

Κεφάλαιο 4

ΕΜΠΕΙΡΙΚΕΣ ΚΑΤΑΝΟΜΕΣ ΣΥΧΝΟΤΗΤΩΝ ΜΙΑΣ ΜΕΤΑΒΛΗΤΗΣ

4.1 Ασυνεχείς και συνεχείς κατανομές συχνοτήτων

Μετά τη συγκέντρωση των στατιστικών στοιχείων που αναφέρονται σε **μια** μόνο ιδιότητα ενός πληθυσμού το πρώτο βήμα είναι η κατάλληλη κατάταξη και η συστηματική ομαδοποίηση των τιμών της μεταβλητής, με αντικειμενικό σκοπό την παρουσίαση των στοιχείων κατά τέτοιο τρόπο που να δίνει πληροφορίες για τη δομή του πληθυσμού που ερευνάται και να διευκολύνει στην εξαγωγή συμπερασμάτων.

Για την ομαδοποίηση των τιμών της μεταβλητής χρησιμοποιούνται ειδικές κατατάξεις που τις ονομάζουμε *κατανομές συχνοτήτων ή πίνακες συχνοτήτων*.

Ο τρόπος κατασκευής των κατανομών συχνοτήτων εξαρτάται από το είδος της μεταβλητής.

α) Όταν η μεταβλητή είναι ασυνεχής

Αν η μεταβλητή είναι ασυνεχής και το εύρος των τιμών της μεταβλητής είναι μικρό τότε αθροίζουμε τις παρατηρήσεις (συχνότητες) που αναφέρονται στην ίδια τιμή της μεταβλητής και παρουσιάζουμε τα δεδομένα σε πίνακα απλής εισόδου όπου στη πρώτη στήλη τοποθετούμε τις τιμές της μεταβλητής (x_i) και στη δεύτερη στήλη των αριθμό των παρατηρήσεων που αντιστοιχεί σε κάθε τιμή της μεταβλητής (f_i).

Παράδειγμα: Καθένας από τους 30 υπαλλήλους μιας επιχείρησης έλαβε για την θερινή άδεια το 1995 των παρακάτω αριθμό ημερών:

25, 26, 24, 22, 21, 25, 26, 26, 24, 22, 21, 26, 26, 25, 27, 24, 24, 27, 22, 23, 23, 26, 25, 25, 24, 25, 25, 27, 24, 25

Ζητείται ο πίνακας κατανομής των παραπάνω παρατηρήσεων σε συχνότητες.

Λύση:

Το πλήθος όλων των παρατηρήσεών μας είναι: $N = 30$. Τοποθετούμε πρώτα τις παρατηρήσεις κατά φυσική αύξουσα τάξη μεγέθους:

21, 21, 22, 22, 22, 23, 23, 24, 24, 24, 24, 24, 24, 25, 25, 25, 25, 25, 25, 25, 26, 26, 26, 26, 26, 27, 27, 27

Κατόπιν υπολογίζουμε το εύρος της μεταβλητής των παρατηρήσεων δηλαδή την διαφορά ανάμεσα στη μεγαλύτερη και τη μικρότερη τιμή, δηλ $R = 27 - 21 = 6$, η παρουσίαση γίνεται με μορφή ασυνεχούς κατανομής ως εξής:

ΠΙΝΑΚΑΣ

Τιμές μεταβλητής (x_i)	Συχνότητες (f_i)
21	2
22	3
23	2
24	6
25	8
26	6
27	3
Σύνολο	30

Στην περίπτωση λοιπών, της ασυνεχούς μεταβλητής, ο πίνακας συχνοτήτων παίρνει τη μορφή του πίνακα :

Τιμές μεταβλητής (x_i)	Αριθμός παρατηρήσεων ή συχνότητες (f_i)
x_1	f_1
x_2	f_2
.	.
.	.
.	.
x_i	f_i
.	.
.	.
.	.
x_k	f_k
Σύνολο	$\sum_{i=1}^k f_i$

Στον πίνακα αυτό:

1. $x_1, x_2, \dots, x_i, \dots, x_k$ είναι οι τιμές της ασυνεχούς μεταβλητής, τις οποίες έχουμε τοποθετήσει κατά φυσική αύξουσα τάξη μεγέθους, από τη μικρότερη ως τη μεγαλύτερη.
2. $f_1, f_2, \dots, f_i, \dots, f_k$ είναι οι αντίστοιχες συχνότητες της μεταβλητής X , δηλαδή πόσες φορές εμφανίζεται, στο συνολικό πληθυσμό, κάθε τιμή της μεταβλητής X .

Όταν η τιμή x_i της μεταβλητής X εμφανίζεται f_i φορές, λέμε ότι το f_i είναι η απόλυτη συχνότητα της τιμής x_i , ενώ η τιμή του πηλίκου $f_i / \sum f_i$ ονομάζεται σχετική συχνότητα της x_i και συχνότερα παρουσιάζεται σε μορφή ποσοστού.

Ο συμβολισμός $\sum f_i$ σημαίνει το άθροισμα όλων των συχνοτήτων f_i δηλαδή:

$$\sum_{i=1}^k f_i = f_1 + f_2 + \dots + f_k = N,$$

όπου με N συμβολίζεται το πλήθος όλων των παρατηρήσεων.

β) Όταν η μεταβλητή είναι συνεχής:

Στην περίπτωση αυτή, που η μεταβλητή X μπορεί να πάρει οποιαδήποτε τιμή σε ένα διάστημα μεταξύ δύο πραγματικών αριθμών $[α, β]$, είναι δηλαδή συνεχής, καταφεύγουμε στην ομαδοποίηση των παρατηρήσεων.

Η ομαδοποίηση αυτή γίνεται αν θεωρήσουμε μερικές μικρότερες ομάδες τιμών της μεταβλητής, που ονομάζονται και **τάξεις** (κλάσεις), έτσι ώστε η κάθε τιμή των παρατηρήσεων μας να ανήκει μόνο σε μία τάξη. Οι τάξεις θα είναι **κλειστά – ανοικτά διαστήματα** πραγματικών αριθμών.

Κάθε τάξη διακρίνεται από τα **όριά της**, δηλ το **κατώτερο όριο** (a_{i-1}) και το **άνωτερο όριο** (a_i). Θα έχουν δηλαδή τη μορφή $[a_{i-1}, a_i)$.

Παρατήρηση: το κάτω όριο της 1^{ης} τάξης είναι το a_0 και το άνω όριο της πρώτης τάξης είναι το a_1 . Η 2^η τάξη έχει ως κάτω όριο το a_1 και ως άνω όριο το a_2 κλπ. Παρατηρούμε δηλαδή, ότι οι τάξεις καλύπτουν όλο το διάστημα στο οποίο παίρνει τιμές η μεταβλητή χωρίς αλληλοεπικάλυψη (γιατί αυτά είναι κλειστά – ανοικτά) και ότι ο δείκτης του άνω ορίου και η αρίθμηση των τάξεων είναι αυτά που συμπίπτουν !!

Οι παρατηρήσεις κάθε τάξης θεωρούνται όμοιες και ότι είναι ομοιόμορφα διασκορπισμένες στο διάστημά τους, οπότε μπορούν να «αντιπροσωπευτούν» από τις **κεντρικές τιμές**, τα κέντρα δηλαδή των τάξεων. Η κεντρική τιμή της κάθε τάξης υπολογίζεται από την σχέση : $x_i = \frac{a_i + a_{i-1}}{2}$ και έχουν την ίδια έννοια και χρήση που έχουν οι τιμές της ασυνεχούς μεταβλητής.

Γενικός κανόνας για τον αριθμό τάξεων μιας κατανομής δεν υπάρχει δίνεται όμως κατά προσέγγιση από τον εμπειρικό τύπο: $K=1+3,322 \log_{10} N$ (κανόνας Sturges) όπου K ο αριθμός τάξεων και N ο αριθμός των παρατηρήσεων.

Στην πράξη όμως ο αριθμός τάξεων προσδιορίζεται με βάση τα κριτήρια της **ομοιογένειας** και της **απλότητας**. Εμπειρικά λοιπόν οι τάξεις είναι από 5 έως 10 ή το πολύ-πολύ 12 (βέλτιστο πλήθος τάξεων είναι 6 έως 8).

Η διαφορά μεταξύ του άνω και του κάτω ορίου της κάθε τάξης ($a_i - a_{i-1}$) ονομάζεται **πλάτος της τάξης** (δ_i). Στις περισσότερες περιπτώσεις το πλάτος (δ) όλων των τάξεων της κατανομής είναι το ίδιο.

Παράδειγμα: Η ωριαία αποζημίωση (σε δραχμές) καθενός από τους 40 υπαλλήλους μιας επιχείρησης έχει ως εξής:

31, 34, 45, 46, 42, 44, 50, 53, 54, 55, 56, 57, 60, 61, 62, 63, 64, 65, 66, 67, 68, 75, 74, 76, 59, 85, 85, 84, 86, 90, 99, 88, 78, 87, 92, 93, 94, 95, 96, 89,

Να γίνει ομαδοποίηση με μορφή κατανομής συχνοτήτων σε τάξεις και να υπολογιστούν οι κεντρικές τιμές.

Λύση:

α. Κατατάσσουμε τα μεγέθη κατά αύξουσα σειρά.

β. Βρίσκουμε τη διαφορά ανάμεσα στη μεγαλύτερη τιμή $M=99$ και τη μικρότερη τιμή $E=31$. Η διαφορά αυτή ονομάζεται **εύρος κατανομής** δηλ $R=M - E = 99 - 31 = 68$. Επειδή

το εύρος (68) είναι μικρό, 7 τάξεις ικανοποιούν το κριτήριο της απλότητας και ομοιογένειας το εύρος κάθε τάξης είναι $\delta=68/7 \approx 10$ κατά συνέπεια θα έχουμε:

ΠΙΝΑΚΑΣ

Τάξεις	Συχνότητες (f_i)	Κεντρικές τιμές των τάξεων (x_i)	Σχετικές συχνότητες
[30- 40)	2	35	5%
[40- 50)	4	45	10%
[50- 60)	7	55	17,5%
[60- 70)	9	65	22,5%
[70- 80)	4	75	10%
[80- 90)	7	85	17,5%
[90-100)	7	95	17,5%
Σύνολο	40		100%

Σημείωση: τα ανώτερα όρια των τάξεων δεν περιέχονται σ' αυτές.

Αν θέλουμε, όπως και περίπτωση των διακριτών μεταβλητών, να υπολογίσουμε την σχετική συχνότητα σε ποσοστό επί της % π.χ. για την 2η τάξη έχουμε:

$$\frac{f_2}{\sum f_i} \cdot 100\% = \frac{4}{40} \cdot 100\% = 10\%.$$
 Θέλει **προσοχή στις στρογγυλοποιήσεις, γιατί το άθροισμα πρέπει να είναι πάντα 100%** είτε πρόκειται για διακριτή μεταβλητή, είτε για συνεχή!!

4.2 Ιστόγραμμα συχνοτήτων

Το ιστόγραμμα χρησιμοποιείται κατά κανόνα για τη γραφική απεικόνιση των κατανομών συχνοτήτων όταν έχουμε διαστήματα τάξεων.

Το διάγραμμα αυτό είναι μια σειρά εφαπτόμενων ορθογωνίων παραλληλογράμμων που έχει βάση τον οριζόντιο άξονα των ορθογωνίων συντεταγμένων, το δε ύψος των ορθογωνίων είναι τέτοιο, ώστε **το εμβαδό των ορθογωνίων να είναι ανάλογο με τη συχνότητα την οποία εκπροσωπούν.**

Για την κατασκευή του ιστογράμματος διακρίνουμε δύο περιπτώσεις:

α) όταν τα διαστήματα τάξεων είναι ίσα

Στην περίπτωση αυτή, παίρνουμε το σύστημα των ορθογωνίων συντεταγμένων και τοποθετούμε στον άξονα των X (τετμημένων) τα διαστήματα τάξεων της κατανομής, αφού διαιρέσουμε τον άξονα των X σε ίσα τμήματα, ενώ στον κατακόρυφο άξονα (τεταγμένων) τοποθετούμε τις συχνότητες (f_i).

Μετά, από τα όρια των τάξεων, τα a_{i-1} και a_i , φέρνουμε κατακόρυφως ύψη ίσα προς τη συχνότητά τους (f_i), αφού για ίσα πλάτη των διαστημάτων θα έχουμε και ίσα πλάτη των ορθογωνίων και άρα το εμβαδόν των ορθογωνίων τους θα εξαρτάται μόνο από το ύψος τους. Προκύπτει έτσι ένας αριθμός αλληλοεφαπτόμενων ορθογωνίων, τα οποία μας δίνουν το ιστόγραμμα τάξεων ίσου πλάτους που θέλουμε να κατασκευάσουμε.

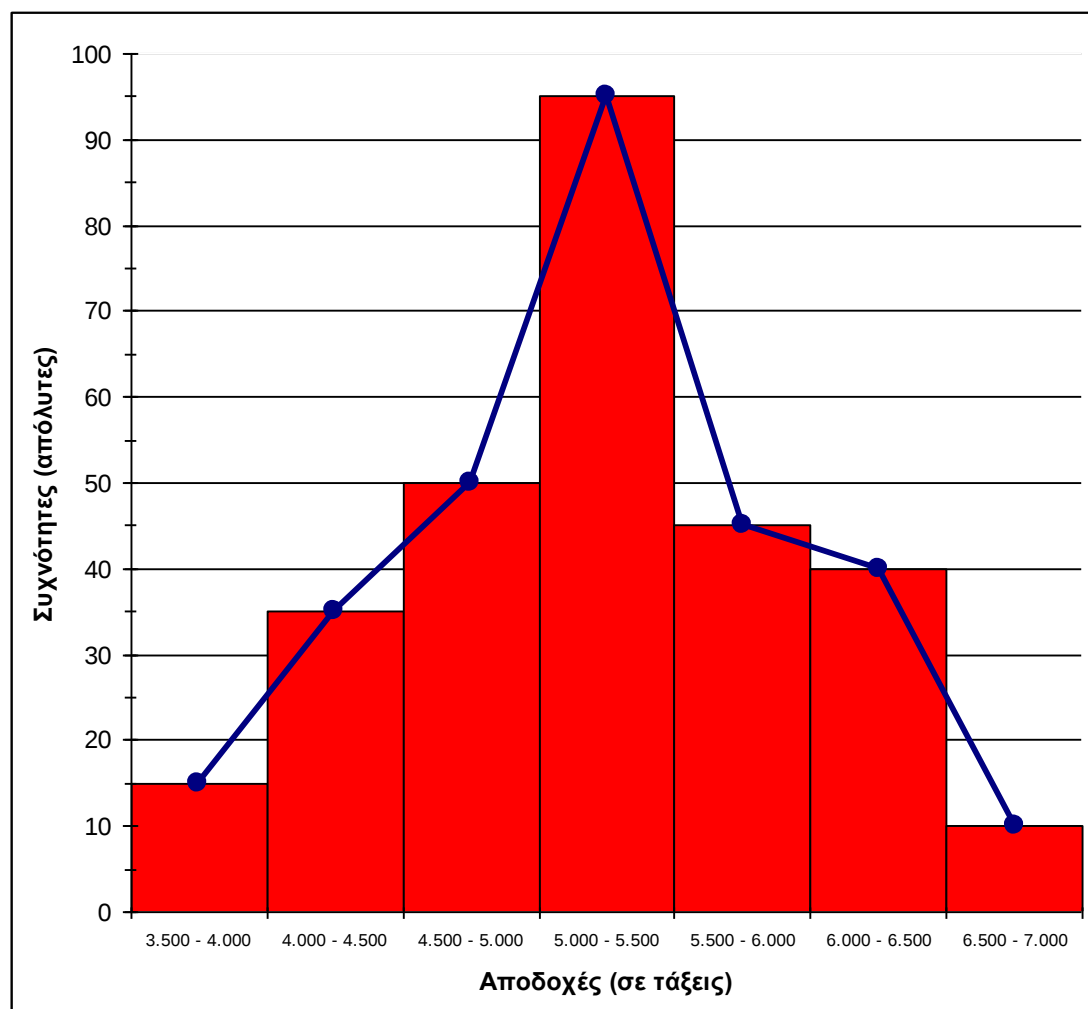
Παράδειγμα: Στον πίνακα που ακολουθεί δίνεται η κατανομή των ημερήσιων αποδοχών 290 εργαζόμενων σε μια επιχείρηση.

ΠΙΝΑΚΑΣ

Τάξεις αποδοχών $[,)$	Συχνότητες (f_i)
3.500 - 4.000	15
4.000 - 4.500	35
4.500 - 5.000	50
5.000 - 5.500	95
5.500 - 6.000	45
6.000 - 6.500	40
6.500 - 7.000	10
Σύνολο	290

Να παρασταθούν τα δεδομένα του πίνακα αυτού σε μορφή ιστογράμματος.

Λύση:



Με την ένωση των μέσων των άνω βάσεων των ορθογωνίων του ιστογράμματος σχηματίζεται μια τεθλασμένη γραμμή που ονομάζεται **πολυγωνική γραμμή ή πολύγωνο συχνοτήτων**.

β) Όταν το πλάτος των τάξεων είναι άνισο

Στην περίπτωση των κατανομών άνισου πλάτους, παίρνουμε στον οριζόντιο άξονα άνισα διαστήματα, ανάλογα με το πλάτος των αντίστοιχων τάξεων, και με βάση τα

άνισα αυτά διαστήματα κατασκευάζουμε ορθογώνια παραλληλόγραμμα, τα ύψη των οποίων είναι ίσα προς το πηλίκο της συχνότητας δια του πλάτους κάθε τάξης ($f_i/\delta_i = f_i/\alpha_i - \alpha_{i-1}$).

Παράδειγμα: Δίνεται η κατανομή του επόμενου πίνακα με άνισα διαστήματα τάξεων.

ΠΙΝΑΚΑΣ

Τάξεις	Συχνότητες (f_i)
0-10	8
10-15	6
15-20	13
20-25	22
25-40	15
40-70	6
Σύνολο	70

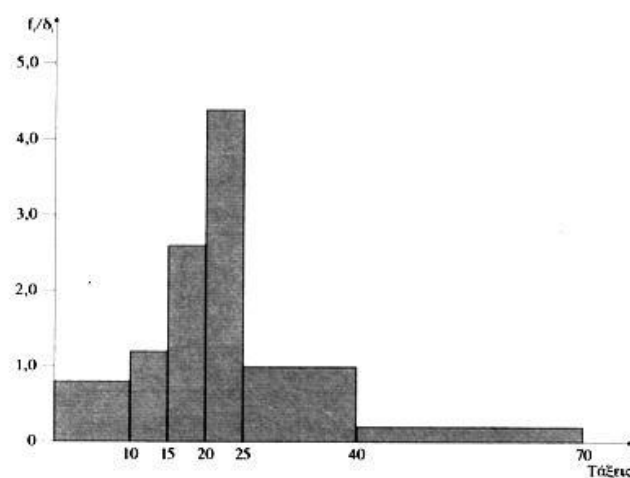
Να σχηματιστεί το ιστόγραμμα συχνοτήτων.

Λύση:

Σε περίπτωση κατά την οποία η κατανομή παρουσιάζει άνισα ταξικά διαστήματα, σχηματίζουμε μια νέα στήλη, η οποία ονομάζεται *πυκνότητα συχνοτήτων* f_i/δ_i και με βάση αυτή κατασκευάζουμε το ιστόγραμμα.

ΠΙΝΑΚΑΣ

Διάστημα τάξεων	Συχνότητες (f_i)	Πυκνότητα συχνοτήτων (f_i/δ_i)
0-10	8	0,8
10-15	6	1,2
15-20	13	2,6
20-25	22	4,4
25-40	15	1,0
40-70	6	0,2
Σύνολο	70	



4.3. Αθροιστικές κατανομές συχνοτήτων

Πολλές φορές χρειάζεται να γνωρίζουμε πόσες ή τι ποσοστό των περιπτώσεων μιας μεταβλητής περιλαμβάνεται μέχρι ενός ορισμένου διαστήματος τάξης, ή ενδιαφερόμαστε να γνωρίσουμε το πλήθος των παρατηρήσεων που οι τιμές τους είναι μικρότερες ή ίσες με ορισμένη τιμή x_i της μεταβλητής X . Απάντηση στο παραπάνω ερώτημα δίνουν οι καλούμενες αθροιστικές συχνότητες F_i . Η αθροιστική συχνότητα ονομάζεται και δεξιόστροφη αθροιστική σειρά.

Αν ο πίνακας συχνοτήτων των τιμών της μεταβλητής X είναι όπως δείχνει ο παρακάτω πίνακας,

ΠΙΝΑΚΑΣ συχνοτήτων

x_i	f_i
x_1	f_1
x_2	f_2
...	...
x_i	f_i
...	...
x_k	f_k

η αθροιστική συχνότητα της τιμής x_i , θα είναι:

$$F_i = f_1 + f_2 + \dots + f_i = \sum_{j=1}^i f_j \quad \text{και} \quad F_k = \sum_{i=1}^k f_i = N$$

Από τον πίνακα αυτόν της κατανομής των συχνοτήτων μπορούμε να κατασκευάσουμε τον πίνακα αθροιστικών συχνοτήτων

ΠΙΝΑΚΑΣ αθροιστικών συχνοτήτων

x_i	f_i	F_i
x_1	f_1	F_1
x_2	f_2	F_2
...	...	...
x_i	f_i	F_i
...	...	...
x_k	f_k	$F_k (= N)$

Για τη γραφική απεικόνιση των αθροιστικών συχνοτήτων χρησιμοποιούνται τα καλούμενα **αθροιστικά διαγράμματα**. Η κατασκευή των διαγραμμάτων αυτών δεν παρουσιάζει καμιά δυσκολία, απαιτείται όμως η διάκριση μεταξύ συνεχών και ασυνεχών κατανομών. Στην περίπτωση ασυνεχούς κατανομής, κάθε τιμή της αθροιστικής F_i απεικονίζεται στην αντίστοιχη τιμή. Στην περίπτωση συνεχούς κατανομής, η κάθε τιμή της αθροιστικής F_i απεικονίζεται στο άνω άκρο του αντίστοιχου διαστήματος.

Παράδειγμα: Δίνεται η κατανομή της ηλικίας 70 ατόμων, όπως φαίνεται στον παρακάτω πίνακα:

Πίνακας

Τάξεις	Αριθμός ατόμων (f_i)
20-25	4
25-30	5
30-35	10
35-40	28
40-45	12
45-50	6
50-55	3
55-60	2
Σύνολο	70

Ζητείται

α) Η αθροιστική σειρά συχνοτήτων και

β) Το αθροιστικό διάγραμμα.

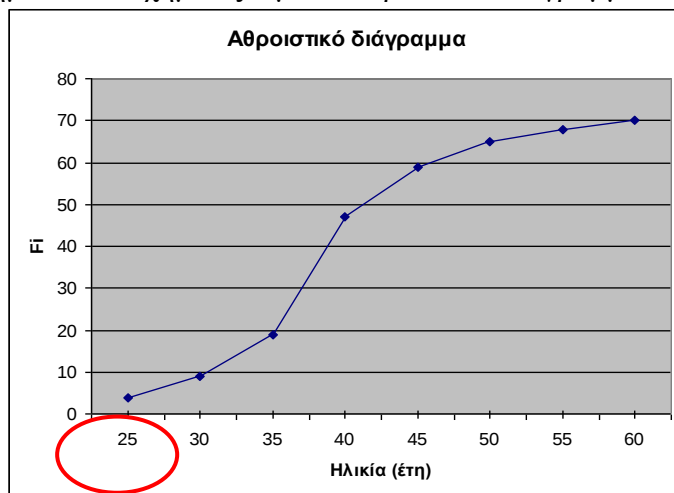
Λύση:

α) Όταν η κατανομή μας εμφανίζεται σε μορφή τάξεων, για να σχηματίσουμε την αθροιστική σειρά εργαζόμαστε όπως φαίνεται στον παρακάτω πίνακα.

Τάξεις (,)	f_i	Αθροιστικές συχνότητες (F_i)
20-25	4	4
25-30	5	9
30-35	10	19
35-40	28	47
40-45	12	59
45-50	6	65
50-55	3	68
55-60	2	70
Σύνολο	70	

β) Το αθροιστικό διάγραμμα έχει ως εξής:

Ενώνουμε τα **δεξιά άκρα** (όχι τα μέσα) των άνω άκρων των ορθογωνίων με ευθύγραμμα τμήματα και σχηματίζουμε το αθροιστικό διάγραμμα.



Συνοψίζοντας:

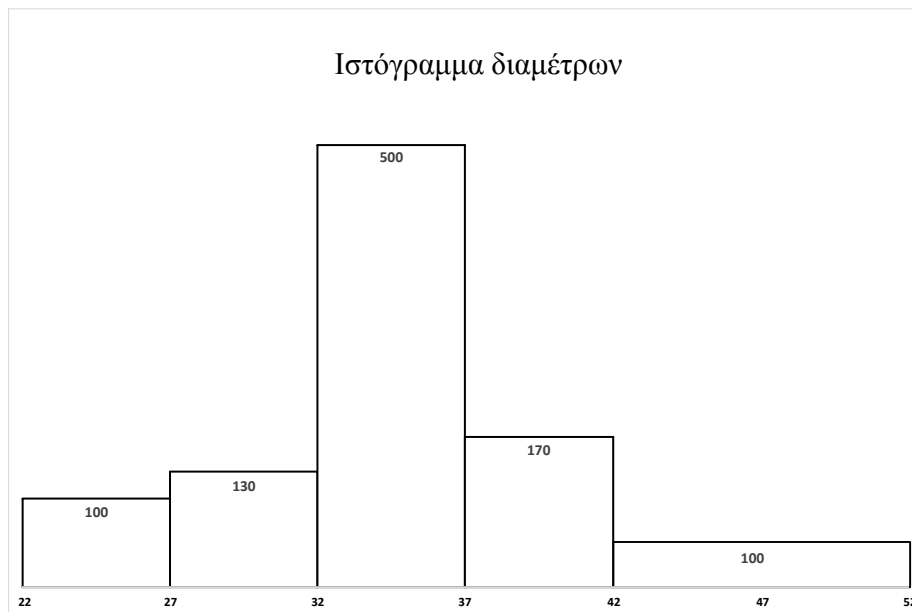
Για την κατασκευή του ιστογράμματος πρέπει όλο το σύνολο των τιμών των αποτελεσμάτων των παρατηρήσεων $x_1, x_2, \dots, x_n$ να χωριστεί σε k διαστήματα τάξεων με τα σημεία $\alpha_0, \alpha_2, \dots, \alpha_k$. Ως συνήθως τα διαστήματα επιλέγονται να είναι ίσα. Μετά προσδιορίζονται οι συχνότητες f_i των τιμών που βρέθηκαν σε κάθε διάστημα $[\alpha_{i-1}, \alpha_i)$ και οι σχετικές συχνότητες $m_i = \frac{f_i}{n}$. Στον άξονα των τετμημένων σημειώνονται τα σημεία $\alpha_0, \alpha_2, \dots, \alpha_k$ και κατασκευάζονται τα ορθογώνια με βάσεις ίσες με τα τμήματα $[\alpha_{i-1}, \alpha_i]$, $i = 1, 2, \dots, k$ και ύψη ίσα με $u_i = \frac{m_i}{\alpha_i - \alpha_{i-1}}$. Σε περίπτωση που τα διαστήματα των τάξεων $[\alpha_{i-1}, \alpha_i)$ είναι ίσα, ως ύψη των ορθογώνιων μπορούν να ληφθούν είτε οι συχνότητες f_i των τιμών, είτε οι σχετικές συχνότητες m_i .

Παράδειγμα:

Οι μετρήσεις των κορμών 1000 ελάτων δίνονται στο παρακάτω πίνακα:

Διάμετρος σε cm	[22, 28)	[27, 32)	[32, 37)	[37, 42)	[42, 52)
πλήθος δέντρων	100	130	500	170	100

Το ιστόγραμμα για τα δεδομένα αυτά είναι το παρακάτω:



(Θέμα παλαιότερων ετών)

- Άσκηση: Δίνεται η παρακάτω κατανομή συχνοτήτων σε πίνακα

ΠΙΝΑΚΑΣ

Τάξεις	Συχνότητες (f_i)
40-60	4
60-80	6
80-100	14
100-120	22
120-140	30
140-160	46
160-180	32
180-200	30

Να κατασκευαστεί

- ιστόγραμμα συχνοτήτων και
- να χαραχτεί η πολυγωνική γραμμή.

• Άσκηση: Δίνονται τα παρακάτω ύψη (σε cm):

170, 180, 178, 165, 170, 168, 175, 175, 173, 162, 160, 170, 167, 177, 180, 170, 182,
178, 165, 178, 156, 175, 172, 173, 167, 187, 170, 180, 178, 191, 176, 169, 167, 166,
179, 178, 180, 164, 170, 173

Να κατασκευαστεί

- το ιστόγραμμα των συχνοτήτων
- το ιστόγραμμα των σχετικών συχνοτήτων και
- να δοθεί το διάγραμμα αθροιστικών συχνοτήτων.