Η.Ζ

# ΣΤΟΙΧΕΙΑ ΠΕΡΙΓΡΑΦΙΚΗΣ ΣΤΑΤΙΣΤΙΚΗΣ

## ΜΕΡΟΣ I


**ΒΙΒΛΙΟΓΡΑΦΙΑ**

i) ΜΑΘΗΜΑΤΙΚΑ & ΣΤΟΙΧΕΙΑ ΣΤΑΤΙΣΤΙΚΗΣ (Ο.Ε.Δ.Β)

ii) ΣΤΑΤΙΣΤΙΚΗ ΕΠΙΧΕΙΡΗΣΕΩΝ – Τ.Ε.Ε (Ο.Ε.Δ.Β)

iii) ΣΤΑΤΙΣΤΙΚΗ (ΠΕΤΡΟΥ ΚΙΟΧΟΥ)


- 2004 -

# ΠΕΡΙΕΧΟΜΕΝΑ

# ΣΤΑΤΙΣΤΙΚΗ

**Στατιστική** είναι ένα σύνολο αρχών και μεθοδολογιών με στόχο

- το σχεδιασμό της διαδικασίας συλλογής δεδομένων,
- τη συνοπτική και αποτελεσματική παρουσίασή τους,
- την ανάλυση και εξαγωγή αντίστοιχων συμπερασμάτων.

Ο όρος "**Στατιστική**" ενδεχομένως να προέρχεται από τη λατινική λέξη "status" (πολιτεία, κράτος) η οποία, χρησιμοποιήθηκε αρχικά για το χαρακτηρισμό αριθμητικών δεδομένων που αναφέρονται κυρίως στον πληθυσμό μιας χώρας. Μπορεί όμως να προέρχεται από την αρχαία ελληνική λέξη στατίζω (τοποθετώ, ταξινομώ, συμπεραίνω).Η Στατιστική σήμερα χρησιμοποιείται ευρύτατα σε όλους σχεδόν τους τομείς της ανθρώπινης δραστηριότητας. Βασικές έννοιες της Στατιστικής έχουν εισχωρήσει και ενσωματωθεί σε όλες τις επιστήμες. Από τις Ανθρωπιστικές, Νομικές και Κοινωνικές Επιστήμες (Αρχαιολογία, Λαογραφία, Κοινωνιολογία, Δημογραφία, ...), τις Φυσικές Επιστήμες (Φυσική, Χημεία, ...), τις Επιστήμες Υγείας (Ιατρική, Φαρμακευτική, Βιολογία, ...), τις Τεχνολογικές Επιστήμες (Μηχανολογία, Τοπογραφία, Ναυπηγική, ...) μέχρι τις Επιστήμες Οικονομίας και Διοίκησης (Οικονομικά, Χρηματιστηριακά, Διαφήμιση, Marketing, ...), βλέπουμε να υπεισέρχεται η Στατιστική είτε με την αρχική περιγραφική μορφή της είτε με τις προηγμένες αναλυτικές τεχνικές της. Η ανάλυση στατιστικών ερευνών είναι το κυριότερο εργαλείο έρευνας σε ένα μεγάλο φάσμα εφαρμογών των παραπάνω επιστημών.

Οι έρευνες των ανθρώπινων πληθυσμών (συχνά αναφερόμενες και ως **δημοσκοπήσεις**) αποτελούν σπουδαίες πηγές βασικής γνώσης των κοινωνικών επιστημών. Οικονομολόγοι, ψυχολόγοι, κοινωνιολόγοι και πολιτικοί επιστήμονες μελετούν ποικίλα θέματα όπως πρότυπα εσόδων-εξόδων των οικογενειών και των επιχειρήσεων, την επίδραση της επαγγελματικής απασχόλησης των γυναικών στην οικογενειακή ζωή, τις συγκοινωνιακές και ταξιδιωτικές συνήθειες των κατοίκων μιας πόλης, τις προτιμήσεις των ψηφοφόρων για τους υποψηφίους και τις θέσεις τους. Πολλά προβλήματα που αντιμετωπίζουν σήμερα οι επιχειρήσεις αφορούν τη διατήρηση, αντικατάσταση ή το κρίσιμο σημείο αντοχής συσκευών ή προσωπικού. Ο διευθυντής μιας βιομηχανίας πρέπει να είναι σε θέση να κατανοεί στατιστικές έρευνες που αφορούν την ποιότητα του προϊόντος και την αποδοτικότητα της παραγωγικής διαδικασίας. Πέρα από όλα αυτά, διαπιστώνουμε ολοένα και περισσότερο να γίνεται χρήση μεθόδων της Στατιστικής για την υποστήριξη διάφορων θέσεων. Ακόμα και σε τηλεοπτικές αντιπαραθέσεις (κυρίως σε προεκλογικές περιόδους) βλέπουμε τους συνομιλητές να κάνουν χρήση αριθμών, στατιστικών στοιχείων, γραφημάτων και διαγραμμάτων, για να δώσουν εγκυρότητα στις απόψεις τους και να πείσουν για τα λεγόμενά τους.

Η μελέτη και η γνώση της Στατιστικής βοηθά όχι μόνο στη σωστή χρήση των γνωστών μεθόδων αλλά και στην ανάπτυξη νέων τεχνικών για την αποτελεσματικότερη εξαγωγή χρήσιμων συμπερασμάτων.

# ΒΑΣΙΚΕΣ ΣΤΑΤΙΣΤΙΚΕΣ ΕΝΝΟΙΕΣ

## ΠΛΗΘΥΣΜΟΣ - ΜΕΤΑΒΛΗΤΕΣ

Όπως αναφέρθηκε και προηγουμένως, αυτό που μας ενδιαφέρει είναι να εξετάσουμε τα στοιχεία ενός συνόλου ως προς ένα ή περισσότερα χαρακτηριστικά τους.

Αυτό συμβαίνει, για παράδειγμα, όταν ενδιαφερόμαστε για:

α) τις προτιμήσεις των ψηφοφόρων εν όψει των προσεχών εκλογών
β) τον αριθμό των υπαλλήλων μιας επιχείρησης
γ) τις συνέπειες του καπνίσματος στην υγεία των καπνιστών κτλ.

Σε καθένα από τα παραδείγματα αυτά έχουμε ένα σύνολο και θέλουμε να εξετάσουμε τα στοιχεία του ως προς ένα ή περισσότερα χαρακτηριστικά τους.
Ένα τέτοιο σύνολο λέγεται **πληθυσμός**.

**ΠΛΗΘΥΣΜΟΣ :** αποτελείται από το σύνολο των τιμών μιας τυχαίας μεταβλητής *X*, η οποία εκφράζει ποσοτικά το υπό μελέτη χαρακτηριστικό

Τα στοιχεία του πληθυσμού συχνά αναφέρονται και ως μονάδες ή άτομα του πληθυσμού. Στο πρώτο παράδειγμα έχουμε το σύνολο των ψηφοφόρων και μας ενδιαφέρει η προτίμησή τους, ποιο "κόμμα" π.χ. υποστηρίζουν.

Τα χαρακτηριστικά ως προς τα οποία εξετάζουμε έναν πληθυσμό λέγονται **μεταβλητές** και τις συμβολίζουμε συνήθως με τα κεφαλαία γράμματα $X, Y, Z, B, \ldots$
Οι δυνατές τιμές που μπορεί να πάρει μια μεταβλητή λέγονται **τιμές της μεταβλητής**.

**ΜΕΤΑΒΛΗΤΕΣ :** είναι τα χαρακτηριστικά ως προς τα οποία εξετάζουμε ένα πληθυσμό και τις συμβολίζουμε με τα γράμματα *X, Y, ...*

Από τη διαδοχική εξέταση των ατόμων του πληθυσμού ως προς ένα χαρακτηριστικό τους προκύπτει μια σειρά από δεδομένα, που λέγονται στατιστικά δεδομένα ή παρατηρήσεις.
Τα στατιστικά δεδομένα δεν είναι κατ' ανάγκη διαφορετικά.
Για παράδειγμα, αν εξετάζουμε την ομάδα αίματος δέκα ατόμων, τα στατιστικά δεδομένα ή παρατηρήσεις που θα προκύψουν μπορεί να είναι: Α, Α, Β, Α, ΑΒ, Ο, ΑΒ, ΑΒ, ΑΒ, Ο, Β.
Οι δυνατές όμως τιμές που μπορεί να πάρει η μεταβλητή "ομάδα αίματος" είναι οι εξής τέσσερις: Α, Β, ΑΒ και Ο.

Τις μεταβλητές τις διακρίνουμε:

1. Σε **ποιοτικές ή κατηγορικές** μεταβλητές, των οποίων οι τιμές τους δεν είναι αριθμοί. Τέτοιες είναι, για παράδειγμα, η ομάδα αίματος (με τιμές *Α*, *Β*, *ΑΒ*, *Ο*), το φύλο (με τιμές αγόρι, κορίτσι), οι συνέπειες του καπνίσματος (με τιμές καρδιακά νοσήματα, καρκίνος κτλ), όπως επίσης και η οικονομική κατάσταση και η υγεία των ανθρώπων (που μπορεί να χαρακτηριστεί ως κακή, μέτρια, καλή ή πολύ καλή), καθώς και το ενδιαφέρον των μαθητών για τη Στατιστική, που μπορεί να χαρακτηριστεί ως υψηλό, μέτριο, χαμηλό ή μηδαμινό.

2. Σε **ποσοτικές** μεταβλητές, των οποίων οι τιμές είναι αριθμοί και διακρίνονται:
   i. Σε **διακριτές ή ασυνεχής** μεταβλητές, που παίρνουν μόνο "μεμονωμένες" τιμές. Τέτοιες μεταβλητές είναι, για παράδειγμα, ο αριθμός των υπαλλήλων μιας επιχείρησης (με τιμές 1,2,…), το αποτέλεσμα της ρίψης ενός ζαριού (με τιμές 1,2,…,6) κτλ.
   ii. Σε **συνεχείς** μεταβλητές, που μπορούν να πάρουν οποιαδήποτε τιμή ενός διαστήματος πραγματικών αριθμών $(\alpha, \beta)$. Τέτοιες μεταβλητές είναι το ύψος και το βάρος των μαθητών της Γ΄ Λυκείου, ο χρόνος που χρειάζονται οι μαθητές να απαντήσουν στα θέματα μιας εξέτασης, η διάρκεια μιας τηλεφωνικής συνδιάλεξης κτλ.

# ΣΥΛΛΟΓΗ ΣΤΑΤΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

**ΑΠΟΓΡΑΦΗ** : Είναι η μέθοδος συλλογής δεδομένων, όταν εξετάζουμε όλα τα στοιχεία ενός πληθυσμού.

**ΔΕΙΓΜΑΤΟΛΗΨΙΑ** : συλλέγει πληροφορίες από κάποια μικρή ομάδα η υποσύνολο του πληθυσμού το οποίο καλείται **δείγμα**, έτσι ώστε κάθε μονάδα του πληθυσμού να έχει την ίδια πιθανότητα να επιλεγεί.

Κατά τη δειγματοληψία ο ερευνητής κάνει τις παρατηρήσεις του στο δείγμα αυτό και στη συνέχεια γενικεύει τα συμπεράσματα του για ολόκληρο τον πληθυσμό.

Το πλήθος των ατόμων που αποτελούν το δείγμα καλείται **τάξη ή μέγεθος** του δείγματος.

**ΑΠΟΓΡΑΦΗ**

Από όλες τις μορφές της απογραφής, η απογραφή του πληθυσμού αποτελεί την σπουδαιότερη γιατί αποτελεί την κύρια πηγή εξαγωγής συμπερασμάτων για τα δημογραφικά, κοινωνικά και οικονομικά προβλήματα του πληθυσμού.
Η απογραφή του πληθυσμού γίνεται από την στατιστική υπηρεσία της κάθε χώρας.
Συνήθως γίνεται κάθε 10 χρόνια και στα έτη που λήγουν σε 0 ή 1.
Η εποχή που γίνεται η απογραφή είναι άνοιξη ή φθινόπωρο. Η προτίμηση αυτή δεν είναι τυχαία κι αυτό γιατί τις εποχές αυτές οι άνθρωποι ασχολούνται με τις γεωργικές καλλιέργειες οπότε οι μετακινήσεις στο εσωτερικό της κάθε χώρας δεν είναι σημαντικές. Ακόμη τις εποχές αυτές δεν υπάρχει μεγάλη τουριστική κίνηση.
Η απογραφή γίνεται συνήθως Κυριακή.

Η μέθοδος της απογραφής παρουσιάζει τα ακόλουθα μειονεκτήματα:

1. Απαιτεί μεγάλο κόστος
2. Επειδή ο πληθυσμός και το πλήθος των πληροφοριών είναι μεγάλο, η δημοσίευση των αποτελεσμάτων καθυστερεί.
3. Η απογραφή δε γίνεται από ειδικευμένο προσωπικό και έτσι μπορεί να υπάρχουν σφάλματα στο πλήθος των ερωτηματολογίων που συμπληρώνονται.

## ΔΕΙΓΜΑΤΟΛΗΨΙΑ

Όταν η απογραφή είναι πρακτικά αδύνατη ή οικονομικά ασύμφορη χρησιμοποιείται η μέθοδος της δειγματοληψίας.
Τα κυριότερα πλεονεκτήματα της δειγματοληπτικής μεθόδου είναι:
Η μεγαλύτερη ταχύτητα πληροφοριών, η μεγαλύτερη ακρίβεια στα αποτελέσματα, η μεγαλύτερη ευχέρεια εφαρμογής, το χαμηλό κόστος, η ολοκληρωτική αδυναμία εφαρμογής της απογραφής
Παράλληλα με τα πλεονεκτήματα της δειγματοληψίας υπάρχουν και μερικά μειονεκτήματα. Αναφέρουμε τα σημαντικότερα

1. Κακή σχεδίαση και εκτέλεση της δειγματοληψίας
2. Μη σωστή επιλογή του δείγματος
3. Μη κατάλληλη μέθοδος διενέργειας της δειγματοληψίας
4. Δειγματοληπτικά σφάλματα

Οι κυριότερες μέθοδοι διενέργειας της δειγματοληψίας είναι:

1. Η απλή-τυχαία δειγματοληψία
2. Δειγματοληψία κατά στρώματα
3. Επιφανειακή δειγματοληψία
4. Δειγματοληψία κατά ομάδες
5. Μεροληπτική δειγματοληψία
6. Συστηματική δειγματοληψία

Η επιλογή του αντιπροσωπευτικού δείγματος είναι "εκ των ων ουκ άνευ". Αποτελεί πολύ σοβαρή και δύσκολη διαδικασία. Ο κακός σχεδιασμός και η εκτέλεση της στατιστικής έρευνας, η μη αντιπροσωπευτικότητα του δείγματος, ο μη σωστός καθορισμός του μεγέθους του δείγματος αποτελούν μερικά βασικά μειονεκτήματα στη διαδικασία επιλογής ενός δείγματος.
Αξίζει να σημειωθεί ότι μία "προσεκτική" επιλογή μικρότερου δείγματος είναι δυνατόν να δώσει καλύτερα αποτελέσματα από ένα μεγαλύτερο δείγμα που δεν έχει εκλεγεί κατάλληλα. Ενδεικτικό είναι το παράδειγμα των προεδρικών εκλογών των ΗΠΑ το 1936. Το περιοδικό Literary Digest χρησιμοποιώντας δείγμα 2.400.000 ατόμων πρόβλεψε νίκη του Landon με ποσοστό 57%. Αντίθετα, το δημοσκοπικό γραφείο του G. Gallup χρησιμοποιώντας δείγμα 50.000 ατόμων πρόβλεψε το σωστό αποτέλεσμα που ήταν νίκη του Roosvelt με ποσοστό 62%! Η παταγώδης αποτυχία της δημοσκόπησης του περιοδικού οφειλόταν στο γεγονός ότι το δείγμα που επελέγη δεν ήταν αντιπροσωπευτικό του πληθυσμού.

# ΠΑΡΟΥΣΙΑΣΗ ΣΤΑΤΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

## ΣΤΑΤΙΣΤΙΚΟΙ ΠΙΝΑΚΕΣ

Μετά τη συλλογή των στατιστικών δεδομένων είναι αναγκαία η κατασκευή συνοπτικών **πινάκων** ή **γραφικών παραστάσεων**, ώστε να είναι εύκολη η κατανόησή τους και η εξαγωγή σωστών συμπερασμάτων. Η παρουσίαση των στατιστικών δεδομένων σε πίνακες γίνεται με την κατάλληλη τοποθέτηση των πληροφοριών σε γραμμές και στήλες, με τρόπο που να διευκολύνεται η σύγκριση των στοιχείων και η καλύτερη ενημέρωση του αναγνώστη σχετικά με τη δομή του πληθυσμού που ερευνάμε.

**Στατιστικοί Πίνακες** είναι οι πίνακες που περιέχουν στατιστικά δεδομένα και αποτελούνται από :

1. τον τίτλο που δηλώνει το περιεχόμενο του πίνακα,
2. τις επικεφαλίδες γραμμών και στηλών που δείχνουν τη φύση και τις μονάδες μέτρησης των δεδομένων,
3. τον κύριο πίνακα που περιέχει τα στατιστικά δεδομένα,
4. και την πηγή προέλευσης των στατιστικών στοιχείων.

Οι πίνακες διακρίνονται στους:
α) **γενικούς πίνακες**, οι οποίοι περιέχουν όλες τις πληροφορίες που προκύπτουν από μία στατιστική έρευνα (συνήθως με αρκετά λεπτομερειακά στοιχεία) και αποτελούν πηγές στατιστικών πληροφοριών στη διάθεση των επιστημόνων-ερευνητών για παραπέρα ανάλυση και εξαγωγή συμπερασμάτων,
β) **ειδικούς πίνακες**, οι οποίοι είναι συνοπτικοί και σαφείς. Τα στοιχεία τους συνήθως έχουν ληφθεί από τους γενικούς πίνακες.

## Πίνακες Κατανομής Συχνοτήτων

Ας υποθέσουμε ότι $x_1, x_2, ..., x_\kappa$ είναι οι τιμές μιας μεταβλητής $X$, που αφορά τα άτομα ενός δείγματος μεγέθους $\nu$, $\kappa \leq \nu$.
Στην τιμή $x_i$ αντιστοιχίζεται η (απόλυτη) **συχνότητα** $\nu_i$, δηλαδή ο φυσικός αριθμός που δείχνει πόσες φορές εμφανίζεται η τιμή $x_i$ της εξεταζόμενης μεταβλητής $X$ στο σύνολο των παρατηρήσεων.
Είναι φανερό ότι το άθροισμα όλων των συχνοτήτων είναι ίσο με το μέγεθος $\nu$ του δείγματος, δηλαδή:

$$\nu_1 + \nu_2 + \ldots + \nu_\kappa = \nu$$

Αν διαιρέσουμε τη συχνότητα $\nu_i$ με το μέγεθος ν του δείγματος, προκύπτει η **σχετική συχνότητα** $f_i$ της τιμής $x_i$, δηλαδή

$$\boxed{f_i = \frac{\nu_i}{\nu}, \qquad i = 1,2,\ldots,\kappa}.$$

Για τη σχετική συχνότητα ισχύουν οι ιδιότητες:

(i) $0 \le f_i \le 1$ για $i = 1,2,\ldots,\kappa$ αφού $0 \le \nu_i \le \nu$.

(ii) $f_1 + f_2 + \ldots + f_\kappa = 1$, αφού

$$f_1 + f_2 + \ldots + f_\kappa = \frac{\nu_1}{\nu} + \frac{\nu_2}{\nu} + \ldots + \frac{\nu_\kappa}{\nu} = \frac{\nu_1 + \nu_2 + \ldots + \nu_\kappa}{\nu} = \frac{\nu}{\nu} = 1.$$

Συνήθως, τις σχετικές συχνότητες $f_i$ τις εκφράζουμε επί τοις εκατό, οπότε συμβολίζονται με $f_i\%$, δηλαδή $f_i\% = 100 f_i$.

Οι ποσότητες $x_i, \nu_i, f_i$ για ένα δείγμα συγκεντρώνονται σε ένα συνοπτικό πίνακα, που ονομάζεται **πίνακας κατανομής συχνοτήτων** ή απλά **πίνακας συχνοτήτων**.
Για μια μεταβλητή, το σύνολο των ζευγών $(x_i, \nu_i)$ λέμε ότι αποτελεί την **κατανομή συχνοτήτων** και το σύνολο των ζευγών $(x_i, f_i)$, ή των ζευγών $(x_i, f_i\%)$, την **κατανομή των σχετικών συχνοτήτων**.

<u>Στην περίπτωση των **ποσοτικών μεταβλητών**</u> εκτός από τις συχνότητες $\nu_i$ και $f_i$ χρησιμοποιούνται συνήθως και οι λεγόμενες **αθροιστικές συχνότητες** $N_i$ και οι **αθροιστικές σχετικές συχνότητες** $F_i$, οι οποίες εκφράζουν το πλήθος και το ποσοστό αντίστοιχα των παρατηρήσεων που είναι μικρότερες ή ίσες της τιμής $x_i$.
Συχνά οι $F_i$ πολλαπλασιάζονται επί 100 εκφραζόμενες έτσι επί τοις εκατό, δηλαδή $F_i\% = 100F_i$.

Αν οι τιμές $x_1, x_2, \ldots, x_\kappa$ μιας ποσοτικής μεταβλητής Χ είναι σε αύξουσα διάταξη, τότε η αθροιστική συχνότητα της τιμής $x_i$ είναι $\boxed{N_i = \nu_1 + \nu_2 + \ldots + \nu_i}$, για $i = 1,2,\ldots,\kappa$.

Όμοια, η αθροιστική σχετική συχνότητα είναι $\boxed{F_i = f_1 + f_2 + \ldots + f_i}$, για $i = 1,2,\ldots,\kappa$.

| |
|---|
| Για μια μεταβλητή Χ το σύνολο των ζευγών $(x_i, \nu_i)$ λέγεται **κατανομή συχνοτήτων** και το σύνολο των ζευγών $(x_i, f_i)$ λέγεται **κατανομή σχετικών συχνοτήτων**. |

Η γενική μορφή ενός στατιστικού πίνακα είναι η παρακάτω :

**ΤΙΤΛΟΣ ΣΤΑΤΙΣΤΙΚΟΥ ΠΙΝΑΚΑ**

| **Μεταβλητές $X_i$** | **Συχνότητα $v_i$** | **Σχετική συχνότητα $f_i$** | **Σχετική συχνότητα % $f_i\%$** | **Αθροιστική συχνότητα $N_i$** | **Αθροιστική σχετ. συχν. $F_i$** | **Αθροιστική σχετ. συχν.% $F_i\%$** |
|---|---|---|---|---|---|---|
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| **Σύνολα** | Μέγεθος Δείγματος | 1 | 100 | | | |

ΠΗΓΗ ΣΤΑΤΙΣΤΙΚΩΝ ΠΛΗΡΟΦΟΡΙΩΝ

Οι στατιστικοί πίνακες της ανωτέρω μορφής ονομάζονται και **πίνακες κατανομής συχνοτήτων**.

# ΕΡΩΤΗΣΕΙΣ – ΑΣΚΗΣΕΙΣ

## Ερωτήσεις του τύπου «Σωστό - Λάθος»

| | | |
|---|---|---|
| 1. Το χρώμα κάθε αυτοκινήτου είναι ποιοτική μεταβλητή. | Σ | Λ |
| 2. Το άθροισμα όλων των συχνοτήτων μιας κατανομής είναι ίσο με 1. | Σ | Λ |
| 3. Η συχνότητα της τιμής $x_i$ μιας μεταβλητής X είναι αρνητικός αριθμός. | Σ | Λ |
| 4. Αν διαιρέσουμε τη συχνότητα $v_i$ μιας μεταβλητής X με το μέγεθος ν του δείγματος, προκύπτει η σχετική συχνότητα $f_i$ της τιμής $x_i$. | Σ | Λ |
| 5. Το άθροισμα όλων των σχετικών συχνοτήτων μιας κατανομής είναι ίσο με το μέγεθος ν του δείγματος. | Σ | Λ |
| 6. Το σύνολο των ζευγών $(x_i, f_i)$, όπου $f_i$ η σχετική συχνότητα της τιμής $x_i$, αποτελεί την κατανομή των σχετικών συχνοτήτων. | Σ | Λ |
| 7. Οι αθροιστικές συχνότητες $N_i$ μιας κατανομής εκφράζουν το πλήθος των παρατηρήσεων που είναι μικρότερες ή ίσες της τιμής $x_i$. | Σ | Λ |
| 8. Οι αθροιστικές σχετικές συχνότητες $F_i$ μιας κατανομής εκφράζουν το ποσοστό των παρατηρήσεων που είναι μεγαλύτερες ή ίσες της τιμής $x_i$. | Σ | Λ |

## Ερωτήσεις πολλαπλής επιλογής

1. Από τις παρακάτω μεταβλητές διακριτή ποσοτική είναι

   Α. το βάρος μαθητών.

   Β. η μηνιαία κατανάλωση ρεύματος.

   Γ. ο χαρακτηρισμός της διαγωγής των μαθητών.

   Δ. ο αριθμός απουσιών.

   Ε. η ποιότητα του περιεχομένου των βιβλίων.

2. Το ζεύγος που αποτελεί την κατανομή συχνοτήτων είναι

Α. $(x_i, v_i)$. Β. $(x_i, f_i)$. Γ. $(v_i, f_i)$. Δ. $(x_i f_i, vx_i)$. Ε. $(vf_i, x_i)$.

3. Σε ένα δείγμα μεγέθους ν με συχνότητα $v_i$ της τιμής $x_i$ μιας μεταβλητής Χ η σχετική συχνότητα $f_i$ ισούται με

Α. $f_i = \dfrac{v}{v_i}$. Β. $f_i = \dfrac{v_i}{v}$. Γ. $f_i = v_i - v$. Δ. $f_i = v_i \cdot v$. Ε. $f_i = \dfrac{100}{v_i}$.

4. Αν $x_1, x_2, \ldots, x_\kappa$ είναι οι τιμές μιας μεταβλητής Χ, που αφορά τα άτομα ενός δείγματος μεγέθους ν, τότε αν στην τιμή $x_i$ αντιστοιχίσουμε τη συχνότητα $v_i$ ισχύει

Α. $v_1 + v_2 + \ldots + v_\kappa = 100$. Β. $v_1 + v_2 + \ldots + v_\kappa = v$.

Γ. $v_1 + v_2 + \ldots + v_\kappa = \kappa$. Δ. $v_1 + v_2 + \ldots + v_\kappa = v\kappa$. Ε. $v_1 + v_2 + \ldots + v_\kappa = 100v$.

5. Αν $x_1, x_2, \ldots, x_\kappa$ είναι οι τιμές μιας μεταβλητής Χ, που αφορά τα άτομα ενός δείγματος μεγέθους ν, $\kappa \leq v$, τότε για τις σχετικές συχνότητες $f_1, f_2, \ldots, f_\kappa$ ισχύει

Α. $f_1 + f_2 + \ldots + f_\kappa = 100$. Β. $f_1 + f_2 + \ldots + f_\kappa = \kappa^2$.

Γ. $f_1 + f_2 + \ldots + f_\kappa = 1$. Δ. $f_1 + f_2 + \ldots + f_\kappa = 100\kappa$. Ε. $f_1 + f_2 + \ldots + f_\kappa = \kappa$.

## Ερωτήσεις συμπλήρωσης - απάντησης

1. Ένα σύνολο στο οποίο εξετάζουμε τα στοιχεία του ως προς ένα ή περισσότερα χαρακτηριστικά του λέγεται ………………………

2. Τα χαρακτηριστικά ως προς τα οποία εξετάζουμε έναν πληθυσμό λέγονται ………………………

3. Οι δυνατές τιμές που μπορεί να πάρει μια μεταβλητή λέγονται ……………

4. Διακρίνουμε τις μεταβλητές σε:

α)......................................., των οποίων οι τιμές δεν είναι αριθμοί και

β) ......................................, των οποίων οι τιμές είναι αριθμοί και διακρίνονται σε:

i) ..............................., που παίρνουν μόνο "μεμονωμένες" τιμές και

ii) ..............................., που μπορούν να πάρουν οποιαδήποτε τιμή ενός διαστήματος πραγματικών αριθμών.

5. Ένας τρόπος για να πάρουμε τις απαραίτητες πληροφορίες που χρειαζόμαστε για κάποιο πληθυσμό είναι να εξετάσουμε όλα τα άτομα του πληθυσμού ως προς το χαρακτηριστικό που μας ενδιαφέρει. Η μέθοδος αυτή συλλογής των δεδομένων ονομάζεται ………………………

6. Μετά τη συλλογή των στατιστικών δεδομένων είναι αναγκαία η κατασκευή συνοπτικών ………………………………, ώστε να είναι εύκολη η κατανόησή τους και η εξαγωγή σωστών συμπερασμάτων.

7. Ας υποθέσουμε ότι $x_1$, $x_2$, …, $x_κ$ είναι οι τιμές μιας μεταβλητής Χ, που αφορά τα άτομα ενός δείγματος μεγέθους ν, $κ \leq ν$. Στην τιμή $x_i$ αντιστοιχίζεται η ………………………, δηλαδή ο φυσικός αριθμός που δείχνει πόσες φορές εμφανίζεται η τιμή $x_i$ της εξεταζόμενης μεταβλητής Χ στο σύνολο των παρατηρήσεων.

8. Οι ποσότητες $x_i$, $ν_i$, $f_i$ για ένα δείγμα συγκεντρώνονται σε ένα συνοπτικό πίνακα, που ονομάζεται ……………………… ή απλά …………………… .

9. Για μια μεταβλητή, το σύνολο των ζευγών $(x_i, ν_i)$ λέμε ότι αποτελεί την ……………………… και το σύνολο των ζευγών $(x_i, f_i)$, ή των ζευγών $(x_i, f_i\%)$, την …………………… .

10. Να συμπληρώσετε τον παρακάτω πίνακα:

| $x_i$ | $ν_i$ | $f_i$ | $N_i$ | $F_i$ | $f_i\,\%$ | $F_i\,\%$ |
|---|---|---|---|---|---|---|
| 1 | 8 | 0,4 | | | | |
| 2 | | | 10 | | | |
| 3 | 5 | 0,25 | 15 | | | |
| 4 | | | | | | 90 |
| 5 | | | | | 10 | |
| Σύνολα | | | | | | |

**11.** Οι παρακάτω αριθμοί παρουσιάζουν τις ενδείξεις ενός ζαριού το οποίο ρίξαμε 30 φορές.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 5 | 6 | 1 | 2 | 5 | 4 | 3 | 2 | 5 |
| 1 | 3 | 5 | 4 | 1 | 3 | 2 | 6 | 5 | 4 |
| 1 | 2 | 6 | 2 | 4 | 3 | 1 | 6 | 4 | 5 |

Να κατασκευάσετε πίνακα:

α) Συχνοτήτων.

β) Αθροιστικών συχνοτήτων.

**12.** Οι αποστάσεις (σε km) των 26 κοινοτήτων ενός νομού από το πλησιέστερο νοσοκομείο είναι:

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 10 | 8 | 8 | 13 | 10 | 4 | 2 | 0 | 16 | 5 | 15 | 9 |
| 6 | 4 | 7 | 5 | 4 | 6 | 7 | 7 | 5 | 8 | 10 | 3 | 9 |

α) Να κατασκευάσετε πίνακα: i) Συχνοτήτων.

ii) Αθροιστικών συχνοτήτων των αποστάσεων.

β) Πόσες κοινότητες απέχουν από το νοσοκομείο περισσότερο από 10 km;

**13.** Ο αριθμός των μαθητών των 16 τμημάτων ενός Λυκείου είναι:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 31 | 27 | 28 | 30 | 29 | 31 | 31 | 27 |
| 29 | 29 | 28 | 28 | 30 | 29 | 27 | 29 |

Να κατασκευάσετε πίνακα: i) Σχετικών συχνοτήτων.

ii) Αθροιστικών σχετικών συχνοτήτων.

# ΓΡΑΦΙΚΗ ΠΑΡΑΣΤΑΣΗ ΚΑΤΑΝΟΜΗΣ ΣΥΧΝΟΤΗΤΩΝ

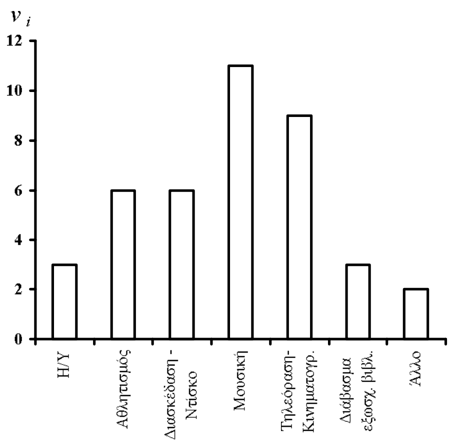
Τα στατιστικά δεδομένα παρουσιάζονται πολλές φορές και υπό μορφή γραφικών παραστάσεων ή διαγραμμάτων. Οι γραφικές παραστάσεις παρέχουν πιο σαφή εικόνα του χαρακτηριστικού σε σχέση με τους πίνακες, είναι πολύ πιο ενδιαφέρουσες και ελκυστικές, χωρίς βέβαια να προσφέρουν περισσότερη πληροφορία από εκείνη που περιέχεται στους αντίστοιχους πίνακες συχνοτήτων.
Επί πλέον με τα διαγράμματα διευκολύνεται η σύγκριση μεταξύ ομοειδών στοιχείων για το ίδιο ή για διαφορετικά χαρακτηριστικά.
Υπάρχουν διάφοροι τρόποι γραφικής παρουσίασης, ανάλογα με το είδος των δεδομένων που έχουμε. Όπως όμως οι στατιστικοί πίνακες έτσι και τα στατιστικά διαγράμματα πρέπει να συνοδεύονται από :
α) τον τίτλο, β) την κλίμακα με τις τιμές των μεγεθών που απεικονίζονται, γ) το υπόμνημα που επεξηγεί συνήθως τις τιμές της μεταβλητής και δ) την πηγή των δεδομένων.

Στη συνέχεια αναφέρονται οι σημαντικότεροι τρόποι γραφικής απεικόνισης των στατιστικών δεδομένων.

## *α) Ραβδόγραμμα*

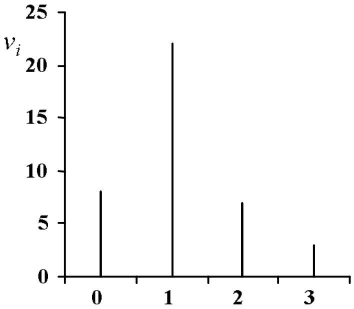
Το **ραβδόγραμμα** χρησιμοποιείται για τη γραφική παράσταση των τιμών ποιοτικών μεταβλητών, ποσοτικών διακριτών μεταβλητών και για την απεικόνιση της διαχρονικής εξέλιξης ενός φαινομένου.
Το ραβδόγραμμα αποτελείται από ορθογώνιες στήλες που οι βάσεις τους βρίσκονται πάνω στον οριζόντιο ή τον κατακόρυφο άξονα.
Σε κάθε τιμή της μεταβλητής $X$ αντιστοιχεί μια ορθογώνια στήλη της οποίας το ύψος είναι ίσο με την αντίστοιχη συχνότητα ή σχετική συχνότητα.
Έτσι έχουμε αντίστοιχα το **ραβδόγραμμα συχνοτήτων** και το **ραβδόγραμμα σχετικών συχνοτήτων**. Τόσο η απόσταση μεταξύ των στηλών όσο και το μήκος των βάσεων τους καθορίζονται αυθαίρετα.

Στο σχήμα παριστάνεται το ραβδόγραμμα συχνοτήτων της "απασχόλησης των μαθητών" .

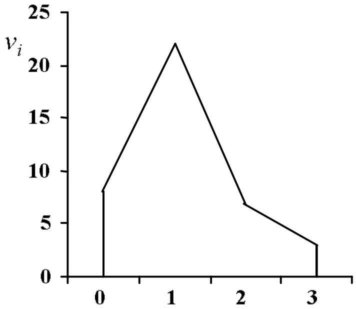


## *β) Διάγραμμα Συχνοτήτων*

Συνήθως στην περίπτωση που έχουμε μια ποσοτική μεταβλητή αντί του ραβδογράμματος χρησιμοποιείται το **διάγραμμα συχνοτήτων** .
Αυτό μοιάζει με το ραβδόγραμμα με μόνη διαφορά ότι αντί να χρησιμοποιούμε συμπαγή ορθογώνια υψώνουμε σε κάθε $x_i$ (υποθέτοντας ότι $x_1 < x_2 < \ldots < x_\kappa$) μία κάθετη γραμμή με μήκος ίσο προς την αντίστοιχη συχνότητα, όπως φαίνεται στο σχήμα (α). Μπορούμε επίσης αντί των συχνοτήτων $v_i$ στον κάθετο άξονα να βάλουμε τις σχετικές συχνότητες $f_i$, οπότε έχουμε το **διάγραμμα σχετικών συχνοτήτων**.
Ενώνοντας τα σημεία $(x_i, v_i)$ ή $(x_i, f_i)$ έχουμε το λεγόμενο **πολύγωνο συχνοτήτων** ή **πολύγωνο σχετικών συχνοτήτων**, αντίστοιχα, που μας δίνουν μια γενική ιδέα για τη μεταβολή της συχνότητας ή της σχετικής συχνότητας όσο μεγαλώνει η τιμή της μεταβλητής που εξετάζουμε, βλέπε σχήμα (β).



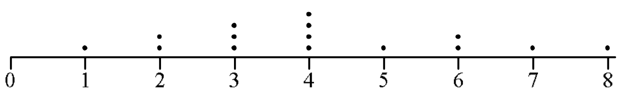

(α)




(β)

## *γ) Σημειόγραμμα*

Όταν έχουμε λίγες παρατηρήσεις, η κατανομή τους μπορεί να περιγραφεί με το **σημειόγραμμα** , στο οποίο οι τιμές παριστάνονται γραφικά σαν σημεία υπεράνω ενός οριζόντιου άξονα.
Στο σχήμα έχουμε το σημειόγραμμα των χρόνων (σε λεπτά) 4,2,3,1,5,6,4,2,3,4,7,4,8,6,3 που χρειάστηκαν δεκαπέντε μαθητές, για να λύσουν ένα πρόβλημα.



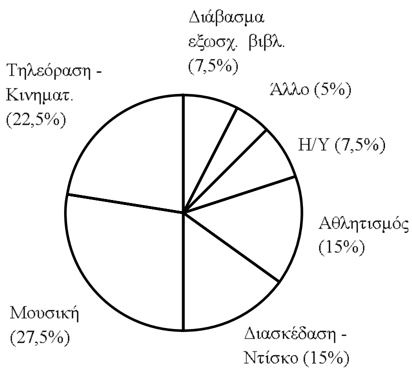

χρόνος (σε λεπτά)

## δ) Κυκλικό Διάγραμμα

Το **κυκλικό διάγραμμα** χρησιμοποιείται για τη γραφική παράσταση τόσο των ποιοτικών όσο και των ποσοτικών δεδομένων, όταν οι διαφορετικές τιμές της μεταβλητής είναι σχετικά λίγες και αναφέρονται σε ορισμένη χρονική στιγμή.
Το κυκλικό διάγραμμα είναι ένας κυκλικός δίσκος χωρισμένος σε κυκλικούς τομείς, τα εμβαδά ή ,ισοδύναμα, τα τόξα των οποίων είναι ανάλογα προς τις αντίστοιχες συχνότητες $v_i$ ή τις σχετικές συχνότητες $f_i$ των τιμών $x_i$ της μεταβλητής.
Αν συμβολίσουμε με $\alpha_i$ το αντίστοιχο τόξο ενός κυκλικού τμήματος στο κυκλικό διάγραμμα συχνοτήτων, τότε

$$\alpha_i = v_i \frac{360^\circ}{v} = 360^\circ f_i \quad \text{για} \quad i = 1,2,\ldots,\kappa .$$

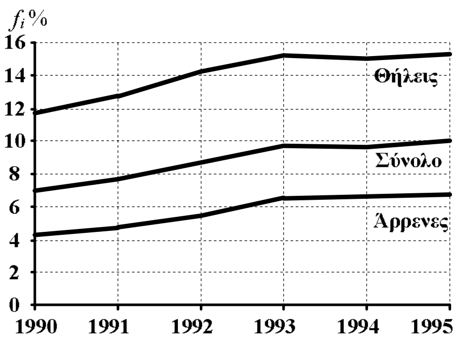
Στο σχήμα παριστάνεται το αντίστοιχο κυκλικό διάγραμμα σχετικών συχνοτήτων της "απασχόλησης των μαθητών" .




## ε) Χρονόγραμμα.

Το χρονόγραμμα ή χρονολογικό διάγραμμα χρησιμοποιείται για τη γραφική απεικόνιση της διαχρονικής εξέλιξης ενός οικονομικού, δημογραφικού ή άλλου μεγέθους. *Ο* οριζόντιος άξονας χρησιμοποιείται συνήθως ως άξονας μέτρησης του χρόνου και ο κάθετος ως άξονας μέτρησης της εξεταζόμενης μεταβλητής. Στο σχήμα έχουμε το χρονόγραμμα του ποσοστού ανεργίας στη χώρα μας από το 1990 έως το 1995. (Πηγή ΕΣΥΕ).

Παρατηρούμε ότι στο γυναικείο πληθυσμό υπάρχει συστηματικά μεγαλύτερο ποσοστό ανεργίας, γύρω στις 8 εκατοστιαίες μονάδες. Στο διάστημα 1993-95 το ποσοστό ανεργίας έχει σταθεροποιηθεί γύρω στο 6,5% για τους άνδρες και γύρω στο 15% για τις γυναίκες.




Ποσοστά ανεργίας στην Ελλάδα

### στ) χαρτογράμματα

Τα χαρτογράμματα είναι γραφικές παραστάσεις στατιστικών στοιχείων πάνω σε γεωγραφικούς ή τοπογραφικούς χάρτες.

### ζ) ειδογράμματα

Τα ειδογράμματα είναι διάφορα σχήματα σε μορφή προσώπων ή πραγμάτων και χρησιμοποιούνται πάρα πολύ γιατί θεωρούνται ως μέσα, πολύ εκφραστικά ενός φαινομένου και μπορεί ο αναγνώστης να συγκρατήσει την εικόνα του ευκολότερα στη μνήμη του.
Παράλληλα με τα ειδογράμματα διευκολύνεται και η σύγκριση δύο η περισσότερων μεγεθών.

## Η πυραμίδα των ηλικιών του ελληνικού πληθυσμού

Η διάρθρωση του πληθυσμού κατά ηλικία παίρνει εκφραστικότερη μορφή αν παρασταθεί γραφικά από το διάγραμμα εκείνο που ονομάζεται πυραμίδα των ηλικιών. Η κατασκευή της πυραμίδας δεν παρουσιάζει καμιά δυσκολία, γιατί στην ουσία είναι ένα διπλό ραβδόγραμμα, όπου στον κάθετο άξονα αντί στον οριζόντιο τοποθετούμε τις ηλικίες , αφού προηγουμένως χωρίσουμε τον άξονα αυτό σε ίσα διαστήματα ηλικιών και στον οριζόντιο τοποθετούμε τις συχνότητες με μορφή ορθογωνίων, το μήκος των οποίων είναι ανάλογο με τον αριθμό των συχνοτήτων που αντιστοιχεί σε κάθε διάστημα ηλικίας περίπου 5 ή 10 χρόνων.

## Στατιστικές εκθέσεις και αναφορές

Εκτός από τους στατιστικούς πίνακες και τις γραφικές παραστάσεις, άλλος τρόπος για την παρουσίαση των στατιστικών στοιχείων, ανάλογα βέβαια με τον σκοπό που επιδιώκουμε είναι και οι εκθέσεις ή αναφορές στο κείμενο των οποίων αναφέρονται τα κυριότερα σημεία των στατιστικών αποτελεσμάτων και σχολιάζονται αυτά.
Ο τρόπος αυτός έχει το μειονέκτημα ότι ο αναγνώστης πρέπει να διαβάσει αρκετές σελίδες για να ενημερωθεί.

# ΕΡΩΤΗΣΕΙΣ – ΑΣΚΗΣΕΙΣ

## Ερωτήσεις του τύπου «Σωστό - Λάθος»

1. Το ραβδόγραμμα χρησιμοποιείται για τη γραφική παράσταση των τιμών μιας ποσοτικής μεταβλητής. Σ Λ
2. Το κυκλικό διάγραμμα χρησιμοποιείται για τη γραφική παράσταση μόνο ποιοτικών δεδομένων. Σ Λ
3. Το κυκλικό διάγραμμα είναι ένας κυκλικός δίσκος χωρισμένος σε κυκλικούς τομείς τα εμβαδά των οποίων είναι αντιστρόφως ανάλογα προς τις αντίστοιχες συχνότητες νi. Σ Λ
4. Το σημειόγραμμα χρησιμοποιείται για τη γραφική απεικόνιση της διαχρονικής εξέλιξης μιας εξεταζόμενης μεταβλητής. Σ Λ
5. Το διπλανό σχήμα είναι ένα χρονόγραμμα. Σ Λ


## Ερωτήσεις πολλαπλής επιλογής

1. Στο κυκλικό διάγραμμα συχνοτήτων αν συμβολίσουμε με $α_i$ το αντίστοιχο τόξο ενός κυκλικού τμήματος τότε το αi ισούται με

   Α. $360° ν_i$. Β. $360° f_i$. Γ. $90° f_i$.

   Δ. $180° ν_i$. Ε. $180° f_i$.

## Ερωτήσεις συμπλήρωσης - σύντομης απάντησης

1. Το …………………… χρησιμοποιείται για τη γραφική παράσταση των τιμών μιας ποιοτικής μεταβλητής. Στην περίπτωση που έχουμε μια ποσοτική μεταβλητή χρησιμοποιείται το διάγραμμα…………………… .

2. Το …………………… διάγραμμα χρησιμοποιείται για τη γραφική παράσταση τόσο των ποιοτικών όσο και των ποσοτικών δεδομένων, όταν οι διαφορετικές τιμές της μεταβλητής είναι σχετικά λίγες.

**Ερωτήσεις ανάπτυξης**

1. Σε ένα κυκλικό διάγραμμα παριστάνονται οι εξαγωγές της χώρας μας αξίας 97.000.000.000 δρχ. κατά το έτος 1980 ανάλογα με το μέσο μεταφοράς. Η γωνία του κυκλικού τομέα για μέσο μεταφοράς "θαλασσίως" είναι 180°. Το 13,917% της αξίας των εξαγωγών έγινε "σιδηροδρομικώς". Οι μεταφορές που έγιναν "οδικώς" ήταν τετραπλάσιες σε αξία από αυτές που έγιναν "αεροπορικώς". Να μετατρέψετε το κυκλικό διάγραμμα σε ραβδόγραμμα σχετικών συχνοτήτων.

2. α) Να συμπληρωθεί ο παρακάτω πίνακας:

| Ήπειρος | Έκταση | $f_i$ % |
|---|---|---|
| Αμερική | 20,8 | |
| Ασία | 44 | |
| Αφρική | 30,5 | |
| Ευρώπη | 10,5 | |
| Ωκεανία | 9 | |
| | 114,8 | |

β) Να σχεδιάσετε το κυκλικό διάγραμμα.

3. Ο παρακάτω πίνακας παρουσιάζει τις εξαγωγές της χώρας μας κατά το 1977, ανάλογα με το μέσο μεταφοράς.

| Μέσο μεταφοράς | *Θαλασσίως* | *Σιδηροδρομικώς* | *Οδικώς* | *Αεροπορικώς* |
|---|---|---|---|---|
| Αξία σε εκατ. δρχ. | 51.000 | 11.000 | 33.000 | 7.000 |

Να κάνετε το αντίστοιχο κυκλικό διάγραμμα.

4. Το βάρος ενός ζώου κατά τους πρώτους 10 μήνες της ζωής του φαίνεται στον πίνακα:

| Μήνες | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Βάρος σε κιλά | 2 | 3 | 4,5 | 5,3 | 6 | 7 | 9 | 10,5 | 13 | 15 | 19 |

Να γράψετε το χρονόγραμμα της εξέλιξης του βάρους του.

# ΟΜΑΔΟΠΟΙΗΣΗ ΤΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ

Οι πίνακες συχνοτήτων και κατ' αναλογία τα αντίστοιχα διαγράμματα είναι δύσκολο να κατασκευαστούν, όταν το πλήθος των τιμών μιας μεταβλητής είναι αρκετά μεγάλο. Αυτό μπορεί να συμβεί είτε στην περίπτωση μιας διακριτής μεταβλητής είτε, πολύ περισσότερο, στην περίπτωση μιας συνεχούς μεταβλητής, όπου αυτή μπορεί να πάρει οποιαδήποτε τιμή στο διάστημα ορισμού της. Σ' αυτές τις περιπτώσεις είναι απαραίτητο να ταξινομηθούν (ομαδοποιηθούν) τα δεδομένα σε μικρό πλήθος ομάδων, που ονομάζονται και **κλάσεις**, έτσι ώστε κάθε τιμή να ανήκει μόνο σε μία κλάση. Τα άκρα των κλάσεων καλούνται **όρια των κλάσεων**.

Συνήθως υιοθετούμε την περίπτωση που μια κλάση περιέχει το κάτω άκρο της (κλειστή αριστερά) αλλά όχι το άνω άκρο της (ανοικτή δεξιά), δηλαδή που οι κλάσεις είναι της μορφής $[\alpha_{i-1}, \alpha_i)$. Οι παρατηρήσεις κάθε κλάσης θεωρούνται όμοιες, οπότε μπορούν να "αντιπροσωπευθούν" από τις **κεντρικές τιμές** $x_i$, τα κέντρα δηλαδή κάθε κλάσης. Η κεντρική τιμή κάθε κλάσης υπολογίζεται από τη σχέση $x_i=(\alpha_{i-1} + \alpha_i)/2$.

Οι κεντρικές τιμές των κλάσεων μιας συνεχούς μεταβλητής έχουν την έννοια και τη χρήση που έχουν οι διακριτές τιμές μιας μεταβλητής.

- Το πρώτο βήμα στην ομαδοποίηση των δεδομένων είναι η εκλογή του αριθμού *κ* των ομάδων ή κλάσεων. Ο αριθμός αυτός συνήθως ορίζεται αυθαίρετα από τον ερευνητή σύμφωνα με την πείρα του. Γενικά όμως μπορεί να χρησιμοποιηθεί ως οδηγός ο παρακάτω πίνακας:
-

| Μέγεθος δείγματος $\nu$ | Αριθμός κλάσεων $\kappa$ | Μέγεθος δείγματος $\nu$ | Αριθμός κλάσεων $\kappa$ |
|---|---|---|---|
| $< 20$ | 5 | $200-400$ | 9 |
| $20-50$ | 6 | $400-700$ | 10 |
| $50-100$ | 7 | $700-1000$ | 11 |
| $100-200$ | 8 | $\geq 1000$ | 12 |

- Το δεύτερο βήμα είναι ο προσδιορισμός του πλάτους των κλάσεων. **Πλάτος μιας κλάσης** ονομάζεται η διαφορά του κατωτέρου από το ανώτερο όριο της κλάσης. Στην πλειονότητα των πρακτικών εφαρμογών οι κλάσεις έχουν το ίδιο πλάτος. Φυσικά υπάρχουν και περιπτώσεις όπου επιβάλλεται οι κλάσεις να έχουν άνισο πλάτος, όπως, για παράδειγμα, στις κατανομές εισοδήματος, ημερών απεργίας κτλ. Για να κατασκευάσουμε ισοπλατείς κλάσεις, χρησιμοποιούμε το **εύρος** *R* του δείγματος, δηλαδή τη διαφορά της μικρότερης παρατήρησης από τη μεγαλύτερη παρατήρηση του συνολικού δείγματος. Τότε υπολογίζουμε το **πλάτος** *c* των κλάσεων διαιρώντας το εύρος *R* δια του αριθμού των κλάσεων *κ*, στρογγυλεύοντας, αν χρειαστεί για λόγους διευκόλυνσης, πάντα προς τα πάνω.

- Το επόμενο βήμα είναι η κατασκευή των κλάσεων. Ξεκινώντας από την μικρότερη παρατήρηση, ή για πρακτικούς λόγους λίγο πιο κάτω από την μικρότερη παρατήρηση, και προσθέτοντας κάθε φορά το πλάτος *c* δημιουργούμε τις *κ* κλάσεις. Αυτονόητο είναι ότι η μεγαλύτερη τιμή του δείγματος θα (πρέπει να) ανήκει οπωσδήποτε στην τελευταία κλάση.

- Τέλος, γίνεται η **διαλογή** των παρατηρήσεων. Το πλήθος των παρατηρήσεων $\nu_i$ που προκύπτουν από τη διαλογή για την κλάση *i* καλείται **συχνότητα της κλάσης** αυτής ή **συχνότητα της κεντρικής τιμής** $x_i$, $i = 1,2,...,\kappa$.

## ΕΦΑΡΜΟΓΗ

Έστω ότι εξετάζουμε το ύψος των 40 μαθητών μιας τάξης ενός σχολείου. Το ύψος όπως έχει καταγραφεί με τη σειρά, δίνεται στον παρακάτω πίνακα.

(Σε αγκύλες έχουμε τη μικρότερη και τη μεγαλύτερη τιμή.)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 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 |

Παρατηρούμε ότι το εύρος του δείγματος είναι $R = 191 - 156 = 35$. Επειδή έχουμε $\nu = 40$ παρατηρήσεις, χρησιμοποιούμε $\kappa = 6$ κλάσεις. Το πλάτος των κλάσεων είναι $c = R/\kappa = 35/6 = 5{,}83 \approx 6$. Αν θεωρήσουμε ως αρχή της πρώτης κλάσης το 156, θα έχουμε τον επόμενο πίνακα 9.

Πρέπει να προσεχτεί ότι:

- Καμία παρατήρηση δεν μπορεί να μείνει έξω από κάποια κλάση.
- Οι κεντρικές τιμές διαφέρουν μεταξύ τους όσο και το πλάτος των κλάσεων, που εδώ είναι ίσο με 6.
- Μία παρατήρηση που συμπίπτει με το άνω άκρο μιας κλάσης θα τοποθετηθεί κατά τη διαλογή στην αμέσως επόμενη κλάση. Για παράδειγμα, ο μαθητής με ύψος 180 θα τοποθετηθεί στην πέμπτη κλάση [180, 186).

Κατανομές συχνοτήτων (απόλυτων, σχετικών, αθροιστικών) για τα δεδομένα του προηγούμενου πίνακα .

<table>
<tr><th>Κλάσεις<br>[ - )</th><th>Κεντρικές<br>τιμές<br>$x_i$</th><th>Διαλογή</th><th>Συχν.<br>$\nu_i$</th><th>Σχετική<br>Συχνότητα<br>$f_i\%$</th><th>Αθρ.συχν.<br>$N_i$</th><th>Αθρ. Σχετ.<br>Συχν.<br>$F_i\%$</th></tr>
<tr><td>156-162</td><td>159</td><td>||</td><td>2</td><td>5,0</td><td>2</td><td>5,0</td></tr>
<tr><td>162-168</td><td>165</td><td>||||||||</td><td>8</td><td>20,0</td><td>10</td><td>25,0</td></tr>
<tr><td>168-174</td><td>171</td><td>||||||||||||</td><td>12</td><td>30,0</td><td>22</td><td>55,0</td></tr>
<tr><td>174-180</td><td>177</td><td>|||||||||||</td><td>11</td><td>27,5</td><td>33</td><td>82,5</td></tr>
<tr><td>180-186</td><td>183</td><td>|||||</td><td>5</td><td>12,5</td><td>38</td><td>95,0</td></tr>
<tr><td>186-192</td><td>189</td><td>||</td><td>2</td><td>5,0</td><td>40</td><td>100,0</td></tr>
<tr><td></td><td>Σύνολο</td><td>—</td><td>40</td><td>100</td><td>—</td><td>—</td></tr>
</table>

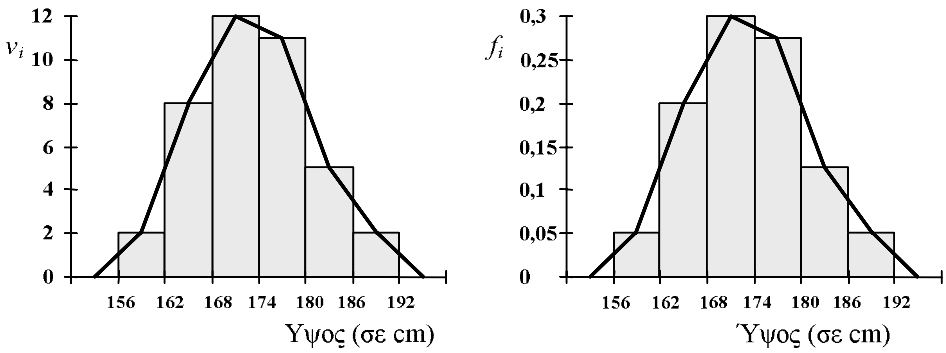
## *Ιστόγραμμα Συχνοτήτων*

Η αντίστοιχη γραφική παράσταση ενός πίνακα συχνοτήτων με ομαδοποιημένα δεδομένα γίνεται με το λεγόμενο **ιστόγραμμα** συχνοτήτων.
Στον οριζόντιο άξονα ενός συστήματος ορθογωνίων αξόνων σημειώνουμε, με κατάλληλη κλίμακα, τα όρια των κλάσεων, ενώ στον κάθετο άξονα τοποθετούμε τις συχνότητες $v_i$.
Στη συνέχεια, κατασκευάζουμε διαδοχικά ορθογώνια (ιστούς), από καθένα από τα οποία έχει βάση ίση με το πλάτος της κλάσης και ύψος ίσο με τη συχνότητα της κλάσης .

Με την ένωση των μέσων των βάσεων των ανωτέρω ορθογωνίων σχηματίζεται μια τεθλασμένη γραμμή που ονομάζεται **πολύγωνο συχνοτήτων**.

## *α) Κλάσεις Ίσου Πλάτους*

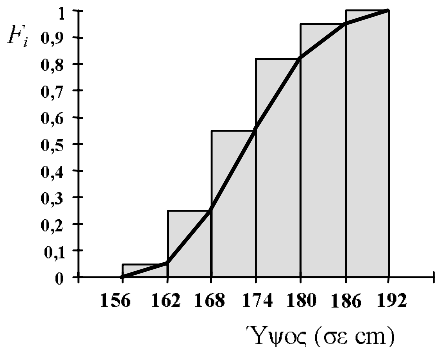
Θεωρώντας το πλάτος ***c*** ως μονάδα μέτρησης του χαρακτηριστικού στον οριζόντιο άξονα, το ύψος κάθε ορθογωνίου είναι ίσο προς τη συχνότητα της αντίστοιχης κλάσης. Αν στα ιστογράμματα συχνοτήτων θεωρήσουμε δύο ακόμη υποθετικές κλάσεις, στην αρχή και στο τέλος, με συχνότητα μηδέν και στη συνέχεια ενώσουμε τα μέσα των άνω βάσεων των ορθογωνίων, σχηματίζεται το λεγόμενο **πολύγωνο συχνοτήτων** .Με ανάλογο τρόπο κατασκευάζεται το **ιστόγραμμα σχετικών συχνοτήτων** και το **πολύγωνο σχετικών συχνοτήτων**, οπότε στον κάθετο άξονα βάζουμε τις σχετικές συχνότητες.




(α) (β)

Ιστόγραμμα και πολύγωνο (α) συχνοτήτων και (β) σχετικών συχνοτήτων για τα δεδομένα του πίνακα των υψών των μαθητών.

Με τον ίδιο τρόπο κατασκευάζονται και τα **ιστογράμματα αθροιστικών συχνοτήτων** και **αθροιστικών σχετικών συχνοτήτων.** Αν ενώσουμε σε ένα ιστόγραμμα αθροιστικών συχνοτήτων τα **δεξιά άκρα** (όχι μέσα) των άνω βάσεων των ορθογωνίων με ευθύγραμμα τμήματα βρίσκουμε το **πολύγωνο αθροιστικών συχνοτήτων** της κατανομής. Στο σχήμα παριστάνεται το ιστόγραμμα και το πολύγωνο αθροιστικών σχετικών συχνοτήτων για το ύψος των μαθητών .




## *β) Κλάσεις Άνισου Πλάτους*

Όπως προαναφέραμε, συνήθως επιλέγουμε κλάσεις ίσου πλάτους. Υπάρχουν όμως και περιπτώσεις που είναι απαραίτητο να έχουμε κλάσεις διαφορετικού πλάτους όπως, για παράδειγμα, στην κατανάλωση νερού και ηλεκτρικού ρεύματος ή ακόμα και περιπτώσεις όπου οι συχνότητες σε κάποιες κλάσεις να είναι πολύ μικρές οπότε γίνεται συγχώνευση κλάσεων. Το αντίστοιχο **ιστόγραμμα συχνοτήτων** κατασκευάζεται πάλι παίρνοντας στον οριζόντιο άξονα τα άνισα πλάτη των κλάσεων όμως στον κάθετο ύψη ίσα με $\upsilon_i = \dfrac{\nu_i}{c_i}$. Με τον τρόπο αυτό το εμβαδόν κάθε ορθογωνίου να ισούται με τη συχνότητα της αντίστοιχης κλάσης.

**<u>ΕΦΑΡΜΟΓΗ</u>**

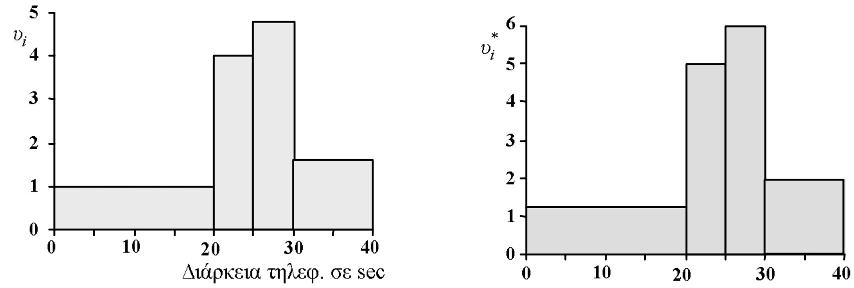
Έστω, για παράδειγμα, η διάρκεια (σε sec) $\nu=80$ τηλεφωνημάτων που έγιναν τυχαία από ένα κινητό τηλέφωνο, η οποία δίνεται στο διπλανό πίνακα συχνοτήτων.

| Διάρκεια τηλεφ. σε sec | Συχνότητα $\nu_i$ |
|---|---|
| 0-20 | 20 |
| 20-25 | 20 |
| 25-30 | 24 |
| 30-40 | 16 |
| Σύνολο | $\nu = 80$ |

Επομένως, για την κατασκευή του ιστογράμματος συχνοτήτων χρειαζόμαστε τα πλάτη των κλάσεων και τα ύψη των ορθογωνίων. Αυτά δίνονται στον επόμενο πίνακα.

| Διάρκεια τηλεφ. σε sec. | Πλάτος κλάσης $c_i$ | Συχνότητα $\nu_i$ | Ύψος $\upsilon_i = \dfrac{\nu_i}{c_i}$ | Ύψος $\upsilon_i^* = \dfrac{f_i\%}{c_i}$ |
|---|---|---|---|---|
| 0-20 | 20 | 20 | 1,0 | 1,25 |
| 20-25 | 5 | 20 | 4,0 | 5,00 |
| 25-30 | 5 | 24 | 4,8 | 6,00 |
| 30-40 | 10 | 16 | 1,6 | 2,00 |

*Με ανάλογο τρόπο κατασκευάζεται και το* ***ιστόγραμμα σχετικών συχνοτήτων****, αρκεί να χρησιμοποιήσουμε ως ύψος των ορθογωνίων το λόγο των σχετικών συχνοτήτων προς το πλάτος των κλάσεων, δηλαδή* $\upsilon_i^* = \dfrac{f_i\%}{c_i}$.

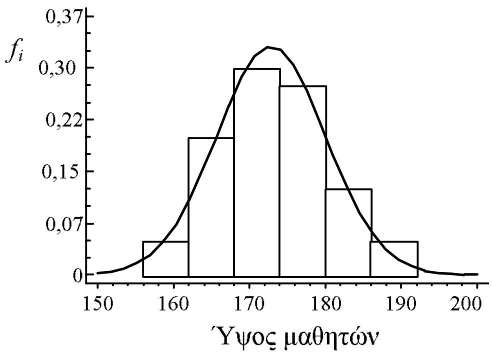



(α) (β)

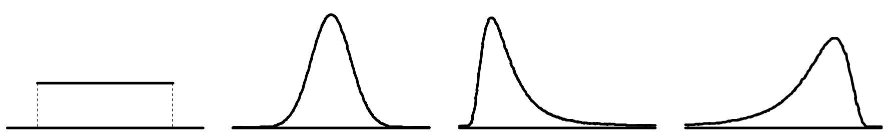
Ιστόγραμμα συχνοτήτων (α) και σχετικών συχνοτήτων (β) της διάρκειας τηλεφωνημάτων.

## *Καμπύλες Συχνοτήτων*

Εάν υποθέσουμε ότι ο αριθμός των κλάσεων για μια συνεχή μεταβλητή είναι αρκετά μεγάλος (τείνει στο άπειρο) και ότι το πλάτος των κλάσεων είναι αρκετά μικρό (τείνει στο μηδέν), τότε η πολυγωνική γραμμή συχνοτήτων τείνει να πάρει τη μορφή μιας ομαλής καμπύλης, η οποία ονομάζεται **καμπύλη συχνοτήτων** , όπως δείχνει το σχήμα . Οι καμπύλες συχνοτήτων έχουν μεγάλη εφαρμογή στη Στατιστική, όπου οι ιδιότητες τους μπορούν να χρησιμοποιηθούν για την εξαγωγή χρήσιμων συμπερασμάτων.




Η μορφή μιας κατανομής συχνοτήτων εξαρτάται από το πώς είναι κατανεμημένες οι παρατηρήσεις σε όλη την έκταση του εύρους τους. Μερικές χαρακτηριστικές καμπύλες συχνοτήτων που συναντάμε συχνά στις εφαρμογές δίνονται στο σχήμα. Η κατανομή (β), με "κωδωνοειδή" μορφή λέγεται **κανονική κατανομή** και παίζει σπουδαίο ρόλο στη Στατιστική. Όταν οι παρατηρήσεις "κατανέμονται" ομοιόμορφα σε ένα διάστημα [$\alpha$, $\beta$], όπως στην κατανομή (α), η κατανομή λέγεται **ομοιόμορφη**. Όταν οι παρατηρήσεις δεν είναι συμμετρικά κατανεμημένες, η κατανομή λέγεται ασύμμετρη με θετική ασυμμετρία όπως στην κατανομή (γ) ή αρνητική ασυμμετρία όπως στην κατανομή (δ).



(α) (β) (γ) (δ)

Μερικές χαρακτηριστικές κατανομές συχνοτήτων

# ΕΡΩΤΗΣΕΙΣ – ΑΣΚΗΣΕΙΣ

**Ερωτήσεις του τύπου «Σωστό - Λάθος»**

1. Όταν το πλήθος των τιμών μιας μεταβλητής είναι αρκετά μεγάλο είναι απαραίτητο να ταξινομηθούν τα δεδομένα σε κλάσεις. **Σ** **Λ**
2. Πλάτος κλάσης ενός δείγματος ονομάζεται το άθροισμα του κατώτερου και του ανώτερου ορίου της κλάσης. **Σ** **Λ**
3. Όταν ο αριθμός των κλάσεων για μια συνεχή μεταβλητή είναι αρκετά μικρός και το πλάτος των κλάσεων είναι αρκετά μεγάλο τότε η πολυγωνική γραμμή συχνοτήτων τείνει να πάρει τη μορφή μιας ομαλής καμπύλης, η οποία ονομάζεται καμπύλη συχνοτήτων. **Σ** **Λ**
4. Σε όλες τις περιπτώσεις οι κλάσεις ενός δείγματος έχουν όλες το ίδιο πλάτος. **Σ** **Λ**
5. Οι παρατηρήσεις κάθε κλάσης ενός δείγματος μπορούν να αντιπροσωπευθούν από τις κεντρικές τιμές τους. **Σ** **Λ**
6. Το κέντρο κάθε κλάσης ενός δείγματος ισούται με την ημιδιαφορά των άκρων της κλάσης. **Σ** **Λ**
7. Η γραφική παράσταση ενός πίνακα συχνοτήτων μιας κατανομής με ομαδοποιημένα δεδομένα γίνεται με το ιστόγραμμα συχνοτήτων. **Σ** **Λ**

**Ερωτήσεις συμπλήρωσης - απάντησης**

1. Σε μια έρευνα μεταξύ 500 ανέργων για το χρόνο σε μήνες που είναι άνεργοι προέκυψε ο παρακάτω πίνακας:

| Χρόνος ανεργίας | $\nu_i$ | $f_i$ % | $F_i$ % |
|---|---|---|---|
| [0, 3) | | 19 | |
| [3, 6) | | 38,6 | |
| [6, 12) | | 24,4 | |
| [12, 24) | | 13,6 | |
| [24, 36) | | 4,4 | |
| | | 100 | |

α) Να συμπληρώσετε τον πίνακα.

β) Να κατασκευάσετε πολύγωνο σχετικών αθροιστικών συχνοτήτων.

2. Ο παρακάτω πίνακας παρουσιάζει τη διάρκεια ζωής 400 οθονών τηλεόρασης από την παραγωγή ενός εργοστασίου.

α) Να συμπληρώσετε τον πίνακα:

| Διάρκεια ζωής σε ώρες λειτουργίας | $\nu_i$ | $f_i$ % | $N_i$ | $F_i$ % |
|---|---|---|---|---|
| [400, 500) | 15 | | | |
| [500, 600) | 45 | | | |
| [600, 700) | 60 | | | |
| [700, 800) | 75 | | | |
| [800, 900) | 70 | | | |
| [900, 1000) | 60 | | | |
| [1000, 1100) | 50 | | | |
| [1100, 1200) | 25 | | | |
| | 400 | | | |

β) Να κατασκευάσετε:

i) Το ιστόγραμμα συχνοτήτων.

ii) Το ιστόγραμμα σχετικών συχνοτήτων.

iii) Το ιστόγραμμα σχετικών αθροιστικών συχνοτήτων.

**3.** Σ' ένα τεστ πήραν μέρος 100 μαθητές προκειμένου ο καθένας να απαντήσει σε 200 ερωτήσεις. Η βαθμολογία είναι 1 ή 0, ανάλογα αν ο μαθητής απαντάει ή όχι στην ερώτηση. Ο επόμενος πίνακας δείχνει τα αποτελέσματα της βαθμολογίας.

| Βαθμοί | Συχνότητα |
|---|---|
| [60, 80) | 5 |
| [80, 100) | 20 |
| [100, 120) | 26 |
| [120, 140) | 30 |
| [140, 160) | 15 |
| [160, 180) | 4 |
| | 100 |

Να κατασκευάσετε: i) Το ιστόγραμμα. ii) Το πολύγωνο των συχνοτήτων.

**5.** Οι χρόνοι (σε λεπτά) που χρειάστηκαν 55 μαθητές να λύσουν ένα πρόβλημα δίνονται παρακάτω:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 3,4 | 13,2 | 6,7 | 1,4 | 1,3 | 3,8 | 3,9 | 2,9 | 13,8 | 3,9 | 2,7 |
| 4,4 | 3,6 | 1,4 | 2,4 | 3,6 | 3,1 | 7,5 | 6,9 | 7,8 | 12,7 | 3,9 |
| 3,3 | 9,7 | 2,0 | 4,4 | 3,3 | 8,7 | 3,9 | 11,6 | 5,6 | 9,0 | 3,4 |
| 1,4 | 3,5 | 2,8 | 10,4 | 11,9 | 12,3 | 2,9 | 2,8 | 1,5 | 4,1 | 5,9 |
| 3,1 | 8,7 | 2,8 | 3,8 | 13,0 | 3,0 | 6,4 | 3,2 | 5,9 | 7,0 | 8,2 |

α) Να ομαδοποιήσετε τα δεδομένα σε κατάλληλο αριθμό κλάσεων.

β) Να κατασκευάσετε τον πίνακα με τις συχνότητες $v_i$, $f_i\%$, $N_i$, $F_i\%$.

γ) Να κατασκευάσετε το πολύγωνο σχετικών συχνοτήτων και αθροιστικών σχετικών συχνοτήτων.

# ΜΕΤΡΑ ΘΕΣΗΣ ΚΑΙ ΔΙΑΣΠΟΡΑΣ

Σκοπός της εμφάνισης των στατιστικών δεδομένων με τη μορφή στατιστικών πινάκων συχνοτήτων είναι ο περιορισμός του όγκου των στοιχείων που συγκεντρώθηκαν και η εύκολη μελέτη και περιγραφή του πληθυσμού που ερευνούμε. Και με τη μορφή όμως αυτή των κατανομών συχνοτήτων, τα αρχικά στατιστικά στοιχεία παρουσιάζουν μια δυσκολία γιατί οι πίνακες αυτοί δεν είναι εύκολο να συγκρατηθούν στη μνήμη ενός μελετητή. Για το λόγο αυτό θεωρείται αναγκαία η παραπέρα συμπύκνωση των πινάκων συχνοτήτων.
Η συμπύκνωση αυτή γίνεται με την αντικατάσταση των πινάκων από ορισμένους αντιπροσωπευτικούς αριθμούς, οι οποίοι ονομάζονται *στατιστικές παράμετροι*, χαρακτηρίζουν τη *θέση*, τη *διασπορά* και τη *μορφολογία* του πληθυσμού που ερευνούμε και διευκολύνουν πάρα πολύ στις συγκρίσεις ομοειδών ομάδων.

Θα ορίσουμε δηλαδή κάποια **μέτρα** (αριθμητικά μεγέθη), που να μας δίνουν α) τη θέση του "κέντρου" των παρατηρήσεων στον οριζόντιο άξονα και β) τη διασπορά των παρατηρήσεων, δηλαδή πόσο αυτές εκτείνονται γύρω από το "κέντρο" τους. Τα πρώτα τα καλούμε **μέτρα θέσης** της κατανομής , ενώ τα δεύτερα **μέτρα διασποράς** ή **μέτρα μεταβλητότητας** .

# ΜΕΤΡΑ ΘΕΣΗΣ

## Α) Μέτρα κεντρικής τάσης

οι παράμετροι αυτές ορίζουν το σημείο γύρω από το οποίο συσσωρεύονται οι διάφορες τιμές ενός πληθυσμού και βρίσκονται στο κέντρο μιας σειράς παρατηρήσεων, εφόσον ο πληθυσμός παρουσιάζει ομοιογένεια. Αντικειμενικός σκοπός αυτών των παραμέτρων είναι να αντιπροσωπεύουν ένα πληθυσμό από την άποψη μιας μεταβλητής κατά τον πιο απλό και σύντομο τρόπο.

**Ι) ΑΡΙΘΜΗΤΙΚΟΣ ΜΕΣΟΣ**

**α) αστάθμητος αριθμητικός μέσος** $(\overline{x})$

Σε μια κατανομή συχνοτήτων, αν $x_1, x_2, \ldots, x_\kappa$ είναι οι τιμές της μεταβλητής *X*, ο αριθμητικός μέσος (ή μέση τιμή) ορίζεται από τη σχέση:

$$\boxed{\overline{x} = \frac{x_1 + x_2 + \ldots + x_\nu}{\nu} = \frac{\sum_{i=1}^{\nu} x_i}{\nu} = \frac{1}{\nu}\sum_{i=1}^{\nu} xi}$$

**β) σταθμικός αριθμητικός μέσος**

Σε μια **ασυνεχή** κατανομή συχνοτήτων, αν $x_1, x_2, \ldots, x_\kappa$ είναι οι τιμές της μεταβλητής *X* με συχνότητες $\nu_1, \nu_2, \ldots, \nu_\kappa$ αντίστοιχα, ο αριθμητικός μέσος (ή μέση τιμή) ορίζεται από τη σχέση:

$$\bar{x}=\frac{x_1v_1+x_2v_2+\ldots+x_\kappa v_\kappa}{v_1+v_2+\ldots+v_\kappa}=\frac{\sum_{i=1}^{\kappa}x_iv_i}{\sum_{i=1}^{\kappa}v_i}=\frac{1}{v}\sum_{i=1}^{\kappa}x_iv_i$$

Η οποία ισοδύναμα γράφεται: $\boxed{\bar{x}=\sum_{i=1}^{\kappa}x_i\frac{v_i}{v}=\sum_{i=1}^{\kappa}x_if_i}$, $f_i$ οι σχετικές συχνότητες.

Στην περίπτωση που η κατανομή είναι **συνεχής** και έχω κλάσεις μεταβλητών, για τον υπολογισμό της μέσης τιμής βρίσκω τις κεντρικές τιμές όλων των κλάσεων και στη συνέχεια τις πολλαπλασιάζω τις κεντρικές τιμές με τις αντίστοιχες συχνότητες κάθε κλάσης, προσθέτουμε τα γινόμενα και διαιρούμε το άθροισμα τους με το άθροισμα των συχνοτήτων τους.

Ο ίδιος τύπος χρησιμοποιείται και στις περιπτώσεις που δίνεται διαφορετική βαρύτητα (έμφαση) στις τιμές $x_1, x_2, \ldots, x_v$ ενός συνόλου δεδομένων. Εάν σε κάθε τιμή $x_1, x_2, \ldots, x_v$ δώσουμε διαφορετική βαρύτητα, που εκφράζεται με τους λεγόμενους συντελεστές στάθμισης (βαρύτητας) $w_1, w_2, \ldots, w_v$, τότε ο σταθμικός μέσος βρίσκεται από τον τύπο:

$$\bar{x}=\frac{x_1w_1+x_2w_2+\ldots+x_vw_v}{w_1+w_2+\ldots+w_v}=\frac{\sum_{i=1}^{v}x_iw_i}{\sum_{i=1}^{v}w_i}$$

## ΙΔΙΟΤΗΤΕΣ ΤΟΥ ΜΕΣΟΥ ΑΡΙΘΜΗΤΙΚΟΥ

1. Αν προσθέσουμε η αφαιρέσουμε στις τιμές μιας μεταβλητής μια σταθερή ποσότητα $x_0$ τότε ο μέσος αριθμητικός των αρχικών τιμών αυξάνεται ή ελαττώνεται κατά $x_0$ αντίστοιχα.

2. Αν πολλαπλασιάσουμε τις στις τιμές μιας μεταβλητής με μια σταθερή ποσότητα $x_0$ τότε ο μέσος αριθμητικός των αρχικών τιμών πολλαπλασιάζεται με $x_0$ και αυτός.

3. Αν α, β είναι δυο πραγματικοί αριθμοί και $\bar{x}$ η μέση αριθμητική τιμή της μεταβλητής x, τότε η μέση τιμή της μεταβλητής y =α+βx θα είναι $\bar{y}$=α+β$\bar{x}$

4. Αν οι τιμές μιας μεταβλητής Χ είναι ίσες μεταξύ τους, τότε η μέση τιμή των τιμών της είναι μια από αυτές τις τιμές.

## II) ΓΕΩΜΕΤΡΙΚΟΣ ΜΕΣΟΣ

### α) αστάθμητος γεωμετρικός μέσος

Αν δοθεί η σειρά των παρατηρήσεων $x_1, x_2, \ldots, x_\kappa$ τότε ο μέσος γεωμετρικός ορίζεται ως η νιοστή ρίζα του γινομένου αυτών. Δηλαδή: $\boxed{G=\sqrt[\kappa]{x_1+x_2+x_3+\ldots+x_k}}$

*Ο υπολογισμός του μέσου γεωμετρικού γίνεται με λογαρίθμιση της προηγούμενης παράστασης δηλαδή :*

$$log\ G=\frac{\log x_1+\log x_2+\log x_3+\ldots+\log x_k}{k},$$ *οπότε με αντιλογαρίθμηση προκύπτει ο* $G$

### β) σταθμικός γεωμετρικός μέσος

Αν δοθεί η σειρά των παρατηρήσεων $x_1, x_2, \ldots, x_\kappa$ με συχνότητες αντίστοιχα $\nu_1, \nu_2, \nu_3, \ldots, \nu_\kappa$ τότε ο μέσος γεωμετρικός ορίζεται από τον τύπο :

$$\boxed{G=\sqrt[\sum \nu_i]{x_1^{\nu_1}+x_2^{\nu_2}+x_3^{\nu_3}+\ldots+x_\kappa^{\nu_\kappa}}}\ ,\ i=1,2,3,\ldots,\kappa$$

*Ο παραπάνω τύπος με τη βοήθεια των λογαρίθμων γράφεται:*

$$log\ G=\frac{\nu_1\log x_1+\nu_2\log x_2+\nu_3\log x_3+\ldots+\nu_\kappa\log x_\kappa}{\sum \nu_i}\ ,\ i=1,2,3,\ldots,\kappa$$

| |
|---|
| Ο κυριότερος σκοπός χρήσης του μέσου γεωμετρικού είναι για την εξακρίβωση της μέσης ετήσιας μεταβολής μιας μεταβλητής . |

**ΕΦΑΡΜΟΓΗ**

Αν η τιμή ενός αγαθού Α σε μια συγκεκριμένη χρονική στιγμή είναι 10δρχ και του αγαθού Β είναι 40 και μετά από παρέλευση χρόνου η τιμή του Α είναι 20 και του Β είναι 20, να βρεθεί αν έχουμε μεταβολή της μέσης τιμής .

Υπολογίζουμε τη μέση τιμή με το μέσο γεωμετρικό οπότε έχουμε:

$G_1=\sqrt{10\cdot 40}=20$ και $G_2=\sqrt{20\cdot 20}=20$. Επομένως δεν έχουμε καμία μεταβολή.

## III) ΑΡΜΟΝΙΚΟΣ ΜΕΣΟΣ

### α) αστάθμητος αρμονικός μέσος

ο μέσος αστάθμητος αρμονικός μιας σειράς παρατηρήσεων $x_1, x_2, \ldots, x_\kappa$ ορίζεται από τη σχέση :

$$\boxed{H=\frac{\kappa}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\cdots+\frac{1}{x_\kappa}}}\ ,\ i=1,2,3,\ldots,\kappa$$

**β) σταθμικός αρμονικός μέσος**

ο μέσος σταθμικός αρμονικός μιας σειράς παρατηρήσεων $x_1, x_2, \ldots, x_\kappa$ με συχνότητες αντίστοιχα $\nu_1, \nu_2, \nu_3, \ldots, \nu_\kappa$ ορίζεται από τη σχέση :

$$\boxed{H = \frac{\sum \nu_i}{\frac{\nu_1}{x_1} + \frac{\nu_2}{x_2} + \frac{\nu_3}{x_3} + \cdots + \frac{\nu_\kappa}{x_\kappa}}}, \; i = 1, 2, 3, \ldots, \kappa$$

Ο μέσος αρμονικός χρησιμοποιείται κυρίως στα προβλήματα κίνησης, όταν οι παρατηρήσεις αναφέρονται σε λόγους δυο ποσοτήτων και στη μέτρηση της μεταβολής των τιμών τους.

Μεταξύ του μέσου αριθμητικού, μέσου γεωμετρικού και μέσου αρμονικού ισχύουν οι σχέσεις: $\bar{x} > G$ και $H > G$

## Β) Παράμετροι κεντρικής θέσης

### Ι) Διάμεσος (δ)

Οι χρόνοι (σε λεπτά) που χρειάστηκαν 9 μαθητές, για να λύσουν ένα πρόβλημα είναι: 3, 5, 5, 36, 6, 7, 4, 7, 8 με μέση τιμή $\bar{x} = 9$. Παρατηρούμε όμως ότι οι οκτώ από τις εννέα παρατηρήσεις είναι μικρότερες του 9 και μία (ακραία τιμή), η οποία επηρεάζει και τη μέση τιμή είναι, αρκετά μεγαλύτερη του 9. Αυτό σημαίνει ότι η μέση τιμή δεν ενδείκνυται ως μέτρο θέσης ("κέντρο") των παρατηρήσεων αυτών. Αντίθετα, ένα άλλο μέτρο θέσης που δεν επηρεάζεται από ακραίες παρατηρήσεις είναι η **διάμεσος** , η οποία ορίζεται ως εξής:

**Διάμεσος** (δ) ενός δείγματος $\nu$ παρατηρήσεων οι οποίες έχουν διαταχθεί σε αύξουσα σειρά ορίζεται ως η μεσαία παρατήρηση, όταν το $\nu$ είναι περιττός αριθμός, ή ο μέσος όρος (ημιάθροισμα) των δύο μεσαίων παρατηρήσεων όταν το $\nu$ είναι άρτιος αριθμός.

Για παράδειγμα, για να βρούμε τη διάμεσο των δεδομένων:

α) 3, 4, 0, 6, 5, 8, 1, 1, 6, 1, 2, 8, 9

β) 3, 4, 0, 6, 5, 8, 1, 1, 6, 1, 2, 8, 9, 9

εργαζόμαστε ως εξής:

α) Έχουμε $\nu = 13$ παρατηρήσεις, οι οποίες σε αύξουσα σειρά είναι:

0 1 1 1 2 3 4 5 6 6 8 8 9.

Άρα, η διάμεσος είναι η μεσαία παρατήρηση (έβδομη στη σειρά), $\delta = 4$.

β) Έχουμε $\nu = 14$ παρατηρήσεις οι οποίες σε αύξουσα σειρά είναι:

0 1 1 1 2 3 4 5 6 6 8 8 9 9.

Άρα, η διάμεσος είναι το ημιάθροισμα των δύο μεσαίων παρατηρήσεων (της έβδομης και όγδοης στη σειρά), δηλαδή $\delta = \frac{4+5}{2} = 4{,}5$.

Παρατηρούμε ότι, η διάμεσος είναι η τιμή που χωρίζει ένα σύνολο παρατηρήσεων σε δύο ίσα μέρη όταν οι παρατηρήσεις αυτές τοποθετηθούν με σειρά τάξης μεγέθους. Ακριβέστερα, η διάμεσος είναι η τιμή για την οποία το πολύ 50% των παρατηρήσεων είναι μικρότερες από αυτήν και το πολύ 50% των παρατηρήσεων είναι μεγαλύτερες από την τιμή αυτήν.

## ΙΙ) Εκατοστημόρια ($P_\kappa$)

Όπως ορίσαμε τη διάμεσο *δ*, έτσι ώστε το πολύ 50% των παρατηρήσεων να είναι μικρότερες του *δ* και το πολύ 50% των παρατηρήσεων να είναι μεγαλύτερες του *δ*, μπορούμε ανάλογα να ορίσουμε και τα **εκατοστημόρια** $P_\kappa$, $\kappa = 1,2,\ldots,99$. Οι τιμές $P_1, P_2, \ldots, P_{99}$ χωρίζουν το σύνολο των παρατηρήσεων σε 100 ίσα μέρη. Επομένως, αναλόγως και με τον ορισμό της διαμέσου, ορίζουμε ως *κ*-εκατοστιαίο σημείο ή $P_\kappa$ εκατοστημόριο ενός συνόλου παρατηρήσεων την τιμή εκείνη για την οποία το πολύ *κ*% των παρατηρήσεων είναι μικρότερες του $P_\kappa$ και το πολύ (100-*κ*)% των παρατηρήσεων είναι μεγαλύτερες από την τιμή αυτήν.

### ΤΕΤΑΡΤΗΜΟΡΙΑ

Ειδική περίπτωση εκατοστημορίων είναι τα $P_{25}$, $P_{50}$, $P_{75}$, τα οποία καλούνται **τεταρτημόρια** και συμβολίζονται με $Q_1$, $Q_2$ και $Q_3$, αντίστοιχα. Για το $Q_1$ έχουμε αριστερά το πολύ 25% των παρατηρήσεων και δεξιά το πολύ 75%. Όμοια για το $Q_3$ έχουμε αριστερά το πολύ 75% των παρατηρήσεων και δεξιά το πολύ 25% των παρατηρήσεων. Προφανώς το $Q_2 = P_{50}$ συμπίπτει και με τη διάμεσο, δηλαδή $Q_2 = \delta$. Τα μέτρα αυτά χρησιμοποιούνται αρκετά συχνά για τη μελέτη ενός συνόλου δεδομένων.

Συχνά για ευκολία ο υπολογισμός των τεταρτημορίων $Q_1$ και $Q_3$ ενός συνόλου δεδομένων γίνεται κατά προσέγγιση υπολογίζοντας τις διάμεσους του πρώτου και του δεύτερου μισού των διατεταγμένων παρατηρήσεων, αντίστοιχα. Για παράδειγμα, προκειμένου να υπολογίσουμε τα τεταρτημόρια των δεδομένων 3, 4, 0, 6, 5, 8, 1, 1, 6, 1, 2, 8, 9, εργαζόμαστε ως εξής:

- Διατάσσουμε τις παρατηρήσεις σε αύξουσα σειρά μεγέθους:

Έχουμε $\nu = 13$ παρατηρήσεις, οι οποίες σε αύξουσα σειρά είναι:

0 1 1 1 2 3 4 5 6 6 8 8 9.

- Υπολογίζουμε τη διάμεσο, όπως προαναφέραμε:

Η διάμεσος είναι η έβδομη στη σειρά παρατήρηση, δηλαδή $\delta = 4$.

- Υπολογίζουμε τη διάμεσο του πρώτου μισού των διατεταγμένων παρατηρήσεων, δηλαδή των παρατηρήσεων που είναι αριστερά του *δ*. Η τιμή αυτή είναι το $Q_1$:

Η διάμεσος των παρατηρήσεων που είναι αριστερά του *δ*, δηλαδή των 0, 1, 1, 1, 2, 3, είναι το $Q_1 = \frac{1+1}{2} = 1$.

- Υπολογίζουμε τη διάμεσο του δεύτερου μισού των διατεταγμένων παρατηρήσεων, δηλαδή των παρατηρήσεων που είναι δεξιά του *δ*. Η τιμή αυτή είναι το $Q_3$.

Η διάμεσος των παρατηρήσεων που είναι δεξιά του *δ*, δηλαδή των 5, 6, 6, 8, 8, 9, είναι το $Q_3 = \frac{6+8}{2} = 7$.



**ΔΕΚΑΤΗΜΟΡΙΑ**

Οι τιμές $P_{10}, P_{20}, P_{30}, \dots, P_{90}$ χωρίζουν τις παρατηρήσεις σε 10 ίσα μέρη και ονομάζονται **δεκατημόρια**.

Η τιμή $P_{10}$ χωρίζει το σύνολο των παρατηρήσεων ώστε το 10% να βρίσκεται κάτω από αυτήν ενώ το 10% να βρίσκεται πάνω από αυτήν. Αντίστοιχα και τα υπόλοιπα δεκατημόρια χωρίζουν το σύνολο των παρατηρήσεων.

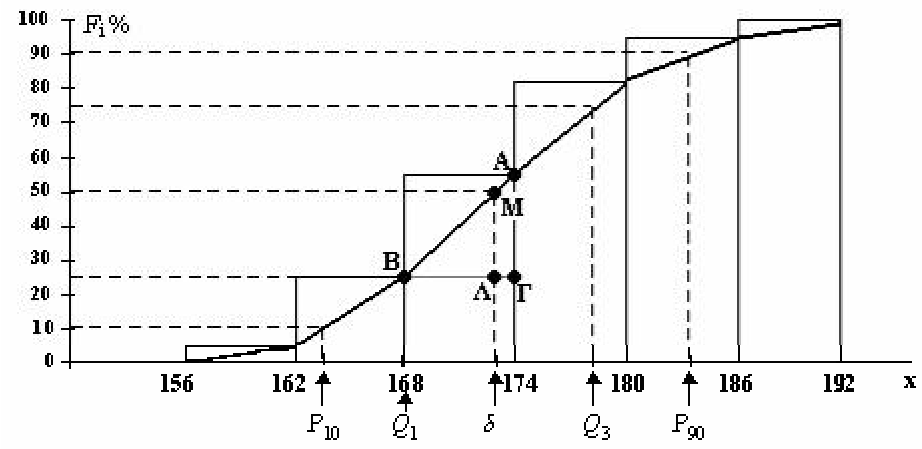
Για τον υπολογισμό των δεκατημορίων προσδιορίζουμε την τιμή $\frac{(\kappa+1)\mathrm{i}}{10}$

όπου κ είναι το σύνολο των παρατηρήσεων και i το δεκατημόριο που θέλουμε.

## Γραφική μέθοδος εύρεσης της διαμέσου και των τεταρτημορίων

Η διάμεσος, όπως ορίστηκε, αντιστοιχεί στην τιμή $x=\delta$ της μεταβλητής $X$ (στον οριζόντιο άξονα), έτσι ώστε το 50% των παρατηρήσεων να είναι μικρότερες ή ίσες του $\delta$. Δηλαδή, η διάμεσος θα έχει αθροιστική σχετική συχνότητα $F_i=50\%$.

Σε ένα πολύγωνο αθροιστικών σχετικών συχνοτήτων, εφόσον στον κάθετο άξονα έχουμε τις αθροιστικές σχετικές συχνότητες, από το σημείο Α (50% των παρατηρήσεων) φέρουμε την $AB // 0x$ και στη συνέχεια τη $B\Gamma \perp 0x$. Τότε, στο σημείο $\Gamma$ αντιστοιχεί η διάμεσος $\delta$ των παρατηρήσεων.




Για να υπολογίσουμε ακριβώς τη διάμεσο πρέπει να υπολογίσουμε το ευθ. τμήμα ΒΛ. Αν $c_i$ είναι το πλάτος της κλάσης στην οποία ανήκει η διάμεσος, $L_i$ το κατώτερο όριο με σχετική αθροιστική συχνότητα $F_{i-1}$ (επί τοις εκατό) και $f_i$ η σχετική συχνότητα (επί τοις εκατό) της κλάσης αυτής τότε από τα όμοια τρίγωνα (έχουν τις γωνίες τους ίσες) **ΒΛΜ** και **ΒΑΓ** έχουμε ότι :

(ιδιότητες αναλογιών) $\dfrac{\mathrm{B\Lambda}}{\mathrm{M\Lambda}}=\dfrac{\mathrm{B\Gamma}}{\mathrm{\Gamma A}} \Leftrightarrow \dfrac{x}{50-F_{i-1}}=\dfrac{c_i}{f_i} \Leftrightarrow x=(50-F_{i-1})\dfrac{c_i}{f_i}$ ,

Άρα η **διάμεσος** είναι : $\boxed{\delta = L_i+(50-F_{i-1})\dfrac{c_i}{f_i}}$

Με ανάλογο τρόπο βρίσκουμε τα σημεία του οριζόντιου άξονα που εκφράζουν τα τεταρτημόρια $Q_1$, $Q_3$ ( $F_i$=25% , $F_i$=75%) και τα δεκατημόρια $P_i$, i=1,…,9 ( $F_i$=10% , 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%))

## Γ) Επικρατούσα τιμή

Στην περίπτωση μη ομαδοποιημένων δεδομένων **επικρατούσα τιμή** ή **κορυφή** $M_0$ ορίζεται ως η παρατήρηση με τη μεγαλύτερη συχνότητα. Είναι προφανές ότι η επικρατούσα τιμή μπορεί να οριστεί και στην περίπτωση ποιοτικών δεδομένων, ενώ τα άλλα μέτρα που είδαμε ορίζονται μόνο για ποσοτικά δεδομένα.

**ΕΦΑΡΜΟΓΗ**

Για να βρούμε την επικρατούσα τιμή των παρατηρήσεων 0 1 1 2 2 2 3 4 4 4 5 5 7 8, κατασκευάζουμε το διπλανό πίνακα συχνοτήτων. Οι τιμές 2 και 4 είναι και οι δύο επικρατούσες τιμές, γιατί καθεμιά έχει συχνότητα 3. Βλέπουμε εδώ ότι η επικρατούσα τιμή μπορεί να μην είναι μοναδική.

| $x_i$ | $\nu_i$ |
|---|---|
| 0 | 1 |
| 1 | 2 |
| 2 | 3 |
| 3 | 1 |
| 4 | 3 |
| 5 | 2 |
| 7 | 1 |
| 8 | 1 |

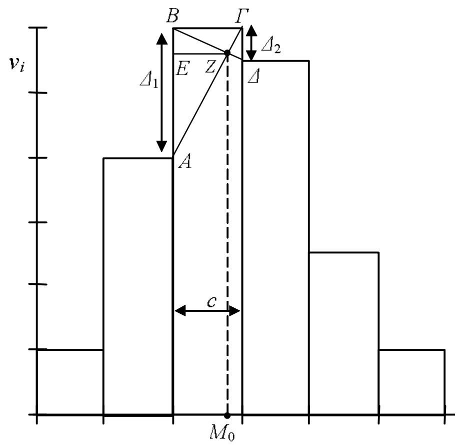
Όταν έχουμε δύο κορυφές, η αντίστοιχη κατανομή συχνοτήτων λέγεται **δικόρυφη**, ενώ όταν έχουμε πολλές κορυφές λέγεται **πολυκόρυφη**.

Όταν όλες οι παρατηρήσεις είναι διαφορετικές, τότε λέμε ότι δεν υπάρχει επικρατούσα τιμή. Έτσι, για τις παρατηρήσεις 0, 1, 2, 7, 8, 9 δεν έχουμε επικρατούσα τιμή.

***Επικρατούσα Τιμή σε Ομαδοποιημένα Δεδομένα***

Όταν έχουμε ομαδοποιημένα (ποσοτικά) δεδομένα σε **ισοπλατείς** κλάσεις, τότε βρίσκουμε πρώτα την **επικρατούσα κλάση** $i$, δηλαδή την κλάση με τη μεγαλύτερη συχνότητα.

Υποθέτοντας, όπως έχουμε ήδη αναφέρει και προηγουμένως, ότι οι παρατηρήσεις στις κλάσεις κατανέμονται ομοιόμορφα, η επικρατούσα τιμή προσδιορίζεται, όπως φαίνεται στο διπλανό σχήμα ως η τετμημένη του σημείου τομής $Z$ των ευθύγραμμων τμημάτων $ΑΓ$ και $ΒΔ$.




Στην περίπτωση της κανονικής κατανομής ο αριθμητικός μέσος , η διάμεσος και η επικρατούσα τιμή συμπίπτουν αν όμως η κατανομή δεν είναι κανονική (είναι δηλαδή ασυμμετρική) τότε οι παραπάνω παράμετροι διαφέρουν μεταξύ τους με τη διάμεσο να βρίσκεται μεταξύ των άλλων δυο παραμέτρων.

# ΤΥΠΟΛΟΓΙΟ

| **ΜΕΤΡΑ ΘΕΣΗΣ** | | |
|---|---|---|
| **Αριθμητικός μέσος** | **Αστάθμητος** : | $\overline{x} = \frac{x_1 + x_2 + \ldots + x_\nu}{\nu} = \frac{\sum_{i=1}^{\nu} x_i}{\nu} = \frac{1}{\nu}\sum_{i=1}^{\nu} xi$ |
| | **Σταθμικός** : | $\overline{x} = \frac{x_1\nu_1 + x_2\nu_2 + \ldots + x_\kappa\nu_\kappa}{\nu_1 + \nu_2 + \ldots + \nu_\kappa} = \frac{\sum_{i=1}^{\kappa} x_i\nu_i}{\sum_{i=1}^{\kappa}\nu_i} = \frac{1}{\nu}\sum_{i=1}^{\kappa} x_i\nu_i$ |
| **Γεωμετρικός μέσος** | **Αστάθμητος** : | $G=\sqrt[\kappa]{x_1 + x_2 + x_3 + \ldots + x_k}$ |
| | **Σταθμικός** : | $G=\sqrt[\sum \nu_i]{x_1^{\nu_1} + x_2^{\nu_2} + x_3^{\nu_3} + \ldots + x_\kappa^{\nu_\kappa}}$ , $i=1,2,3,\ldots,\kappa$ |
| **Αρμονικός μέσος** | **Αστάθμητος** : | $H=\frac{\kappa}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\cdots+\frac{1}{x_\kappa}}$ , $i=1,2,3,\ldots,\kappa$ |
| | **Σταθμικός** : | $H=\frac{\sum \nu_i}{\frac{\nu_1}{x_1}+\frac{\nu_2}{x_2}+\frac{\nu_3}{x_3}+\cdots+\frac{\nu_\kappa}{x_\kappa}}$ , $i=1,2,3,\ldots,\kappa$ |

# ΕΡΩΤΗΣΕΙΣ – ΑΣΚΗΣΕΙΣ

**Ερωτήσεις του τύπου «Σωστό - Λάθος»**

1. Η επικρατούσα τιμή των παρατηρήσεων 0, 1, 2, 2, 3, 3, 5, 6, 6, 6, 7 είναι ο αριθμός 7. Σ Λ
2. Διάμεσος (δ) ενός δείγματος ν παρατηρήσεων είναι η τιμή για την οποία το πολύ 50% των παρατηρήσεων είναι μικρότερες και το πολύ 50% των παρατηρήσεων είναι μεγαλύτερες από την τιμή αυτή. Σ Λ
3. Διάμεσος (δ) ενός δείγματος ν παρατηρήσεων οι οποίες έχουν διαταχθεί σε αύξουσα σειρά ορίζεται ως η μεσαία παρατήρηση, όταν ο ν είναι περιττός. Σ Λ
4. Διάμεσος (δ) ενός δείγματος ν παρατηρήσεων οι οποίες έχουν διαταχθεί σε αύξουσα σειρά ορίζεται η ημιδιαφορά των δύο μεσαίων παρατηρήσεων, όταν ο ν είναι άρτιος αριθμός. Σ Λ
5. Η μέση τιμή ενός συνόλου ν παρατηρήσεων είναι ένα μέτρο θέσης. Σ Λ
6. Επικρατούσα τιμή ενός δείγματος ν παρατηρήσεων ορίζεται η τιμή με τη μεγαλύτερη σχετική συχνότητα. Σ Λ
7. Ορίζουμε ως κ εκατοστιαίο σημείο ή $P_κ$ εκατοστημόριο ενός συνόλου παρατηρήσεων την τιμή εκείνη για την οποία το πολύ κ% των παρατηρήσεων είναι μικρότερες του $P_κ$ και το πολύ (100 - κ) % των παρατηρήσεων είναι μεγαλύτερες από την τιμή αυτή. Σ Λ
8. Για το $Q_1$ τεταρτημόριο ενός συνόλου παρατηρήσεων έχουμε αριστερά το πολύ 75% των παρατηρήσεων και δεξιά το πολύ 25% των παρατηρήσεων. Σ Λ
9. Το $Q_2$ τεταρτημόριο ενός συνόλου παρατηρήσεων ισούται με τη διάμεσο. Σ Λ
10. Έχουμε τις παρατηρήσεις: 0, 1, 1, 2, 2, 2, 3, 4, 5.
    i) Η διάμεσος είναι 2. Σ Λ
    ii) Το τεταρτημόριο $Q_1$ είναι ίσο με 1. Σ Λ
    iii) Το τεταρτημόριο $Q_3$ είναι ίσο με 4. Σ Λ

## Ερωτήσεις πολλαπλής επιλογής

1. Σε ένα δείγμα μεγέθους ν αν οι παρατηρήσεις μιας μεταβλητής Χ είναι $t_1, t_2, \ldots, t_\nu$. Τότε η μέση τιμή $\overline{x}$ ισούται με

   Α. $\frac{1}{\nu^2}\sum_{i=1}^{\nu} t_i$.   Β. $\frac{1}{\nu}\sum_{i=1}^{\nu} t_i$.   Γ. $\frac{1}{\nu^2}\sum_{i=1}^{\nu} t_i^2$.

   Δ. $\frac{1}{\nu}\sum_{i=1}^{\nu} t_i^2$.   Ε. $\frac{1}{\nu-1}\sum_{i=1}^{\nu} t_i$.

2. Αν σε κάθε τιμή $x_1, x_2, \ldots, x_\nu$ ενός συνόλου δεδομένων δώσουμε διαφορετική βαρύτητα που εκφράζεται με τους συντελεστές στάθμισης (βαρύτητας) $w_1, w_2, \ldots, w_\nu$, τότε ο σταθμικός μέσος βρίσκεται από τον τύπο

   Α. $\overline{x} = \frac{\sum_{i=1}^{\nu} x_i w_i}{\nu}$.   Β. $\overline{x} = \frac{\sum_{i=1}^{\nu} x_i w_i}{\sum_{i=1}^{\nu} \nu_i}$.   Γ. $\overline{x} = \frac{\sum_{i=1}^{\nu} x_i w_i}{\sum_{i=1}^{\nu} w_i}$.

   Δ. $\overline{x} = \frac{\sum_{i=1}^{\nu} x_i w_i}{\sum_{i=1}^{\nu} x_i}$.   Ε. $\overline{x} = \frac{\sum_{i=1}^{\nu} x_i w_i}{\sum_{i=1}^{\nu} \nu_i^2}$.

3. Στις παρατηρήσεις 0, 1, 2, 2, 3, 4, 5, 6 η επικρατούσα τιμή είναι

   Α. 1.   Β. 2.   Γ. 3.   Δ. 4.   Ε. 6.

4. Στις παρατηρήσεις 0, 1, 2, 3, 4, 5 η επικρατούσα τιμή είναι

   Α. 0.   Β. 1.   Γ. 2.   Δ. 3.   Ε. καμία από τις παραπάνω.

## Ερωτήσεις συμπλήρωσης - σύντομης απάντησης

1. Τα …………………… μας δίνουν τη θέση του "κέντρου" των παρατηρήσεων στον οριζόντιο άξονα και τα …………………… ή …………………… μας δείχνουν πόσο οι παρατηρήσεις εκτείνονται γύρω από το "κέντρο" τους.

2. Ειδική περίπτωση εκατοστημορίων είναι τα $P_{25}$, $P_{50}$, $P_{75}$ τα οποία καλούνται …………………… και συμβολίζονται με ……………………, αντίστοιχα.

3. Τα μέτρα που χρησιμοποιούνται για την περιγραφή της θέσης ενός συνόλου δεδομένων πάνω στον οριζόντιο άξονα Οx είναι:

   α) …………………… β)…………………… γ)……………………

   δ) …………………… ε) ……………………

**Ερωτήσεις ανάπτυξης**

1. Εξετάστηκε ένα δείγμα 400 οικογενειών ως προς τον αριθμό των παιδιών τους και προέκυψε ο παρακάτω πίνακας:

| Αριθμός παιδιών ($x_i$) | Αριθμός ($v_i$) οικογενειών | $f_i$ | $f_i$ % | $v_i x_i$ | $N_i$ |
|---|---|---|---|---|---|
| 0 | 135 | | | | |
| 1 | 220 | | | | |
| 2 | 8 | | | | |
| 3 | 15 | | | | |
| 4 | 12 | | | | |
| 5 | 10 | | | | |
| | 400 | | | | |

α) Να συμπληρώσετε τον πίνακα. β) Να κάνετε το διάγραμμα συχνοτήτων.

γ) Να υπολογίσετε: i) Τη μέση τιμή. ii) Τη διάμεσο της κατανομής.

2. Οι παρακάτω αριθμοί δίνουν τα ύψη (σε cm) ενός δείγματος 41 μαθητών ενός σχολείου.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 159 | 168 | 162 | 183 | 180 | 179 | 153 | 168 | 170 | 170 | |
| 172 | 175 | 175 | 181 | 165 | 166 | 171 | 185 | 169 | 180 | |
| 180 | 182 | 160 | 157 | 175 | 167 | 162 | 174 | 174 | 187 | |
| 192 | 166 | 172 | 167 | 187 | 177 | 178 | 174 | 171 | 177 | 172 |

α) Να υπολογίσετε τη διάμεσο.

β) Να ομαδοποιήσετε τα αναστήματα σε κλάσεις πλάτους 5 cm και να προσδιορίσετε γραφικά τη διάμεσο από το διάγραμμα σχετικών αθροιστικών συχνοτήτων.

γ) Να συγκρίνετε τα δύο αποτελέσματα.

**3.** Ο αριθμός των μαθητών των 16 τμημάτων ενός Λυκείου είναι:

31, 27, 28, 30, 29, 31, 21, 27, 29, 29, 28, 28, 30, 29, 27, 29.

Να υπολογίσετε τη μέση τιμή της μεταβλητής "αριθμός μαθητών ανά τμήμα".

**4.** Να υπολογίσετε τη μέση τιμή της μεταβλητής του παρακάτω πίνακα:

| **Ηλικία σε χρόνια** | $\nu_i$ |
|---|---|
| [0, 4) | 3 |
| [4, 8) | 5 |
| [8, 12) | 6 |
| [12, 16) | 6 |
| [16, 20) | 2 |
| | 22 |

**5.** Τα ύψη 8 αθλητών μιας ομάδας καλαθοσφαίρισης (μπάσκετ μπωλ) είναι (σε cm):

172, 175, 183, 177, 190, 193, 189, 195.

α) Να βρείτε: i) Το μέσο ύψος των αθλητών. ii) Τη διάμεσο των υψών.

β) Επίσης, σε καθεμιά από τις παρακάτω τρεις περιπτώσεις, να βρείτε :

i) Το μέσο ύψος των αθλητών. ii) Τη διάμεσο των υψών.

*Περίπτωση 1:* Φεύγει ο αθλητής με το ύψος 172 cm.

*Περίπτωση 2:* Έρχεται ακόμα ένας αθλητής με ύψος 197 cm.

*Περίπτωση 3:* Φεύγει ο αθλητής με το ύψος 195 cm και έρχεται ένας αθλητής με ύψος 198 cm.

**6.** Η βαθμολογία ενός μαθητή στα τέσσερα τεστ ενός μαθήματος ήταν (σε εκατοντα-βάθμια κλίμακα): 38, 67, 43, 72. Η βαρύτητα σε καθένα ήταν αντίστοιχα 1, 2, 2 και 3. Να βρείτε τη μέση επίδοση του μαθητή στα τεστ.

**7.** Η βαθμολογία στα 10 μαθήματα ενός μαθητή είναι: 13,9,6,10,15,12,11,0,20,14.

Να υπολογίσετε: α) Τη μέση τιμή. β) Τη διάμεσο. γ) Τα τεταρτημόρια $Q_1$, $Q_2$, $Q_3$.

**8.** Η αντοχή 100 ηλεκτρικών συσκευών δίνεται από τον επόμενο πίνακα:

| Χρόνος αντοχής σε ώρες | Αριθμός συσκευών | $f_i\%$ | $F_i\%$ |
|---|---|---|---|
| [1000, 1200) | 8 | | |
| [1200, 1400) | 16 | | |
| [1400, 1600) | 28 | | |
| [1600, 1800) | 32 | | |
| [1800, 2000) | 12 | | |
| [2000, 2200) | 4 | | |
| [2200, 2400) | 0 | | |
| ΣΥΝΟΛΑ | 100 | | |

α) Να συμπληρώσετε τον πίνακα.

β) i) Να κατασκευάσετε το ιστόγραμμα και το πολύγωνο συχνοτήτων.

ii) Να βρείτε την επικρατούσα τιμή.

γ) i) Να κατασκευάσετε το πολύγωνο σχετικών αθροιστικών συχνοτήτων.

ii) Να βρείτε τη διάμεσο. iii) Το πρώτο και τρίτο τεταρτημόριο.

**9.** Έξι διαδοχικοί άρτιοι αριθμοί έχουν μέση τιμή 15. Να βρείτε τους αριθμούς και τη διάμεσό τους.

**10.** Έχουμε ένα δείγμα $\nu = 10$ παρατηρήσεων, όπου κάθε παρατήρηση μπορεί να είναι 1, 2 ή 3. Είναι δυνατό η μέση τιμή να είναι α) 1 β) 4 γ) 1,8;

**11.** Η βαθμολογία δέκα μαθητών σε ένα διαγώνισμα ήταν: 7,11,10,13,15,3,12,11,4,14. Υπολογίστε: α) την επικρατούσα τιμή και τη διάμεσο, β) τα $Q_1$, και $Q_3$

**12.** Το μέσο ύψος 9 καλαθοσφαιριστών μιας ομάδας είναι 205 cm.

α) Για να "ψηλώσει" την ομάδα ο προπονητής πήρε έναν ακόμη παίκτη με ύψος 216 cm. Ποιο είναι το μέσο ύψος της ομάδας τώρα;

β) Εάν ήθελε να "ψηλώσει" την ομάδα στα 208 cm, πόσο ύψος έπρεπε να έχει ο καλαθοσφαιριστής που πήρε;

**13.** Το μέσο ύψος των 30 μαθητών και μαθητριών μιας τάξης είναι 170 cm. Ποιο θα είναι το μέσο ύψος της τάξης:

α) αν φύγει ένας μαθητής με ύψος 180 cm,

β) αν έρθει μια νέα μαθήτρια με ύψος 170 cm,

γ) αν φύγει ένας μαθητής με ύψος 180 cm και έλθει μαθήτρια με ύψος 170 cm;

14. Η επίδοση ενός μαθητή σε πέντε μαθήματα είναι 12, 10, 16, 18, 14.
α) Να βρείτε τη μέση επίδοση.
β) Αν τα μαθήματα είχαν συντελεστές στάθμισης 2, 3, 1, 1 και 3, ποια θα ήταν η μέση επίδοση; Σε ποια μαθήματα έπρεπε να δώσει ιδιαίτερη προσοχή ο μαθητής;

15. Η μέση ηλικία 18 αγοριών και 12 κοριτσιών μιας τάξης είναι 15,4 χρόνια. Εάν η μέση ηλικία των αγοριών είναι 15,8 χρόνια, να βρείτε αυτή των κοριτσιών.

16. Ένα εργοστάσιο απασχολεί 5 υπαλλήλους στο Τμήμα Α με μέσο (μηνιαίο) μισθό 249 χιλ. δρχ., 6 υπαλλήλους στο Τμήμα Β με μέσο μισθό 280 χιλ. δρχ. και 4 υπαλλήλους στο Τμήμα Γ με μέσο μισθό 360 χιλ. δρχ. Ποιος είναι ο μέσος μισθός όλων των υπαλλήλων;

17. Η μέση τιμή και η διάμεσος πέντε αριθμών είναι 6. Οι τρεις από αυτούς είναι οι 5, 8, 9. Να βρείτε τους άλλους δύο.

18. Στο διπλανό πίνακα δίνονται οι τιμές μιας μεταβλητής $X$ με τις αντίστοιχες συχνότητές τους. Η πέμπτη συχνότητα χάθηκε! Μπορείτε να την "ανακαλύψετε", εάν γνωρίζετε ότι
α) η μέση τιμή είναι 4,4,
β) η διάμεσος είναι το 4,5,
γ) υπάρχουν δύο επικρατούσες τιμές;

| $x_i$ | $\nu_i$ |
|---|---|
| 2 | 1 |
| 3 | 3 |
| 4 | 1 |
| 5 | 2 |
| 6 | ; |
| 7 | 1 |

19. Να υπολογίσετε τη μέση τιμή και τη διάμεσο για τα παρακάτω δείγματα δεδομένων και να σχολιάσετε τα αποτελέσματα:
α) 1 2 6 β) 2 4 12 γ) 11 12 16 δ) 12 14 22.

20. Ο παρακάτω πίνακας δίνει τον αριθμό των επισκέψεων 40 μαθητών σε διάφορα μουσεία της χώρας κατά τη διάρκεια ενός έτους.

| Επισκέψεις | Συχνότητα |
|---|---|
| [0-2) | 8 |
| 2-4 | 12 |
| 4-6 | 10 |
| 6-8 | 6 |
| 8-10 | 4 |

Να υπολογιστούν:
α) η μέση τιμή,
β) η επικρατούσα τιμή,
γ) η διάμεσος,
δ) το πρώτο και τρίτο τεταρτημόριο.

# ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ

Γνωρίζουμε από τα προηγούμενα ότι: **Τα μέτρα θέσης είναι στατιστικές παράμετροι οι οποίες καθορίζουν μια τιμή της μεταβλητής, γύρω από την οποία τείνουν να συγκεντρωθούν οι παρατηρήσεις μας.**

Στον επόμενο πίνακα φαίνονται οι βαθμοί δυο μαθητών στις Πανελλήνιες εξετάσεις

| ΜΑΘΗΜΑΤΑ | ΑΡΧΑΙΑ | ΜΑΘΗΜΑΤΙΚΑ | ΦΥΣΙΚΗ | ΧΗΜΕΙΑ | ΒΙΟΛΟΓΙΑ | ΝΕΑ ΕΛΛΗΝΙΚΑ | ΕΚΘΕΣΗ | ΛΑΤΙΝΙΚΑ | | ΣΥΝΟΛΟ | | Μ. ΟΡΟΣ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ΜΑΘΗΤΗΣ Α. | 8 | 8 | 9 | 18 | 19 | 9 | 8 | 16 | | 95 | | 11,875 |
| ΜΑΘΗΤΗΣ Β. | 11 | 11 | 12 | 12 | 13 | 12 | 12 | 12 | | 95 | | 11,875 |

Παρατηρούμε ότι οι μέσοι όροι των δυο μαθητών είναι ίδιοι, όμως οι βαθμοί του πρώτου μαθητή παρουσιάζουν μεγάλες αποκλίσεις από το μέσο όρο, ενώ αυτοί του δεύτερου μαθητή είναι συγκεντρωμένοι κοντά στο μέσο όρο της βαθμολογίας τους.

Είναι λοιπόν φανερό ότι δεν μπορούμε να καταλήξουμε σε αξιόπιστα συμπεράσματα για τα δεδομένα μας και να τα συγκρίνουμε με άλλα ομοειδή δεδομένα γνωρίζοντας μόνο τη μέση τιμή, αλλά είναι απαραίτητο να χρησιμοποιήσουμε και άλλες παραμέτρους που να φανερώνουν τον τρόπο διασποράς των τιμών της μεταβλητής.

**Αυτές οι παράμετροι ονομάζονται μέτρα διασποράς και δείχνουν το πόσο είναι "απλωμένα" τα δεδομένα γύρω από κάποιο μέτρο θέσης.**

Τα μέτρα διασποράς που θα μελετήσουμε είναι:

- Το εύρος μεταβολής
- Το ημιενδοτεταρτημοριακό εύρος
- Η μέση απόκλιση
- Η διακύμανση
- Η τυπική απόκλιση
- Ο συντελεστής μεταβολής.

### α) *Εύρος* μεταβολής (*R*)

Το απλούστερο από τα μέτρα διασποράς είναι το **εύρος μεταβολής** ή **κύμανση** (R), που ορίζεται ως η διαφορά της ελάχιστης παρατήρησης από τη μέγιστη παρατήρηση, δηλαδή:

$$\boxed{(\text{Εύρος})\quad R = \text{Μεγαλύτερη παρατήρηση} - \text{Μικρότερη παρατήρηση}}$$

Όταν έχουμε ομαδοποιημένα δεδομένα, το εύρος δίνεται από τη διαφορά του κατώτερου ορίου της πρώτης κλάσης από το ανώτερο όριο της τελευταίας κλάσης.
Το εύρος είναι ένα αρκετά απλό μέτρο, που υπολογίζεται εύκολα δε θεωρείται όμως αξιόπιστο μέτρο διασποράς, γιατί βασίζεται μόνο στις δυο ακραίες παρατηρήσεις.
Το εύρος χρησιμοποιείται κυρίως στη μετεωρολογία και στα χρηματιστήρια.

### *β) Το ημιενδοτεταρτημοριακό εύρος (Q)*

Η παράμετρος αυτή ορίζεται ως το ημιάθροισμα της διαφοράς μεταξύ του τρίτου και του πρώτου τεταρτημορίου, δηλαδή : $\boxed{Q=\frac{Q_3-Q_1}{2}}$, όσο μικρότερη είναι η τιμή του, τόσο μεγαλύτερη είναι η συγκέντρωση των τιμών της μεταβλητής. Το ημιενδοτεταρτημοριακό εύρος δεν επηρεάζεται από τις ακραίες θέσεις όπως συμβαίνει με το εύρος μεταβολής.

### *γ) Μέση απόκλιση*

Ως μέση απόκλιση θεωρούμε το μέτρο διασποράς το οποίο ορίζεται ως εξής:
Αν μια μεταβλητή Χ παίρνει ν τιμές: $x_1, x_2, x_3, \ldots, x_\nu$, που έχουν μέση τιμή $\bar{x}$, και σχηματίσουμε τις διαφορές $x_1-\bar{x}, x_2-\bar{x}, \ldots x_\nu-\bar{x}$, τότε ως **μέση απόκλιση της μεταβλητής Χ** ορίζεται η μέση τιμή των απολύτων τιμών των διαφορών αυτών, δηλαδή:

$$M.A. = \frac{|x_1-\bar{x}|+|x_2-\bar{x}|+\ldots+|x_\nu-\bar{x}|}{\nu} = \frac{\sum_{i=1}^{\nu}|x_i-\bar{x}|}{\nu} = \frac{\sum|x-\bar{x}|}{\nu}$$

- Στις ομαδοποιημένες παρατηρήσεις ακολουθούμε παρόμοια διαδικασία μόνο που ως $x_i$ παίρνουμε τις κεντρικές τιμές των κλάσεων και η μέση απόκλιση δίνεται από τον τύπο : $M.A. = \frac{\sum f_i|x_i-\bar{x}|}{\sum f_i}$

### *δ) Διακύμανση ($s^2$)*

**Διακύμανση** ή **διασπορά των τιμών της μεταβλητής X** ορίζουμε τη μέση τιμή των τετραγώνων των αποκλίσεων των τιμών της μεταβλητής από το μέσο αριθμητικό της ($\bar{x}$). Τη συμβολίζουμε με $s^2$ και δίνεται από τον τύπο:

$$\boxed{s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_v - \bar{x})^2}{v} = \frac{\sum_{i=1}^{v}(x_i - \bar{x})^2}{v} = \frac{\sum (x - \bar{x})^2}{v}}$$

Όταν έχουμε πίνακα συχνοτήτων ή ομαδοποιημένα δεδομένα, η διακύμανση ορίζεται

από την : $\boxed{s^2 = \frac{1}{v}\sum_{i=1}^{\kappa}(x_i - \bar{x})^2 v_i}$ ή ισοδύναμα : $\boxed{s^2 = \frac{1}{v}\left\{\sum_{i=1}^{\kappa} x_i^2 v_i - \frac{\left(\sum_{i=1}^{\kappa} x_i v_i\right)^2}{v}\right\}.}$

όπου $x_1, x_2, \ldots, x_\kappa$ οι τιμές της μεταβλητής (ή τα κέντρα των κλάσεων) με αντίστοιχες συχνότητες $v_1, v_2, \ldots, v_\kappa$.

Για παράδειγμα, η διακύμανση της βαθμολογίας των μαθητών ενός τμήματος είναι

$$s_A^2 = \frac{(13-15)^2 + (13-15)^2 + (14-15)^2 + \ldots + (18-15)^2}{10} = \frac{20}{10} = 2.$$

Ομοίως, η διακύμανση του ύψους των μαθητών για τα ομαδοποιημένα δεδομένα του επόμενου πίνακα, υπολογίζεται σύμφωνα με τον τελευταίο τύπο, όπως φαίνεται:

| Κλάσεις [ - ) | Κεντρικές τιμές $x_i$ | Συχνότητα $v_i$ | $x_i^2$ | $x_i v_i$ | $x_i^2 v_i$ |
|---|---|---|---|---|---|
| 156-162 | 159 | 2 | 25281 | 318 | 50562 |
| 162-168 | 165 | 8 | 27225 | 1320 | 217800 |
| 168-174 | 171 | 12 | 29241 | 2052 | 350892 |
| 174-180 | 177 | 11 | 31329 | 1942 | 344619 |
| 180-186 | 183 | 5 | 33489 | 915 | 167445 |
| 186-192 | 189 | 2 | 35721 | 378 | 71442 |
| Σύνολο: | | $v = 40$ | — | $\Sigma x_i v_i = 6930$ | $\Sigma x_i^2 v_i = 1202760$ |

Επομένως: $s^2 = \frac{1}{v}\left\{\sum_{i=1}^{\kappa} x_i^2 v_i - \frac{\left(\sum_{i=1}^{\kappa} x_i v_i\right)^2}{v}\right\} = \frac{1}{40}\left\{1202776 - \frac{6930^2}{40}\right\} = 53{,}4$

## ε) Τυπική Απόκλιση (s)

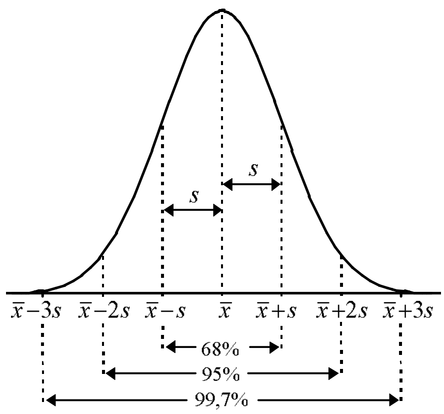
Η διακύμανση είναι μια αξιόπιστη παράμετρος διασποράς, αλλά έχει ένα μειονέκτημα. Δεν εκφράζεται με τις μονάδες με τις οποίες εκφράζονται οι παρατηρήσεις.
Για παράδειγμα, αν οι παρατηρήσεις εκφράζονται σε cm, η διακύμανση εκφράζεται σε $\mathrm{cm}^2$. Αν όμως πάρουμε τη θετική τετραγωνική ρίζα της διακύμανσης, θα έχουμε ένα μέτρο διασποράς που θα εκφράζεται με την ίδια μονάδα μέτρησης του χαρακτηριστικού, όπως ακριβώς είναι και όλα τα άλλα μέτρα θέσης, που εξετάσαμε έως τώρα. Η ποσότητα αυτή λέγεται **τυπική απόκλιση**, συμβολίζεται με *s* και δίνεται από τη σχέση:

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

Όσο μικρότερες είναι οι τιμές της διακύμανσης και της τυπικής απόκλισης τόσο πιο συγκεντρωμένες γύρω από τη μέση τιμή $\bar{x}$ βρίσκονται οι τιμές της μεταβλητής X, με αποτέλεσμα η μέση τιμή $(\bar{x})$ να αποτελεί αντιπροσωπευτικό στατιστικό μέτρο για την κατανομή της μεταβλητής.

Η τυπική απόκλιση για το ύψος των μαθητών του πίνακα από το προηγούμενο παράδειγμα είναι $s = \sqrt{53{,}4} = 7{,}3\ \mathrm{cm}$.

Αξίζει να σημειωθεί ότι αν η καμπύλη συχνοτήτων για το χαρακτηριστικό που εξετάζουμε είναι κανονική ή περίπου κανονική, τότε η τυπική απόκλιση *s* έχει τις παρακάτω ιδιότητες:




i) το 68% περίπου των παρατηρήσεων βρίσκεται στο διάστημα
$$(\bar{x} - s,\ \bar{x} + s)$$
ii) το 95% περίπου των παρατηρήσεων βρίσκεται στο διάστημα
$$(\bar{x} - 2s,\ \bar{x} + 2s)$$
iii) το 99,7% περίπου των παρατηρήσεων βρίσκεται στο διάστημα
$$(\bar{x} - 3s,\ \bar{x} + 3s)$$
iv) το εύρος ισούται περίπου με έξι τυπικές αποκλίσεις, δηλαδή $R \approx 6s$.

## ***<u>Ιδιότητες της διακύμανσης</u>***

Έστω ένα σύνολο αριθμών $x_1, x_2, \ldots, x_\nu$ με μέση τιμή $\overline{x}$ και τυπική απόκλιση s.

- *Αν σε κάθε τιμή της μεταβλητής **προσθέσουμε ή αφαιρέσουμε** ένα σταθερό αριθμό κ, τότε η διακύμανση παραμένει αμετάβλητη.* $s_y^2 = s_x^2$
- *Αν κάθε τιμή της μεταβλητής Χ **πολλαπλασιασθεί (ή διαιρεθεί)** με ένα σταθερό αριθμό (α), τότε η διακύμανση τιμών της μεταβλητής Υ = αΧ* ($Y = \frac{X}{a}, a \neq 0$) *πολλαπλασιάζεται (ή διαιρείται) με το τετράγωνο του αριθμού αυτού, δηλαδή:*

$s_y^2 = \alpha^2 s_x^2$ $\qquad$ $(s_y^2 = \frac{s_x^2}{a^2})$

- *Αν οι τιμές της μεταβλητής Χ είναι **ίσες** μεταξύ τους δηλαδή:*
  $x_1 = x_2 = \ldots = x_\nu = \kappa$, *τότε* $s_x^2 = 0$

## ***<u>Διόρθωση της διακύμανσης κατά Sheppard</u>***

Όταν έχουμε κατανομή συχνοτήτων σε μορφή κλάσεων, για να υπολογίσουμε το μέσο αριθμητικό ή τη διακύμανση, υποθέτουμε ότι οι παρατηρήσεις που ανήκουν σε κάθε κλάση είναι συγκεντρωμένες στο κέντρο της, δηλαδή δεχόμαστε ότι η τιμή του μέσου κάθε κλάσης ισούται με το μέσο των παρατηρήσεων που ανήκουν σε κάθε κλάση.
Η υπόθεση αυτή όμως, ότι όλες οι τιμές που ανήκουν σε μια κλάση συμπίπτουν με το μέσο της δεν είναι σωστή.
Αν η κατανομή είναι συμμετρική η παραπάνω υπόθεση δεν οδηγεί σε σφάλμα όταν υπολογίζουμε το μέσο αριθμητικό, γιατί οι αποκλίσεις μεταξύ του μέσου κάθε κλάσης και του αριθμητικού μέσου των παρατηρήσεων που βρίσκονται στην κλάση αυτή αλλοσυμψηφίζονται, γιατί τα θετικά σφάλματα συμψηφίζονται με τα αρνητικά. Όταν όμως τις αποκλίσεις τις υψώνουμε στο τετράγωνο, όπως συμβαίνει στη διακύμανση, έχουμε ένα σφάλμα, γιατί το τετράγωνο της απόκλισης του μέσου κάθε κλάσης από τη μέση τιμή της κατανομής είναι μεγαλύτερο από το τετράγωνο της απόκλισης του πραγματικού μέσου της τάξης από τη μέση τιμή της κατανομής. Επομένως το άθροισμα των τετραγώνων των αποκλίσεων που προκύπτει από τα δεδομένα που έχουμε ταξινομήσει σε μορφή κλάσεων είναι μεγαλύτερο του αθροίσματος των τετραγώνων των αποκλίσεων που θα προέκυπτε από μη ομαδοποιημένα κατά συχνότητα δεδομένα.
Μια εμπειρική διόρθωση στο παραπάνω θέμα έγινε από τον Sheppard.

*"Αν μια κατανομή συχνοτήτων είναι συμμετρική και έχει ίσα διαστήματα κλάσεων, τότε η διόρθωση της διακύμανσης γίνεται από τη σχέση :*

$$s_*^2 = s^2 - \frac{c^2}{12}$$ *, όπου c είναι το πλάτος των 12 κλάσεων."*

Αν μια κατανομή είναι ασυμμετρική δεν ισχύει η παραπάνω σχέση.
Στην περίπτωση που το πλάτος των παρατηρήσεων είναι πάνω από 500 η έχουμε άνισα διαστήματα κλάσεων πρέπει να αποφεύγεται η διόρθωση κατά Sheppard.

## *ζ) Συντελεστής Μεταβολής (CV)*

Έστω ότι από ένα δείγμα είκοσι μαθητών της Α΄ Γυμνασίου βρήκαμε μέσο βάρος $\bar{x}_A = 40\,\text{kgr}$ και τυπική απόκλιση $s_A = 6\,\text{kgr}$, ενώ από ένα δεύτερο δείγμα τριάντα μαθητών της Γ΄ Λυκείου βρήκαμε μέσο βάρος $\bar{x}_B = 75\,\text{kgr}$ και τυπική απόκλιση $s_B = 6\,\text{kgr}$. Όπως αντιλαμβανόμαστε, είναι λάθος να πούμε ότι το βάρος των μαθητών του Λυκείου έχει τον ίδιο βαθμό μεταβλητότητας με το βάρος των μαθητών του Γυμνασίου, καθόσον η βαρύτητα που έχουν τα 6 kgr στο μέσο βάρος των 40 kgr είναι διαφορετική από αυτήν που έχουν στο μέσο βάρος των 75 kgr.
Ακόμη, ας υποθέσουμε ότι ο μέσος μισθός των υπαλλήλων μιας εταιρείας Α είναι $\bar{x}_A = 250.000$ δρχ. με τυπική απόκλιση $s_A = 42.000$ δρχ., ενώ για τους υπαλλήλους μιας εταιρείας Β είναι $\bar{x}_B = \$1.400$ με τυπική απόκλιση $s_B = \$350$. Στην περίπτωση αυτή έχουμε διαφορετικές μονάδες μέτρησης του μισθού, επομένως οι διασπορές των παρατηρήσεων δεν είναι άμεσα συγκρίσιμες.
Ένα μέτρο με το οποίο μπορούμε να ξεπεράσουμε τις παραπάνω δυσκολίες και το οποίο μας βοηθά στη σύγκριση ομάδων τιμών, που είτε εκφράζονται σε διαφορετικές μονάδες μέτρησης είτε εκφράζονται στην ίδια μονάδα μέτρησης, αλλά έχουν σημαντικά διαφορετικές μέσες τιμές, είναι ο **συντελεστής μεταβολής** ή **συντελεστής μεταβλητότητας** , ο οποίος ορίζεται από το λόγο:

$$\boxed{CV = \frac{\text{τυπική απόκλιση}}{\text{μέση τιμή}} \cdot 100\% = \frac{s}{\bar{x}} \cdot 100\%.}$$

Ο συντελεστής μεταβολής εκφράζεται επί τοις εκατό, είναι συνεπώς ανεξάρτητος από τις μονάδες μέτρησης και παριστάνει ένα μέτρο **σχετικής διασποράς** των τιμών και όχι της απόλυτης διασποράς, όπως έχουμε δει έως τώρα. Εκφράζει, δηλαδή, τη μεταβλητότητα των δεδομένων απαλλαγμένη από την επίδραση της μέσης τιμής. Για το πρώτο παράδειγμα του βάρους έχουμε συντελεστή μεταβολής για τις δύο ομάδες μαθητών:

$$CV_A = \frac{s_A}{\bar{x}_A} \cdot 100\% = \frac{6}{40} \cdot 100\% = 15\% \quad \text{και} \quad CV_B = \frac{s_B}{\bar{x}_B} \cdot 100\% = \frac{6}{75} \cdot 100\% = 8\%$$

δηλαδή, ο βαθμός διασποράς του βάρους των μαθητών Γυμνασίου είναι μεγαλύτερος από το βαθμό διασποράς του βάρους των μαθητών Λυκείου.
Ανάλογα συμπεράσματα βγάζουμε και για το δεύτερο παράδειγμα, όπου βρίσκουμε $CV_A = 12\%$ και $CV_B = 25\%$. Παρ' όλο που η τυπική απόκλιση των μισθών στην εταιρεία Α είναι μεγαλύτερη από την τυπική απόκλιση στην εταιρεία Β, ο συντελεστής μεταβολής δίνει μεγαλύτερη σχετική διασπορά στην εταιρεία Β. Αυτό μεταφράζεται στο να λέμε ότι έχουμε μεγαλύτερη **ομοιογένεια** μισθών στην εταιρεία Α παρά στη Β.
Γενικά δεχόμαστε ότι ένα δείγμα τιμών μιας μεταβλητής θα είναι ομοιογενές, εάν ο συντελεστής μεταβολής δεν ξεπερνά το 10%.

## ΕΦΑΡΜΟΓΗ

Ο διπλανός πίνακας συχνοτήτων δίνει την κατανομή του χρόνου $X$ (σε sec) 60 μαθητών που χρειάστηκαν, για να τρέξουν μια δεδομένη απόσταση. Να υπολογιστούν:

α) ο μέσος, ο διάμεσος και ο επικρατέστερος χρόνος για την κάλυψη της συγκεκριμένης απόστασης,

β) η τυπική απόκλιση, γ)σε πόσο χρόνο από της στιγμή της εκκίνησης κάλυψε την απόσταση το 25% των μαθητών.

| $x_i$ | $v_i$ |
|---|---|
| 50 | 4 |
| 55 | 6 |
| 60 | 8 |
| 65 | 12 |
| 70 | 14 |
| 75 | 10 |
| 80 | 6 |

*ΛΥΣΗ*

α) • Για τον υπολογισμό της μέσης τιμής συμπληρώνουμε τις τρεις πρώτες στήλες του παρακάτω πίνακα:

| $x_i$ | $v_i$ | $x_i v_i$ | $x_i^2 v_i$ |
|---|---|---|---|
| 50 | 4 | 200 | 10000 |
| 55 | 6 | 330 | 18150 |
| 60 | 8 | 480 | 28800 |
| 65 | 12 | 780 | 50700 |
| 70 | 14 | 980 | 68600 |
| 75 | 10 | 750 | 56250 |
| 80 | 6 | 480 | 38400 |
| Σύνολο | $v$=60 | $\Sigma x_i v_i$ =4000 | $\Sigma x_i^2 v_i$ =270900 |

Επομένως, ο μέσος χρόνος για την κάλυψη της συγκεκριμένης απόστασης είναι:

$$\bar{x} = \frac{\sum x_i v_i}{v} = \frac{4000}{60} \approx 66{,}67 \text{ sec.}$$

• Έχουμε $v = 60$ παρατηρήσεις σε αύξουσα σειρά, άρα η διάμεσος είναι ο μέσος όρος της 30ής και 31ης παρατήρησης, δηλαδή ο μέσος όρος των παρατηρήσεων 65 και 70, άρα $\delta = \frac{65+70}{2} = 67{,}5 \text{ sec}$.

• Η επικρατούσα τιμή είναι η τιμή με τη μεγαλύτερη συχνότητα, άρα $M_0 = 70 \text{ sec}$.

β) Για τον υπολογισμό της τυπικής απόκλισης είναι προτιμότερο να εφαρμόσουμε τη σχέση (4), μιας και η μέση τιμή δεν είναι ακέραιος αριθμός. Με βάση τον παραπάνω πίνακα η διακύμανση της μεταβλητής $X$ είναι:

$$s^2 = \frac{1}{v}\left\{\sum x_i^2 v_i - \frac{\left(\sum x_i v_i\right)^2}{v}\right\} = \frac{1}{60}\left\{270900 - \frac{4000^2}{60}\right\} = 70{,}56 \text{ sec}^2$$ και η τυπική απόκλιση

$s = \sqrt{70{,}56} = 8{,}4 \text{ sec}$.

γ) Θέλουμε να υπολογίσουμε το πρώτο τεταρτημόριο, $Q_1$. Αριστερά της διαμέσου $\delta = 67{,}5$ έχουμε 30 παρατηρήσεις. Η διάμεσος αυτών των 30 πρώτων παρατηρήσεων είναι το ημιάθροισμα της 15ης και 16ης παρατήρησης, δηλαδή $Q_1 = (60+60)/2 = 60 \text{ sec}$. Δηλαδή, ύστερα από μία ώρα από τη στιγμή της εκκίνησης το 25% των μαθητών κάλυψαν τη συγκεκριμένη απόσταση.

# ΤΥΠΟΛΟΓΙΟ

| **ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ** |
|---|
| **Εύρος :** R = Μεγαλύτερη παρατήρηση-Μικρότερη παρατήρηση |
| **Μέση απόκλιση :** $\text{M.A.} = \dfrac{\sum f_i \lvert x_i - \bar{x} \rvert}{\sum f_i}$ |
| **Ενδοτεταρτημοριακό εύρος :** $Q = \dfrac{Q_3 - Q_1}{2}$ |
| **Διακύμανση :** $s^2 = \dfrac{1}{v}\sum_{i=1}^{\kappa}(x_i - \bar{x})^2 v_i$ ή ισοδύναμα $\boxed{s^2 = \dfrac{1}{v}\left\{\sum_{i=1}^{\kappa} x_i^2 v_i - \dfrac{\left(\sum_{i=1}^{\kappa} x_i v_i\right)^2}{v}\right\}.}$ |
| **Τυπική απόκλιση :** $s = \sqrt{s^2}$ |
| **Συντελεστής μεταβολής :** $CV = \dfrac{\text{τυπική απόκλιση}}{\text{μέση τιμή}} \cdot 100\% = \dfrac{s}{\bar{x}} \cdot 100\%$ |

# ΕΡΩΤΗΣΕΙΣ – ΑΣΚΗΣΕΙΣ

**Ερωτήσεις του τύπου «Σωστό - Λάθος»**

| | | |
|---|---|---|
| 1. Ο συντελεστής μεταβολής ή συντελεστής μεταβλητότητας (CV) είναι ανεξάρτητος από τις μονάδες μέτρησης. | Σ | Λ |
| 2. Ο συντελεστής μεταβλητότητας εκφράζει τη μεταβλητότητα των δεδομένων απαλλαγμένη από την επίδραση της μέσης τιμής. | Σ | Λ |
| 3. Ο συντελεστής μεταβλητότητας (CV) παριστάνει ένα μέτρο απόλυτης διασποράς και όχι σχετικής διασποράς. | Σ | Λ |
| 4. Ένα δείγμα τιμών μιας μεταβλητής είναι ομοιογενές αν ο συντελεστής μεταβολής ξεπερνά το 10%. | Σ | Λ |
| 5. Αν οι παρατηρήσεις εκφράζονται σε cm και η διακύμανση εκφράζεται σε cm. | Σ | Λ |
| 6. Τα μέτρα διασποράς εκφράζουν τις αποκλίσεις των τιμών μιας μεταβλητής γύρω από τα μέτρα κεντρικής τάσης. | Σ | Λ |
| 7. Το εύρος ή κύμανση (R) ενός δείγματος ν παρατηρήσεων ορίζεται ως το άθροισμα της μεγαλύτερης και της μικρότερης παρατήρησης. | Σ | Λ |
| 8. Το εύρος ενός δείγματος βασίζεται στις δύο ακραίες παρατηρήσεις και είναι αξιόπιστο μέτρο διασποράς. | Σ | Λ |
| 9. Το ενδοτεταρτημοριακό εύρος είναι η διαφορά του πρώτου τεταρτημορίου $Q_1$ από το τρίτο τεταρτημόριο $Q_3$. | Σ | Λ |
| 10. Η διακύμανση ή διασπορά είναι ο μέσος όρος των τετραγώνων των αποκλίσεων των παρατηρήσεων $t_i$ από τη μέση τιμή τους $\bar{x}$. | Σ | Λ |
| 11. Τα μέτρα θέσης δίνουν τη θέση του “κέντρου” των παρατηρήσεων στον κατακόρυφο άξονα Oy. | Σ | Λ |
| 12. Τα μέτρα διασποράς μας δίνουν πόσο επεκτείνονται οι παρατηρήσεις γύρω από το “κέντρο” τους. | Σ | Λ |

**Ερωτήσεις πολλαπλής επιλογής**

**1.** Αν η καμπύλη συχνοτήτων για το χαρακτηριστικό που εξετάζουμε είναι κανονική, τότε το εύρος ισούται περίπου με

**Α.** 2 τυπικές αποκλίσεις. **Β.** 3 τυπικές αποκλίσεις. **Γ.** 4 τυπικές αποκλίσεις. **Δ.** 5 τυπικές αποκλίσεις. **Ε.** 6 τυπικές αποκλίσεις.

**2.** Αν η καμπύλη συχνοτήτων για το χαρακτηριστικό που εξετάζουμε είναι κανονική, τότε το 68% περίπου των παρατηρήσεων βρίσκεται στο διάστημα

**Α.** $(\bar{x} + s, \bar{x} + 2s)$. **Β.** $(\bar{x} - s, \bar{x} + 2s)$. **Γ.** $(\bar{x} - s, \bar{x} + s)$. **Δ.** $(\bar{x} - 2s, \bar{x} + 2s)$. **Ε.** $(\bar{x} - 3s, \bar{x} + 3s)$.

**3.** Αν η καμπύλη συχνοτήτων για το χαρακτηριστικό που εξετάζουμε είναι κανονική, τότε το 95% περίπου των παρατηρήσεων βρίσκεται στο διάστημα

**Α.** $(\bar{x} - s, \bar{x} + s)$. **Β.** $(\bar{x} - 2s, \bar{x} + s)$. **Γ.** $(\bar{x} - 2s, \bar{x} + 2s)$. **Δ.** $(\bar{x} - s, \bar{x} + 3s)$. **Ε.** $(\bar{x} - 3s, \bar{x} + 3s)$.

**4.** Η μέση τιμής μιας κανονικής κατανομής είναι 25 και η τυπική απόκλιση είναι 5. Το ποσοστό των παρατηρήσεων που είναι μεταξύ 20 και 30 είναι περίπου

**Α.** 34%. **Β.** 65%. **Γ.** 68%. **Δ.** 95%. **Ε.** 99,7%.

**5.** Η μέση τιμή μιας κανονικής κατανομής είναι 20 και η τυπική απόκλιση είναι 3. Το ποσοστό των παρατηρήσεων που είναι μεταξύ 14 και 26 είναι περίπου

**Α.** 34%. **Β.** 47,5%. **Γ.** 68%. **Δ.** 95%. **Ε.** 99,7%.

**6.** Η μέση τιμής μιας κανονικής κατανομής είναι 30 και η τυπική απόκλιση είναι 3. Το ποσοστό των παρατηρήσεων που είναι μεταξύ 30 και 33 είναι περίπου

**Α.** 34%. **Β.** 47,5%. **Γ.** 68%. **Δ.** 95%. **Ε.** 99,7%.

**7.** Ένα εργοστάσιο κατασκευάζει μεταλλικούς δίσκους για τη λειτουργία μιας μηχανής. Η κατανομή συχνοτήτων ως προς τη διάμετρό τους είναι κανονική με μέση τιμή (διάμετρο) 32 cm και τυπική απόκλιση 0,2 cm.

i) Αν αγοράσουμε ένα τέτοιο δίσκο η διάμετρός του είναι σχεδόν βέβαιο ότι θα βρίσκεται στο διάστημα μεταξύ

**Α.** 33,5 cm και 35,2 cm. **Β.** 31,4 cm και 32,6 cm. **Γ.** 29,2 cm και 31,4 cm. **Δ.** 32,6 cm και 35,5 cm. **Ε.** 20,7 cm και 22,3 cm.

ii) Αν διαλέξουμε ένα τέτοιο δίσκο στην τύχη, πρέπει να ελέγξουμε τη λειτουργία της μηχανής για πιθανή βλάβη, όταν η διάμετρός του είναι

**Α**. 31,5 cm. **Β**. 31,7 cm. **Γ**. 31,2cm.
**Δ**. 31,9 cm. **Ε**. 32,5 cm.

**8.** Σε ένα δείγμα μεγέθους ν, αν $x_1, x_2, \ldots, x_\kappa$ είναι οι τιμές της μεταβλητής Χ με συχνότητες αντίστοιχα $\nu_1, \nu_2, \ldots, \nu_\kappa$ και αν $f_i$ είναι οι σχετικές συχνότητες, ποια (ή ποιες) από τις παρακάτω σχέσεις **δεν** ορίζει τη μέση τιμή $\overline{x}$ του δείγματος

**Α**. $\overline{x} = \frac{1}{\nu}\sum_{i=1}^{\kappa} x_i \nu_i$. **Β**. $\overline{x} = \frac{1}{\nu}\sum_{i=1}^{\kappa} x_i^2 f_i$. **Γ**. $\overline{x} = \frac{1}{\nu}\sum_{i=1}^{\kappa} x_i f_i$. **Δ**. οι σχέσεις Α και Γ.

**9.** Ο συντελεστής μεταβολής εκφράζεται από το λόγο

**Α**. $\frac{s^2}{\overline{x}}100\%$. **Β**. $\frac{s}{\overline{x}}100\%$. **Γ**. $\frac{\overline{x}}{s}100\%$. **Δ**. $\frac{\overline{x}}{s^2}100\%$. **Ε**. $\frac{\overline{x}^2}{s}100\%$.

**Ερωτήσεις συμπλήρωσης - σύντομης απάντησης**

**1.** Τα σπουδαιότερα μέτρα διασποράς ή μεταβλητότητας είναι:

α) ……………………… β) ………………………
γ) ……………………… δ) ………………………

**2.** Το μέτρο το οποίο μας βοηθά στη σύγκριση ομάδων τιμών, που είτε εκφράζονται σε διαφορετικές μονάδες μέτρησης είτε εκφράζονται στην ίδια μονάδα μέτρησης, αλλά έχουν σημαντικά διαφορετικές μέσες τιμές, είναι ο ……………………………… .

## Ερωτήσεις ανάπτυξης

1. Η τυπική απόκλιση μιας μεταβλητής Χ, είναι ίση με το μηδέν. Αν $t_1, t_2, \ldots, t_\nu$ είναι οι τιμές της Χ και $\bar{x}$ η μέση τιμή, δείξτε ότι $t_1 = t_2 = \ldots = t_\nu = \bar{x}$.

2. Η βαθμολογία στα 10 μαθήματα ενός μαθητή είναι: 13, 9, 6, 10, 15, 12, 11, 0, 20, 14. Να υπολογίσετε:
α) Τη μέση τιμή. β) Τη διακύμανση. γ) Την τυπική απόκλιση. δ) Τη διάμεσο.
ε) Τα τεταρτημόρια $Q_1$, $Q_2$, $Q_3$. στ) Το ενδοτεταρτημοριακό εύρος των βαθμών.
ζ) Το εύρος (R). η) Το συντελεστή μεταβολής (CV).

3. Οι μαθητές του Γ2 ξόδεψαν σε μια μέρα κατά μέσο όρο 625 δρχ. αγοράζοντας διάφορα τρόφιμα από το κυλικείο του σχολείου. Εάν ο συντελεστής μεταβολής είναι 27,2%, να βρείτε την τυπική απόκλιση. Εάν επιπλέον γνωρίζετε ότι το $\sum x_i^2 = 11.746.700$, πόσοι είναι οι μαθητές του Γ2;

4. Δίνεται ο πίνακας:

| Κλάσεις | Κέντρο κλάσης ($x_i$) | $\nu_i$ | $x_i\nu_i$ | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ | $\nu_i (x_i - \bar{x})^2$ |
|---|---|---|---|---|---|---|
| [4, 6) | | 7 | | | | |
| [6, 8) | | 13 | | | | |
| [8, 10) | | 17 | | | | |
| [10, 12) | | 18 | | | | |
| [12, 14) | | 29 | | | | |
| [14, 16) | | 11 | | | | |
| [16, 18) | | 5 | | | | |
| | **ΣΥΝΟΛΑ** | 100 | | | | |

α) Να συμπληρώσετε τον πίνακα.
β) Να υπολογίσετε: i) Τη μέση τιμή. ii) Τη διακύμανση.
ii) Την τυπική απόκλιση iii)Το συντελεστή μεταβολής.

5. Οι μηνιαίες αποδοχές ενός δείγματος 70 υπαλλήλων ενός οργανισμού δίνονται στον επόμενο πίνακα:

| Αποδοχές σε χιλιάδες δρχ. | Κεντρικές τιμές $x_i$ | $v_i$ | $x_i^2$ | $x_i\ v_i$ | $x_i^2\ v_i$ |
|---|---|---|---|---|---|
| [30, 35) | | 8 | | | |
| [35, 40) | | 10 | | | |
| [40, 45) | | 16 | | | |
| [45, 50) | | 15 | | | |
| [50, 55) | | 10 | | | |
| [55, 60) | | 8 | | | |
| [60, 65) | | 3 | | | |
| ΣΥΝΟΛΑ | | 70 | | | |

α) Να συμπληρώσετε τον πίνακα.

β) Να υπολογίσετε: i)Τη μέση τιμή. ii)Τη διακύμανση. iii)Την τυπική απόκλιση iv) Το συντελεστή μεταβολής.

6. α) Να συμπληρώσετε τον πίνακα στον οποίο παρουσιάζονται οι απουσίες 75 μαθητών μιας τάξης ενός Λυκείου, αν γνωρίζουμε ότι ο μέσος όρος των απουσιών είναι 12.

| Απουσίες $x_i$ | *Μαθητές* $v_i$ |
|---|---|
| 10 | x |
| 20 | y |
| 30 | 5 |
| | 75 |

β) Να υπολογίσετε: i) Τη διακύμανση $s^2$. ii) Το συντελεστή μεταβλητότητας (CV).

7. Η μέση τιμή και η διακύμανση των 5 τιμών ενός δείγματος είναι $\bar{x} = 4$ και $s^2 = 10$, αντίστοιχα. Εάν, για τις τέσσερις τιμές ισχύει $\sum_{i=1}^{4}(x_i - \bar{x})^2 = 14$, να βρεθεί η πέμπτη τιμή.

**8.** Στο διπλανό πίνακα δίνεται η κατανομή της ηλικίας των ατόμων μιας πόλης. Να υπολογίσετε:

α)Την τυπική απόκλιση και το συντελεστή μεταβολής.

β)Το ημιενδοτεταρτημοριακό εύρος.

| Ηλικία (σε έτη) | Συχνότητα (σε χιλιάδες) |
|---|---|
| 0-20 | 12 |
| 20-40 | 14 |
| 40-60 | 20 |
| 60-80 | 10 |
| 80-100 | 4 |