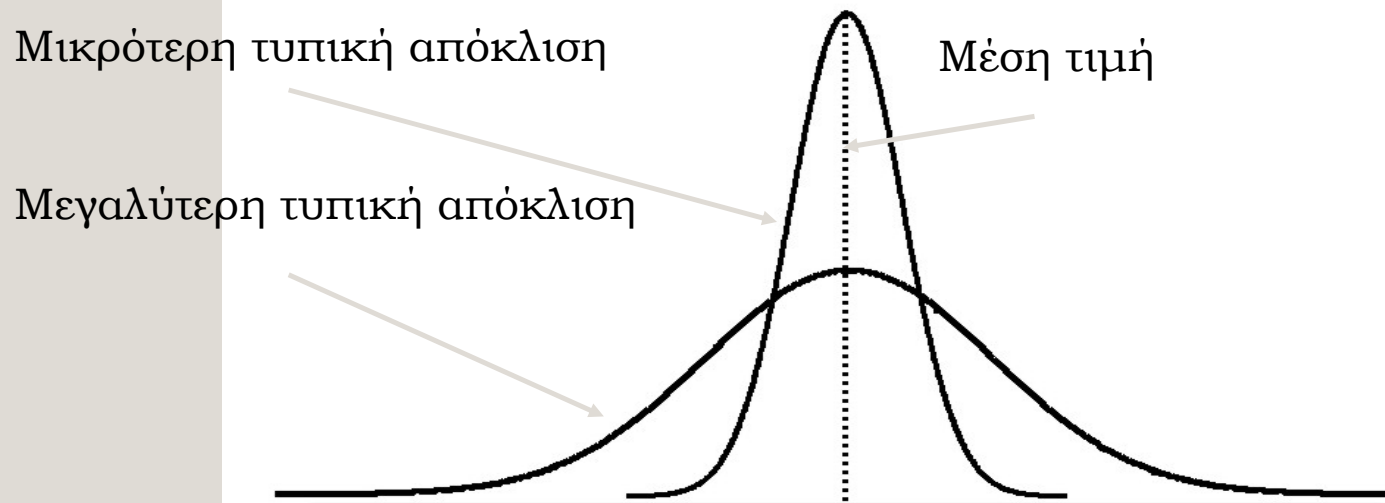
Μέση τιμή = 15.5
 $S = 0.926$

Δεδομένα C



Μέση τιμή = 15.5
 $S = 4.567$

Μέτρα Μεταβλητότητας: Σύγκριση Τυπικών Αποκλίσεων



Μέτρα Μεταβλητότητας: Ο Συντελεστής Μεταβλητότητας

- Μετρά τη **σχετική μεταβλητότητα**
- Εκφράζεται **ως ποσοστό (%)**
- Δείχνει τη μεταβλητότητα **σε σχέση με τη μέση τιμή**
- Μπορεί να χρησιμεύσει για τη σύγκριση της μεταβλητότητας μεταξύ δύο ή περισσότερων συνόλων δεδομένων που εκφράζονται σε διαφορετικές μονάδες μέτρησης

$$CV = \left(\frac{S}{\bar{X}} \right) \cdot 100\%$$

Μέτρα Μεταβλητότητας:

Σύγκριση Συντελεστών Μεταβλητότητας

- Μετοχή Α:

- Μέση τιμή προηγούμενου έτους = \$50
- Τυπική απόκλιση = \$5

$$CV_A = \left(\frac{S}{\bar{X}} \right) \cdot 100\% = \frac{\$5}{\$50} \cdot 100\% = 10\%$$

Η μετοχή C έχει πολύ μικρότερη τυπική απόκλιση αλλά πολύ μεγαλύτερο συντελεστή μεταβλητότητας

- Μετοχή C:

- Μέση τιμή προηγούμενου έτους = \$8
- Τυπική απόκλιση = \$2

$$CV_C = \left(\frac{S}{\bar{X}} \right) \cdot 100\% = \frac{\$2}{\$8} \cdot 100\% = 25\%$$

Μέτρα Μεταβλητότητας: Σύγκριση Συντελεστών Μεταβλητότητας

- Μετοχή Α:

- Μέση τιμή προηγούμενου έτους = \$50
- Τυπική απόκλιση = \$5

$$CV_A = \left(\frac{S}{\bar{X}} \right) \cdot 100\% = \frac{\$5}{\$50} \cdot 100\% = 10\%$$

- Μετοχή D:

- Μέση τιμή προηγούμενου έτους = 50 euro
- Τυπική απόκλιση = 2 euro

$$CV_D = \left(\frac{S}{\bar{X}} \right) \cdot 100\% = \frac{2 \text{ euro}}{50 \text{ euro}} \cdot 100\% = 4\%$$

Διαφορετικές μονάδες μέτρησης της τιμής της μετοχής. Οι διασπορές των μετοχών Α και D δεν είναι άμεσα συγκρίσιμες.

2ο Πρόβλημα

Στον παρακάτω πίνακα παρουσιάζονται οι μισθοί 5 δημοσίων υπαλλήλων.

A/A	Μισθός (X)
1	1.000
2	880
3	1.500
4	1.800
5	2.050

Να υπολογιστεί η μέση τιμή και να συγκριθεί με τη μεταβολή που θα προκύψει στις εξής περιπτώσεις:

- a) Αν δοθεί σε όλους επίδομα 100€.
- b) Αν γίνει περικοπή σε όλους 100€.
- c) Αν γίνει αύξηση 50% στους μισθούς.
- d) Αν γίνει περικοπή 50%

Λύση

A/A	Μισθός (X)
1	1.000
2	880
3	1.500
4	1.800
5	2.050
ΣΥΝΟΛΟ	7230

Αύξηση 100€

A/A	Μισθός (X)
1	1.100
2	980
3	1.600
4	1.900
5	2.150
ΣΥΝΟΛΟ	7730

Μείωση 100€

A/A	Μισθός (X)
1	900
2	780
3	1.400
4	1.700
5	1.950
ΣΥΝΟΛΟ	6730

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{1000+880+1500+1800+2050}{5} = \frac{7230}{5} = 1446$$

$$E(X \pm a) = a \pm E(X)$$

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{1100+980+1600+1900+2150}{5} = \frac{7730}{5} = 1546$$

$$(1446 + 100)$$

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{900+780+1400+1700+1950}{5} = \frac{6730}{5} = 1346$$

$$(1446 - 100)$$

Λύση

A/A	Μισθός (X)
1	1.000
2	880
3	1.500
4	1.800
5	2.050
ΣΥΝΟΛΟ	7230

Αύξηση 50%

A/A	Μισθός (X)
1	1500
2	1320
3	2250
4	2700
5	3075
ΣΥΝΟΛΟ	10845

Μείωση 50%

A/A	Μισθός (X)
1	500
2	440
3	750
4	900
5	1025
ΣΥΝΟΛΟ	3615

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{1000+880+1500+1800+2050}{5} = \frac{7230}{5} = 1446$$

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{2169}{5} = 2169$$

$$(1,5 \cdot 1446)$$

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{3615}{5} = 723$$

$$(0,5 \cdot 1446)$$

$$E(aX) = a \cdot E(X)$$

2ο Πρόβλημα

Στον παρακάτω πίνακα παρουσιάζονται οι μισθοί 5 δημοσίων υπαλλήλων.

A/A	Μισθός (X)
1	1.000
2	880
3	1.500
4	1.800
5	2.050

Να υπολογιστεί η τυπική απόκλιση και να συγκριθεί με τη μεταβολή που θα προκύψει στις εξής περιπτώσεις:

- a) Αν δοθεί σε όλους επίδομα 100€.
- b) Αν γίνει περικοπή σε όλους 100€.
- c) Αν γίνει αύξηση 50% στους μισθούς.
- d) Αν γίνει περικοπή 50%

Λύση

A/A	Μισθός (X)	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
1	1.000	-446	198.916
2	880	-566	320.356
3	1.500	54	2.916
4	1.800	354	125.316
5	2.050	604	364.816
ΣΥΝΟΛΟ			1.012.320

Αύξηση 100€

A/A	Μισθός (X)	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
1	1.100	-446	198.916
2	980	-566	320.356
3	1.600	54	2.916
4	1.900	354	125.316
5	2.150	604	364.816
ΣΥΝΟΛΟ			1.012.320

Μείωση 100€

A/A	Μισθός (X)	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
1	900	-446	198.916
2	780	-566	320.356
3	1.400	54	2.916
4	1.700	354	125.316
5	1.950	604	364.816
ΣΥΝΟΛΟ			1.012.320

$$S_X = \sqrt{\frac{\sum_{i=1}^5 (x_i - \bar{X})^2}{5 - 1}}$$

$$\bar{X} = 1446$$

$$S_X = 503,07$$

$$\bar{X} = 1546$$

$$S_X = 503,07$$

$$\bar{X} = 1346$$

$$S_X = 503,07$$

$$\text{Var}(X \pm a) = \text{Var}(X)$$

$$\text{sd}(X \pm a) = \text{sd}(X)$$

Λύση

A/A	Μισθός (X)
1	1.000
2	880
3	1.500
4	1.800
5	2.050
ΣΥΝΟΛΟ	7230

Αύξηση 50%

A/A	Μισθός (X)	Xi-Xbar	(Xi-Xbar)^2
1	1.500	-669	447.561
2	1.320	-849	720.801
3	2.250	81	6.561
4	2.700	531	281.961
5	3.075	906	820.836
ΣΥΝΟΛΟ			2.277.720

Μείωση 50%

A/A	Μισθός (X)	Xi-Xbar	(Xi-Xbar)^2
1	500	-223	49.729
2	440	-283	80.089
3	750	27	729
4	900	177	31.329
5	1.025	302	91.204
ΣΥΝΟΛΟ			253.080

$$\bar{X} = 1446$$

$$S_X = 503,07$$

$$S_X = \sqrt{\frac{\sum_{i=1}^5 (x_i - \bar{X})^2}{5 - 1}}$$

$$\bar{X} = 2169$$

$$S_X = 754,61$$

$$\text{Var}(aX) = a^2 \cdot \text{Var}(X)$$

$$\text{sd}(aX) = a \cdot \text{sd}(X)$$

$$(1,5 \cdot 503,07) = 754,61$$

$$\bar{X} = 723$$

$$S_X = 251,54$$

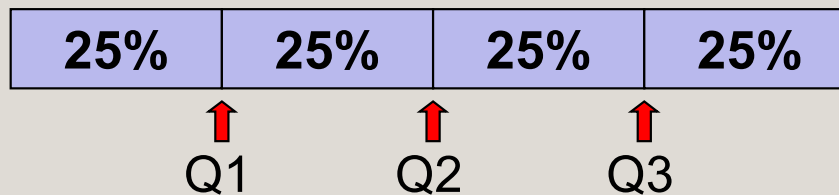
$$(0,5 \cdot 503,07) = 251,54$$

Μέτρα σχετικής θέσης



Μέτρα σχετικής θέσης: Τεταρτημόρια

- Τα **τεταρτημόρια** (quartiles) χωρίζουν τις **ιεραρχημένες κατά αύξουσα τάξη** τιμές ενός δείγματος σε **4** ίσα μέρη με ίσο αριθμό τιμών ανά τεταρτημόριο



- Το πρώτο τεταρτημόριο, Q_1 , χωρίζει το **κάτω 25%** των τιμών από το υπόλοιπο 75% που είναι μεγαλύτερες.
- Το δεύτερο τεταρτημόριο, Q_2 , συμπίπτει με τη διάμεσο (50% των τιμών είναι μικρότερες και 50% μεγαλύτερες).
- Το 25% μόνο των τιμών είναι μεγαλύτερες από το τρίτο τεταρτημόριο.

Σημαντική παρατήρηση

Για τον υπολογισμό των τεταρτημορίων έχουν προταθεί διαφορετικοί τρόποι στη βιβλιογραφία. Εδώ αναφέρουμε δύο από αυτούς.

Η μέθοδος που **θα πρέπει να ακολουθήσετε** για τους υπολογισμούς και τη δημιουργία διαγράμματος box-plot αναφέρεται στις διαφάνειες ως **ΥΠΟΛΟΓΙΣΜΟΙ όπως το ΒΙΒΛΙΟ** ^[1]

Στη μέθοδο αυτή χρησιμοποιείται γραμμική παρεμβολή.

Σε άλλα εγχειρίδια ^[2] στρογγυλοποιούν τη θέση στον πλησιέστερο ακέραιο και μόνο αν η θέση είναι δεκαδικός αριθμός με δεκαδικό ,5 (πχ 1,5 ή 3,5 κ.ο.) θεωρούν ως τεταρτημόριο το μέσο όρο των παρατηρήσεων που αντιστοιχούν στις δύο γειτονικές τάξεις.

Θέση

$$i * \frac{(n+1)}{4}, i = 1, 2, 3$$

Τιμή

$$Q_i = X_{(A_Q)} + \Delta_Q [X_{(A_Q+1)} - X_{(A_Q)}]$$

$$A_Q : \text{το ακέραιο μέρος του } i * \frac{n+1}{4} \quad \Delta_Q : \text{το δεκαδικό μέρος του } i * \frac{n+1}{4}$$

^[1] Αγγελής, Β, Δημάκη, Αικ. ΣΤΑΤΙΣΤΙΚΗ ΤΟΜΟΣ Α. Εκδόσεις ΣΟΦΙΑ & Χαλκιάς Ι. Στατιστική Μέθοδοι Ανάλυσης για Επιχειρηματικές Αποφάσεις. 3^η Έκδοση, Rossili

^[2] Berenson, M. L., Levine, D. M., Szabat, K. A., & Stephan, D. F. (2020). *Βασικές Αρχές Στατιστικής για Επιχειρήσεις-Έννοιες και Εφαρμογές*. BROKEN HILL PUBLISHERS LTD.

0	0	0	1	1	1	1	1	1	2	2	2	2	2	2	2	3	4	4	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

ΥΠΟΛΟΓΙΣΜΟΙ όπως το
ΒΙΒΛΙΟ

2^ο Βήμα: Εύρεση θέσης Τεταρτημορίων & Τιμής

Θέση

$$i * \frac{(n + 1)}{4}, i = 1,2,3$$

Τιμή

$$Q_i = X_{(A_Q)} + \Delta_Q [X_{(A_Q+1)} - X_{(A_Q)}]$$

A_Q : το ακέραιο μέρος του $i * \frac{n+1}{4}$

Δ_Q : το δεκαδικό μέρος του $i * \frac{n+1}{4}$

1^ο τεταρτημόριο:
i=1

$$i * \frac{(n + 1)}{4} = 1 * \frac{21}{4} = 5,25$$

A_Q : το ακέραιο μέρος του 5, 25 = 5

Δ_Q : το δεκαδικό μέρος του 5, 25 = 0, 25

$$X_{(5)} = 1$$

$$X_{(5+1)} = 1$$

$$Q_1 = 1 + 0,25 * (1 - 1) = 1$$

Εύρεση Τεταρτημορίων → 1^ο Βήμα: Ταξινόμηση δεδομένων κατά αύξουσα σειρά

0	0	0	1	1	1	1	1	1	1	1	2	2	2	2	2	2	3	4	4
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

ΥΠΟΛΟΓΙΣΜΟΙ όπως το
ΒΙΒΛΙΟ

2^ο Βήμα: Εύρεση θέσης Τεταρτημορίων & Τιμής

Θέση

$$i * \frac{(n + 1)}{4}, i = 1,2,3$$

Τιμή

$$Q_i = X_{(A_Q)} + \Delta_Q [X_{(A_Q+1)} - X_{(A_Q)}]$$

A_Q : το ακέραιο μέρος του $i * \frac{n+1}{4}$ Δ_Q : το δεκαδικό μέρος του $i * \frac{n+1}{4}$

$$i * \frac{(n + 1)}{4} = 3 * \frac{21}{4} = 15,75$$

3^ο τεταρτημόριο:
 $i=3$

A_Q : το ακέραιο μέρος του 15, 75 = 15

Δ_Q : το δεκαδικό μέρος του 15, 75 = 0, 75

$$X_{(15)} = 2$$

$$X_{(15+1)} = 2$$

$$Q_3 = 2 + 0,75 * (2 - 2) = 2$$

0	0	0	1	1	1	1	1	1	1	1	2	2	2	2	2	2	3	4	4
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

ΥΠΟΛΟΓΙΣΜΟΙ όπως το
ΒΙΒΛΙΟ

2^ο Βήμα: Εύρεση θέσης Τεταρτημορίων & Τιμής

Θέση

$$i * \frac{(n + 1)}{4}, i = 1,2,3$$

Τιμή

$$Q_i = X_{(A_Q)} + \Delta_Q [X_{(A_Q+1)} - X_{(A_Q)}]$$

A_Q : το ακέραιο μέρος του $i * \frac{n+1}{4}$ Δ_Q : το δεκαδικό μέρος του $i * \frac{n+1}{4}$

$$i * \frac{(n + 1)}{4} = 2 * \frac{21}{4} = 10,5$$

2^ο τεταρτημόριο:
 $i=2$

A_Q : το ακέραιο μέρος του 10, 5 = 10

Δ_Q : το δεκαδικό μέρος του 10, 5 = 0, 5

$X_{(10)} = 1$

$X_{(10+1)} = 2$

$Q_2 = 1 + 0,5 * (2 - 1)=1,5$



Εύρεση Τεταρτημορίων

→ 1^ο Βήμα: Ταξινόμηση δεδομένων κατά αύξουσα σειρά

20 παρατηρήσεις

0	0	0	1	1	1	1	1	1	1	2	2	2	2	2	2	3	4	4
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	20

ΥΠΟΛΟΓΙΣΜΟΙ όπως το
ΒΙΒΛΙΟ

2^ο Βήμα: Εύρεση θέσης Τεταρτημορίων

Θέση

Τιμή

1^ο τεταρτημόριο

$$i * \frac{(n + 1)}{4} = 1 * \frac{21}{4} = 5,25$$

→

$$Q_1 = 1 + 0,25 * (1 - 1) = 1$$

3^ο τεταρτημόριο

$$i * \frac{(n + 1)}{4} = 3 * \frac{21}{4} = 15,75$$

→

$$Q_3 = 2 + 0,75 * (2 - 2) = 2$$

2^ο τεταρτημόριο -
Διάμεσος:

$$i * \frac{(n + 1)}{4} = 2 * \frac{21}{4} = 10,5$$

→

$$Q_2 = 1 + 0,5 * (2 - 1) = 1,5$$

Σύνοψη

Minimum	0
Q1	1
Median	1,5
Q3	2
Maximum	4

Εύρεση Τεταρτημορίων

→ 1^ο Βήμα: Ταξινόμηση δεδομένων κατά αύξουσα σειρά

20 παρατηρήσεις

0	0	0	1	1	1	1	1	1	1	2	2	2	2	2	2	3	4	4	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

ΆΛΛΟΣ ΤΡΟΠΟΣ
ΥΠΟΛΟΓΙΣΜΩΝ

2^ο Βήμα: Εύρεση θέσης Τεταρτημορίων

Θέση

Τιμή

1^ο τεταρτημόριο:

$$\frac{n+1}{4} = \frac{21}{4} = 5,25$$

Στρογγυλοποιώ προς τα κάτω

$Q_1 = 1$

3^ο τεταρτημόριο

$$\frac{3 * (n+1)}{4} = \frac{3 * 21}{4} = 15,75$$

Στρογγυλοποιώ προς τα πάνω

$Q_3 = 2$

2^ο τεταρτημόριο -
Διάμεσος:

Ο μέσος όρος των δύο γειτονικών παρατηρήσεων

$$\frac{n+1}{2} = \frac{21}{2} = 10,5$$

→

$$Q_2 = \frac{1+2}{2} = 1,5$$

Σύνοψη

Minimum	0
Q1	1
Median	1,5
Q3	2
Maximum	4

0	0	0	1	1	1	1	1	1	1	2	2	2	2	2	2	2	3	4	4
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

20 παρατηρήσεις

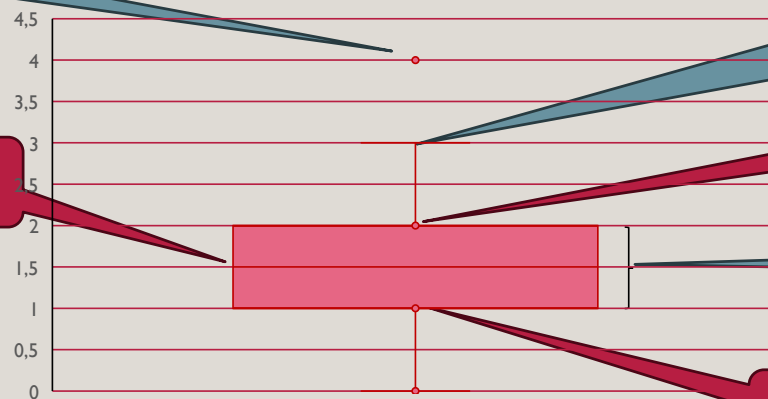
Ένα σημείο δεδομένων θεωρείται **ακραίο** εάν υπερβαίνει μια απόσταση 1,5 φορές το IQR κάτω από το 1ο τεταρτημόριο ($Q_1 - 1,5 * IQR = 1 - 1,5 * 1 = -0,5$) ή 1,5 φορές το IQR πάνω από το 3ο τεταρτημόριο ($Q_3 + 1,5 * IQR = 2 + 1,5 * 1 = 3,5$).

Σύνοψη 5 αριθμών

Minimum	0
Q1	1
Median	1,5
Q3	2
Maximum	4

Ακραία τιμή

Box-Plot



$Q_2=1,5$

το ανώτερο στέλεχος εκτείνεται στη μεγαλύτερη τιμή (3) αγνοώντας την τιμή 4

$Q_3=2$

$IQR=Q_3-Q_1=1$

$Q_1=1$

Minimum=0

0	0	0	1	1	1	1	1	1	1	2	2	2	2	2	2	3	4	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

ΥΠΟΛΟΓΙΣΜΟΙ όπως το
ΒΙΒΛΙΟ

2^ο Βήμα: Εύρεση θέσης Τεταρτημορίων

Θέση

Τιμή

1^ο τεταρτημόριο:

$$i * \frac{(n + 1)}{4} = 1 * \frac{20}{4} = 5 \longrightarrow Q_1 = 1 + 0 * (1 - 1) = 1$$

3^ο τεταρτημόριο

$$i * \frac{(n + 1)}{4} = 3 * \frac{20}{4} = 15 \longrightarrow Q_3 = 2 + 0 * (2 - 2) = 2$$

2^ο τεταρτημόριο -
Διάμεσος:

$$i * \frac{(n + 1)}{4} = 2 * \frac{20}{4} = 10 \longrightarrow Q_2 = 1 + 0 * (2 - 1) = 1$$

Σύνοψη

Minimum	0
Q1	1
Median	1
Q3	2
Maximum	4

Εύρεση Τεταρτημορίων

→ 1^ο Βήμα: Ταξινόμηση δεδομένων κατά αύξουσα σειρά

19 παρατηρήσεις

0	0	0	1	1	1	1	1	1	1	2	2	2	2	2	2	2	3	4
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

ΆΛΛΟΣ ΤΡΟΠΟΣ
ΥΠΟΛΟΓΙΣΜΩΝ

2^ο Βήμα: Εύρεση θέσης Τεταρτημορίων

Θέση

Τιμή

1^ο τεταρτημόριο:

$$\frac{n + 1}{4} = \frac{19 + 1}{4} = \frac{20}{4} = 5 \longrightarrow Q_1 = 1$$

3^ο τεταρτημόριο

$$\frac{3(n + 1)}{4} = \frac{3 * 20}{4} = \frac{60}{4} = 15 \longrightarrow Q_3 = 2$$

2^ο τεταρτημόριο -
Διάμεσος:

$$\frac{n + 1}{2} = \frac{20}{2} = 10 \longrightarrow Q_2 = 1$$

Σύνοψη

Minimum	0
Q1	1
Median	1
Q3	2
Maximum	4

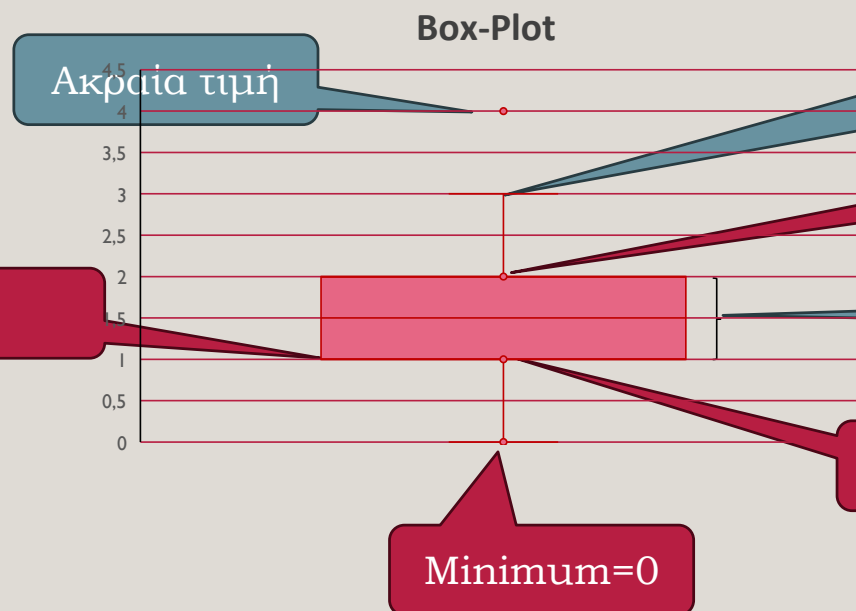
0	0	0	1	1	1	1	1	1	1	2	2	2	2	2	2	2	3	4
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

19 παρατηρήσεις

Σύνοψη 5
αριθμών

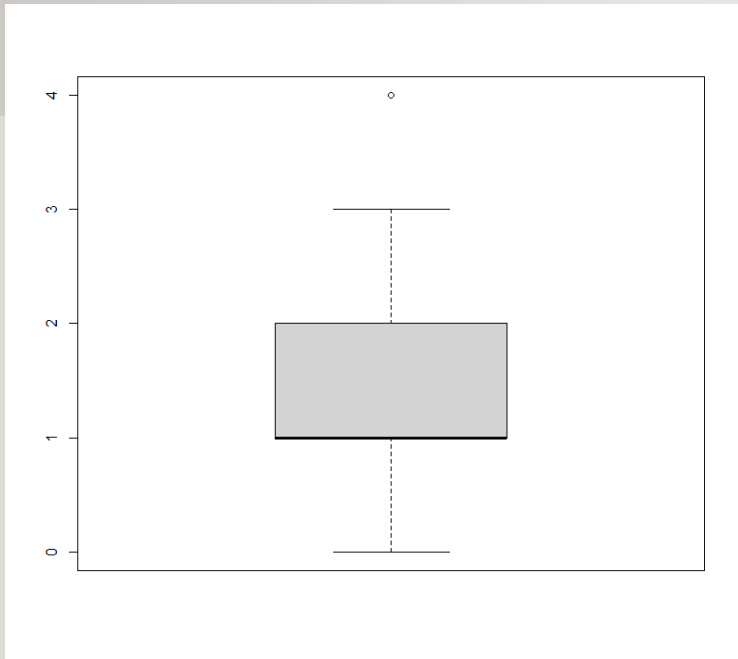
Ένα σημείο δεδομένων θεωρείται **ακραίο** εάν υπερβαίνει μια απόσταση 1,5 φορές το IQR **κάτω** από το 1ο τεταρτημόριο : ($Q_1 - 1,5 * IQR = 1 - 1,5 * 1 = -0,5$) ή 1,5 φορές το IQR **πάνω** από το 3ο τεταρτημόριο : ($Q_3 + 1,5 * IQR = 2 + 1,5 * 1 = 3,5$).

το ανώτερο στέλεχος
εκτείνεται στη
μεγαλύτερη τιμή (3)
αγνοώντας την τιμή 4

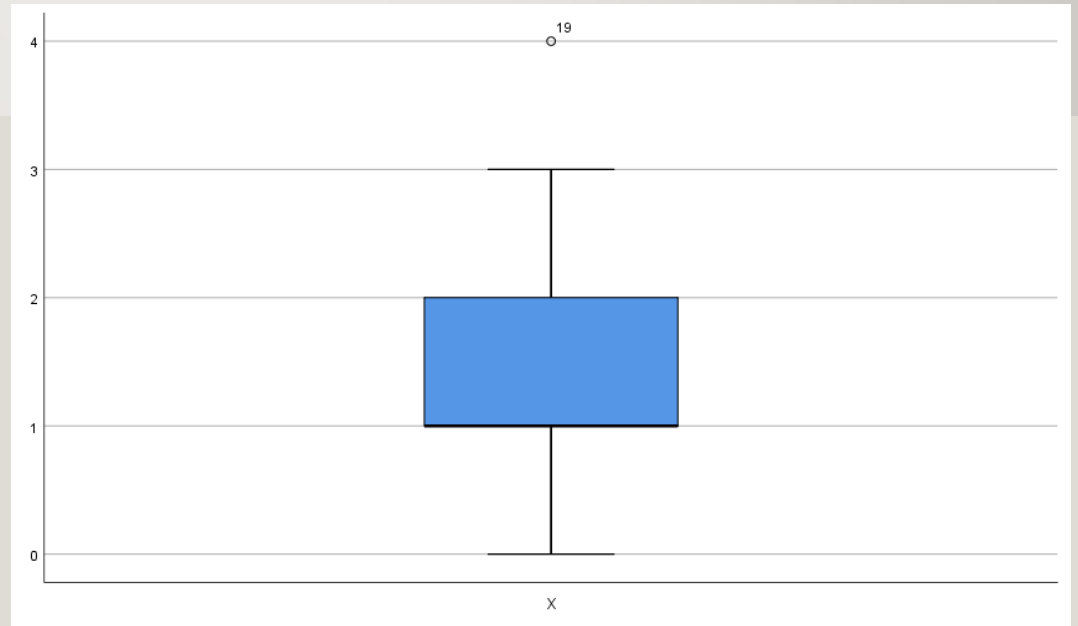


Minimum	0
Q1	1
Median	1
Q3	2
Maximum	4

0	0	0	1	1	1	1	1	1	1	2	2	2	2	2	2	2	3	4
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19



Box-plot από R



Box-plot από SPSS

20 παρατηρήσεις
Αλλαγή σε 3

Εύρεση **Τεταρτημορίων** → 1^ο Βήμα: Ταξινόμηση δεδομένων κατά αύξουσα σειρά

0	0	0	1	1	1	1	1	1	1	2	2	2	2	2	2	3	3	4	4
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

ΥΠΟΛΟΓΙΣΜΟΙ όπως το
ΒΙΒΛΙΟ

2^ο Βήμα: Εύρεση θέσης Τεταρτημορίων

$$i * \frac{(n + 1)}{4}, i = 1, 2, 3$$

Θέση

Τιμή

1^ο τεταρτημόριο:

$$i * \frac{(n + 1)}{4} = 1 * \frac{21}{4} = 5,25 \longrightarrow Q_1 = 1 + 0,25 * (1 - 1) = 1$$

3^ο τεταρτημόριο

$$i * \frac{(n + 1)}{4} = 3 * \frac{21}{4} = 15,75 \longrightarrow Q_3 = 2 + 0,75 * (3 - 2) = 2,75$$

2^ο τεταρτημόριο -
Διάμεσος:

$$i * \frac{(n + 1)}{4} = 2 * \frac{21}{4} = 10,5 \longrightarrow Q_2 = 1 + 0,5 * (2 - 1) = 1,5$$

Σύνοψη

Minimum	0
Q1	1
Median	1,5
Q3	2,75
Maximum	4

Εύρεση Τεταρτημορίων → 1^ο Βήμα: Ταξινόμηση δεδομένων κατά αύξουσα σειρά

20 παρατηρήσεις
Αλλαγή σε 3

0	0	0	1	1	1	1	1	1	1	1	2	2	2	2	2	3	3	3	4	4
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	

ΆΛΛΟΣ ΤΡΟΠΟΣ
ΥΠΟΛΟΓΙΣΜΩΝ

2^ο Βήμα: Εύρεση θέσης Τεταρτημορίων

Θέση

Τιμή

1^ο τεταρτημόριο:

$$\frac{n+1}{4} = \frac{21}{4} = 5,25$$

Στρογγυλοποιώ προς τα κάτω

→

$Q_1 = 1$

3^ο τεταρτημόριο

$$\frac{3(n+1)}{4} = \frac{3 \cdot 21}{4} = \frac{63}{4} = 15,75$$

Στρογγυλοποιώ προς τα πάνω

→

$Q_3 = 3$

2^ο τεταρτημόριο -
Διάμεσος:

Ο μέσος όρος των δύο γειτονικών παρατηρήσεων

$$\frac{n+1}{2} = \frac{21}{2} = 10,5$$

→

$Q_2 = \frac{1+2}{2} = 1,5$

Σύνοψη

Minimum	0
Q1	1
Median	1,5
Q3	3
Maximum	4

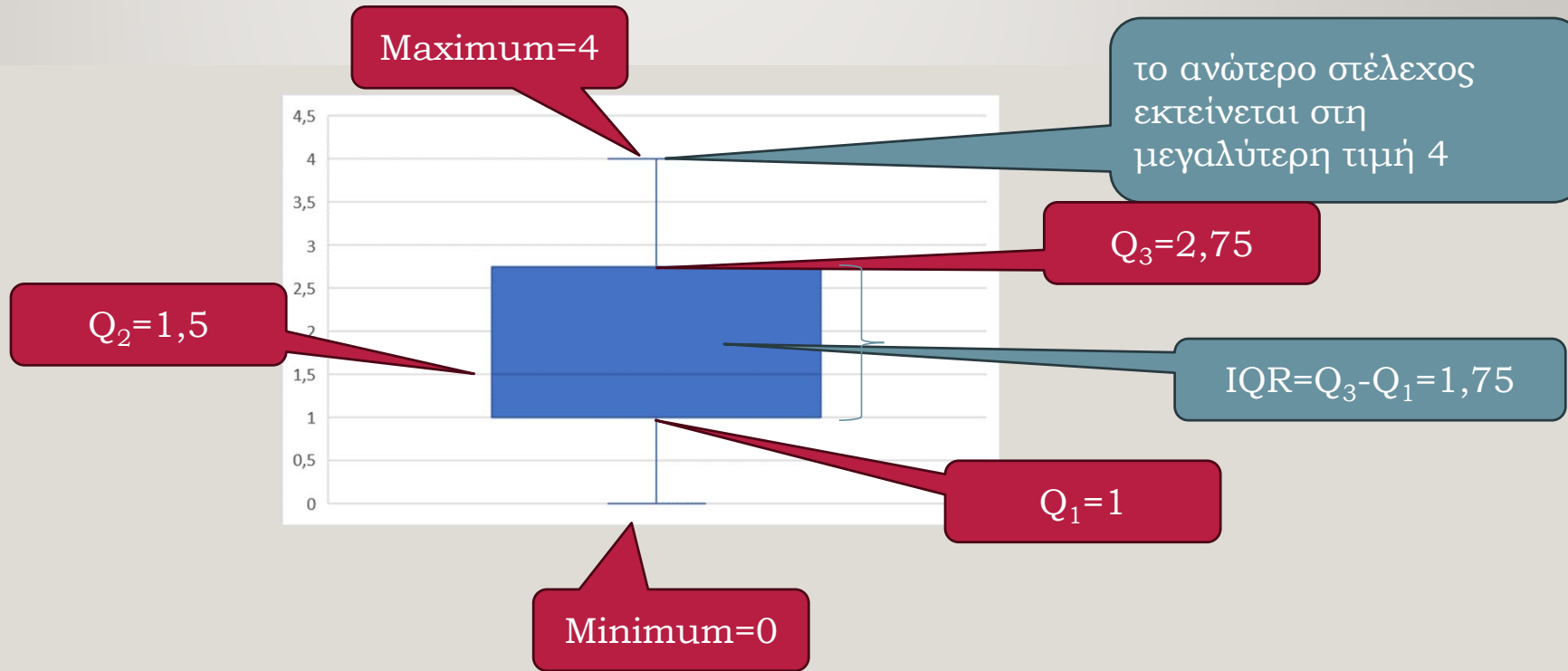
0	0	0	1	1	1	1	1	1	1	2	2	2	2	2	3	3	3	4	4
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Ένα σημείο δεδομένων θεωρείται **ακραίο** εάν υπερβαίνει μια απόσταση 1,5 φορές το IQR κάτω από το 1ο τεταρτημόριο ($Q_1 - 1,5 * IQR = 1 - 1,5 * 1,75 = -1,625$) ή 1,5 φορές το IQR πάνω από το 3ο τεταρτημόριο ($Q_3 + 1,5 * IQR = 2,25 + 1,5 * 1,75 = 4,9$).

20 παρατηρήσεις
Αλλαγή σε 3

Σύνοψη 5 αριθμών

Minimum	0
Q1	1
Median	1,5
Q3	2,75
Maximum	4

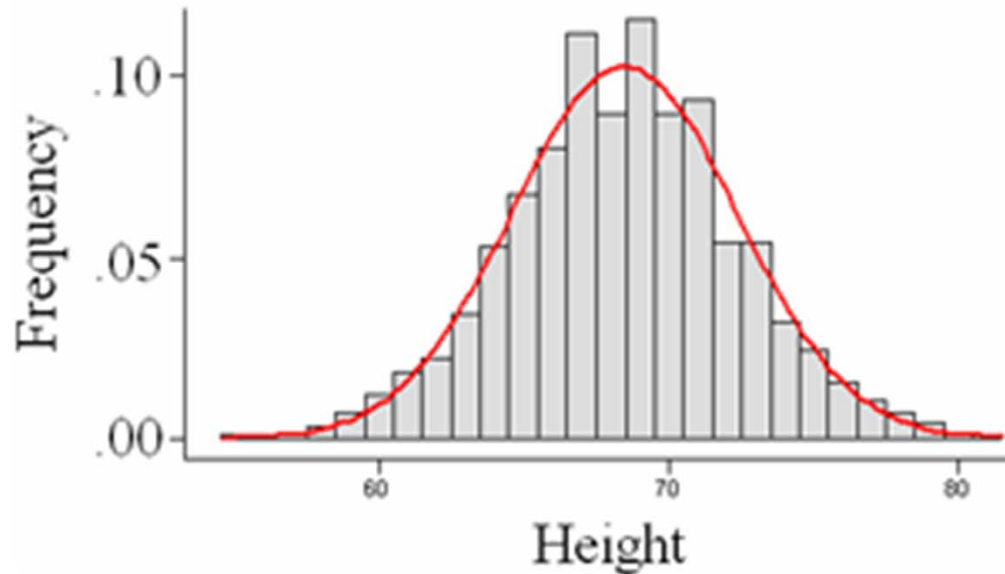


Μορφή Της Κατανομής

- Εάν υποθέσουμε ότι ο αριθμός των διαστημάτων τιμών (τάξεων) για μια αριθμητική μεταβλητή είναι αρκετά μεγάλος (τείνει στο άπειρο) και ότι το πλάτος των τάξεων είναι αρκετά μικρό (τείνει στο μηδέν), τότε η πολυγωνική γραμμή συχνοτήτων τείνει να πάρει τη μορφή μιας **ομαλής καμπύλης**, η οποία ονομάζεται **καμπύλη συχνοτήτων** (frequency curve)

Μορφή Της Κατανομής

Η μορφή μιας κατανομής συχνοτήτων εξαρτάται από το πώς είναι κατανεμημένες οι παρατηρήσεις σε όλη την έκταση του εύρους τους.



Μέτρα Περιγραφής Της Μορφής Της Κατανομής

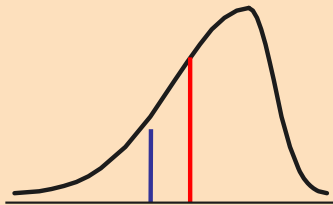
- Περιγράφουν τον τρόπο με τον οποίο κατανέμονται οι παρατηρήσεις
- Τα δυο χρήσιμα στατιστικά μέτρα είναι:
 - Η **ασυμμετρία** (*skewness*)
 - Μετρά τον βαθμό κατά τον οποίο οι τιμές **δεν** είναι συμμετρικές γύρω από την μέση τιμή
 - Η **κύρτωση** (*kurtosis*)
 - Η κύρτωση επηρεάζει την κύρτωση της καμπύλης κατανομής- δηλ. το κατά πόσο η καμπύλη ανεβαίνει απότομα όσο πλησιάζουμε στο κέντρο της κατανομής.

Μορφή Μιας Κατανομής (Ασυμμετρία)

Μετρά τον βαθμό κατά τον οποίο οι τιμές **δεν** είναι συμμετρικές γύρω από την μέση τιμή

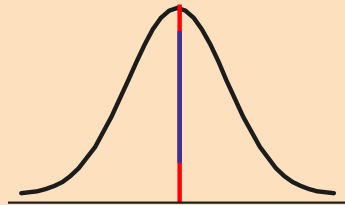
Ασύμμετρη προς τα αριστερά
Ουρά τιμών προς τα αριστερά

Μέση τιμή < Διάμεσος



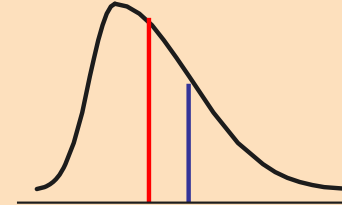
Συμμετρική

Μέση τιμή = Διάμεσος



Ασύμμετρη προς τα δεξιά
Ουρά τιμών προς τα δεξιά

Διάμεσος < Μέση τιμή



Στατιστική
Ασυμμετρία

< 0

0

> 0

Παρακάτω παρουσιάζονται τα ιστογράμματα για το κόστος της ηλεκτρικής ενέργειας (σε δολάρια) του μηνός Ιουλίου 2013, για δύο τυχαία δείγματα 50 διαμερισμάτων ενός δωματίου σε δύο μεγάλες πόλεις.



Μπορείτε να περιγράψετε το σχήμα της κατανομής του κόστους ηλεκτρικής ενέργειας;

Παρακάτω παρουσιάζονται τα ιστογράμματα για το κόστος της ηλεκτρικής ενέργειας (σε δολάρια) του μηνός Ιουλίου 2013, για δύο τυχαία δείγματα 50 διαμερισμάτων ενός δωματίου σε δύο μεγάλες πόλεις.



Μπορείτε να περιγράψετε το σχήμα της κατανομής του κόστους ηλεκτρικής ενέργειας;

Λοξή δεξιά

Λίγα διαμερίσματα έχουν **μεγάλη** κατανάλωση ρεύματος

Παρακάτω παρουσιάζονται τα ιστογράμματα για το κόστος της ηλεκτρικής ενέργειας (σε δολάρια) του μηνός Ιουλίου 2013, για δύο τυχαία δείγματα 50 διαμερισμάτων ενός δωματίου σε δύο μεγάλες πόλεις.



Μπορείτε να περιγράψετε το σχήμα της κατανομής του κόστους ηλεκτρικής ενέργειας;

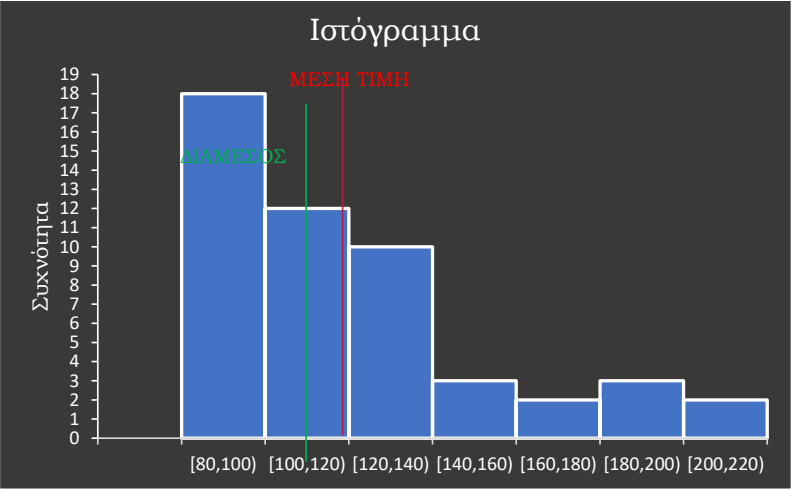
Λοξή αριστερά

Λίγα διαμερίσματα έχουν **μικρή** κατανάλωση ρεύματος

126,10	121,39	120,51	153,14	133,53
83,07	93,43	145,82	155,15	138,70
111,84	185,97	89,85	92,23	138,45
118,13	130,59	93,85	94,36	136,56
115,04	173,87	106,43	109,71	109,63
105,16	177,46	86,18	84,01	88,31
119,15	179,62	119,71	95,80	98,31
185,58	85,69	80,12	102,97	103,73
218,36	212,76	87,32	98,80	109,60
89,29	80,04	86,62	102,01	103,81

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = 119,73$$

Minimum	80,04
Q1	93,43
Median	109,6
Q3	136,56
Maximum	218,36



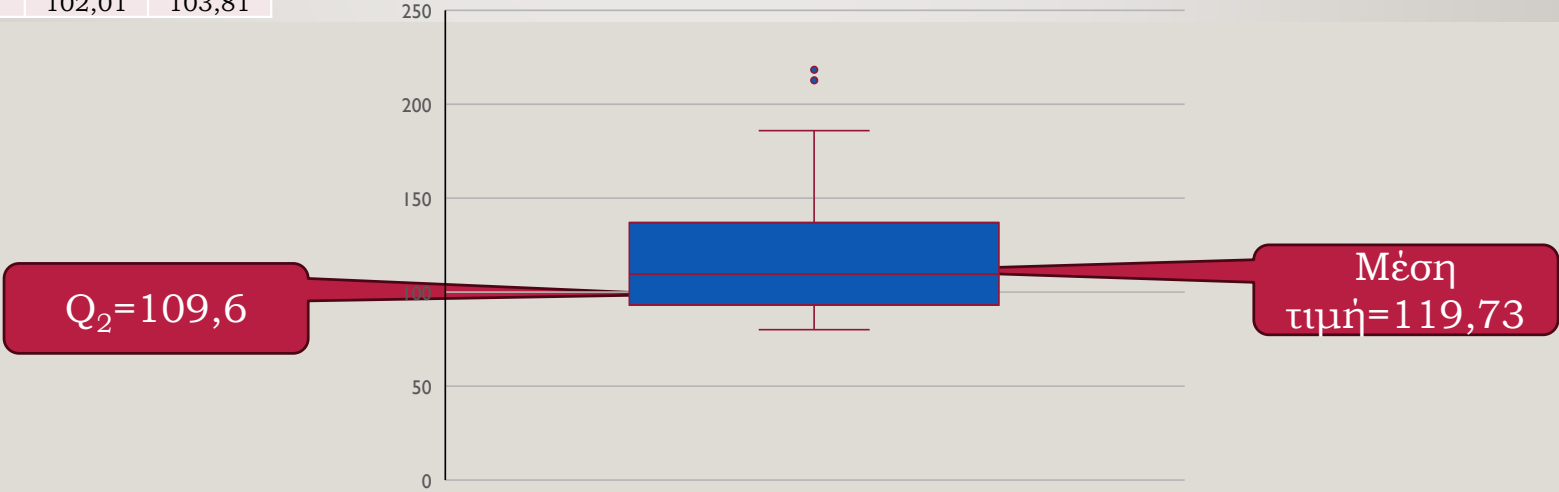
Διάμεσος < Μέση τιμή

Ασύμμετρη προς τα δεξιά
Ουρά τιμών προς τα δεξιά

126,10	121,39	120,51	153,14	133,53
83,07	93,43	145,82	155,15	138,70
111,84	185,97	89,85	92,23	138,45
118,13	130,59	93,85	94,36	136,56
115,04	173,87	106,43	109,71	109,63
105,16	177,46	86,18	84,01	88,31
119,15	179,62	119,71	95,80	98,31
185,58	85,69	80,12	102,97	103,73
218,36	212,76	87,32	98,80	109,60
89,29	80,04	86,62	102,01	103,81

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = 119,73$$

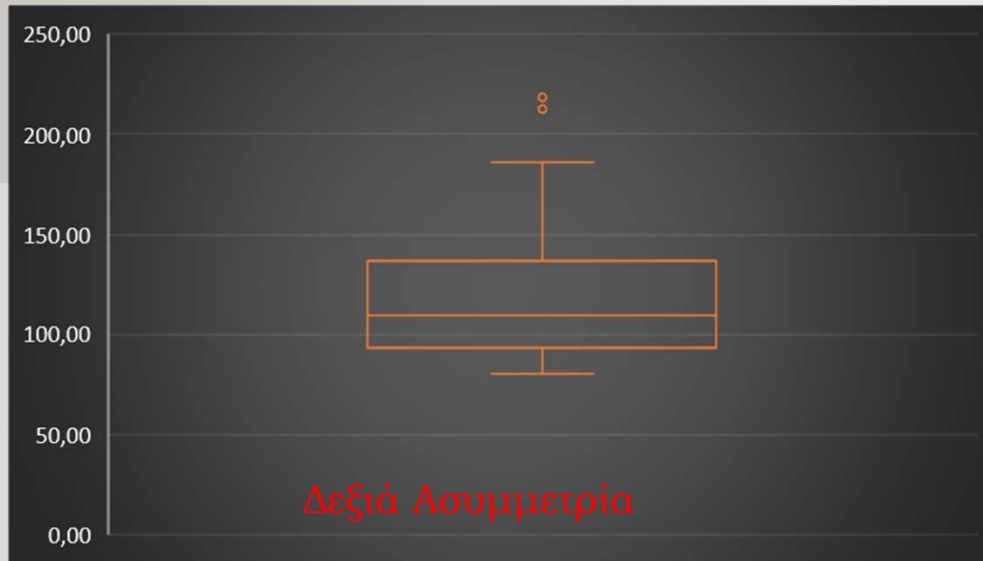
Minimum	80,04
Q1	93,43
Median	109,6
Q3	136,56
Maximum	218,36



$$\lambda = \frac{E(X - \bar{X})^4}{S^3} = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^3}{S^3} = 1,213 > 0$$

Δεξιά Ασυμμετρία

Ποιο διάγραμμα πλαισίου-απολήξεων αντιπροσωπεύει δεδομένα με δεξιά ασυμμετρία και ποιο με αριστερή?

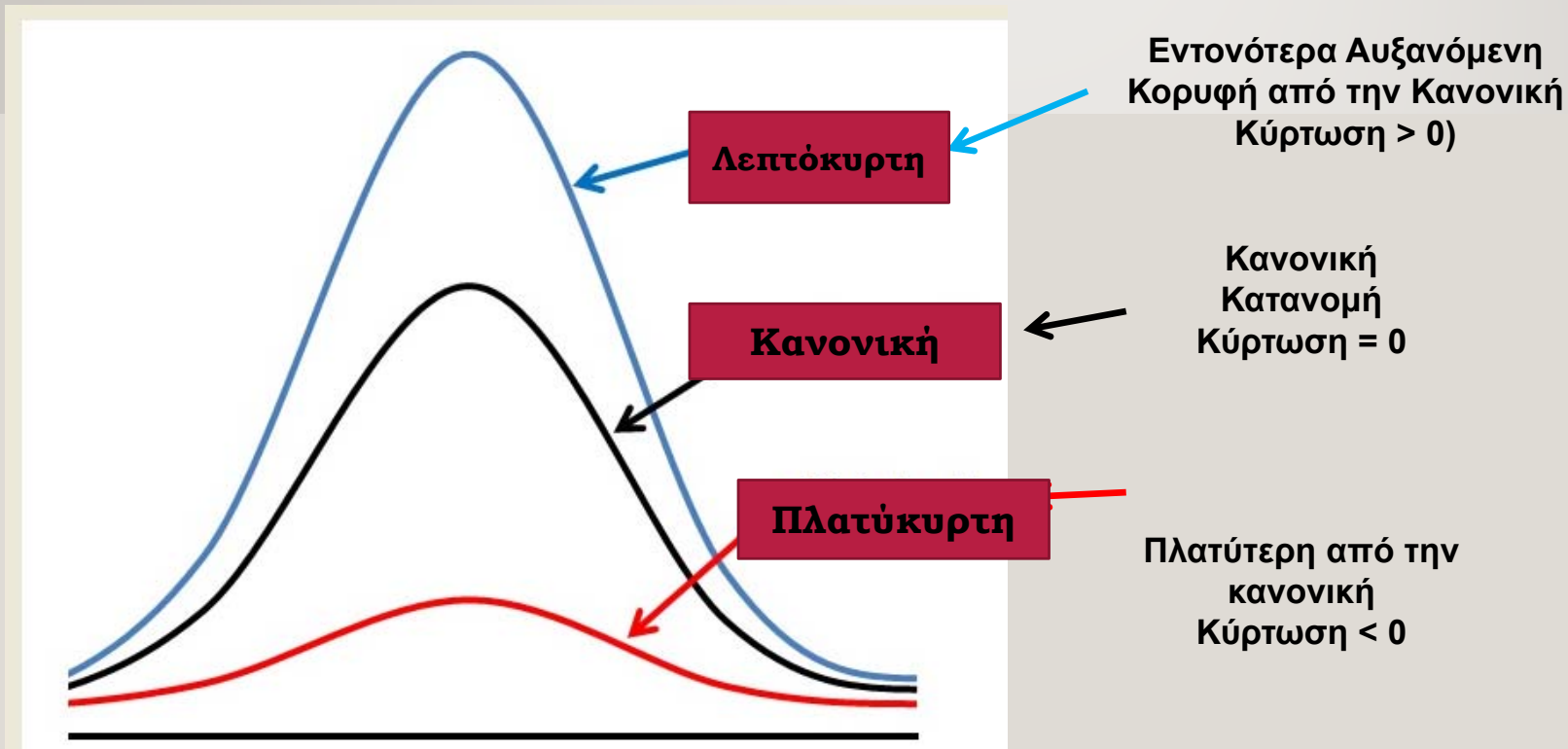


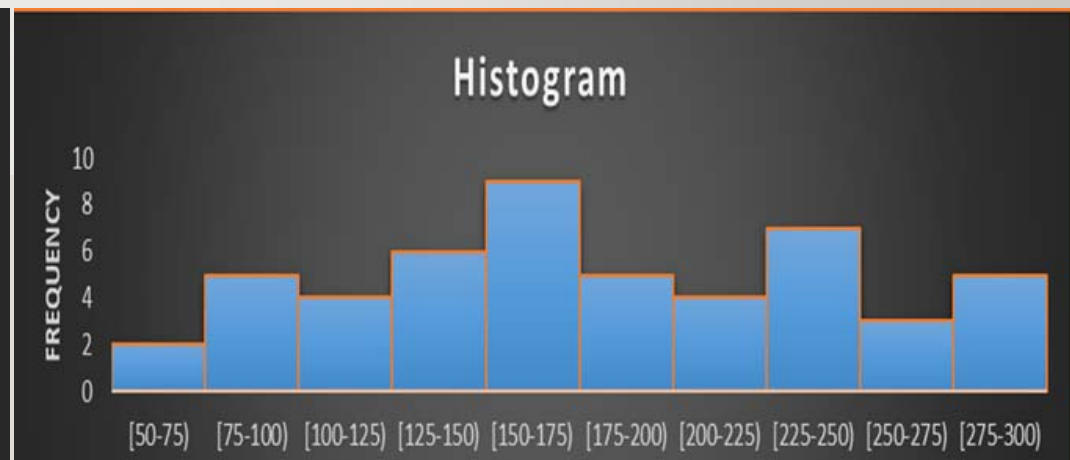
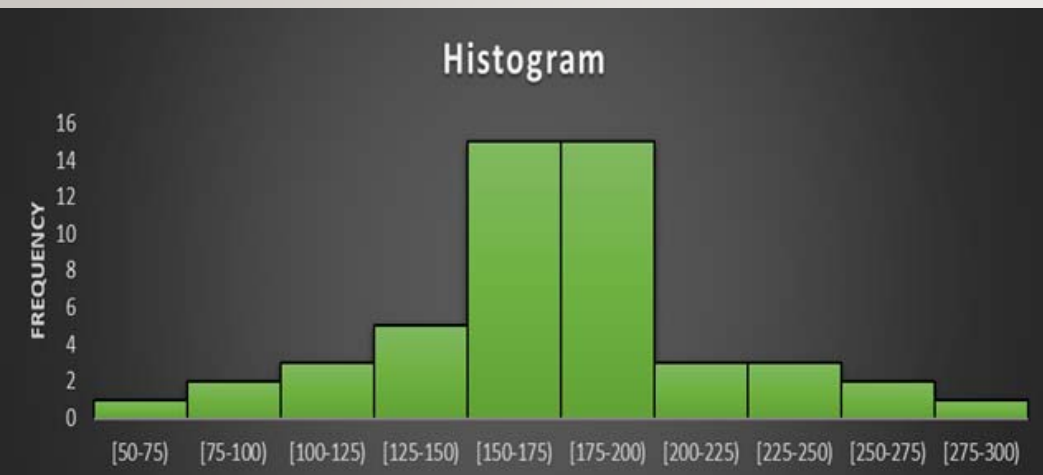
- Η διάμεσος είναι πιο κοντά στο **Q1**.
- Το διάστημα Q1 έως τη διάμεσο είναι μικρότερο από το διάστημα από τη διάμεσο έως το Q3.
- Αυτό σημαίνει ότι οι περισσότερες τιμές βρίσκονται στο **κάτω μέρος** της κατανομής και υπάρχουν λίγες ακραίες τιμές στα δεξιά (μεγάλες τιμές).



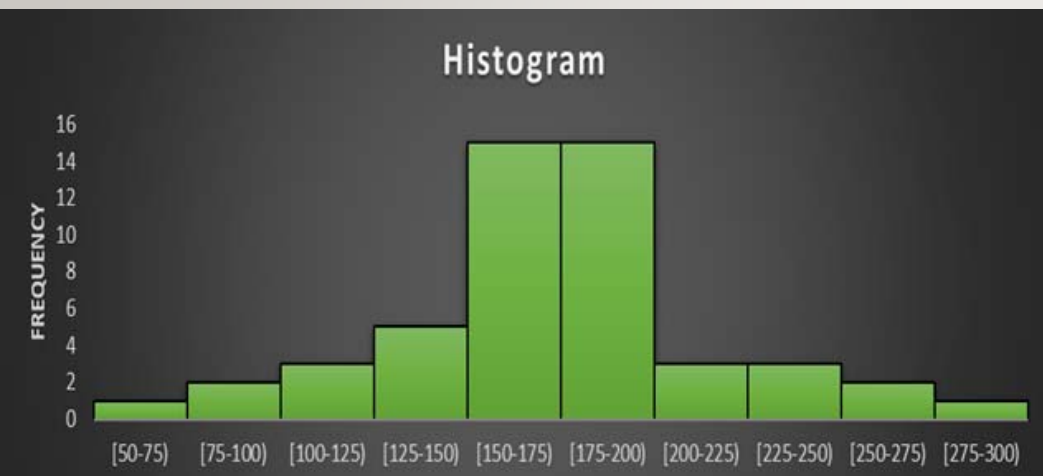
- Η διάμεσος είναι πιο κοντά στο **Q3**.
- Το διάστημα Q1 έως τη διάμεσο είναι μεγαλύτερο από το διάστημα από τη διάμεσο έως το Q3.
- Αυτό σημαίνει ότι οι περισσότερες τιμές βρίσκονται στο **πάνω μέρος** της κατανομής και υπάρχουν λίγες ακραίες τιμές στα αριστερά (μικρές τιμές).

Μορφή Κατανομής: Κύρτωση, Η Καμπύλη Ανεβαίνει Όσο Πλησιάζουμε Στο Κέντρο Της Κατανομής





Περιγράψτε τη μορφή της κατανομής ως προς την κύρτωσή της

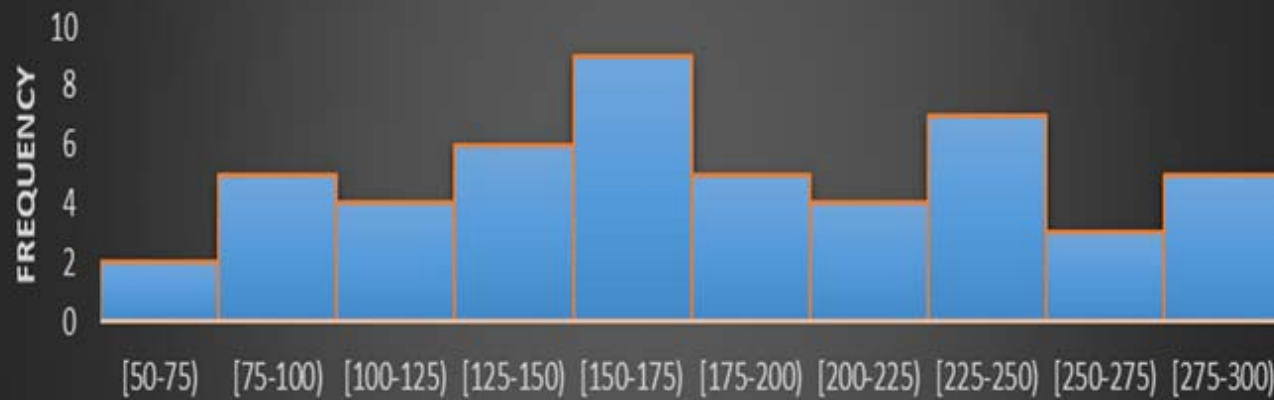


Λεπτόκυρτη

$$\kappa = \frac{E(X - \bar{X})^4}{S^4} = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4}{S^4} = 4,20 > 3$$

$$\kappa' = \frac{E(X - \bar{X})^4}{S^4} - 3 = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4}{S^4} - 3 = 1,20 > 0$$

Histogram

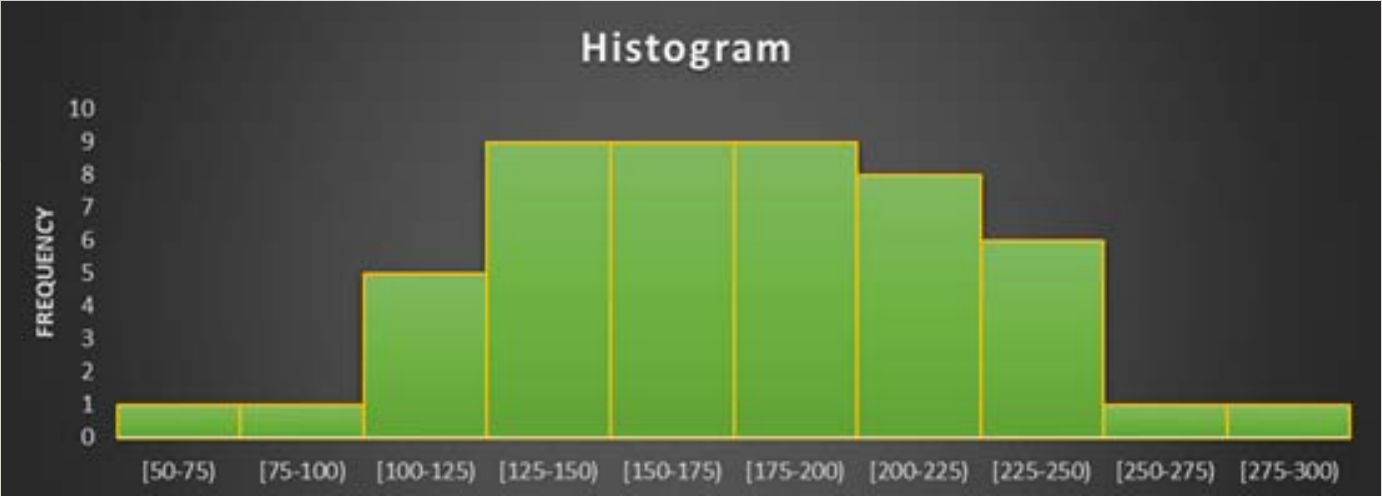


$$\kappa = \frac{E(X - \bar{X})^4}{S^4} = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4}{S^4} = 2,13 < 3$$

$$\kappa' = \frac{E(X - \bar{X})^4}{S^4} - 3 = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4}{S^4} - 3 = -0,87 < 0$$

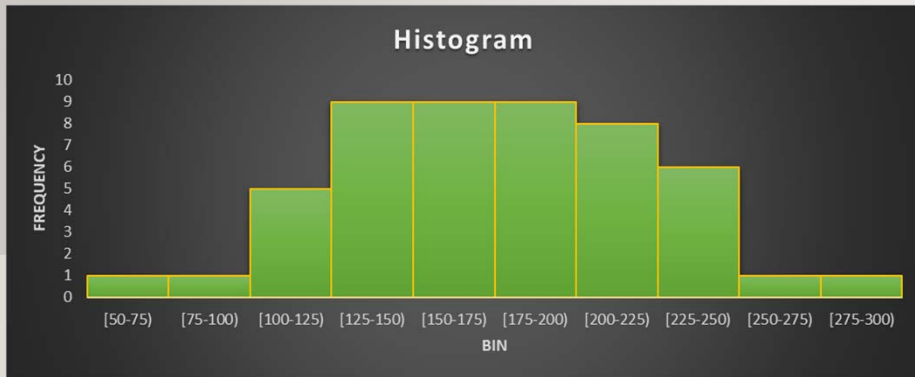
↓
Πλατύκυρτη

Παρακάτω παρουσιάζεται τα ιστογράμματα για το ετήσιο εισόδημα (σε δολάρια) του μηνός , για ένα τυχαίο δείγμα 50 στελεχών από διάφορες πολυεθνικές εταιρίες.



- A. Τι περιμένω για τη λοξότητα;
- B. Τι περιμένω για την κύρτωση;
- Γ. Ποια είναι η σχέση μέσης τιμής και διαμέσου;

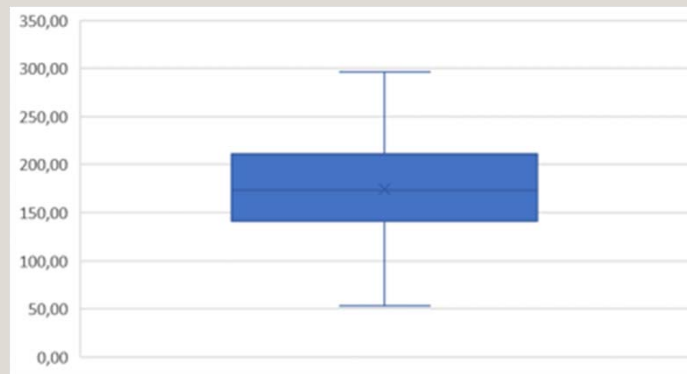
Λύση



Κλάση	Συχνότητα
[50-75)	1
[75-100)	1
[100-125)	5
[125-150)	9
[150-175)	9
[175-200)	9
[200-225)	8
[225-250)	6
[250-275)	1
[275-300)	1

Minimum	53,00
Q1	143,65
Median	173,62
Q3	210,25
Maximum	296,00
St. Deviation	49,01
Average	174,06
Range	243,00
skew	-0,02
kurtosis	2,99
excess kurtosis	-0,01

Παρατηρούμε ότι η μέση τιμή είναι σχεδόν ίση με τη διάμεσο, ενώ η κύρτωση είναι σχεδόν ίση με 3 και η ασυμμετρία σχεδόν ίση με το μηδέν



Οι Τιμές Z (Z-score)

Εντοπισμός Ακραίων Τιμών

- Για να υπολογίσετε το Z-Score μίας συγκεκριμένης τιμής **αφαιρείτε** τη μέση τιμή από την τιμή και **διαίρεíte** με την **τυπική απόκλιση**.
- Το Z-score είναι ένα αριθμός που δείχνει **πόσες τυπικές αποκλίσεις** απέχει μια τιμή από τη μέση τιμή.
- Μια τιμή θεωρείται ως ακραία αν οι τιμές Z είναι μικρότερες από -3.0 ή μεγαλύτερες από +3.0.
- Όσο μεγαλύτερη είναι η απόλυτη τιμή του Z-score, τόσο μεγαλύτερη η απόσταση της τιμής από τη μέση τιμή.

Οι Τιμές Z (Z-Score)

$$Z = \frac{X - \mu}{\sigma}$$

Όπου X η τιμή της μεταβλητής
μ η μέση τιμή του πληθυσμού
σ η τυπική απόκλιση του πληθυσμού

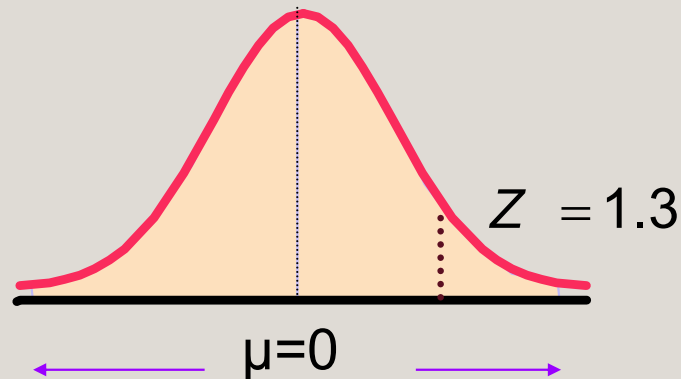
Οι Τιμές Z (Z-Score)

- Έστω ότι η μέση βαθμολογία στην εξέταση των μαθηματικών είναι 490, με τυπική απόκλιση 100.
- Να υπολογιστεί το Z-Score για μια βαθμολογία 620.

$$Z = \frac{X - \mu}{\sigma} = \frac{620 - 490}{100} = \frac{130}{100} = 1.3$$

Οι Τιμές Z (Z-Score)

Έστω ότι η μέση βαθμολογία στην εξέταση των μαθηματικών είναι 490, με τυπική απόκλιση 100.



Μια βαθμολογία 620 είναι 1.3 **τυπικές αποκλίσεις πάνω από τη μέση τιμή**
Δεν μπορεί να θεωρηθεί ως ακραία.

Οι Τιμές Z (Z-score)

- Μία φοιτήτρια στην εξέταση στο μάθημα της Στατιστικής βαθμολογήθηκε με 86 ενώ στο μάθημα των Μαθηματικών με 82.
- Σε ποιο μάθημα είχε καλύτερη επίδοση;
- Σύμφωνα με τα αποτελέσματα των καθηγητών οι βαθμοί στη Στατιστική και τα Μαθηματικά έχουν **κανονική κατανομή** με 80 και 76 και τυπική απόκλιση 6 και 4, αντίστοιχα.

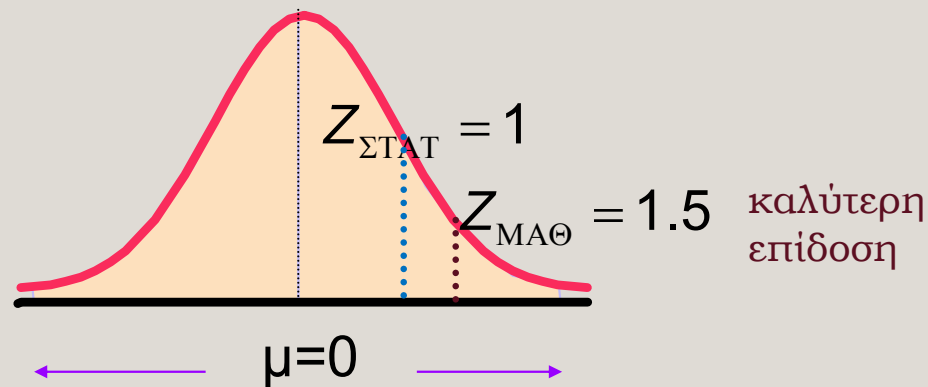
$$Z_{\text{ΣΤΑΤ}} = \frac{X - \mu}{\sigma} = \frac{86 - 80}{6} = \frac{6}{6} = 1$$

τυπικές αποκλίσεις πάνω από τη μέση τιμή

$$Z_{\text{ΜΑΘ}} = \frac{X - \mu}{\sigma} = \frac{82 - 76}{4} = \frac{6}{4} = 1.5$$

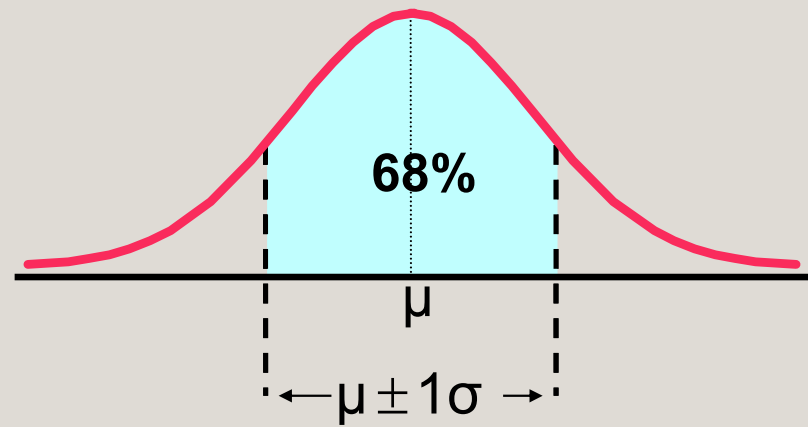
Οι Τιμές Z (Z-score)

- Μία φοιτήτρια στην εξέταση στο μάθημα της Στατιστικής βαθμολογήθηκε με 86 ενώ των **Μαθηματικών** με 82. Σε ποιο μάθημα είχε καλύτερη επίδοση; Σύμφωνα με τα αποτελέσματα των καθηγητών οι βαθμοί στη Στατιστική και τα Μαθηματικά έχουν **κανονική κατανομή** με μέση τιμή 80 και 76 και τυπική απόκλιση 6 και 4, αντίστοιχα.



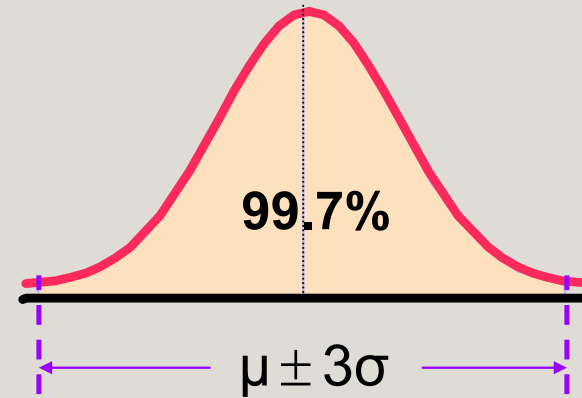
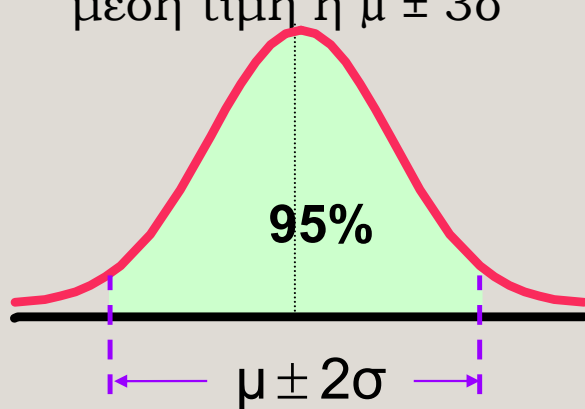
Ο Εμπειρικός Κανόνας

- Ο εμπειρικός κανόνας προσεγγίζει την διακύμανση των δεδομένων σε μια **Κανονική** κατανομή
- Το 68% περίπου των δεδομένων σε μια **Κανονική** κατανομή βρίσκεται εντός 1 τυπικής απόκλισης από τη μέση τιμή ή $\mu \pm 1\sigma$



Ο Εμπειρικός Κανόνας

- Το 95% περίπου των δεδομένων σε μια **Κανονική** κατανομή βρίσκεται εντός 2 τυπικών αποκλίσεων από τη μέση τιμή ή $\mu \pm 2\sigma$
- Το 99,7% περίπου των δεδομένων σε μια **Κανονική** κατανομή βρίσκεται εντός 3 τυπικών αποκλίσεων από τη μέση τιμή ή $\mu \pm 3\sigma$



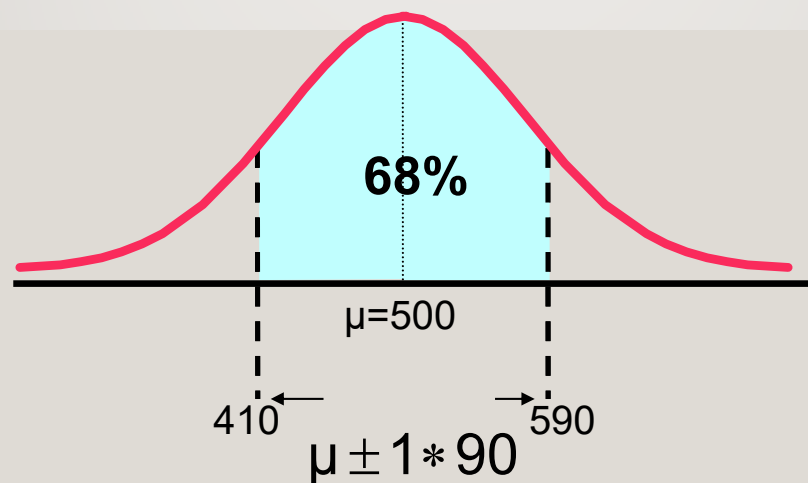
Χρήση Του Εμπειρικού Κανόνα

- Έστω ότι τα αποτελέσματα ενός διαγωνισμού μαθηματικών ακολουθούν **κανονική κατανομή** (bell-shaped), με μέση τιμή 500 και τυπική απόκλιση 90.
- Αυτό σημαίνει ότι:



Χρήση του Εμπειρικού Κανόνα

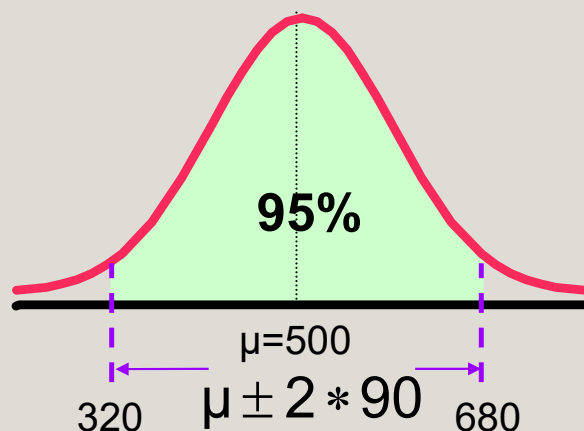
▪ Έστω ότι τα αποτελέσματα ενός διαγωνισμού μαθηματικών ακολουθούν **κανονική κατανομή** (bell-shaped), με μέση τιμή 500 και τυπική απόκλιση 90. Αυτό σημαίνει ότι:



▪ Το **68%** όλων των διαγωνιζομένων βαθμολογήθηκε μεταξύ 410 and 590 (500 ± 90).

Χρήση του Εμπειρικού Κανόνα

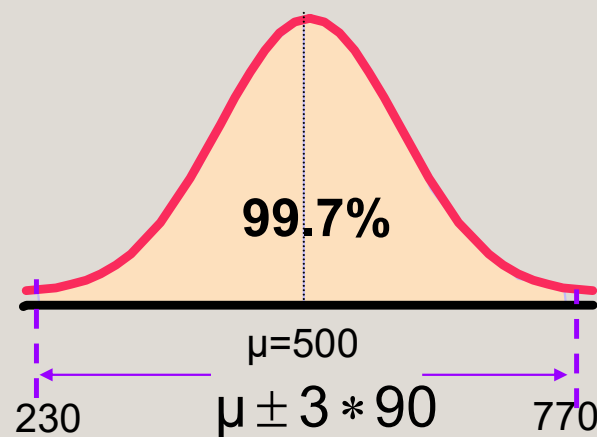
- Έστω ότι τα αποτελέσματα ενός διαγωνισμού μαθηματικών ακολουθούν **κανονική κατανομή** (bell-shaped), με μέση τιμή 500 και τυπική απόκλιση 90. Αυτό σημαίνει ότι:



- Το **95%** όλων των διαγωνιζομένων βαθμολογήθηκε μεταξύ 320 και 680 (500 ± 180).

Χρήση του Εμπειρικού Κανόνα

- Έστω ότι τα αποτελέσματα ενός διαγωνισμού μαθηματικών ακολουθούν **κανονική κατανομή** (bell-shaped), με μέση τιμή 500 και τυπική απόκλιση 90. Αυτό σημαίνει ότι:



- Το **99.7%** όλων των διαγωνιζομένων βαθμολογήθηκε μεταξύ 230 και 770 (500 ± 270).

Χρήση του Εμπειρικού Κανόνα

- Έστω ότι τα αποτελέσματα ενός διαγωνισμού μαθηματικών ακολουθούν **κανονική κατανομή** (bell-shaped), με μέση τιμή 500 και τυπική απόκλιση 90.
 - Τι ποσοστό των διαγωνιζομένων βαθμολογήθηκε μεταξύ 410 και 590
 - $410 = 500 - 1 \cdot 90$ (τυπική απόκλιση)
 - $590 = 500 + 1 \cdot 90$

Το ποσοστό των διαγωνιζομένων που βαθμολογήθηκε μεταξύ 410 και 590, με βάση τον εμπειρικό κανόνα είναι 68%

Ο Κανόνας Chebyshev

- Ο κανόνας ορίζει ότι για **οποιοδήποτε σύνολο δεδομένων** και **ανεξάρτητα** από το σχήμα της κατανομής, τουλάχιστον ένα ποσοστό $(1 - 1/k^2) \times 100\%$ των τιμών συγκεντρώνονται σε απόσταση k τυπικών αποκλίσεων γύρω από τη μέση τιμή (όπου $k > 1$)
- Παραδείγματα:

Τουλάχιστον	Εντός
$(1 - 1/2^2) \times 100\% = 75\%$	$k=2 \quad (\mu \pm 2\sigma)$
$(1 - 1/3^2) \times 100\% = 88.89\%$	$k=3 \quad (\mu \pm 3\sigma)$

Παρακάτω παρουσιάζονται οι τριμηνιαίες εισπράξεις λιανικών πωλήσεων (σε χιλιάδες \$), από το σύνολο των 50 επιχειρήσεων μιας μεγάλης πόλης.

10,3	11,1	9,6	9	14,5	13	6,7	11	8,4	10,3
8	11,2	7,3	5,3	12,5	8	11,8	8,7	10,6	9,5
11,1	10,2	11,1	9,9	9,8	11,6	15,1	12,5	6,5	7,5
10	12,9	9,2	10	12,8	12,5	9,3	10,4	12,7	10,5
9,3	11,5	10,7	11,6	7,8	10,5	7,6	10,1	8,9	8,6

A. Τι ποσοστό των 50 επιχειρήσεων έχει τριμηνιαίες εισπράξεις σε διάστημα ± 1 , ± 2 ή ± 3 τυπικές αποκλίσεις από τη μέση τιμή (οι τιμές ακολουθούν κανονική κατανομή);

B. Συγκρίνετε τα ευρήματά σας με τα ποσοστά που προκύπτουν από τον εμπειρικό κανόνα.



Οι τιμές ακολουθούν **κανονική** κατανομή

Για να υπολογίσουμε πόσες τυπικές αποκλίσεις απέχει μία παρατήρηση από τη μέση τιμή του δείγματος, χρησιμοποιούμε τα z-scores.

$$Z = \frac{X - \bar{X}}{S}$$

Συνεπώς, είναι αναγκαίος ο υπολογισμός της μέσης τιμής και της τυπικής απόκλισης

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{509}{50} = 10,18$$

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{(10,3-10,18)^2 + \dots + (8,6-10,18)^2}{50-1}} = 2,031$$

Παράδειγμα μετατροπής της πρώτης παρατήρησης σε z-score:

$$Z = \frac{X - \bar{X}}{S} = \frac{10,3 - 10,18}{2,031} = 0,06$$

Συνεπώς η πρώτη παρατήρηση είναι πολύ κοντά στη μέση τιμή.

Επαναλαμβάνουμε τη διαδικασία για όλες τις παρατηρήσεις και προκύπτει ο ακόλουθος πίνακας:



Z-Scores									
-2,4	-1,81	-1,71	-1,42	-1,32	-1,27	-1,17	-1,07	-1,07	-0,88
-0,78	-0,73	-0,63	-0,58	-0,48	-0,43	-0,43	-0,33	-0,29	-0,19
-0,14	-0,09	-0,09	-0,04	0,01	0,06	0,06	0,11	0,16	0,16
0,21	0,26	0,4	0,45	0,45	0,45	0,5	0,65	0,7	0,7
0,8	1,14	1,14	1,14	1,24	1,29	1,34	1,39	2,13	2,42

±1 Τυπικές αποκλίσεις απέχουν από τη μέση τιμή **32** παρατηρήσεις, δηλαδή το 64% των παρατηρήσεων.

±2 Τυπικές αποκλίσεις απέχουν από τη μέση τιμή 47 παρατηρήσεις, δηλαδή το 94% των παρατηρήσεων.

±3 Τυπικές αποκλίσεις απέχουν από τη μέση τιμή 50 παρατηρήσεις, δηλαδή το 100% των παρατηρήσεων.

Τα αντίστοιχα ποσοστά από των εμπειρικό κανόνα είναι 68%, 95% και 99,7%.

Συνεπώς βλέπουμε στο συγκεκριμένο παράδειγμα να επιβεβαιώνεται προσεγγιστικά ο εμπειρικός κανόνας.

Προσοχή! Δεν σημαίνει ότι αν σε κάποια δεδομένα επιβεβαιώνεται ο εμπειρικός κανόνας, ότι τα δεδομένα μου προέρχονται από την κανονική κατανομή.

Έστω ότι το ύψος των κατοίκων μιας χώρας ακολουθεί την **κανονική κατανομή** με μέση τιμή 175cm και τυπική απόκλιση 10cm.

A. Τι ποσοστό των κατοίκων έχουν ύψος μεταξύ 165cm και 185cm;

B. Τι ποσοστό των κατοίκων έχουν ύψος μεταξύ 145cm και 205cm;

Λύση

A. $165 = 175 - 10$ δηλαδή μέση τιμή μείον μία τυπική απόκλιση

$185 = 175 + 10$ δηλαδή μέση τιμή συν μία τυπική απόκλιση

Άρα το ποσοστό των κατοίκων που έχουν ύψος μεταξύ 165cm και 185cm είναι 68%. [Εμπειρικός Κανόνας]

B. $145 = 175 - 30$ δηλαδή μέση τιμή μείον τρεις τυπικές αποκλίσεις

$205 = 175 + 30$ δηλαδή μέση τιμή συν τρεις τυπικές αποκλίσεις

Άρα το ποσοστό των κατοίκων που έχουν ύψος μεταξύ 145cm και 205cm είναι 99,7%. [Εμπειρικός Κανόνας]

Έστω ότι κατανάλωση των κατοίκων μιας χώρας έχει μέση τιμή 175m^3 και τυπική απόκλιση 10m^3 . Τι ποσοστό των κατοίκων καταναλώνουν νερό μεταξύ 159m^3 και 191m^3 ;

Λύση

Δεν αναφέρεται ότι η μεταβλητή κατανάλωση ακολουθεί την **κανονική κατανομή** ~

Δεν μπορώ να χρησιμοποιήσω Εμπειρικό Κανόνα

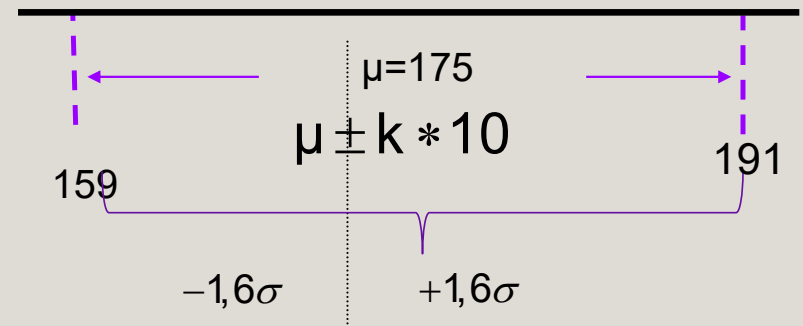
Από τον κανόνα του **Chebyshev** έχουμε ότι τουλάχιστον ένα ποσοστό $\left(1 - \frac{1}{k^2}\right) \cdot 100\%$ των τιμών πρέπει να συγκεντρώνονται σε απόσταση k $k \geq 2$ τυπικών αποκλίσεων **γύρω** από τη μέση τιμή.

Υπολογίζουμε τις αποστάσεις των ορίων από τη μέση τιμή

σε μονάδες τυπικών αποκλίσεων:

Κάτω όριο(159): $(175 - 159)/10 = 1,6$

Άνω όριο(191): $(191 - 175)/10 = 1,6$



$$1 - \frac{1}{1,6^2} = 1 - \frac{1}{2,56} = 60,94\%$$

τουλάχιστον 60,94% των κατοίκων καταναλώνουν νερό στο διάστημα (159,191)