

# Μάθημα 3<sup>ο</sup>

Περιγραφική Στατιστική



# Η Στατιστική είναι...

- Μια τυποποιημένη σειρά αναλυτικών μεθόδων, οι οποίες χρησιμοποιούνται από τον εκάστοτε ερευνητή για την ανάλυση των διαθέσιμων δεδομένων.
- Υπάρχουν δύο κύριοι τύποι στατιστικών τεχνικών
  - Η περιγραφική στατιστική (*descriptive*)
  - Η επαγωγική στατιστική (*inferential*)

# ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

- Χρησιμοποιείται για την οργάνωση και παρουσίαση των στατιστικών στοιχείων
  - Πίνακες και γραφήματα
  - Αριθμητικά περιγραφικά μέτρα
- Περιγράφονται τα κύρια χαρακτηριστικά των μεμονωμένων μεταβλητών
  - π.χ., ο υπολογισμός της μέσης ηλικίας ενός δείγματος ανθρώπων

# Πίνακες και γραφήματα

- Αποτελούν τις πιο αποτελεσματικές μεθόδους για την οργάνωση και την παρουσίαση των στοιχείων
  - Γρήγορη εμποπτική εικόνα των στοιχείων που πρόκειται να αναλυθούν
- Πίνακες: παρουσιάζονται βασικά στοιχεία, π.χ., πλήθος στοιχείων, συχνότητα, αριθμητικά περιγραφικά μέτρα
- Γραφήματα: οπτική αναπαράσταση των στοιχείων που περιέχονται στους πίνακες

# Πίνακες συχνοτήτων

- Συχνότητα ( $f$ ): το πλήθος των παρατηρήσεων που έχουν ένα συγκεκριμένο χαρακτηριστικό
  - Εάν  $x_i$  ( $i=1, \dots, n$ ) οι τιμές του χαρακτηριστικού  $x$  για τα  $n$  άτομα του δείγματος και  $y_j$  ( $j=1, \dots, k$ ,  $k \leq n$ ) οι  $k$  διαφορετικές τιμές από τα  $x_i$  τότε: συχνότητα  $f$  είναι το πλήθος των  $x_i$  που είναι ίσα με  $y_j$ 
    - π.χ., σε δείγμα 293 ατόμων, το πλήθος των συμμετεχόντων που είναι άνδρες είναι 108

➤ Σχετική συχνότητα ( $f_i$ ): το πηλίκο της συχνότητας εμφάνισης ενός χαρακτηριστικού προς το συνολικό αριθμό των μελών του δείγματος

- $f_i = v_i / n$

- π.χ., σε δείγμα 293 ατόμων η σχετική συχνότητα των ανδρών είναι 0,37 (ή το 37% του δείγματος αποτελείται από άνδρες)

- Αθροιστική συχνότητα: το **πλήθος** των παρατηρήσεων που το χαρακτηριστικό τους έχει τιμή μικρότερη ή ίση κάποιας συγκεκριμένης τιμής
- Αθροιστική σχετική συχνότητα: το **ποσοστό** των παρατηρήσεων που το χαρακτηριστικό τους έχει τιμή μικρότερη ή ίση κάποιας συγκεκριμένης τιμής

# Παράδειγμα

Προκύπτει από το άθροισμα των δύο προηγούμενων συχνοτήτων (δλδ., 6 άτομα στην κλίμακα 1 & 20 άτομα στην κλίμακα 2)

Κατανομή συχνοτήτων συμμετεχόντων σύμφωνα με το σκορ σωματικής άσκησης				
	Συχνότητα	Αθροιστική συχνότητα	Σχετική συχνότητα	Σχετική αθροιστική συχνότητα
Κλίμακα σωματικής άσκησης				
1	6	6	2.04778157	2.04778157
2	20	26	6.82593857	8.873720137
3	24	50	8.19112628	17.06484642
4	38	88	12.9692833	30.03412969
5	58	146	19.7952218	49.82935154
6	44	190	15.0170648	64.84641638
7	41	231	13.9931741	78.83959044
8	36	267	12.2866894	91.12627986
9	10	277	3.41296928	94.53924915
10	16	293	5.46075085	100
Σύνολο	293		100	



# Παράδειγμα

Προκύπτει ως το πηλίκο της συχνότητας προς το σύνολο (δλδ., 6 άτομα στην κλίμακα 1/ 293 συμμετέχοντες = 2.05)

Κατανομή συχνοτήτων συμμετεχόντων σύμφωνα με το σκορ σωματικής άσκησης

	Συχνότητα	Αθροιστική συχνότητα	Σχετική συχνότητα	Σχετική αθροιστική συχνότητα
Κλίμακα σωματικής άσκησης				
1	6	6	2.04778157	2.04778157
2	20	26	6.82593857	8.873720137
3	24	50	8.19112628	17.06484642
4	38	88	12.9692833	30.03412969
5	58	146	19.7952218	49.82935154
6	44	190	15.0170648	64.84641638
7	41	231	13.9931741	78.83959044
8	36	267	12.2866894	91.12627986
9	10	277	3.41296928	94.53924915
10	16	293	5.46075085	100
Σύνολο	293		100	

# Παράδειγμα

Προκύπτει ως το πηλίκο της αθροιστικής συχνότητας προς το σύνολο (δλδ., 26 άτομα στην κλίμακα 1 & 2/ 293 συμμετέχοντες = 8.87)

Κατανομή συχνοτήτων συμμετεχόντων σύμφωνα με το σκορ σωματικής άσκησης

	Συχνότητα	Αθροιστική συχνότητα	Σχετική συχνότητα	Σχετική αθροιστική συχνότητα
Κλίμακα σωματικής άσκησης				
1	6	6	2.04778157	2.04778157
2	20	26	6.82593857	8.873720137
3	24	50	8.19112628	17.06484642
4	38	88	12.9692833	30.03412969
5	58	146	19.7952218	49.82935154
6	44	190	15.0170648	64.84641638
7	41	231	13.9931741	78.83959044
8	36	267	12.2866894	91.12627986
9	10	277	3.41296928	94.53924915
10	16	293	5.46075085	100
Σύνολο	293		100	

# Γραφικές μέθοδοι

- Προτιμώνται έναντι των πινάκων για την παρουσίαση μεγάλου όγκου πληροφοριών
  - Οι πίνακες προτείνονται για την παρουσίαση αριθμητικών δεδομένων
  - Τα γραφήματα προτείνονται για την παρουσίαση πληροφοριών
- Ανάλογα με το είδος των δεδομένων που έχει ο κάθε ερευνητής στη διάθεσή του υπάρχουν και διάφοροι γραφικοί τρόποι παρουσίασης των δεδομένων

# ΜΕΤΑΒΛΗΤΕΣ

Συνεχείς  
μεταβλητές

Ιστογράμματα,  
θηκογράμματα

Κατηγορικές  
μεταβλητές

Κυκλικά  
διαγράμματα  
(μικρός αριθμός  
κατηγοριών)

Ραβδογράμματα  
(μεγάλος αριθμός  
κατηγοριών)

Διατάξιμες ή  
διακριτές  
μεταβλητές

Κυκλικά  
διαγράμματα  
(μικρός αριθμός  
κατηγοριών)

Ραβδογράμματα  
(μεγάλος αριθμός  
κατηγοριών)

# Ιστόγραμμα συχνοτήτων

- Χρησιμοποιείται για την παρουσίαση ποσοτικών συνεχών μεταβλητών
- Αποτελείται από διαδοχικά ορθογώνια, το ύψος των οποίων επιλέγεται έτσι ώστε το εμβαδόν του ορθογωνίου να είναι ίσο με την αντίστοιχη συχνότητα της τιμής που αναφέρεται
  - Το συνολικό εμβαδό είναι ίσο με το μέγεθος του δείγματος  $n$

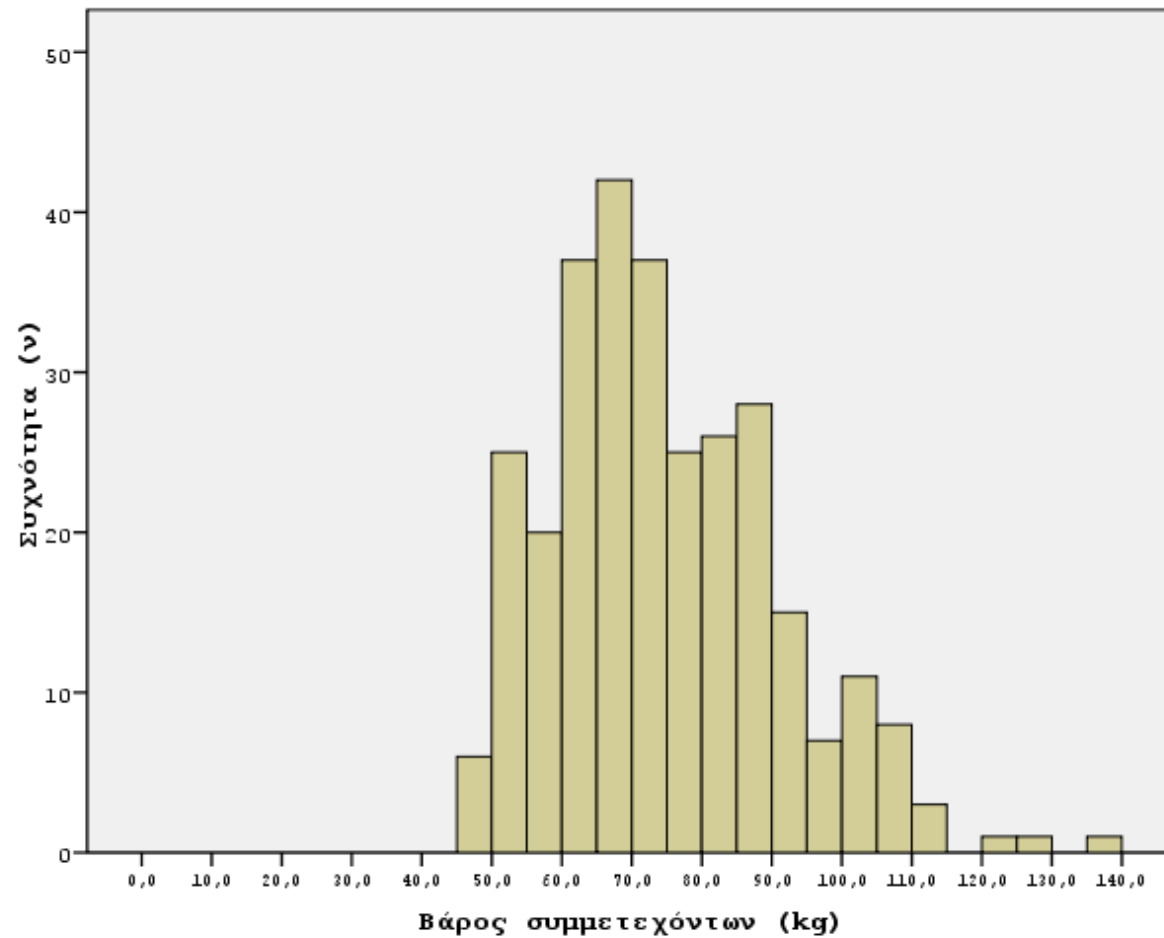
# Κυκλικό διάγραμμα

- Τα κυκλικά διαγράμματα (piecharts) χρησιμοποιούνται για την παράσταση των δεδομένων ένα κύκλο χωρισμένα σε κυκλικά τμήματα, τόσα όσα και οι κατηγορίες της μεταβλητής που χρησιμοποιήθηκε
- Κάθε κυκλικό τμήμα αναφέρεται σε μια κατηγορία και έχει τόξο ανάλογο της αντίστοιχης συχνότητας

# Ραβδόγραμμα

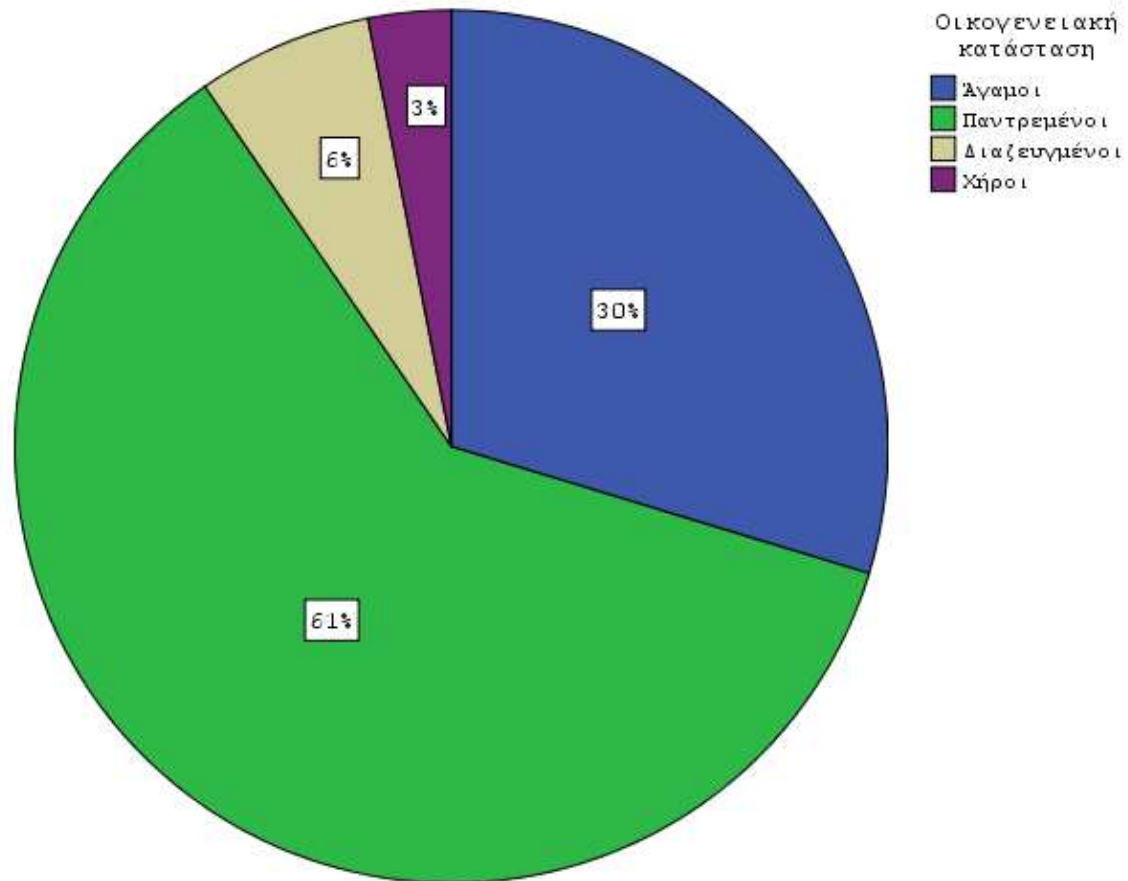
- Στο ραβδόγραμμα (barchart) οι κατηγορίες της μεταβλητής παρίστανται στον οριζόντιο άξονα ως ισομήκη διαστήματα (με κενά μεταξύ τους), ενώ στον κατακόρυφο παρίστανται οι συχνότητες
  - Αποτελείται από ορθογώνια παραλληλεπίπεδα ίσης βάσης με ύψη ίσα με τις συχνότητες της εκάστοτε κατηγορίας

Ιστογράμμα κατανομής συχνοτήτων του βάρους 293  
συμμετεχόντων στη μελέτη.

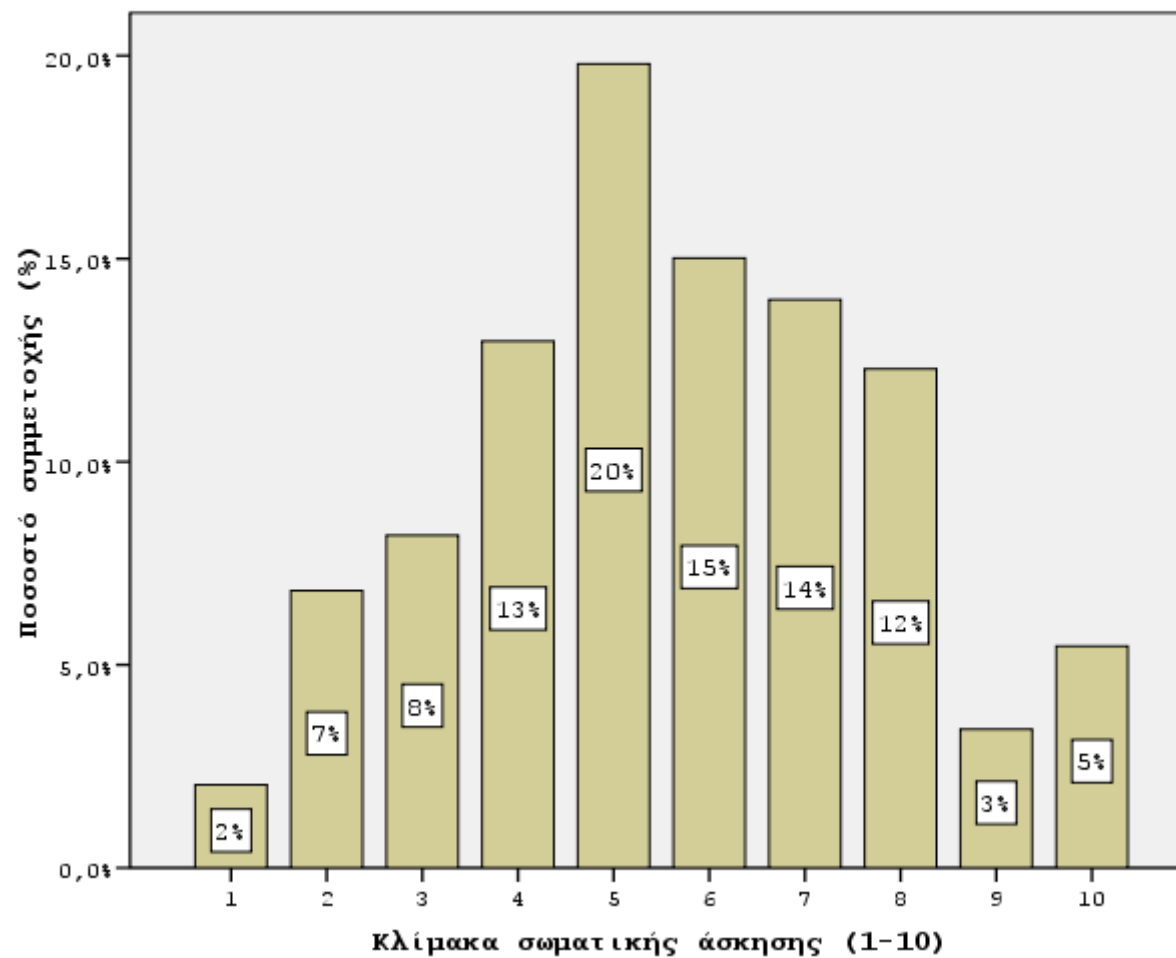




Κυκλικό διάγραμμα κατανομής της οικογενειακής κατάστασης των 293 συμμετεχόντων στη μελέτη.



Ραβδόγραμμα σχετικών συχνοτήτων της κατανομής της σωματικής άσκησης των 293 συμμετεχόντων



# Αριθμητικά περιγραφικά μέτρα

- Χρησιμοποιούνται για να δώσουμε μια συνοπτική εικόνα των διαθέσιμων δεδομένων χρησιμοποιώντας μικρό πλήθος αριθμητικών στοιχείων
  - Μέτρα κεντρικής τάσης
  - Μέτρα διασποράς
  - Μέτρα μεταβλητότητας

# Μέτρα κεντρικής τάσης

- Χρήσιμα για την περιγραφή της θέσης της κατανομής από όπου προέρχονται τα δεδομένα
- Πρόκειται για μεμονωμένες αριθμητικές τιμές οι οποίες θεωρούνται ως αντιπροσωπευτικές του συνόλου των τιμών μιας μεταβλητής
  - Μέση τιμή
  - Διάμεσος
  - Επικρατούσα τιμή

# Μέση τιμή (*mean*)

- Ορίζεται ως ο μέσος όρος όλων των τιμών της μεταβλητής στο δείγμα
  - Μαθηματικά είναι το άθροισμα των τιμών των παρατηρήσεων του δείγματος διαιρεμένο δια του πλήθους των παρατηρήσεων του δείγματος:

$$\bar{x}_i = \frac{1}{\nu} \sum_{i=1}^{\nu} x_i$$

## Πλεονεκτήματα

- Η αναμενόμενη τιμή που έχει κάποιο χαρακτηριστικό (μεταβλητή) στο δείγμα
- Καθορίζεται με αριθμητικές διαδικασίες
- Είναι μόνο μια σε κάθε κατανομή

## Μειονεκτήματα

- Δεν θεωρείται αντιπροσωπευτική όταν τα δεδομένα ακολουθούν μη συμμετρική κατανομή
- Επηρεάζεται από τις ακραίες παρατηρήσεις
- Δεν εφαρμόζεται σε ποιοτικές μεταβλητές

# Διάμεσος (median)

- Εκείνη η τιμή μιας μεταβλητής ( $\delta$ ) η οποία χωρίζει το δείγμα σε δύο ίσα μέρη έτσι ώστε ο αριθμός των παρατηρήσεων που είναι μικρότερες ή ίσες από την τιμή  $\delta$  να είναι ίσος με τον αριθμό των παρατηρήσεων που είναι μεγαλύτερες ή ίσες από την τιμή  $\delta$ 
  - Αν  $n$ : περιττός τότε  $\delta$ =κεντρική τιμή σε διατεταγμένα δεδομένα
  - Αν  $n$ : άρτιος τότε  $\delta$ =ημιάθροισμα των δύο κεντρικών τιμών σε διατεταγμένα δεδομένα

## Πλεονεκτήματα

- Σταθερό μέτρο κεντρικής θέσης
- Δεν επηρεάζεται από ακραίες παρατηρήσεις
- Είναι μια για κάθε κατανομή

## Μειονεκτήματα

- Χρονοβόρος τρόπος υπολογισμού
- Δεν ερμηνεύεται εύκολα όταν υπάρχουν δεδομένα που έχουν την ίδια τιμή με τη διάμεσο



# Επικρατούσα τιμή (mode)

- Εκείνη η τιμή η οποία εμφανίζεται με τη μεγαλύτερη συχνότητα στο δείγμα
  - Αν όλες οι τιμές είναι διαφορετικές τότε δεν υπάρχει επικρατούσα τιμή
  - Αν δύο ή περισσότερες τιμές εμφανίζονται με την ίδια συχνότητα τότε υπάρχουν δύο ή περισσότερες επικρατούσες τιμές

## Πλεονεκτήματα

- Χρήσιμο για την περιγραφή ποιοτικών στοιχείων, όπου ο μέσος και η διάμεσος δεν ορίζονται
- Εύκολος & γρήγορος υπολογισμός

## Μειονεκτήματα

- Δεν υπολογίζεται μαθηματικά
  - Δεν προσφέρεται σε μαθηματικούς χειρισμούς
- Μπορεί να μην υπάρχει ή να είναι περισσότερες από μια

# Μέτρα διασποράς

- Απαραίτητα για την πλήρη περιγραφή των χαρακτηριστικών ενός πληθυσμού
- Αντανakλούν τη διαφορά στις τιμές μιας μεταβλητής από άτομο σε άτομο
- Εκφράζουν τις αποκλίσεις των τιμών μιας μεταβλητής γύρω από τα μέτρα κεντρικής θέσης
  - Διακύμανση
  - Τυπική απόκλιση
  - Ενδοτεταρτημοριακό εύρος
  - Εύρος μεταβολής

# Διακύμανση (variance)

- Πρόκειται για το πιο διαδεδομένο μέτρο διασποράς
  - Όταν οι τιμές ενός συνόλου παρατηρήσεων δε διαφέρουν πολύ από τη μέση τιμή τότε η διασπορά είναι μικρή
  - Όταν οι τιμές είναι σκορπισμένες σε μεγάλη απόσταση από τη μέση τιμή τότε η διασπορά μεγαλώνει
- Ορίζεται από τη σχέση: 
$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

# Τυπική απόκλιση (standard deviation)

- Η διακύμανση ορίζεται σε μονάδα ίση με το τετράγωνο της αρχικής μονάδας μέτρησης του χαρακτηριστικού
- Η τυπική απόκλιση ορίζεται ως η τετραγωνική ρίζα της διακύμανσης
- Μέτρο μεταβλητότητας που εκφράζεται στη μονάδα μέτρησης του χαρακτηριστικού

# Σημειώνεται ότι...

- Στο διάστημα  $[\mu - \sigma, \mu + \sigma]$  περιλαμβάνεται περίπου το 68% των παρατηρήσεων
- Στο διάστημα  $[\mu - 2\sigma, \mu + 2\sigma]$  περιλαμβάνεται περίπου το 95% των παρατηρήσεων
- Στο διάστημα  $[\mu - 3\sigma, \mu + 3\sigma]$  περιλαμβάνεται περίπου το 99% των παρατηρήσεων

# Ενδοτεταρτημοριακό εύρος (Interquartile range, IQR)

- Η διαφορά του 1<sup>ου</sup> (Q1) από το 3<sup>ο</sup> (Q3) τεταρτημόριο
  - Q1: η τιμή μιας μεταβλητής κάτω από την οποία βρίσκεται το 25% των παρατηρήσεων
  - Q3: η τιμή μιας μεταβλητής κάτω από την οποία βρίσκεται το 75% των παρατηρήσεων
- Μεταξύ των δύο περιέχεται το 50% των παρατηρήσεων
  - Μικρό διάστημα → μεγάλη συγκέντρωση τιμών → μικρή διασπορά των τιμών της μεταβλητής

# Γίνεται φανερό ότι...

- Η διακύμανση και η τυπική απόκλιση ενός συνόλου τιμών είναι μέτρα διασποράς γύρω από το μέσο, ενώ η ενδοτεταρτημοριακή απόσταση είναι μέτρο διασποράς γύρω από τη διάμεσο



# Εύρος μεταβολής (range)

- Ορίζεται ως η διαφορά της ελάχιστης τιμής από τη μέγιστη τιμή, των τιμών μιας μεταβλητής
  - $R = X_{\max} - X_{\min}$
- Βασίζεται μόνο στις ακραίες τιμές

# Μέτρα μεταβλητότητας

- Συντελεστής μεταβλητότητας (coefficient of variance, CV)
- Μέτρο σχετικής μεταβλητότητας
- Εκφράζεται ως ποσοστό:  
$$CV = (\text{τυπική απόκλιση} / \text{μέση τιμή}) * 100\%$$
- Μπορεί να χρησιμοποιηθεί για συγκρίσεις ομάδων τιμών οι οποίες είτε εκφράζονται σε διαφορετικές μονάδες μέτρησης, είτε έχουν ίδιες μονάδες μέτρησης, αλλά διαφορετικές μέσες τιμές

Πραγματοποίηση αυτών στο  
SPSS...

