ΣΤΑΤΙΣΤΙΚΗ

ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ

Πληθυσμός λέγεται ένα σύνολο που θέλουμε να εξετάσουμε τα στοιχεία του ως προς ένα ή περισσότερα χαρακτηριστικά τους.

Μεταβλητές λέγονται τα χαρακτηριστικά ως προς τα οποία εξετάζουμε έναν πληθυσμό και τις συμβολίζουμε συνήθως με τα κεφαλαία γράμματα $X, Y, Z, B, \dots$

Τιμές της μεταβλητής λέγονται οι δυνατές τιμές που μπορεί να πάρει μια μεταβλητή.

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).
2. Σε **ποσοτικές** μεταβλητές, των οποίων οι τιμές είναι αριθμοί και διακρίνονται:

i) Σε **διακριτές** 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 αποιαδήποτε τιμή ενός διαστήματος πραγματικών αριθμών (α, β). 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

Απογραφή καλείται η μέθοδος συλλογής των δεδομένων όπου εξετάζουμε όλα τα άτομα (στοιχεία) του πληθυσμού ως προς το χαρακτηριστικό που μας ενδιαφέρει.

Δείγμα καλείται ένα υποσύνολο του πληθυσμού.

ΠΑΡΟΥΣΙΑΣΗ ΣΤΑΤΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Πίνακες Κατανομής Συχνοτήτων

Ας υποθέσουμε ότι $x_1, x_2, \dots, x_\kappa$ είναι οι τιμές μιας μεταβλητής **Χ**, που αφορά τα άτομα ενός δείγματος μεγέθους ***ν***, $\kappa \le \nu$.

Συχνότητα (απόλυτη) ν_i της τιμής x_i είναι ο φυσικός αριθμός που δείχνει πόσες φορές εμφανίζεται η τιμή x_i της εξεταζόμενης μεταβλητής **Χ** στο σύνολο των παρατηρήσεων.

Είναι φανερό ότι το άθροισμα όλων των συχνοτήτων είναι ίσο με το μέγεθος *ν* του δείγματος, δηλαδή:

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

Η σχετική συχνότητα f_i της τιμής x_i, προκύπτει αν διαιρέσουμε τη συχνότητα ν_i με το μέγεθος *ν* του δείγματος, δηλαδή

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

Για τη σχετική συχνότητα ισχύουν οι ιδιότητες:

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

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

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

Οι ποσότητες x_i, ν_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, ν_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**.**

Αθροιστικές Συχνότητες

Στην περίπτωση των **ποσοτικών μεταβλητών εκτός** από τις συχνότητες ν_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 $x_1, x_2, ..., x_\kappa$ μιας ποσοτικής μεταβλητής X είναι σε αύξουσα διάταξη, τότε η αθροιστική συχνότητα της τιμής x_i είναι $N_i = \nu_1 + \nu_2 + ... + \nu_i$.
Όμοια, η αθροιστική σχετική συχνότητα είναι $F_i = f_1 + f_2 + ... + f_i$, για $i = 1,2,...,\kappa$.
Είναι φανερό ότι ισχύουν οι σχέσεις:
$\nu_1 = N_1,\ \nu_2 = N_2 - N_1, ..., \nu_\kappa = N_\kappa - N_{\kappa-1}$
και
$f_1 = F_1,\ f_2 = F_2 - F_1, ..., f_\kappa = F_\kappa - F_{\kappa-1}$.

Γραφική Παράσταση Κατανομής Συχνοτήτων

Ποιοτικές μεταβλητές

α) Ραβδόγραμμα

Σε κάθε τιμή της μεταβλητής X αντιστοιχεί μια ορθογώνια στήλη της οποίας το ύψος είναι ίσο με την αντίστοιχη συχνότητα ή σχετική συχνότητα.

Ποσοτικές μεταβλητές

β) Διάγραμμα Συχνοτήτων

Αυτό μοιάζει με το ραβδόγραμμα με μόνη διαφορά ότι αντί να χρησιμοποιούμε συμπαγή ορθογώνια υψώνουμε σε κάθε x_i (υποθέτοντας ότι $x_1 < x_2 < ... < x_\kappa$) μία κάθετη γραμμή με μήκος ίσο προς την αντίστοιχη συχνότητα. Μπορούμε επίσης αντί των συχνοτήτων ν_i στον κάθετο άξονα να βάλουμε τις σχετικές συχνότητες f_i, οπότε έχουμε το **διάγραμμα σχετικών συχνοτήτων**.
Ενώνοντας τα σημεία (x_i, ν_i) ή (x_i, f_i) έχουμε το λεγόμενο **πολύγωνο συχνοτήτων** ή **πολύγωνο σχετικών συχνοτήτων**.

γ) 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.
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 ν_i ή τις σχετικές συχνότητες f_i των τιμών x_i της μεταβλητής.
Αν συμβολίσουμε με α_i το αντίστοιχο τόξο ενός κυκλικού τμήματος στο κυκλικό διάγραμμα συχνοτήτων, τότε

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

Ομαδοποίηση των Παρατηρήσεων

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 [,). Οι παρατηρήσεις κάθε κλάσης θεωρούνται όμοιες, οπότε μπορούν να "αντιπροσωπευθούν" από τις **κεντρικές τιμές**, τα κέντρα δηλαδή κάθε κλάσης.

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

Μέγεθος δείγματος ν	Αριθμός κλάσεων κ	Μέγεθος δείγματος ν	Αριθμός κλάσεων κ
< 20	5	$200-400$	9
$20-50$	6	$400-700$	10
$50-100$	7	$700-1000$	11
$100-200$	8	≥ 1000	12

- **Πλάτος c μιας κλάσης** ονομάζεται η διαφορά του κατωτέρου από το ανώτερο όριο της κλάσης. Στην πλειονότητα των πρακτικών εφαρμογών οι κλάσεις έχουν το ίδιο πλάτος. Για να κατασκευάσουμε ισοπλατείς

κλάσεις, χρησιμοποιούμε το **εύρος R** του δείγματος, δηλαδή τη διαφορά της μικρότερης παρατήρησης από τη μεγαλύτερη παρατήρηση του συνολικού δείγματος.
Υπολογίζουμε το πλάτος $c = R / \kappa$ στρογγυλεύοντας, αν χρειαστεί για λόγους διευκόλυνσης, πάντα προς τα πάνω.
Πρέπει να προσεχτεί ότι:

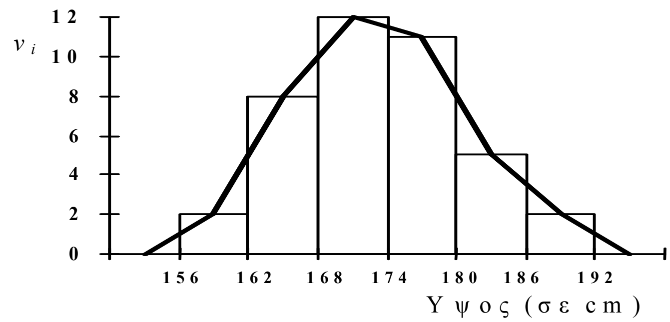
- Καμία παρατήρηση δεν μπορεί να μείνει έξω από κάποια κλάση.
- Οι κεντρικές τιμές διαφέρουν μεταξύ τους όσο και το πλάτος των κλάσεων, που εδώ είναι ίσο με 6.
- Μία παρατήρηση που συμπίπτει με το άνω άκρο μιας κλάσης θα τοποθετηθεί κατά τη διαλογή στην αμέσως επόμενη κλάση. Για παράδειγμα, ο μαθητής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

Ιστόγραμμα Συχνοτήτων

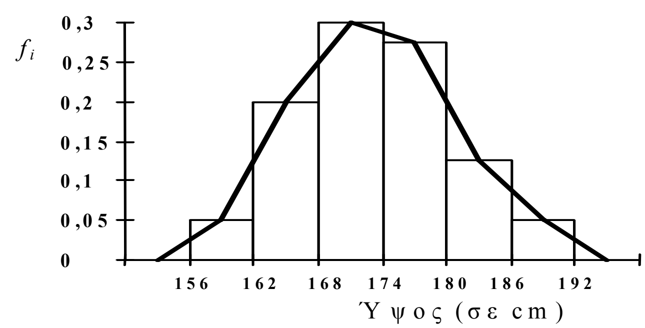
Η αντίστοιχη γραφική παράσταση ενός πίνακα συχνοτήτων με ομαδοποιημένα δεδομένα γίνεται με το λεγόμενο **ιστόγραμμα** συχνοτήτων.
Στον οριζόντιο άξονα ενός συστήματος ορθογωνίων αξόνων σημειώνουμε, με κατάλληλη κλίμακα, τα όρια των κλάσεων. Στη συνέχεια, κατασκευάζουμε διαδοχικά ορθογώνια (ιστούς), από καθένα από τα οποία έχει βάση ίση με το πλάτος της κλάσης και ύψος τέτοιο, ώστε το **εμβαδόν του ορθογωνίου να ισούται με τη συχνότητα της κλάσης αυτής**.

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, έτσι ώστε να ισχύει πάλι ότι το εμβαδόν των ορθογωνίων είναι ίσο με τις αντίστοιχες συχνότητες.
Επομένως, στον κατακόρυφο άξονα σε ένα ιστόγραμμα συχνοτήτων βάζουμε τις συχνότητες.
Με ανάλογο τρόπο κατασκευάζεται και το **ιστόγραμμα σχετικών συχνοτήτων**, οπότε στον κάθετο άξονα βάζουμε τις σχετικές συχνότητες.
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**.

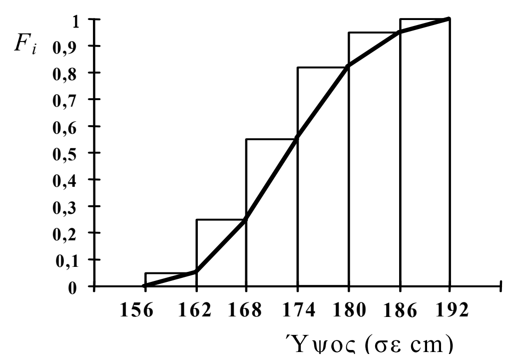


Το **εμβαδόν** του χωρίου που ορίζεται από το πολύγωνο συχνοτήτων και τον οριζόντιο άξονα είναι **ίσο με το άθροισμα των συχνοτήτων, δηλαδή με το μέγεθος του δείγματος *ν*.**



Από το ιστόγραμμα σχετικών συχνοτήτων και το **πολύγωνο σχετικών συχνοτήτων με εμβαδόν ίσο με 1.**

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**



ΜΕΤΡΑ ΘΕΣΗΣ

Τα μέτρα Θέσης χρησιμοποιούνται για την περιγραφή της θέσης ενός συνόλου δεδομένων πάνω στον οριζόντιο άξονα ox, εκφράζοντας την "κατά μέσο όρο" απόστασή τους από την **αρχή των αξόνων.**

α) Μέση Τιμή $(\bar{x})$

Η μέση τιμή $\bar{x}$ ενός συνόλου **ν** παρατηρήσεων ορίζεται ως το άθροισμα των παρατηρήσεων διά του πλήθους των παρατηρήσεων.

Όταν σε ένα δείγμα μεγέθους **ν** οι παρατηρήσεις μιας μεταβλητής **X** είναι $\mathbf{t_1, t_2, ..., t_\nu}$, τότε η μέση τιμή συμβολίζεται με $\bar{\mathbf{x}}$ και δίνεται από τη σχέση:

$$\bar{x} = \frac{t_1 + t_2 + ... + t_\nu}{\nu} = \frac{\sum_{i=1}^{\nu} t_i}{\nu} = \frac{1}{\nu}\sum_{i=1}^{\nu} t_i$$

όπου το σύμβολο $\sum_{i=1}^{\nu} t_i = t_1 + t_2 + ... + t_\nu$ και διαβάζεται "**άθροισμα των t_i από $i = 1$ έως ν**".

Συχνά, όταν δεν υπάρχει πρόβλημα σύγχυσης, συμβολίζεται και ως $\sum t_i$ ή ακόμα πιο απλά με $\sum t$.

Σε μια κατανομή συχνοτήτων, αν $\mathbf{x_1, x_2, ..., x_\kappa}$ είναι οι τιμές της μεταβλητής **X** με συχνότητες $\mathbf{\nu_1, \nu_2, ..., \nu_\kappa}$ αντίστοιχα, η μέση τιμή ορίζεται ισοδύναμα από τη σχέση:

$$\bar{x} = \frac{x_1\nu_1 + x_2\nu_2 + ... + x_\kappa\nu_\kappa}{\nu_1 + \nu_2 + ... + \nu_\kappa} = \frac{\sum_{i=1}^{\kappa} x_i\nu_i}{\sum_{i=1}^{\kappa} \nu_i} = \frac{1}{\nu}\sum_{i=1}^{\kappa} x_i\nu_i$$

Η παραπάνω σχέση ισοδύναμα γράφεται: $\bar{x} = \sum_{i=1}^{\kappa} x_i \frac{\nu_i}{\nu} = \sum_{i=1}^{\kappa} x_i f_i$ όπου f_i οι σχετικές συχνότητες.

Η **μέση τιμή επηρεάζεται** από τις **ακραίες** παρατηρήσεις

β) Σταθμικός Μέσος

Στις περιπτώσεις που δίνεται **διαφορετική βαρύτητα** στις τιμές $\mathbf{x_1, x_2, ..., x_\nu}$ ενός συνόλου δεδομένων, τότε αντί του αριθμητικού μέσου χρησιμοποιούμε τον **σταθμισμένο αριθμητικό μέσο** ή **σταθμικό μέσο**.

Εάν σε κάθε τιμή $\mathbf{x_1, x_2, ..., x_\nu}$ δώσουμε διαφορετική βαρύτητα, που εκφράζεται με τους λεγόμενους συντελεστές στάθμισης (βαρύτητας) $\mathbf{w_1, w_2, ..., w_\nu}$, τότε ο σταθμικός μέσος βρίσκεται από τον τύπο:

$$\bar{x} = \frac{x_1w_1 + x_2w_2 + ... + x_\nu w_\nu}{w_1 + w_2 + ... + w_\nu} = \frac{\sum_{i=1}^{\nu} x_i w_i}{\sum_{i=1}^{\nu} w_i}$$

.γ) Διάμεσος (δ)

Διάμεσος **(δ)** ενός δείγματος **ν** παρατηρήσεων οι οποίες έχουν διαταχθεί σε <u>**αύξουσα**</u> σειρά ορίζεται ως :

1. Η **μεσαία παρατήρηση**, όταν το **ν** είναι **περιττός** αριθμός $\delta = t_{\frac{\nu+1}{2}}$
2. Ο μέσος όρος (ημιάθροισμα) των δύο μεσαίων παρατηρήσεων όταν το **ν** είναι **άρτιος** αριθμός $\delta = \dfrac{t_{\frac{\nu}{2}} + t_{\frac{\nu}{2}+1}}{2}$

Η **διάμεσος δεν επηρεάζεται** από τις **ακραίες** παρατηρήσεις

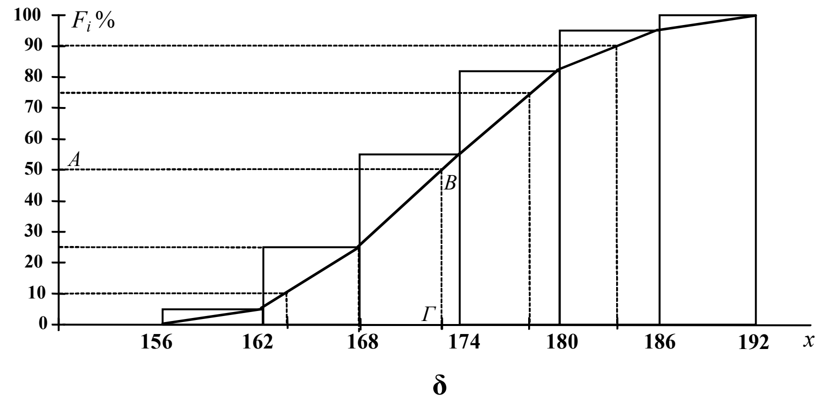
Η **διάμεσος** είναι η τιμή που χωρίζει ένα σύνολο παρατηρήσεων σε **δύο ίσα μέρη** όταν οι παρατηρήσεις αυτές τοποθετηθούν με σειρά τάξης μεγέθους.

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.

Διάμεσος σε Ομαδοποιημένα Δεδομένα

Η διάμεσος, όπως ορίστηκε, αντιστοιχεί στην τιμή $\mathbf{x = \delta}$ της μεταβλητής **X** (στον οριζόντιο άξονα), έτσι ώστε το **50%** των παρατηρήσεων να είναι μικρότερες ή ίσες του **δ.**

Δηλαδή, η διάμεσος θα έχει αθροιστική σχετική συχνότητα $F_i = 50\%$. 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 Γ αντιστοιχεί η διάμεσος δ των παρατηρήσεων. Δηλαδή, $\delta \approx 173$.



ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ

Μέτρα διασποράς λέγονται τα μέτρα που εκφράζουν τις αποκλίσεις των τιμών μιας μεταβλητής γύρω από τα μέτρα κεντρικής τάσης.

α) Εύρος (R)

Το **εύρος** ή **κύμανση** (R), που ορίζεται ως η διαφορά της ελάχιστης παρατήρησης από τη μέγιστη παρατήρηση, δηλαδή:

Εύρος R = Μεγαλύτερη παρατήρηση-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.

γ) Διακύμανση (s^2)

Ένας άλλος τρόπος για να υπολογίσουμε τη διασπορά των παρατηρήσεων $\mathbf{t_1, t_2, ..., t_\nu}$ μιας μεταβλητής **X** θα ήταν να αφαιρέσουμε τη μέση τιμή $\mathbf{\bar{x}}$ από κάθε παρατήρηση και να βρούμε τον αριθμητικό μέσο των διαφορών αυτών, δηλαδή τον αριθμό:

$$\frac{(\mathbf{t_1}-\bar{\mathbf{x}})+(\mathbf{t_2}-\bar{\mathbf{x}})+...+(\mathbf{t_\nu}-\bar{\mathbf{x}})}{\nu} = \frac{\sum_{i=1}^{\nu}(\mathbf{t_i}-\bar{\mathbf{x}})}{\nu} = \frac{\mathbf{t_1}+\mathbf{t_2}+...+\mathbf{t_\nu}}{\nu} - \frac{\nu\bar{\mathbf{x}}}{\nu} = \bar{\mathbf{x}} - \bar{\mathbf{x}} = \mathbf{0}.$$

Ο αριθμός όμως αυτός είναι ίσος με μηδέν, γι' αυτό, ως ένα μέτρο διασποράς παίρνουμε τον **μέσο όρο** των **τετραγώνων** των **αποκλίσεων** των $\mathbf{t_i}$ από τη μέση τιμή τους $\bar{\mathbf{x}}$.

Το μέτρο αυτό καλείται **διακύμανση** ή **διασπορά** και ορίζεται $\mathbf{s^2 = \frac{1}{\nu}\sum_{i=1}^{\nu}(t_i-\bar{x})^2}$

Όταν η μέση τιμή $\bar{x}$ δεν είναι ακέραιος αριθμός ο τύπος γίνεται: $\mathbf{s^2 = \frac{1}{\nu}\left\{\sum_{i=1}^{\nu}t_i^2 - \frac{\left(\sum_{i=1}^{\nu}t_i\right)^2}{\nu}\right\}}$

Όταν έχουμε **πίνακα συχνοτήτων ή ομαδοποιημένα** δεδομένα, $s^2 = \frac{1}{\nu}\sum_{i=1}^{\kappa}(x_i-\bar{x})^2\nu_i$

ή την ισοδύναμη μορφή: $\mathbf{s^2 = \dfrac{1}{\nu}\left\{\sum_{i=1}^{\kappa} x_i^2 \nu_i - \dfrac{\left(\sum_{i=1}^{\kappa} x_i \nu_i\right)^2}{\nu}\right\}}$

όπου $\mathbf{x_1, x_2, ..., x_\kappa}$ οι τιμές της μεταβλητής (ή τα κέντρα των κλάσεων) με αντίστοιχες συχνότητες $\mathbf{\nu_1, \nu_2, ..., \nu_\kappa}$.

δ) 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 cm^2.

Τυπική απόκλιση λέγεται η θετική τετραγωνική ρίζα της διακύμανσης, και εκφράζεται με την ίδια μονάδα μέτρησης του χαρακτηριστικού, όπως ακριβώς είναι και όλα τα άλλα μέτρα θέσης, που εξετάσαμε έως τώρα.

$s = \sqrt{s^2}$

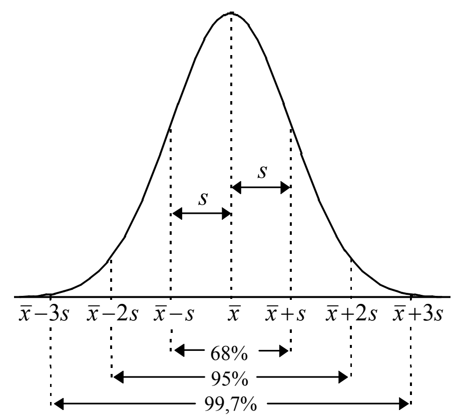
Αξίζει να σημειωθεί ότι αν η καμπύλη συχνοτήτων για το χαρακτηριστικό που εξετάζουμε είναι κανονική ή περίπου κανονική, τότε η τυπική απόκλιση s έχει τις παρακάτω ιδιότητες:

i) το 68% περίπου των παρατηρήσεων βρίσκεται στο διάστημα $(\bar{x} - s,\ \bar{x} + s)$ 15

ii) το 95% περίπου των παρατηρήσεων βρίσκεται στο διάστημα $(\bar{x} - 2s,\ \bar{x} + 2s)$

iii) το 99,7% περίπου των παρατη-ρήσεων βρίσκεται στο διάστημα $(\bar{x} - 3s,\ \bar{x} + 3s)$

iv) το εύρος ισούται περίπου με έξι τυπικές αποκλίσεις, δηλαδή $R \approx 6s$.



Συντελεστής Μεταβολής (CV)

Ένα μέτρο με το οποίο μπορούμε να συγκρίνουμε ομάδες 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:

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

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.

Γενικά δεχόμαστε ότι ένα δείγμα τιμών μιας μεταβλητής θα είναι ομοιογενές, εάν ο συντελεστής μεταβολής **δεν ξεπερνά το 10%.**

ΕΦΑΡΜΟΓΕΣ

Έστω $x_1, x_2, ..., x_\nu$ ν παρατηρήσεις με μέση τιμή $\bar{x}$ και τυπική απόκλιση s_x.

α) Αν $y_1, y_2, ..., y_\nu$ είναι οι παρατηρήσεις που προκύπτουν αν προσθέσουμε σε καθεμιά από τις $x_1, x_2, ..., x_\nu$ μια σταθερά c, να δειχτεί ότι:

i) $\bar{y} = \bar{x} + c$, ii) $s_y = s_x$

β) Αν $y_1, y_2, ..., y_\nu$ είναι οι παρατηρήσεις που προκύπτουν αν πολλαπλασιάσουμε τις $x_1, x_2, ..., x_\nu$ επί μια σταθερά c, να αποδειχτεί ότι:

i) $\bar{y} = c\,\bar{x}$, ii) $s_y = |c|\, s_x$