

**ΣΤΑΤΙΣΤΙΚΗ ΕΠΙΧΕΙΡΗΣΕΩΝ**  
**ΘΟΔΩΡΟΣ ΠΑΠΑΔΟΓΓΟΝΑΣ**

**Σετ Διαφανειών 1: Περιγραφική  
Στατιστική, Μέτρα θέσης - Διασποράς,  
Κύρτωση - Ασυμμετρία**

# Βασικές Έννοιες

- **Περιγραφική Στατιστική:** συνοπτική παρουσίαση των δεδομένων. Δεν προβαίνει σε γενικεύσεις.
- **Επαγωγική Στατιστική:** με βάση περιορισμένα διαθέσιμα στοιχεία (συνήθως ενός δείγματος) προβαίνει στην εξαγωγή και ανάλυση συμπερασμάτων.
- **Πληθυσμός:** το σύνολο των μονάδων που διαθέτουν κάποιο χαρακτηριστικό
- **Δείγμα:** ένα υποσύνολο ενός πληθυσμού
- **Παράμετρος:** ένα χαρακτηριστικό του πληθυσμού
- **Στατιστικό μέτρο (statistic):** ένα χαρακτηριστικό του δείγματος

# Τύποι Στατιστικών Δεδομένων

- **Ποιοτικά ή κατηγορικά δεδομένα**, οι τιμές των οποίων δεν είναι αριθμοί (π.χ. χρώμα αυτοκινήτων)
- **Ποσοτικά ή μετρήσιμα δεδομένα**, οι τιμές των οποίων είναι αριθμοί. Δύο επιμέρους κατηγορίες:
  - **διακριτές μεταβλητές**, οι τιμές των οποίων είναι ακέραιοι αριθμοί (π.χ. αριθμός εργαζομένων)
  - **συνεχείς μεταβλητές**, οι τιμές των οποίων είναι οποιοιδήποτε αριθμοί (π.χ. μισθοί εργαζομένων)

ΜΕΡΟΣ Α

*ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ*

# Σκοπός

- Να παρουσιάσουμε τα δεδομένα με απλό και κατανοητό τρόπο, με τη χρήση κατάλληλων περιγραφικών μέτρων και διαγραμμάτων, συνοψίζοντας τις πληροφορίες που περιέχουν.
- Έτσι λαμβάνουμε μια εικόνα των δεδομένων και μπορούμε να κάνουμε σύγκριση ανάμεσα σε διαφορετικά σύνολα δεδομένων.

# ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

- Γραφικές Μέθοδοι
- Αριθμητικές Μέθοδοι

# ΓΡΑΦΙΚΕΣ ΜΕΘΟΔΟΙ

- Πίνακας Κατανομής Συχνοτήτων
- Ραβδόγραμμα
- Ιστόγραμμα
- Κυκλικό Διάγραμμα

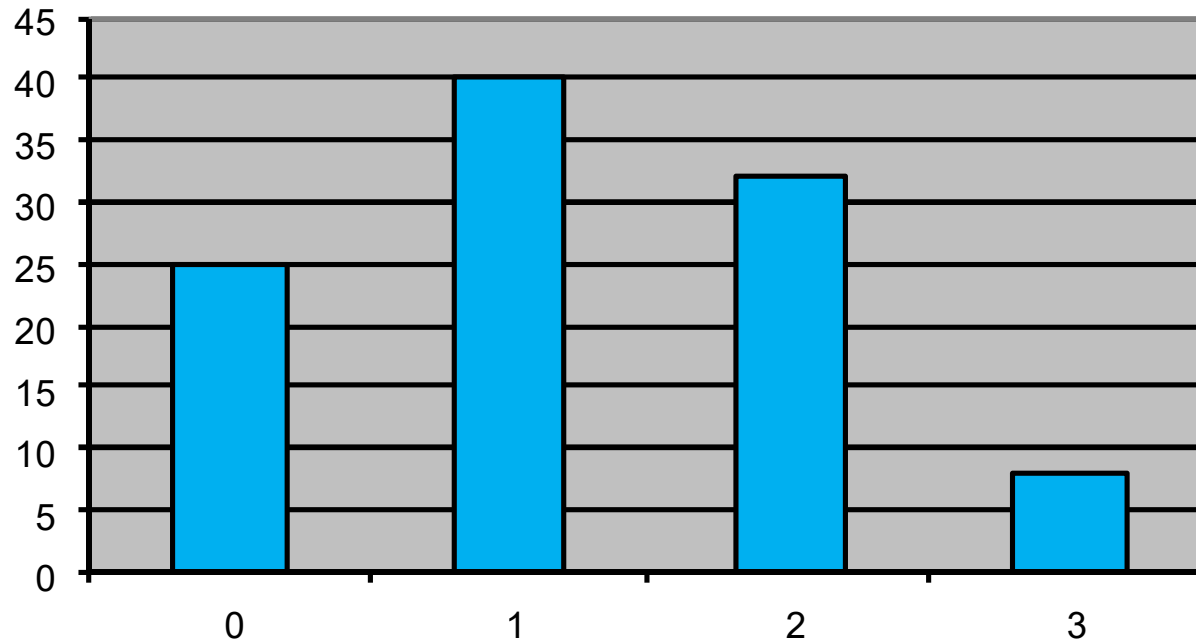
# Πίνακας Κατανομής Συχνοτήτων

	Τιμές	Συχνότητα	Σχετ. Συχν.
	$x_i$	$v_i$	$f_i$
Παιδιά στην οικογένεια	0	25	25/100=0,25
	1	40	0,40
	2	32	0,32
	3	8	0,08
Σύνολο		100	1,0

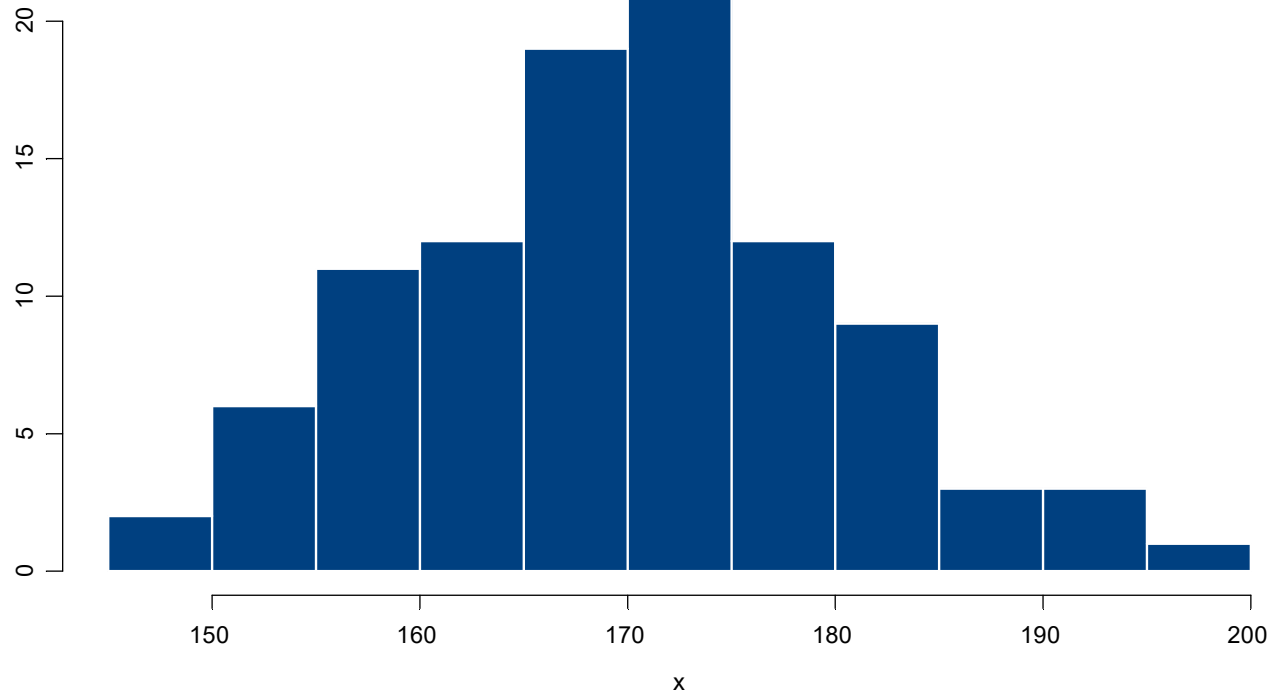


# Ραβδόγραμμα

Παιδιά στην οικογένεια



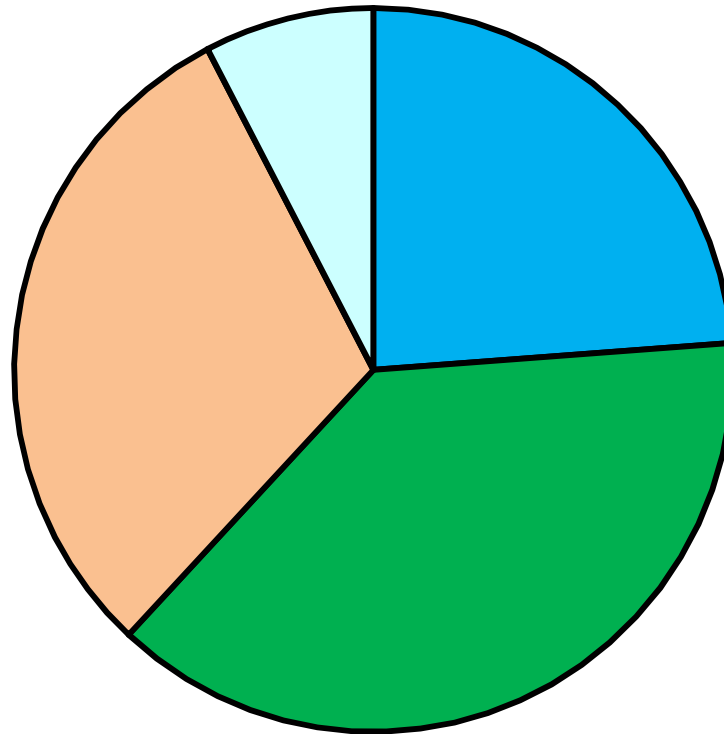
# Ιστόγραμμα



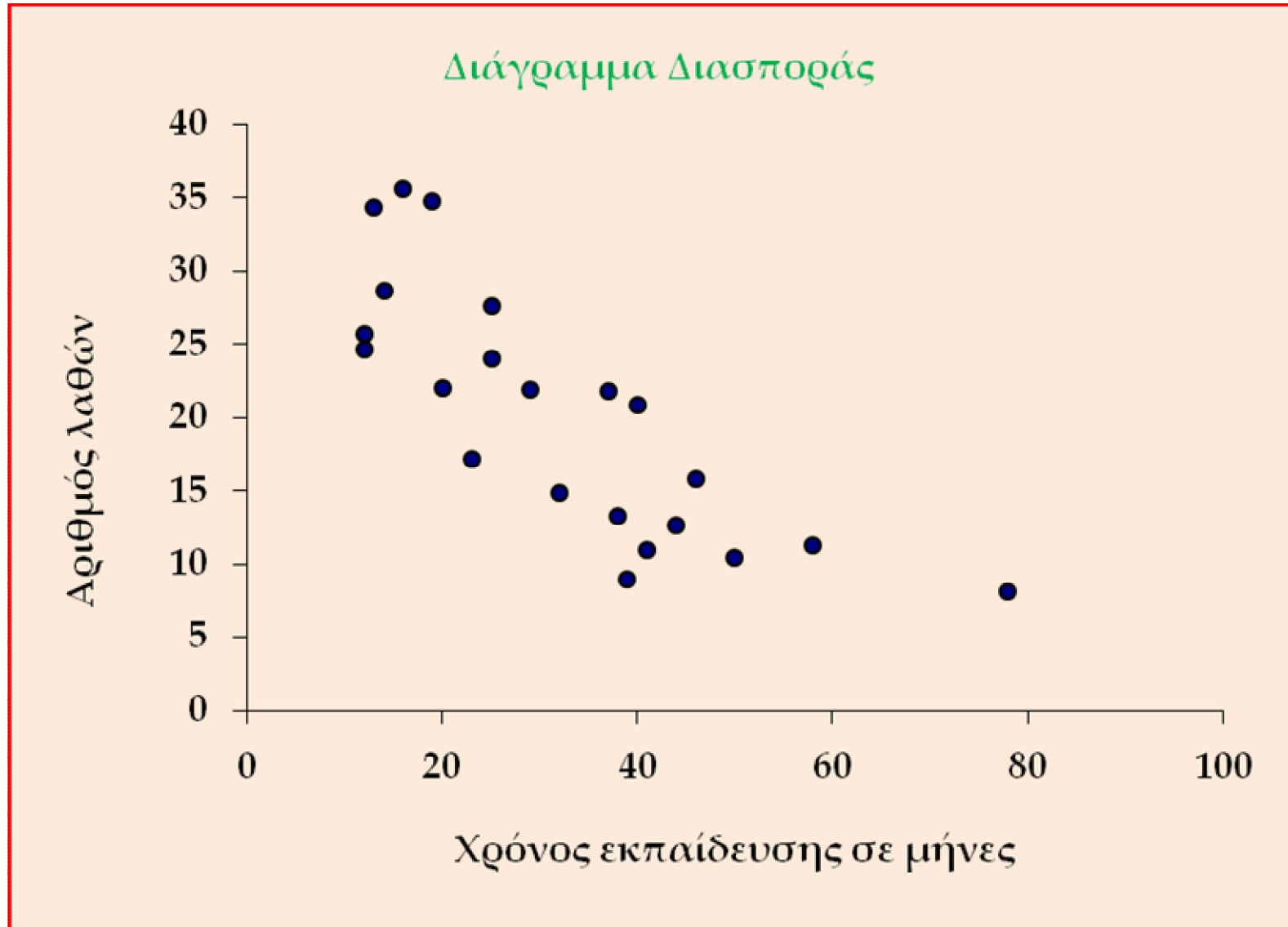
Ιστόγραμμα τιμών ύψους 200 μαθητών (ποσοστά)

# Κυκλικό Διάγραμμα

Παιδιά ανά οικογένεια

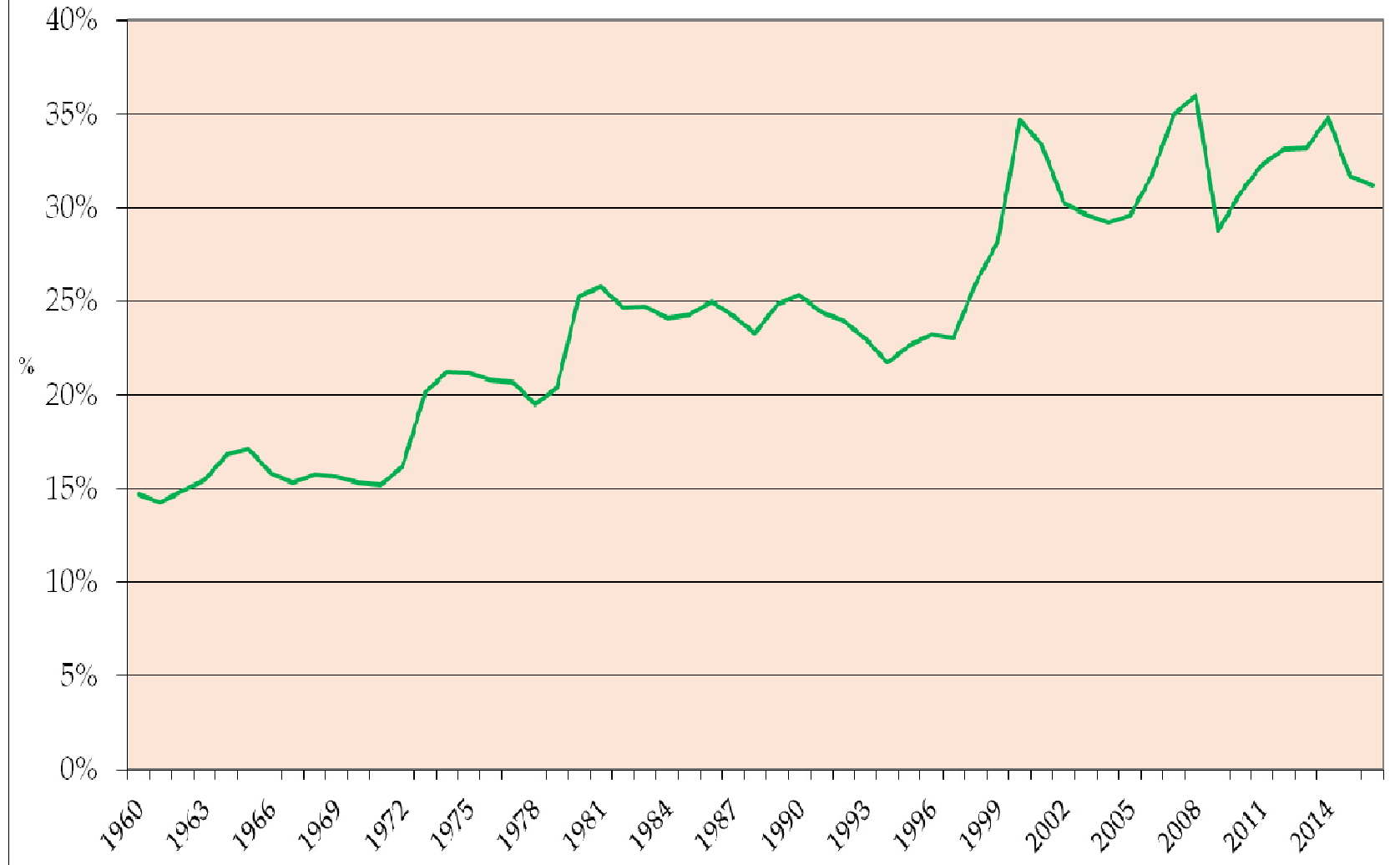


# Διάγραμμα διασποράς

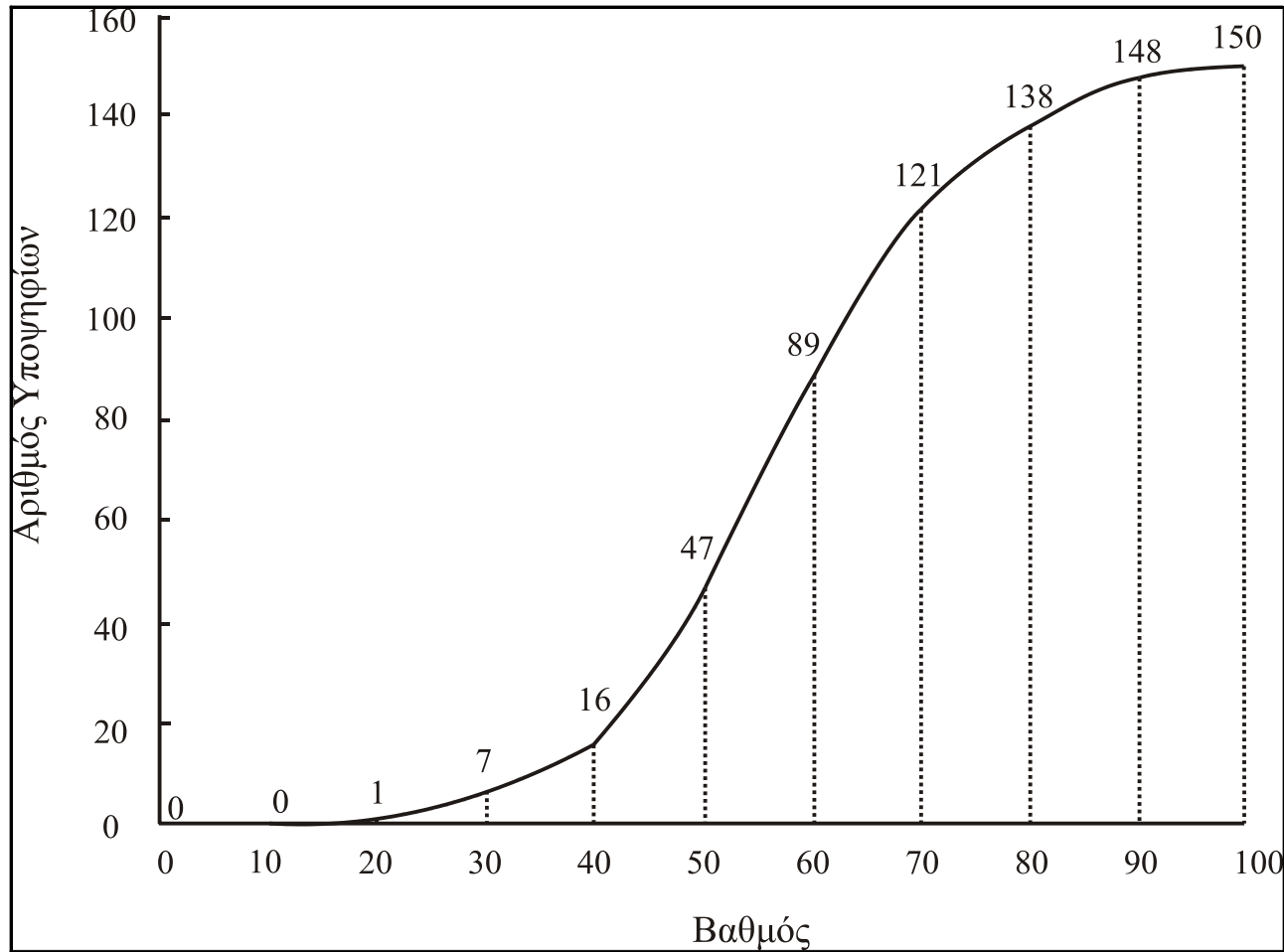


# Γραμμογράφημα

Εξέλιξη των Εισαγωγών ως ποσοστό του ΑΕΠ της Ελλάδας



# Αθροιστική κατανομή συχνοτήτων



# Αριθμητικές Μέθοδοι

- Μέτρα Κεντρικής Τάσης
- Μέτρα Σχετικής Θέσης
- Μέτρα Διασποράς
- Μέτρα Ασυμμετρίας

# Μέτρα Κεντρικής Τάσης

- Αριθμητικός Μέσος  
Ο μέσος όρος

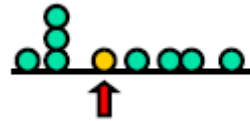
$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$



# Μέτρα Κεντρικής Τάσης

- Διάμεσος

Η παρατήρηση που βρίσκεται ακριβώς στο μέσο των διατεταγμένων δεδομένων

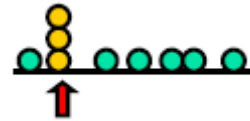


$$n = 2k + 1 \quad M = X_{\frac{n+1}{2}}$$

$$n = 2k \quad M = \frac{1}{2} \left( X_{\frac{n}{2}+1} + X_{\frac{n}{2}} \right)$$

# Μέτρα Κεντρικής Τάσης

- Επικρατούσα Τιμή  
Η παρατήρηση με την  
υψηλότερη συχνότητα



$$T_0$$

# Παράδειγμα 1

## Τιμές Ακινήτων

1000000

400000

300000

150000

100000

100000

50000

**Άθροισμα: 2100000**

# Παράδειγμα 1

## Τιμές Ακινήτων

1000000

400000

300000

150000

100000

100000

50000

**Άθροισμα:** 2100000

**Αρ. Μέσος:**  $2100000 / 7 = 300000$

# Παράδειγμα 1

## Τιμές Ακινήτων

1000000

400000

300000

150000

100000

100000

50000

**Άθροισμα:** 2100000

**Αρ. Μέσος:**  $2100000 / 7 = 300000$

**Διάμεσος** = 150000

**Επ. Τιμή** = 100000

## Παράδειγμα 2

Δίνονται παρακάτω οι τελικές βαθμολογίες ενός τυχαίου δείγματος 20 φοιτητών του ΕΚΠΑ που έχουν εγγραφεί στο μάθημα «Διοίκηση Επιχειρησεων» κατά το ακαδημαϊκό έτος 2020-21:

1,5 – 5 – 3 – 6 – 3 – 6,5 – 6 – 1 – 2 – 2,5 – 4 – 5 – 3 –  
2 – 7 – 2 – 5 – 3 – 8 – 2,5

1,5 - 5 - 3 - 6 - 3 - 6,5 - 6 - 1 - 2 - 2,5 - 4 - 5 - 3 - 2 - 7 - 2 -  
5 - 3 - 8 - 2,5

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{78}{20} = 3,9$$

1,5 - 5 - 3 - 6 - 3 - 6,5 - 6 - 1 - 2 - 2,5 - 4 - 5 - 3 - 2 - 7 - 2 -  
5 - 3 - 8 - 2,5

1	1,5	2	2	2	2,5	2,5	3	3	3	3	4	5	5	5	6	6	6,5	7	8
---	-----	---	---	---	-----	-----	---	---	---	---	---	---	---	---	---	---	-----	---	---

$$M = \frac{1}{2} \left( X_{\frac{n}{2}+1} + X_{\frac{n}{2}} \right) = (3 + 3) = 3$$

$$T_0 = 3$$



# Μέτρα Σχετικής Θέσης

## Τεταρτημόρια

Τα σημεία που χωρίζουν τα δεδομένα σε 4 ίσα μέρη ( $Q_1, Q_2, Q_3$ )

$Q_1$  = η τιμή εκείνη για την οποία το 25% των παρατηρήσεων είναι μικρότερες από την τιμή αυτή και το 75% των παρατηρήσεων είναι μεγαλύτερες από την τιμή αυτή

$Q_2$  = η τιμή εκείνη για την οποία το 50% των παρατηρήσεων είναι μικρότερες από την τιμή αυτή και το 50% των παρατηρήσεων είναι μεγαλύτερες από την τιμή αυτή (ΔΙΑΜΕΣΟΣ).

$Q_3$  = η τιμή εκείνη για την οποία το 75% των παρατηρήσεων είναι μικρότερες από την τιμή αυτή και το 25% των παρατηρήσεων είναι μεγαλύτερες από την τιμή αυτή.

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

Όπου  $A_Q$  ( $\Delta_Q$ ) είναι το ακέραιο (δεκαδικό) μέρος του πηλίκου  $i(n+1)/4$  ( $i = 1, 2, 3$ )

1	1,5	2	2	2	2,5	2,5	3	3	3	3	4	5	5	5	6	6	6,5	7	8
---	-----	---	---	---	-----	-----	---	---	---	---	---	---	---	---	---	---	-----	---	---

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

Το 1ο τεταρτημόριο είναι η τιμή που βρίσκεται στη θέση

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

Στη συγκεκριμένη περίπτωση  $A_Q = 5$  και  $\Delta_Q = 0,25$

$$Q_1 = X_5 + 0,25 [2,5 - 2] = 2 + 0,125 = 2,125$$

1	1,5	2	2	2	2,5	2,5	3	3	3	3	4	5	5	5	6	6	6,5	7	8
---	-----	---	---	---	-----	-----	---	---	---	---	---	---	---	---	---	---	-----	---	---

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

Το 2ο τεταρτημόριο είναι η τιμή που βρίσκεται στη θέση  
 $(20+1)*2/4=10,5$

$$Q_2 = X_{10} + 0,5[3 - 3] = 3 + 0 = 3$$

Το 3ο τεταρτημόριο είναι η τιμή που βρίσκεται στη θέση  
 $(20+1)*3/4=15,75$

$$Q_3 = X_{15} + 0,75[6 - 5] = 5 + 0,75 = 5,75$$

# ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ

## Εύρος

Το εύρος των παρατηρήσεων

$$R = X_{\max} - X_{\min}$$

## Ενδοτεταρτημοριακό Εύρος

Το εύρος του “κεντρικού” 50% των παρατηρήσεων

$$IR = Q_3 - Q_1$$

## Διακύμανση

Σταθμισμένη τετραγωνική απόσταση παρατηρήσεων από το μέσο

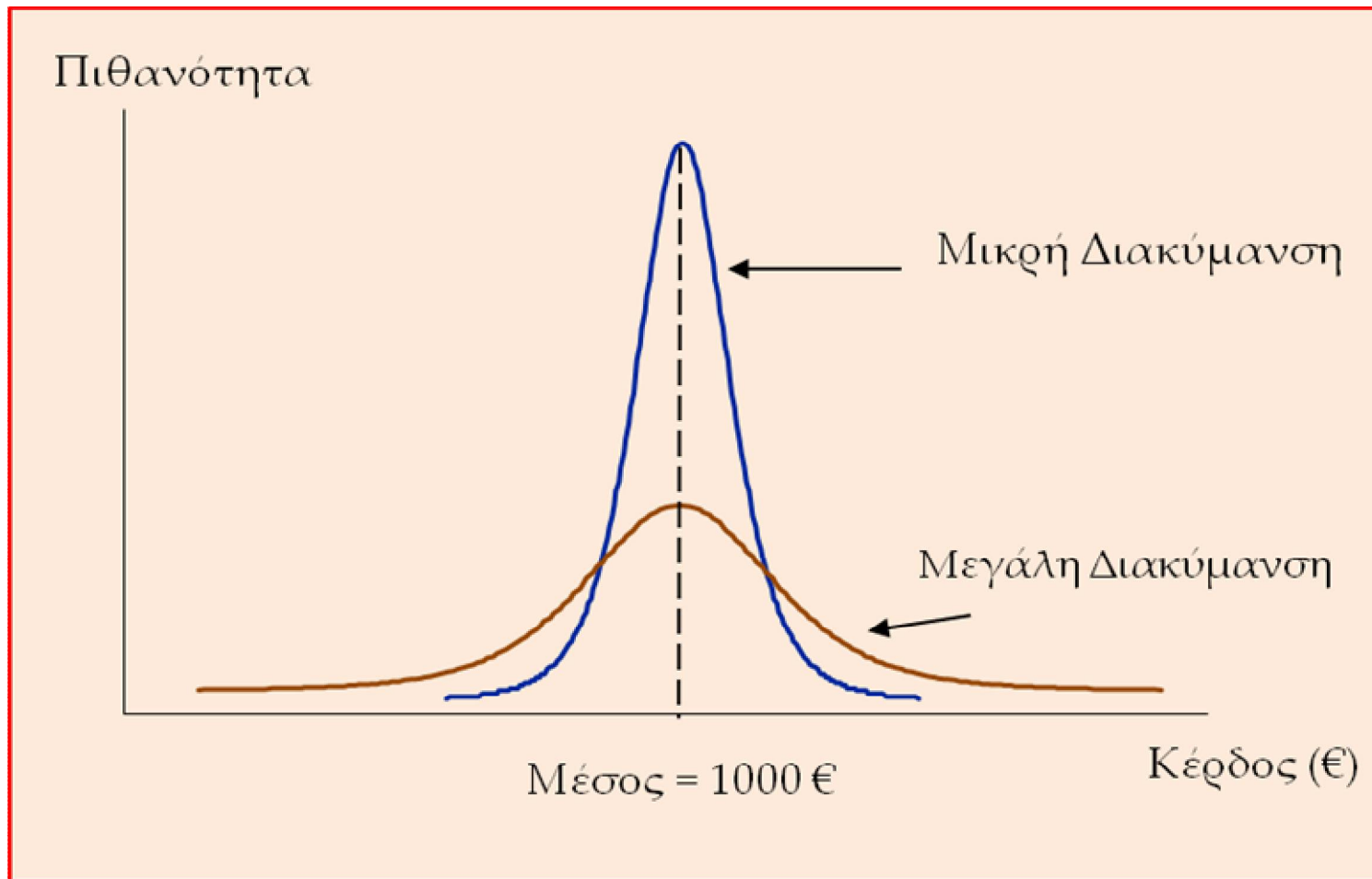
$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

## Τυπική Απόκλιση

Θετική τετραγωνική ρίζα της διακύμανσης

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

# ΔΙΑΣΠΟΡΑ



Όσο μικρότερη είναι η διακύμανση τόσο πιο συγκεντρωμένες είναι οι τιμές της μεταβλητής γύρω από τον μέσο

1	1,5	2	2	2	2,5	2,5	3	3	3	3	4	5	5	5	6	6	6,5	7	8
---	-----	---	---	---	-----	-----	---	---	---	---	---	---	---	---	---	---	-----	---	---

$$R = X_{\max} - X_{\min} = 8 - 1 = 7$$

$$IR = Q_3 - Q_1 = 5,75 - 2,125 = 3,625$$

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{77,8}{19} = 4,09$$

$$S = +\sqrt{S^2} = \sqrt{4,09} = 2,02$$

# Συντελεστής μεταβλητότητας

$$CV = \frac{S}{\bar{X}}$$

- Μέτρο σχετικής μεταβλητότητας
- Μετρά τη διασπορά των δεδομένων σε σχέση με το μέσο αυτών
- Εκφράζεται συνήθως ως ποσοστό (%)
- Κατάλληλο για σύγκριση μεταβλητών που έχουν διαφορετικούς μέσους και διακυμάνσεις

## Σύγκριση συντελεστών μεταβλητότητας

<b>Μετοχή</b>	Μέση Τιμή	Τυπική Απόκλιση
<b>A</b>	200€	40€
<b>B</b>	\$500	\$80

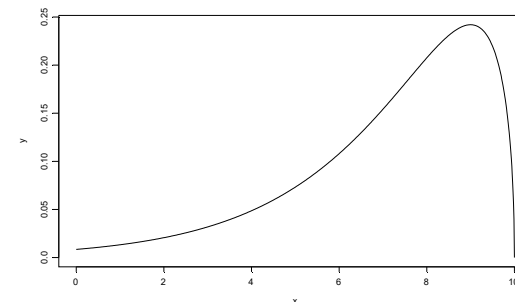
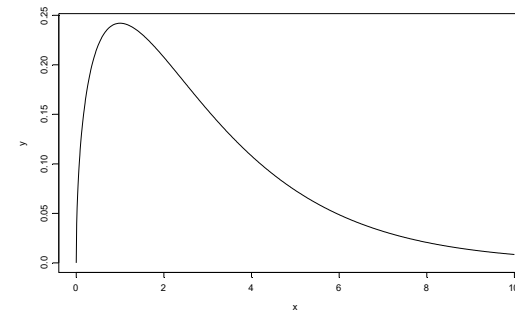
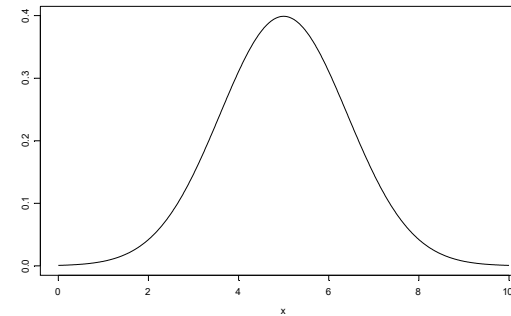


## Σύγκριση συντελεστών μεταβλητότητας

<b>Μετοχή</b>	Μέση Τιμή	Τυπική Απόκλιση	<b>100×CV</b>
<b>A</b>	200€	40€	$100 \times 40\text{€}/200\text{€}$ <b>= 20%</b>
<b>B</b>	\$500	\$80	$100 \times \$80/\$500$ <b>= 16%</b>

# Συμμετρία - Ασυμμετρία

- Συμμετρική κατανομή
- Κατανομή με δεξιά ασυμμετρία
- Κατανομή με αριστερή ασυμμετρία



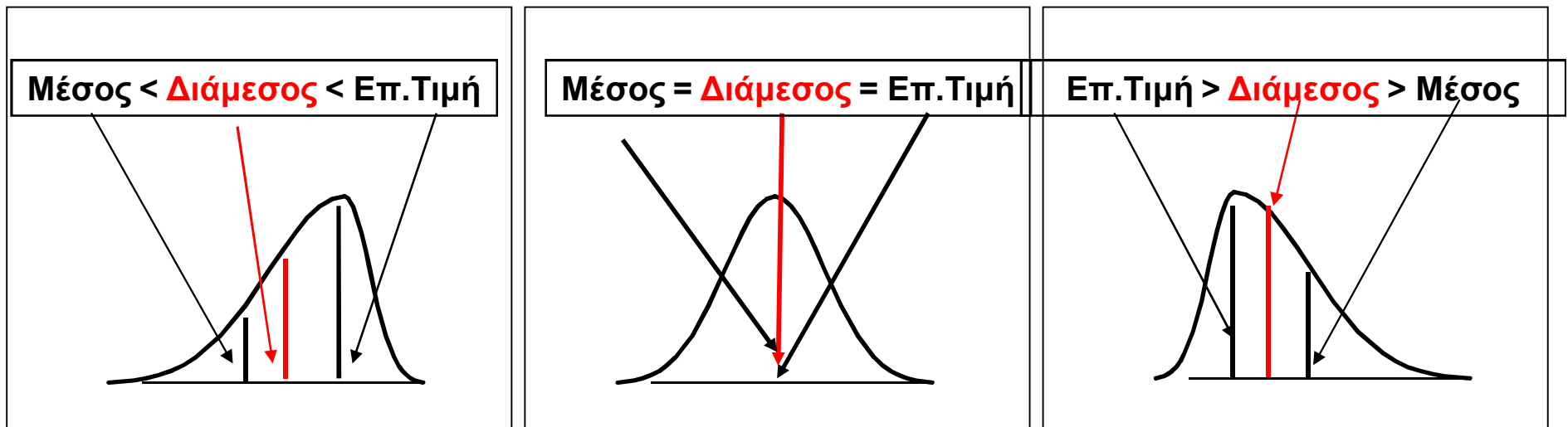
# Συντελεστής ασυμμετρίας

- Περιγράφει τον τρόπο που κατανέμονται οι τιμές της κατανομής. 3 περιπτώσεις:
  - Συμμετρική ή ασύμμετρη κατανομή
  - Αν η κατανομή είναι ασύμμετρη ενδέχεται να παρουσιάζει είτε αρνητική ασυμμετρία ή θετική ασυμμετρία

Αρνητική Ασυμμετρία

Συμμετρική Κατανομή

Θετική Ασυμμετρία



# Συντελεστής ασυμμετρίας

$$\beta_3 = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^3}{S^3}$$

**Spearman**

$$S_p = \frac{\bar{X} - T_0}{S}$$

$$S_p = \frac{3(\bar{X} - M)}{S}$$

- Θετική ασυμμετρία, εάν  $\beta_3$  ή  $S_p > 0$
- Συμμετρία, εάν  $\beta_3$  ή  $S_p = 0$
- Αρνητική ασυμμετρία, εάν  $\beta_3$  ή  $S_p < 0$

# Συντελεστής κύρτωσης

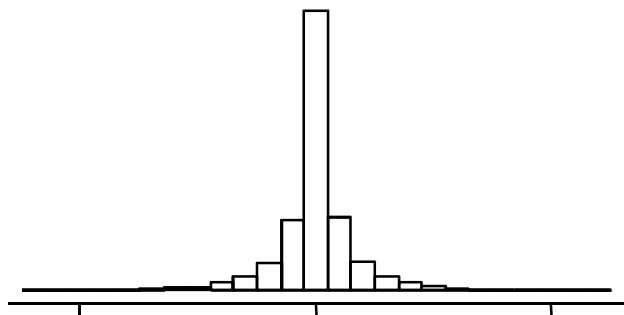
- Μετράει την “οξύτητα” της κορυφής της κατανομής, δηλαδή πόσο μεγάλη είναι η κορυφή (στο σημείο της επικρατούσας τιμής) σε σχέση με τις υπόλοιπες τιμές

$$Kyrtnosis = \beta_4 = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4}{S^4}$$

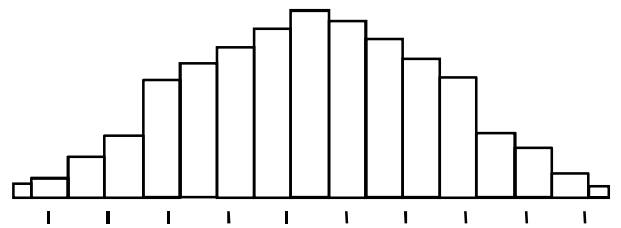
- Εάν  $\beta_4 < 3$  τότε έχουμε πλατύκυρτη κατανομή (χαμηλή κορυφή)
- Εάν  $\beta_4 = 3$  τότε έχουμε μεσόκυρτη κατανομή (κανονική κατανομή)
- Εάν  $\beta_4 > 3$  τότε έχουμε λεπτόκυρτη κατανομή (οξεία κορυφή)

# Συντελεστής κύρτωσης

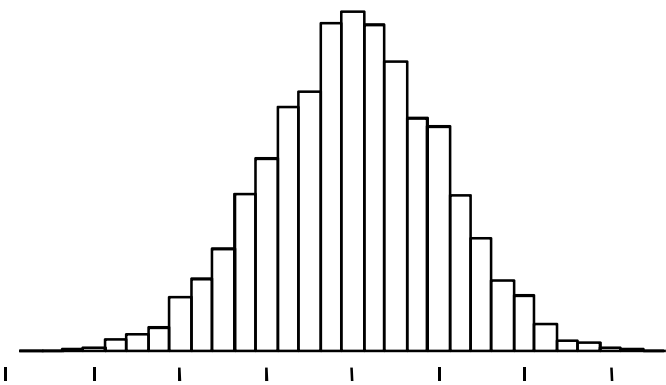
Λεπτόκυρτη κατανομή



Πλατύκυρτη κατανομή



Μεσόκυρτη κατανομή



# Υπολογισμός μέσω του Excel

Εργαλεία → Ανάλυση Δεδομένων → Περιγραφικά Στατιστικά →  
Περιληπτικά Στατιστικά.

Μέσος	3.9
Τυπικό σφάλμα	0.452478554
Διάμεσος	3
Επικρατούσα τιμή	3
Μέση απόκλιση τετραγώνου	2.023545612
Διακύμανση	4.094736842
Κύρτωση	1.879929512
Ασυμμετρία	0.505403877
Εύρος	7
Ελάχιστο	1
Μέγιστο	8
Άθροισμα	78
Πλήθος	20

(επισημαίνεται ότι το Excel χρησιμοποιεί διαφορετικές συναρτήσεις για την ασυμμετρία και την κύρτωση από αυτές που περιγράφηκαν παραπάνω)

# Ομαδοποιημένα Δεδομένα

- Όταν το πλήθος των τιμών μιας μεταβλητής είναι μεγάλο τα δεδομένα ταξινομούνται σε μικρό αριθμό ομάδων, που ονομάζονται τάξεις. Τα άκρα των τάξεων καλούνται όρια των τάξεων.
- Αν έχουμε ορίσει  $m$  τάξεις τότε  $c_i$  είναι το κέντρο της κάθε τάξης και  $f_i$  είναι η συχνότητα των παρατηρήσεων στη τάξη  $i$



# Ομαδοποιημένα δεδομένα - Παράδειγμα

Με βάση τα δεδομένα του παρακάτω πίνακα να υπολογιστούν ο αριθμητικός μέσος και η τυπική απόκλιση

Ορια τάξης	(συχνότητα) $f_i$
16-20	38
21-25	57
26-30	56
31-35	45
36-40	42
41-45	40
46-50	38
51-55	15
56-60	18

# Ομαδοποιημένα δεδομένα - Παράδειγμα

Για τον υπολογισμό των μέτρων αυτών χρειαζόμαστε τα εξής βοηθητικά στοιχεία:

Ορια τάξης	(συχνότητα) $f_i$	(κέντρο τάξης) $c_i$	$f_i c_i$	$c_i^2$	$f_i c_i^2$
16-20	38	18	684	324	12312
21-25	57	23	1311	529	30153
26-30	56	28	1568	784	43904
31-35	45	33	1485	1089	49005
36-40	42	38	1596	1444	60648
41-45	40	43	1720	1849	73960
46-50	38	48	1824	2304	87552
51-55	15	53	795	2809	42135
56-60	18	58	1044	3364	60552
<b>Σύνολο</b>	<b>349</b>		<b>12027</b>	<b>14496</b>	<b>460221</b>

# Ομαδοποιημένα δεδομένα - Παράδειγμα

$$\bar{X} = \frac{\sum_{i=1}^m f_i c_i}{n} \quad \bar{X} = \frac{12027}{349} = 34,46$$

$$S^2 = \frac{\sum_{i=1}^m f_i c_i^2}{n} - \bar{X}^2 \quad S^2 = \frac{460221}{349} - 34,46^2 = 131,1$$

$$s = +\sqrt{S^2} = 11,45$$