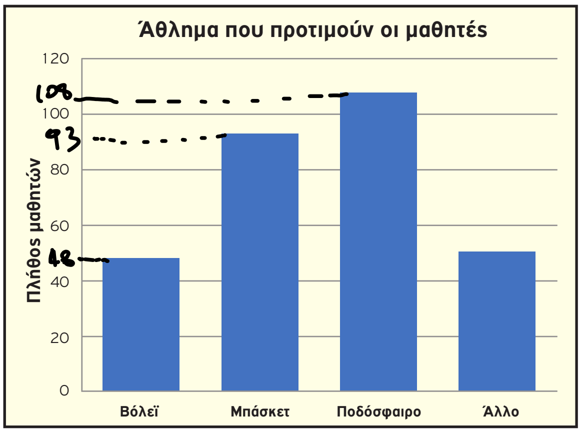
ΕΝΟΤΗΤΑ 3.2:
ΠΑΡΟΥΣΙΑΣΗ ΣΤΑΤΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Διερεύνηση

Μια εταιρεία αθλητικών ειδών, πριν επενδύσει σε αθλητικά είδη που προτιμούν οι μαθητές, αποφάσισε να κάνει μια έρευνα. Για το λόγο αυτό επέλεξε, με τυχαίο τρόπο, δείγμα τριακοσίων μαθητών απ' όλη την Ελλάδα. Ο υπεύθυνος που έκανε την έρευνα, μετά την επεξεργασία των στοιχείων που συγκέντρωσε, παρουσίασε στο διευθυντή της εταιρείας τον παρακάτω πίνακα και τα δυο διαγράμματα.

X = **Άθλημα**	**Πλήθος μαθητών**	**Ποσοστό**
x_1 = Βόλεϊ	48 $= \nu_1$	$\frac{48}{300}$ = 16% $= f_1\%$
x_2 = Μπάσκετ	93 $= \nu_2$	31% $= f_2\%$
x_3 = Ποδόσφαιρο	108 $= \nu_3$	36% $= f_3\%$
x_4 = Άλλο	51 $= \nu_4$	17% $= f_4\%$
Σύνολο	**300** $= \nu$	**100%**



1) Ποια είναι η μεταβλητή της έρευνας και ποιο το είδος της;

2) Πως προκύπτουν τα αντίστοιχα ποσοστά για κάθε άθλημα;

3) Ποιο είναι το ύψος της κάθε μπάρας στο ραβδόγραμμα;

4) Ποια είναι η γωνία του κάθε κυκλικού τομέα στο κυκλικό διάγραμμα;

Βασικές μαθηματικές έννοιες - Ιδέες - Διεργασίες

Ορισμοί

τιμές   πόσες φορές x_i

- Έστω Χ μια μεταβλητή με τιμές $x_1, x_2, x_3, \ldots, x_\kappa$. **Συχνότητα ν_i** μιας τιμής x_i λέγεται ο φυσικός αριθμός που εκφράζει πόσες φορές εμφανίζεται η τιμή αυτή στο σύνολο των παρατηρήσεων του δείγματος. Είναι φανερό ότι το άθροισμα όλων των συχνοτήτων μας δίνει το μέγεθος ν του δείγματος. Δηλαδή:

$$\boxed{\nu_1 + \nu_2 + \nu_3 + \ldots \nu_\kappa = \nu}$$ μέγεθος δείγματος

- **Η σχετική συχνότητα f_i** μιας τιμής x_i ορίζεται ως ο λόγος της αντίστοιχης συχνότητας ν_i προς το μέγεθος ν του δείγματος. Δηλαδή:

$$\boxed{f_i = \frac{\nu_i}{\nu} \quad \text{για } i = 1, 2, 3, \ldots, \kappa}$$

Η σχετική συχνότητα μπορεί να εκφραστεί και ως ποσοστό $f_i\%$.

- Σύμφωνα με τους παραπάνω ορισμούς, ο πίνακας του παραδείγματος που αναφέραμε με μεταβλητή «το είδος της μουσικής που προτιμούν οι μαθητές» μετατρέπεται, όπως φαίνεται στη συνέχεια και λέγεται **πίνακας κατανομής συχνοτήτων και σχετικών συχνοτήτων**.

Άθλημα x_i	Συχνότητα ν_i	Σχετική συχνότητα f_i	Σχετική συχνότητα % $f_i\%$
Βόλεϊ	48	0,16	16
Μπάσκετ	93	0,31	31
Ποδόσφαιρο	108	0,36	36
Άλλο	51	0,17	17
Σύνολο	**300**	$\sum_{i=1}^{\nu} f_i$ **1,00**	**100**

$f_1 + f_2 + f_3 + f_4 = \sum_{i=1}^{4} f_i$

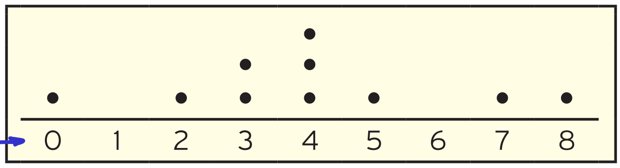
- Οι πλέον συνηθισμένοι τρόποι γραφικής παρουσίασης ποιοτικών αλλά και ποσοτικών διακριτών δεδομένων είναι το ραβδόγραμμα συχνοτήτων και το κυκλικό διάγραμμα.
  - **Το ραβδόγραμμα συχνοτήτων** αποτελείται από ορθογώνιες στήλες, μια για κάθε τιμή της μεταβλητής, όπου το ύψος της κάθε στήλης είναι ίσο με την αντίστοιχη συχνότητα. Αν αντί για τις συχνότητες έχουμε τις σχετικές συχνότητες, τότε λέγεται **ραβδόγραμμα σχετικών συχνοτήτων**.
  - **Το κυκλικό διάγραμμα** χρησιμοποιείται επίσης για τη γραφική παράσταση δεδομένων, κυρίως όταν αυτά παίρνουν λίγες τιμές. Η γωνία α_i του κάθε κυκλικού

τομέα, είναι ανάλογη της αντίστοιχης σχετικής συχνότητας. Δηλαδή:

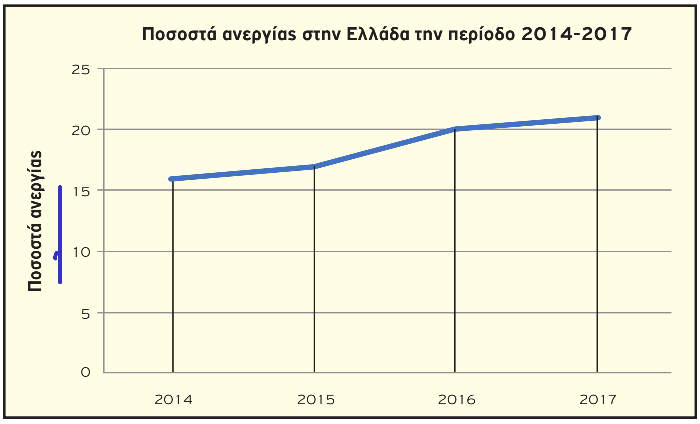
$$\boxed{\alpha_i = 360^o \cdot f_i \ \text{ για } i = 1, 2, 3, \ldots, \kappa}$$

- Όταν έχουμε λίγες παρατηρήσεις, τότε η κατανομή τους μπορεί να περιγραφεί με το **σημειόγραμμα** στο οποίο οι τιμές παριστάνονται με σημεία υπεράνω ενός άξονα.

Χ



- Το **χρονόγραμμα** χρησιμοποιείται για την γραφική απεικόνιση της εξέλιξης σε σχέση με το χρόνο ενός μεγέθους, συνήθως οικονομικού ή δημογραφικού.

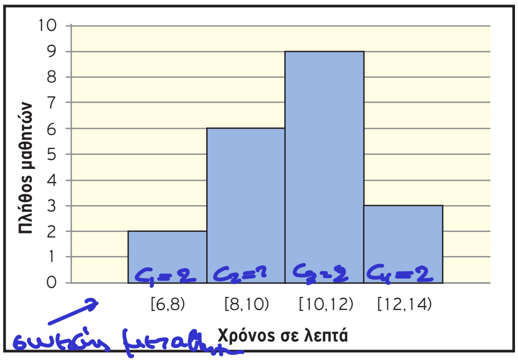


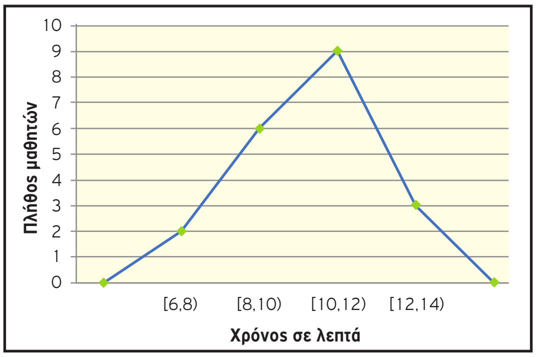
- Στη περίπτωση που έχουμε ποσοτικά συνεχή δεδομένα με πολλές διαφορετικές τιμές, τότε, για την καλύτερη παρουσίασή τους, γίνεται ομαδοποίηση αυτών σε κλάσεις, συνήθως ίσου πλάτους. Σχετικός είναι ο παρακάτω πίνακας με τους χρόνους που χρειάστηκαν οι 20 μαθητές ενός τμήματος για να απαντήσουν σε ένα πρόβλημα.

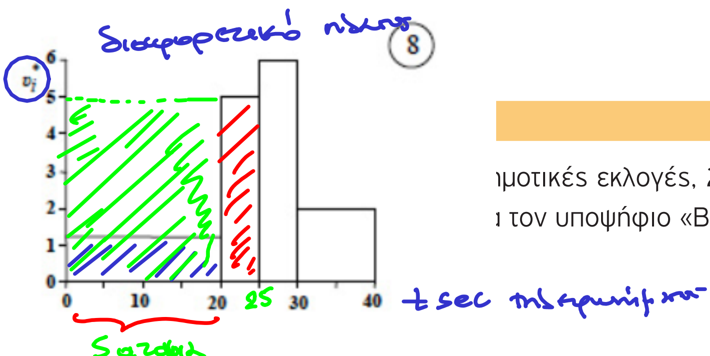
ΚΛΑΣΕΙΣ

Χρόνος σε λεπτά των μαθητών	Συχνότητα ν_i	Σχετική συχνότητα f_i	Σχετική συχνότητα % $f_i\%$
[6,8)	2	0,10	10
[8,10)	6	0,30	30
[10,12)	9	0,45	45
[12,14)	3	0,15	15
Σύνολο	**20**	**1,00**	**100**

- Η γραφική παρουσίαση ομαδοποιημένων στατιστικών δεδομένων γίνεται με το **ιστόγραμμα συχνοτήτων**. Στο παρακάτω σχήμα φαίνεται η γραφική παρουσίαση των δεδομένων του προηγουμένου πίνακα. Αν αντί για τις συχνότητες έχουμε τις σχετικές συχνότητες, τότε λέγεται **ιστόγραμμα σχετικών συχνοτήτων**.
- Αν θεωρήσουμε δυο επιπλέον κλάσεις ίσου πλάτους, μια στη αρχή και μια στο τέλος, με συχνότητα 0 και ενώσουμε τα μέσα των άνω βάσεων, τότε προκύπτει το λεγόμενο **πολύγωνο συχνοτήτων**. Σχετικό είναι το παρακάτω σχήμα. Αν, αντί για τις συχνότητες έχουμε τις σχετικές συχνότητες, τότε λέγεται **πολύγωνο σχετικών συχνοτήτων**.







ιμοτικές εκλογές, 280 άτομα απάντησαν ότι
ι τον υποψήφιο «Β» και 200 άτομα τον υπο-

2) Να κάνετε πίνακα συχνοτήτων και σχετικών συχνοτήτων.

3) Να παραστήσετε τα δεδομένα με ραβδόγραμμα αλλά και με κυκλικό διάγραμμα.

Απάντηση

1) Το πλήθος αυτών που απάντησαν είναι:

$$\nu = 280 + 320 + 200 = 800$$

Επομένως το μέγεθος του δείγματος είναι 800 άτομα.

2) Οι συχνότητες είναι: $\nu_1 = 280$, $\nu_2 = 320$ και $\nu_3 = 200$. Οι σχετικές συχνότητες είναι:

$$f_1 = \frac{\nu_1}{\nu} = \frac{280}{800} = 0{,}35,\ f_2 = \frac{\nu_2}{\nu} = \frac{320}{800} = 0{,}40 \text{ και } f_3 = \frac{\nu_3}{\nu} = \frac{200}{800} = 0{,}25$$

Ο πίνακας κατανομής συχνοτήτων και σχετικών συχνοτήτων είναι:

Υποψήφιος δήμαρχος	Συχνότητα ν_i	Σχετική συχνότητα f_i	Σχετική συχνότητα % $f_i\%$
Υποψήφιος «Α»	280	0,35	35
Υποψήφιος «Β»	320	0,40	40
Υποψήφιος «Γ»	200	0,25	25
Σύνολο	**800**	**1,00**	**100**

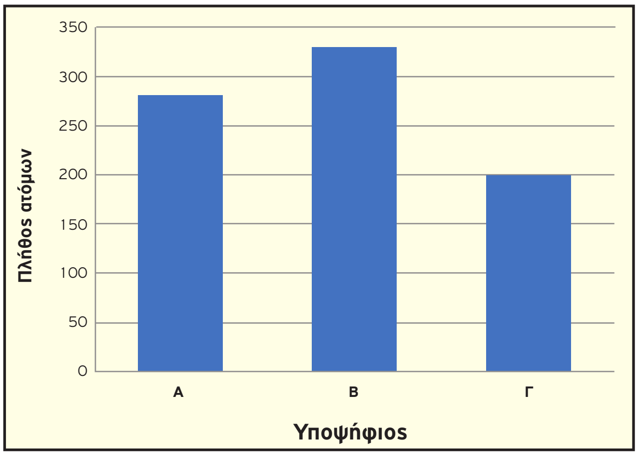
3) Για τις γωνίες των κυκλικών τομέων στο κυκλικό διάγραμμα έχουμε:

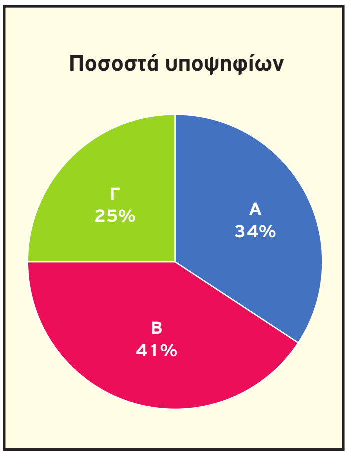
$$\alpha_1 = 360^\circ \cdot f_1 = 360^\circ \cdot 0{,}35 = 126^\circ$$

$$\alpha_2 = 360^\circ \cdot f_2 = 360^\circ \cdot 0{,}40 = 144^\circ$$

$$\alpha_3 = 360^\circ \cdot f_3 = 360^\circ \cdot 0{,}25 = 90^\circ$$

Παρακάτω φαίνονται το ραβδόγραμμα συχνοτήτων και το κυκλικό διάγραμμα.





Εφαρμογή 2η

Παρακάτω δίνονται οι χρόνοι, στρογγυλοποιούμενοι στο δέκατο του δευτερολέπτου, που απαιτήθηκαν για να τρέξουν 50 αθλητές έναν αγώνα δρόμου 1000 m.

52,1	55,3	50	56,4	59,1	54,2	56,7	54,4	57,1	53,7
55,2	55,1	58	59,2	56	55,5	52,5	56,5	58,5	55
55,2	57,3	54,3	53,5	57,9	53	55,4	55,6	52,4	54,5
56,4	59,1	54,2	56,7	55,3	52,4	56,4	54,1	54,3	56,7
54,3	51,5	57	53,2	54,9	55,6	52	55,3	55,1	54,7

1) Ποιο είναι το είδος της μεταβλητής, ποιος ο μικρότερος και ποιος ο μεγαλύτερος χρόνος;
2) Τι έχετε να παρατηρήσετε για το πλήθος των τιμών των παρατηρήσεων σε σχέση με το πλήθος των παρατηρήσεων;
3) Ξεκινώντας από το μικρότερο χρόνο και με βήμα 2 sec, ποιες κλάσεις της μορφής [α,β) δημιουργούνται στις οποίες περιέχονται όλες οι παρατηρήσεις;
4) Να παρουσιαστούν οι παρατηρήσεις, ομαδοποιημένες στις παραπάνω κλάσεις ίσου πλάτους, σε έναν πίνακα συχνοτήτων και σχετικών συχνοτήτων.
5) Να παραστήσετε τα δεδομένα με ιστόγραμμα συχνοτήτων και με πολύγωνο συχνοτήτων.

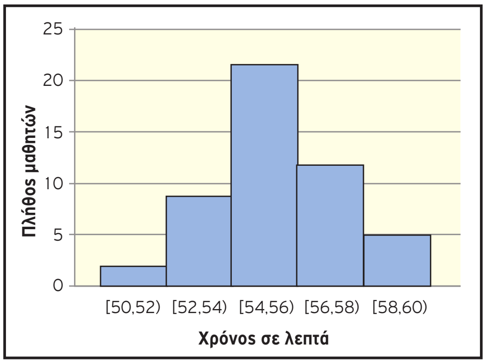
Απάντηση

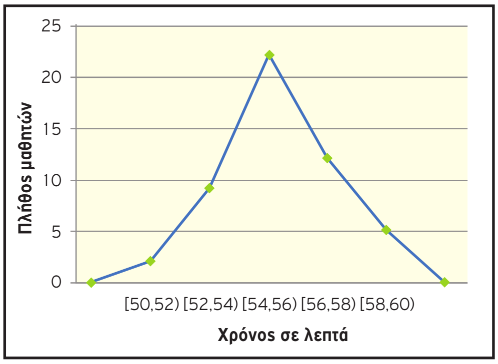
1) Η μεταβλητή είναι ποσοτική συνεχής. Ο μικρότερος χρόνος είναι 50 sec και ο μεγαλύτερος 59,2 sec.
2) Το πλήθος των τιμών των παρατηρήσεων είναι 37 σε σύνολο 50 παρατηρήσεων. Για να παρουσιαστούν τα δεδομένα, ως έχουν, σε ένα πίνακα συχνοτήτων θα χρειαστούμε 37 γραμμές. Είναι φανερό ότι ένας τέτοιος πίνακας είναι δύσχρηστος και οι πληροφορίες δεν παρουσιάζονται συνοπτικά.
3) Οι κλάσεις που δημιουργούνται είναι: [50,52), [52,54), [54,56), [56,58) και [58,60)

4) Μετά τη διαλογή ο πίνακας συχνοτήτων και σχετικών συχνοτήτων είναι ο παρακάτω.

Κλάσεις με χρόνους σε sec	Συχνότητα v_i	Σχετική συχνότητα f_i	Σχετική συχνότητα % f_i%
[50,52)	2	0,04	4
[52,54)	9	0,18	18
[54,56)	22	0,44	44
[56,58)	12	0,24	24
[58,60)	5	0,10	10
Σύνολο	**50**	**1,00**	**100**

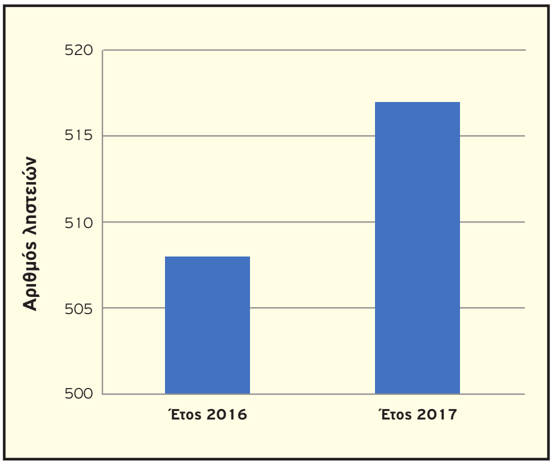
5) Παρακάτω φαίνονται το ιστόγραμμα και το πολύγωνο συχνοτήτων.





Εφαρμογή 3η

Σε ένα τηλεοπτικό κανάλι, ένας δημοσιογράφος σχολίασε την διπλανή γραφική παράσταση ως εξής: *«Η γραφική παράσταση δείχνει ότι σημειώθηκε τεράστια αύξηση του αριθμού των ληστειών το έτος 2017 σε σχέση με το έτος 2016»*.



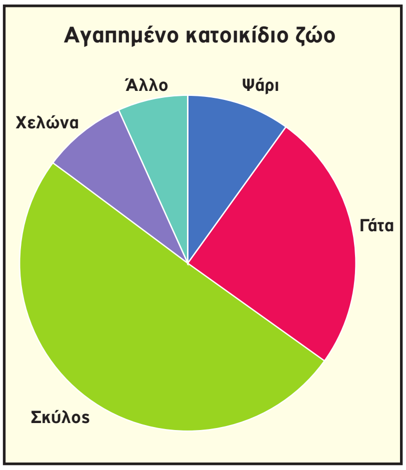
Νομίζετε ότι ο δημοσιογράφος του καναλιού αυτού ερμήνευσε σωστά την γραφική παράσταση; Να γράψετε ένα επιχείρημα που να τεκμηριώνει την απάντησή σας.

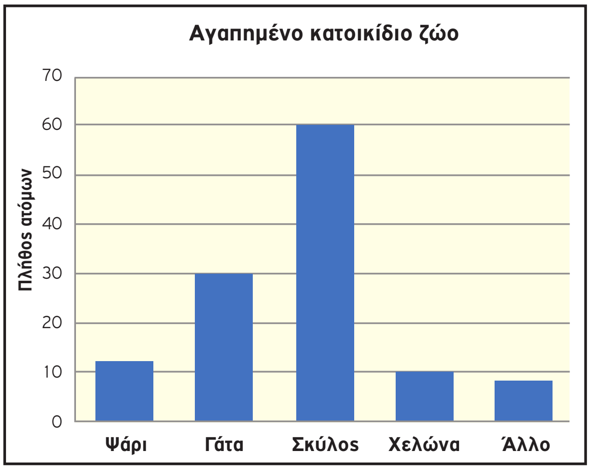
Απάντηση

Για το λόγο ότι η δεύτερη στήλη φαίνεται περίπου διπλάσια στο ύψος από την πρώτη στήλη, δόθηκε η λανθασμένη ερμηνεία. Όμως αυτό που πραγματικά φαίνεται είναι το άνω τμήμα των στηλών, αφού ο άξονας των τεταγμένων αρχίζει από το 500. Έτσι, το πραγματικό ύψος της πρώτης στήλης είναι 508, ενώ της δεύτερης είναι 517. Επομένως ο αριθμός των ληστειών αυξήθηκε μόνο κατά 9 σε σύνολο 500 και πλέον ληστειών. Το γεγονός αυτό δεν δικαιολογεί την παραπάνω ερμηνεία.

Εφαρμογή 4η

Ρωτήθηκαν 120 άτομα για το αγαπημένο τους κατοικίδιο ζώο και τα αποτελέσματα παρουσιάζονται με ένα κυκλικό διάγραμμα και ένα ραβδόγραμμα.





1) Να κάνετε μια ερώτηση που μπορεί να απαντηθεί πιο εύκολα από το ραβδόγραμμα.

2) Να κάνετε μια ερώτηση που μπορεί να απαντηθεί πιο εύκολα από το κυκλικό διάγραμμα.

Απάντηση

1) Για το ραβδόγραμμα μια ερώτηση με εύκολη απάντηση μπορεί να είναι: «Πόσες φορές είναι περισσότερα τα άτομα που το αγαπημένο τους κατοικίδιο ζώο είναι ο σκύλος σε σχέση με αυτά που είναι η χελώνα»;

2) Για το κυκλικό διάγραμμα μια ερώτηση με εύκολη απάντηση μπορεί να είναι: «Ποιο είναι το ποσοστό των ατόμων που το αγαπημένο τους κατοικίδιο ζώο είναι ο σκύλος»;