ΘΕΜΑΤΙΚΉ ΕΝΌΤΗΤΑ 5

ΜΈΤΡΑ ΚΕΝΤΡΙΚΉΣ ΤΆΣΗΣ

Μέτρα κεντρικής τάσης

Μέτρα κεντρικής τάσης (central tendency measures) ή

Μέτρα θέσης (measures of location)

Περιγράφουν την τιμή γύρω από την οποία συγκεντρώνονται οι τιμές των δεδομένων που μελετώνται, δηλαδή τα δεδομένα περιγράφονται με έναν αριθμό.

Ερευνητικό Παράδειγμα

5	10	6	7	6	3	4	1	9	6
8	5	7	4	1	5	8	5	6	7
7	8	5	4	6	9	6	7	3	5
4	6	9	5	7	5	3	4	6	8
6	7	4	8	5	10	6	8	5	6

Ο πίνακας παρουσιάζει την επίδοσή των μαθητών σε ένα τεστ, σε κλίμακα 1-10.

Δεσπόζουσα τιμή (mode)

Κατανομή συχνοτήτων

Τιμές	Συχνότητα (*f*)
1	2
2	0
3	3
4	6
5	10
6	11
7	7
8	6
9	3
10	2
Σύνολο	N = 50

Δεσπόζουσα τιμή (δσπ) = 6

Η Δεσπόζουσα τιμή είναι η τιμή που εμφανίζεται τις περισσότερες φορές σε μία κατανομή, δηλαδή η τιμή με τη μεγαλύτερη συχνότητα.

- Σε μία γραφική παράσταση αντιστοιχεί στην κορυφή της κατανομής.
- Δεν επηρεάζεται από ακραίες τιμές
- Υπολογίζεται εύκολα.
- Δεν μπορεί να οδηγήσει σε υπολογισμό παραμέτρων του πληθυσμού

Δεσπόζουσα τιμή

Κατανομή συχνοτήτων

Τιμές	Συχνότητα (f)
1	2
2	0
3	3
4	6
5	11
6	11
7	6
8	6
9	3
10	2
Σύνολο	N = 50

Δύο γειτονικές θέσεις με μέγιστη συχνότητα

Δσπ = (5 + 6) / 2 = 5.5

Δεσπόζουσα τιμή

Ομαδοποιημένη κατανομή

Δσπ: η κεντρική τιμή στο αντίστοιχο διάστημα

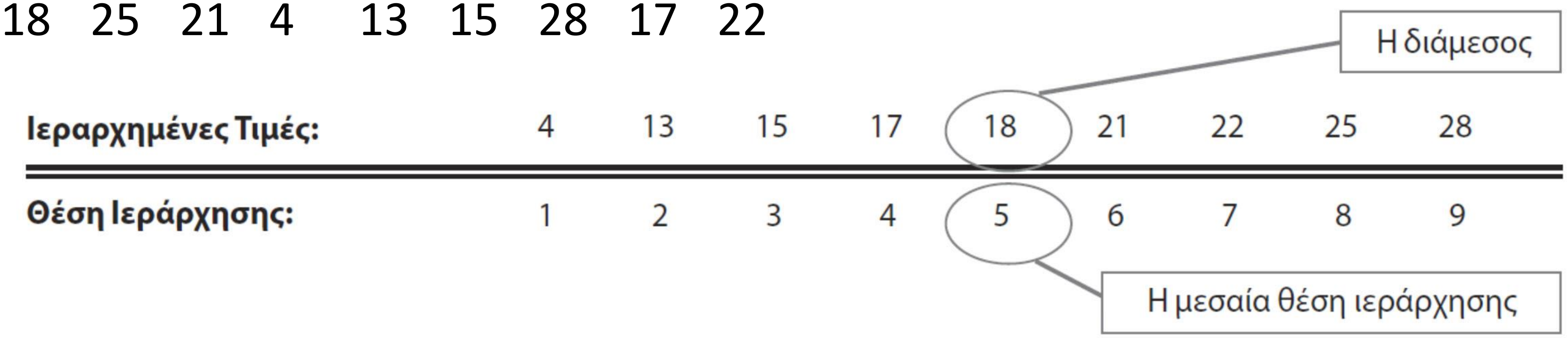
Διαστήματα Κλίμακας	Μέση τιμή	Συχνότητα (*f*)
50 – 54	52	11
45 – 49	47	14
40 – 44	42	23
35 – 39	37	29
30 – 34	32	21
25 – 29	27	16
20 – 24	22	10
Σύνολο		**N =124**

Διάμεσος (median)

- Είναι η μεσαία τιμή όταν ταξινομηθούν όλες οι τιμές της κατανομής
- Χωρίζει την κατανομή σε δύο ίσα τμήματα, ονομάζεται 50ό εκατοστημόριο
- Δεν επηρεάζεται από τις ακραίες τιμές, άρα είναι καλύτερο μέτρο για ασύμμετρη κατανομή
- Δεν μπορεί να χρησιμοποιηθεί για να υπολογιστούν παράμετροι πληθυσμού
- Χρειάζεται αρκετά μεγάλο πλήθος τιμών για να έχει νόημα

Παράδειγμα (Δμ = 18)

18 25 21 4 13 15 28 17 22



Διάμεσος (median)

Άρτιο πλήθος τιμών

Παράδειγμα (Δμ = 17.5)

18 25 21 4 13 15 17 22

Διάμεσος (median)

Άρτιο πλήθος τιμών, επαναλαμβανόμενες τιμές

Παράδειγμα

13 21 17 4 13 21 17 21 28 30

Η διάμεσος είναι ο μέσος όρος των δύο τιμών, δηλαδή 19.

Ιεραρχημένες Τιμές:	4	13	13	17	17	21	21	21	28	30
Θέση Ιεράρχησης:	1	2,5	2,5	4,5	4,5	7	7	7	9	10
	1	2	3	4	5	6	7	8	9	10

Η μεσαία θέση ιεράρχησης βρίσκεται ανάμεσα σε αυτές τις δύο θέσεις, δηλαδή 5,5.

Αριθμητικός Μέσος Όρος

Ο Μέσος Όρος (mean) είναι το πιο γνωστό και περισσότερο χρησιμοποιούμενο μέτρο κεντρικής τάσης. Εκφράζει την πιο αντιπροσωπευτική τιμή της κατανομής

- Αντικατοπτρίζει πιο πιστά την κεντρική τιμή της κατανομής σε σχέση με άλλα μέτρα
- Μπορεί να χρησιμοποιηθεί για τον υπολογισμό παραμέτρων του πληθυσμού
- Είναι πολύ ευαίσθητος στις ακραίες τιμές

$$\bar{X} = \frac{\sum X}{N}, \ \bar{X} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{x_1 + x_2 + x_3 + \ldots + x_N}{N}$$

Παράδειγμα

15 18 16 24 27 19 22

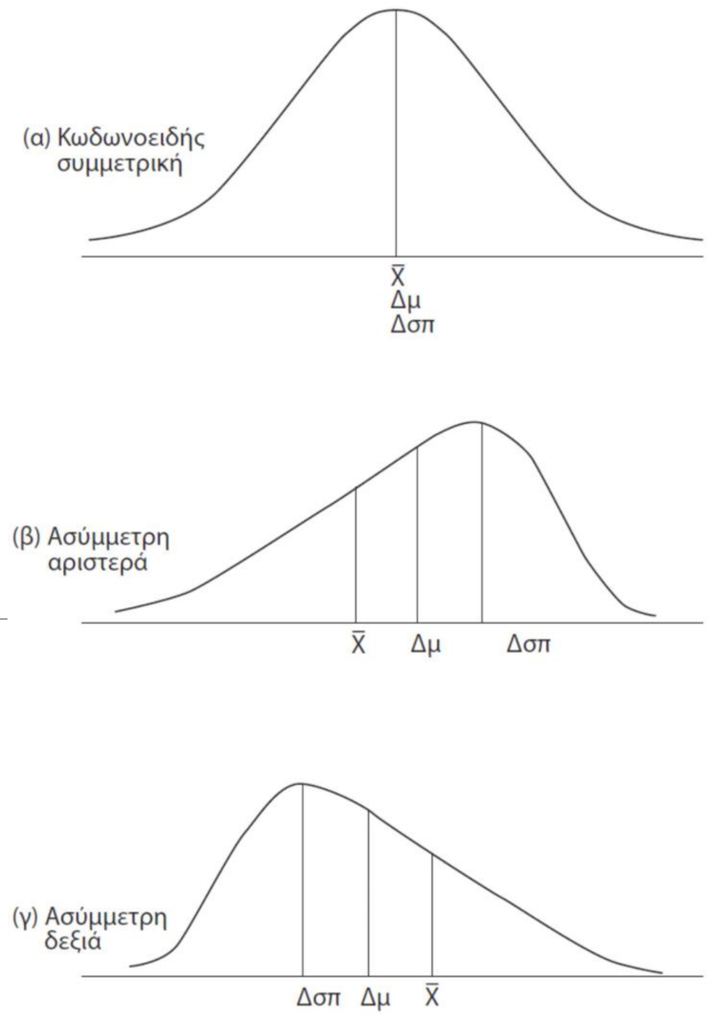
$$\bar{X} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{15 + 18 + 16 + 24 + 27 + 19 + 22}{7} = 20.14$$

Επιλογή Μέτρου Κεντρικής Τάσης

Η επιλογή του κατάλληλου μέτρου κεντρικής τάσης εξαρτάται από το σχήμα της κατανομής και από την κλίμακα Μέτρησης

- Κατηγορική κλίμακα: Δεσπόζουσα τιμή
- Ιεραρχική κλίμακα: Διάμεσος
- Ίσων διαστημάτων – Αναλογική

  Και τα τρία μέτρα – κυρίως Μ.Ο.



Η Έννοια της Διασποράς

Τα μέτρα κεντρικής τάσης δεν είναι αρκετά από μόνα τους να περιγράψουν μία κατανομή.
Εκτός από το πού βρίσκονται συγκεντρωμένες οι τιμές, μάς ενδιαφέρει και πόσο διασπείρονται γύρω από το «κέντρο».
Μας ενδιαφέρει δηλαδή η «Διασπορά».

Παράδειγμα
Επίδοση Μαθητών

Ποιο τμήμα έχει καλύτερη επίδοση?

Δσπ: 17
Δμ: 17
Μ.Ο.: 16.24

Τμήμα Α				Τμήμα Β			
17	18	19	14	20	16	14	17
18	17	13	16	17	20	9	16
20	13	17	17	19	16	17	20
14	17	15	18	15	17	9	16
16	12	19	17	17	10	20	17
14				19			

Φυλλογράφημα

Τμήμα Α		Τμήμα Β
0	2	0000
99	1	99
888	1	
777777	1	777777
66	1	6666
5	1	5
444	1	4
33	1	
2	1	
	1	
	1	0
	0	99

Εύρος (Range)

Εύρος L είναι το μήκος της κλίμακας από τη μικρότερη μέχρι τη μεγαλύτερη τιμή, L.

- Υπολογίζεται εύκολα
- Περιλαμβάνει και τις ακραίες τιμές της κατανομής
- Επηρεάζεται πολύ από τις ακραίες τιμές
- Δεν παρέχει πληροφορία για ενδιάμεσες τιμές (πόσο συγκεντρωμένες είναι οι τιμές γύρω από το Μέσο Όρο).

Παράδειγμα υπολογισμού Εύρους

Για παράδειγμα οι τιμές των βαθμών σε δύο τμήματα δίνονται στον παρακάτω πίνακα. Υπολογίζοντας το εύρος σε κάθε περίπτωση:

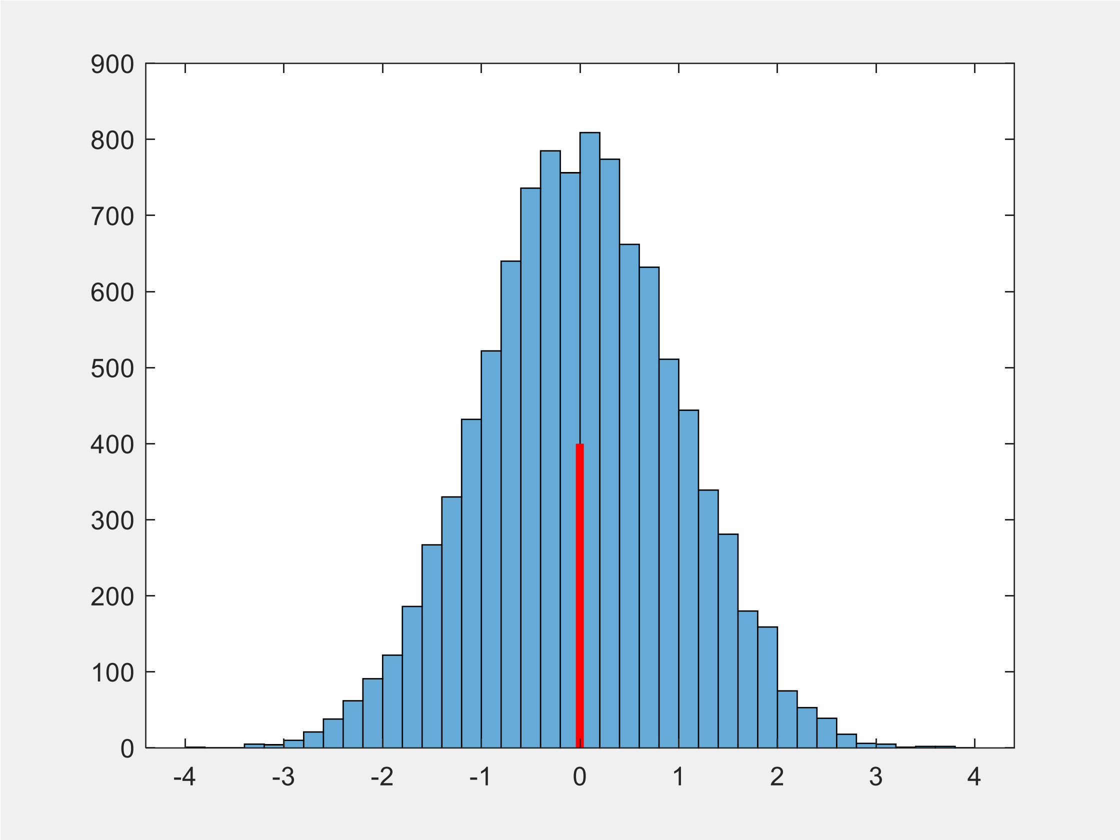
$L_A = 20 – 12 = 8$

$L_B = 20 – 9 = 11$

Παρατήρηση: το εύρος των τιμών στο δεύτερο τμήμα είναι μεγαλύτερο άρα υπάρχει μεγαλύτερη διασπορά των τιμών.

Τμήμα Α				Τμήμα Β			
17	18	19	14	20	16	14	17
18	17	13	16	17	20	9	16
20	13	17	17	19	16	17	20
14	17	15	18	15	17	9	16
16	12	19	17	17	10	20	17
14				19			

Διάμεσος



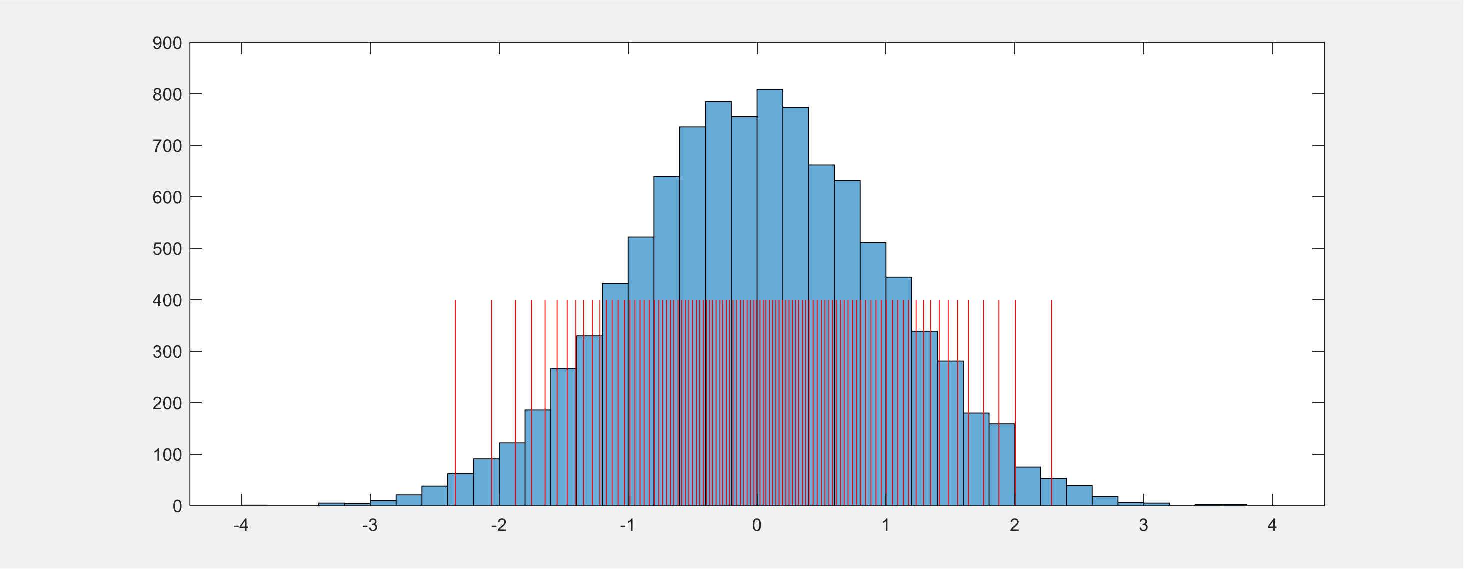
Διάμεσος (δ) ενός δείγματος *ν* παρατηρήσεων οι οποίες έχουν διαταχθεί σε αύξουσα σειρά ορίζεται ως η μεσαία παρατήρηση, όταν το *ν* είναι περιττός αριθμός, ή ο μέσος όρος (ημιάθροισμα) των δύο μεσαίων παρατηρήσεων όταν το *ν* είναι άρτιος αριθμός.

Η διάμεσος χωρίζει την κατανομή σε δύο ίσα μέρη, το 50% των δεδομένων βρίσκονται αριστερά και το υπόλοιπο 50% δεξιά.

Εκατοστημόρια

Ορίζουμε ως *κ*-εκατοστιαίο σημείο ή $P_κ$ εκατοστημόριο ενός συνόλου παρατηρήσεων την τιμή εκείνη για την οποία το πολύ *κ*% των παρατηρήσεων είναι μικρότερες του $P_κ$ και το πολύ (100-*κ*)% των παρατηρήσεων είναι μεγαλύτερες από την τιμή αυτήν.

Τα ε**κατοστημόρια** (percentiles) $P_κ$, $κ = 1,2,...,99$, χωρίζουν τη συνολική συχνότητα σε 100 ίσα μέρη.



Τεταρτημόρια

Ειδική περίπτωση εκατοστημορίων είναι τα P_{25}, P_{50}, P_{75}, τα οποία καλούνται **τεταρτημόρια** (quartiles) και συμβολίζονται με Q_1, Q_2 και Q_3, αντίστοιχα.

Q_1: αριστερά το πολύ 25% των παρατηρήσεων και δεξιά το πολύ 75%.

Q_3: αριστερά το πολύ 75% των παρατηρήσεων και δεξιά το πολύ 25% των παρατηρήσεων.

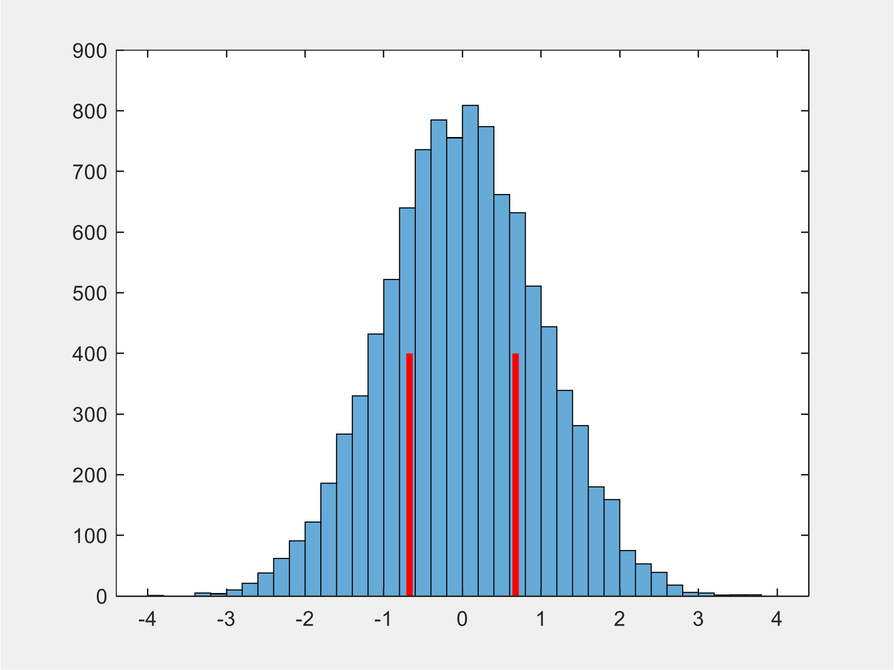
$Q_2 = P_{50}$ συμπίπτει και με τη διάμεσο, δηλαδή $Q_2 = \delta$.

Τα μέτρα αυτά χρησιμοποιούνται αρκετά συχνά για τη μελέτη ενός συνόλου δεδομένων.

Ενδοτεταρτημοριακό Εύρος

Inter-Quartile Range = IQR = $Q3$ - $Q1$

- Είναι σχετικά εύκολο στον υπολογισμό
- Δεν επηρεάζεται από τις ακραίες τιμές
- Είναι αντιπροσωπευτικό των κεντρικών τιμών της κατανομής
- Δεν επιτρέπει την εκτίμηση παραμέτρου του πληθυσμού



Απόκλιση και Μέση αριθμητική απόκλιση

Απόκλιση: $d_i = x_i - \bar{X}$ ή $d_i = x_i - \bar{x}$

Υπολογίζει την απόσταση των τιμών από το Μέσο Όρο και μπορεί να πάρει θετικές ή αρνητικές τιμές.

Μέση αριθμητική απόκλιση $MD = \sum_i \frac{d_i}{N} = \sum_i \frac{x_i - \bar{x}}{N}$

Διακύμανση (s^2)

Διακύμανση (Variance): Υπολογίζει το τετράγωνο των αποκλίσεων, ώστε να αποφύγουμε την εξουδετέρωση των προσήμων.

$$s^2 = \sum_i \frac{d_i^2}{N}$$

Τυπική απόκλιση (*s*)

Τυπική απόκλιση (standard deviation): η τετραγωνική ρίζα της διακύμανσης

$$s = \sqrt{\sum_i \frac{d_i^2}{N}} = \sqrt{\sum_i \frac{(x_i - \bar{x})^2}{N}}$$

Όταν θέλουμε να εκτιμήσουμε την τυπική απόκλιση του πληθυσμού διαιρούμε με *N*-1 αντί για *N*

$$\sigma = \sqrt{\sum_i \frac{d_i^2}{N-1}} = \sqrt{\sum_i \frac{(x_i - \bar{x})^2}{N-1}}$$

- Λαμβάνει υπόψιν όλες τις τιμές της κατανομής
- Μπορεί να χρησιμοποιηθεί για τον υπολογισμό παραμέτρων του πληθυσμού
- Είναι το πιο ευαίσθητο από τα μέτρα διασποράς και είναι ευαίσθητο στις ακραίες τιμές

Στο παράδειγμα

Περιγραφικό στατιστικό μέτρο	Τμήμα Α	Τμήμα Β
Δεσπόζουσα τιμή	17	17
Διάμεσος	17	17
Μέσος Όρος	16,24	16,24
Εύρος	8	11
Ενδοτεταρτημοριακό εύρος	3,49	3,31
Τυπική απόκλιση	2,16	3,28

Περισσότερες έκτοπες τιμές στο Τμήμα Β, πιο ομογενείς επιδόσεις στο Τμήμα Α

Επιλογή μέτρου διασποράς

- Κατηγορική κλίμακα: Δεν έχει νόημα
- Ιεραρχική κλίμακα: εύρος
- Κλίμακα ίσων διαστημάτων ή αναλογική:
  - Όταν δεν υπάρχουν πολύ ακραίες τιμές: Μ.Ο. και τυπική απόκλιση
  - Όταν υπάρχουν αρκετές ακραίες τιμές: Διάμεσος και IQR
  - Χονδροειδής πληροφόρηση: Δεσπόζουσα τιμή και εύρος
- Τυπική απόκλιση: Οδηγεί στην εκτίμηση παραμέτρων του πληθυσμού