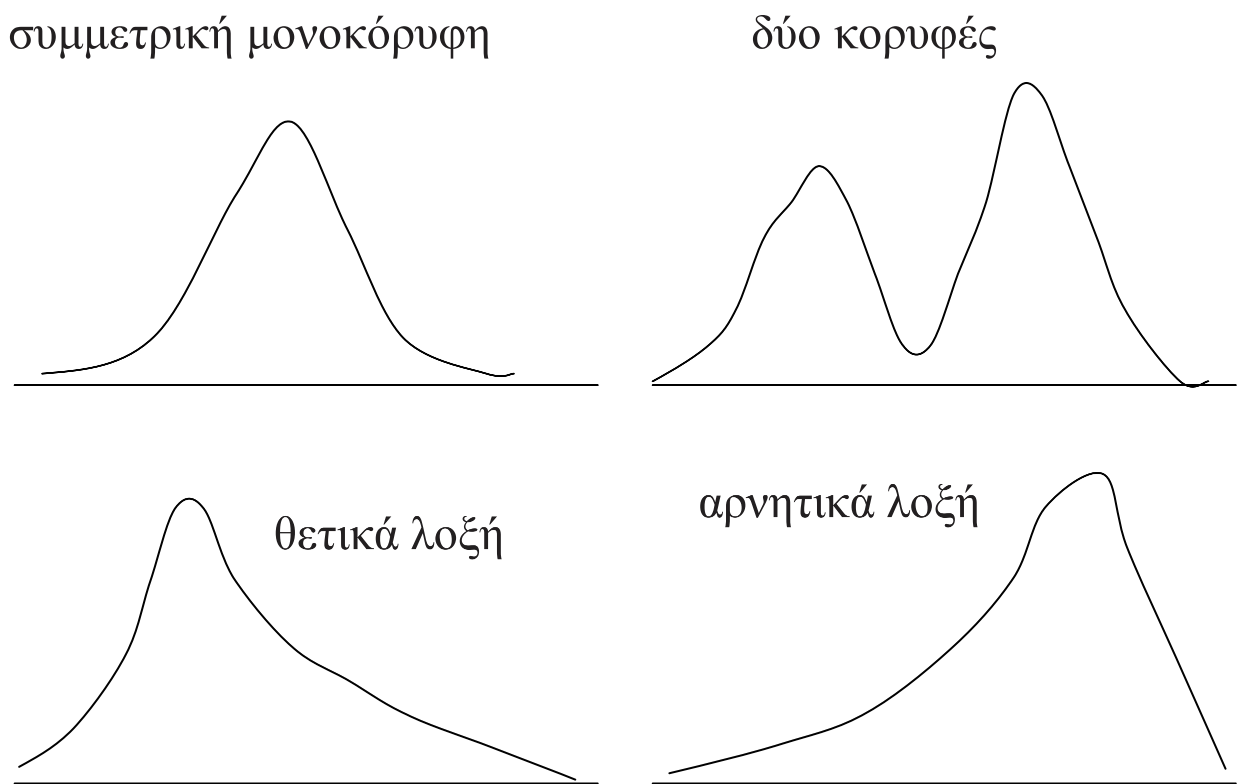
Κατανομές συνεχών τυχαίων μεταβλητών



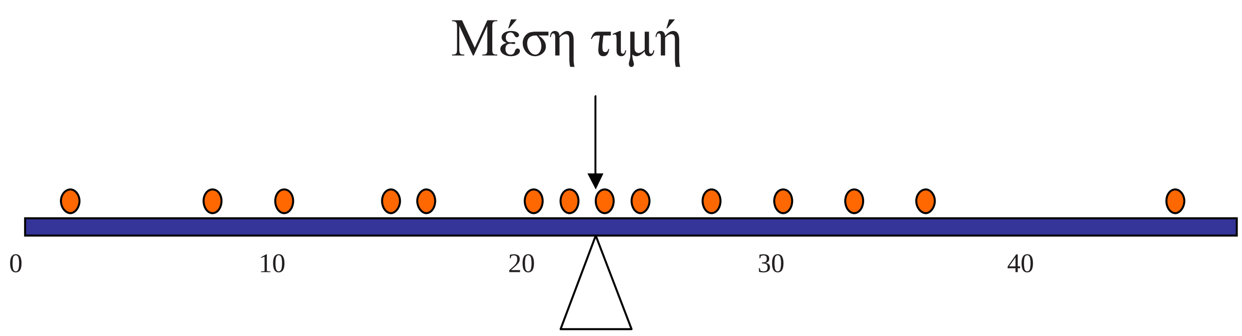
Περιγραφικά μέτρα

Μέτρα θέσης (μέτρα κεντρικής τάσης)

$x_1, x_2, \ldots, x_n$, παρατηρήσεις του δείγματος για τ.μ. X

- **δειγματική μέση τιμή** $\bar{x}$

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$



- **δειγματική διάμεσος** $\tilde{x}$

$x_1, x_2,\ldots, x_n$, σε αύξουσα σειρά

n περιττός ➔ $\tilde{x}$ είναι η τιμή στη θέση $(n+1)/2$,

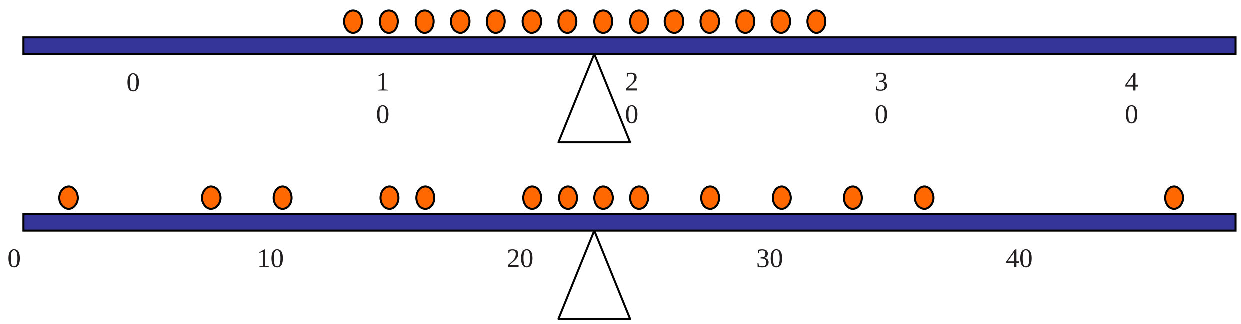
n άρτιος ➔ $\tilde{x}$ είναι το ημιάθροισμα των τιμών στις θέσεις $n/2$ και $n/2+1$

- **δειγματική επικρατούσα τιμή**

τιμή που εμφανίζεται με τη μεγαλύτερη συχνότητα.

- ☆ Η σημαντικότερη είναι η $\bar{x}$
- ☆ $\bar{x}$ χρησιμοποιεί τις τιμές $x_1, x_2,\ldots, x_n$,
- ☆ $\tilde{x}$ χρησιμοποιεί την τάξη των $x_1, x_2,\ldots, x_n$,
- ☆ $\bar{x}$ επηρεάζεται από μακρινές τιμές, όχι η $\tilde{x}$

Μέτρα μεταβλητότητας



- δειγματικό εύρος *R*= *xmax* - *xmin*,

- δειγματική διακύμανση (διασπορά) s^2
δειγματική τυπική απόκλιση s

απόκλιση μιας παρατήρησης x_i από $\bar{x}$: $x_i - \bar{x}$

Άθροισμα αποκλίσεων:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

Άθροισμα τετραγώνων αποκλίσεων

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

$$\boxed{s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)}$$

- **εκατοστιαία σημεία**

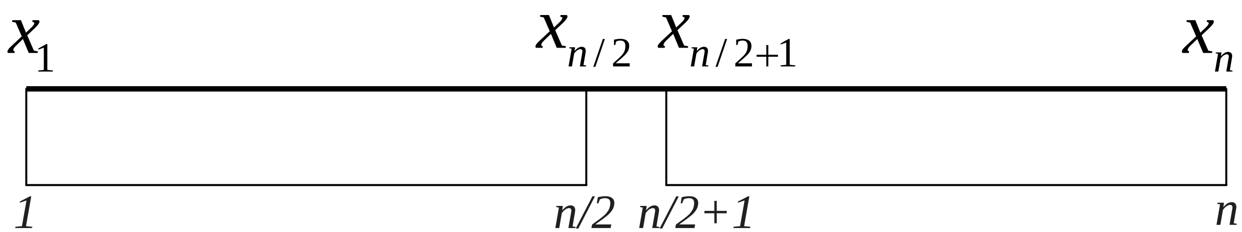
$\tilde{x}$ είναι το 50-εκατοστιαίο σημείο

Q_1: 25-εκατοστιαίο σημείο, **πρώτο τεταρτομόριο**

Q_3: 75-εκατοστιαίο σημείο, **τρίτο τεταρτομόριο**

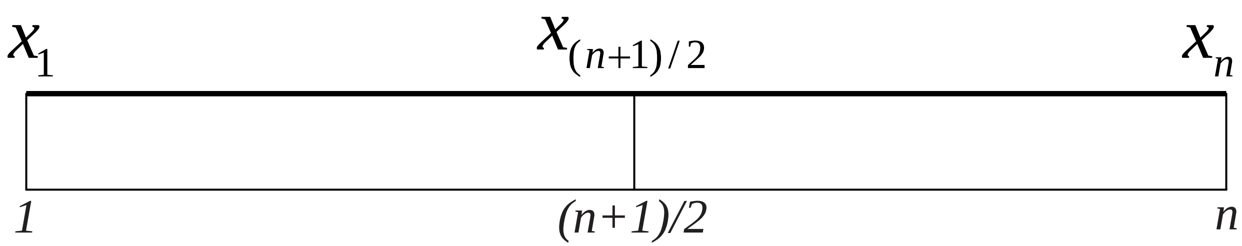
$I = Q_1 - Q_3$: **ενδοτεταρτομοριακό εύρος**

$n = 2k$



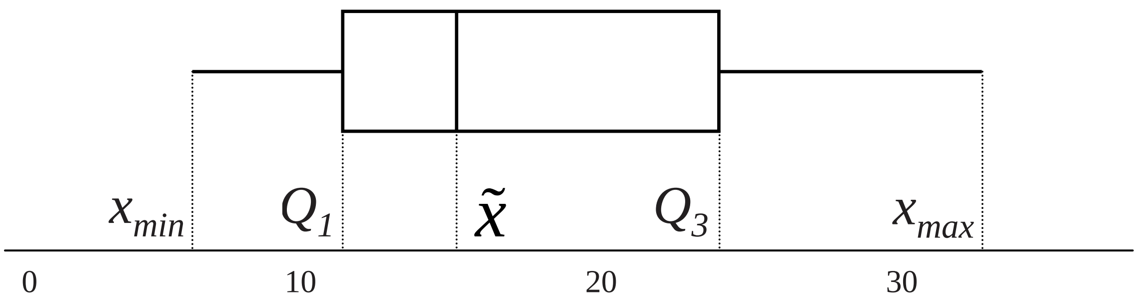
Q_1: διάμεσος των $x_1,\ldots,x_{n/2}$ Q_3: διάμεσος των $x_{n/2+1},\ldots,x_n$

$n = 2k+1$



Q_1: διάμεσος των $x_1,\ldots,x_{(n+1)/2}$ Q_3: διάμεσος των $x_{(n+1)/2},\ldots,x_n$

Σύνοψη 5 αριθμών, θηκόγραμμα



μακρινή τιμή x_i	
ύποπτη ακραία, απόμακρη τιμή (outlier)	SPSS → o
ακραία τιμή (extreme)	SPSS → *

Παράδειγμα:

Για την αρχιτεκτονική τοπίου ορίστηκε μια περιοχή με 10 αγροτεμάχια και η έκταση τους είναι (σε m^2)

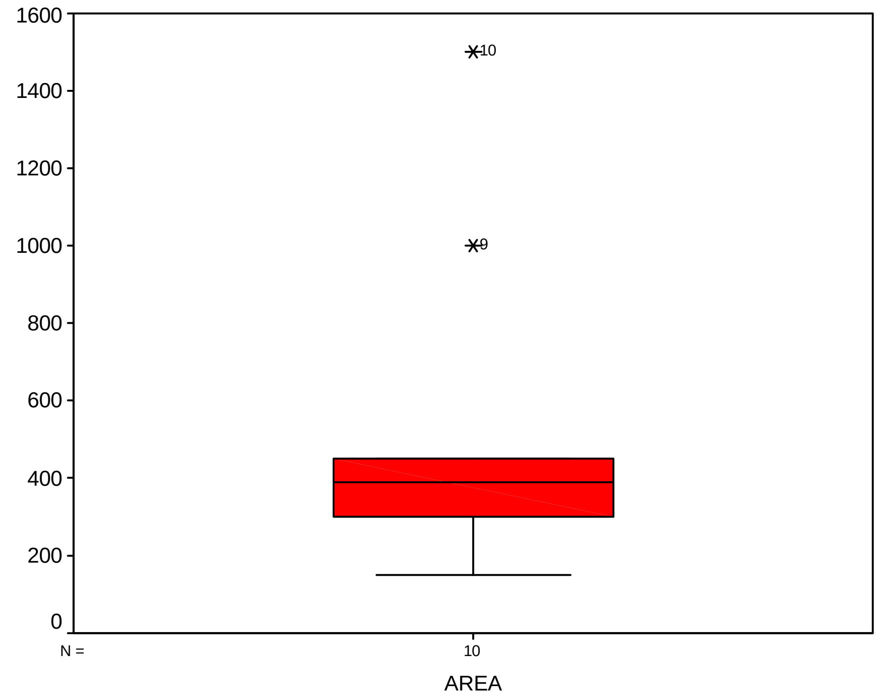
150 200 300 340 380 400 400 450 1000 1500

$$\bar{x} = \frac{150+200+300+340+380+400+400+450+1000+1500}{10} = 512$$

$$\tilde{x} = \frac{x_{n/2} + x_{n/2+1}}{2} = \frac{x_5 + x_6}{2} = \frac{380+400}{2} = 390$$

x_{min} =150m^2 x_{max} =1500m^2

R =1350m^2



$$\sum_{i=1}^{n} x_i^2 = 150^2 + 200^2 + 300^2 + 340^2 + 380^2 + 400^2 +$$

$$400^2 + 450^2 + 1000^2 + 1500^2 = 4185000$$

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) = \frac{1}{9}\left(4185000 - 10 \cdot 512^2\right) = 173730$$

$$s = \sqrt{173730} = 416.8$$

Παράδειγμα:

Χρόνος ανάφλεξης (δευτερόλεπτα) δύο υλικών ταπετσαρίας

Υλικό 1 (30 δεδομένα)

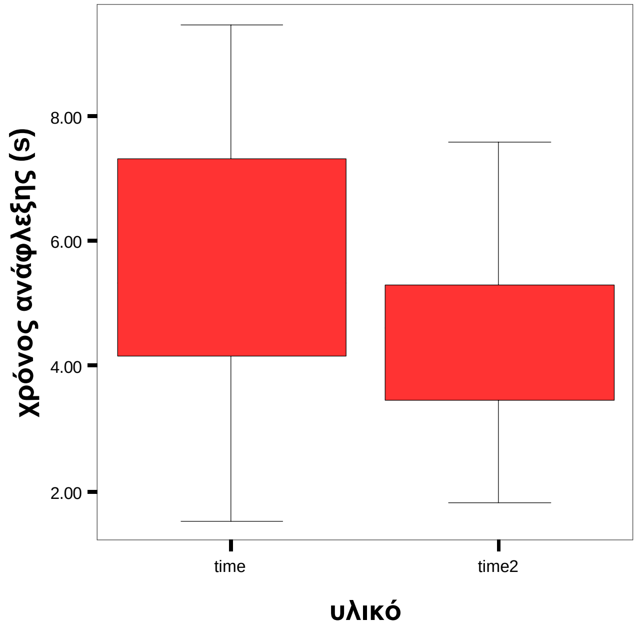
1. 52	6. 34	4. 48	4. 89	5. 56	4. 98
9. 45	7. 58	8. 4	6. 2	5. 01	6. 72
5. 09	8. 67	7. 34	7. 32	5. 98	3. 65
3. 48	3. 4	6. 89	4. 44	3. 8	2. 35
4. 16	5. 9	9. 01	3. 02	4. 5	7. 65

Υλικό 2 (20 δεδομένα)

2. 35	3. 56	4. 8	6. 7
4. 55	2. 88	6. 05	3. 4
3. 47	3. 45	5. 9	4. 8
7. 57	5. 6	3. 45	4. 6
1. 8	4. 6	4. 69	5

Υλικό 1		Υλικό 2
Φύλλο	Μίσχος	Φύλλο
5	1.	8
3	2.	38
86440	3.	44445
985441	4.	566688
99500	5.	069
8732	6.	07
6533	7.	5
64	8.	
40	9.	

Μέτρο	*Υλικό 1 (μεταβλητή X_1)*	*Υλικό 2 (μεταβλητή X_2)*
Μέση τιμή	$\bar{x}_1 = 5.59$ s	$\bar{x}_2 = 4.46$ s
Διάμεσος	$\tilde{x}_1 = 5.32$ s	$\tilde{x}_1 = 4.60$ s
Εύρος	$R_1 = 7.93$ s	$R_2 = 5.77$ s
Διασπορά	$s_1^2 = 4.12$ s^2	$s_2^2 = 2.11$ s^2
Τυπική απόκλιση	$s_1 = 2.03$ s	$s_2 = 1.45$ s
Πρώτο τεταρτημόριο	$Q_1 = 4.1$ s	$Q_1 = 3.4$ s
Τρίτο τεταρτημόριο	$Q_1 = 7.3$ s	$Q_3 = 5.3$ s
Ενδοτεταρτημοριακό εύρος	$I_1 = 3.2$ s	$I_2 = 1.9$ s



Παρατηρήσεις:

-η μεταβλητότητα (δειγματική διασπορά, εύρος και ενδοτεταρτημοριακό εύρος) του χρόνου ανάφλεξης του πρώτου υλικού είναι μεγαλύτερη από αυτή του δεύτερου υλικού,

- κεντρική τάση (δειγματική μέση τιμή και διάμεσος) του χρόνου ανάφλεξης του πρώτου υλικού είναι μεγαλύτερη από αυτή του δεύτερου υλικού.

Σημείωση:

Δεδομένα: διακεκριμένες τιμές x_i, $i=1,\ldots,n$, για μικρό n,
f_i: αντίστοιχη συχνότητα

$$N = \sum_{i=1}^{n} f_i$$

Υπολογισμός π.χ. δειγματική μέση τιμή

$$\bar{x} = \frac{\overbrace{x_1 + x_1 + \cdots x_1}^{f_1} + \cdots + \overbrace{x_n + x_n + \cdots x_n}^{f_n}}{f_1 + \cdots f_n} = \frac{f_1 x_1 + \cdots + f_n x_n}{N} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i$$

Παράδειγμα (δωμάτια διαμερίσματος)

Δωμάτια x_i	1	2	3	4	5	6	7	8	9
Συχνότητα f_i	19	29	31	21	13	3	2	1	1
Αθροιστική συχνότητα	19	48	79	100	113	116	118	119	120

Δειγματική μέση τιμή

$$\bar{x} = \frac{1}{120}\sum_{i=1}^{9} f_i x_i = \frac{19\cdot 1 + 29\cdot 2 + 31\cdot 3 + 21\cdot 4 + 13\cdot 5 + 3\cdot 6 + 2\cdot 7 + 1\cdot 8 + 1\cdot 9}{120} = 3.07$$

Δειγματική διάμεσος $\tilde{x}$?

Δειγματική διασπορά

$$s^2 = \frac{1}{N-1}\left(\sum_{i=1}^{n} f_i x_i^2 - N\overline{x}^2\right) = \frac{1}{119}\left(\sum_{i=1}^{9} f_i x_i^2 - 119\overline{x}^2\right) = 2.5$$

Δειγματική τυπική απόκλιση είναι $s = \sqrt{2.5} = 1.58$

Q1=2 *Q1=4* ➔ *I=Q3 - Q1=4-2=2*