

Ποσοτικές Μέθοδοι

Δρ. Χαίδω Δριτσάκη

MSc Τραπεζική & Χρηματοοικονομική



Τεχνολογικό Εκπαιδευτικό Ίδρυμα Δυτικής Μακεδονίας
Western Macedonia University of Applied Sciences
Κοίλα Κοζάνης 50100 | Kozani GR 50100

Εισαγωγή σε βασικές έννοιες Στατιστικής

Ορισμός

Στατιστική (Statistics) είναι ένα σύνολο αρχών και μεθοδολογιών για:

- A. το σχεδιασμό της διαδικασίας συλλογής δεδομένων
- B. τη συνοπτική και αποτελεσματική παρουσίασή τους
- Γ. την εξαγωγή αντίστοιχων συμπερασμάτων

Η Στατιστική επιτυγχάνει τη συλλογή, επεξεργασία, παρουσίαση και ανάλυση των στατιστικών στοιχείων (αριθμητικών δεδομένων) με την εφαρμογή κατάλληλων για κάθε περίπτωση στατιστικών μεθόδων, οι οποίες και συνιστούν το περιεχόμενό της.

Πληθυσμός (Population)

Κάθε σύνολο αντικειμένων ή ατόμων που έχουν κάποιο κοινό μετρήσιμο χαρακτηριστικό

Δείγμα (Sample)

Κάθε υποσύνολο του πληθυσμού.

Τυχαίο δείγμα (random sample) είναι το δείγμα του πληθυσμού, όπου τα άτομα διαλέγονται το ένα μετά το άλλο, με κύριο χαρακτηριστικό, ότι τα υπόλοιπα άτομα του πληθυσμού κάθε φορά, έχουν τις ίδιες πιθανότητες να περιληφθούν στο τυχαίο δείγμα.

Η δειγματοληψία μας επιτρέπει τη συλλογή πληροφοριών γρήγορα και με μικρό κόστος. Είναι φανερό ότι η γενίκευση των πληροφοριών του δείγματος σε ολόκληρο τον πληθυσμό συνεπάγεται κάποιο σφάλμα το οποίο ελαχιστοποιείται όταν το δείγμα είναι **αντιπροσωπευτικό**.

- Οι όροι πληθυσμός και δείγμα μπορεί να αναφέρονται είτε στα άτομα, είτε στις μετρήσεις του κοινού χαρακτηριστικού τους. Τότε υπάρχει μια **κατανομή** των μετρήσεων του δείγματος, η οποία συνήθως μελετάται και μια κατανομή των μετρήσεων όλου του πληθυσμού που συνήθως υπάρχει αλλά είναι δύσκολο να προσδιοριστεί.
- Ένα ενδιαφέρον πρόβλημα είναι η απόφαση για το τι είδος πληροφορίες δύναται να εξαχθούν για την κατανομή του πληθυσμού, από την παρατήρηση και τη μελέτη της κατανομής του τυχαίου δείγματος.

- Τα χαρακτηριστικά ή ιδιότητες των στατιστικών μονάδων ως προς τα οποία εξετάζουμε έναν πληθυσμό ονομάζονται **μεταβλητές (variables)**.
- **Διάκριση των μεταβλητών**
 - A. Ποιοτικές:** Χαρακτηρίζονται οι μεταβλητές που δεν επιδέχονται αριθμητική μέτρηση
 - B. Ποσοτικές**
 - i. Ασυνεχείς (ή Διακριτές) (Discrete variables) είναι εκείνες που μπορούν να λάβουν πεπερασμένο ή αριθμήσιμο πλήθος τιμών (η ένδειξη ενός ζαριού).
 - ii. Συνεχείς (Continuous variables) είναι εκείνες που μπορούν να πάρουν όλες τις τιμές ενός διαστήματος (βάρος, ύψος).

Περιγραφή Δεδομένων- Τρόποι Παρουσίασης

Τα στατιστικά δεδομένα πρέπει να παρουσιάζονται με τρόπο απλό και σαφή, έτσι ώστε να είναι εύκολη η κατανόησή τους από τον κάθε ενδιαφερόμενο. Η παρουσίαση μπορεί να γίνει με μορφή

A. Πινάκων

B. Γραφικών Παραστάσεων

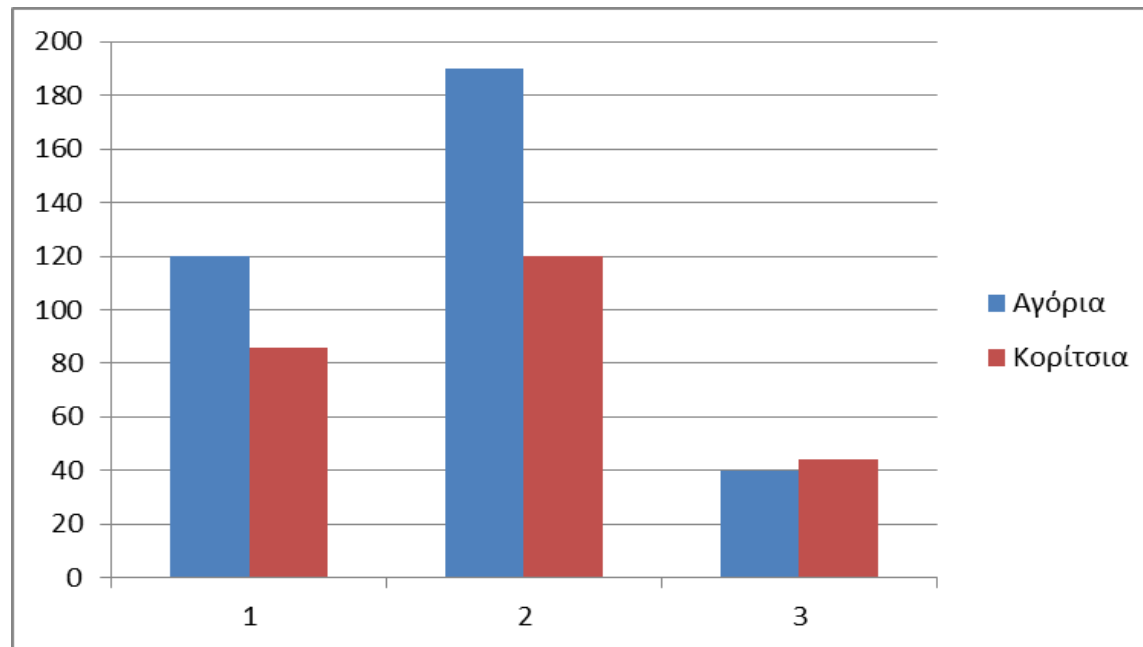
- Οι Πίνακες μπορεί να είναι **απλής εισόδου** ή **διπλής εισόδου**.

Οι **πίνακες απλής εισόδου** χρησιμοποιούνται όταν οι μονάδες του εξεταζόμενου πληθυσμού ερευνώνται ως προς ένα ποιοτικό ή ποσοτικό χαρακτηριστικό.

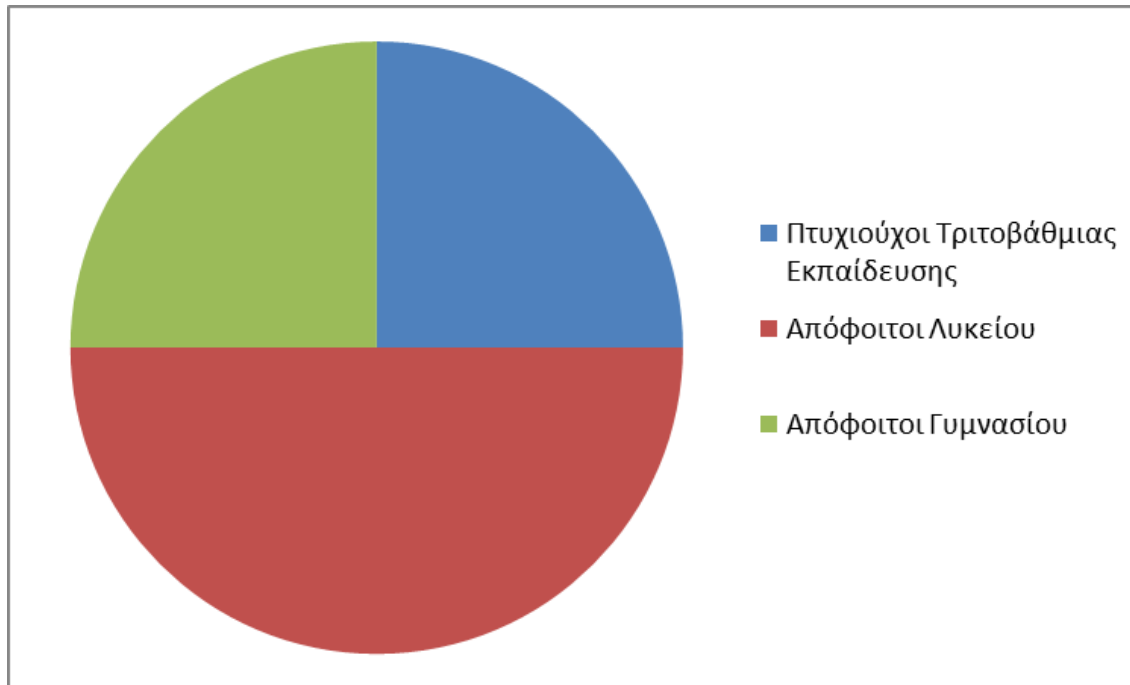
Ενώ οι **πίνακες διπλής εισόδου** όταν οι μονάδες του εξεταζόμενου πληθυσμού μελετώνται ταυτοχρόνως ως προς δυο ποιοτικά ή ποσοτικά χαρακτηριστικά.

- Γραφικές Παραστάσεις

- 1) Ακιδωτά διαγράμματα ή Ραβδογράμματα (Bar Charts)



2) Κυκλικά Διαγράμματα (Pie Charts)

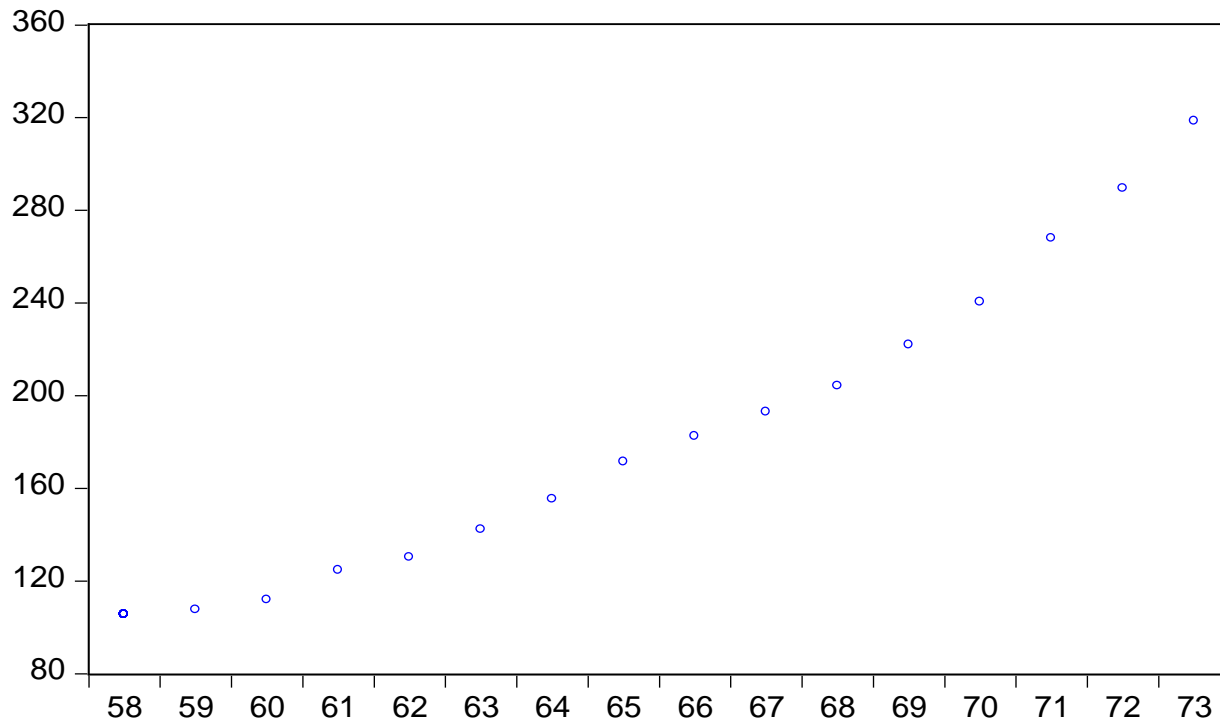


3) Ιστογράμματα

Χρησιμοποιούνται για τη γραφική απεικόνιση ποσοτικών κατανομών, αποτελούνται από διαδοχικά ορθογώνια, που έχουν βάσεις ίσες με τα διαστήματα των τάξεων τοποθετημένες πάνω στον οριζόντιο άξονα. Το εμβαδόν κάθε ορθογωνίου ισούται με τη συχνότητα της αντίστοιχης τάξεως. Επίσης, αν ενώσουμε τα μέσα των επάνω βάσεων των ορθογωνίων ενός ιστογράμματος, σχηματίζεται τεθλασμένη γραμμή, που λέγεται πολύγωνο συχνοτήτων (frequency polygon)

- Διάγραμμα σημείων (Scatter Plot)

x



Μέτρα θέσης και μεταβλητότητας

- Μέτρα θέσης καθορίζουν τη θέση των τιμών στο χώρο.
- Μέτρα μεταβλητότητας ή διασποράς καθορίζουν πώς μεταβάλλονται οι τιμές της μεταβλητής

Μέτρα Θέσης

- Αριθμητικός Μέσος ή μέση τιμή (mean) ορίζεται ως το πηλίκο των τιμών της μεταβλητής δια το πλήθος των τιμών της.

$$\mu = \frac{\sum_{i=1}^N x_i}{N}$$

- Διάμεσος (median) είναι η τιμή εκείνη της μεταβλητής που χωρίζει το σύνολο των τιμών σε δυο ίσα μέρη.
- (όταν το πλήθος των τιμών της μεταβλητής είναι άρτιος αριθμός τότε η διάμεσος ορίζεται ως ο μέσος αριθμητικός των τιμών των δύο κεντρικών όρων)
- (όταν το πλήθος των τιμών είναι περιττός αριθμός τότε ως διάμεσο παίρνουμε την τιμή εκείνη της μεταβλητής που βρίσκεται στο κέντρο, αφού προηγουμένως οι τιμές τοποθετηθούν κατά αύξουσα τάξη μεγέθους)
- Επικρατούσα Τιμή (mode) είναι η τιμή της μεταβλητής με τη μεγαλύτερη συχνότητα
- Τεταρτημόρια (quartiles) οι τιμές της μεταβλητής που χωρίζουν το σύνολο των τιμών της σε 4 ισοπληθείς ομάδες.

Πληθυσμιακός μέσος

$$\begin{aligned}\mu &= \frac{\sum X}{N} = \frac{X_1 + X_2 + X_3 + \dots + X_N}{N} \\ &= \frac{24 + 13 + 19 + 26 + 11}{5} \\ &= \frac{93}{5} \\ &= 18.6\end{aligned}$$

Δειγματικός μέσος

$$\begin{aligned}\bar{X} &= \frac{\sum X}{n} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n} \\ &= \frac{57 + 86 + 42 + 38 + 90 + 66}{6} \\ &= \frac{379}{6} \\ &= 63.167\end{aligned}$$

Διάμεσος: Παράδειγμα με μονό αριθμό στοιχείων

Το διατεταγμένο σύνολο περιλαμβάνει:

3 4 5 7 8 9 11 14 15 16 16 17 19 19 20 21 22

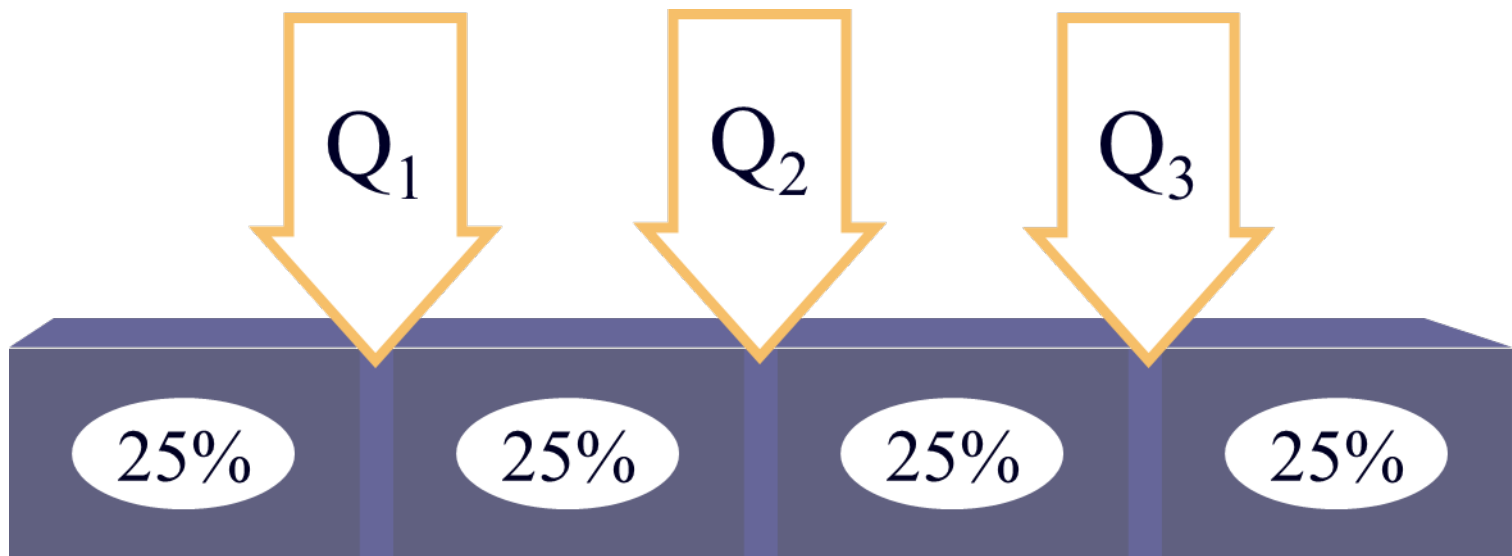
- Υπάρχουν 17 στοιχεία στο διατεταγμένο σύνολο.
- Θέση της Διαμέσου = $(n+1)/2 = (17+1)/2 = 9$
- Η διάμεσος είναι το 9^ο στοιχείο, δηλ. το 15.

Επικρατούσα τιμή - παράδειγμα

- Η επικρατούσα τιμή είναι 44.
- Η τιμή 44 εμφανίζεται περισσότερες φορές από οποιαδήποτε άλλη τιμή.

35	41	44	45
37	41	44	46
37	43	44	46
39	43	44	46
40	43	44	46
40	43	45	48

Τεταρτημόρια-παράδειγμα



Το Q_1 είναι ίσο με το 25^ο εκατοστημόριο

Το Q_2 τοποθετείται στο 50^ο εκατοστημόριο και είναι ίσο με τη διάμεσο

Το Q_3 είναι ίσο με το 75^ο εκατοστημόριο

Μέτρα Θέσης

- Στην περίπτωση που μια κατανομή είναι συμμετρική, οι τιμές του αριθμητικού μέσου, διαμέσου και επικρατούσας τιμής συμπίπτουν μεταξύ τους.

Μέτρα Διασποράς

- Εύρος (range) δίνει τη διαφορά ανάμεσα στη μεγαλύτερη και την μικρότερη τιμή μιας σειράς παρατηρήσεων.
- Διακύμανση (variance) είναι ο μέσος αριθμητικός των τετραγώνων των διαφορών των τιμών μιας μεταβλητής από το μέσο αριθμητικό της. Συμβολίζεται με σ^2 (όταν αναφερόμαστε στον πληθυσμό) ή με s^2 (όταν αναφερόμαστε στο δείγμα)

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N} \qquad s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

- Δηλώνει πόσο μακριά από τη μέση τιμή απέχουν οι παρατηρήσεις.
- Όταν οι παρατηρήσεις απέχουν πολύ από τη μέση τιμή, η διασπορά είναι μεγάλη.

- Επειδή η διακύμανση εκφράζεται μέσω του τετραγώνου της μεταβλητής, γι' αυτό παίρνουμε την τετραγωνική ρίζα της διακύμανσης που ονομάζεται τυπική απόκλιση (standard deviation)

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

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

- Παράδειγμα: Εύρος =
= $48 - 35 = 13$

35	41	44	45
37	41	44	46
37	43	44	46
39	43	44	46
40	43	44	46
40	43	45	48

Πληθυσμιακή Διακύμανση- Παράδειγμα

X	$X - \mu$	$(X - \mu)^2$
5	-8	64
9	-4	16
16	+3	9
17	+4	16
18	<u>+5</u>	<u>25</u>
	0	130

$$\begin{aligned}\sigma^2 &= \frac{\sum (X - \mu)^2}{N} \\ &= \frac{130}{5} \\ &= 26.0\end{aligned}$$

Πληθυσμιακή Τυπική απόκλιση

$$\begin{aligned}\sigma^2 &= \frac{\sum (X - \mu)^2}{N} \\ &= \frac{130}{5} \\ &= 26.0 \\ \sigma &= \sqrt{\sigma^2} \\ &= \sqrt{26.0} \\ &= 5.1\end{aligned}$$

Δειγματική Διακύμανση

X	$X - \bar{X}$	$(X - \bar{X})^2$
2,398	625	390,625
1,844	71	5,041
1,539	-234	54,756
<u>1,311</u>	<u>-462</u>	<u>213,444</u>
7,092	0	663,866

$$\begin{aligned} S^2 &= \frac{\sum (X - \bar{X})^2}{n - 1} \\ &= \frac{663,866}{3} \\ &= 221,288.67 \end{aligned}$$

Δειγματική τυπική απόκλιση

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

$$= \frac{663,866}{3}$$

$$= 221,288.67$$

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

$$= \sqrt{221,288.67}$$

$$= 470.41$$

Μέτρα μεταβλητότητας

- Όσο μικρότερες οι τιμές της διασποράς και της τυπικής απόκλισης, τόσο πιο συγκεντρωμένες γύρω από το μ βρίσκονται οι τιμές της μεταβλητής.
- Συντελεστής μεταβλητικότητας (coefficient of variation) όταν θέλουμε να συγκρίνουμε δύο κατανομές οι οποίες εκφράζονται σε διαφορετικές μονάδες τότε χρησιμοποιούμε το CV και είναι το πηλίκο της τυπικής απόκλισης μιας κατανομής προς τον αριθμητικό μέσο αυτής. Εκφράζεται ως ποσοστό.

$$CV = \frac{\sigma}{\mu}$$

$$\mu_1 = 29$$

$$\sigma_1 = 4.6$$

$$C.V._1 = \frac{\sigma_1}{\mu_1}(100)$$

$$= \frac{4.6}{29}(100)$$

$$= 15.86$$

$$\mu_2 = 84$$

$$\sigma_2 = 10$$

$$C.V._2 = \frac{\sigma_2}{\mu_2}(100)$$

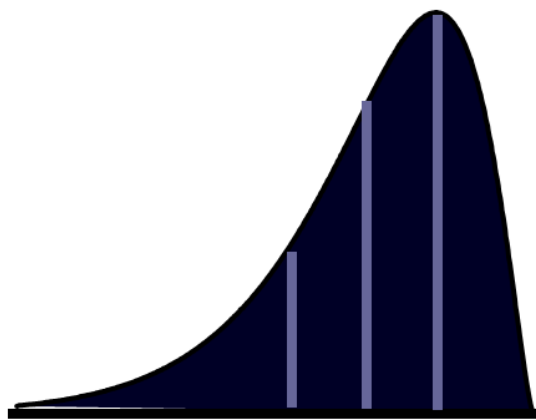
$$= \frac{10}{84}(100)$$

$$= 11.90$$

Ασυμμετρία & Κύρτωση

- Συμμετρική είναι μια κατανομή όταν οι τιμές της τοποθετούνται συμμετρικά γύρω από το μέσο.
- Ο συντελεστής ασυμμετρίας (skewness) μας πληροφορεί για το αν οι τιμές της μεταβλητής τοποθετούνται συμμετρικά γύρω από το μέσο ή όχι και πόσο.

- Αν $S < 0$, η κατανομή είναι αρνητικά ασύμμετρη (ασύμμετρη στα αριστερά).
- Αν $S = 0$, η κατανομή είναι συμμετρική (όχι ασύμμετρη).
- Αν $S > 0$, η κατανομή είναι θετικά ασύμμετρη (ασύμμετρη στα δεξιά).

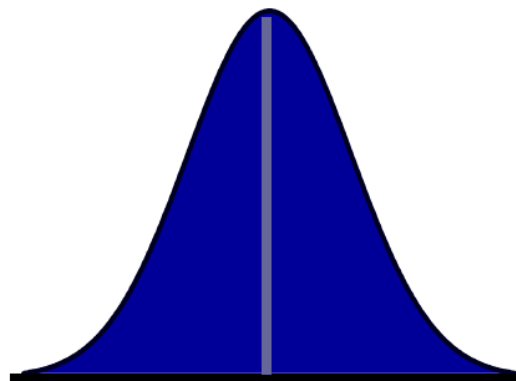


Μέσος

Επικρατούσα
Τιμή

Διάμεσος

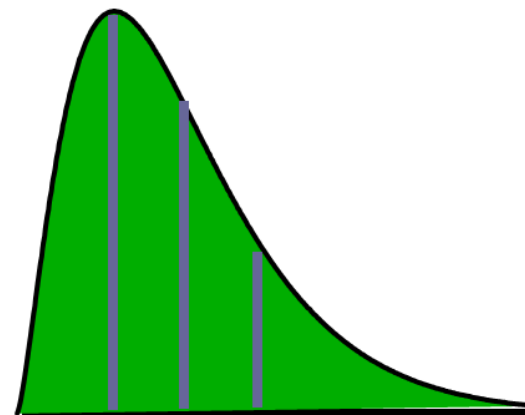
**Αρνητική
Ασυμμετρία**



Μέσος
Διάμεσος

Επικρατούσα Τιμή

Συμμετρία



Επικρατούσα
Τιμή

Διάμεσος

Μέσος

**Θετική
Ασυμμετρία**

Ασυμμετρία & Κύρτωση

- Η κύρτωση (kyrtosis) μετράει το βαθμό της συγκέντρωσης των τιμών της μεταβλητής στην περιοχή του αριθμητικού μέσου και προς τα άκρα του μέσου.



Λεπτόκυρτη

Μεσόκυρτη

Πλατόκυρτη

Πίνακες κατανομής συχνοτήτων

- Οι πίνακες αυτοί συντάσσονται με κατάλληλη κατάταξη και συστηματική ομαδοποίηση των τιμών της μεταβλητής που εξετάζεται.
- Αν τα χαρακτηριστικά είναι ασυνεχή και τα αποτελέσματα της μέτρησης σχετικά λίγα τότε ο πίνακας παίρνει την ακόλουθη μορφή:

Δυνατές τιμές της μεταβλητής x	Αριθμός φορών που παρατηρήθηκε η κάθε τιμή (συχνότητα)
x_1	f_1
x_2	f_2
x_3	f_3
...
Σύνολο	$\sum_{i=1}^k f = n$

- Τα $x_1, x_2, x_3, \dots, x_k$ είναι τιμές της διακριτής μεταβλητής X , οι οποίες τοποθετούνται κατά τη φυσική τους σειρά, από τη μικρότερη στη μεγαλύτερη. Τα $f_1, f_2, f_3, \dots, f_k$ εκφράζουν πόσες φορές εμφανίζεται στο συνολικό πληθυσμό κάθε τιμή της μεταβλητής.

- Αν τα χαρακτηριστικά είναι συνεχή ή διακριτά με μεγάλο πλήθος δυνατών τιμών, τότε δυσχεραίνεται η μορφή του πίνακα, οπότε κρίνεται απαραίτητη η ομαδοποίηση των παρατηρήσεων. Η ομαδοποίηση αυτή πραγματοποιείται με το χωρισμό του διαστήματος μεταβολής (α_0, α_1) της μεταβλητής X σε υποδιαστήματα που ονομάζονται κλάσεις ή τάξεις ή ομάδες.

- Τα άκρα των τάξεων καλούνται κατώτερο όριο και ανώτερο όριο. Η διαφορά των δυο ορίων καλείται πλάτος της τάξεως και συμβολίζεται με δ . Το ημιάθροισμα των ορίων της κάθε τάξης καλείται κεντρική τιμή της τάξεως.