ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν Πανεπιστήμιον Αθηνών
ΙΔΡΥΘΕΝ ΤΟ 1837

ΓΕΩΠΟΝΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
AGRICULTURAL UNIVERSITY OF ATHENS

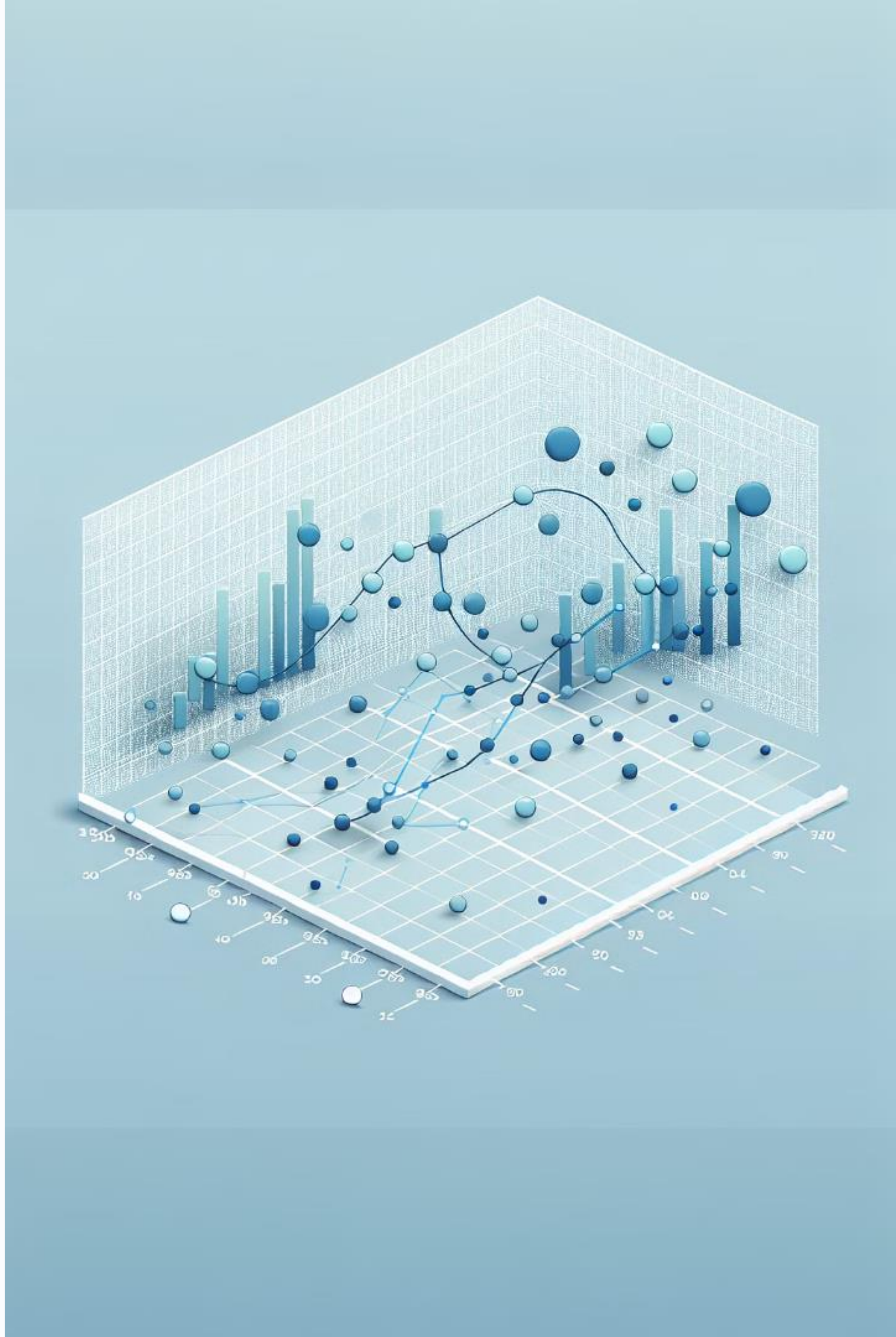
ΔΠΜΣ «ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ ΔΙΟΙΚΗΣΗΣ»

Μάθημα: Ποσοτικές Μέθοδοι

Διδάσκων: Δρ. Ευάγγελος Δημ. Μακρυβέλιος

1η Εκπαιδευτική Συνάντηση:
Περιγραφική Στατιστική

Αθήνα, 2024



Περιγραφική και επαγωγική στατιστική

Στατιστική

Η επιστήμη του **σχεδιασμού** μελετών και πειραμάτων, της **συγκέντρωσης δεδομένων** και της **οργάνωσης, σύνοψης, παρουσίασης, ανάλυσης** και **ερμηνείας** αυτών των δεδομένων και στη συνέχεια της **εξαγωγής συμπερασμάτων** βάσει αυτών

Περιγραφική Στατιστική

Κλάδος της στατιστικής που ασχολείται με την **οργάνωση, την παρουσίαση και την περιγραφή** ποσοτικών πληροφοριών (όχι γενικεύσεις)

Επαγωγική Στατιστική

Κλάδος της στατιστικής που ασχολείται με την **εξαγωγή συμπερασμάτων** για ολόκληρους πληθυσμούς **με βάση τα δεδομένα ενός δείγματος** (προβλέψεις, προβολές, έλεγχος υποθέσεων, λήψεων αποφάσεων)

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν Πανεπιστήμιον Αθηνών
ΙΔΡΥΘΕΝ ΤΟ 1837

ΓΕΩΠΟΝΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
AGRICULTURAL UNIVERSITY OF ATHENS

Ορισμοί

Πληθυσμός (population)

Περιλαμβάνει το ***σύνολο*** των μονάδων που διαθέτουν κάποιο χαρακτηριστικό που μας ενδιαφέρει

Δείγμα (sample)

είναι ένα ***υποσύνολο*** ενός πληθυσμού

Απογραφή

Η απογραφή είναι **μια διαδικασία συλλογής δεδομένων** από **<u>όλα</u>** τα μέλη **ενός πληθυσμού**

Δεδομένα (data)

Τα δεδομένα είναι οι μεμονωμένες τιμές που συλλέγονται για μια μεταβλητή

Ορισμοί (συνέχεια)

Παράμετρος (parameter): Ένα ***χαρακτηριστικό/ιδιότητα*** του πληθυσμού

Μεταβλητή (variable): *Ιδιότητα* ενός δείγματος, κάθε χαρακτηριστικό που μπορεί να μετρηθεί ή να παρατηρηθεί, π.χ. αύξηση βάρους σώματος

Στατιστικό μέτρο (statistic): Είναι ένα ***χαρακτηριστικό*** του δείγματος

Παρατήρηση (observation): **Η *τιμή της μεταβλητής*** για ένα συγκεκριμένο αντικείμενο που μελετάται, π.χ. η αύξηση βάρους της 12ης εγκύου κατά τη διάρκεια της κύησης ήταν 4Kg

Τύποι στατιστικών δεδομένων/μεταβλητών

Ποιοτικά ή κατηγορικά δεδομένα (qualitative data), οι τιμές των οποίων ***δεν είναι αριθμοί*** (π.χ. χρώμα αυτοκινήτων)

Ποσοτικά ή μετρήσιμα δεδομένα (quantitative data), οι τιμές των οποίων ***είναι αριθμοί***

Κατηγορικές (ποιοτικές) μεταβλητές (qualitative variables)
Εκφράζουν έμμεσα μία ***ιδιότητα του πληθυσμού***, κατατάσσοντας σε μία κατηγορία. Το φύλο (άνδρας/γυναίκα), οικογενειακή κατάσταση (άγαμος/ έγγαμος),το αποτέλεσμα μιας θεραπείας (ναι/όχι), η ποσότητα λήψης φαρμάκου (μικρή/μεσαία/μεγάλη). **Μετριέται** με τη χρήση ακέραιων αριθμών

Αριθμητικές (ποσοτικές) μεταβλητές (quantitative variables)
Αντιπροσωπεύουν μία ***μετρήσιμη ιδιότητα*** του πληθυσμού (π.χ. θερμοκρασία, βάρος, ύψος, η ηλικία κτλ.). Χρησιμοποιείται οποιαδήποτε φυσική τιμή. Δύο ***επιμέρους κατηγορίες***:

- **Διακριτές μεταβλητές (discrete variables)**, οι τιμές των οποίων είναι ***ακέραιοι αριθμοί*** (π.χ. αριθμός εργαζομένων, αριθμός παιδιών, ελαττωματικά προϊόντα ανά εβδομάδα παραγωγής(καταμέτρηση)
- **Συνεχείς μεταβλητές (continuous variables)**, οι τιμές των οποίων είναι ***οποιοιδήποτε αριθμοί*** (π.χ. μισθοί εργαζομένων, βάρος) (μέτρηση)

Τύποι μεταβλητών

- **Ερώτηση:** Έχετε προφίλ στο Instagram;
- **Απάντηση:** Ναι ή Όχι
- **Τύπος μεταβλητής :** *Κατηγορική (ποιοτική)*

- **Ερώτηση:** Πόσους Followers έχετε αποκτήσει την τελευταία εβδομάδα;
- **Απάντηση:** 2.000
- **Τύπος μεταβλητής :** *Αριθμητική (διακριτή)*

- **Ερώτηση:** Πόση ώρα έκανε η φωτογραφία να ανέβει στο προφίλ σας;
- **Απάντηση:** 3.2 δευτερόλεπτα
- **Τύπος μεταβλητής :** *Αριθμητική (συνεχής)*

Πηγές δεδομένων

1 **Πρωτογενείς πηγές (primary sources):** Ο ***συλλέκτης δεδομένων*** είναι αυτός που χρησιμοποιεί ***ο ίδιος*** τα

δεδομένα για ανάλυση.

- Δεδομένα από έρευνα
- Δεδομένα που συλλέγονται από ένα πείραμα
- Δεδομένα από παρατήρηση

(άρθρα περιοδικών και επιθεωρήσεων, πρακτικά συνεδρίων, αναφορές και εκθέσεις, επίσημες εκδόσεις ευρεσιτεχνίας, κατάλογοι, προδιαγραφές και ευρετήρια, δεδομένα από τον πληθυσμό)

2 **Δευτερογενείς πηγές (secondary sources):** Το άτομο που πραγματοποιεί

ανάλυση δεδομένων ***δεν είναι ο συλλέκτης δεδομένων***

- Ανάλυση δεδομένων απογραφής
- Εξέταση δεδομένων από έντυπα περιοδικά ή δεδομένων που δημοσιεύονται στο διαδίκτυο

(μονογραφίες, εγχειρίδια, επετηρίδες, ευρετήρια δημοσιεύσεων, ενημερωτικά δημοσιεύματα, βάσεις δεδομένων, έρευνες στο διαδίκτυο κ.α.)

3 **Τριτογενείς πηγές (tertiary sources)**

(οδηγοί, οδηγοί στη χρήση ειδικευμένων βιβλιογραφιών, γενικές & ειδικές βιβλιογραφίες, εγκυκλοπαίδειες)

Πηγές δεδομένων

Οι πηγές δεδομένων εμπίπτουν σε πέντε κατηγορίες:

- Δεδομένα **που διανέμονται** από έναν οργανισμό ή ένα άτομο

  (οικονομικά δεδομένα επιχείρησης, τιμές μετοχών, καιρικές συνθήκες, στατιστικά για τον αθλητισμό)

- Αποτελέσματα ενός σχεδιασμένου **πειράματος**

  (δοκιμές των καταναλωτών για διάφορες εκδοχές ενός προϊόντος, έλεγχος αγοράς για εναλλακτικές προωθητικές ενέργειες)

- Απαντήσεις από **μια έρευνα**

  (πολιτικές δημοσκοπήσεις, ικανοποίησή τους σχετικά με μια πρόσφατη εμπειρία προϊόντος ή υπηρεσίας)

- Αποτελέσματα της διενέργειας μιας **μελέτης παρατήρησης**

  (ερευνητές της αγοράς που χρησιμοποιούν ομάδες εστίασης για να αποσπάσουν μη δομημένες απαντήσεις σε ανοιχτές ερωτήσεις, μέτρηση του χρόνου που απαιτείται για την εξυπηρέτηση των πελατών)

- Δεδομένα που συλλέγονται από την **συνεχή δραστηριότητα** οντοτήτων

  (δεδομένα παρακολούθησης για να αξιολογήσουν την αποτελεσματικότητα ενός ιστότοπου)

Περιγραφική στατιστική

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

Μέθοδοι παρουσίασης στατιστικών δεδομένων

Βασικοί τρόποι σύνοψης στατιστικών δεδομένων

❖ **Γραφικοί μέθοδοι** (πίνακας κατανομής συχνοτήτων, ραβδόγραμμα, ιστόγραμμα, κυκλικό διάγραμμα)

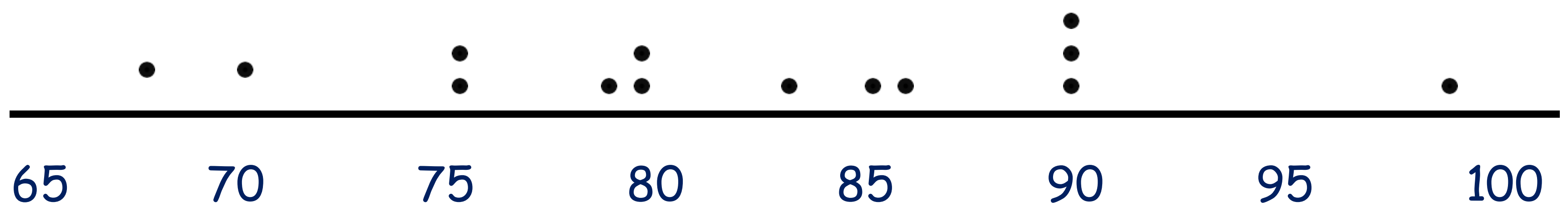
❖ **Αριθμητικοί μέθοδοι** (μέτρα κεντρικής τάσης, μέτρα σχετικής θέσης, μέτρα διασποράς, μέτρα ασυμμετρίας)

Διάγραμμα Σημείων

- Το **διάγραμμα σημείων** (dot diagram) χρησιμοποιείται για την **απεικόνιση ποσοτικών δεδομένων** όταν έχουμε **μικρό αριθμό** παρατηρήσεων (μικρό πλήθος)
- Πρόκειται για την **τοποθέτηση** όλων των διαθέσιμων τιμών πάνω σε ένα ευθύγραμμο τμήμα.Αν *υπάρχουν περισσότερες από μια τιμές οι οποίες* ***συμπίπτουν, τοποθετούνται*** *η μία πάνω στην άλλη*

Παράδειγμα: Η βαθμολογία 20 φοιτητών στο μάθημα της Στατιστικής δίνεται στον πιο κάτω πίνακα

70 90 82 90 75 85 86 79 84 98 75 80 90 80 68



Ορισμοί Συχνότητας

1 **Συχνότητα (frequency)**: Ο **αριθμός** ο οποίος υποδηλώνει το **πλήθος** των μονάδων οι οποίες είτε εντάσσονται σε **κάποια κατηγορία** (περίπτωση ποιοτικών δεδομένων),είτε αντιστοιχούν σε **κάποια τιμή** της μεταβλητής (περίπτωση ποσοτικών δεδομένων)

2 **Σχετική συχνότητα (relative frequency)**: ορίζεται ως ο **λόγος** της **συχνότητας προς το σύνολο** των μονάδων του δείγματος

3 **Αθροιστική συχνότητα (cumulative frequencies)** μια τιμής μιας μεταβλητής είναι ο φυσικός αριθμός ο οποίος υποδηλώνει το **πλήθος των παρατηρήσεων**, οι οποίες είναι **μικρότερες ή ίσες** από τη συγκεκριμένη τιμής της μεταβλητής
.

4 **Σχετική αθροιστική συχνότητα (cummulative relative frequencies)** μια τιμής της μεταβλητής ορίζεται ως **ο λόγος της αθροιστικής συχνότητας**, που αντιστοιχεί στη συγκεκριμένη τιμής της μεταβλητής **προς το σύνολο των μονάδων** του δείγματος. Σε πολλές περιπτώσεις οι σχετικές αθροιστικές συχνότητες εκφράζονται ως **ποσοστά επί τοις εκατό**

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

- Είναι ένα εργαλείο που χρησιμοποιείται για να **οργανώσει και να παρουσιάσει** τα δεδομένα με μια πιο εύληπτη μορφή. Μας βοηθά να δούμε την **κατανομή των δεδομένων** και να **κατανοήσουμε την συχνότητα εμφάνισης** των διαφορετικών τιμών
- Ο πίνακας περιέχει μια λίστα των διαφορετικών τιμών των δεδομένων, **μαζί με τον αριθμό των φορών που κάθε τιμή εμφανίζεται στο σύνολο των δεδομένων (vi).** Επιπλέον, ο πίνακας περιλαμβάνει τη **σχετική συχνότητα (fi), η οποία αντιπροσωπεύει το ποσοστό εμφάνισης κάθε τιμής στο σύνολο των δεδομένων.** Η **σχετική συχνότητα υπολογίζεται διαιρώντας τη συχνότητα κάθε τιμής με τον συνολικό αριθμό των τιμών**

Κατανομή Συχνότητας

Παράδειγμα

Η βαθμολογία (με άριστα το 100) σε ένα τεστ 150 υποψηφίων για θέσεις γραμματέων σε μια μεγάλη εταιρεία φαίνεται στον πίνακα. Να κατασκευαστεί η κατανομή συχνοτήτων των δεδομένων.

27	79	69	40	51	88	55	48	36	61
53	44	94	51	65	42	58	55	69	63
70	48	61	55	60	25	47	78	61	54
57	76	73	62	36	67	40	51	59	68
27	46	62	43	54	83	59	13	72	57
82	45	54	52	71	53	82	69	60	35
41	65	62	75	60	42	55	34	49	45
49	64	40	61	73	44	59	46	71	86
43	69	54	31	56	51	75	44	66	53
80	71	53	56	91	60	41	29	56	57
35	54	43	39	56	27	62	44	85	61
59	89	60	51	71	53	58	26	77	68
62	57	48	69	76	52	49	45	54	41
33	61	80	57	42	45	59	44	68	73
55	70	39	58	69	51	85	46	55	67

Κατανομή Συχνότητας

Για την **ταξινόμηση** σε κατανομή συχνοτήτων (frequency distribution) ακολουθούμε **τρία** βήματα:

- Δημιουργία **Τάξεων** (classes) για τη μεταβλητή που εξετάζουμε
- **Ένταξη των δεδομένων** στις τάξεις
- Αναγραφή του **πλήθους των τιμών** της μεταβλητής που εμπίπτουν στην κάθε τάξη

Μέγεθος δείγματος (ν)	Αριθμός κλάσεων (κ)	Μέγεθος δείγματος (ν)	Αριθμός κλάσεων (κ)
<20	5	200-400	9
20-50	6	400-700	10
50-100	7	700-1.000	11
100-200	8	≥ 1.000	12

Για το πρώτο βήμα πρέπει να προσδιορίσουμε κάθε φορά το **πλήθος των τάξεων** και το **πλάτος της κάθε τάξης**. Για να το πετύχουμε αυτό λαμβάνουμε υπόψη μας τα ακόλουθα:

- Σχεδόν ποτέ οι κατανομές έχουν λιγότερες από 6 ή περισσότερες από 16 τάξεις
- Η κάθε τιμή της μεταβλητής εντάσσεται σε **μία μόνο** τάξη
- Επιδιώκουμε οι τάξεις να έχουν όλες το ίδιο πλάτος
- Συνήθως οι τάξεις είναι διαστήματα ανοιχτά προς τα πάνω

Κατανομή Συχνότητας

Οι βαθμολογίες κυμαίνονται από 13 ως και 94, επιλέγουμε τις εννέα τάξεις 10-20, 20-30,...,90-100

Εντάσσουμε τις τιμές της μεταβλητής (βαθμολογία υποψηφίων) στις δημιουργηθείσες τάξεις

Συχνότητα (frequency) = το πλήθος των τιμών της μεταβλητής που εντοπίζονται σε μια δεδομένη τάξη

Κάθε **τάξη ορίζεται από δύο αριθμούς**. Ο μικρότερος από αυτούς ονομάζεται **κατώτερο όριο** και ο μεγαλύτερος **ανώτερο όριο**

Ανώτατο όριο=20
Κατώτατο όριο=10

ΤΑΞΕΙΣ	ΠΕΡΙΠΤΩΣΕΙΣ	ΣΥΧΝΟΤΗΤΑ
10-20	I	1
20-30	IIIII I	6
30-40	IIIII IIII	9
40-50	IIIII IIIII IIIII IIIII IIIII IIIII I	31
50-60	IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII II	42
60-70	IIIII IIIII IIIII IIIII IIIII IIIII II	32
70-80	IIIII IIIII IIIII II	17
80-90	IIIII IIIII	10
90-100	II	2
		150

Πλάτος Τάξεως (δ) =
Η διαφορά μεταξύ δύο διαδοχικών κάτω (ή πάνω) ορίων
δ=40-30=10

Κεντρικός όρος= ημιάθροισμα του κατώτερου και ανώτερου ορίου κάθε τάξης

$$\frac{50+60}{2} = 55$$

Οι τάξεις είναι **διαστήματα ανοικτά** προς τα πάνω.
π.χ. [10,20)

Κατανομή Συχνότητας

Το ποσοστό όλων των τιμών της μεταβλητής που βρίσκονται σε μία συγκεκριμένη τάξη ονομάζεται **σχετική συχνότητα** (relative frequency)

ΤΑΞΕΙΣ	ΣΧΕΤΙΚΗ ΣΥΧΝΟΤΗΤΑ	ΣΧΕΤΙΚΗ ΣΥΧΝΟΤΗΤΑ ΩΣ ΠΟΣΟΣΤΟ
10-20	1/150=0,007	0,007*100= 0,7
20-30	6/150=0,040	4,0
30-40	9/150=0,060	6,0
40-50	31/150=0,207	20,7
50-60	42/150=0,280	28,0
60-70	32/150=0,213	21,3
70-80	17/150=0,113	11,3
80-90	10/150=0,067	6,7
90-100	2/150=0,013	1,3
	1,000	**100,0**

Αθροιστική Κατανομή Συχνοτήτων

Τα δεδομένα πολλές φορές παρουσιάζονται με **αθροιστική κατανομή συχνοτήτων** (cumulative frequency distribution)

ΤΑΞΕΙΣ	ΣΥΧΝΟΤΗΤΑ	ΑΘΡΟΙΣΤΙΚΗ ΣΥΧΝΟΤΗΤΑ	ΣΧΕΤΙΚΗ ΑΘΡΟΙΣΤΙΚΗ ΣΥΧΝΟΤΗΤΑ	ΣΧΕΤΙΚΗ ΑΘΡΟΙΣΤΙΚΗ ΣΥΧΝΟΤΗΤΑ ΩΣ %
10-20	1	*1*	*1/150=0,007*	*0,7*
20-30	6	1+6=7	7/150=0,047	4,7
30-40	9	7+9=*16*	16/150=0,106	10,6
40-50	31	16+31=*47*	47/150=0,313	31,3
50-60	42	47+42=*89*	89/150=0,593	59,3
60-70	32	89+32=*121*	121/150=0,806	80,6
70-80	17	121+17=*138*	138/150=0,92	92
80-90	10	138+10=*148*	148/150=0,986	98,6
90-100	2	148+2=*150*	150/150=1	100
	150			

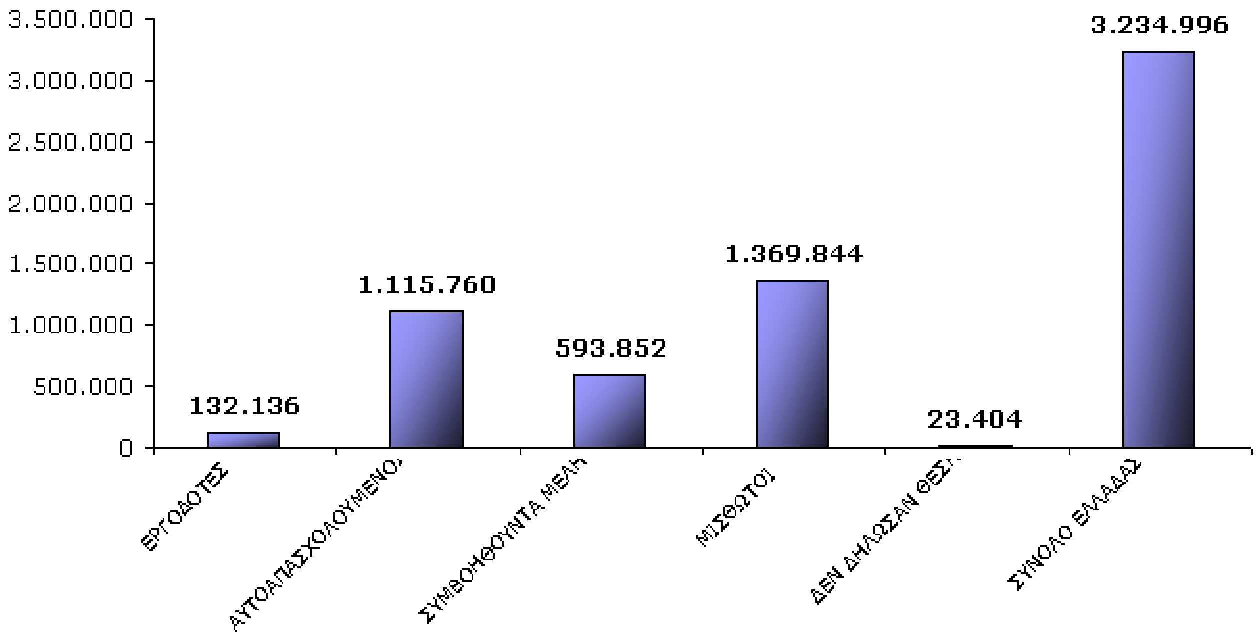
Ραβδόγραμμα-Bar Chart

Το ραβδόγραμμα **απεικονίζει τη συχνότητα των κατηγοριών** στις οποίες είναι ταξινομημένα τα δεδομένα ως ένα ορθογώνιο κάθετο στον οριζόντιο ή τον κατακόρυφο άξονα

Όλα τα ορθογώνια έχουν το ίδιο (αυθαίρετο πλάτος), ενώ το **μήκος** κάθε ορθογωνίου είναι **ανάλογο της συχνότητας** της αντίστοιχης κατηγορίας δεδομένων

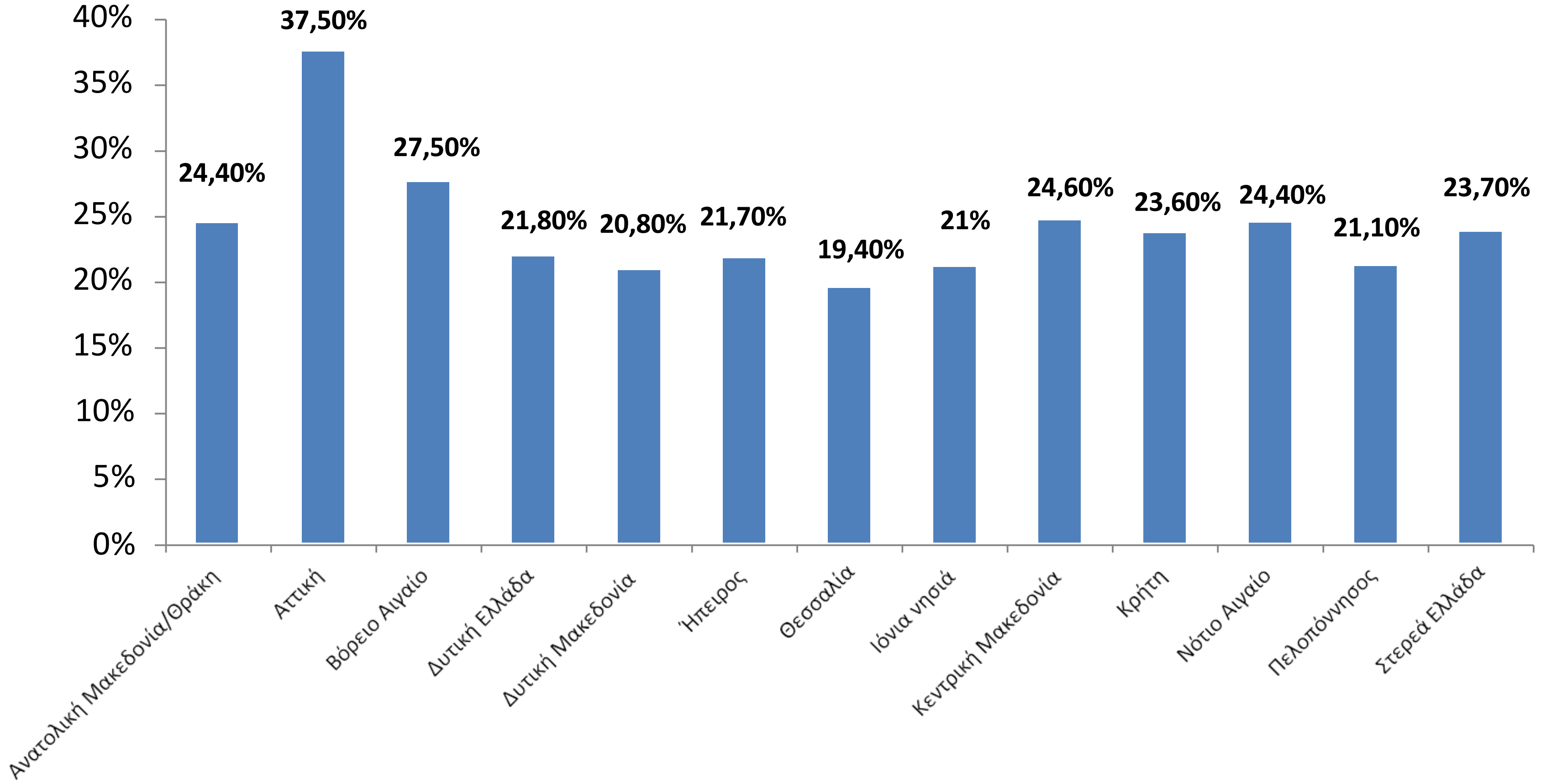
Παράδειγμα

ΘΕΣΗ ΣΤΟ ΕΠΑΓΓΕΛΜΑ	ΣΥΝΟΛΟ
ΕΡΓΟΔΟΤΕΣ	132.136
ΑΥΤΟΑΠΑΣΧΟΛΟΥΜΕΝΟΙ	1.115.760
ΣΥΜΒΟΗΘΟΥΝΤΑ ΜΕΛΗ	593.852
ΜΙΣΘΩΤΟΙ	1.369.844
ΔΕΝ ΔΗΛΩΣΑΝ ΘΕΣΗ	23.404
ΣΥΝΟΛΟ ΕΛΛΑΔΑΣ	3.234.996



Ραβδογράμματα

Χρησιμοποιούνται κυρίως όταν θέλουμε **να εξετάσουμε λεπτομερέστερα το μέγεθος μεταβολών** για κάθε κατηγορία

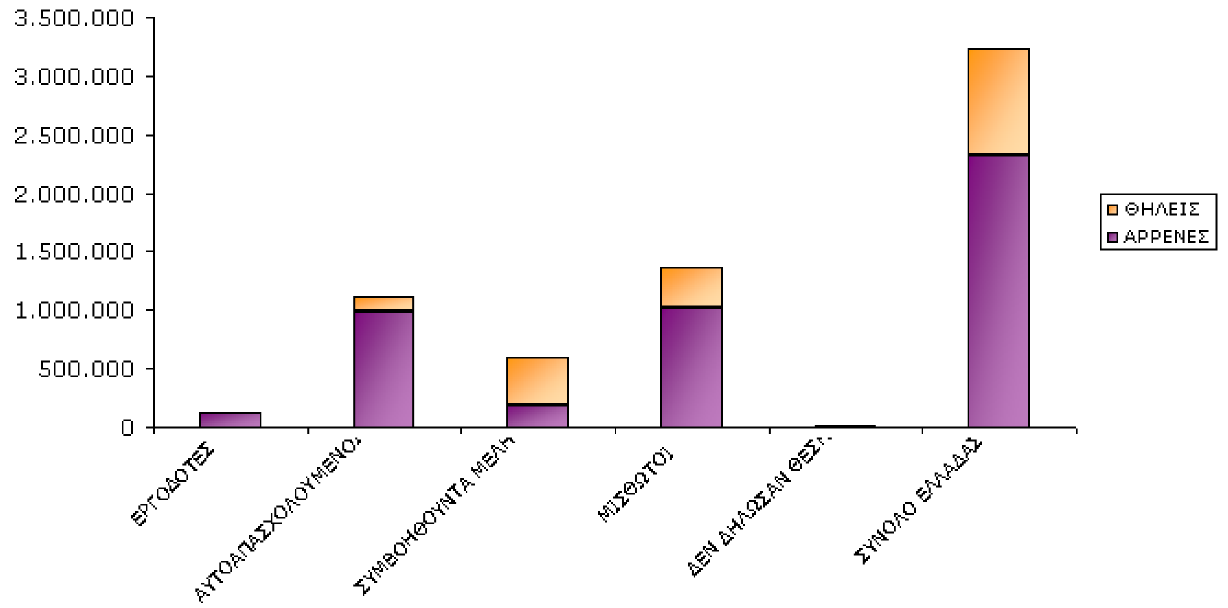


Ραβδόγραμμα Συνιστωσών

Στο ραβδόγραμμα συνιστωσών (component bar chart) **κάθε** ορθογώνιο χωρίζεται σε τμήματα (συνιστώσες) που το καθένα αντιστοιχεί σε μία **υποδιαίρεση της κατηγορίας** που αντιπροσωπεύει το ορθογώνιο

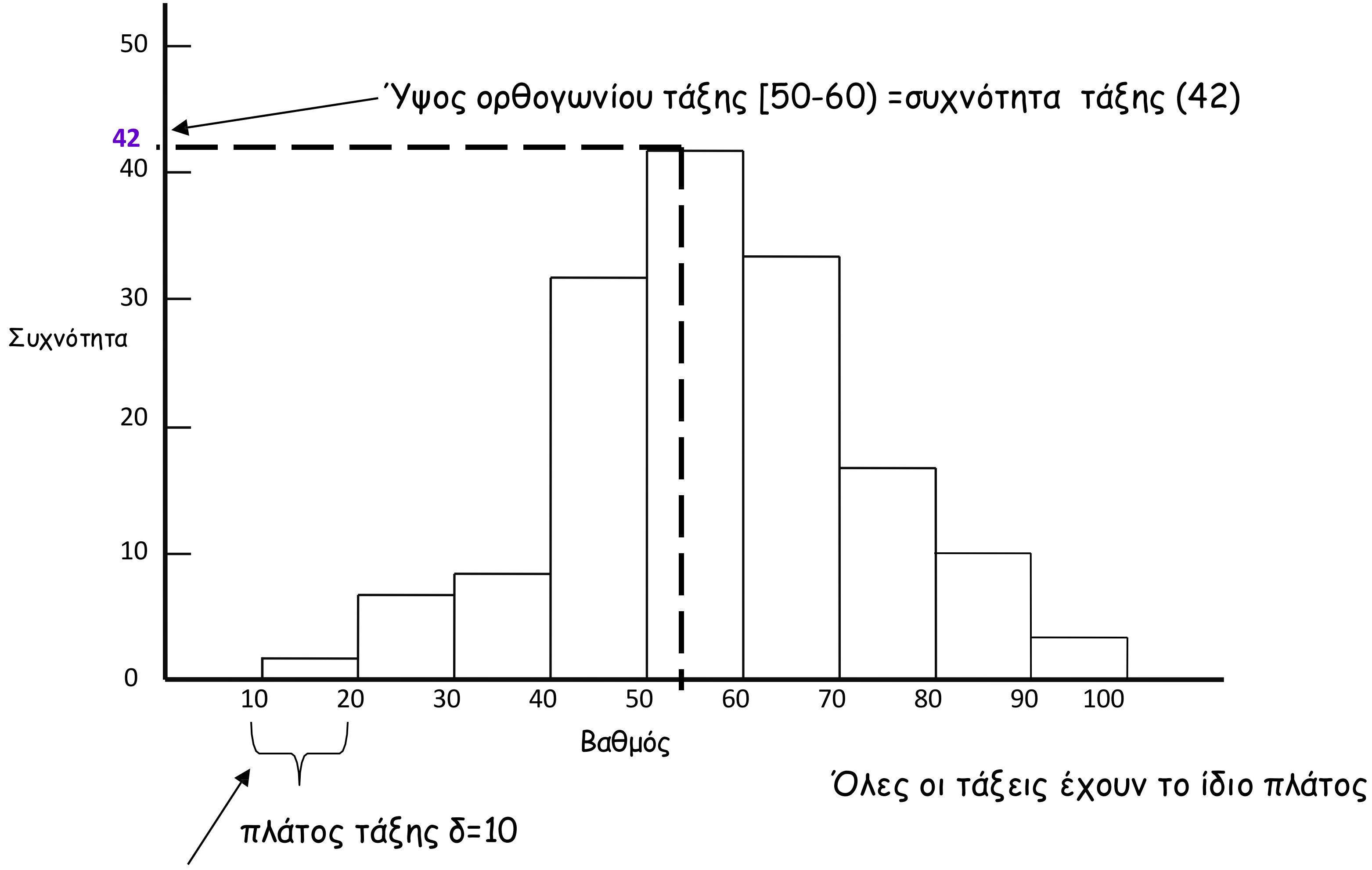
Το **μήκος** κάθε συνιστώσας είναι ανάλογο της **συχνότητας** (ή σχετικής συχνότητας) της αντίστοιχης υποδιαίρεσης

ΘΕΣΗ ΣΤΟ ΕΠΑΓΓΕΛΜΑ	ΣΥΝΟΛΟ	ΑΡΡΕΝΕΣ	ΘΗΛΕΙΣ
ΕΡΓΟΔΟΤΕΣ	132.136	120.832	11.304
ΑΥΤΟΑΠΑΣΧΟΛΟΥΜΕΝΟΙ	1.115.760	982.280	133.480
ΣΥΜΒΟΗΘΟΥΝΤΑ ΜΕΛΗ	593.852	183.696	410.156
ΜΙΣΘΩΤΟΙ	1.369.844	1.026.236	343.608
ΔΕΝ ΔΗΛΩΣΑΝ ΘΕΣΗ	23.404	16.544	6.860
ΣΥΝΟΛΟ ΕΛΛΑΔΑΣ	3.234.996	2.329.588	905.408



Ιστόγραμμα (histogram)

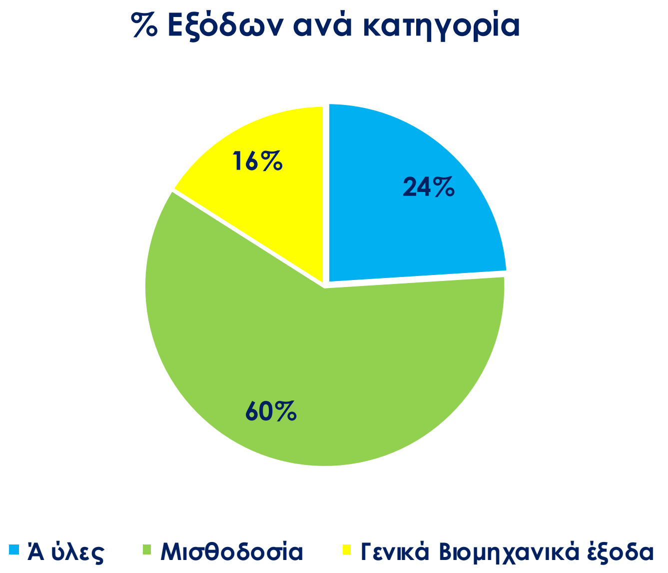
- Στον οριζόντιο άξονα παίρνουμε **ίσα** διαδοχικά ευθύγραμμα τμήματα τα οποία *αντιστοιχούν* στις **τάξεις** της κατανομής
- Οι **συχνότητες** της κατανομής τοποθετούνται στον κατακόρυφο άξονα
- Στη συνέχεια κατασκευάζουμε ορθογώνια παραλληλόγραμμα με **βάσεις** τα ευθύγραμμα τμήματα που αντιστοιχούν στις **τάξεις** της κατανομής και **ύψη** ανάλογα των **συχνοτήτων** της κάθε τάξης



Κυκλικό Διάγραμμα

- Τα κυκλικά διαγράμματα χρησιμοποιούνται για τη **γραφική απεικόνιση δεδομένων** που αποτελούν **υποδιαιρέσεις ενός συνόλου**
- Αν οι υποδιαιρέσεις **εκφράζονται σε ποσοστά** το διάγραμμα αναφέρεται **ως κυκλικό διάγραμμα ποσοστών** (percentage pie chart)
- Για να κατασκευάσουμε ένα κυκλικό διάγραμμα χαράσσουμε καταρχήν ένα κύκλο με αυθαίρετη ακτίνα
- Τον χωρίζουμε σε κυκλικούς τομείς οι οποίοι αντιστοιχούν στις διάφορες κατηγορίες των παρατηρήσεων και το εμβαδόν τους είναι ανάλογο με τη συχνότητα της κάθε κατηγορίας

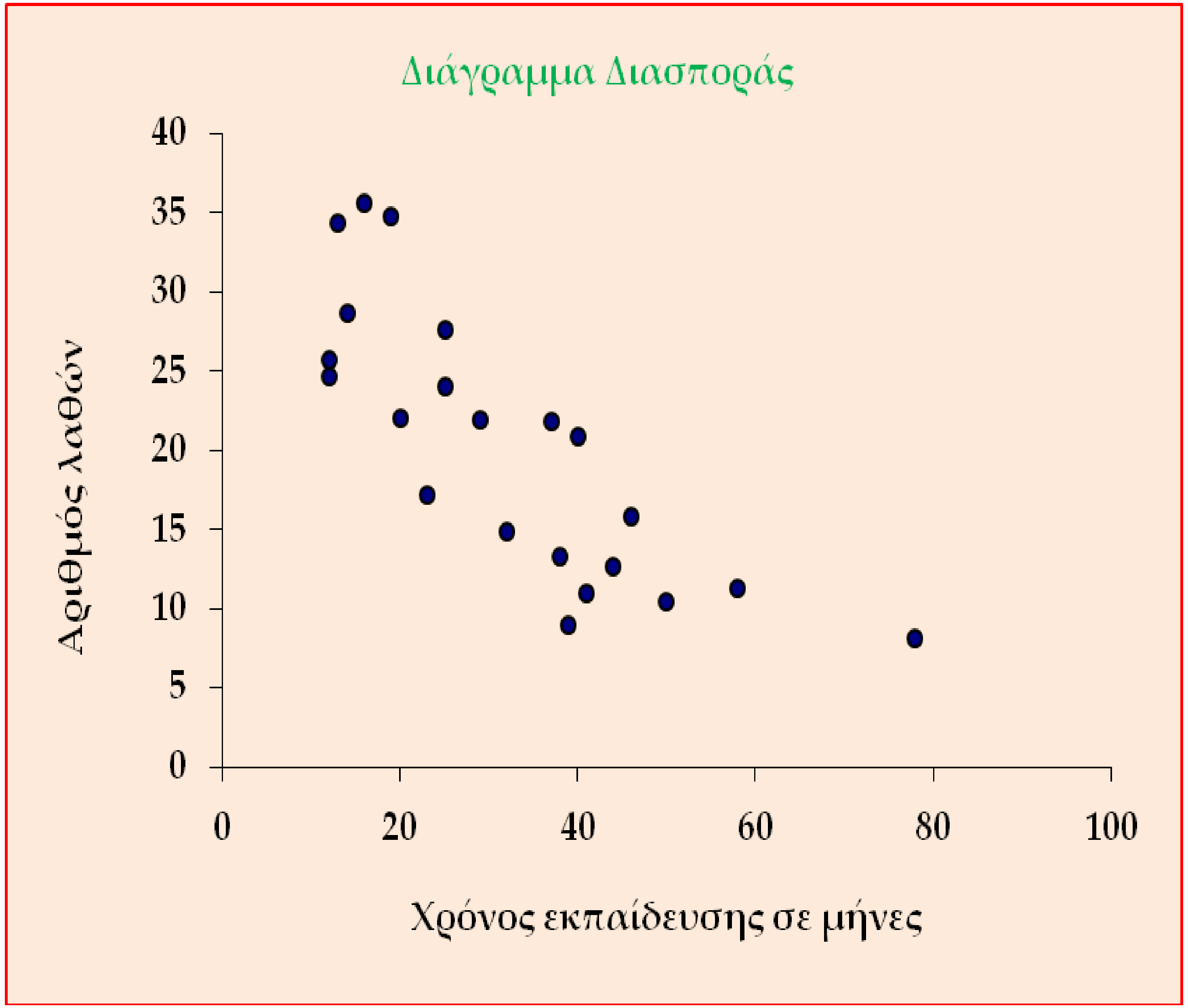
Κατηγορίες Εξόδων Επιχείρησης	Ποσά σε ευρώ/έτος	%
Ά ύλες	60.000	24%
Μισθοδοσία	150.000	60%
Γενικά Βιομηχανικά έξοδα	40.000	16%
Σύνολο	250.000	100%



Διάγραμμα διασποράς

Χρησιμοποιείται για να **απεικονίσει τη σχέση μεταξύ δύο μεταβλητών**. Στις περισσότερες εφαρμογές, η ανεξάρτητη μεταβλητή απεικονίζεται στον οριζόντιο άξονα (άξονας x) και η εξαρτώμενη μεταβλητή στον κατακόρυφο άξονα (άξονας y). Μπορεί να μας βοηθήσει να προσδιορίσουμε εάν υπάρχει συσχέτιση (σύνδεση) μεταξύ των δύο μεταβλητών και ποιος είναι ο τύπος της συσχέτισης (θετική, αρνητική, μη γραμμική, κλπ.).

Η γραμμή που προκύπτει από την απεικόνιση των δεδομένων σε ένα διάγραμμα διασποράς, **ονομάζεται γραμμή τάσης.** Η γραμμή τάσης μπορεί να οριστεί μέσω μιας γραμμικής συνάρτησης, η οποία **ορίζει με ακρίβεια τη σχέση μεταξύ των δύο μεταβλητών.** Όσο **πιο κοντά βρίσκονται τα σημεία δεδομένων στη γραμμή τάσης, τόσο πιο ισχυρή είναι η συσχέτιση μεταξύ των δύο μεταβλητών.**



Γραμμογράφημα

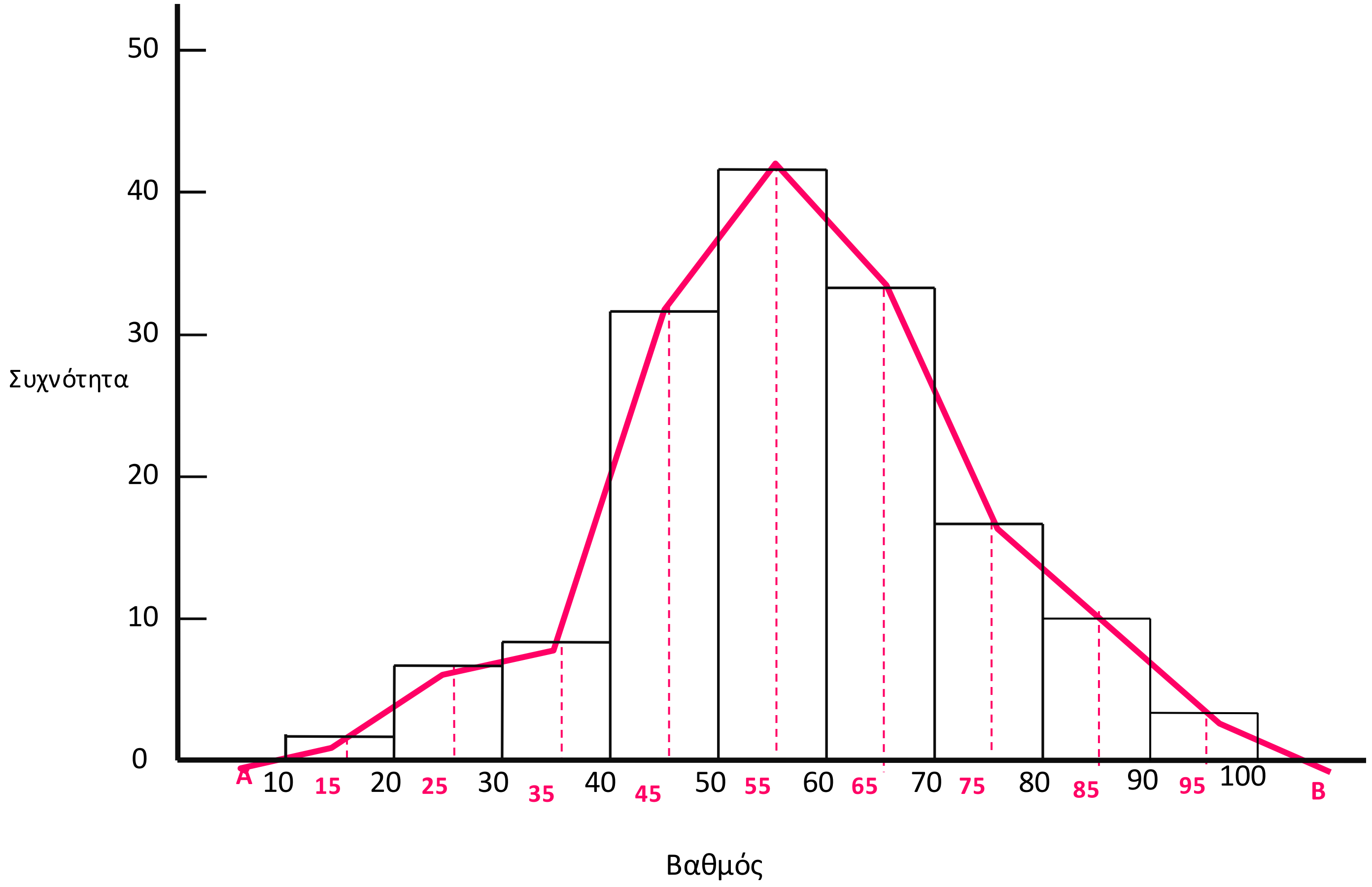
- Τα γραμμογραφήματα χρησιμοποιούνται κυρίως όταν τα δεδομένα μας είναι **χρονολογικές σειρές**
- Ο **οριζόντιος** άξονας εκφράζει τις **χρονικές στιγμές** στις οποίες έχουμε κατατάξει τα δεδομένα και ο **κατακόρυφος** τις αντίστοιχες **συχνότητες**
- Για να κατασκευάσουμε ένα γραμμογράφημα προσδιορίζουμε τις συχνότητες για κάθε κατηγορία (χρονική στιγμή) και τις ενώνουμε με ευθύγραμμα τμήματα

Πολύγωνο Συχνοτήτων

Κατασκευή πολυγωνική γραμμής

❖ Τοποθετούμε στον οριζόντιο άξονα τους κεντρικούς όρους των τάξεων της κατανομής. Υψώνουμε στα σημεία αυτά κάθετα ευθύγραμμα τμήματα ίσα προς τις αντίστοιχες συχνότητες των τάξεων. Ενώνουμε με ευθύγραμμα τμήματα τα διαδοχικά σημεία που έχουν προκύψει. Η πολυγωνική γραμμή είναι μία κλειστή τεθλασμένη.

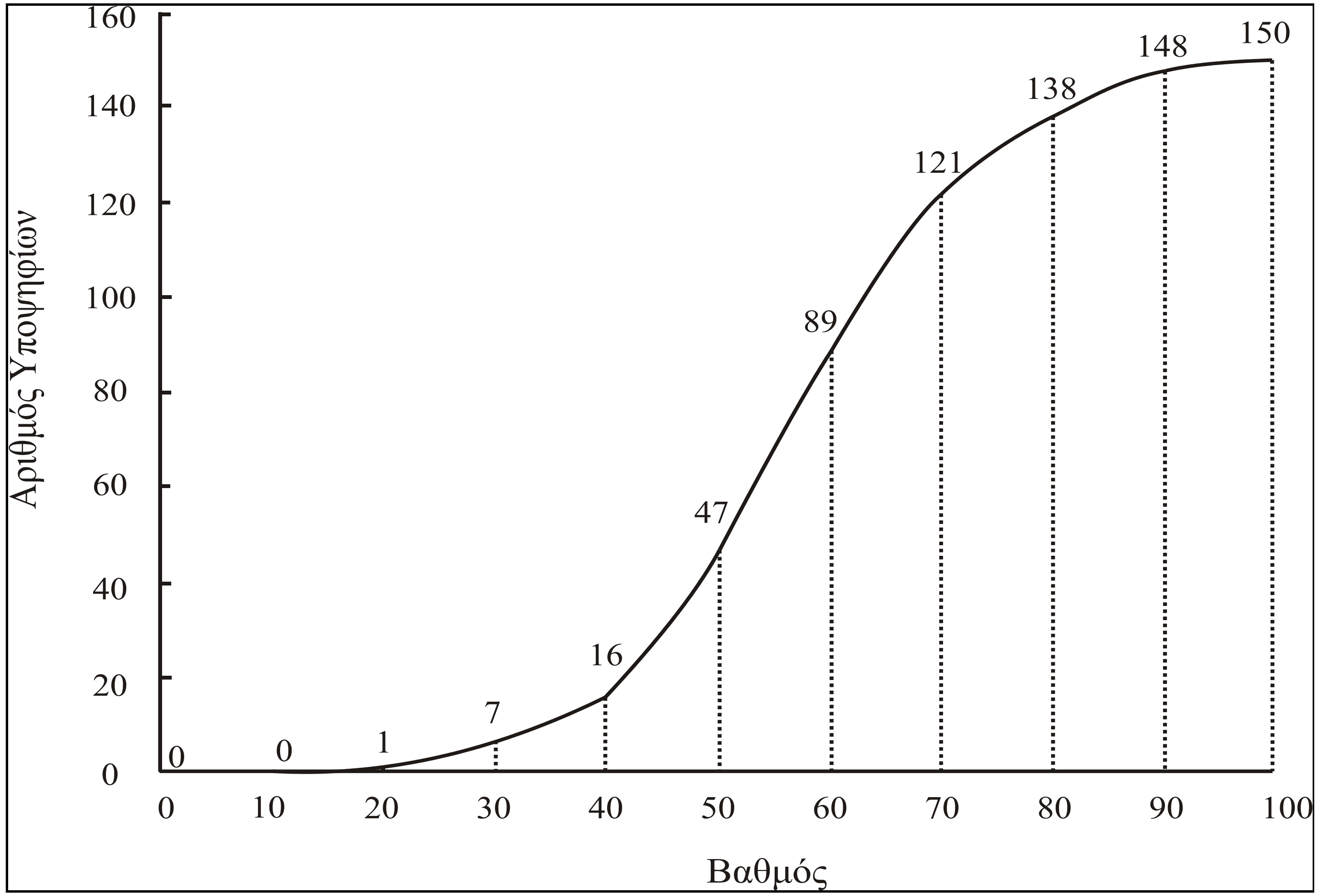
❖ Για να το πετύχουμε αυτό θεωρούμε δύο υποθετικές τάξεις με συχνότητα **μηδέν στα δύο άκρα** της κατανομής και συνδέουμε τα άκρα **Α και Β** της τεθλασμένης που έχει ήδη προκύψει με τους κεντρικούς όρους των δύο αυτών υποθετικών τάξεων.



Αθροιστική κατανομή συχνοτήτων

Η **αθροιστική κατανομή συχνοτήτων** απεικονίζει τον αριθμό των παρατηρήσεων που έχουν **τιμή ίση ή μικρότερη** από μια συγκεκριμένη τιμή. Η αθροιστική κατανομή συχνοτήτων **δημιουργείται προσθέτοντας τις συχνότητες της κατανομής συχνοτήτων**

Η **αθροιστική κατανομή συχνοτήτων** χρησιμοποιείται για να βρεθούν οι **συσσωρευμένες συχνότητες για καθεμία από τις τιμές της μεταβλητής**



Αριθμητικές Μέθοδοι Σύνοψης Δεδομένων

Τα **αριθμητικά μέτρα** που αναφέρονται σε πληθυσμό ονομάζονται **παράμετροι**, ενώ τα αντίστοιχα μέτρα που αναφέρονται **σε δείγμα** υπολογίζονται ως **τιμές στατιστικών συναρτήσεων**

Με τη βοήθεια των δειγματικών δεδομένων υπολογίζουμε στατιστικές συναρτήσεις που δίνουν μια θεωρητική εικόνα της δειγματικής κατανομής αποτελούν τη βάση της στατιστικής συμπερασματολογίας για τον πληθυσμό

Τα αριθμητικά μέτρα χωρίζονται σε **πέντε** κατηγορίες:

- Μέτρα Κεντρικής Τάσης
- Μέτρα Διασποράς
- Μέτρα Ασυμμετρίας
- Μέτρα Κύρτωσης
- Μέτρα Συγκέντρωσης

Μέτρα Κεντρικής Τάσης: Αριθμητικός μέσος

Τα μέτρα κεντρικής τάσης αποτελούν **μια ένδειξη της τιμής** γύρω από την οποία **τείνουν να συσσωρεύονται τα δεδομένα**

Ο **αριθμητικός μέσος**:

- Μπορεί να υπολογισθεί για οποιοδήποτε σύνολο δεδομένων
- Κάθε σύνολο δεδομένων έχει έναν και μοναδικό μέσο αριθμητικό
- Επιδέχεται μαθηματικούς χειρισμούς
- Λαμβάνει υπόψη του όλα τα δεδομένα
- Όμως επηρεάζεται από τυχόν ακραίες τιμές των δεδομένων

Δειγματικός Α.Μ

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

Α.Μ. Πληθυσμού

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

x_i	Την τιμή της i μονάδας του πληθυσμού
X_i	Την τιμή της i παρατήρησης του δείγματος
N	Το πλήθος των μονάδων του πληθυσμού
n	Το πλήθος των παρατηρήσεων του δείγματος

Σταθμικός Αριθμητικός Μέσος

Για τον υπολογισμό του **σταθμικού αριθμητικού μέσου** (*weighted arithmeticmean*) οι παρατηρήσεις αθροίζονται αφού **πρώτα πολλαπλασιασθούν** με κάποιους **συντελεστές στάθμισης**, οι οποίοι εκφράζουν **την σπουδαιότητα** κάθε παρατήρησης στον προσδιορισμό της μέσης τιμής τους, και το άθροισμα τους **διαιρείται με το άθροισμα των συντελεστών στάθμισης**

Δειγματικός Σταθμικός Α.Μ Πληθυσμού

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

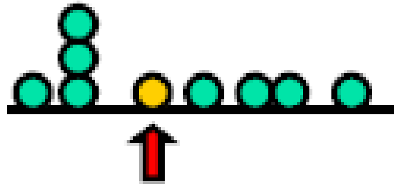
Σταθμικός Α.Μ.

$$\overline{X} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

wi = 1, 2, … είναι οι συντελεστές στάθμισης

Μέτρα Κεντρικής Τάσης

Η **Διάμεσος (median) M** ενός συνόλου μετρήσεων είναι η **τιμή** εκείνη που όταν οι παρατηρήσεις τοποθετηθούν σε **σειρά τάξης μεγέθους** τις χωρίζει σε **δύο μέρη** έτσι ώστε το πολύ **50% των μετρήσεων** είναι **μικρότερες** από την τιμή αυτή και το **πολύ 50% των μετρήσεων μεγαλύτερες** από την τιμή αυτή



Προσδιορισμός διαμέσου

Διάταξη του συνόλου των παρατηρήσεων κατά αύξουσα σειρά μεγέθους

- Αν n περιττός,

M: η τιμή της παρατήρησης στη θέση $\frac{n}{2}+\frac{1}{2}$ δηλαδή $M = X_{\left(\frac{n}{2}+\frac{1}{2}\right)}$.

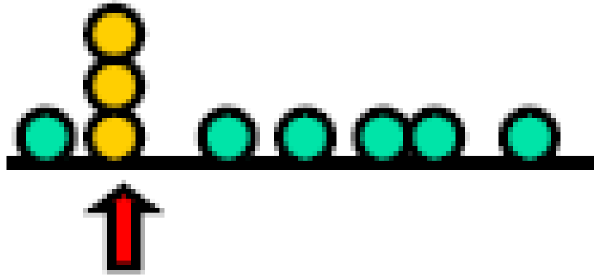
- Αν n άρτιος,

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

Μέτρα Κεντρικής Τάσης

Ως **επικρατούσα τιμή** (mode) T_o ενός συνόλου δεδομένων χαρακτηρίζουμε την παρατήρηση με την υψηλότερη συχνότητα εμφάνισης

$$T_0$$



- Όταν τα δεδομένα έχουμε **μια** **επικρατούσα τιμή**, η κατανομή τους ονομάζεται **μονοκόρυφη**

- Στην περίπτωση που υπάρχουν **δύο ή περισσότερες** τιμές με την ίδια συχνότητα εμφάνισης τότε λέμε ότι τα δεδομένα έχουν δύο ή περισσότερες επικρατούσες τιμές. Η κατανομή τους τότε ονομάζεται **δικόρυφη** (bimodal) ή **πολυκόρυφη** (multimodal) αντίστοιχα

Παράδειγμα

Έχουμε τις συγκεκριμένες τιμές ακινήτων (σε ευρώ):1.000.000, 400.000, 300.000, 150.000, 100.000, 100.000 και 50.000

Άθροισμα: 1.000.000+ 400.000+300.000+150.000+ 100.000+ 100.000+50.000=2.100.000

Αριθμητικός μέσος: Άθροισμα/πλήθος παρατηρήσεων= 2.100.000/7=300.000

Επικρατούσα τιμή: 100.000

Διάμεσος: 50.000, 100.000, 100.000, **150.000**, 300.000, 400.000, 1.000.000

Άσκηση

Δίνονται παρακάτω οι τελικές βαθμολογίες ενός τυχαίου δείγματος **20 φοιτητών** του ΕΚΠΑ που έχουν εγγραφεί στο μάθημα «Διοίκηση Επιχειρήσεων» 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

Να βρεθεί ο μέσος όρος της τελικής βαθμολογίας των φοιτητών

Λύση:

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

Ερώτηση

Δίνονται οι παρατηρήσεις:

3, 5, 8, 15, 12, 14, 20, 25, 26

Αρχικά διατάσσουμε τις παρατηρήσεις σε αύξουσα σειρά

3, 5, 8, 12, 14, 15, 20, 25, 26

Στη συνέχεια παρατηρούμε ότι το πλήθος των παρατηρήσεων είναι **περιττό**, οπότε η **διάμεσος** δίνεται από τη **μεσαία** παρατήρηση

Άσκηση

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

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

$T_0 = 3$

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 (**διάμεσος**)

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

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

Παράδειγμα

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

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

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

Το 1ο τεταρτημόριο είναι η τιμή που βρίσκεται στη θέση

Στη συγκεκριμένη περίπτωση **AQ = 5** και **ΔQ = 0,25**

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

Παράδειγμα

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

$$\mathrm{Q}_2 = \mathrm{X}_{10} + 0,5[3-3] = 3+0 = 3$$

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

$$\mathrm{Q}_3 = \mathrm{X}_{15} + 0,75[6-5] = 5+0,75 = 5,75$$

Διακύμανση

- Η **Διακύμανση** (*Variance*) βασίζεται στην έννοια της **απόστασης** μιας παρατήρησης από **τον μέσο** αριθμητικό των παρατηρήσεων

- *Ξεπερνά το πρόβλημα του μηδενικού αθροίσματος αποκλίσεων χρησιμοποιώντας τα τετράγωνα των αποκλίσεων τα οποία έχουν πάντοτε μη αρνητικές τιμές*

Διακύμανση Πληθυσμού

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

Διακύμανση Δείγματος

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

Εδώ χρησιμοποιήσουμε ως διαιρέτη το n-1 αντί του n γιατί η τιμή του s^2 στην οποία καταλήγουμε είναι καλύτερη εκτίμηση, από την προηγούμενη

$$\boxed{s^2 = \frac{\sum_{i=1}^{N}(X_i - \overline{X})^2}{n-1}}$$

Μέτρα διασποράς

Τα μέτρα κεντρικής τάσης δίνουν μια ένδειξη της τιμής **γύρω από την οποία τείνουν να συσσωρεύονται τα δεδομένα**. Είναι απαραίτητο να γνωρίζουμε και το **πώς κατανέμονται** οι παρατηρήσεις **γύρω** από ένα μέτρο κεντρικής τάσης. Την πληροφορία αυτή μας την παρέχουν τα **μέτρα διασποράς**

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

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

Τυπική Απόκλιση

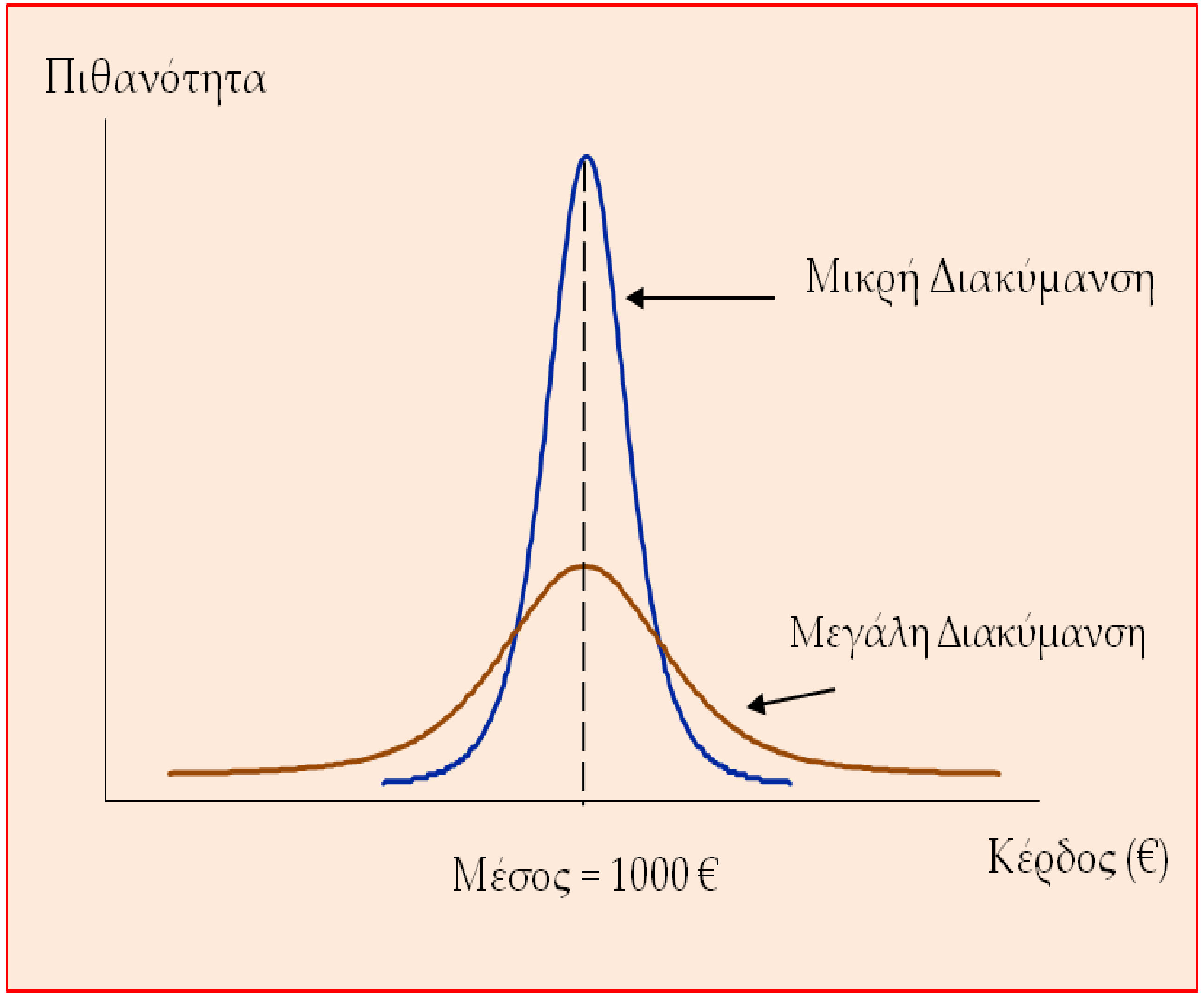
Θετική τετραγωνική ρίζα της διακύμανσης

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

Διασπορά

Η διακύμανση είναι ένα **μέτρο της διασποράς** των τιμών μιας μεταβλητής **γύρω από τον μέσο όρο της**. Όσο **μικρότερη** είναι η διασπορά, **τόσο πιο συγκεντρωμένες είναι οι τιμές της μεταβλητής γύρω από τον μέσο όρο**. Αυτό σημαίνει ότι οι **τιμές είναι πιο κοντά στον μέσο όρο και η διακύμανση είναι μικρότερη**.

Για παράδειγμα, εάν έχουμε μια κατανομή τιμών με μικρή διασπορά, οι περισσότερες τιμές θα βρίσκονται κοντά στον μέσο όρο. Αντίθετα, εάν η διασπορά είναι μεγάλη, οι τιμές θα είναι πιο διάσπαρτες και οι τιμές θα απέχουν περισσότερο από τον μέσο όρο. Η διασπορά επηρεάζεται από την συμμετρία της κατανομής.



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

Παράδειγμα

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

Ο **συντελεστής μεταβλητότητας** (*coefficient of variation*) συμβολίζεται με **CV** και εκφράζεται ως ποσοστό της τυπικής απόκλισης προς τον αντίστοιχο αριθμητικό μέσο

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

- ❖ Μέτρο σχετικής μεταβλητότητας των δεδομένων
- ❖ 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
- ❖ **Όσο μεγαλύτερη η τιμή του CV**, τόσο **μεγαλύτερη η μεταβλητότητα των τιμών την κατανομής**
- ❖ Ο CV είναι ανεξάρτητος από την μονάδα μέτρησης του χαρακτηριστικού και εκφράζει την τυπική απόκλιση ως ποσοστό του αριθμητικού μέσου

Συντελεστής μεταβλητότητας

Σύγκριση συντελεστών μεταβλητότητας

	Μέση τιμή	Τυπική απόκλιση
Μετοχή Α	€200	€40
Μετοχή Β	$500	$80

Συντελεστής μεταβλητότητας

Σύγκριση συντελεστών μεταβλητότητας

	Μέση τιμή	Τυπική απόκλιση	100 * CV
Μετοχή Α	€200	€40	100*€40/€200=20%
Μετοχή Β	$500	$80	100*$80/$500=16%

Η μετοχή Α είναι πιο μεταβλητή από τη μετοχή Β

Μέτρα ασυμμετρίας

- Μια κατανομή θεωρείται **συμμετρική** εάν η μία πλευρά της είναι καθρέφτης της άλλης. Ο αριθμός των τιμών δεδομένων **κάτω από το μέσο όρο είναι ίσος με τον αριθμό των τιμών δεδομένων πάνω από το μέσο όρο**

- Ένα μέτρο **ασυμμετρίας** μας βοηθάει να προσδιορίσουμε **πόσο** και **προς ποια κατεύθυνση** αποκλίνει η κατανομή μας από την συμμετρική ως προ του αριθμητικό μέσο κατανομή.

- Η **ασυμμετρία** αναφέρεται σε μια κατανομή που **δεν είναι συμμετρική**. Μπορεί να είναι είτε δεξιά είτε αριστερή. Μια κατανομή **με δεξιά ασυμμετρία έχει περισσότερες τιμές δεδομένων κάτω από το μέσο όρο από ό,τι πάνω από το μέσο όρο.** Μια κατανομή με **αριστερή ασυμμετρία** έχει περισσότερες τιμές δεδομένων πάνω από το μέσο όρο από ό,τι κάτω από το μέσο όρο.

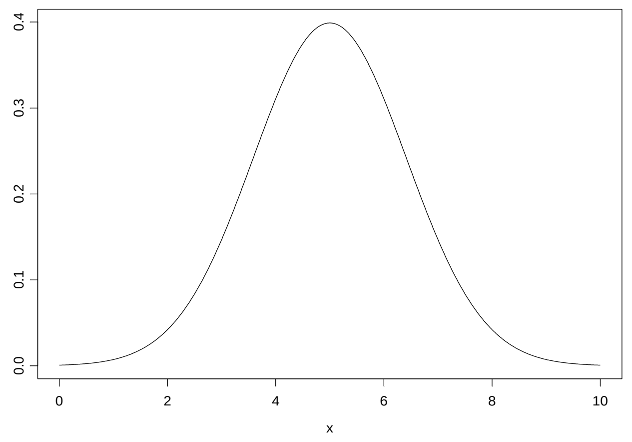
- Η ασυμμετρία μια κατανομής μπορεί να είναι **θετική** ή **αρνητική**

- **Θετικά ασυμμετρική** είναι μία κατανομή όταν η **μεγάλη συγκέντρωση** των παρατηρήσεων της βρίσκεται στις **μικρές τιμές** της μεταβλητής. Αντίθετα στην **αρνητικά ασυμμετρική** η **μεγάλη συγκέντρωση** των παρατηρήσεων βρίσκεται στις **μεγάλες τιμές** της μεταβλητής

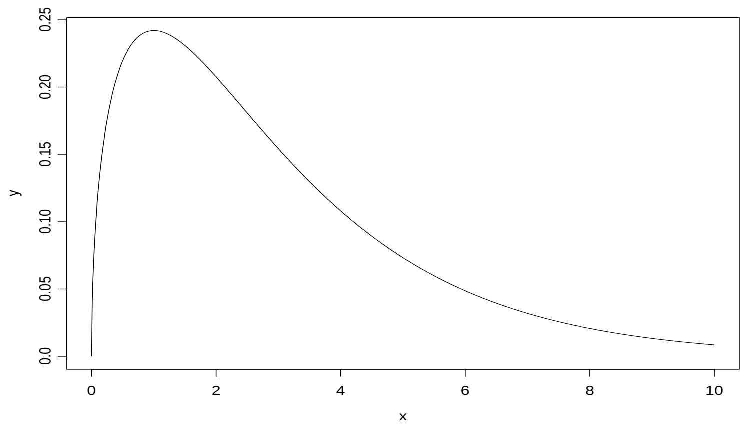
Συμμετρία - Ασυμμετρία

Συμμετρική κατανομή

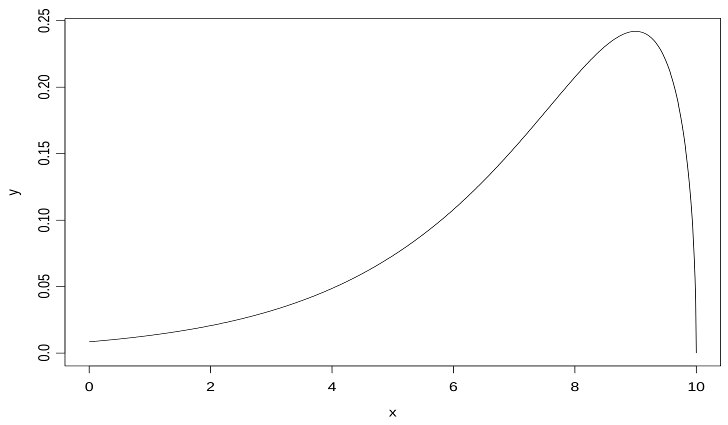
Στατιστικά, μια **συμμετρική κατανομή** σημαίνει ότι **τα δεδομένα είναι ομοιόμορφα κατανεμημένα προς την αριστερή και δεξιά πλευρά του μέσου όρου.** Η μορφή της κατανομής είναι σαν μια καμπύλη με ένα κέντρο που αντιστοιχεί στον μέσο όρο.



Κατανομή με δεξιά ασυμμετρία



Κατανομή με αριστερή ασυμμετρία



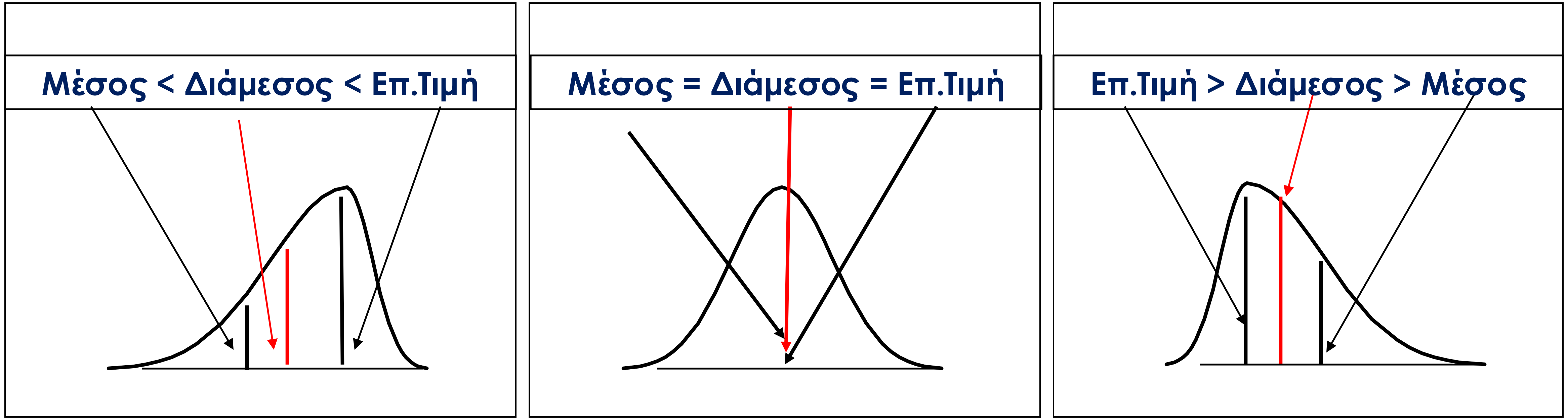
Η **ασυμμετρία μιας κατανομής** υποδηλώνει πόσο διαφορετικά είναι τα δεδομένα από μια συμμετρική κατανομή. Μια **θετική ασυμμετρία** σημαίνει **ότι η ουρά της κατανομής εκτείνεται προς τα δεξιά του μέσου όρου, ενώ μια αρνητική ασυμμετρία υποδηλώνει ότι η ουρά εκτείνεται προς τα αριστερά**

Συντελεστής ασυμμετρίας

Περιγράφει τον τρόπο που κατανέμονται οι τιμές της κατανομής. Έχουμε 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

- ❖ **Μέτρο ασυμμετρίας βασιζόμενο στον αριθμητικό μέσο και στην επικρατούσα τιμή**
- ❖ **Ο συντελεστής αυτός μας βοηθά να κατανοήσουμε τον βαθμό ασυμμετρίας μίας κατανομής**

Συντελεστής ασυμμετρίας Pearson

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

- ❖ **Θετική** ασυμμετρία, εάν ***Sp* > 0**
- ❖ **Συμμετρία**, εάν ***Sp*= 0**
- ❖ **Αρνητική** ασυμμετρία, εάν ***Sp*< 0**

Συντελεστής ασυμμετρίας *β₃* του Pearson

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

- ❖ **Θετική** ασυμμετρία, εάν β_3 **> 0**
- ❖ **Συμμετρία**, εάν β_3**= 0**
- ❖ **Αρνητική** ασυμμετρία, εάν β_3 **< 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.

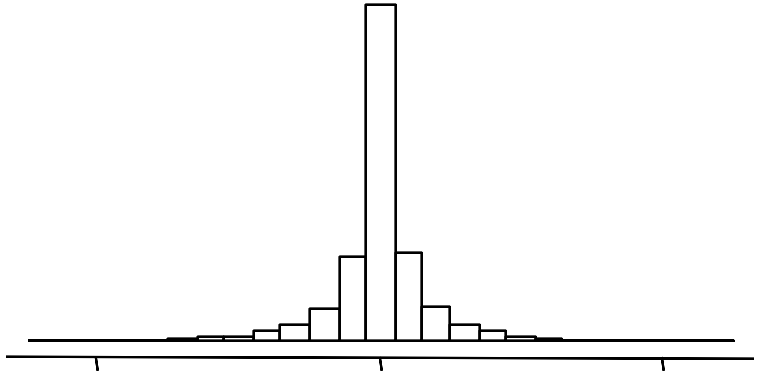
Ως μέτρο κύρτωσης της κατανομής **n** παρατηρήσεων ο Κ. Pearson τον συντελεστή:

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

- Εάν β_4 **< 3** τότε έχουμε **πλατύκυρτη** κατανομή (χαμηλή κορυφή)
- Εάν β_4 **= 3** τότε έχουμε **μεσόκυρτη** κατανομή (κανονική κατανομή)
- Εάν β_4 **> 3** τότε έχουμε **λεπτόκυρτη** κατανομή (οξεία κορυφή)

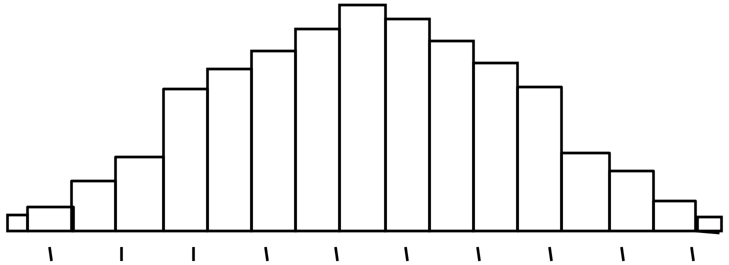
Συντελεστής κύρτωσης

Λεπτόκυρτη κατανομή



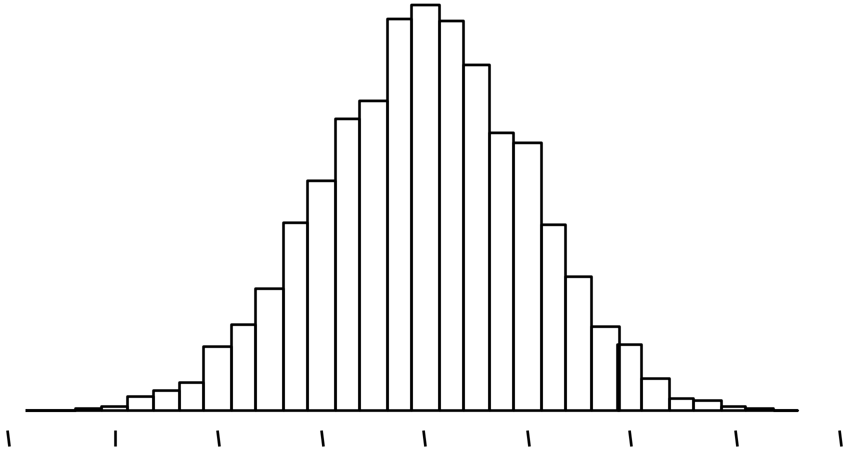
Στις λεπτόκυρτες κατανομές, **οι τιμές είναι πιο συγκεντρωμένες γύρω από το μέσο**, με μια πιο οξεία κορυφή. **Ο συντελεστής κύρτωσης έχει θετική τιμή**, δηλαδή **μεγαλύτερη από 0.**

Πλατύκυρτη κατανομή



Στις πλατύκυρτες κατανομές, **οι τιμές είναι πιο διασκορπισμένες γύρω από το μέσο**. Η κορυφή της κατανομής **είναι πιο επίπεδη σε σύγκριση με την μεσόκυρτη**. Ο συντελεστής κύρτωσης **έχει αρνητική τιμή**, δηλαδή **λιγότερο από 0.**

Μεσόκυρτη κατανομή



Μία μεσόκυρτη κατανομή έχει κορυφή που **μοιάζει με την κορυφή ενός κουδουνιού**. Η κορυφή της κατανομής **είναι συγκεντρωμένη στο μέσο**, **ενώ οι τιμές που απέχουν από το μέσο μειώνονται σταδιακά.** **Ο συντελεστής κύρτωσης σε αυτή την περίπτωση είναι 0**.

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)

Ερώτηση

Αν η διακύμανση ενός πληθυσμού είναι **4**, τότε η τυπική απόκλιση είναι:

α) 2
β) 4
γ) 16
δ) 8

Η σωστή απάντηση είναι η **(α)**

Η τυπική απόκλιση είναι η τετραγωνική ρίζα της διακύμανσης

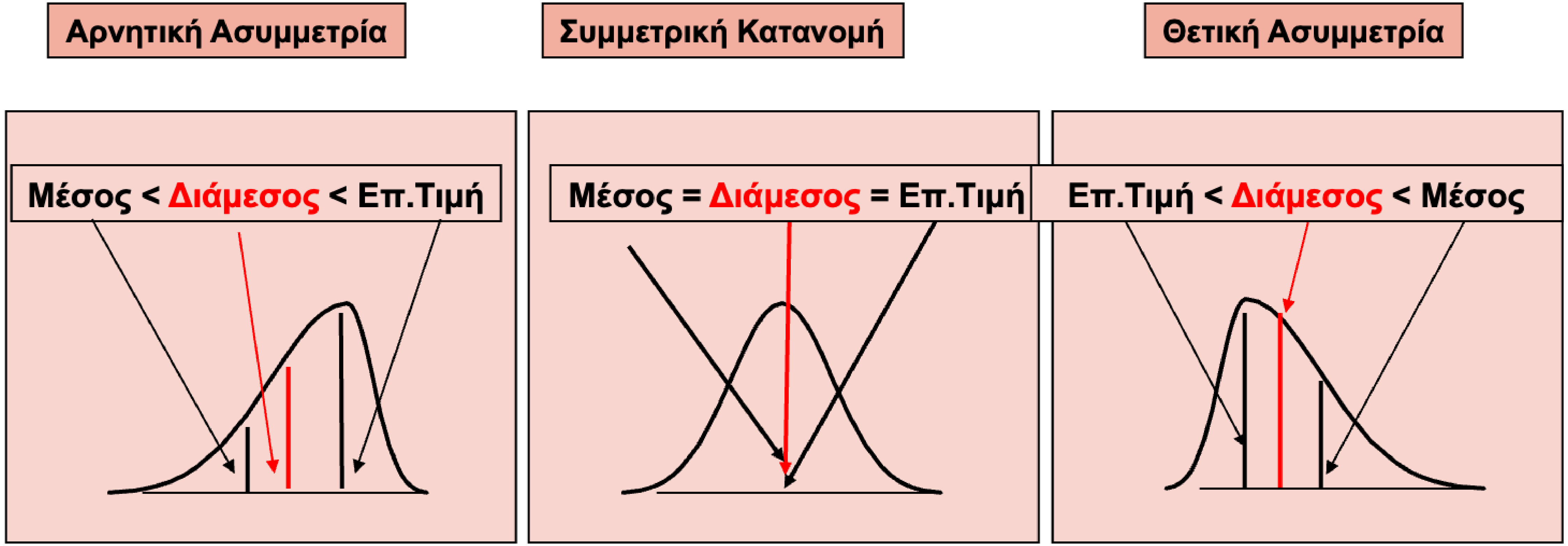
$$s = +\sqrt{s^2}$$

Ερώτηση

Ο **μέσος** μιας κατανομής είναι **7**, η **διάμεσος** είναι **9** και η **επικρατούσα τιμή** είναι **12**. Τότε η κατανομή αυτή:

α) παρουσιάζει θετική ασυμμετρία
β) παρουσιάζει αρνητική ασυμμετρία
γ) είναι συμμετρική
δ) ακολουθεί την κανονική κατανομή

Απάντηση



Η σωστή απάντηση είναι η **(β)**

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 **ci** είναι το **κέντρο** της κάθε τάξης και **fi** 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

Όρια 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:

Όρια 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	12.312
21-25	57	23	1.311	529	30.153
26-30	56	28	1.568	784	43.904
31-35	45	33	1.485	1.089	49.005
36-40	42	38	1.596	1.444	60.648
41-45	40	43	1.720	1.849	73.960
46-50	38	48	1.824	2.304	87.552
51-55	15	53	795	2.809	42.135
56-60	18	58	1.044	3.364	60.552
Σύνολο	**349**		**12.027**	**14.496**	**460.221**

Ομαδοποιημένα δεδομένα - παράδειγμα

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

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

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

Για περισσότερη μελέτη και εξάσκηση

Από το βιβλίο του κ. Θ. Παπαδόγγονα, «Ποσοτικές μέθοδοι ανάλυσης επιχειρηματικών αποφάσεων». Ερωτήσεις πολλαπλής επιλογής και ασκήσεις, σελίδες: 38-42 και 359-373

Βιβλιογραφικές αναφορές

Θ. Παπαδόγγονας, «Ποσοτικές μέθοδοι ανάλυσης επιχειρηματικών αποφάσεων", 2η Έκδοση, Εκδόσεις Τσότρας, 2022