ΚΕΦΑΛΑΙΟ 1ο

Στοιχεία Περιγραφικής Στατιστικής

Εισαγωγή

Μεταβλητές και Παρατηρήσεις

Διαλογή – Κατανομή Συχνοτήτων – Σχετικών Συχνοτήτων

Αθροιστικές συχνότητες

Ομαδοποίηση των Παρατηρήσεων

Διάγραμμα Συχνοτήτων

Σκοπός των περιγραφικών στατιστικών

Τεταρτημόρια

Μέτρα διασποράς

Συμμετρικά και ασύμμετρα δεδομένα

Παραδείγματα με χρήση υπολογιστή

ΚΕΦΑΛΑΙΟ 1ο
ΣΤΟΙΧΕΙΑ ΠΕΡΙΓΡΑΦΙΚΗΣ ΣΤΑΤΙΣΤΙΚΗΣ

1.1 Εισαγωγή

Καθημερινά ακούμε ή διαβάζουμε τα αποτελέσματα διαφόρων δημοσκοπήσεων ή μελετών. Οι μελέτες αυτές γίνονται από εταιρείες που ενδιαφέρονται για τη διάθεση των προϊόντων τους, από πολιτικά κόμματα που ανυπομονούν να μάθουν για την εκλογική τους δύναμη, από έντυπα ποικίλης ύλης ή πιθανώς και από κάποια ομάδα μαθητών που ενδιαφέρεται να καταγράψει και αναλύσει χαρακτηριστικά των συμμαθητών της όπως: ύψος βάρος, ομάδα αίματος, φύλο. **Το κοινό στις μελέτες αυτές είναι ότι όλες έχουν στόχο τη μελέτη των χαρακτηριστικών μιας συγκεκριμένης ομάδας**. Η ομάδα αυτή που μπορεί να αποτελείται από έμβια όντα, αντικείμενα, φαινόμενα ή μετρήσεις σαφώς καθορισμένων χαρακτηριστικών των μονάδων ή ατόμων της αποτελεί ένα στατιστικό **πληθυσμό** .

ΠΛΗΘΥΣΜΟΣ

Το σύνολο των μετρήσεων ή παρατηρήσεων που αναφέρονται σε κάποιο χαρακτηριστικό ή σε κάποια ιδιότητα των μονάδων του συνόλου που εξετάζουμε.

Στο βιβλίο των Μαθηματικών Γενικής Παιδείας της Γ' Λυκείου εξετάζονται διεξοδικά βασικά στοιχεία και έννοιες της Στατιστικής Επιστήμης, η οποία σήμερα βρίσκει εφαρμογή σ' όλους σχεδόν τους τομείς της ανθρώπινης δραστηριότητας.

Η μελέτη των χαρακτηριστικών ενός στατιστικού πληθυσμού θα ήταν επιθυμητή, αλλά είναι πολύ δύσκολη. Για παράδειγμα σαφή γνώση για τα χαρακτηριστικά του Ελληνικού πληθυσμού μπορούμε να έχουμε μόνο με απογραφή που γίνεται κάθε 10 χρόνια. Αντί αυτής επιλέγεται μια μικρή ομάδα ή ένα υποσύνολο του πληθυσμού το οποίο καλείται **δείγμα,** απ' όπου αντλούνται πληροφορίες για τα χαρακτηριστικά του συγκεκριμένου πληθυσμού.

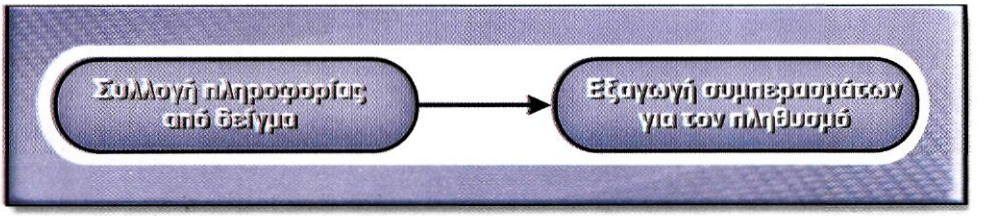
ΔΕΙΓΜΑ

Κάθε τμήμα του πληθυσμού

Ο σωστός τρόπος επιλογής του δείγματος σκοπό έχει να μας οδηγήσει σε αξιόπιστα συμπεράσματα για ολόκληρο τον πληθυσμό από τον οποίο πήραμε το δείγμα .

Η αναγκαιότητα και χρησιμότητα του δείγματος γίνεται φανερή στην πράξη π.χ. εταιρίες ερευνών συλλέγοντας πληροφορίες από μικρό σχετικά αριθμό ψηφοφόρων μπορούν να προβλέψουν με ακρίβεια τα αποτελέσματα των Εθνικών εκλογών. Δηλαδή από τη μελέτη της συμπεριφοράς ενός μικρού τμήματος των ψηφοφόρων (δείγμα) προσδιορίζεται η συμπεριφορά του συνόλου των ψηφοφόρων (πληθυσμός).

Η διαδικασία αυτή μπορεί να αποδοθεί σχηματικά ως εξής:



1.2 Μεταβλητές και παρατηρήσεις

Ο όρος ***μεταβλητή*** αναφέρεται σε χαρακτηριστικό του πληθυσμού που μελετάμε και συμβολίζεται με τα κεφαλαία γράμματα Χ, Ψ, Ζ... και ο όρος ***παρατηρηθείσα τιμή ή παρατήρηση*** χρησιμοποιείται για την αριθμητική ή άλλη συμβολική έκφρασή της.

Από τη μελέτη των ατόμων του πληθυσμού, ως προς κάποιο χαρακτηριστικό τους, προκύπτουν πληροφορίες ή παρατηρήσεις που λέγονται **στατιστικά δεδομένα** και είναι κατάλληλα για επικοινωνία, ερμηνεία και επεξεργασία .

Μερικά παραδείγματα μεταβλητών και παρατηρήσεων παρατίθενται στον πίνακα (1.1) που ακολουθεί :

Πίνακας (1.1)

Μεταβλητή	Παρατηρηθείσα τιμή
Ταχύτητα αυτοκινήτου	26°C
Βάρος ενός ατόμου	Καλή
Θερμοκρασία περιβάλλοντος	Μαύρο
Κατάσταση υγείας	110 km/h
Χρώμα μαλλιών	80 kp

Υπάρχουν δύο τύποι μεταβλητών:

1. Οι ***ποιοτικές***, των οποίων οι τιμές δεν είναι αριθμητικές, αλλά αποτελούν περιγραφές με τη χρήση ονομάτων.

   Οι τιμές τους δεν είναι αριθμητικές ή δεν είναι πλήρως καθοριστικές αλλά μόνο ενδεικτικές ή συγκριτικές π.χ. οι προτιμήσεις των ψηφοφόρων. Οι βλαβερές συνέπειες του καπνού, η κατάσταση της υγείας, η οικογενειακή κατάσταση, η καταγωγή κ.λ.π.

2. **Οι *ποσοτικές***, των οποίων οι τιμές είναι αριθμητικές και επιδέχονται μέτρηση π.χ. ο ετήσιος αριθμός των τροχαίων ατυχημάτων, ο αριθμός των παιδιών σε μια οικογένεια, οι μισθοί των δημοσίων υπαλλήλων κλπ.

   Οι ποσοτικές μεταβλητές διακρίνονται σε ***διακριτές*** και ***συνεχείς***

α) Οι ***διακριτές*** παίρνουν μόνο "μεμονωμένες" αριθμητικές τιμές, είναι δηλαδή στοιχεία ενός συνόλου τα οποία μπορούν να αντιστοιχηθούν ένα προς ένα με στοιχεία του συνόλου των θετικών ακέραιων αριθμών. Τέτοια δεδομένα είναι π.χ. ο αριθμός των παιδιών σε μία οικογένεια, ο αριθμός των δωματίων μιας κατοικίας, το νούμερο γυναικείων παπουτσιών, που μπορεί να πάρει τις τιμές: $35\frac{1}{2}$, 36, $36\frac{1}{2}$, 37, $37\frac{1}{2}$, 38, ...

β) Οι ***συνεχείς*** μπορούν να πάρουν αριθμητικές τιμές που καλύπτουν ολόκληρο διάστημα τιμών των πραγματικών αριθμών (α, β), όπου $-\infty < \alpha < \beta < +\infty$. Π.χ. η ηλικία, η διάρκεια μιας τηλεφωνικής συνδιάλεξης, η θερμοκρασία κλπ.

Εκτός της διάκρισης των μεταβλητών σε ποιοτικές, και ποσοτικές είναι δυνατή και η ταξινόμηση τους αναλόγως της ***κλίμακας μέτρησής*** τους. Σύμφωνα με τη διάκριση αυτή έχουμε τα ακόλουθα είδη μεταβλητών:

Ονομαστικές μεταβλητές επιδέχονται μόνο αυθαίρετη κατάταξη π.χ. φύλο, φυλή, θρησκεία. κλπ.

Διατάξιμες μεταβλητές διαφέρουν από τις ονομαστικές διότι επιδέχονται μέτρηση ανωτέρου επιπέδου που επιτρέπει την ιεράρχησή τους. Όπως π.χ. κατάσταση υγείας, χαρακτηρισμός πτυχίου (άριστα, λίαν καλώς, καλώς), ποιοτική διαβάθμιση προϊόντος.

Οι ονομαστικές και οι διατάξιμες μεταβλητές χαρακτηρίζονται ως ποιοτικές μεταβλητές.

Μεταβλητές διαστήματος οι τιμές των οποίων είναι αριθμητικά μετρήσιμες και κατά συνέπεια διατάξιμες, όπως επίσης μετρήσιμη είναι και η διαφορά μεταξύ των τιμών τους. Παραδείγματα αποτελούν η θερμοκρασία, ο βαθμός πτυχίου, ο αριθμός μονάδων εισαγωγής σε Α.Ε.Ι. – Τ.Ε.Ι.

Αναλογικές μεταβλητές η αριθμητική σειρά και οι διαφορές ανάμεσα στις τιμές τους είναι σαφώς καθορισμένες όπως στις μεταβλητές διαστήματος. Ο λόγος των τιμών τους δίνει επιπλέον σημαντικές συγκριτικές πληροφορίες γύρω από το μετρούμενο μέγεθος. Παραδείγματα αποτελούν το βάρος, το ύψος, η διάρκεια σπουδών, η ταχύτητα, η χιλιομετρική απόσταση.

Οι μεταβλητές διαστήματος και οι αναλογικές μεταβλητές χαρακτηρίζονται ως ποσοτικές μεταβλητές.

Ο πίνακας που ακολουθεί δίνει συνοπτικά και συγκριτικά τα βασικά χαρακτηριστικά των μεταβλητών ως προς την κλίμακα μέτρησής τους.

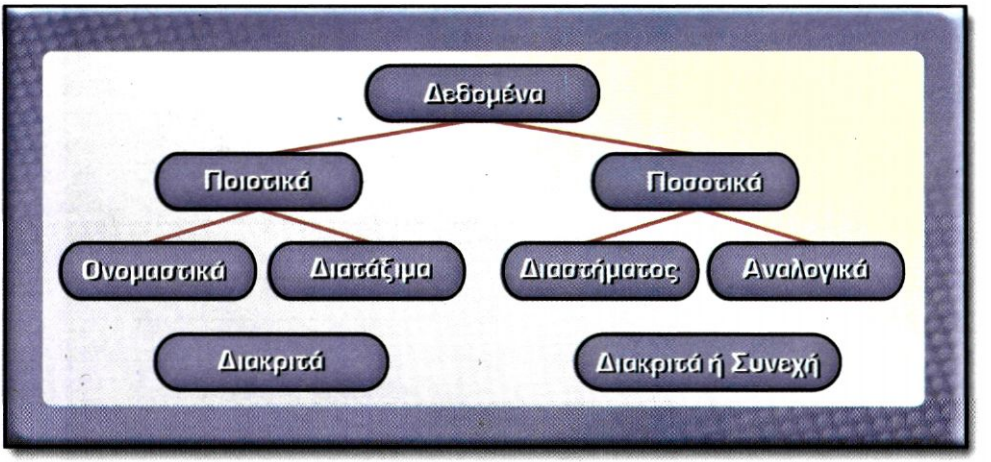
Πίνακας (1.2)

Είδος Δεδομένων	Αρχή Μέτρησης	Νόημα Διάταξης	Ορισμός Απόστασης	Ερμηνεία του λόγου
Ονομαστικά	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ
Διατάξιμα	ΟΧΙ	ΝΑΙ	ΟΧΙ	ΟΧΙ
Διαστήματος	Αυθαίρετη	ΝΑΙ	ΝΑΙ	ΝΑΙ
Αναλογικά	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ

Αναλόγως του είδους της μεταβλητής είναι και τα δεδομένα, που αποτελούν τις μετρήσεις της μεταβλητής σε ένα σύνολο ατόμων.

Μία ταξινόμηση των δεδομένων δίνεται στο διάγραμμα που ακολουθεί.

Διάγραμμα Ταξινόμησης Δεδομένων



ΣΗΜΕΙΩΣΗ :

Ο διαχωρισμός μεταξύ διακριτών και συνεχών δεδομένων δυσχεραίνεται στην πράξη από τους περιορισμούς που επιβάλλονται από τα όργανα μέτρησης. Έτσι π.χ. η μέτρηση του ύψους ενός συνόλου ατόμων με ακρίβεια εκατοστού του μέτρου καταγράφεται με τη χρήση διακριτών τιμών, όπως 174 cm, 186 cm, 172 cm, κ.λ.π., αν και η μεταβλητή "ύψος" είναι συνεχής.

1.3 Διαλογή – Κατανομή Συχνοτήτων –Σχετικών συχνοτήτων

Πήραμε δείγμα 30 πτυχιούχων οδοντιάτρων και ζητήσαμε να μάθουμε το χρόνο που απαιτήθηκε για την απόκτηση πτυχίου.

Οι απαντήσεις έδωσαν τους ακόλουθους χρόνους σε έτη:

7, 8, 6, 8, 6, 6, 6, 6, 6, 6, 6, 6, 8, 7, 8

6, 6, 6,5, 5, 6, 5, 8, 7, 5, 6, 5, 6, 6,5

Ένας απλός τρόπος συνοπτικής παρουσίασης των δεδομένων είναι η **διαλογή** που γίνεται ως εξής:

Σε μία στήλη (την στήλη "Παρατηρήσεις") τοποθετούμε όλες τις δυνατές τιμές των παρατηρήσεων, από μία φορά την κάθε τιμή.

Δίπλα στη στήλη "Παρατηρήσεις" τοποθετούμε τη στήλη "Διαλογή" και στη στήλη αυτή απέναντι από κάθε παρατήρηση σημειώνουμε μία κατακόρυφη γραμμή για κάθε φορά που η παρατήρηση εμφανίζεται στο δείγμα.

Η συμπλήρωση πέντε εμφανίσεων μιας συγκεκριμένης παρατήρησης σημειώνεται με διαγραφή των τεσσάρων κατακόρυφων γραμμών.

Ο πίνακας διαλογής των δεδομένων του χρόνου που απαιτήθηκε από τους 30 πτυχιούχους για την απόκτηση πτυχίου είναι:

Πίνακας (1.3)

Παρατήρηση	Διαλογή
5	卌 I
6	卌 卌 卌 I
7	III
8	卌

Το πλήθος των γραμμών που αντιστοιχούν σε κάθε παρατήρηση δείχνει πόσες φορές αυτή εμφανίζεται στο συγκεκριμένο δείγμα και καλείται ***συχνότητα*** της παρατήρησης, π.χ. η συχνότητα του 5 είναι 6 και του 7 είναι 3.

Αξίζει να σημειωθεί ότι η χρήση πινάκων διαλογής σε περιπτώσεις δεδομένων μεγάλου εύρους (R), όπου $R = \max\{x_j\} - \min\{x_j\}$, δεν ενδείκνυται.

Στην περίπτωση αυτή ο πίνακας διαλογής μπορεί να αντικατασταθεί από τον πίνακα:

Πίνακας (1.4)

Παρατηρήσεις (x_j)	5	6	7	8
Συχνότητα(v_j)	6	16	3	5

ο οποίος ονομάζεται ***πίνακας κατανομής συχνοτήτων***.

Για να σχηματίσουμε ***πίνακα συχνοτήτων της μεταβλητής X,*** καταγράφουμε το φυσικό αριθμό v_j ($v_j \leq v$), που δείχνει πόσες φορές εμφανίζεται η τιμή x_j της εξεταζόμενης μεταβλητής X στο σύνολο των παρατηρήσεων και τον ονομάζουμε **συχνότητα** της τιμής x_j .

Το σύνολο των ζευγών της μορφής (x_j, v_j) αποτελεί την ***κατανομή συχνοτήτων***.

Π.χ. η κατανομή συχνοτήτων του παραδείγματος είναι:

$$\{ (5,6) \quad (6,16) \quad (7,3) \quad (8,5) \}$$

Είναι φανερό ότι το άθροισμα όλων των συχνοτήτων μας δίνει το μέγεθος **ν** του δείγματος.

$$\text{Δηλαδή :.} \nu_1+\nu_2+\nu_3+\ldots+\nu_\kappa=\nu$$

Σε πολλές μελέτες και ειδικότερα όταν το δείγμα είναι πολυπληθές, για να έχουμε μια σαφή συγκριτική εικόνα για το μέρος του συνολικού πλήθους του δείγματος που καταλαμβάνει κάθε τιμή της μεταβλητής που εξετάζουμε, χρησιμοποιούμε τη σχετική συχνότητα (f_j) ή την επί τοις εκατό σχετική συχνότητα (f_j %) .

Η Σχετική συχνότητα (f_j) της τιμής x_j ορίζεται ως ο λόγος της συχνότητας ν_j ως προς το μέγεθος του δείγματος, δηλαδή

$$f_j = \nu_j / \nu.$$

Η «επί τοις εκατό» σχετική συχνότητα (f_j %) της τιμής x_j ορίζεται ως εξής:

$$f_j\,\% = (\nu_j / \nu) \cdot 100$$

Άμεση συνέπεια του ορισμού της σχετικής συχνότητας είναι οι ιδιότητες της:

α) $0 \leq f_j \leq 1$, για $j = 1,2,\ldots\kappa$ και

β) $f_1+f_2+\ldots+f_\kappa = 1$

Σύμφωνα με τα παραπάνω θα συμπληρώσουμε τον πίνακα συχνοτήτων του παραδείγματος που αναφέραμε με τις σχετικές συχνότητες των τιμών της μεταβλητής «διάρκεια σπουδών» ή χρόνος που απαιτήθηκε για την απόκτηση πτυχίου.

Πίνακας (1.5)

Διάρκεια σπουδών (x_i)	Συχνότητα (ν_i)	Σχετική συχνότητα (f_i)	Η «επί τοις εκατό» σχετική συχνότητα (f_i %)
5	6	0,20	20
6	16	0,53	53
7	3	0,10	10
8	5	0,17	17
Σύνολο	**30**	**1,00**	**100**

1.4 Αθροιστικές συχνότητες

Πολλές φορές μας ενδιαφέρει να μάθουμε το πλήθος ή το ποσοστό των παρατηρήσεων, που οι τιμές τους είναι μικρότερες ή ίσες ορισμένης τιμής x_j της ***ποσοτικής*** μεταβλητής Χ. π.χ αν θέλαμε να απαντήσουμε στο ερώτημα «πόσοι οδοντίατροι του προηγούμενου παραδείγματος πήραν πτυχίο *το πολύ* σε 7 χρόνια », θα υπολογίζαμε σύμφωνα με τον πίνακα (1.4) τη λεγόμενη ***αθροιστική συχνότητα*** : $N_3 = \nu_1 + \nu_2 + \nu_3 = 6 + 16 + 3 = 25$. Οι 25, λοιπόν, οδοντίατροι πήραν πτυχίο το πολύ σε 7 χρόνια, ή σε όρους σχετικής αθροιστικής συχνότητας

$F_3\,\% = f_1\,\% + f_2\,\% + f_3\,\% = 20 + 53 + 10 = 83\%$ των οδοντιάτρων.

Γενικεύοντας, ορίζουμε την:

Αθροιστική συχνότητα $N_j = \sum_{j=1}^{k} \nu_j = \nu_1 + \nu_2 + \nu_3 + \ldots + \nu_k \quad , 1 \le j \le k$

και την

Σχετική αθροιστική συχνότητα $F_j = \sum_{j=1}^{k} f_j = f_1 + f_2 + f_3 + \ldots + f_k \quad , 1 \le j \le \kappa$

με τις τιμές $x_1, x_2, x_3, \ldots x_k$ της ποσοτικής μεταβλητής Χ διατεταγμένες σε αύξουσα τάξη.

Η επί τοις εκατό σχετική αθροιστική συχνότητα ισούται με $F_j\% = 100\, F_j$

Παρατηρούμε λοιπόν ότι μπορούμε να αντλήσουμε περισσότερες πληροφορίες για τη διάρκεια σπουδών των οδοντιάτρων, αν συμπληρώσουμε τον πίνακα (1.4).

Πίνακας (1.6)

Αθροιστική συχνότητα (N_i)	Σχετική αθροιστική συχνότητα (F_i)
$N_1 = \nu_1 = 6$	$F_1 = f_1 = 0{,}20$
$N_2 = \nu_1 + \nu_2 = 22$	$F_2 = f_1 + f_2 = 0{,}73$
$N_3 = \nu_1 + \nu_2 + \nu_3 = 25$	$F_3 = f_1 + f_2 + f_3 = 0{,}83$
$N = N_4 = \nu_1 + \nu_2 + \nu_3 + \nu_4 = 30$	$F = F_4 = f_1 + f_2 + f_3 + f_4 = 1{,}00$

1.5 Ομαδοποίηση των Παρατηρήσεων

Στην περίπτωση διακριτής μεταβλητής, όταν το πλήθος των τιμών της είναι μεγάλο, αλλά πολύ περισσότερο σε συνεχή μεταβλητή Χ που μπορεί να πάρει οποιαδήποτε τιμή στο διάστημα ορισμού της, ταξινομούμε τα δεδομένα σε μικρό πλήθος ομάδων που ονομάζονται **κλάσεις ή τάξεις έτσι, ώστε κάθε τιμή να ανήκει μόνο σε μία κλάση. Ομαδοποιούμε** τις παρατηρήσεις χωρίζοντας το διάστημα ορισμού (α_o, α_κ) της μεταβλητής Χ σε κλάσεις, δηλαδή σε υποδιαστήματα της μορφής $[\alpha_{j-1}, \alpha_j)$. Τα άκρα των κλάσεων ονομάζονται **όρια των κλάσεων** και η διαφορά **:**

$c = \alpha_{j-1} - \alpha_j$ = (ανώτερο όριο – κατώτερο όριο) ονομάζεται πλάτος της $j^{ης}$ κλάσης

Στην πράξη συνήθως χρησιμοποιούμε κλάσεις ίσου πλάτους, εκτός βέβαια από τις περιπτώσεις εκείνες που η χρήση κλάσεων άνισου πλάτους κρίνεται απαραίτητη. Διεξοδική αναφορά στις κλάσεις άνισου πλάτους γίνεται στο βιβλίο γενικής παιδείας.

Αν συμβολίσουμε με **R** το **εύρος του συνολικού δείγματος** όπου

R=(τιμή της μεγαλύτερης παρατήρησης – τιμή της μικρότερης παρατήρησης)

υπολογίζουμε το πλάτος c των κλάσεων διαιρώντας το εύρος R δια του αριθμού των κλάσεων, δηλαδή $c = R/\kappa$, όπου κ είναι το πλήθος των κλάσεων

Για να προσδιορίσουμε τον αριθμό των κλάσεων κ χρησιμοποιούμε τον εμπειρικό τύπο, **$\kappa = 1 + 3{,}32 \log \nu$, γνωστό ως κανόνα του** Sturges όπου ν είναι το πλήθος των παρατηρήσεων.

Γενικά, αν πρόκειται για ομαδοποιημένα δεδομένα, ο πίνακας κατανομών συχνοτήτων έχει την παρακάτω μορφή:

Πίνακας (1.7)

Αύξων αριθμό κλάσης	Κλάσεις	Κέντρο Κλάσεων x_i	Συχνότητα ν_i	Αθροιστική Συχνότητα N
1.	α_0-α_1	x_1	ν_1	N_1
2.	α_1-α_2	x_2	ν_2	N_2
3.	α_2-α_3	x_3	ν_3	N_3
...	...	...	...	...
i.	α_{i-1}-α_i	x_i	ν_i	N_i
...	...	...	...	...
k.	α_{k-1}-α_k	x_k	ν_k	N_k
		Σύνολο	N	

Επειδή οι παρατηρήσεις σε κάθε κλάση θεωρείται ότι κατανέμονται ομοιόμορφα, μπορούν να αντιπροσωπευθούν από την **κεντρική τιμή** κάθε κλάσης που υπολογίζεται από τη σχέση:

$$x^*_j = (\alpha_{j-1} + \alpha_j)/2, \text{ για } j = 1, 2, \ldots, k$$

1.6 Διαγράμματα συχνοτήτων

Τις περισσότερες φορές είναι πολύ χρήσιμο να έχουμε ένα είδος «εικόνας» των δεδομένων. Η πληροφορία που παίρνουμε από το σχήμα της κατανομής τους είναι σημαντική για την κατανόηση του χαρακτηριστικού που μελετάμε.

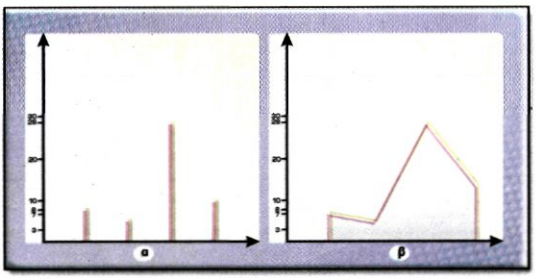
Μια τέτοια εικόνα αποτελεί το ***διάγραμμα συχνοτήτων***, το ***πολύγωνο συχνοτήτων*** (Σχήμα 1.1), ***το κυκλικό διάγραμμα*** (Σχήμα 1.2), ***το ραβδόγραμμα*** (Σχήμα 1.3), κ.λ.π.

Στην περίπτωση πίνακα συχνοτήτων με ομαδοποιημένα δεδομένα, η αντίστοιχη εικόνα είναι το ***ιστόγραμμα*** συχνοτήτων (Σχήμα 1.4). Στο ιστόγραμμα οι ιστοί εφάπτονται, το εμβαδόν δε κάθε ιστού είναι ανάλογο της συχνότητας των δεδομένων με τιμές που βρίσκονται μεταξύ των δύο άκρων του διαστήματος της βάσης του ιστού. Το ιστόγραμμα αποκτά νόημα στην περίπτωση που ο οριζόντιος άξονας αναφέρεται σε χαρακτηριστικό με τιμές αύξουσες, όπως π.χ. ο χρόνος.

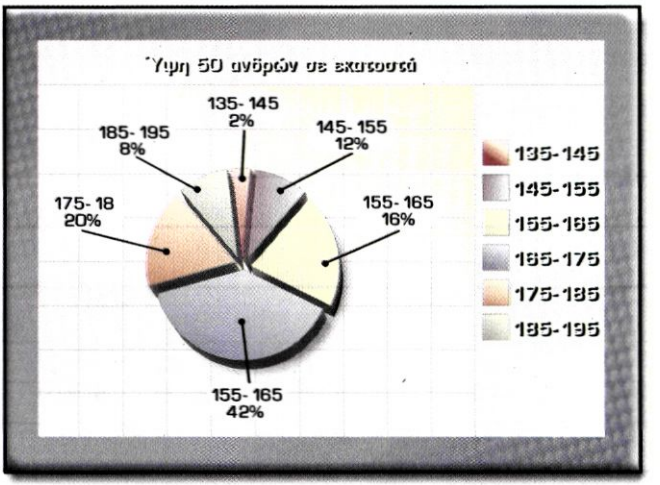
Επεξεργαζόμενοι την εικόνα που παρέχει το ιστόγραμμα μπορούμε ενώνοντας τα κεντρικά σημεία των άνω πλευρών των ιστών του να δημιουργήσουμε το ***συχνοπολύγωνο*** της κατανομής ή το ***πολύγωνο της κατανομής συχνοτήτων***.

Εύκολα αποδεικνύεται ότι τα συνολικά εμβαδά ιστογράμματος και συχνοπολυγώνου είναι ίσα. Η τελική επεξεργασία της εικόνας που παρέχει το συχνοπολύγωνο επιτυγχάνεται με την εξομάλυνση των γωνιών του με συνέπεια τούτο να μπορεί να αντικατασταθεί από καμπύλη, εμβαδού ίσου με αυτό του συχνοπολυγώνου, άρα και του ιστογράμματος.

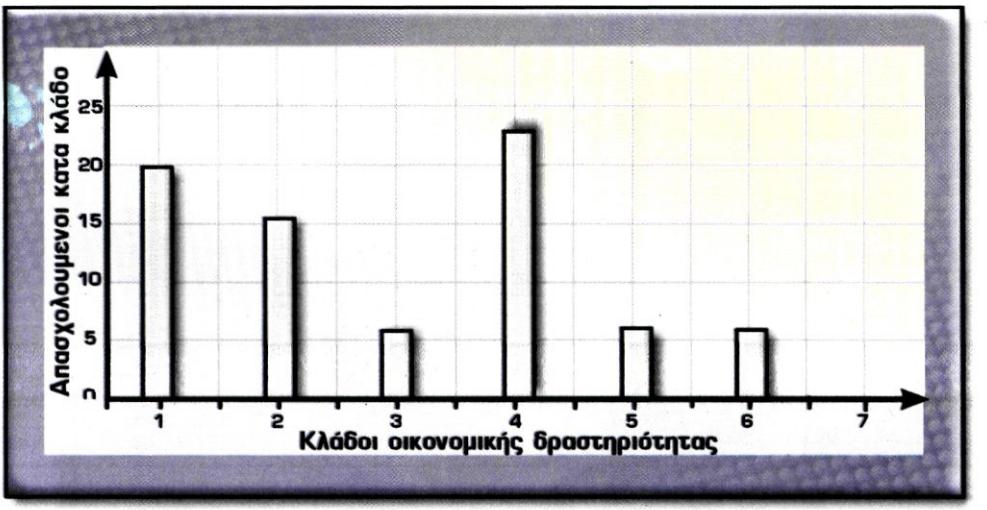
Η χρήση καμπυλών στην στατιστική είναι ευρύτατα διαδεδομένη, διότι πολλές κατανομές συχνοτήτων μπορούν να προσεγγισθούν ή και να απεικονιστούν πλήρως με χρήση καμπυλών που έχουν γνωστή μαθηματική έκφραση.



Σχήμα (1.1) α) Διάγραμμα συχνοτήτων και β) πολύγωνο συχνοτήτων της ομάδας αίματος 46 ασθενών



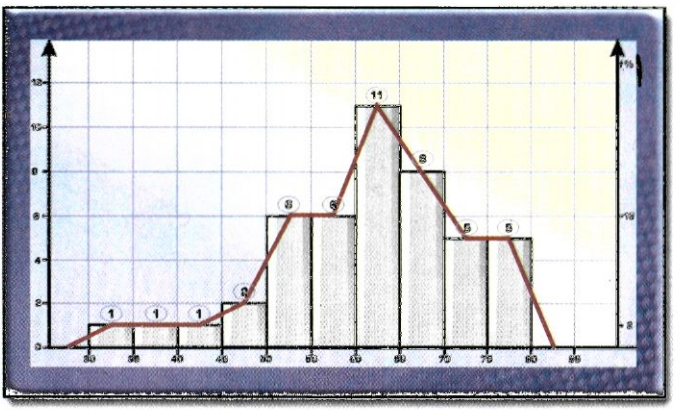
Σχήμα (1.2) Κυκλικό διάγραμμα της «επί τοις εκατό» κατανομής σχετικών συχνοτήτων των «υψών» 50 ανδρών.



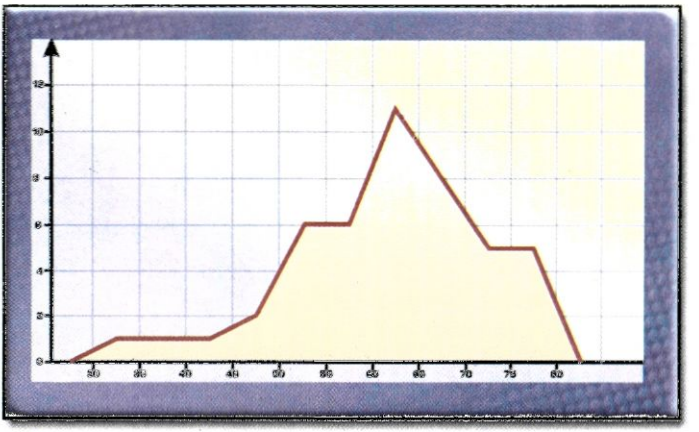
Σχήμα (1.3) Ραβδόγραμμα της κατανομής συχνοτήτων των «απασχολουμένων κατά κλάδο οικονομικής δραστηριότητας»

Τιμές της μεταβλητής: «Κλάδοι οικονομικής δραστηριότητας»

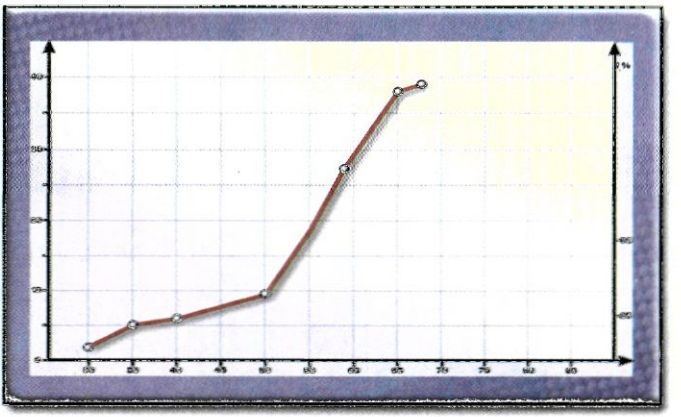
1	Γεωργία, Κτηνοτροφία, Δάση	4	Εμπόριο, Ξενοδοχειακές Επιχ/σεις
2	Ορυχεία, Ηλεκτρισμός	5	Μεταφορές, Επικοινωνίες
3	Οικοδομήσεις, Δημ. Έργα	6	Τράπεζες, Ασφάλειες
7 Λοιπές Υπηρεσίες			



Σχήμα (1.4α) Ιστόγραμμα κατανομής συχνοτήτων ηλικιών 46 ασθενών σε πενταετείς ομάδες.



Σχήμα (1.4β) Πολύγωνο συχνοτήτων κατανομής ηλικιών 46 ασθενών σε πενταετείς ομάδες



Σχ. (1.4γ) Αντίστοιχο διάγραμμα αθροιστικών και σχετικών αθροιστικών συχνοτήτων

Παράδειγμα 1°:

Σε 155 αυτοκίνητα μετρήθηκε ο χρόνος που απαιτήθηκε, ώστε από θέση στάσης (0 km/h) να αναπτύξουν στιγμιαία ταχύτητα 100 km/h.

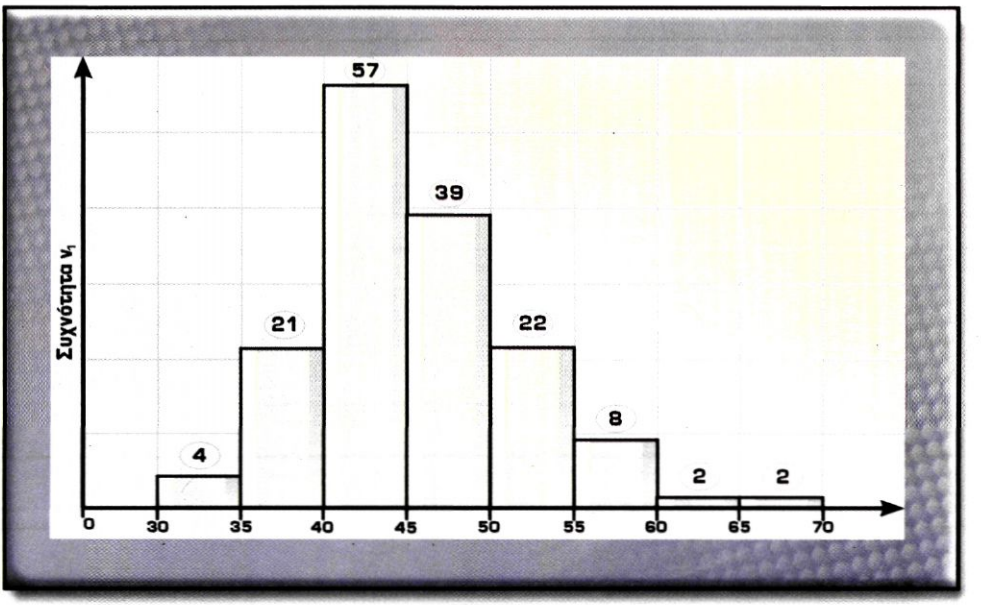
Απάντηση

Οι χρόνοι των 155 αυτοκινήτων ταξινομήθηκαν σε 8 κλάσεις (με βάση τον τύπο του Sturges $k=1+3,32\log_{10}v = 1+3,32\log_{10}155 = 1+7,27 = 8,27$). Η κατανομή συχνοτήτων που προέκυψε δίνεται στον πίνακα (1.8) που ακολουθεί:

Πίνακας (1.8) Κατανομή συχνοτήτων της επιτάχυνσης 155 αυτοκινήτων

Κλάσεις t (sec)	Συχνότητα v_i	Κέντρο Κλάσεων x_i	Σχετική Συχνότητα v_i/v	Αθροιστική Συχνότητα N_i	Αθροιστική Σχετική Συχνότητα $F_i=N_i/v$
10-12	4	11	0,0258	4	0,0258
12-14	21	13	0,1355	25	0,1613
14-16	57	15	0,3677	82	0,5290
16-18	39	17	0,2516	121	0,7806
18-20	22	19	0,1419	143	0,9225
20-22	8	21	0,0516	151	0,9741
22-24	2	23	0,0129	153	0,9870
24-26	2	25	0,0129	155	1,000
Σύνολα	155=v		1,000		

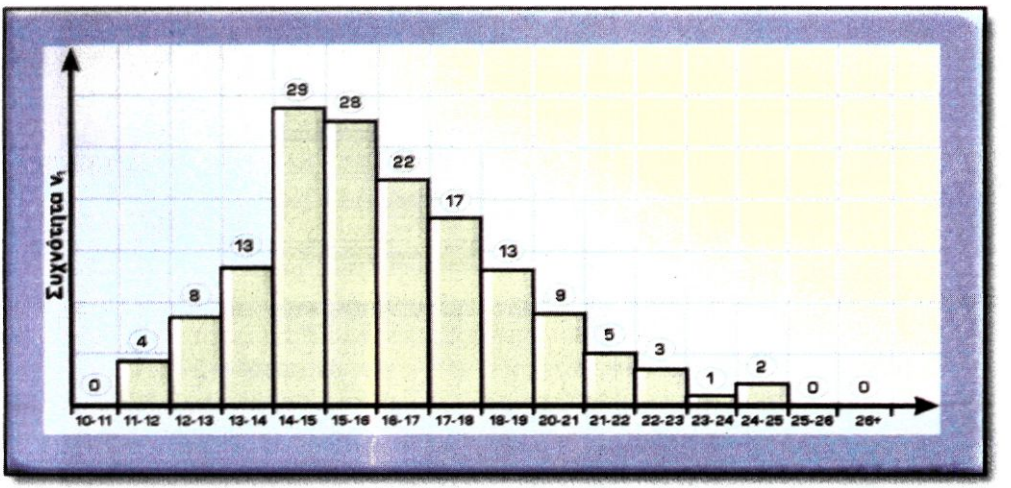
Το αντίστοιχο ιστόγραμμα Σχ. (1.5) των οκτώ κλάσεων δίνει τη μορφή της κατανομής των δεδομένων μας (δηλαδή του χρόνου επιτάχυνσης από 0 σε 100 km/h) στο δείγμα των 155 αυτοκινήτων.



Σχ. (1.5) Ιστόγραμμα της κατανομής συχνοτήτων του χρόνου επίτευξης στιγμιαίας ταχύτητας 100 km/h 155 αυτοκινήτων

Από το ιστόγραμμα αυτό και με τη βοήθεια του πίνακα κατανομής συχνοτήτων μπορούμε, όπως γνωρίζουμε, να υπολογίσουμε το ποσοστό των αυτοκινήτων με χρόνο επιτάχυνσης μικρότερο των 12, 14, 16, 18, 26 sec. Έτσι μπορούμε να πούμε ότι ποσοστό 78,06% των αυτοκινήτων έχουν χρόνο επιτάχυνσης μικρότερο των 18 sec ή ότι ποσοστό 61,93% των αυτοκινήτων έχουν χρόνο επιτάχυνσης μεταξύ 14 και 18 sec (όπως προκύπτει αν από το 0,7806 αφαιρέσουμε το 0,1613 και πολλαπλασιάσουμε το αποτέλεσμα επί 100).

Πρόβλημα εμφανίζεται στην περίπτωση που θέλουμε να υπολογίσουμε ποσοστά αυτοκινήτων με χρόνους επιτάχυνσης, που δεν είναι άκρα των κλάσεων, π.χ. 13, 15 κ.λ.π. Τα προβλήματα αυτά αντιμετωπίζονται με τη δημιουργία κλάσεων όλο και λεπτότερου εύρους δ, που έχουν τα σημεία αυτά ως άκρα τους.Για την περίπτωση αυτή ταξινομούμε τα δεδομένα με τον τρόπο που εμφανίζεται στον πίνακα (1.9). Αυτός ο πίνακας δίνει τη δυνατότητα υπολογισμού του ποσοστού των αυτοκινήτων με χρόνο επιτάχυνσης μικρότερο των 15 sec, πράγμα που στην προηγούμενη κλασικοποίηση δεν ήταν δυνατό, όπως επίσης το ποσοστό των αυτοκινήτων με χρόνο επιτάχυνσης μικρότερο των 17 sec και μεγαλύτερο των 13 sec κ.ο.κ.



Σχ. (1.6) Ιστόγραμμα της κατανομής συχνοτήτων των χρόνων επίτευξης στιγμιαίας ταχύτητας 100 km/h των 155 αυτοκινήτων (16 κλάσεις)

Στον πίνακα (1.9) ο αριθμός των κλάσεων είναι διπλάσιος του αριθμού των κλάσεων του πίνακα (1.8), ενώ το εύρος των κλάσεων του πίνακα (1.9) είναι το μισό του εύρους των κλάσεων του πίνακα (1.8).

Πίνακας (1.9) Κατανομής Συχνοτήτων των επιταχύνσεων των 155 αυτοκινήτων του προηγούμενου παραδείγματος με διπλάσιο αριθμό κλάσεων

Αριθμός Κλάσης i	Κλάσεις	Κέντρα Κλάσεων	Συχνότητα v_i	Σχετική Συχνότητα f%	Αθροιστική Συχνότητα N_i
1	10-11	10,5	0	0	0
2	11-12	11,5	4	0,02581	4
3	12-13	12,5	8	0,05161	12
4	13-14	13,5	13	0,08387	25
5	14-15	14,5	29	0,18710	54
6	15-16	15,5	28	0,18065	82
7	16-17	16,5	22	0,14194	104
8	17-18	17,5	17	0,10968	121
9	18-19	18,5	13	0,08387	134
10	20-21	19,5	9	0,05806	143
11	21-22	20,5	5	0,03226	148
12	22-23	21,5	3	0,01935	151
13	23-24	22,5	1	0,00645	152
14	24-25	23,5	2	0,01290	155
15	25-26	24,5	0	0	155
16	26+	25,5	0	0	155
Σύνολα			**155**	**1,000**	

Χρησιμοποιώντας τον τελευταίο πίνακα μπορούμε να υπολογίσουμε, όπως είπαμε και προηγουμένως, το ποσοστό των αυτοκινήτων με χρόνο επιτάχυνσης μικρότερο ή ίσο των 15 sec. Για το σκοπό αυτό αρκεί να προσθέσουμε τις σχετικές συχνότητες των πέντε πρώτων κλάσεων, οπότε βρίσκουμε ότι:

$$(0 + 0{,}02581 + 0{,}05161 + 0{,}08387 + 0{,}18710) = 0{,}34839$$

που είναι η αναλογία των αυτοκινήτων με χρόνο επιτάχυνσης μικρότερο ή ίσο των 15 sec ή ποσοστό 34,839%.

1.7 Σκοπός των περιγραφικών στατιστικών

Καλούμε στατιστικό περιγραφικό μέτρο τον αριθμό που συνοψίζει βασικά χαρακτηριστικά των παρατηρήσεων του συνόλου των δεδομένων που εξετάζουμε.

Σκοπός των στατιστικών μέτρων είναι η αντικατάσταση μιας μεγάλης μάζας (αριθμητικών) δεδομένων από ένα ή δύο αριθμούς, που από κοινού μεταφέρουν το μεγαλύτερο ποσοστό της βασικής πληροφορίας που περιέχεται στα δεδομένα.

Στην περίπτωση μονομεταβλητών δεδομένων (δεδομένων δηλαδή, που προέρχονται από μετρήσεις ενός μεμονωμένου χαρακτηριστικού π.χ. βάρους) έχουμε δύο τύπους περιληπτικών στατιστικών μέτρων:

Τα στατιστικά μέτρα θέσης ή και μέτρα θέσης
και
τα στατιστικά μέτρα διασποράς ή και μέτρα διασποράς

Τα στατιστικά μέτρα θέσης δίνουν πληροφορία για το μέγεθος των τιμών των δεδομένων, ενώ τα στατιστικά μέτρα διασποράς για το μέγεθος της μεταβολής τους.

Α) Μέτρα θέσης

α) ***Απλά δεδομένα***

Σε ένα τεστ μαθηματικών συμμετείχαν 11 μαθητές.

Ο μαθητής Φ. Γ. πήρε για πρώτη φορά σε τεστ μαθηματικών βαθμό κάτω από τη βάση. Σε συζήτηση που είχε με το συμμαθητή του Α. Κ. παραπονέθηκε για την αυστηρότητα του καθηγητή των μαθηματικών Ν. Σ.

Η βαθμολογία στο τεστ (με άριστα το 100) ήταν:

100, 100, 100, 63, 62, 60, 12, 12, 6, 2, 0

Ο Φ. Γ. είπε παραπονούμενος ότι ο μέσος όρος βαθμολογίας είναι 47 μονάδες, βαθμός σχετικά μικρός. Ο φίλος του Α. Κ. παρατήρησε ότι η μεσαία βαθμολογία είναι 60, ενώ τέλος ο καθηγητής Ν. Σ. ισχυρίστηκε ότι υπάρχουν περισσότερα άριστα (100) από οποιαδήποτε άλλη βαθμολογία.

Κάθε ένα από τα τρία άτομα αναζητούσε ένα τρόπο για να περιγράψει τη γενική τάση της βαθμολογίας. Κάθε τέτοιος αριθμός καλείται ***μέτρο θέσης*** ή ***μέτρο κεντρικής τάσης*** των δεδομένων.

Συγκεκριμένα, ο Φ. Γ. χρησιμοποίησε το ***μέσο*** ή ***μέσο αριθμητικό***, ο οποίος στην περίπτωση **απλών δεδομένων** $x_1, x_2, \ldots, x_\nu$ ορίζεται από την έκφραση:

$$\bar{x} = \frac{1}{\nu}\sum_{j=1}^{\nu} x_j$$

δηλαδή ***από το άθροισμα όλων των τιμών διαιρεμένων δια του συνολικού πλήθους τους.***

Ο Α. Κ. χρησιμοποίησε τη ***διάμεση τιμή ή διάμεσο(δ), που ορίζεται ως η μεσαία παρατήρηση ενός συνόλου διατεταγμένων τιμών κατά αύξουσα τάξη, όταν το ν είναι περιττός αριθμός , ή το ημιάθροισμα των δύο μεσαίων παρατηρήσεων όταν το ν είναι άρτιος***

Στο διατεταγμένο σύνολο των ν = 11 βαθμολογιών του τεστ, η τιμή που κατέχει την $\frac{\nu+1}{2} = \frac{11+1}{2} = 6$η θέση αποτελεί τη ***διάμεσο***.

Τέλος, ο καθηγητής Ν.Σ. χρησιμοποίησε για την περιγραφή της βαθμολογίας την ***επικρατούσα τιμή*** ή ***κορυφή*** **(M_0)** ***που αποτελεί τη συχνότερα εμφανιζόμενη τιμή της μεταβλητής X***, στην προκειμένη περίπτωση το 100.

Κάθε ένας από τους αριθμούς (μέσος, διάμεσος, επικρατούσα τιμή) που χρησιμοποιήθηκαν για την περιγραφή της γενικής τάσης της βαθμολογίας καλείται ***μέτρο θέσης*** ή ***μέτρο κεντρικής τάσης***.

β) Ομαδοποιημένα Δεδομένα

Αν τα δεδομένα είναι ***ομαδοποιημένα*** σε κατανομή συχνοτήτων, τότε ο ***μέσος*** ορίζεται από την έκφραση:

$$\bar{x} = \frac{\sum_{j=1}^{k} \nu_j x_j}{\sum_{j=1}^{k} \nu_j}$$

όπου $x^*_j = \frac{\alpha_{j-1}+\alpha_j}{2}$ η ***κεντρική τιμή*** της j-οστής κλάσης.

Αξίζει να παρατηρήσουμε ότι στον υπολογισμό του μέσου συμμετέχουν όλες ανεξαιρέτως οι παρατηρήσεις.

Η διάμεσος ομαδοποιημένων δεδομένων προσδιορίζεται με τη βοήθεια του ιστογράμματος των αθροιστικών συχνοτήτων. Συγκεκριμένα, αν έχουμε τον πίνακα συχνοτήτων σε γενική μορφή (πίνακας 1.7), τότε η διάμεσος βρίσκεται στη $\frac{\nu}{2}$ θέση.

Με βάση την πληροφορία αυτή μπορούμε να εντοπίσουμε την κλάση στην οποία ανήκει η διάμεσος, αφού η τιμή $\frac{\nu}{2}$ θα βρίσκεται μεταξύ δύο διαδοχικών αθροιστικών συχνοτήτων π.χ. N_{j-1}, N_j. Στην περίπτωση αυτή η κλάση της διαμέσου είναι η $[\alpha_{j-1} - \alpha_j)$.

Υποθέτοντας ότι οι παρατηρήσεις σε κάθε κλάση κατανέμονται ομοιόμορφα, αποδεικνύεται (με απλή μέθοδο των τριών) ότι ο τύπος που δίνει τη διάμεσο σε ομαδοποιημένα δεδομένα είναι :

$$\delta = \alpha_{j-1} + \frac{\frac{\nu}{2} - N_{j-1}}{\nu_j} \cdot c$$

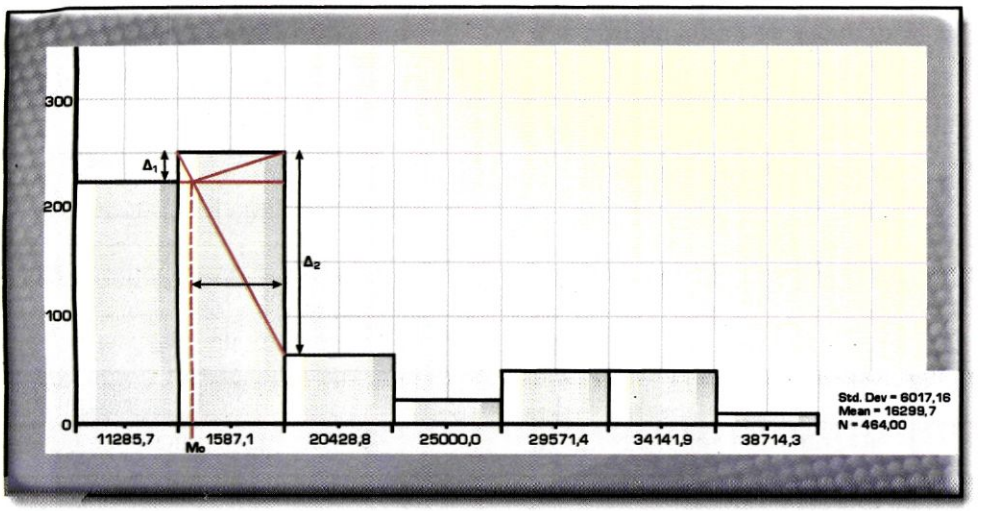
όπου:

α_{j-1} το κατώτερο όριο της κλάσης που περιέχει τη διάμεσο
ν_J η συχνότητα της κλάσης
c το πλάτος της κλάσης (αναφερόμαστε σε κλάσεις ίσου πλάτους)
N_{j-1} η αθροιστική συχνότητα της προηγούμενης κλάσης και
ν το πλήθος των παρατηρήσεων.

Η ***διάμεσος*** είναι προτιμότερη του μέσου σε περιπτώσεις ασύμμετρων κατανομών (Σχ.1.8 και 1.9) ή όταν υπάρχουν παρατηρήσεις σε μεγάλη απόσταση από τον κύριο όγκο των δεδομένων (τέτοιες παρατηρήσεις λέγονται έκτροπες).

Τέλος, η ***επικρατούσα τιμή*** ή ***κορυφή(M_0)*** στην περίπτωση ομαδοποιημένων δεδομένων σε ***κλάσεις ίσου πλάτους*** εντοπίζεται στην κλάση με την μεγαλύτερη συχνότητα, την ***επικρατούσα τάξη.*** Υποθέτοντας και πάλι ότι οι παρατηρήσεις σε κάθε κλάση κατανέμονται ομοιόμορφα υπολογίζουμε την επικρατούσα τιμή μιας ομαδοποιημένης κατανομής ως εξής :

Όπως παρατηρούμε στο σχήμα (1.7), αν α_{j-1} είναι το αριστερό άκρο της επικρατούσας κλάσης, Δ_1 και Δ_2 είναι οι διαφορές των συχνοτήτων των γειτονικών κλάσεων, τότε έχουμε:



Σχ (1.7) Γεωμετρικός τρόπος προσδιορισμού της επικρατούσας τιμής

$$\frac{M_0 - \alpha_{j-1}}{c - (M_0 - \alpha_{j-1})} = \frac{\Delta_1}{\Delta_2}$$ από την οποία προκύπτει ότι:

$$M_0 = \alpha_{j-1} + \frac{c \cdot \Delta_1}{\Delta_1 + \Delta_2}$$

Στην περίπτωση κατανομής συχνοτήτων με δύο ή περισσότερες κλάσεις μεγίστων συχνοτήτων μιλάμε για ***δικόρυφες*** ή ***πολυκόρυφες κατανομές*** αντιστοίχως.

1.8 Τεταρτημόρια

Σε σύνολο διατεταγμένων παρατηρήσεων, ορίζουμε το πρώτο τεταρτημόριο Q_1 ως την τιμή, αριστερά της οποίας βρίσκεται το πολύ το 25% του συνολικού αριθμού των δεδομένων.

Το δεύτερο τεταρτημόριο Q_2 ορίζεται αναλόγως και συμπίπτει με τη διάμεσο δ ων δεδομένων.

λος, το τρίτο τεταρτημόριο Q_3 ορίζεται ως η τιμή, αριστερά της οποίας βρίσκεται το πολύ το 75% του συνολικού αριθμού των δεδομένων.

Σύμφωνα με τους ανωτέρω ορισμούς, αν ν είναι το πλήθος των διατεταγμένων παρατηρήσεων, τότε η θέση που κατέχει το πρώτο τεταρτημόριο Q_1 είναι (ν+1)/4 και η θέση που κατέχει το τρίτο τεταρτημόριο Q_3 είναι 3(ν+1)/4.

Ο υπολογισμός των τεταρτημορίων σε ομαδοποιημένα δεδομένα γίνεται με τη βοήθεια του ιστογράμματος των αθροιστικών συχνοτήτων.

Ένα τρόπο παρουσίασης της θέσης των τεταρτημορίων στο σύνολο των δεδομένων αποτελεί το **θηκόγραμμα,** που δίνεται στα παραδείγματα με χρήση Η/Υ στο τέλος του κεφαλαίου.

1.9 Μέτρα διασποράς

Τα μέτρα διασποράς είναι ενδεικτικά του βαθμού μεταβλητότητας των δεδομένων.

Το εύρος R του συνολικού δείγματος είναι το απλούστερο των μέτρων διασποράς και επηρεάζεται σημαντικά από ακραίες τιμές.

Η ***διακύμανση*** s^2 αποτελεί μέτρο της διασποράς των δεδομένων γύρω από τον αριθμητικό τους μέσο. Για τον υπολογισμό της χρησιμοποιούνται όλες οι τιμές των δεδομένων.

Η έκφραση που χρησιμοποιούμε για τον υπολογισμό της διακύμανσης δίνεται από τη σχέση:

$$s^2 = \frac{1}{\nu}\sum_{j=1}^{\nu}\left(x_j - \bar{x}\right)^2 = \frac{\sum_{j=1}^{\nu} x_j^2}{\nu} - \bar{x}^2$$

όπου $\bar{x} = \frac{1}{\nu}\sum_{j=1}^{\nu} x_j$ ο **δειγματικός μέσος**, που χρησιμοποιείται για την εκτίμηση του αγνώστου **πληθυσμιακού μέσου μ.**

Στην προσπάθεια μας να εκτιμήσουμε τη διασπορά των δεδομένων γύρω από τον μέσο τους, φαίνεται λογική η διαίρεση του αθροίσματος των τετραγωνικών αποκλίσεων, $(x_j - \bar{x})^2$ δια του πλήθους τους ν.

Ο παραπάνω τύπος χρησιμοποιείται όταν:

1. Οι τιμές $x_1, x_2, \ldots, x_\nu$ αποτελούν συνολικό πληθυσμό.
2. Οι τιμές $x_1, x_2, \ldots, x_\nu$ αποτελούν δείγμα από πληθυσμό και ενδιαφερόμαστε για τη διασπορά μέσα στο ίδιο το δείγμα.

Για ***ομαδοποιημένα δεδομένα*** οι αντίστοιχες εκφράσεις για τη διακύμανση είναι:

$$s^2 = \frac{\sum_{j=1}^{k}(x_j - \bar{x})^2}{\sum_{j=1}^{k}\nu_j} = \frac{\sum_{i=1}^{k}\nu_j x_j^2}{\sum_{j=1}^{k}\nu_j} - \bar{x}^2$$

όπου $\sum_{j=1}^{k} v_j = \nu$ το άθροισμα των συχνοτήτων, των k κλάσεων.

Να θυμίσουμε ότι η τετραγωνική ρίζα ($\sqrt{s^2}$) της διακύμανσης είναι η τυπική απόκλιση (s), ο ρόλος της οποίας θα αναδειχθεί στα όσα θα ακολουθήσουν.

Σημείωση : Αν η διακύμανση s^2 του δείγματος $x_1, x_2, \ldots, x_\nu$ χρησιμοποιείται για την εκτίμηση της διακύμανσης σ^2 του πληθυσμού, από τον οποίο προήλθε το δείγμα, τότε χρησιμοποιείται η έκφραση :

$$s^2 = \frac{1}{\nu-1} \cdot \sum_{j=1}^{\nu} (x_j - \bar{x})^2 = \frac{1}{\nu-1}\left[\sum_{j=1}^{\nu} x_j^2 - \nu \bar{x}^2\right]$$

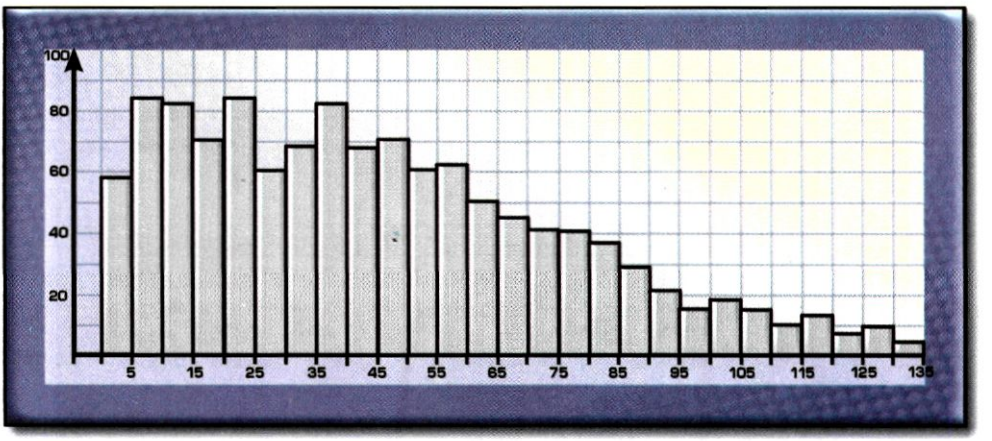
1.10 Συμμετρικά και ασύμμετρα δεδομένα

Αν ένας πληθυσμός είναι κατά προσέγγιση συμμετρικός, τότε σε δείγμα μεγέθους ($\nu \geq 30$) οι τιμές του μέσου και της διαμέσου δε διαφέρουν πολύ μεταξύ τους.

Αν τα δεδομένα έχουν και κορυφή, τότε η τιμή της είναι κοντά στην τιμή του μέσου και της διαμέσου.

Πληθυσμός που δεν είναι συμμετρικός λέγεται ασύμμετρος.

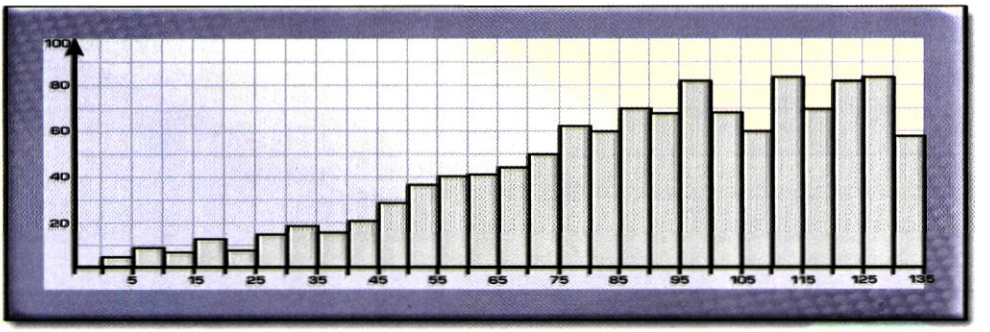
Αν η κατανομή εμφανίζει μακριά ουρά στην περιοχή των μεγάλων τιμών, ονομάζεται θετικά ασύμμετρη. Βλέπε σχήμα (1.8).



Σχ.(1.8) Ιστόγραμμα θετικά ασύμμετρης κατανομής

Στην περίπτωση αυτή, η τιμή της κορυφής είναι μικρότερη της τιμής της διαμέσου, που είναι μικρότερη της τιμής του μέσου.

Πράγματι, για τα δεδομένα του παραδείγματος η κορυφή έχει τιμή 18,670, η διάμεσος 42,967 και ο μέσος 46,951.



Σχ.(1.9) Ιστόγραμμα αρνητικά ασύμμετρης κατανομής

Αν η μακριά ουρά εμφανίζεται στην περιοχή των μικρών τιμών, η κατανομή ονομάζεται αρνητικά ασύμμετρη. Βλέπε σχήμα (1.9).

Μεταξύ των μέτρων που χρησιμοποιούνται για τη μέτρηση της ασυμμετρίας μιας κατανομής θα αναφέρουμε δύο συντελεστές που οφείλονται στον Pearson.

1. **Συντελεστής ασυμμετρίας του Pearson**

(Μέσος – Κορυφή) / Τυπική Απόκλιση

Στην περίπτωση που η κορυφή είναι άγνωστη ή υπάρχουν περισσότερες των δύο κορυφών χρησιμοποιείται εναλλακτική έκφραση για τη μέτρηση του βαθμού ασυμμετρίας της κατανομής στην οποία αντί της κορυφής χρησιμοποιείται η διάμεσος.

2. Εναλλακτική έκφραση του συντελεστή ασυμμετρίας του Pearson

3(Μέσος – Διάμεσος) / Τυπική Απόκλιση

ΕΦΑΡΜΟΓΕΣ

Παράδειγμα 2ο

Ένας αγρότης σημείωσε τον αριθμό των αυγών που συγκέντρωσε σε μια περίοδο 150 ημερών. Η κατανομή συχνοτήτων των αυγών που μάζεψε παρουσιάζεται στον πίνακα που ακολουθεί:

Πίνακας (1.10)

Αριθμός Αυγών	15	16	17	18	19	20	21	22	23	24	25
Συχνότητα	1	2	4	6	9	13	18	22	35	30	10

Να βρεθεί η επικρατούσα τιμή του αριθμού των αυγών που συγκεντρώθηκαν σε μια μέρα.

Να υπολογιστούν:

α) η διάμεσος του αριθμού των αυγών που συγκεντρώθηκαν ανά ημέρα.

β) η μέση τιμή και η απόκλιση του αριθμού των αυγών που συγκεντρώθηκαν ανά ημέρα.

γ) Να σχεδιαστεί διάγραμμα συχνοτήτων που να απεικονίζει τα δεδομένα και να σχολιαστεί η μορφή του.

Απάντηση

Η επικρατούσα τιμή είναι 23. Παρατηρούμε ότι η παρατήρηση αυτή εμφανίζει τη μεγαλύτερη συχνότητα.

α) ν=150 άρα η διάμεσος αντιστοιχεί στην (150 + 1) /2 = 75.5^{η} τιμή.

Ο πίνακας αθροιστικών συχνοτήτων είναι ο ακόλουθος:

Πίνακας (1.11)

Αριθμός αυγών x	Συχνότητα (ν)	Αθροιστική συχνότητα (N)
15	1	1
16	2	3
17	4	7
18	6	13
19	9	22
20	13	35
21	18	53
22	22	75
23	35	110
24	30	140
25	10	150

Παρατηρούμε ότι η 75η τιμή είναι 22 και η 76η τιμή είναι 23, επομένως η 75,5η τιμή είναι 22,5. Άρα η διάμεσος είναι 22,5 αυγά.

β) Για να υπολογίσουμε το μέσο και την τυπική απόκλιση, χρησιμοποιούμε τους τύπους:

i) $$\bar{x} = \frac{\sum \nu_j x_j}{\sum \nu_j} \frac{3.291}{150} = 21,94$$

ii) $$s^2 = \frac{\sum \nu_j x_j^2}{\sum \nu_j} - \bar{x}^2 = \frac{72.917}{150} - (21,94)^2 = 4,749$$

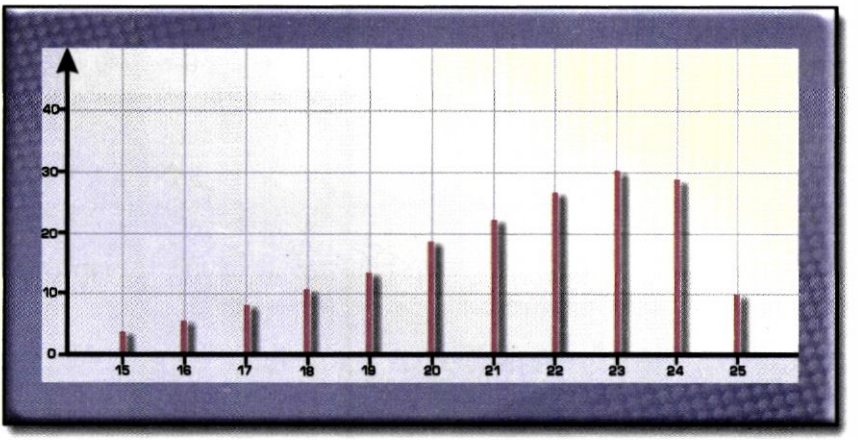
και $$s = \sqrt{4,749} = 2,179$$

Άρα, η μέση τιμή του αριθμού των αυγών είναι 21,94 και η τυπική απόκλιση είναι 2,179.

γ) Από το διάγραμμα συχνοτήτων παρατηρούμε ότι η κατανομή εμφανίζει αρνητική ασυμμετρία. Ο συντελεστής ασυμμετρίας του Pearson υπολογίζεται ως εξής :

$$\frac{\text{μέσος} - \text{κορυφή}}{\text{τυπικήαπόκλιση}} = \frac{21,94 - 23}{2,179} = -0,486$$

Το αποτέλεσμα αυτό δείχνει ότι πράγματι, η κατανομή είναι αρνητικά ασύμμετρη, όπως προκύπτει και από το ραβδόγραμμα της κατανομής συχνοτήτων Σχ. (1.10)



Σχ(1.10) Ραβδόγραμμα της κατανομής συχνοτήτων του παραδείγματος 2.

Παράδειγμα 3°

Ηλεκτρικές ασφάλειες των 30Α ελέγχθηκαν και σημειώθηκε η τιμή της έντασης του ρεύματος για την οποία καταστρέφονται. Τα αποτελέσματα αυτού του ελέγχου σε δείγμα 125 ασφαλειών παρουσιάζονται στον παρακάτω πίνακα :

Πίνακας (1.12)

Ένταση ηλεκτρικού ρεύματος (x σε Α)	Αριθμός ασφαλειών
[27,28)	6
[28,29)	12
[29,30)	27
[30,31)	30
[31,32)	18
[32,33)	14
[33,34)	9
[34,35)	4
[35,36)	5

Να κατασκευαστεί ιστόγραμμα συχνοτήτων που να περιγράφει τα δεδομένα.

Για το συγκεκριμένο δείγμα να υπολογιστούν:

α) η διάμεσος της έντασης του ρεύματος

β) η μέση τιμή της έντασης του ρεύματος

γ) η τυπική απόκλιση.

Να υπολογιστεί η τιμή του συντελεστή ασυμμετρίας των δεδομένων χρησιμοποιώντας τον εναλλακτικό τύπο του Pearson. Να εξηγηθεί συνοπτικά με ποιο τρόπο η ασυμμετρία της κατανομής γίνεται εμφανής από τη μορφή του ιστογράμματος.

Απάντηση

Σύμφωνα με τον πίνακα (1.12), κατασκευάζουμε τον πίνακα (1.13) κατανομής συχνοτήτων και αθροιστικών συχνοτήτων:

Πίνακας (1.13)

Ένταση ρεύματος (x σε A)	Κέντρο κλάσης (x_i)	Συχνότητα (v_i)	Αθροιστική συχνότητα (N_i)
[27,28)	27,5	6	6
[28,29)	28,5	12	18
[29,30)	29,5	27	45
[30,31)	30,5	30	75
[31,32)	31,5	18	93
[32,33)	32,5	14	107
[33,34)	33,5	9	116
[34,35)	34,5	4	120
[35,36)	35,5	5	125

α) Ως γνωστόν, για ομαδοποιημένα δεδομένα, η διάμεσος (δ) αντιστοιχεί στην τιμή x=δ της μεταβλητής, έτσι ώστε το πολύ το 50% των παρατηρήσεων να είναι μικρότερες ή ίσες του δ. Δηλαδή σε δείγμα μεγέθους ν=125, η διάμεσος θα βρίσκεται στην $125 \cdot \frac{1}{2} = 62{,}5$ θέση.

Από την αθροιστική συχνότητα παρατηρούμε ότι 45 ασφάλειες καίγονται, όταν η ένταση του ρεύματος είναι μικρότερη από 30A και 75 ασφάλειες σε ένταση ρεύματος μικρότερη από 31A. Χρησιμοποιούμε τον τύπο (3) υπολογισμού της διαμέσου και έχουμε ότι:

$$\delta = 30 + \frac{62{,}5 - 45}{30} = 30{,}58$$

Δηλαδή, η διάμεσος της έντασης του ρεύματος είναι περίπου ίση με 30,6Α

β) Η μέση τιμή της έντασης του ρεύματος υπολογίζεται από τον τύπο : $\bar{x} = \frac{\sum \nu_j x_j}{\sum \nu_j}$ δηλαδή $\bar{x} = \frac{3857,5}{125} = 30,86$ και η διακύμανση από τον τύπο:

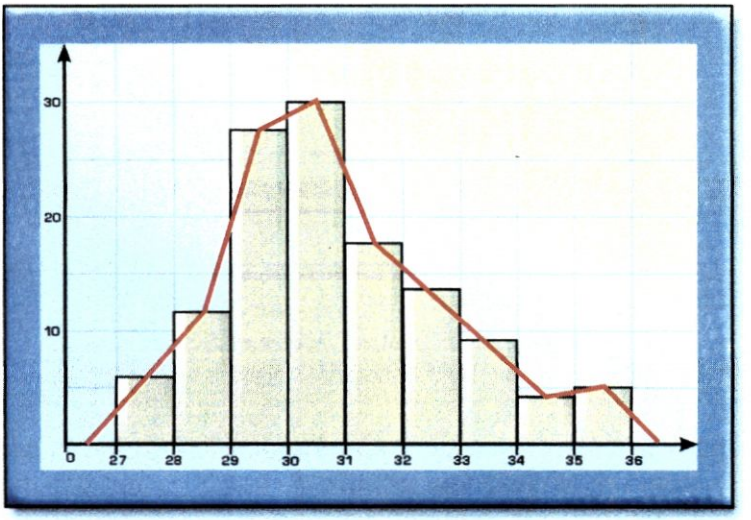
γ) $s^2 = \frac{\sum \nu_j x_j^2}{\sum \nu_j} - \bar{x}^2 = \frac{119499,25}{125} - 30,86^2 = 955,99 - 952,34 = 3,65,$

άρα $s \approx 1,91$.

Ο εναλλακτικός τύπος του συντελεστή ασυμμετρίας δίνεται από τον τύπο :

$$\frac{3(\text{μέσος - διάμεσο})}{\text{τυπική απόκλιση}} = \frac{3(30,86 - 30,58)}{1,91} = 0,438$$

Επειδή ο συντελεστής ασυμμετρίας είναι θετικός, η κατανομή εμφανίζει θετική ασυμμετρία όπως παρατηρούμε και από τη μορφή του ιστογράμματος, του Σχ. (1.11)



Σχ. (1.11) Ιστόγραμμα και συχνοπολύγωνο της μεταβλητής "ένταση ρεύματος"

ΠΑΡΑΔΕΙΓΜΑΤΑ ΜΕ ΧΡΗΣΗ ΥΠΟΛΟΓΙΣΤΗ

Φυλλογράφημα

1) Από τα αρχεία Πανεπιστημίου επιλέξαμε δείγμα 30 πτυχιούχων του έτους 1998 των οποίων ο βαθμός πτυχίου κυμαίνεται από [5.9-7,9). Με ακρίβεια προσέγγισης δεκάτου οι βαθμοί ήταν οι εξής:

6,6	6,0	6,0	6,1	6,1	6,1	6,1	6,1	5,9	6,5
6,1	6,2	6,2	6,3	6,4	6,4	6,4	6,6	6,6	6,8
6,8	6,8	6,9	6,8	7,0	7,9	7,5	7,6	7,6	6,6

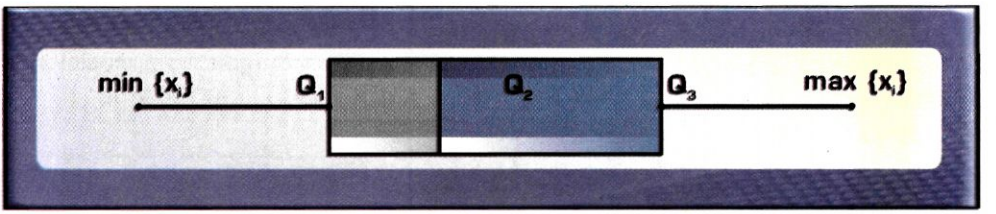
Προτείνεται λοιπόν ένας άλλος τρόπος ομαδοποίησης των δεδομένων που διατηρεί τις ακριβείς τιμές της μεταβλητής «βαθμός πτυχίου», γνωστός ως φυλλογράφημα (stem and leaf) σύμφωνα με τον οποίο ο μίσχος (stem) παριστάνει το σημαντικότερο ψηφίο του αριθμού (π.χ. μονάδες) και τα φύλλα (leaf) παριστάνουν τα λιγότερο σημαντικά ψηφία (π.χ. δέκατα)

Στην προκειμένη περίπτωση η μεταβλητή «βαθμός πτυχίου» έχει εύρος R= 7,9-5,9 =2 και το φυλλογράφημα που ακολουθεί αντιστοιχεί στην κατανομή συχνοτήτων των βαθμών πτυχίου, 30 πτυχιούχων.

Συχνότητα	Κορμός	Φύλλα
1	5	9
14	6	00111111223444
10	6	5666688889
1	7	0
4	7	5669

Θηκόγραμμα

Το θηκόγραμμα αποτελεί γραφικό τρόπο παρουσίασης πέντε περιληπτικών μέτρων μιας κατανομής ομαδοποιημένων δεδομένων, με συνδυασμό των οποίων είναι δυνατή η άντληση περισσότερων πληροφοριών από αυτήν που περιέχεται στα πέντε αυτά μέτρα.



Σχ. (1.12) Τυπική μορφή θηκογράμματος.

Όπου min{x_j}, max{x_j} η ελάχιστη και η μέγιστη τιμή συνόλου ομαδοποιημένων δεδομένων και Q_1, Q_2, Q_3 το πρώτο, δεύτερο και τρίτο τεταρτημόριο αντιστοίχως.

Από το θηκόγραμμα προκύπτουν τα ακόλουθα :

1. Μεταξύ min{x_j} και πρώτου τεταρτημορίου Q_1, περιέχεται το πρώτο τέταρτο (25%) του συνολικού αριθμού των δεδομένων.
2. Μεταξύ Q_3 και max{x_j} περιέχεται το τελευταίο τέταρτο (25%) του συνολικού αριθμού των δεδομένων.
3. Μεταξύ Q_1 και Q_3 περιέχεται το 50% των δεδομένων, που στην στατιστική ορολογία καλείται ενδοτεταρτημοριακό εύρος.
4. Στο ορθογώνιο που ορίζεται από το ενδοτεταρτημοριακό εύρος σημειώνεται έντονα η θέση του δεύτερου τεταρτημορίου που ταυτίζεται με τη θέση της διαμέσου.
5. Τα θηκογράμματα αποτελούν βοηθητικά μέσα σύγκρισης κατανομών συχνοτήτων.

Παραδείγματα

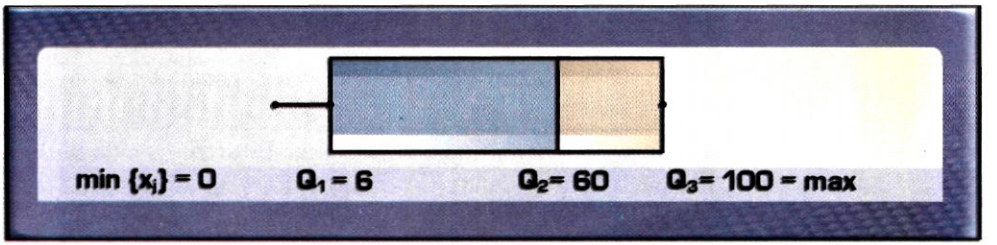
1) Στην παράγραφο 1.7 έχουμε δείγμα των βαθμών που πέτυχαν σε τεστ των Μαθηματικών 11 μαθητές ενός Λυκείου.

0, 2, 6, 12, 12, 60, 62, 63, 100, 100, 100

Η κατασκευή του θηκογράμματος απαιτεί τον υπολογισμό των:

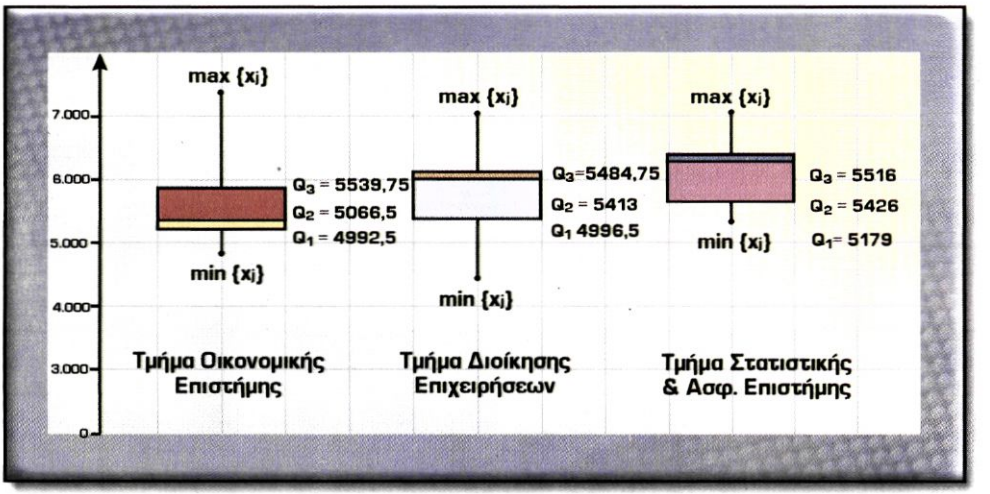
min {x_j} = 0, Q_1 = 6 Q_2 = 60, Q_3 = 100, max {x_j} = 100

μετά την εύρεση των οποίων είναι δυνατός ο σχεδιασμός του.



Σχ. (1.13) Θηκόγραμμα για την περιγραφή δείγματος ν = 11 βαθμών στα Μαθηματικά

2) Σύγκριση με τη βοήθεια θηκογραμμάτων των κατανομών βαθμολογιών εισαγωγής στο Πανεπιστήμιο στα τμήματα Διοίκησης Επιχειρήσεων, Οικονομικής Επιστήμης και Στατιστικής και Ασφαλιστικής Επιστήμης αντι στοίχως.



Σχήμα (1.14) Σύγκριση βαθμολογιών εισαγωγής σε 3 πανεπιστημιακά τμήματα

Τα θηκογράμματα δίνουν ένα εύκολο τρόπο για τη σύγκριση κατανομών, όπως φαίνεται από τα σχήματα. (1.13) και (1.14).

Από το Σχ. (1.13) είναι φανερό ότι το δείγμα εμφανίζει αρνητική ασυμμετρία γεγονός που προκύπτει από τη θέση που κατέχει το Q_2 μεταξύ των Q_1 και Q_3, η τιμή του οποίου ταυτίζεται με το $\max\{x_j\}$.

Το **ενδοτεταρτημοριακό εύρος** $I = Q_3 - Q_1 = 100 - 6 = 94$ είναι πολύ μεγάλο, ίσο σχεδόν με το εύρος $R = \max\{x_j\} - \min\{x_j\} = 100 - 0 = 100$ (Τι μπορεί να σημαίνει αυτό;)

Από το Σχ. (1.14) διαπιστώνουμε ότι:

α) Ο μικρότερος βαθμός εισαγωγής σημειώθηκε από επιτυχόντα στο Τμήμα Διοίκησης Επιχειρήσεων.

β) Ο μεγαλύτερος βαθμός εισαγωγής σημειώθηκε από επιτυχόντα στο Τμήμα Οικονομικής Επιστήμης.

γ) Το 50% των εισαχθέντων στο Τμήμα Οικονομικής Επιστήμης είχε βαθμολογία ≤ 5066,5 μορίων. Στα Τμήματα Διοίκησης Επιχειρήσεων και Στατιστικής και Ασφαλιστικής Επιστήμης το 50% των εισαχθέντων είχαν βαθμολογία (≤ 5413) και (≤ 5426) μορίων αντιστοίχως.

δ) Τη μεγαλύτερη βάση εισαγωγής είχε το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης.

ε) Η κατανομή των μορίων εισαγωγής είναι ασύμμετρη και στις τρεις περιπτώσεις.

Στο μεν Τμήμα Οικονομικής Επιστήμης η ασυμμετρία είναι θετική στα δε άλλα τμήματα αρνητική.

Από την τελευταία αυτή παρατήρηση μπορούμε να συμπεράνουμε ότι ο απαιτούμενος αριθμός μορίων εισαγωγής στο Οικονομικό Τμήμα είναι μικρότερος από ότι στα άλλα δύο τμήματα (γιατί ;)

ΕΡΩΤΗΣΕΙΣ - ΑΣΚΗΣΕΙΣ

ΕΡΩΤΗΣΕΙΣ ΚΑΤΑΝΟΗΣΗΣ

1. Συνδέστε με μια γραμμή κάθε μεταβλητή της στήλης (Α) με τον αντίστοιχο χαρακτηρισμό της στήλης (Β).

Στήλη Α	Στήλη Β
1)Αριθμός τροχαίων ατυχημάτων (σε ένα Σαββατοκύριακο)	α) Κατηγορική μεταβλητή
2) Το μηνιαίο εισόδημα της οικογένειας	β) Ποιοτική μεταβλητή
3) Το επάγγελμα	γ) Συνεχής μεταβλητή
4) Υγεία	δ) Διακριτή μεταβλητή

2. Αν ο μέσος των αριθμών 12, 18, 21, x, 13 είναι 17, τότε η τιμή του x είναι :

Α. 24 ❑ Β. 21 ❑ Γ. 17 ❑ Δ. 15 ❑

3. Σ' ένα εργοστάσιο κατασκευής τροφίμων πρόκειται να αγοραστεί μία καινούρια μηχανή συσκευασίας (Β) για να αντικαταστήσει την μηχανή (Α), η οποία θεωρείται πλέον παλαιού τύπου. Έγινε μία δοκιμή για να ελεγχθεί η αποτελεσματικότητα της νέας μηχανής.
Επελέγησαν 10 πακέτα τροφίμων από κάθε μηχανή και μετρήθηκαν τα αντίστοιχα βάρη

Μηχανή Α (βάρος σε g)	196	198	198	199	200	200	201	201	202	205
Μηχανή Β (βάρος σε g)	192	194	195	198	200	201	203	204	206	207

Χαρακτηρίστε τις παρακάτω προτάσεις

α) $\bar{x}_A = \bar{x}_B = 200$ Σ ❑ Λ ❑

$\bar{x}_A < \bar{x}_B$ Σ ❑ Λ ❑

β) Η τυπική απόκλιση για τη μηχανή είναι :

A : 5,6 ❑ 2,37 ❑ 4,89 ❑ καμία από τις προηγούμενες τιμές ❑

B : 4,9 ❑ 24 ❑ 2,37 ❑ καμία από τις προηγούμενες τιμές ❑

Αν είσαστε εσείς ο ιδιοκτήτης του εργοστασίου θα αντικαθιστούσατε τη μηχανή Α με τη μηχανή Β ; Δικαιολογήστε την απάντησή σας.

ΑΣΚΗΣΕΙΣ

ΟΜΑΔΑ Α

1. Υποθέτουμε ότι ο δήμαρχος της πόλης όπου ζείτε θέλει να μάθει τη γνώμη των δημοτών για την οικοδόμηση ενός νέου γυμναστηρίου. Αν ρωτήσει 100 άτομα που μένουν στη γειτονιά σας, θα μπορούσε να αποκτήσει μία καλή ιδέα για τη γνώμη των δημοτών; Εξηγείστε την απάντησή σας χρησιμοποιώντας τους όρους «πληθυσμός» και «δείγμα».

2. Ο μέσος ν αριθμών ισούται με 5. Αν προσθέσουμε τον αριθμό 13 στους ν αριθμούς, ο νέος μέσος ισούται με 6. Να βρεθεί το πλήθος των αριθμών.

3. Τα βάρη των 30 μελών μιας ομάδας αθλητών είναι οι ακόλουθες :
   74, 52, 67, 68, 71, 76, 86, 81, 73, 68, 64, 75, 71, 57, 67, 57, 59, 72, 79,
   64, 70, 74, 77, 79, 65, 68, 76, 83, 61, 63
   α) Να κατασκευαστεί ο πίνακας κατανομής συχνοτήτων των βαρών. Να χρησιμοποιηθεί εύρος διαστήματος 5kp.
   β) Να κατασκευαστεί το ιστόγραμμα και το αντίστοιχο πολύγωνο συχνοτήτων.
   γ) Να υπολογιστεί ο μέσος και η διάμεσος των βαρών.
   δ) Να κατασκευαστεί πίνακας αθροιστικών συχνοτήτων και να σχεδιαστεί το αντίστοιχο διάγραμμα.
   ε) Πόσοι αθλητές έχουν βάρος λιγότερο από 77 kp;

4. Οι αριθμοί α, β, 8, 5, 7 έχουν μέσο 6 και διακύμανση 2. Να βρεθούν οι τιμές των α και β, αν το α > β.

5. Από τις πληροφορίες που δίνονται για κάθε μία από τις ακόλουθες κατανομές συχνοτήτων, να συμπληρώσετε τα στοιχεία του πίνακα που λείπουν.

	$\Sigma v_i = v$	$\Sigma v_i x_i$	$\Sigma v_i x_i^2$	$\Sigma v_i (x_i - \bar{x})^2$	$\bar{x}$	s
Α	20	563	16143			
Β		270		160	27	
Γ	50				10	3
Δ	30		1025	182,3		
Ε		240	5100		20	

ΟΜΑΔΑ Β

1. Ένα δοχείο περιέχει 5 σφαιρίδια που το καθένα φέρει τους αριθμούς 1, 2, 3, 4, 5 αντίστοιχα. Κάθε φορά που επιλέγεται ένα σφαιρίδιο από το δοχείο, ο αριθμός του σημειώνεται και το σφαιρίδιο επανατοποθετείται στο δοχείο. Το πείραμα επαναλαμβάνεται 50 φορές και τα αποτελέσματα καταγράφηκαν στον παρακάτω πίνακα:

Αριθμός	1	2	3	4	5
Συχνότητα	x	11	y	8	9

Αν ο μέσος είναι 2,7 προσδιορίστε τις τιμές του x και του y.

2. Για κάθε μία από τις παρακάτω κατανομές, ποιο είναι το καταλληλότερο μέτρο κεντρικής τάσης : ο μέσος, η διάμεσος ή η επικρατούσα τιμή ; Εξηγείστε την επιλογή σας.
   α) Οι μισθοί όλων των καθηγητών του σχολείου σας.
   β) Οι χρόνοι μέσα στους οποίους 10 άλογα διανύουν μία δεδομένη απόσταση.
   γ) Την ώρα με ακρίβεια λεπτού, όπως την δείχνουν 10 ρολόγια.
   δ) Οι βαθμοί του τμήματός σας στο πρώτο διαγώνισμα στατιστικής.

3. Η ομαδοποιημένη κατανομή συχνοτήτων των ηλικιών 358 υπαλλήλων ενός εργοστασίου παρουσιάζεται στον πίνακα που ακολουθεί:

ΗΛΙΚΙΑ (με βάση τα τελευταία γενέθλια)	ΑΡΙΘΜΟΣ ΥΠΑΛΛΗΛΩΝ
16-20	36
21-25	56
26-30	58
31-35	42
36-40	46
41-45	38
46-50	36
51-55	18
56-60	18

Να εκτιμηθεί, στον πλησιέστερο μήνα, ο μέσος και η τυπική απόκλιση των ηλικιών των υπαλλήλων. Με γραφική ή άλλη μέθοδο να εκτιμήσετε :
α) τη διάμεσο των ηλικιών με προσέγγιση στον πλησιέστερο μήνα.
β) το ποσοστό, με ακρίβεια πρώτου δεκαδικού ψηφίου, των υπαλλήλων που έχουν ηλικία μεγαλύτερη των 26 ετών και μικρότερη των 56 ετών.

4. Οι αριθμοί 4, 6, 12, 4, 10, 12, 3, x, y έχουν μέσο 7 και επικρατούσα τιμή 4. Να βρεθούν :
   α) οι τιμές των x και y
   β) η διάμεσος του συνόλου των 9 αριθμών.
   Όταν προστεθούν δύο επιπλέον αριθμοί 7+n και 7-n, η τυπική απόκλιση των 11 πλέον αριθμών είναι ίση με 4. Να γραφεί ο μέσος των 11 αριθμών και να υπολογιστεί ο αριθμός n.

5. Ο μέσος των παρακάτω αριθμών : 3, 1, 7, 2, 11, 7, x, y όπου οι αριθμοί x, y είναι μονοψήφιοι, θετικοί ακέραιοι, είναι 4. Να δείξετε ότι x+y=14.
   Να βρείτε την επικρατούσα τιμή αυτού του συνόλου των αριθμών, όταν :

   α) $x = y$ και β) $x \neq y$

   Αν η τυπική απόκλιση είναι $\frac{1}{3} \cdot \sqrt{76}$, να βρεθεί το x, y δοθέντος ότι $x < y$.

ΔΡΑΣΤΗΡΙΟΤΗΤΕΣ

Η εκμάθηση των βασικών εννοιών της στατιστικής γίνεται πιο ενδιαφέρουσα, όταν μπορούμε να εργαστούμε με πληροφορίες που συλλέγονται από εμάς και αποθηκεύονται σε έναν υπολογιστή. Η απλούστερη συλλογή δεδομένων μπορεί να αφορά πληροφορίες για τους συμμαθητές σας. Η έρευνα που ακολουθεί μπορεί να σας δώσει αρκετά δεδομένα με τα οποία μπορείτε να απαντήσετε σε μερικές ενδιαφέρουσες ερωτήσεις. Αποθηκεύστε τα δεδομένα που θα συλλέξετε! Στο τέλος του κάθε κεφαλαίου θα υπάρχουν ερωτήσεις που θα αφορούν τα δεδομένα αυτά.

(Σημειώστε με το σύμβολο Χ αν δεν γνωρίζεται την απάντηση)

1. Φύλο
2. Ηλικία
3. Ύψος (σε cm)
4. Ύψος του πατέρα σας
5. Ύψος της μητέρας σας
6. Το πέμπτο ψηφίο του αριθμού της ταυτότητάς σας
7. Το τελευταίο ψηφίο του αριθμού ενός αυτοκινήτου της οικογένειας σας
8. Χρώμα μαλλιών
9. Χρώμα ματιών
10. Είστε δεξιόχειρας, αριστερόχειρας ή αμφίχειρας ;
11. Τους βαθμούς σας στα μαθήματα κατεύθυνσης
12. Προσθέστε μία ή περισσότερες ερωτήσεις που σας ενδιαφέρουν.

Στη συνέχεια παρουσιάζεται ένας εύκολος τρόπος συλλογής των δεδομένων, ώστε κάθε μαθητής να έχει από ένα αντίγραφο.

- Κάθε μαθητής να απαντήσει τις ερωτήσεις σε μία σελίδα
- Μοιράστε ένα έντυπο ερωτηματολόγιου, όπου κάθε μαθητής θα γράφει τις απαντήσεις του. Στο τέλος θα πρέπει να υπάρχει κάτι παρόμοιο με το υπόδειγμα που ακολουθεί
- Να δοθεί ένα αντίγραφο σε κάθε μαθητή.

ΥΠΟΔΕΙΓΜΑ

α/α	1 Φύλο	2 Ηλικία	3 Ύψος	4 Ύψος πατέρα	4 Ύψος μητέρας	6 Σε μηνών	7 τελευταίο μηνών	8 Μάθημα	9 Μάθημα	10 Χέρι	11 Μ	Φ	Χ
1	Θ	19	167	178	158	8	9	Κ	Μ	Α	77	86	69
2	Θ	17	160	175	163	4	2	Κ	Κ	Δ	54	76	80
3	Α	18	170	184	168	0	2	Μ	Μ	Δ	32	56	90
4	Θ	17	173	175	170	4	2	Κ	Μ	Δ	95	73	48
5	Θ	19	158	182	168	6	6	Ξ	Κ	Α	66	48	54

Επεξεργαστείτε τα δεδομένα που συγκεντρώσατε με κατάλληλο στατιστικό πρόγραμμα και

α) προσδιορίστε το είδος κάθε μεταβλητής

β) κατασκευάστε τον πίνακα συχνοτήτων, σχετικών συχνοτήτων και αθροιστικών συχνοτήτων.

γ) κατασκευάστε τα αντίστοιχα διαγράμματα ανάλογα με το είδος της μεταβλητής

δ) και υπολογίστε τα μέτρα θέσης και διασποράς όπου είναι δυνατόν.