

ΟΜΑΔΟΠΟΙΗΣΗ ΤΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ

- ◆ Πότε κάνουμε ομαδοποίηση των παρατηρήσεων;

Απάντηση:

Όταν το πλήθος των τιμών μιας μεταβλητής είναι αρκετά μεγάλο κάνουμε **ομαδοποίηση** των παρατηρήσεων. Αυτό συμβαίνει είτε στην περίπτωση μιας **διακριτής** μεταβλητής πολύ δε περισσότερο όταν η μεταβλητή είναι **συνεχής** οπότε μπορεί να πάρει να πάρει οποιαδήποτε τιμή μέσα σ' ένα διάστημα.

- ◆ Να δώσετε του ορισμούς των παρακάτω όρων:

- | | |
|---------------------|----------------------|
| i. Κλάσεις | ii. Όρια των κλάσεων |
| iii. Κεντρική τιμή | iv. πλάτος κλάσης |
| v. συχνότητα κλάσης | |

Απάντηση:

- i. **Κλάσεις** ονομάζονται τα **διαστήματα** της μορφής $[\alpha, \beta)$. Μια παρατήρηση ανήκει στην κλάση $[\alpha, \beta)$ αν η τιμή της x είναι:

$$\alpha \leq x < \beta$$

- ii. Τα **όρια των κλάσεων** είναι τα άκρα του διαστήματος της κλάσης. Αν η κλάση είναι η $[\alpha, \beta)$, τότε τα όρια της είναι οι αριθμοί **α** και **β** .

- iii. **Κεντρική τιμή** μιας κλάσης είναι το **ημιάθροισμα των ορίων** της. Αν για παράδειγμα η κλάση είναι η $[\alpha, \beta)$, η κεντρική τιμή που της αντιστοιχεί είναι:

$$x = \frac{\alpha + \beta}{2}$$

- iv. **Πλάτος** μιας κλάσης C λέγεται η **διαφορά** του **κάτω άκρου** (ορίου) από το **άνω άκρο**. Για την κλάση $[\alpha, \beta)$ το πλάτος είναι:

$$C = \beta - \alpha$$

- v. **Συχνότητα** μιας κλάσης είναι ο **φυσικός αριθμός** που δείχνει πόσες παρατηρήσεις ανήκουν στην κλάση.

- ◆ Πως κατασκευάζουμε κλάσεις ίσου πλάτους;

Απάντηση:

Βρίσκουμε το πραγματικό εύρος R δείγματος των παρατηρήσεων που είναι η διαφορά της μεγαλύτερης τιμής M της μεταβλητής από τη μικρότερη m ($R = M - m$). Αν k είναι ο αριθμός των κλάσεων που θα κατασκευάσουμε τότε το πλάτος των κλάσεων θα είναι:

$$C = \frac{R}{k}$$

Αν ο αριθμός $\frac{R}{k}$ δεν μας εξυπηρετεί για την πρακτική εφαρμογή τότε τον στρογγυλοποιούμε προς τα «πάνω».

Παράδειγμα:

$$\text{Αν } \frac{R}{k} = 7,81 \text{ τότε } C = 8$$

Ξεκινώντας από τη μικρότερη παρατήρηση m (ή λίγο πιο κάτω) και προσθέτοντας κάθε φορά το C , δημιουργούμε της k κλάσεις. Δηλαδή οι κλάσεις θα είναι διαστήματα της μορφής:

$$[m, m + C), [m + C, m + 2C), [m + 2C, m + 3C), \dots$$

Παρατηρήσεις:

1. Για το πλήθος των κλάσεων k χρησιμοποιείται ο τύπος Sturges:

$$k = 1 + 3,32 \cdot \log n$$

όπου n το μέγεθος του δείγματος. Στις ασκήσεις συνήθως το k δίνεται.

2. Καμία παρατήρηση δεν μπορεί να μείνει έξω από κάποια κλάση.
3. Οι κεντρικές τιμές διαφέρουν μεταξύ τους όσο και το πλάτος C των κλάσεων.
4. Μια παρατήρηση που συμπίπτει με το άνω άκρο μιας κλάσης, πρέπει να τοποθετηθεί κατά τη διαλογή στην επόμενη κλάση, εκτός και αν δεν υπάρχει επόμενη κλάση.

ΙΣΤΟΓΡΑΜΜΑ, ΠΟΛΥΓΩΝΑ ΚΑΙ ΚΑΜΠΥΛΕΣ ΣΥΧΝΟΤΗΤΩΝ

- Τι είναι το **ιστόγραμμα** και το **πολύγωνο συχνοτήτων**;

Απάντηση:

Η γραφική παράσταση ενός πίνακα συχνοτήτων με ομαδοποιημένα δεδομένα γίνεται με **ιστόγραμμα**.

Στον οριζόντιο άξονα ενός συστήματος αξόνων σημειώνουμε, με κατάλληλη κλίμακα, τα όρια των κλάσεων και κατασκευάζουμε ορθογώνια

με βάση δυο διαδοχικά όρια και ύψος τη συχνότητα της κλάσης, για κλάσεις ίσου πλάτους.

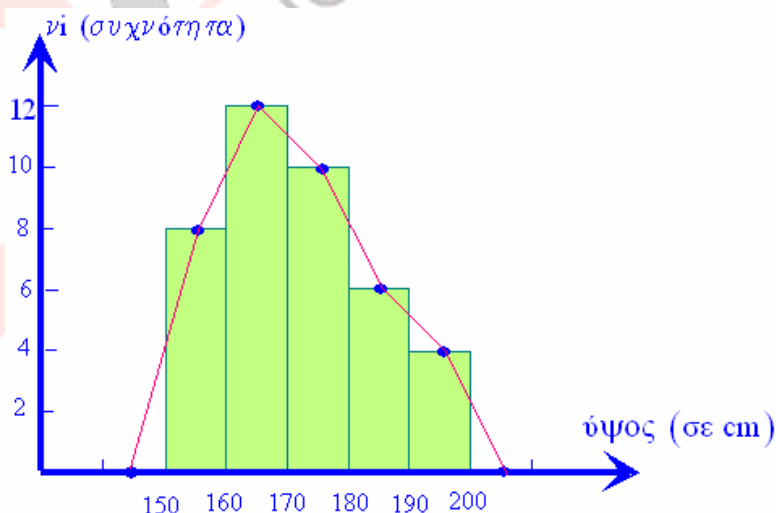
Για να κατασκευάσουμε το πολύγωνο **συχνοτήτων**, θεωρούμε μια υποθετική κλάση του ίδιου πλάτους στην αρχή και μια στο τέλος με συχνότητα μηδέν. Σημειώνουμε τα μέσα των άνω βάσεων των ορθογωνίων και τα ενώνουμε με ευθύγραμμα τμήματα. Η τεθλασμένη γραμμή που προκύπτει λέγεται **πολύγωνο συχνοτήτων**.

Παράδειγμα:

Ο παρακάτω πίνακας συχνοτήτων αφορά τα ύψη 40 μαθητών ενός λυκείου σε cm.

Ύψη (κλάσεις)	Συχνότητες
[150, 160)	8
[160, 170)	12
[170, 180)	10
[180, 190)	6
[190, 200)	4
Σύνολο	40

Το ιστόγραμμα και το αντίστοιχο πολύγωνο συχνοτήτων φαίνεται στο παρακάτω σχήμα:



Παρατηρήσεις:

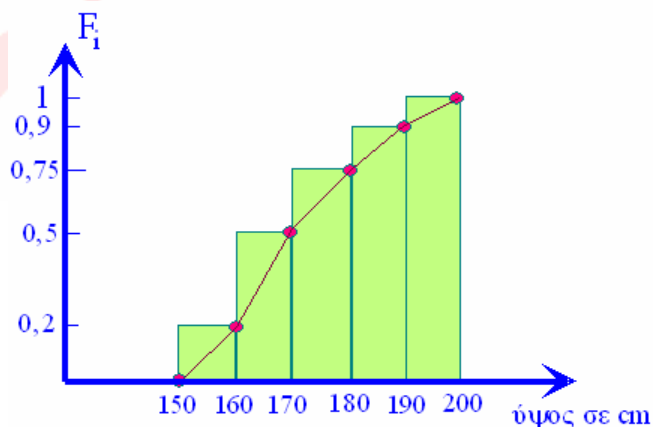
1. Αν θεωρηθεί ως μονάδα το πλάτος των κλάσεων, το **εμβαδόν** όλων των ορθογωνίων καθώς και το **εμβαδόν** χωρίου που περικλείεται από το πολύγωνο συχνοτήτων και τον οριζόντιο άξονα είναι ίσο με το **μέγεθος** του δείγματος.
2. Ανάλογα κατασκευάζονται το **ιστόγραμμα σχετικών συχνοτήτων**, όπου τα ορθογώνια θα έχουν ύψη ίσα με την αντίστοιχη σχετική συχνότητα και το **πολύγωνο των σχετικών συχνοτήτων**.
3. Με τον ίδιο τρόπο μπορούμε να κατασκευάσουμε το ιστόγραμμα αθροιστικών ή αθροιστικών σχετικών συχνοτήτων. Τα πολύγωνα αυτών κατασκευάζονται χωρίς βοηθητικές κλάσεις. Η τεθλασμένη, σ' αυτή την περίπτωση αρχίζει από την αριστερή κάτω κορυφή του πρώτου ορθογωνίου και ενώνοντας όλες τις δεξιά άνω κορυφές καταλήγει στην τελευταία.

Παράδειγμα:

Με βάση τα δεδομένα του προηγούμενου παραδείγματος, έχουμε:

Ύψη [...,...)	Συχνότητες v_i	Σχετικές συχνότητες f_i	Αθροιστικές συχνότητες F_i
150–160	8	0,2	0,2
160–170	12	0,3	0,5
170–180	10	0,25	0,75
180–190	6	0,15	0,9
190–200	4	0,1	1
Σύνολα	40	1	

Το **ιστόγραμμα των αθροιστικών σχετικών συχνοτήτων** και το **πολύγωνο αθροιστικών σχετικών συχνοτήτων** φαίνεται στα παρακάτω σχήμα:



4. Αν σε κάθε κλάση οι παρατηρήσεις **κατανέμονται ομοιόμορφα**, τότε μπορούμε να υπολογίσουμε το ποσοστό ή τον αριθμό των παρατηρήσεων που ανήκουν σ' ένα υποδιάστημα αυτής της κλάσης.

Αν για παράδειγμα στο προηγούμενο παράδειγμα μας ζητούν το ποσοστό των μαθητών με ύψος κάτω από 165 cm και τα ύψη ήταν ομοιόμορφα κατανεμημένα στην κλάση [160,170) με συχνότητα 12, επειδή το 165 αποτελεί το μέσον της κλάσης τα 6 παιδιά (τα μισά δηλαδή) αναμένεται να έχουν ύψος στο διάστημα [160,165). Άρα κάτω από 165 cm ύψος θα έχουν συνολικά:

$$8 + 6 = 14 \text{ παιδιά}$$

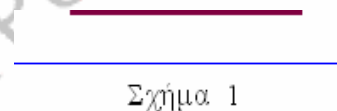
- Τι ονομάζουμε **καμπύλη συχνότητας**;

Απάντηση:

Όταν το πλήθος των κλάσεων είναι πάρα πολύ μεγάλο (τείνει στο άπειρο) και το πλάτος πολύ μικρό (τείνει στο μηδέν) τότε το **πολύγωνο συχνότητας** τείνει να πάρει τη μορφή μιας **λείας καμπύλης** που λέγεται **καμπύλη συχνότητας**.

Παρατηρήσεις:

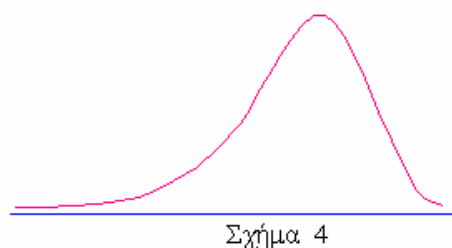
1. Όταν όλες οι παρατηρήσεις εμφανίζονται με τον ίδιο αριθμό φορών, δηλαδή έχουν την ίδια συχνότητα, τότε η καμπύλη συχνότητας έχει τη μορφή που φαίνεται στο σχήμα 1. Αυτή η κατανομή λέγεται **ομοιόμορφη**.



2. Όταν η καμπύλη συχνότητας έχει τη μορφή του σχήματος 2 (κωδωνοειδής) τότε η κατανομή λέγεται **κανονική**.



3. Οι δυο προηγούμενες κατανομές έχουν καμπύλες που παρουσιάζουν κάποια συμμετρία. Όταν η καμπύλη συχνότητας δεν παρουσιάζει συμμετρία η κατανομή λέγεται **ασύμμετρη**.



Η καμπύλη του σχήματος 3 αφορά κατανομή με **θετική ασυμμετρία**, ενώ στο σχήμα 4, έχουμε κατανομή με **αρνητική ασυμμετρία**.