ΜΑΘΗΜΑΤΙΚΑ

Γ

Μαθηματικά Ι

ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

ΕΠΑ.Λ.

Ε π ι μ έ λ ε ι α

Κ Ο Λ Λ Α Σ   Α Ν Τ Ω Ν Η Σ

Βασικές έννοιες & ορισμοί

Στατιστική
είναι ο κλάδος των μαθηματικών, ο οποίος ως έργο έχει τη συγκέντρωση στοιχείων, την ταξινόμησή τους και την παρουσίασή τους σε κατάλληλη μορφή, ώστε να μπορούν να αναλυθούν και να ερμηνευθούν για την εξυπηρέτηση διαφόρων σκοπών.

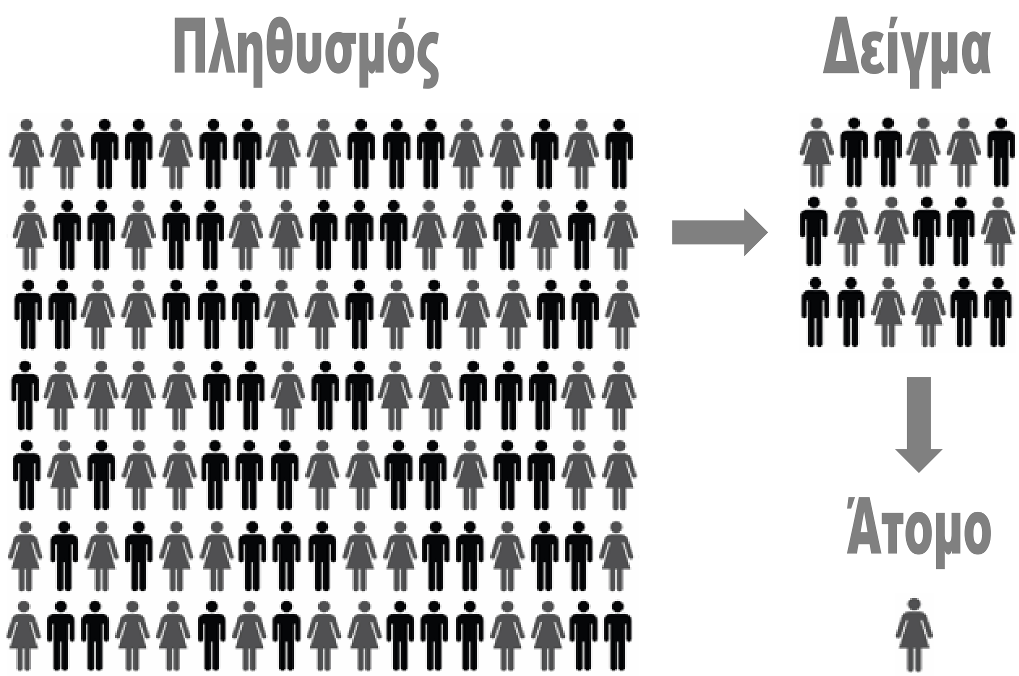
Πληθυσμός
είναι το σύνολο των αντικειμένων (έμψυχων ή άψυχων) για τα οποία συλλέγονται στοιχεία.

Άτομο
ονομάζεται κάθε στοιχείο ενός πληθυσμού ή ενός δείγματος.

Δείγμα
είναι ένα μέρος (υποσύνολο) του πληθυσμού, που είναι αντιπροσωπευτικό του πληθυσμού και από την εξέταση του οποίου βγάζουμε συμπεράσματα για ολόκληρο τον πληθυσμό.

Δειγματοληψία
είναι η εξέταση ενός δείγματος, κάποιου πληθυσμού.

Μέγεθος (ν)
ενός πληθυσμού (ή ενός δείγματος) ονομάζεται το πλήθος των ατόμων του.



Μεταβλητή

είναι το χαρακτηριστικό ενός πληθυσμού, ως προς το οποίο αυτός εξετάζεται.

Οι μεταβλητές διακρίνονται στις εξής κατηγορίες:

Ποσοτικές μεταβλητές

είναι εκείνες που μπορούν να μετρηθούν, πχ. ύψος, μισθός, ώρες εργασίας, τιμή, κλπ.

Ποιοτικές μεταβλητές

είναι εκείνες που δεν επιδέχονται μέτρηση, πχ. χρώμα ματιών, μόρφωση, θρήσκευμα, κλπ.

Διακριτές μεταβλητές

είναι εκείνες, στις οποίες κάθε άτομο του πληθυσμού μπορεί να πάρει μόνο διακεκριμένες τιμές, πχ. αριθμός παιδιών, μέρες διακοπών, κλπ.

Συνεχείς μεταβλητές

είναι εκείνες, στις οποίες κάθε άτομο του πληθυσμού μπορεί να πάρει οποιαδήποτε πραγματική τιμή, που ανήκει σε διάστημα (ή ένωση διαστημάτων) πραγματικών αριθμών, πχ. ύψος, βάρος, κλπ.

Συχνότητες παρατηρήσεων

Συχνότητα (ν_i)

της τιμής x_i μιας μεταβλητής Χ ονομάζεται το πλήθος των ατόμων του πληθυσμού (ή του δείγματος) για τα οποία η μεταβλητή παίρνει την τιμή x_i και συμβολίζεται με ν_i.

Προφανώς, αν η μεταβλητή Χ παρουσιάζει **κ** διαφορετικές τιμές $\mathbf{x_i}$ με αντίστοιχες συχνότητες $\boldsymbol{\nu_i}$, τότε:

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

Σχετική συχνότητα (f_i)

της τιμής $\mathbf{x_i}$ μιας μεταβλητής **Χ** ονομάζεται ο λόγος της συχνότητας προς το μέγεθος του δείγματος και συμβολίζεται με $\mathbf{f_i}$.

$$\boxed{f_i = \frac{\nu_i}{\nu}}$$

Ισχύει ότι:

$$f_1 + f_2 + \ldots + f_\kappa = 1$$

ΠΡΟΣΟΧΗ !!! Αν η τιμή f_i είναι γνωστή και χρειάζεται να τη συμπληρώσουμε στον τύπο (γιατί πχ. αναζητούμε κάποιο απ' τα ν_i ή ν) τότε **ΔΕΝ** χρησιμοποιούμε την τιμή **επί τοις εκατό (%)**, αλλά το δεκαδικό αριθμό < 1 που βρίσκεται στη στήλη f_i. Αν η τελευταία δεν υπάρχει, τότε διαιρούμε απλώς την τιμή της στήλης % με το 100.

Αθροιστική συχνότητα (N_i)

της τιμής x_i μιας ποσοτικής μεταβλητής Χ λέγεται το άθροισμα των συχνοτήτων ν_i των τιμών που είναι μικρότερες ή ίσες με την τιμή αυτή.

Σχετική αθροιστική συχνότητα (F_i)

της τιμής $\mathbf{x_i}$ μιας ποσοτικής μεταβλητής **Χ** λέγεται το άθροισμα των σχετικών συχνοτήτων $\mathbf{f_i}$ των τιμών που είναι μικρότερες ή ίσες με την τιμή αυτή.

Ισχύει επίσης ότι:

$$\boxed{F_i = \frac{N_i}{\nu}}$$

Παράμετροι θέσης

Επικρατούσα τιμή (M_o)

μιας μεταβλητής ονομάζεται η τιμή με τη μεγαλύτερη συχνότητα. Αν δύο ή περισσότερες τιμές έχουν τη μέγιστη συχνότητα τότε υπάρχουν περισσότερες από μία επικρατούσες τιμές.

Μέση τιμή ($\overline{X}$)

διαφόρων τιμών είναι το πηλίκο του αθροίσματος των τιμών προς το πλήθος τους.

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

Αν οι μεταβλητές είναι ταξινομημένες σε πίνακα συχνοτήτων, τότε καταλληλότερος είναι ο παρακάτω τύπος:

$$\overline{X} = \frac{\sum_{i=1}^{\nu} x_i \cdot \nu_i}{\nu} = \frac{\nu_1 x_1 + \nu_2 x_2 + \ldots + \nu_\nu x_\nu}{\nu}$$

Διάμεσος (δ)

ενός δείγματος ν παρατηρήσεων που έχουν διαταχθεί σε αύξουσα σειρά ονομάζεται:

- Η **μεσαία** παρατήρηση αν το πλήθος των παρατηρήσεων είναι περιττό.

πχ. x_1 x_2 x_3 x_4 x_5 x_6 x_7

- Το **ημιάθροισμα** των μεσαίων παρατηρήσεων αν το πλήθος των παρατηρήσεων είναι άρτιο.

πχ. x_1 x_2 x_3 x_4 x_5 x_6

$$\delta = \frac{x_3 + x_4}{2}$$

Παράμετροι διασποράς

Εύρος

των τιμών μιας μεταβλητής είναι η διαφορά της μικρότερης τιμής από τη μεγαλύτερη.

$$\mathbf{R = x_{max} - x_{min}}$$

Διακύμανση (s^2)

μιας μεταβλητής **X** που παίρνει **ν** το πλήθος τιμές $\mathbf{x_1, x_2, \dots, x_\nu}$ με μέση τιμή $\overline{X}$ ονομάζεται το πηλίκο:

$$s^2 = \frac{\sum_{i=1}^{\nu}(\overline{X} - x_i)^2}{\nu} = \frac{(\overline{X} - x_1)^2 + (\overline{X} - x_2)^2 + \dots + (\overline{X} - x_\nu)^2}{\nu}$$

Αν οι μεταβλητές είναι ταξινομημένες σε πίνακα συχνοτήτων, τότε καταλληλότερος είναι ο παρακάτω τύπος:

$$s^2 = \frac{\sum_{i=1}^{\nu}\nu_i(\overline{X} - x_i)^2}{\nu} = \frac{\nu_1(\overline{X} - x_1)^2 + \nu_2(\overline{X} - x_2)^2 + \dots + \nu_\kappa(\overline{X} - x_\kappa)^2}{\nu}$$

Τυπική απόκλιση (s)

μιας μεταβλητής **X** που παίρνει **ν** το πλήθος τιμές $\mathbf{t_1, t_2, \dots, t_\nu}$ με μέση τιμή $\overline{X}$ ονομάζεται το:

$$s = \sqrt{\frac{\nu_1(\overline{X} - x_1)^2 + \nu_2(\overline{X} - x_2)^2 + \dots + \nu_\kappa(\overline{X} - x_\kappa)^2}{\nu}}$$

Με άλλα λόγια, η τυπική απόκλιση είναι η τετραγωνική ρίζα της διακύμανσης:

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

Συντελεστής μεταβολής ή μεταβλητότητας (CV)

μιας ποσοτικής μεταβλητής **X** που παρουσιάζει μέση τιμή $\bar{X}$ και τυπική απόκλιση **s** ονομάζεται το πηλίκο:

$$CV = \frac{s}{\bar{X}} \cdot 100\% = \frac{\text{Τυπική απόκλιση}}{\text{Μέση τιμή}}$$

Αν **CV < 10%** τότε λέμε ότι ο πληθυσμός (ή το δείγμα) είναι **ομοιογενής**.
Αν **CV ≥ 10%** τότε λέμε ότι ο πληθυσμός (ή το δείγμα) είναι **ανομοιογενής**.

Εφαρμογή 1 - Πίνακας συχνοτήτων

Η απλούστερη άσκηση είναι να μας δίνουν ένα σύνολο δεδομένων, συνήθως αριθμητικές μετρήσεις κάποιου μεγέθους, και να μας ζητούν να φτιάξουμε έναν πίνακα συχνοτήτων. Παρακάτω θα δούμε, καταρχάς, πώς κατασκευάζεται ένας τέτοιος πίνακας και, κατά δεύτερο λόγο, τι είδους ερωτήσεις μας βοηθάει αυτός ν' απαντήσουμε.

Ας υποθέσουμε ότι τα παρακάτω δεδομένα δίνουν τις μέσες θερμοκρασίες μιας πόλης, κατά τη διάρκεια 20 ημερών:

15	**15**	**18**	**17**	**19**	**10**	**17**	**16**	**14**	**17**
18	**19**	**16**	**15**	**17**	**18**	**15**	**14**	**16**	**17**

Στόχος μας είναι να παρουσιάσουμε τα δεδομένα σε έναν πίνακα συχνοτήτων, στον οποίο να περιλαμβάνονται επιπλέον οι σχετικές συχνότητες, οι αθροιστικές συχνότητες και οι σχετικές αθροιστικές συχνότητες.

Πίνακας συχνοτήτων

Τιμές x_i	Συχνότητα ν_i
10	1
14	2
15	4
16	3
17	5
18	3
19	2
Σύνολο	**20**

- Μετράμε πόσες φορές εμφανίζεται ο αριθμός 10
- Μετράμε πόσες φορές εμφανίζεται ο αριθμός 14
- Μετράμε πόσες φορές εμφανίζεται ο αριθμός 15
- Μετράμε πόσες φορές εμφανίζεται ο αριθμός 16
- Μετράμε πόσες φορές εμφανίζεται ο αριθμός 17
- Μετράμε πόσες φορές εμφανίζεται ο αριθμός 18
- Μετράμε πόσες φορές εμφανίζεται ο αριθμός 19

Προφανώς, το άθροισμα όλων των συχνοτήτων μας δίνει το μέγεθος (ν) του δείγματος.

Σχετική συχνότητα

Για να συμπληρώσουμε τη στήλη των σχετικών συχνοτήτων, γενικά, εφαρμόζουμε τον τύπο που είναι γνωστός από τη θεωρία, έτσι:

$$\mathbf{f_1} = \frac{\mathbf{v_1}}{\mathbf{v}} = \frac{1}{20} = 0{,}05 \qquad \mathbf{f_2} = \frac{\mathbf{v_2}}{\mathbf{v}} = \frac{2}{20} = 0{,}10 \qquad \mathbf{f_3} = \frac{\mathbf{v_3}}{\mathbf{v}} = \frac{4}{20} = 0{,}20 \qquad \text{κλπ...}$$

Τιμές x_i	Συχνότητα v_i	Σχετική συχνότητα f_i	
10	1	0,05	$f_1 = v_1 / v = 1 / 20 = 0,05$
14	2	0,10	$f_2 = v_2 / v = 2 / 20 = 0,05$
15	4	0,20	$f_3 = v_3 / v = 4 / 20 = 0,05$
16	3	0,15	$f_4 = v_4 / v = 3 / 20 = 0,05$
17	5	0,25	$f_5 = v_5 / v = 5 / 20 = 0,05$
18	3	0,15	$f_6 = v_6 / v = 3 / 20 = 0,05$
19	2	0,10	$f_7 = v_7 / v = 2 / 20 = 0,05$
Σύνολο	**20**	**1**	

Συνήθως, διευκολύνει ιδιαίτερα την ανάλυσή μας αν στον πίνακα προσθέσουμε και μια στήλη επιπλέον, με τη σχετική συχνότητα **επί τοις εκατό (%)**. Ο πίνακας τότε γίνεται:

Τιμές x_i	Συχνότητα v_i	Σχετική συχνότητα f_i	f_i (%)	
10	1	0,05	5	$= f_1 \cdot 100$
14	2	0,10	10	$= f_2 \cdot 100$
15	4	0,20	20	$= f_3 \cdot 100$
16	3	0,15	15	$= f_4 \cdot 100$
17	5	0,25	25	$= f_5 \cdot 100$
18	3	0,15	15	$= f_6 \cdot 100$
19	2	0,10	10	$= f_7 \cdot 100$
Σύνολο	**20**	**1**	**100**	

Πρακτικός τρόπος

Πολύ συχνά – αλλά μόνο αν είναι βολικό – το σύνολο **ν** χωράει ακριβώς, έστω **κ φορές**, στον αριθμό **100**, πχ. το 20 χωράει στο 100 ακριβώς 5 φορές. Τότε μπορούμε να συμπληρώσουμε, εξαιρετικά εύκολα τη στήλη **fi (%)** πολλαπλασιάζοντας κατευθείαν κάθε συχνότητα $\mathbf{v_i}$ με τον αριθμό **κ**. Στη συνέχεια, η συμπλήρωση της στήλης $\mathbf{f_i}$ γίνεται παιχνιδάκι, διαιρώντας απλώς με το 100, κάθε τιμή της **fi (%)**.

x_i	v_i		f_i (%)		f_i
10	1	· 5	5	: 100	0,05
14	2	· 5	10	: 100	0,10
15	4	· 5	20	: 100	0,20
16	3	· 5	15	: 100	0,15
17	5	· 5	25	: 100	0,25
18	3	· 5	15	: 100	0,15
19	2	· 5	10	: 100	0,10
Σύνολο	**20**	· 5	**100**		**1**

Αθροιστική συχνότητα

Αν φτιάξουμε μια νέα στήλη, όπου κάθε σειρά της θα αποτελείται από τα αθροίσματα όλων των προηγούμενων συχνοτήτων, μέχρι και τη σειρά αυτή, τότε θα έχουμε φτιάξει τη στήλη των αθροιστικών συχνοτήτων.

Τιμές x_i	Συχνότητα v_i	N_i
10	1	1
14	2	3
15	4	7
16	3	10
17	5	15
18	3	18
19	2	20
Σύνολο	**20**	

$N_1 = 1$

$N_2 = 1+2 = 3$

$N_3 = 1+2+4 = 7$

$N_4 = 1+2+4+3 = 10$

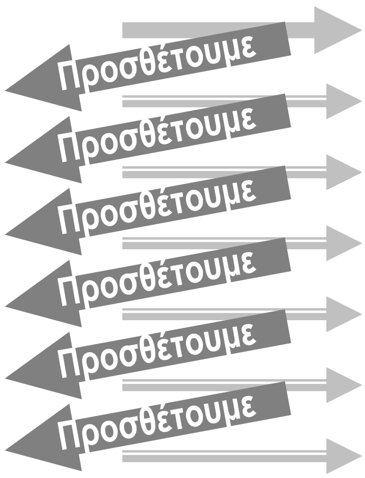
$N_5 = 1+2+4+3+5 = 15$

$N_6 = 1+2+4+3+5+3 = 18$

$N_7 = 1+2+4+3+5+3+2 = 20$

Εναλλακτικά, μπορούμε να σκεφτούμε και ως εξής:

Συχνότητα vi	Αθροιστική Συχνότητα Ni
1	1
2	3
4	7
3	10
5	15
3	18
2	20



Αθροιστική σχετική συχνότητα

Όπως ακριβώς υπολογίζουμε τα διάφορα $\mathbf{N_i}$ με βάση τα $\mathbf{v_i}$, με εντελώς ανάλογο τρόπο υπολογίζουμε τη στήλη $\mathbf{F_i}$ με βάση τα $\mathbf{f_i}$. Συνήθως, συμπληρώνουμε με μια επιπλέον στήλη $\mathbf{F_i}$ **(%)**, δηλαδή με το ποσοστό **επί τοις εκατό** της $\mathbf{F_i}$.

Έτσι τελικά, ο ολοκληρωμένος πίνακας συχνοτήτων έχει ως εξής:

Τιμές x_i	Συχνότητα ν_i	Σχετική συχνότητα f_i	f_i %	Αθροιστική συχνότητα N_i	Αθροιστική σχετική συχνότητα F_i	F_i %
10	1	0,05	5,00	1	0,05	5,00
14	2	0,10	10,00	3	0,15	15,00
15	4	0,20	20,00	7	0,35	35,00
16	3	0,15	15,00	10	0,50	50,00
17	5	0,25	25,00	15	0,75	75,00
18	3	0,15	15,00	18	0,90	90,00
19	2	0,10	10,00	**20**	**1**	**100**
Σύνολο	**20**	**1**	**100**			

Εφαρμογή 2 - Πληροφορίες πίνακα συχνοτήτων

Η κατασκευή του πίνακα συχνοτήτων μας βοηθάει να απαντούμε εύκολα σε μια σειρά από τυπικά ερωτήματα των παρακάτω τύπων:

Συχνότητα	**Αντίστοιχη μαθηματική σχέση**
• ίση με κ	$= \mathbf{\kappa}$
• μικρότερη / λιγότερη / κάτω από κ	$< \mathbf{\kappa}$
• το πολύ / έως και κ	$\leq \mathbf{\kappa}$
• τουλάχιστον κ	$\geq \mathbf{\kappa}$
• μεγαλύτερη / περισσότερη / πάνω από κ	$> \mathbf{\kappa}$
• από κ έως και λ	$\mathbf{\kappa \leq \ldots \leq \lambda}$

Με βάση το προηγούμενο παράδειγμα, λοιπόν, μπορούμε να απαντήσουμε σε ερωτήσεις, όπως «Να βρεθεί πόσες μέρες η μέση θερμοκρασία ήταν...»

α. ίση με 15 βαθμούς.
β. μικρότερη από 15 βαθμούς.
γ. το πολύ 16 βαθμούς.
δ. τουλάχιστον 17 βαθμούς.
ε. μεγαλύτερη από 17 βαθμούς.
στ. μεταξύ 15 και 17 βαθμών (περιλαμβανομένων).
ζ. ζυγός αριθμός.
η. η υψηλότερη.

Λύσεις με τη βοήθεια της στήλης των απλών συχνοτήτων (να προτιμάται)

α. Πηγαίνουμε απευθείας στη στήλη των 15 βαθμών. Η συχνότητά της είναι $\mathbf{\nu = 4}$.

β. Σημαίνει 10 ή 14 βαθμούς, άρα αθροίζουμε τις αντίστοιχες συχνότητες. Θα είναι $\nu = 1 + 2 \Rightarrow \mathbf{\nu = 3}$.

γ. Σημαίνει έως και 16 βαθμούς. Αθροίζοντας τις αντίστοιχες συχνότητες βρίσκουμε $\nu = 1 + 2 + 4 + 3 \Rightarrow \mathbf{\nu = 10}$.

δ. Σημαίνει από 17 βαθμούς και πάνω, άρα $\nu = 5 + 3 + 2 \Rightarrow \mathbf{\nu = 10}$.

ε. Ότι και το δ αλλά οι 17 βαθμοί δεν περιλαμβάνονται, δηλαδή $\nu = 3 + 2 \Rightarrow \mathbf{\nu = 5}$.

στ. Αθροίζουμε τις κατάλληλες συχνότητες. Θα είναι $\nu = 4 + 3 + 5 \Rightarrow \mathbf{\nu = 12}$.

ζ. Μιλάμε για τις θερμοκρασίες των 10, 14, 16 και 18 βαθμών. Από τις αντίστοιχες συχνότητες βρίσκουμε $\nu = 1 + 2 + 3 + 3 \Rightarrow \mathbf{\nu = 9}$.

η. Η υψηλότερη θερμοκρασία είναι εκείνη των 19 βαθμών, άρα $\mathbf{\nu = 2}$.

Λύσεις με τη βοήθεια της στήλης των αθροιστικών συχνοτήτων

Η αθροιστική συχνότητα, επειδή συγκεντρώνει την πληροφορία από όλες τις προηγούμενες τιμές, απαντάει άμεσα σε ερωτήσεις του τύπου *«το πολύ μέχρι»*. Για κάθε άλλη περίπτωση χρειάζονται κατάλληλοι υπολογισμοί.

α. Από την αθροιστική συχνότητα των 15 βαθμών αφαιρούμε τις συχνότητες που αφορούν σε μικρότερες θερμοκρασίες, δηλαδή την προηγούμενη αθροιστική. Άρα $\nu = N(15) - N(14) = 7 - 3 \Rightarrow \boldsymbol{\nu = 4}$.

β. Μικρότερη από 15 βαθμούς σημαίνει έως και 14, άρα $\nu = N(14) \Rightarrow \boldsymbol{\nu = 3}$.

γ. Απευθείας $\nu = N(16) \Rightarrow \boldsymbol{\nu = 10}$.

δ. Εδώ σκεφτόμαστε λίγο διαφορετικά. Από το σύνολο αφαιρούμε όλες εκείνες τις θερμοκρασίες που δεν πληρούν τις προϋποθέσεις, δηλαδή κάτω από 17 βαθμούς. Τελικά $\nu = 20 - N(16) = 20 - 10 \Rightarrow \boldsymbol{\nu = 10}$.

ε. Αναλόγως, από το σύνολο αφαιρούμε τις θερμοκρασίες από 17 και κάτω. Άρα $\nu = 20 - N(17) = 20 - 15 \Rightarrow \boldsymbol{\nu = 5}$.

στ. Από την αθροιστική συχνότητα των 17 βαθμών αφαιρούμε τις θερμοκρασίες που είναι κάτω κι από 15, άρα $\nu = N(17) - N(14) = 15 - 3 \Rightarrow \boldsymbol{\nu = 12}$.

Οι 2 τελευταίες ερωτήσεις δεν είναι δυνατόν να απαντηθούν με τη βοήθεια της αθροιστικής συχνότητας.

Μετατροπή της στήλης των αθροιστικών συχνοτήτων σε απλές

Αν δίνεται μόνο η στήλη των $\mathbf{N_i}$ και η λογική της φαντάζει δυσνόητη, τότε χωρίς πολύ κόπο μπορούμε από αυτήν να υπολογίσουμε την στήλη των απλών $\boldsymbol{\nu_i}$ κι έτσι να δουλέψουμε σύμφωνα με τον πρώτο τρόπο, που είναι ιδιαίτερα απλός. Το μόνο που έχουμε να κάνουμε είναι σε κάθε σειρά να αφαιρούμε το αντίστοιχο $\mathbf{N_i}$ από εκείνο της προηγούμενης σειράς. Προφανώς, ξεκινάμε από τη 2η σειρά, εφόσον ο πρώτος αριθμός παραμένει ο ίδιος.

Τιμές x_i	N_i
10	1
14	3
15	7
16	10
17	15
18	18
19	**20**
Σύνολο	

	Συχνότητα ν_i
	1
3 – 1 =	2
7 – 3 =	4
10 – 7 =	3
15 – 10 =	5
18 – 15 =	3
20 – 18 =	2
	20

Εφαρμογή 3 - Συμπλήρωση πίνακα συχνοτήτων

Σε άλλες ασκήσεις ζητείται να συμπληρωθεί ένας πίνακας συχνοτήτων, ο οποίος μπορεί να περιέχει ελάχιστα μόνο στοιχεία. Αυτό που πρέπει να κατανοήσουμε είναι πως τα δεδομένα όσο λίγα κι αν είναι, σίγουρα είναι όσα χρειαζόμαστε. Το μόνο που χρειάζεται είναι να ανακαλύψουμε τον τρόπο που συνδέονται μεταξύ τους. Ορίστε μερικές χρήσιμες συμβουλές:

Τιμές x_i	**Συχνότητα** ν_i	**Σχετική συχνότητα** f_i	f_i **%**	**Αθροιστική συχνότητα** N_i	**Αθροιστική σχετική συχνότητα** F_i	F_i **%**
x_1				Ίδιο με ν_1	Ίδιο με f_1	
x_2	•	•				
x_3						
x_4				**ν**	**1**	**100**
Σύνολο	**ν**	**1**	**100**			

Εδώ πάντοτε συμπληρώνουμε τον αριθμό 1.

Εδώ τα τελικά αθροίσματα συμπίπτουν πάντα με τα αντίστοιχα σύνολα των πρώτων 2 στηλών.

Οι αριθμοί αυτοί συνδέονται πάντα με τη σχέση $f_i = \nu_i / \nu$. Αυτό σημαίνει πως αν γνωρίζουμε 2 από αυτούς μπορούμε εύκολα να υπολογίσουμε και τον τρίτο.

Εδώ πάντοτε συμπληρώνουμε τον αριθμό 100.

Τα υπόλοιπα στοιχεία υπολογίζονται εύκολα εφαρμόζοντας τη λογική κάθε στήλης ευθέως ή αντίστροφα σκεπτόμενοι. Για παράδειγμα, ας συμπληρώσουμε τον παρακάτω πίνακα:

x_i	ν_i	f_i (%)	N_i	F_i (%)
x_1				
x_2	100		150	
x_3				67,5
x_4		10		
x_5			400	
Σύνολο				

Συμπληρώνουμε με κλειστά μάτια τα εξής:

x_i	v_i	f_i (%)	N_i	F_i (%)
x_1				
x_2	100		150	
x_3				67,5
x_4		10		
x_5			400	**100**
Σύνολο	**400**	**100**		

- Γνωρίζουμε το ν = 400 και το $ν_2$ = 100, άρα απ' τον τύπο υπολογίζουμε $f_2 = \frac{ν_2}{ν} \Rightarrow f_2 = \frac{100}{400} = 0{,}25$ ή 25%.

- Επίσης από ν = 400 και f_4 = 10 υπολογίζουμε $f_4 = \frac{ν_4}{ν} \Rightarrow 0{,}10 = \frac{ν_4}{400} \Rightarrow$ $ν_4 = 400 \cdot 0{,}10 \Rightarrow ν_4 = 40$. Προσέξτε ότι στον τύπο δεν αντικαταστήσαμε όπου f_4 το 10 αλλά το 0,10.

- Αφού το $N_2 = 150$ και $ν_2 = 100$ τότε $N_1 = N_2 - ν_2 = 150 - 100 = 50$. Συνεπώς και το $ν_1 = 50$.

- Τέλος, εύκολα υπολογίζουμε και το $F_4 = F_3 + f_4 = 67{,}5 + 10 = 77{,}5$.

x_i	v_i	f_i (%)	N_i	F_i (%)
x_1	**50**		**50**	
x_2	100	**25**	150	
x_3				67,5
x_4	**40**	10		**77,5**
x_5			400	100
Σύνολο	400	100		

- Στη στήλη Fi είναι προφανές ότι από το 77,5 για να φτάσουμε στο 100 χρειάζεται 100 – 77,5 = 22,5. Άρα $f_5 = 22{,}5$.

- Αν όμως $f_5 = 22{,}5$ τότε $f_5 = \frac{ν_5}{ν} \Rightarrow 0{,}225 = \frac{ν_5}{400} \Rightarrow ν_5 = 400 \cdot 0{,}225 \Rightarrow ν_5 = 90$.

- Τελικά, στη στήλη $ν_i$ απομένει μονάχα ένα κενό στοιχείο. Εύκολα, σκεφτόμαστε: $50 + 100 + ν_3 + 40 + 90 = 400 \Rightarrow ν_3 = 400 - 280 = 120$.

x_i	ν_i	f_i (%)	N_i	F_i (%)
x_1	50		50	
x_2	100	25	150	
x_3	**120**			67,5
x_4	40	10		77,5
x_5	**90**	**22,5**	400	100
Σύνολο	400	100		

Εφόσον, έχει πλέον ολοκληρωθεί η στήλη των συχνοτήτων ν_i η συνέχεια είναι απλούστατη. Τελικά, ο πίνακας καταλήγει:

x_i	ν_i	f_i (%)	N_i	F_i (%)
x_1	50	**12,5**	50	**12,5**
x_2	100	25	150	**37.5**
x_3	120	**30**	**270**	67,5
x_4	40	10	**310**	77,5
x_5	90	22,5	400	100
Σύνολο	400	100		

Εφαρμογή 4 - Παράμετροι θέσης

Επικρατούσα τιμή

Πρόκειται για τον ευκολότερο υπολογισμό. Στη στήλη v_i, αναζητούμε τη μεγαλύτερη συχνότητα. Η τιμή x_i που αντιστοιχεί σε αυτή είναι η επικρατούσα τιμή. Στο παράδειγμά μας, η μεγαλύτερη συχνότητα είναι 5 και αντιστοιχεί σε μέση θερμοκρασία 17 βαθμών. Άρα η επικρατούσα τιμή είναι το 17. **ΠΡΟΣΟΧΗ !!!** Πολύ συχνά γίνεται η παρανόηση να εκλαμβάνεται ως επικρατούσα τιμή η συχνότητα πχ. το 5, αντί του ορθού 17.

Να μην ξεχνάμε, επίσης, ότι κάποιες φορές μπορεί να υπάρχουν περισσότερες της μίας επικρατούσες τιμές, άρα δε χρειάζεται να τα χάνουμε.

Μέση τιμή

Η εύρεσή της είναι απλή εφαρμογή. Για λίγα δεδομένα συμφέρει η άμεση εφαρμογή του τύπου. Διαφορετικά, ταξινομούμε πρώτα τα δεδομένα μας σε πίνακα συχνοτήτων. Στο παράδειγμα με τις μέσες θερμοκρασίες θα ήταν:

$$\overline{X} = \frac{15+15+18+17+19+10+17+16+14+17+18+19+16+15+17+18+15+14+16+17}{20} \Rightarrow$$

$$\overline{X} = \frac{323}{20} \Rightarrow \mathbf{\overline{X} = 16{,}15}$$

Ωστόσο, είναι προφανές ότι ακόμη και για μονάχα 20 απλούς ακέραιους αριθμούς ο τύπος είναι δύσχρηστος. Για να απλοποιήσουμε τη διαδικασία, συμπληρώνουμε τον πίνακα συχνοτήτων με τη στήλη $\mathbf{x_1 \cdot v_1}$. Το άθροισμα της στήλης είναι ο αριθμητής του προηγούμενου κλάσματος.

Τιμές x_i	Συχνότητα v_i	$x_i \cdot v_i$
10	1	10
14	2	28
15	4	60
16	3	48
17	5	85
18	3	54
19	2	38
Σύνολο	**20**	**323**

Άρα, εύκολα λέμε: $\overline{X} = \frac{323}{20} \Rightarrow \mathbf{\overline{X} = 16{,}15}$

Διάμεσος (ν = άρτιος)

Για την εύρεση της διαμέσου, προκειμένου για λίγες, διακριτές τιμές η διαδικασία είναι σαν παιχνίδι. Καταρχάς, πρώτα απ' όλα ταξινομούμε τα δεδομένα κατά αύξουσα σειρά. Στο παράδειγμά μας, επειδή το πλήθος ν = 20 είναι ζυγός αριθμός, η διάμεσος είναι το ημιάθροισμα των 2 μεσαίων παρατηρήσεων. Για να βρούμε τις τελευταίες, σκεφτόμαστε 20 : 2 = 10, άρα η διάμεσος βρίσκεται ανάμεσα στην **10η** και **11η** θέση.

10 14 14 15 15 15 15 16 16 16 17 17 17 17 17 18 18 18 19 19

$$\boldsymbol{\delta} = \frac{16+17}{2} = \frac{33}{2} = \mathbf{16{,}5}$$

Λίγο δυσκολότερη γίνεται η εύρεση της διαμέσου, από τον πίνακα συχνοτήτων. Τότε αθροίζουμε διαδοχικά τις συχνότητες $\mathbf{\nu_i}$ μέχρι να φτάσουμε το 10. Μπορούμε να κάνουμε το ίδιο συντομότερα, αναζητώντας στη στήλη $\mathbf{N_i}$, αν υπάρχει, τον αριθμό 10. Βλέπουμε ότι στη θέση αυτή αντιστοιχεί ο αριθμός 16. Στην αμέσως επόμενη θέση, δηλαδή την 11η, αντιστοιχεί ο αριθμός 17. Συνεπώς:

$$\delta = \frac{16+17}{2} = \frac{33}{2} = 16{,}5$$

Τιμές x_i	Συχνότητα ν_i	N_i	
10	1	1	1η θέση
14	2	3	3η θέση
15	4	7	7η θέση
16	3	10	10η θέση
17	5	15	15η θέση
18	3	18	18η θέση
19	2	20	20η θέση
Σύνολο	**20**		

Στην περίπτωση που στο άθροισμα δε βρίσκουμε ακριβώς τον αριθμό 10 αλλά τον ξεπερνάμε, τότε αυτό σημαίνει ότι και στις 2 μεσαίες θέσεις βρίσκεται ο ίδιος αριθμός, ο οποίος είναι τελικά και η διάμεσος. Για παράδειγμα, στον παρακάτω πίνακα διάμεσος είναι ο αριθμός που βρίσκεται ανάμεσα στην 60η και 61η θέση, που είναι ο ίδιος δηλαδή **δ = 4** (δε χρειάζονται άλλοι υπολογισμοί).

Τιμές x_i	Συχνότητα ν_i	N_i	
1	11	11	11η θέση
2	17	28	28η θέση
3	20	48	48η θέση
4	25	73	73η θέση
5	22	95	95η θέση
6	16	111	111η θέση
7	9	120	120η θέση
Σύνολο	**120**		

Διάμεσος (ν = περιττός)

Σε περιττό πλήθος παρατηρήσεων, θα υπάρχει πάντα ένας αριθμός ο οποίος βρίσκεται θα ακριβώς στη μέση των (ταξινομημένων πάντα) δεδομένων. Ο αριθμός αυτός είναι η διάμεσος. Για να υπολογίσουμε κατευθείαν τη θέση που βρίσκεται, αν **ν** το μέγεθος του δείγματος τότε **(ν + 1) : 2** η θέση της διαμέσου. Στο παρακάτω παράδειγμα, είναι ν = 21, άρα η διάμεσος βρίσκεται στην 22 : 2 = 11η θέση:

1 1 1 2 2 3 3 3 3 4 5 5 6 6 6 7 7 8 8 9 10

δ = **5**

Εφαρμογή 5 - Παράμετροι διασποράς

Εύρος

Πολύ απλά:

$$\mathbf{R} = x_{max} - x_{min} = 19 - 10 = \mathbf{9}$$

Διακύμανση

Στην περίπτωση των παραμέτρων διασποράς, η άμεση εφαρμογή των τύπων είναι χρονοβόρα και δύσχρηστη. Για τον λόγο αυτό, συμπληρώνουμε τον πίνακα συχνοτήτων με τις παρακάτω στήλες:

x_i	ν_i	$x_i \cdot \nu_i$		$\bar{X} - x_i$		$(\bar{X} - x_i)^2$		$\nu_i \cdot (\bar{X} - x_i)^2$
10	1	10	16,15 – 10 =	6,15	$6{,}15^2$ =	37,82	1 · 37,82 =	37,82
14	2	28	16,15 – 14 =	2,15	$2{,}15^2$ =	4,62	2 · 4,62 =	9,24
15	4	60	16,15 – 15 =	1,15	$1{,}15^2$ =	1,32	4 · 1,32 =	5,29
16	3	48	16,15 – 16 =	0,15	$0{,}15^2$ =	0,02	3 · 0,02 =	0,07
17	5	85	16,15 – 17 =	-0,85	$(-0{,}85)^2$ =	0,72	5 · 0,72 =	3,61
18	3	54	16,15 – 18 =	-1,85	$(-1{,}85)^2$ =	3,42	3 · 3,42 =	10,27
19	2	38	16,15 – 19 =	-2,85	$(-2{,}85)^2$ =	8,12	2 · 8,12 =	16,25
Σύνολο	**20**	**323**						**82,55**

Το άθροισμα της τελευταίας στήλης δεν είναι παρά ο αριθμητής από τον τύπο της διακύμανσης, δηλαδή:

$$s^2 = \frac{82{,}55}{20} = \mathbf{4{,}13}$$

Προσοχή! Είναι προφανές ότι για τους παραπάνω υπολογισμούς απαιτείται πρωτύτερα η εύρεση της μέσης τιμής $\bar{X}$.

Τυπική απόκλιση

Πολύ εύκολα, από τον τύπο:

$$s = \sqrt{s^2} = \sqrt{4{,}13} \cong \mathbf{2{,}03}$$

Συντελεστής μεταβλητότητας

Πάλι με απλή χρήση του τύπου:

$$\mathbf{CV} = \frac{s}{\overline{\overline{X}}} = \frac{4{,}13}{16{,}15} \cong 0{,}26 \text{ ή } \mathbf{26\%}$$

Επειδή **CV > 10%** το δείγμα των μέσων θερμοκρασιών **δεν είναι ομοιογενές**.

Εφαρμογή 6 - Ομαδοποίηση σε κλάσεις

Πολύ συχνά, αν μετράμε μια συνεχή μεταβλητή η οποία παίρνει απεριόριστες ενδιάμεσες τιμές ή αν, γενικότερα, τα δεδομένα μας παρουσιάζουν μεγάλη διαφοροποίηση τιμών, τότε ο κλασικός πίνακας συχνοτήτων μπορεί να περιέχει μεγάλο αριθμό σειρών και να καθίσταται δύσχρηστος. Πχ. Να κατασκευάσετε πίνακα συχνοτήτων για τις παρακάτω 40 μετρήσεις:

177	**181**	**183**	**187**	**192**	**194**	**171**	**180**	**190**	**192**
179	**182**	**184**	**187**	**191**	**193**	**194**	**190**	**186**	**183**
179	**176**	**178**	**180**	**183**	**186**	**185**	**191**	**187**	**178**
170	**176**	**177**	**181**	**174**	**192**	**194**	**178**	**188**	**193**

Αν κατασκευάσουμε τον πίνακα με τον «κλασικό» τρόπο, θα προκύψει ένα αποτέλεσμα που δε μας γλιτώνει απ’ το μεγάλο πλήθος υπολογισμών, αυτό δηλαδή που θέλουμε να αποφύγουμε:

x_i	ν_i
170	1
171	1
174	1
176	2
177	2
178	3
179	2
180	2
181	2
182	1
183	3
184	1
185	1
186	2
187	3
188	1
190	2
191	2
192	3
193	2
194	3
Σύνολο	**20**

Παρατηρούμε ότι το πλήθος των σειρών είναι τόσο μεγάλο, ώστε τελικά είναι σχεδόν σα να μην κάναμε καν πίνακα! Στην περίπτωση αυτή, συμφέρει ν’ ακολουθήσουμε διαφορετική τακτική.

Ομαδοποιούμε τα δεδομένα μας σε κατηγορίες που ονομάζονται **κλάσεις** και οι οποίες περιλαμβάνουν περισσότερες της μίας μετρήσεις. Κάθε κλάση έχει ένα συγκεκριμένο εύρος αριθμών που περιλαμβάνει, το οποίο ονομάζεται **πλάτος** της κλάσης. Το πλάτος μια κλάσης υπολογίζεται εύκολα αν αφαιρέσουμε από το μεγαλύτερο όριό της το μικρότερο.

Για το πρηγούμενο παράδειγμα, μια κλάση θα μπορούσε να ομαδοποιήσει, ας πούμε, όλες τις μετρήσεις από το 175 έως και το 180 κι έτσι να εργαστούμε πιο άνετα. Ας προσέξουμε λίγο το συμβολισμό…

Μια κλάση συμβολίζεται σαν ένα κλειστό–ανοιχτό διάστημα, πχ.

[175, 180)

Η αγκύλη [διαβάζεται «κλειστό» και σημαίνει ότι ο αριθμός 175 περιλαμβάνεται στο διάστημα.

Η παρένθεση) διαβάζεται «ανοικτό» και σημαίνει ότι ο αριθμός 180 ΔΕΝ περιλαμβάνεται στο διάστημα. Μην ανησυχείτε, θα περιληφθεί στην επόμενη κλάση: [180, 185).

Η κλάση αυτή θα έχει πλάτος = 180 − 175 = **5**. Μετρώντας διαπιστώνουμε ότι στην κλάση αυτή ανήκουν 9 δεδομένα: 176, 176, 177, 177, 178, 178, 178, 179, 179. Τότε λέμε ότι η κλάση αυτή έχει συχνότητα **ν_i = 9**.

Ας υποθέσουμε τώρα ότι θέλουμε να ομαδοποιήσουμε τους παραπάνω 40 αριθμούς σε κλάσεις πλάτους ίσου με 5 (αυτό θα δίνεται συνήθως από την εκφώνηση της άσκησης). Χρησιμοποιούμε όσες κλάσεις χρειάζονται, έως ότου καλύψουμε όλα τα δεδομένα μας. Στη συγκεκριμένη περίπτωση, οι κλάσεις που θα χρειαστούν είναι οι:

[170, 175) [175, 180) [180, 185) [185, 190) [190, 195)

Παρατηρούμε! ότι κάθε επόμενη κλάση ξεκινά από τον ίδιο ακριβώς αριθμό που τελειώνει η προηγούμενη κλάση, έτσι ώστε να καλύπτονται όλες οι μετρήσεις. Στη συνέχεια, μετράμε τα δεδομένα μας ένα–ένα, ταξινομώντας τα στις κατάλληλες κλάσεις.

170, 171, 174	176, 176, 177, 177, 178, 178, 178, 179, 179	180, 180, 181, 181, 182, 183, 183, 183, 184	185, 186, 186, 187, 187, 187, 188	190, 190, 191, 191, 192, 192, 192, 193, 193, 194, 194, 194

[170, 175)	**[175, 180)**	**[180, 185)**	**[185, 190)**	**[190, 195)**
Συχνότητα = 3	Συχνότητα = 9	Συχνότητα = 9	Συχνότητα = 7	Συχνότητα = 12

Έτσι, μπορούμε να φτιάξουμε ένα νέο πίνακα συχνοτήτων, με κλάσεις αυτή τη φορά, και προφανώς πολύ περισσότερο «συμπυκνωμένο» από τον αρχικό πίνακα:

Κλάση	$\mathbf{v_i}$
[170, 175)	3
[175, 180)	9
[180, 185)	9
[185, 190)	7
[190, 195)	12

Παράμετροι θέσης & διασποράς σε κλάσεις

Η εύρεση των $\overline{\mathbf{X}}$, $\mathbf{s^2}$ και **s** πραγματοποιείται με τον ίδιο ακριβώς τρόπο. Το πρόβλημα έκγειται στο γεγονός ότι προκειμένουν να κατασκευάσουμε τις βοηθητικές μας στήλες, πχ. $\mathbf{x_i \cdot v_i}$ μας λείπουν οι τιμές x_i, εφόσον τα δεδομένα μας είναι ταξινομημένα σε κλάσεις. Στην περίπτωση αυτή, επιλέγουμε σαν «αντιπρόσωπο» κάθε κλάσης τη μέση τιμή των άκρων της, δηλαδή τον αριθμό που βρίσκεται ακριβώς στο **κέντρο** της κλάσης και τον οποίο συμβολίζουμε με $\mathbf{\kappa_i}$. Γενικά:

$$[\alpha, \beta) \rightarrow \kappa_i = \frac{\alpha + \beta}{2}$$

Κλάση	$\mathbf{\kappa_i}$	$\mathbf{v_i}$
[170, 175)	**172,5**	3
[175, 180)	**177,5**	9
[180, 185)	**182,5**	9
[185, 190)	**187,5**	7
[190, 195)	**192,5**	12

Οι τιμές κ_i παίζουν τον ίδιο ακριβώς ρόλο με τα x_i ενός απλού πίνακα συχνοτήτων.

Η **επικρατούσα τιμή** και η **διάμεσος** υπολογίζονται μόνο γραφικά για την εξεταστέα ύλη αυτής της τάξης. Ένα παράδειγμα δίνεται στην **εφαρμογή 8**.

Εφαρμογή 7 - Γραφική απεικόνιση

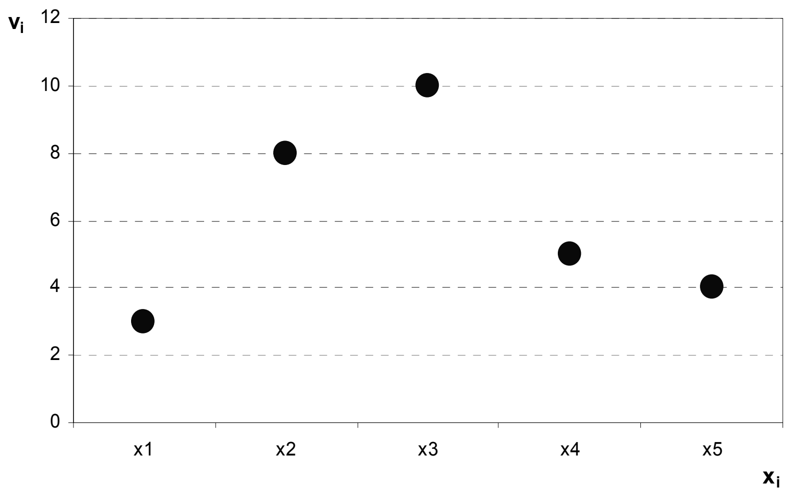
Προκειμένου να έχουμε μια καλύτερη αντίληψη, συνηθίζουμε να τα απεικονίζουμε τα δεδομένα μας σε διάφορα διαγράμματα, το καθένα με το δικό του ιδιαίτερο χαρακτήρα και τα πλεονεκτήματα ή μειονεκτήματα.

Απλή Γραφική Παράσταση

Θεωρούμε ένα ορθογώνιο σύστημα αξόνων. Στον **οριζόντιο άξονα** x΄x εκφράζουμε τις διάφορες **τιμές x_i** της μεταβλητής X, ενώ στον **κάθετο άξονα** y΄y τις αντίστοιχες **συχνότητες v_i**.

Έστω ο παρακάτω τυχαίος πίνακας συχνοτήτων, σαν αφορμή για το παράδειγμά μας.

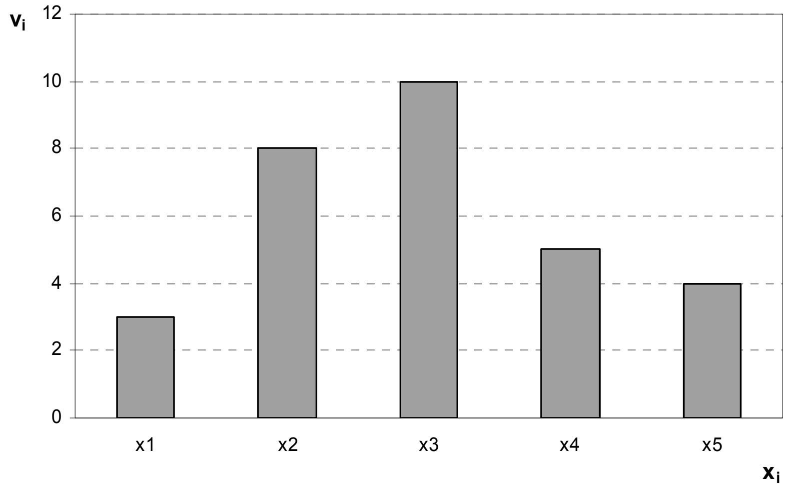
x_i	v_i	f_i
x_1	3	0,10
x_2	8	0,27
x_3	10	0,33
x_4	5	0,17
x_5	4	0,13



Κατακόρυφο Ραβδόγραμμα

Αν στο προηγούμενο τύπο διαγράμματος, αντικαταστήσουμε τις τελείες με κάθετα ορθογώνια (μπάρες), τότε δημιουργείται αυτό που ονομάζουμε κατακόρυφο ραβδόγραμμα.

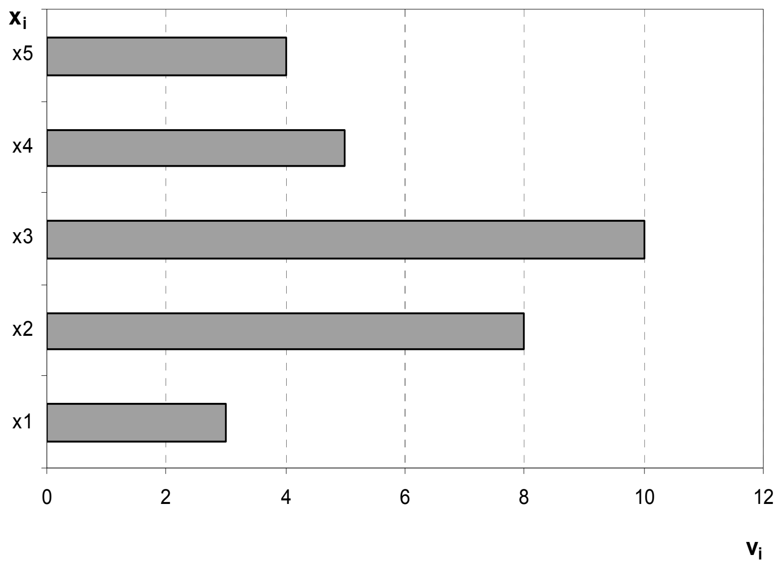
x_i	v_i	f_i
x_1	3	0,10
x_2	8	0,27
x_3	10	0,33
x_4	5	0,17
x_5	4	0,13



Οριζόντιο Ραβδόγραμμα

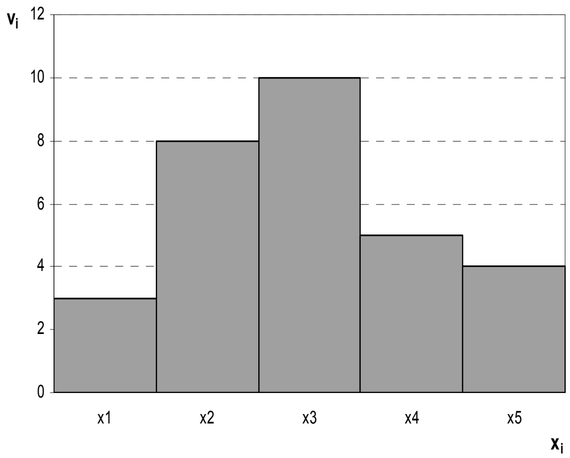
Αν στο κατακόρυφο ραβδόγραμμα εναλλάξουμε τη θέση των δύο αξόνων, δηλαδή μεταφέρουμε τις τιμές x_i στον άξονα y΄y ενώ τις αντίστοιχες συχνότητες στον x΄x, τότε παίρνουμε το οριζόντιο ραβδόγραμμα.

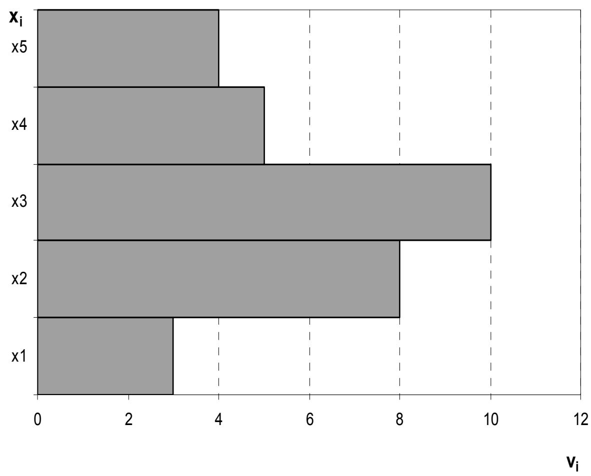
x_i	v_i	f_i
x_1	3	0,10
x_2	8	0,27
x_3	10	0,33
x_4	5	0,17
x_5	4	0,13



Παραλλαγή

Τόσο στο κατακόρυφο, όσο και στο οριζόντιο ραβδόγραμα, είναι δυνατόν να «γεμίσουμε» τις μπάρες έτσι ώστε να καταλαμβάνουν όλο το χώρο ανάμεσά τους.



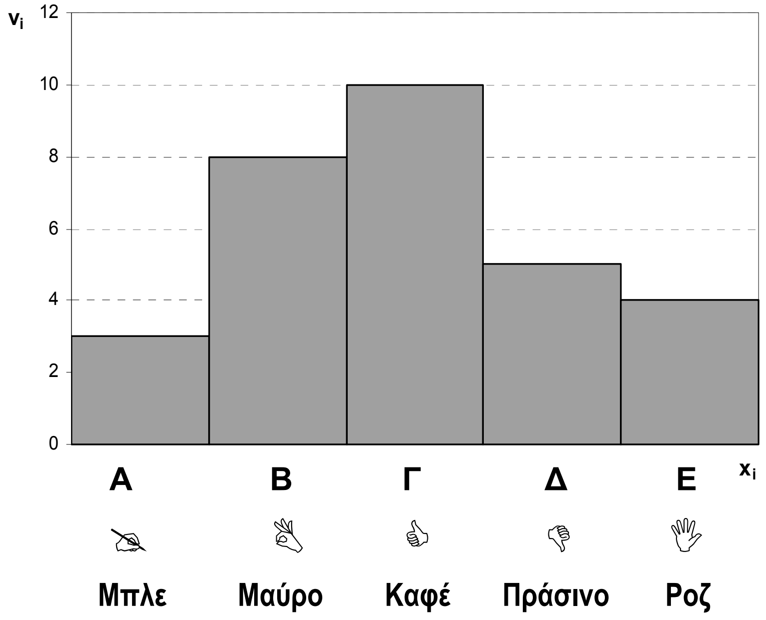


ΠΡΟΣΟΧΗ!

Πρέπει να γίνει κατανοητό ότι ο άξονας των τιμών x_i **ΔΕΝ** είναι ένας αριθμητικός άξονας, όπως ο άξονας των συχνοτήτων. Πρόκειται απλά για ένα άξονα με διαδοχικές «θέσεις» στις οποίες τοποθετούνται οι διάφορες τιμές x_i, συνεπώς **ΔΕΝ** παίζει ρόλο η ορθή διάταξη (ασχέτως αν μας εξυπηρετεί όταν έχουμε αριθμητικές τιμές). Συνεπώς μπορούμε να φτιάχνουμε ραβδογράμματα ακόμα και για τιμές οι οποίες δεν είναι αριθμητικές, όπως ας πούμε σε μια ποιοτική μεταβλητή.

Θα μπορούσαμε, για παράδειγμα, στον προηγούμενο πίνακα να έχουμε γράμματα, εικόνες ή απλές λέξεις και το ραβδόγραμμα να μην άλλαζε στο παραμικρό...

Τιμές x_i	Τιμές x_i	Τιμές x_i	Συχνότητα ν_i
Α	✍	Μπλε	3
Β	👌	Μαύρο	8
Γ	👍	Καφέ	10
Δ	👎	Πράσινο	5
Ε	✋	Ροζ	4



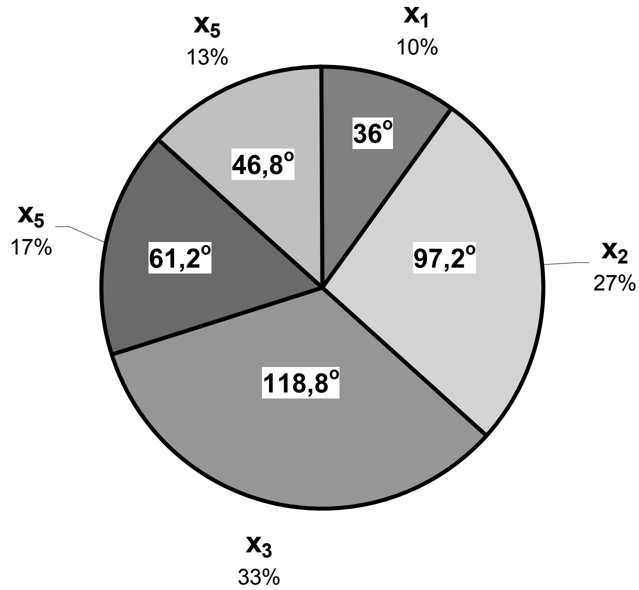
Κυκλικό διάγραμμα

Αν πάλι χωρίζαμε έναν κύκλο σε κομμάτια «πίτσας», έτσι ώστε κάθε **επίκεντρη γωνία** να καταλαμβάνει μέρος του κύκλου ίσο με το **ποσοστό** της αντίστοιχης τιμής, την οποία συμβολίζει, τότε κατασκευάζουμε αυτό που λέμε «κυκλικό διάγραμμα». Για να βρούμε πόσες μοίρες πρέπει να σχεδιάσουμε κάθε γωνία φ_i χρησιμοποιούμε τον παρακάτω τύπο:

$$\boxed{\hat{\varphi}_i = 360 \cdot f_i}$$

Για το παράδειγμά μας, έχουμε:

$\varphi_1 = 360^o \cdot f_1 = 360^o \cdot 0{,}10 = \mathbf{36^o}$
$\varphi_2 = 360^o \cdot f_2 = 360^o \cdot 0{,}27 = \mathbf{97{,}2^o}$
$\varphi_3 = 360^o \cdot f_3 = 360^o \cdot 0{,}33 = \mathbf{118{,}8^o}$
$\varphi_4 = 360^o \cdot f_4 = 360^o \cdot 0{,}17 = \mathbf{61{,}2^o}$
$\varphi_5 = 360^o \cdot f_5 = 360^o \cdot 0{,}13 = \mathbf{46{,}8^o}$

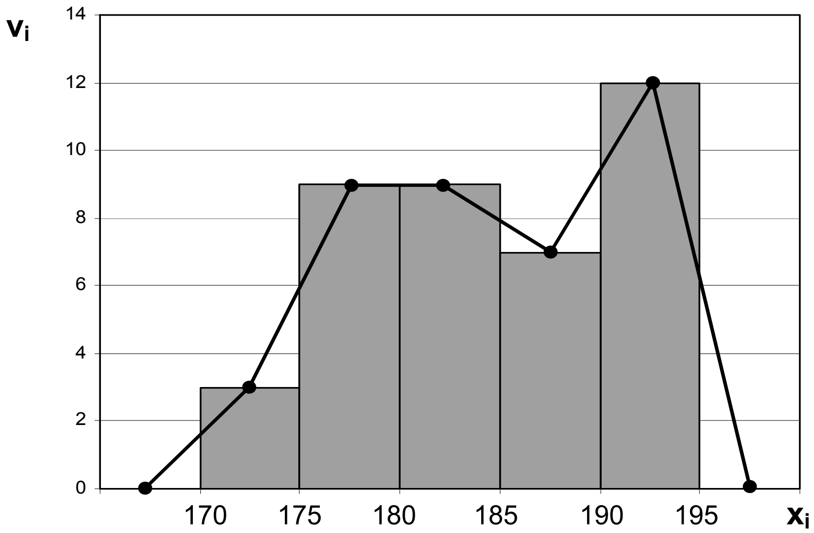


Ιστόγραμμα συχνοτήτων

Το ραβδόγραμμα που αντιστοιχεί σε ένα πίνακα χωρισμένο σε κλάσεις λέγεται ιστόγραμμα. Η ουσιαστικότερη διαφορά του ιστογράμματος από το ραβδόγραμμα είναι πως στο πρώτο ο οριζόντιος άξονας των x_i είναι κι αυτός επίσης **αριθμητικός άξονας** και για το λόγο αυτό συνεχής και διατεταγμένος. Για το λόγο αυτό, τοποθετούμε τα όρια των κλάσεων **πάνω** στις γραμμές του άξονα και **όχι ανάμεσά** τους!

Έτσι για παράδειγμα στον πίνακα της εφαρμογής 6 αντιστοιχεί το παρακάτω ιστόγραμμα:

Κλάση	ν_i	N_i
[170, 175)	3	3
[175, 180)	9	12
[180, 185)	9	21
[185, 190)	7	28
[190, 195)	12	40

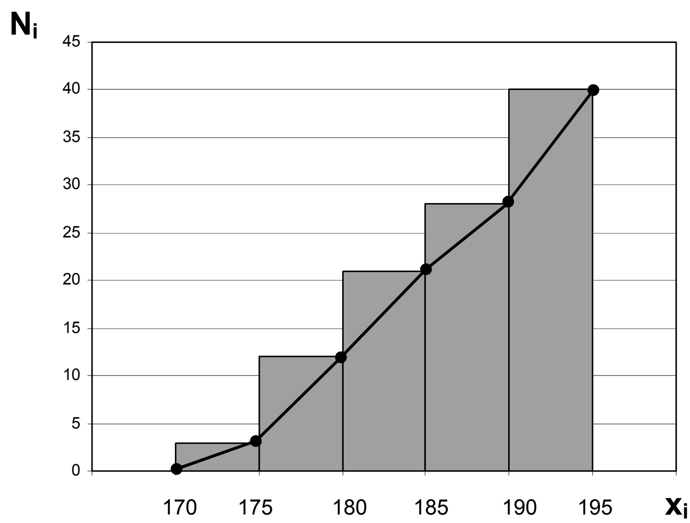


Πολύγωνο συχνοτήτων

Η τεθλασμένη γραμμή που παρατηρούμε στο προηγούμενο ιστόγραμμα λέγεται πολύγωνο συχνοτήτων. Όταν καλούμαστε να το σχεδιάσουμε, απλά **ενώνουμε τα μέσα** των κορυφών κάθε ορθογωνίου. Επιπρόσθετα, προκειμένου να «κλείσουν» οι άκρες του πολυγώνου, εισάγουμε **δύο επιπλέον κλάσεις** στην αρχή και στο τέλος κι ενώνουμε με παρόμοιο τρόπο.

Πολύγωνο αθροιστικών (σχετικών) συχνοτήτων

Ενώ το ιστόγραμμα αθροιστικών συχνοτήτων δεν παρουσιάζει ιδιαίτερη δυσκολία, χρειάζεται ωστόσο προσοχή στην κατασκευή του πολυγώνου, γιατί διαφέρει απ' την προηγούμενη περίπτωση. Ξεκινάμε φέρνοντας τη **διαγώνιο** του πρώτου ορθογωνίου και συνεχίζουμε με παρόμοια τρόπο διαγωνίως.

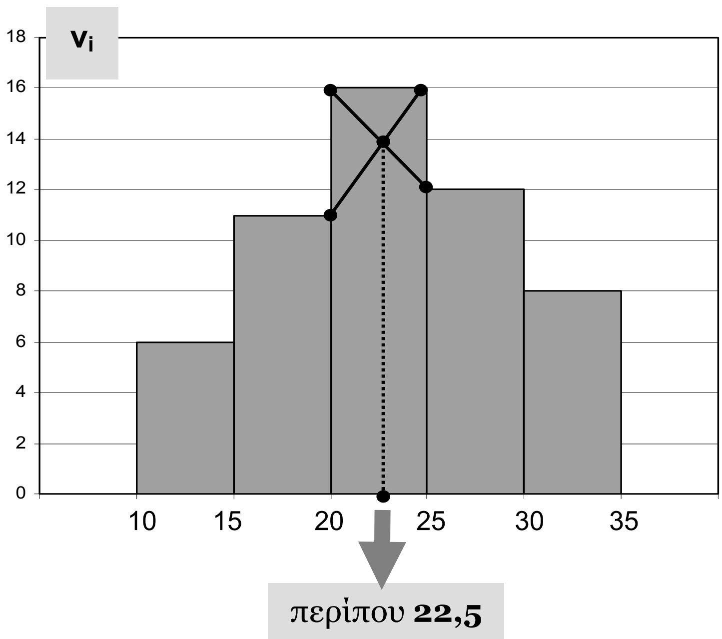


Εφαρμογή 8

Επικρατούσα τιμή σε πίνακα κλάσεων

Χρησιμοποιώντας ως παράδειγμα τον παρακάτω πίνακα συχνοτήτων κάποιας μεταβλητής, κατασκευάζουμε το ιστόγραμμα συχνοτήτων. Επιλέγουμε **την κλάση με τη μεγαλύτερη συχνότητα** (δηλ. το ψηλότερο ορθογώνιο) και φέρνουμε τις διαγώνιες γραμμές όπως ακριβώς φαίνονται στο σχήμα. Η **τετμημένη** του σημείου τομής των δύο διαγωνίων γραμμών μας δίνει **κατά προσέγγιση** την επικρατούσα τιμή του δείγματος.

Κλάση	ν_i	N_i
[10, 15)	6	3
[15, 20)	11	14
[20, 25)	16	30
[25, 30)	12	42
[30, 35)	8	50



Διάμεσος σε πίνακα κλάσεων

Για τη διάμεσο κατασκευάζουμε το ιστόγραμμα αθροιστικών (ή σχετικών αθροιστικών) συχνοτήτων και μαζί με αυτό και το αντίστοιχο πολύγωνο συχνοτήτων. Στη συνέχεια από το μέσο του άξονα $\mathbf{N_i}$ ή $\mathbf{F_i}$ – δηλαδή από τον αριθμό **ν:2** ή το **50** αντίστοιχα – φέρουμε μια οριζόντια γραμμή που τέμνει το πολύγωνο σε κάποιο σημείο. Η **τετμημένη** του σημείου αυτού μας δίνει **κατά προσέγγιση** τη διάμεσο του δείγματος.