Γ΄ΛΥΚΕΙΟΥ

ΑΣΚΗΣΕΙΣ ΣΤΗΝ ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Τουρναβίτης Στέργιος

Σκοπός της εργασίας αυτής, είναι να παρουσιάσει κάποιες ασκήσεις που λύνονται με την βοήθεια στατιστικών πινάκων, διαγραμμάτων και να βοηθήσει ταυτόχρονα στην περαιτέρω κατανόηση κάποιων στατιστικών όρων (συχνότητα, σχετική συχνότητα κ.λ.π.) που εμπλέκονται στη λύση αυτών των ασκήσεων.

1. Ο τρόπος μεταφοράς ενός δείγματος φοιτητών σ' ένα Πανεπιστήμιο, βρέθηκε ότι ήταν:

Μεταφορικό μέσο (x_i)	Συχνότητα (ν_i)
Περπάτημα	12
Ποδήλατο	8
Λεωφορείο ή τρόλεϊ	26
Τρένο	33
Αυτοκίνητο	11
Σύνολο	---

Μπορείτε να βοηθήσετε την ομάδα σύνταξης των φοιτητών, να παρουσιάσει τα δεδομένα αυτά του διπλανού πίνακα, σ' ένα κυκλικό διάγραμμα με σκοπό να τα εκθέσει στο μηνιαίο περιοδικό του Πανεπιστημίου;

<u>Λύση</u>:

Το σύνολο των φοιτητών είναι: $12+8+26+11=90$

Αν $\omega_i, i=1,...,5$ οι γωνίες του κυκλικού διαγράμματος που θα φτιάξουμε τότε:

$\omega_i = f_i \cdot 360^\circ, i=1,...,5$,όπου $f_i, i=1,...,5$ η σχετική συχνότητα κάθε μεταφορικού μέσου. $\left(f_i = \frac{\nu_i}{\nu}, i=1,...,5\right)$. Άρα: $\omega_1 = \frac{12}{90}\cdot 360^\circ = 48^\circ$.

Με την ίδια διαδικασία μπορείτε να υπολογίσετε και τις υπόλοιπες γωνίες ώστε να συμπληρώσετε στη συνέχεια την τρίτη στήλη του πίνακα.

x_i	ν_i	ω_i (σε μοίρες)
Περπάτημα	12	48
Ποδήλατο	8	___
Λεωφορείο	26	___
Τρένο	33	___
Αυτοκίνητο	11	___
Σύνολο	90	___

Αν έχετε κάνει σωστά τους υπολογισμούς σας, με τη βοήθεια (του «εξορισμένου» από τις γεωμετρικές κατασκευές μοιρογνωμονίου), το διάγραμμά σας θα πρέπει να μοιάζει με αυτό της παρακάτω εικόνας:

2. Ο παρακάτω (ελλιπής) πίνακας συχνοτήτων και το αντίστοιχο κυκλικό διάγραμμα, αναφέρονται στις προτιμήσεις ενός δείγματος ν μαθητών από διάφορα Λύκεια της Ελλάδας όσον αφορά το αγαπημένο τους άθλημα.

x_i	ν_i	ω_i (σε μοίρες)
Χάντμπολ		50
Μπάσκετ		80
Κολύμπι		60
Βόλεϊ	220	55
Ποδόσφαιρο		
Σύνολο		

α) Πόσοι μαθητές πήραν μέρος στην δειγματοληψία;

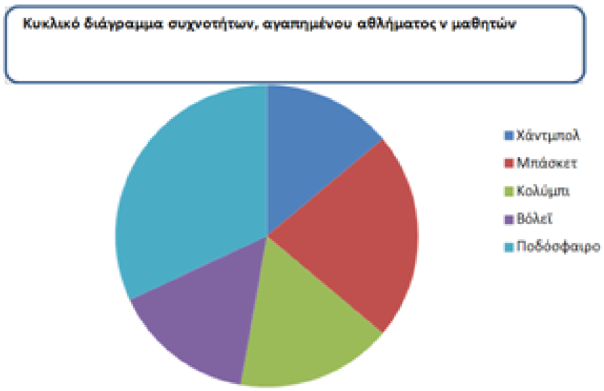
β) Πόσοι από αυτούς είναι φίλοι του ποδοσφαίρου;

γ) Μπορείτε να συμπληρώσετε τα κενά της στήλης των συχνοτήτων και για τα άλλα αθλήματα;

Λύση:

α) Για να βρούμε το σύνολο των μαθητών του δείγματος, ξεκινάμε από την γραμμή του πίνακα συχνοτήτων που είναι συμπληρωμένη. Αυτή δεν είναι άλλη από την προτίμηση στο βόλεϊ. Έχουμε όπως και προηγουμένως: $\omega_5 = f_5 \cdot 360^\circ$.

Άρα: $55^\circ = \dfrac{220}{\nu} \cdot 360^\circ \Leftrightarrow \nu = 1440$



β, γ) Για τους φίλους του ποδοσφαίρου, δεν γνωρίζουμε ούτε πόσοι είναι (**ν_5**), αλλά ούτε και την γωνία (**ω_5**) που αντιστοιχεί στο «κομμάτι της πίτας τους». Μπορούμε ωστόσο να την βρούμε, αν από το **360** (το σύνολο των μοιρών του κύκλου) αφαιρέσουμε το άθροισμα όλων των γωνιών από τα υπόλοιπα αθλήματα που δίνονται.

Έχουμε: $\omega_5 = 360 - (50 + 80 + 60 + 55) = 115$

Έχοντας ξετυλίξει αρκετά το κουβάρι της λύσης, μπορούμε να υπολογίσουμε όχι μόνο τους «λάτρεις» του ποδοσφαίρου, αλλά και αυτούς που συμπαθούν και τα υπόλοιπα αθλήματα από τον τύπο:

$$\omega_i = \frac{\nu_i}{\nu} \cdot 360^\circ \Leftrightarrow \nu_i = \frac{\nu \cdot \omega_i}{360}, i = 1, \ldots, 5$$

π.χ. $\nu_1 = \dfrac{1440 \cdot 50}{360} = 200$

Έτσι αν ολοκληρώσετε σωστά τους υπολογισμούς σας, ο πίνακάς σας θα πρέπει να μοιάζει με τον παρακάτω:

x_i	ν_i	ω_i (σε μοίρες)
Χάντμπολ	200	50
Μπάσκετ	320	80
Κολύμπι	240	60
Βόλεϊ	220	55
Ποδόσφαιρο	460	115
Σύνολο	1440	360

4. Στις προηγούμενες ασκήσεις, μπορείτε να προσθέσετε μία επιπλέον στήλη. Τη στήλη των σχετικών συχνοτήτων επί τοις εκατό $(f_i\%)$

Κρίμα που οι μεταβλητές είναι ποιοτικές και δεν μπορούμε να προσθέσουμε και 4η στήλη, την στήλη των σχετικών αθροιστικών συχνοτήτων επί τοις εκατό $(F_i\%)$, αλλά ούτε καν τη στήλη των αθροιστικών συχνοτήτων (N_i). Οι αθροιστικές συχνότητες έχουν νόημα, μόνο για ποσοτικές μεταβλητές…

Λύση:

Ως γνωστόν ισχύει:

$$f_i\% = f_i \cdot 100 = \frac{\nu_i}{\nu} \cdot 100, i = 1, 2, 3, 4, 5$$

π.χ. $f_1\% = \dfrac{12}{90} \cdot 100 \approx 13,33$.

Την τελευταία σχετική συχνότητα επειδή κάνουμε προσέγγιση στους υπολογισμούς μας θα την βρούμε αν αφαιρέσουμε από το 100 το σύνολο των υπολοίπων.

Στην άσκηση 1:

x_i	ν_i	ω_i (σε μοίρες)	$f_i\%$
Περπάτημα	12	48	13,33
Ποδήλατο	8	32	8,89
Λεωφορείο	26	104	28,89
Τρένο	33	132	36,67
Αυτοκίνητο	11	44	12,22
Σύνολο	90	360	100

Στην άσκηση 2:

x_i	ν_i	ω_i (σε μοίρες)	$f_i\%$
Χάντμπολ	200	50	13,89
Μπάσκετ	320	80	22,22
Κολύμπι	240	60	16,67
Βόλεϊ	220	55	15,28
Ποδόσφαιρο	460	115	31,94
Σύνολο	1440	360	100

Ας δούμε τώρα τι γίνεται και με τις ποσοτικές – διακριτές μεταβλητές, όπως τη ρίψη ενός ζαριού, όπου μπορούμε να ασχοληθούμε και με δύο ακόμη συχνότητες (απόλυτες και σχετικές) αλλά αθροιστικές…

5. Ο παρακάτω ελλιπής πίνακας, είναι το αποτέλεσμα (άγνωστου αριθμού, ν) ρίψεων ενός ζαριού σ' ένα επιτραπέζιο παιχνίδι (φιδάκι).

x_i	$ν_i$	N_i	f_i	F_i	$f_i\%$	$F_i\%$
1		8				
2				0,4		
3	6			0,52		
4		35				
5						
6			0,22			
Σύνολο						

Α) Μπορείτε να συμπληρώσετε τα άδεια κελιά του παραπάνω πίνακα;

Β) Από τον πίνακα αυτόν, να εκτιμήσετε το ποσοστό των ρίψεων που είναι:

i) μικρότερο του 3,

ii) 5 ή 6,

iii) τουλάχιστον 2, αλλά το πολύ 5.

Γ) Στη συνέχεια να κατασκευάσετε το πολύγωνο των σχετικών συχνοτήτων $f_i\%$.

Λύση:

Α) Αρχίζουμε από, $ν_1 = N_1 = 8$, $F_2\% = F_2 \cdot 100 = 0{,}4 \cdot 100 = 40$. Όμοια, $F_3\% = F_3 \cdot 100 = \ldots = 52$. Στη συνέχεια μπορούμε να εκμεταλλευτούμε τις δύο συνεχόμενες τιμές F_2 και F_3 (στη στήλη των αθροιστικών συχνοτήτων που δίνονται) και από το γνωστό «τριγωνάκι» να υπολογίσουμε την άγνωστη συχνότητα f_3.

	0,4
f_3	0,52

Έχουμε: $0{,}4 + f_3 = 0{,}52 \Rightarrow f_3 = 0{,}52 - 0{,}40 = 0{,}12$.

Ο υπολογισμός αυτής της σχετικής συχνότητας f_3, σε συνδυασμό με την $ν_3 = 6$, θα μας επιτρέψει να βρούμε το μέγεθος $ν$ του δείγματος. Χρησιμοποιώντας έτσι τον ορισμό της υπολογισθείσας σχετικής συχνότητας f_3, έχουμε: $f_3 = \frac{ν_3}{ν} \Rightarrow ν = \cdots = 50$. Αυτό με την σειρά του θα «ξεκλειδώσει» μια σειρά από άγνωστες συχνότητες, όπως την $f_1 = \frac{ν_1}{ν} \Rightarrow f_1 = \cdots = 0{,}16$. Χρησιμοποιώντας ορισμούς και τεχνικές όπως περιγράφηκαν παραπάνω, ο πίνακάς μας θα μοιάζει όπως ο παρακάτω:

Με **έντονο μωβ χρώμα**, έχουμε σημειώσει τα κελιά του πίνακα που έχουμε υπολογίσει.

x_i	$ν_i$	N_i	f_i	F_i	$f_i\%$	$F_i\%$
1	**8**	8	**0,16**	**0,16**	**16**	**16**
2	**12**	**20**	**0,24**	0,4	**24**	**40**
3	6	**26**	**0,12**	0,52	**12**	**52**
4	**9**	35	**0,18**	**0,7**	**18**	**70**
5						
6	**11**		0,22			
Σύνολο	**50**		**1**		**100**	

Στο σημείο αυτό είναι σκόπιμο να αναφέρουμε την «τεχνική» υπολογισμού της άγνωστης συχνότητας $ν_5$, που βρίσκεται σε μία γραμμή που όλα της τα κουτάκια είναι άγνωστα, αξιοποιώντας ξανά το σύνολο **ν** των ρίψεων που υπολογίσαμε προηγουμένως.

$ν_1 + ν_2 + ν_3 + ν_4 + ν_5 + ν_6 = ν \Rightarrow ν_5 = 50 - \ldots = 4$.

Η συνέχεια δεν θα μας δυσκολέψει αφού γίνεται με παρόμοιες διαδικασίες. Ο ολοκληρωμένος

x_i	$ν_i$	N_i	f_i	F_i	$f_i\%$	$F_i\%$
1	**8**	8	**0,16**	**0,16**	**16**	**16**
2	**12**	**20**	**0,24**	0,4	**24**	**40**
3	6	**26**	**0,12**	0,52	**12**	**52**
4	**9**	35	**0,18**	**0,7**	**18**	**70**
5	**4**	**39**	**0,08**	**0,78**	**8**	**78**
6	**11**	**50**	0,22	**1**	**22**	**100**
Σύνολο	**50**		**1**		**100**	

πίνακας θα πρέπει να μοιάζει με τον παρακάτω:

Β) i) Παρατηρούμε ότι οι φορές που ήρθε το ζάρι 1 ή 2 είναι **8+12 = 20** και το αντίστοιχο ποσοστό είναι: $\frac{20}{50} = 40\%$. Αυτό μπορούμε να το βρούμε και κατευθείαν από την $F_2\% = 40$.

ii) Το ποσοστό των ρίψεων που είναι 5 ή 6, βρίσκεται αν υπολογίσουμε το άθροισμα, $f_5\% + f_6\% = 8 + 22 = 30$.

iii) Παρόμοια βρίσκουμε το ποσοστό των ρίψεων που είναι από 2 μέχρι και 5. Επιχειρώντας μία διαφορετική προσέγγιση από τις προηγούμενες, μπορούμε να πούμε ότι αυτό το ποσοστό μπορεί να υπολογισθεί και από τον τύπο: $F_5\% - F_1\% = \ldots = 62$.

Γ)

Αυτά με τις κατανομές, ποιοτικές και ποσοτικές που το χαρακτηριστικό, οι τιμές x_i της μεταβλητής είναι μικρό. Τι γίνεται όμως σε μία έρευνα, όταν έχουμε ***ποσοτική*** *διακριτή και πολύ περισσότερο συνεχή μεταβλητή* και το πλήθος των τιμών της είναι αρκετά μεγάλο; Σε τέτοιου είδους περιπτώσεις, είναι απαραίτητο να κάνουμε μία ομαδοποίηση των παρατηρήσεων προκειμένου να κατασκευάσουμε τα διαγράμματά μας, αλλά και να εξάγουμε γενικά τα συμπεράσματά μας.

6. Οι χρόνοι (σε λεπτά) που παρατηρήθηκαν σε 25 δρομολόγια τρένων, δίνονται από τον διπλανό πίνακα.

2	4	6	7	3
11	4	5	6	8
7	9	6	8	4
8	4	7	9	11
5	7	6	2	10

Α) Να ομαδοποιήσετε τα δεδομένα σε 5 κλάσεις ίσου πλάτους και να κατασκευάσετε τον πίνακα κατανομής ν_i, N_i, f_i, F_i, $f_i\%$, $F_i\%$.
Β) Να κατασκευάσετε το ιστόγραμμα και το πολύγωνο,
Β1) συχνοτήτων,
Β2) σχετικών αθροιστικών συχνοτήτων.
Γ) Να βρείτε το μέσο χρόνο καθυστερήσεων των δρομολογίων.
Δ) Ποιο είναι το ποσοστό των δρομολογίων που είχαν καθυστέρηση τουλάχιστον 7 λεπτά;
<u>Λύση</u>:

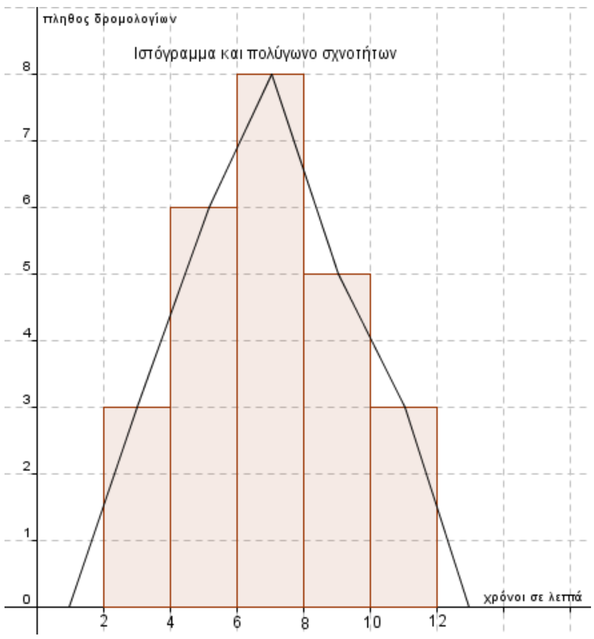
Α) Το πλάτος c κάθε κλάσης το βρίσκουμε αν διαιρέσουμε το εύρος **R** (*Range*) της διαφοράς της μικρότερης από την μεγαλύτερη τιμή, δια του πλήθους κ των κλάσεων. Στην άσκηση μας είναι $\boldsymbol{R} = \mathbf{11} - \mathbf{2}$ και $\boldsymbol{\kappa} = \mathbf{5}$.

Έτσι, $\boldsymbol{c} = \frac{\boldsymbol{R}}{\boldsymbol{\kappa}} = \mathbf{1{,}8} \cong \mathbf{2}$. Αρχίζοντας από *την μικρότερη παρατηρούμενη καθυστέρηση **2** και προσθέτοντας κάθε φορά το **2**, δημιουργούμε τις **5** κλάσεις*. Αν υπολογίσουμε κατά τα γνωστά και τα υπόλοιπα στοιχεία, καταλήγουμε στον παρακάτω πίνακα:

	x_i	ν_i	N_i	f_i	F_i	$f_i\%$	$F_i\%$	$x_i \cdot \nu_i$
[2,4)	3	3	3	0,12	0,12	12	12	9
[4,6)	5	6	9	0,24	0,36	24	36	30
[6,8)	7	8	17	0,32	0,68	32	68	56
[8,10)	9	5	22	0,2	0,88	20	88	45
[10,12)	11	3	25	0,12	1	12	100	33
Σύνολο		25		1		100		173

Οι κεντρικές τιμές είναι οι Μέσοι Όροι κάθε κλάσης.

Β1 & Β2) Για το πολύγωνο συχνοτήτων θεωρούμε στην αρχή και στο τέλος δύο υποθετικές κλάσεις μηδενικής συχνότητας και στη συνέχεια ενώνουμε με ευθύγραμμα τμήματα τα μέσα των άνω βάσεων των ορθογωνίων.



Για να κατασκευάσουμε το ιστόγραμμα των σχετικών αθροιστικών συχνοτήτων, ενώνουμε αυτή τη φορά τα δεξιά άκρα (και όχι τα μέσα) των άνω βάσεων των ορθογωνίων.

Γ) Ο Μέσος χρόνος των καθυστερήσεων είναι ουσιαστικά η Μέση Τιμή της ομαδοποιημένης κατανομής. Στον υπολογισμό της θα μας βοηθήσει η στήλη των κεντρικών τιμών κάθε κλάσης που αντιπροσωπεύουν κάθε παρατήρηση της κλάσης τους. Σύμφωνα με τον τύπο,

$\bar{x} = \dfrac{\sum_{i=1}^{5} x_i \cdot \nu_i}{25}$, έχουμε: $\bar{x} = \dfrac{173}{25} = 6,92$.

Δ) Παρατηρούμε (*από τον αρχικό πίνακα*) ότι υπάρχουν **4** δρομολόγια τραίνων που καθυστέρησαν ακριβώς **7** λεπτά και **5 + 3** δρομολόγια (*από τον πίνακα συχνοτήτων* $\mathbf{\nu_i}$) με καθυστέρηση από ***8*** *λεπτά και πάνω*.

Το αντίστοιχο ποσοστό θα είναι: $\frac{4+5+3}{25} = \mathbf{0,48}$.