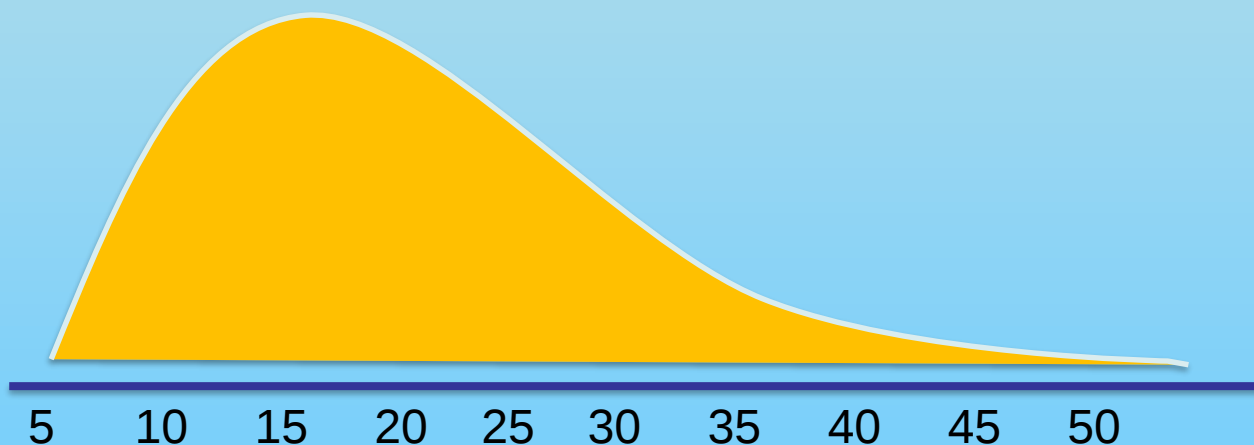
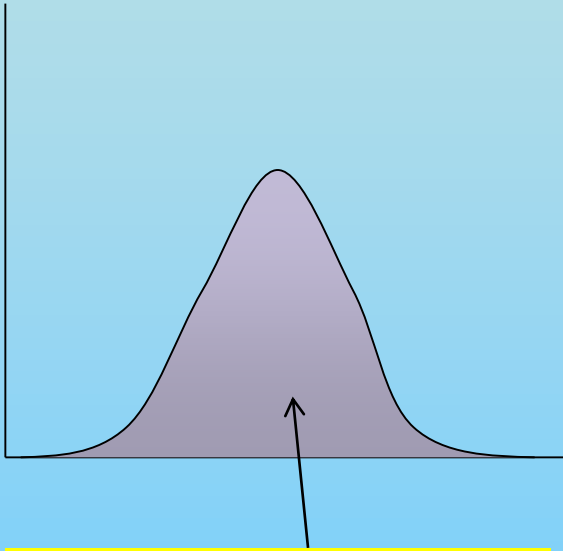


Κατανομή συχνοτήτων για ένα ποσοτικό χαρακτηριστικό (συνεχές)

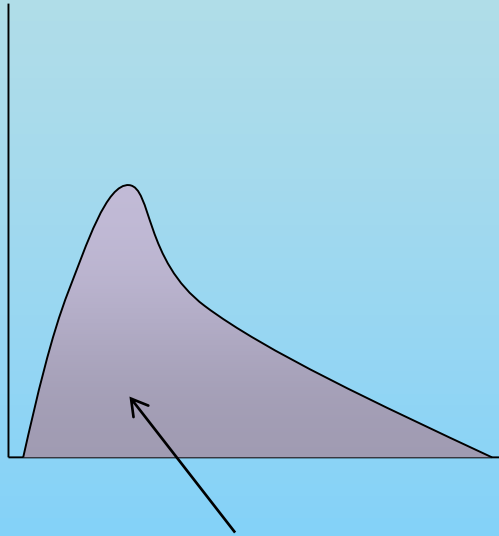


Τιμές του X
(π.χ. για κάποιο ΤΕΣΤ...)

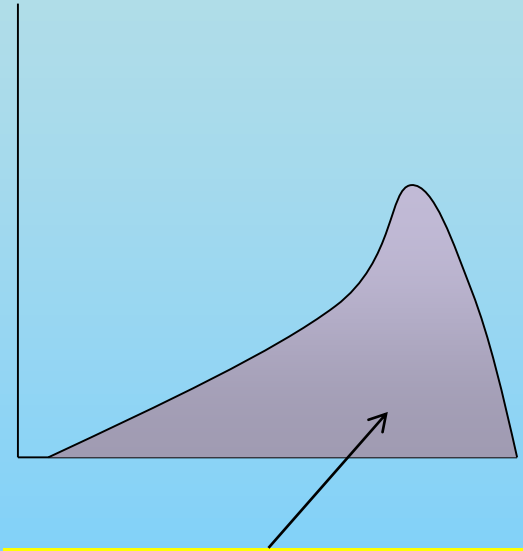
συμμετρικές και μη συμμετρικές κατανομές



Πολλές τιμές στη μέση,
λίγες μεγάλες τιμές και
λίγες μικρές τιμές



Πολλές μικρές τιμές,
κάποιες τιμές στη μέση
και λίγες μεγάλες τιμές



Πολλές μεγάλες τιμές,
κάποιες τιμές στη μέση
και λίγες μικρές τιμές

Μέτρα μεταβλητότητας ή διασποράς

Σε τι διαφέρουν αυτές οι κατανομές;



Μικρή διασπορά

Μεγάλη διασπορά

Παράδειγμα

Σε ένα τέστ πήραμε τα αποτελέσματα

45 46 46 50 48 47 48

(ο μέσος όρος βγήκε 47,1)

Ενώ σε άλλο τεστ είχαμε

32 48 53 62 72 29 34

(ο μέσος όρος και εδώ, βγήκε 47,1)

τα τεστ αυτά διαφέρουν – το πρώτο είναι πιο «ομοιογενές» από το δεύτερο.

1) Εύρος – Ενδοτεταρτομοριακό Εύρος

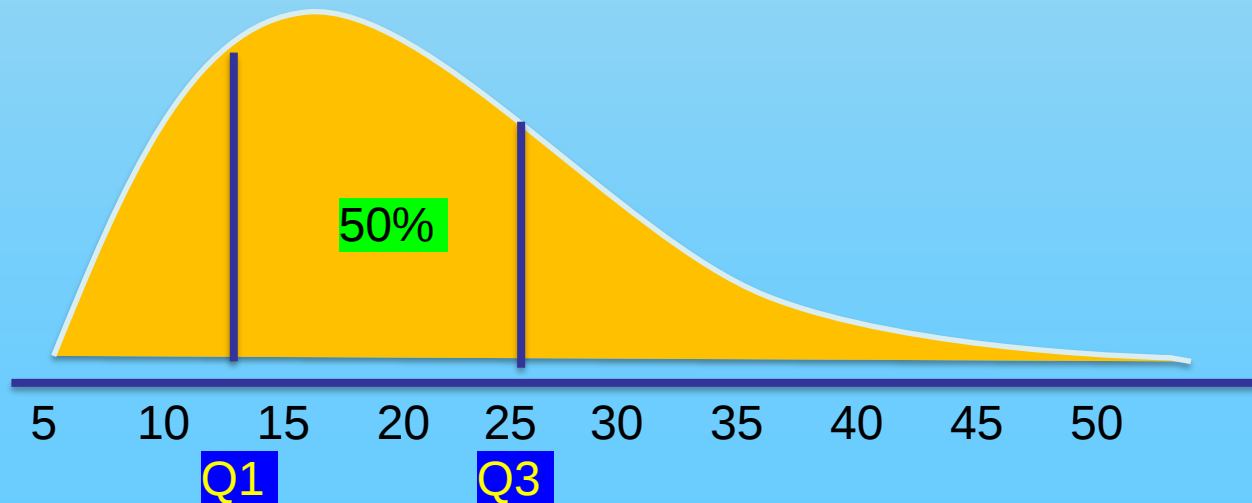
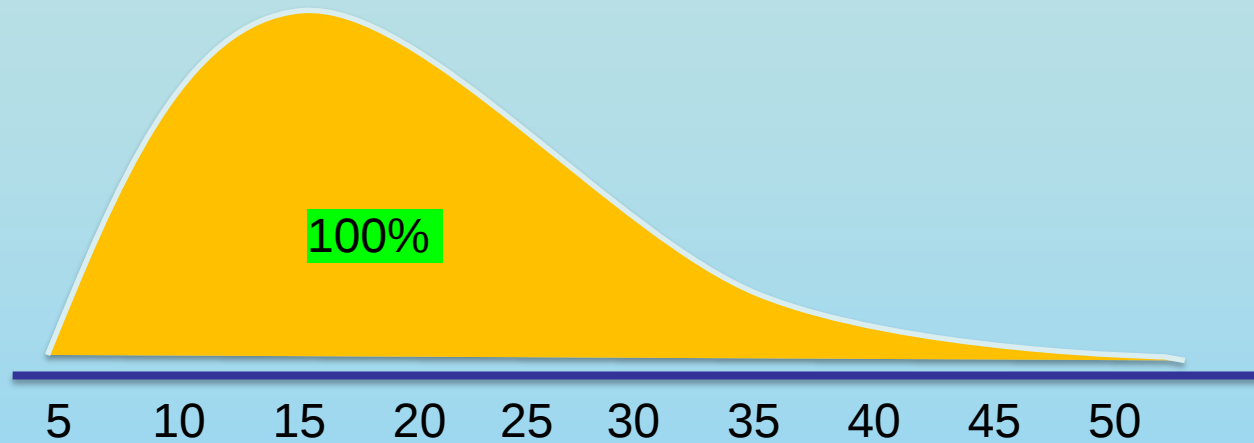
A) Εύρος τιμών (R)

Μεγαλύτερη – Μικρότερη Τιμή
(Max – Min)

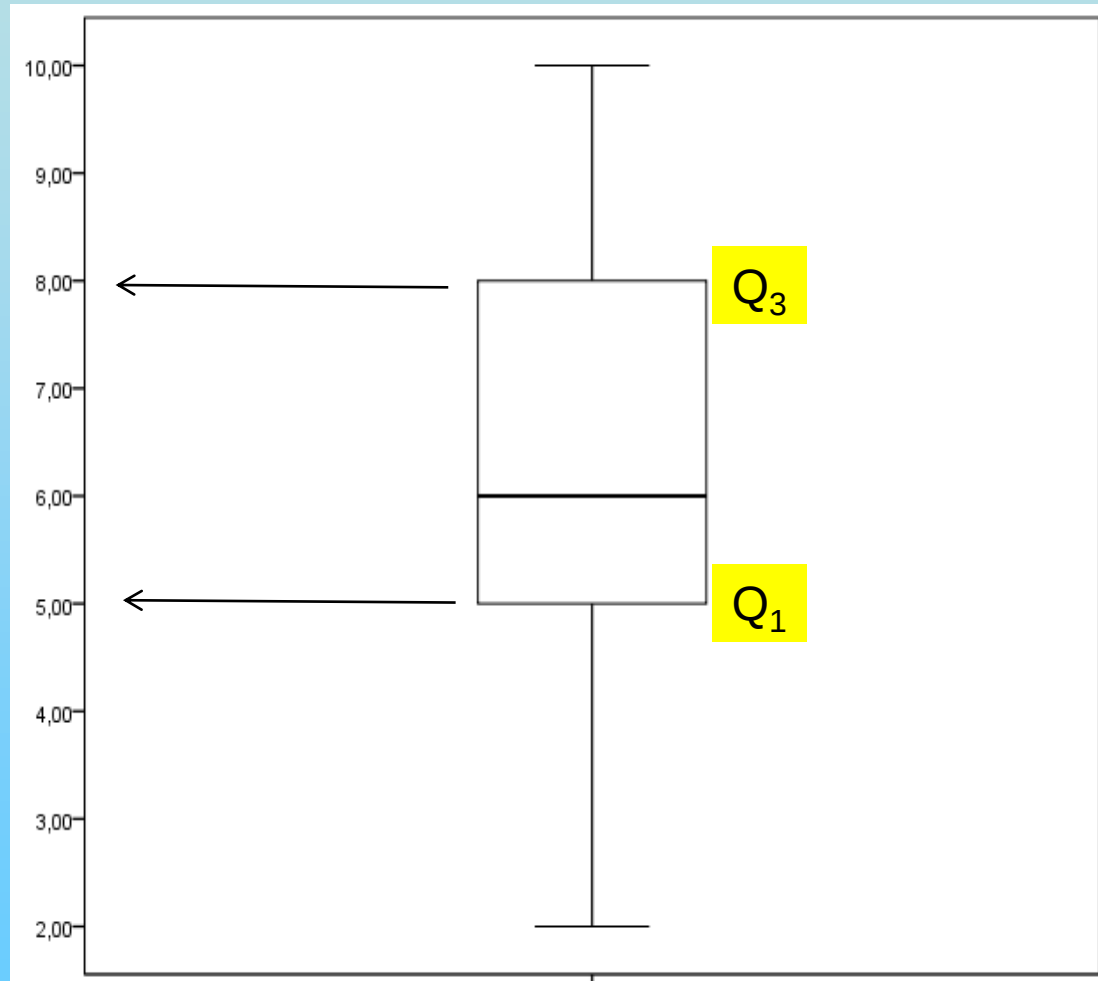
B) Ενδοτεταρτομοριακό Εύρος

Είναι η διαφορά $Q_3 - Q_1$

Εύρος – Ενδοτεταρτομοριακό Εύρος



Θηκόγραμμα



Και τα δύο χρειάζονται σε μια παρουσίαση αποτελεσμάτων, συνήθως δηλαδή γράφουμε σε μια παρουσίαση

Min=...

Max=

Q1=...

Q3=...

αλλά εκείνο το μέτρο που ταιριάζει περισσότερο στην έννοια της μεταβλητότητας (ποικιλότητας) είναι η

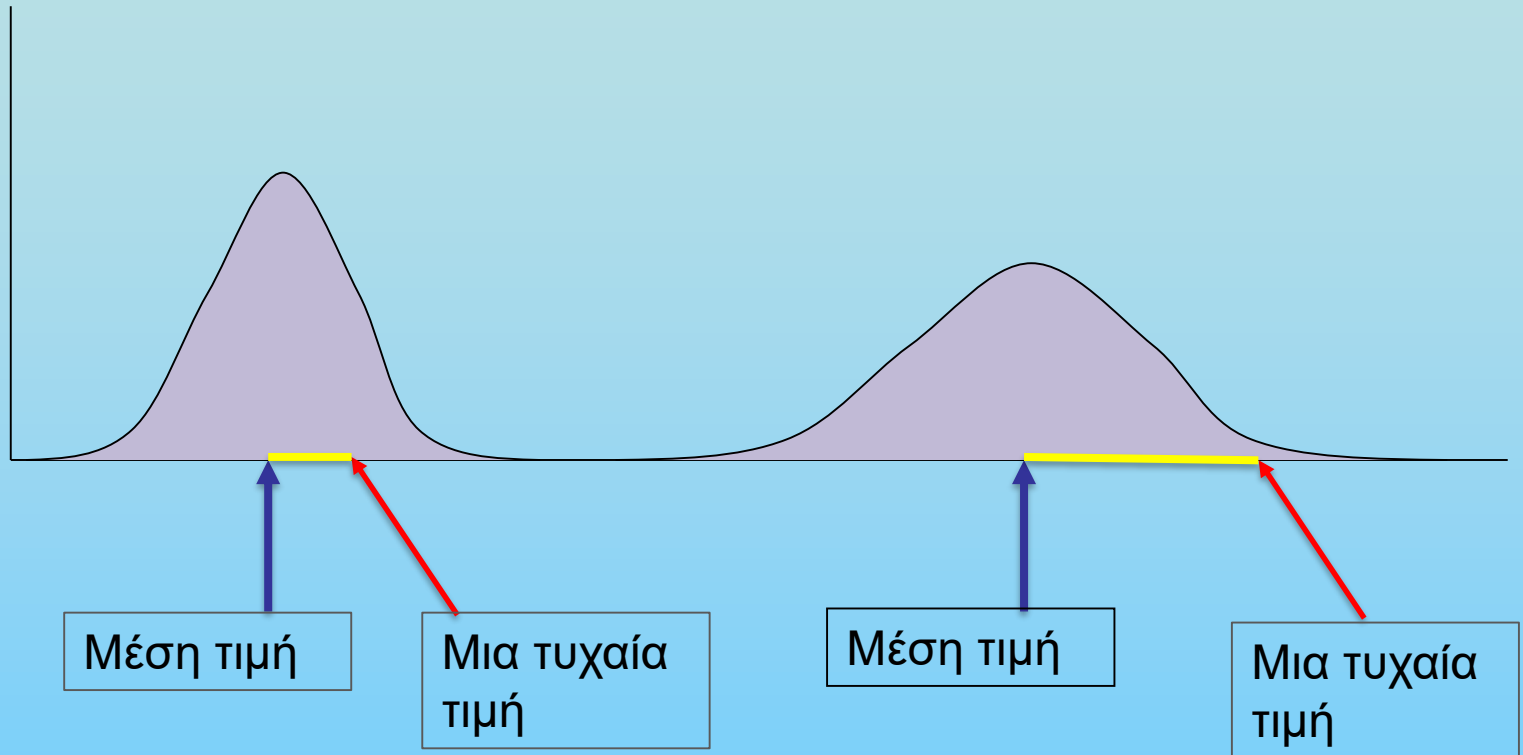
2) Η Διακύμανση

Η διακύμανση μας δείχνει πόσο μακριά (ή κοντά) από τη μέση τιμή, είναι οι άλλες τιμές που είχαμε.

Όταν οι τιμές είναι μακριά από το μέσο όρο, η διακύμανση είναι μεγάλη,

Όταν όμως οι τιμές είναι κοντά στο μέσο όρο, η διακύμανση είναι μικρή

Διακύμανση



Η διακύμανση βασίζεται στις **αποστάσεις** που έχουν οι διάφορες τιμές του δείγματος μας, από την μέση τιμή

Πως υπολογίζεται η διακύμανση

α) Ξεκινάμε με τα παρακάτω δεδομένα (6 τιμές)

6 8 8 9 12 14

β) Η μέση τιμή είναι $= 9,5$

γ) Από όλες τις τιμές, αφαιρώ τη μέση τιμή (9,5)
και βρίσκω τις διαφορές αυτές:

-3.5 -1.5 -1.5 -0.5 +2.5 +4.5

Δυστυχώς η μέση τιμή των διαφορών αυτών
είναι πάντα ίση με μηδέν, γιατί έχουμε
αρνητικά και θετικά! Αυτό δεν είναι τυχαίο,
συμβαίνει πάντα.

δ) Για το λόγο αυτό τις υψώνουμε στο τετράγωνο
ώστε να γίνουν θετικές

$$(-3.5)^2, (-1.5)^2, (-1.5)^2, (-0.5)^2, (+2.5)^2, (+4.5)^2$$

$$\text{δηλ. } 12.25 \quad 2.25 \quad 2.25 \quad 0.25 \quad 6.25 \quad 20.25$$

ε) τις προσθέτουμε και βρίσκουμε 43,5
και τελικά

στ) διαιρούμε με το πλήθος των παρατηρήσεων
μείον ένα (δηλαδή με το $n-1$, δηλαδή με το 5) και
φτάνουμε στη ΔΙΑΚΥΜΑΝΣΗ

$$43,5 / 5 = 8,7$$

2) Η Διακύμανση (variance)

- Είναι η μέση τιμή των ‘τετραγώνων’ των αποκλίσεων των τιμών από τη μέση τιμή τους

$$s^2 = \frac{\sum (X - \bar{X})^2}{n - 1}$$

Συμβολισμός

Παράδειγμα

α) Για το πρώτο τέστ

45 46 46 50 48 47 48

β) Βρήκαμε μέση τιμή **47,1**

Τα επόμενα βήματα:

	γ	δ
τιμές	τιμές-μέση τιμή	στο τετράγωνο
45	-2,1	4,41
46	-1,1	1,21
46	-1,1	1,21
50	2,9	8,41
48	0,9	0,81
47	-0,1	0,01
48	0,9	0,81

ε) προσθέτουμε και βγάζουμε 16,87

στ) το πλήθος είναι 7, οπότε το $n-1$ είναι 6, οπότε διαιρούμε το 16,87 με το 6, οπότε η διακύμανση είναι

$$16,87 / 6 = \mathbf{2,81}$$

*Για το δεύτερο τεστ, με τους ίδιους υπολογισμούς βρίσκουμε διακύμανση ίση με **267,5***

Πρόβλημα;

Ποιες είναι οι μονάδες μέτρησης της διακύμανσης;

Αν π.χ. μετράμε βάρος σε Κιλά, η μέση τιμή εκφράζει Κιλά, η Διακύμανση τι εκφράζει?

Κιλά στο τετράγωνο ή $(\text{Κιλά})^2$!!!!!

Αν μετράμε αριθμό παιδιών, η μέση τιμή είναι αριθμός παιδιών, αλλά η διακύμανση είναι $(\text{αριθμός παιδιών})^2$!!!!!

3) Τυπική Απόκλιση (Standard Deviation)

$$S = \sqrt{S^2}$$

Είναι η τετραγωνική ρίζα
της διακύμανσης

$$\text{τυπική απόκλιση} = \sqrt{\text{διακύμανση}}$$

Η τυπική απόκλιση έχει τις **ίδιες μονάδες μέτρησης** με τις αρχικές μετρήσεις. Για το λόγο αυτό χρησιμοποιείται πιο συχνά στην παρουσίαση των αποτελεσμάτων μας.

4) Συντελεστής μεταβλητότητας (Coefficient Of Variation)

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

Χρησιμοποιείται όταν συγκρίνουμε την μεταβλητότητα χαρακτηριστικών,

- 1) Με διαφορετικές μονάδες μέτρησης (gr, m, °C, lt, ...)
- 2) Με διαφορετικές κλίμακες (1000m, 10m or 0.01 m)

Παράδειγμα

Έγιναν δύο τεστ σε κάποιο δείγμα.

Στο πρώτο τεστ είχαμε

μέση τιμή = 32 και τυπική απόκλιση = 6

Στο δεύτερο τεστ είχαμε

μέση τιμή = 84 και τυπική απόκλιση = 12

Σε ποιο τεστ τα αποτελέσματα μας παρουσίασαν μεγαλύτερη μεταβλητότητα;

- $CV1 = 6/32 \cdot 100 = 18,75\%$
- $CV2 = 12/84 \cdot 100 = 14,29\%$

Το πρώτο τεστ αν και είχε πιο μικρή τυπική απόκλιση, παρουσίασε πιο μεγάλη μεταβλητότητα

Επειδή τα τεστ ήταν σε διαφορετικές κλίμακες, ο μόνος τρόπος για να συγκρίνουμε μεταβλητότητες είναι ο CV

Παρατήρηση: Όταν ο CV είναι πιο μικρός από 10% το δείγμα μας είναι αρκετά Ομοιογενές

Εκτίμηση των πληθυσμιακών μεγεθών

Η πληθυσμιακή μέση τιμή
συμβολίζεται με μ

Η δειγματική μέση τιμή
EKTIMAEI
την πληθυσμιακή μέση τιμή

Η πληθυσμιακή διακύμανση
συμβολίζεται με σ^2

Η δειγματική διακύμανση S^2
EKTIMAEI
την πληθυσμιακή διακύμανση

Η πληθυσμιακή τυπική
απόκλιση
συμβολίζεται με σ

Η δειγματική τυπική S απόκλιση
EKTIMAEI
την πληθυσμιακή τυπική απόκλιση

Όλοι οι εκτιμητές αυτοί είναι “αμερόληπτοι”

Βαθμοί Ελευθερίας (β.ε / d.f.)

- $n-1$ είναι οι “**βαθμοί ελευθερίας**” της διακύμανσης
- Η διαίρεση με το $n-1$ αντί με το μέγεθος δείγματος n , κάνει την εκτίμηση της πληθυσμιακής διακύμανσης “**αμερόληπτη**”
- Αν διαιρέσουμε με το n υπο-εκτιμάμε την πραγματική πληθυσμιακή διακύμανση