ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ (ΚΕΦ. 2ο) !

Παρακάτω δίνονται τα αποτελέσματα της μέτρησης του βάρους γεννήσεως 25 βρεφών σε γραμμάρια

2700	4200	4100	3350	3100	2950	3550	1900	3300
2500	2600	2800	2000	3450	2850	3540	3470	2800
3040	4070	2850	3290	3850	3150	3150	2050	

i) Να ομαδοποιήσετε τα δεδομένα σε 5 ομαδοποιημένες μετρήσεις ίσου μήκους.
ii) Να κατασκευάσετε τον πίνακα κατανομής συχνοτήτων, σχετικών συχνοτήτων, αθροιστικών συχνοτήτων και αθροιστικών σχετ. συχνοτήτων
iii) Να κατασκευάσετε το πολύγωνο σχετικών συχνοτήτων και αθροιστικών σχετικών συχνοτήτων επί τις εκατό (%)

ΛΥΣΗ

i) Έχουμε το εύρος των παρατηρήσεων

$R = \max x_i - \min x_i = 4200 - 1900 = 2300$

Επίσης, $k=5$ ομαδ. μετρήσεις

Άρα, το μήκος κάθε ομάδας είναι:

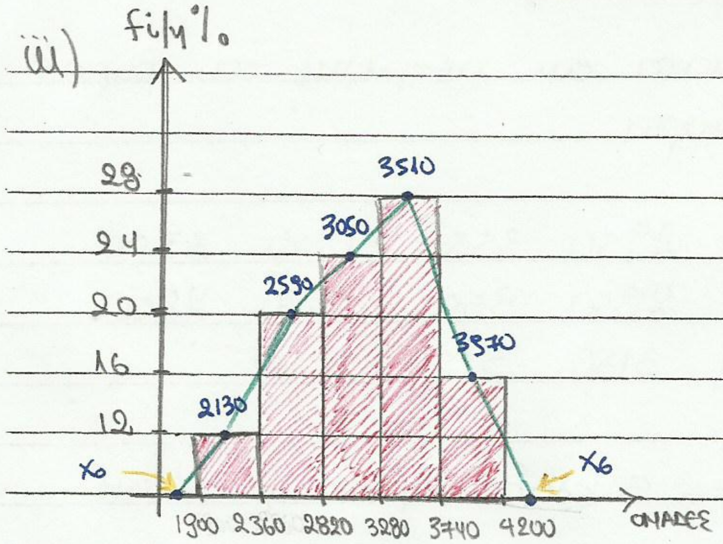
$d = \frac{R}{K} = \frac{2300}{5} = 460$

ii) ΠΙΝΑΚΑΣ

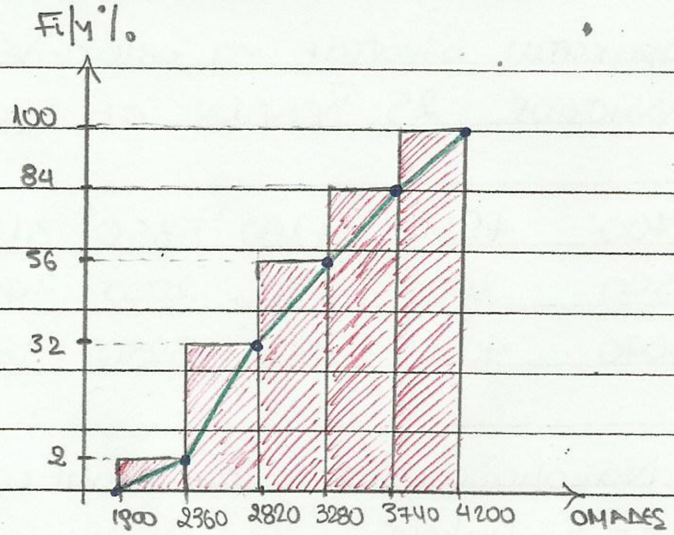
κεντρική τιμή

i-ομάδα	$[L_i, U_i]$ όρια	X_i	f_i	f_i/n	$f_i/n\%$	F_i	$F_i/n\%$
1	1900 - 2360	2130	3	0.12	12	3	12
2	2360 - 2820	2590	5	0.20	20	8	32
3	2820 - 3280	3050	6	0.24	24	14	56
4	3280 - 3740	3510	7	0.28	28	21	84
5	3740 - 4200	3970	4	0.16	16	25	100
Σύνολο			25	1.00	100		

iii)



ΠΟΛΥΓΩΝΟ ΣΧΕΤ. ΣΥΧΝΟΤΗΤΩΝ



ΠΟΛΥΓΩΝΟ ΑΘΡΟΙΣΤ. ΣΧΕΤ. ΣΥΧΝΟΤΗΤΩΝ

■ Να βρεθεί η μέση τιμή των παρακάτω παρατηρήσεων:

i. −2, 3, 2, 1, 6

ii. που δίνονται από τους πίνακες

(α)

x_i	−2	0	1	2
f_i	3	6	6	5

, (β)

x_i	−3	−2	0	1
f_i/n	0.1	0.3	0.4	0.2

και

(γ)

ΟΜΑΔΕΣ	3−5	5−7	7−9	9−11
f_i/n %	15	20	35	30

ΛΥΣΗ

i. $\bar{x} = \dfrac{-2+3+2+1+6}{5} = \dfrac{10}{5} = 2$

ii. (α) $\bar{x}_1 = \dfrac{-2\cdot3+0\cdot6+1\cdot6+2\cdot5}{3+6+6+5} = \dfrac{10}{20} = \dfrac{1}{2}$

(β) $\bar{x}_2 = \dfrac{-3\cdot0{,}1-2\cdot0{,}3+0\cdot0{,}4+1\cdot0{,}2}{0{,}1+0{,}3+0{,}4+0{,}2} = -\dfrac{7}{10}$

(γ) $\bar{x}_3 = \sum_{i=1}^{4} x_i \dfrac{f_i}{n} \Rightarrow 100\,\bar{x}_3 = \sum_{i=1}^{4} x_i \dfrac{f_i}{n}\% \Rightarrow$

$\Rightarrow 100\,\bar{x}_3 = 4\cdot15+6\cdot20+8\cdot35+10\cdot30 \Rightarrow \bar{x}_3 = 7{,}6$

(όπου $x_i = 4, 6, 8, 10$ είναι οι τιμές του δείγματος)

Να βρείτε τη διάμεσο στα δείγματα που έχουν τους εξής πίνακες:

i)

x_i	3	4	5	6
f_i	6	7	3	5

ii)

x_i	1	0	5	3
f_i	15	5	8	12

iii)

x_i	-1	2	3	5
$f_i/\nu\%$	20	30	35	15

iv)

x_i	0	1	3	4
$f_i/\nu\%$	10	20	45	25

ΛΥΣΗ

i)

x_i	f_i
3	6 → 6ος όρος
4	7 → 13ος όρος
5	3 → 16ος όρος
6	5 → 21ος όρος

$f = 21$ περιττός

$\frac{f}{2} = \frac{21}{2} = 10,5 \rightarrow \delta = 11^{ος}$ όρος

Άρα, $\delta = 4$

ii)

Διάταξη

x_i	f_i
0	5 → 5ος όρος
1	15 → 20ος όρος
3	12 → 32ος όρος
5	8 → 40ος όρος

$f = 40$ άρτιος

$\frac{f}{2} = 20 \rightarrow \delta = \frac{20^{ος} + 21^{ος}}{2} = \frac{1+3}{2} = 2$

Άρα, $\delta = 2$

iii)

x_i	$f_i/\nu\%$	$F_i/\nu\%$
-1	20	20
2	30	50
3	35	85
5	15	100

$F_2/\nu\% = 50$ Άρα, εφόσον δ είναι το σημείο με την ιδιότητα 50% των μετρήσεων να 'ναι μικρότερες ή ίσες από αυτό και τουλάχιστον 50% να είναι μεγαλύτερες από αυτό. Τότε, $\delta = \frac{2+3}{2} = 2,5$

iv)

x_i	$f_i/\nu\%$	$F_i/\nu\%$
0	10	10
1	20	30 ← 50%
3	45	75
4	25	100

Έχουμε ότι $F_2/\nu\% = 30 < 50 < 75 = F_3/\nu\%$

Άρα, η διάμεσος έχει τιμή που αντιστοιχεί το $F_3/\nu\%$.

Άρα, $\delta = 3$.

Να βρείτε τη διάμεσο που έχουν οι εξής πίνακες:

i)

ΟΜΑΔΕΣ	$f_i/\nu\%$
[0,3)	15
[3,6)	35
[6,9)	40
[9,12)	10

ii)

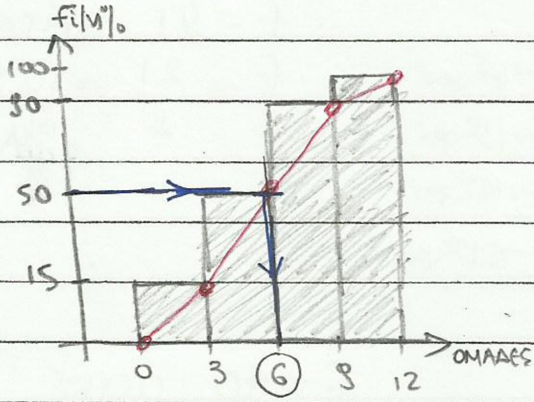
ΟΜΑΔΕΣ	$f_i/\nu\%$
[1,3)	10
[3,5)	20
[5,7)	40
[7,9)	30

ΛΥΣΗ.

i)

$F_i/\nu\%$
15
50
90
100

τότε



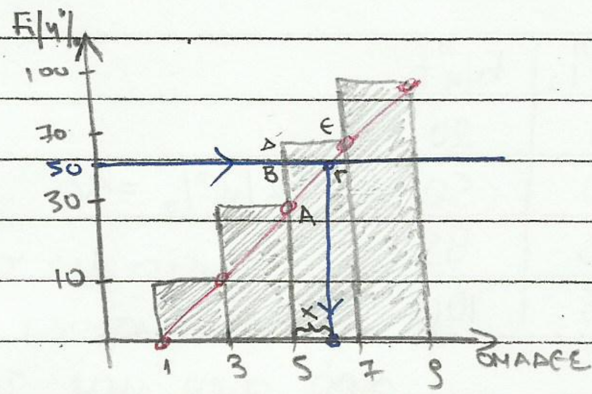
$\delta = 6.$

(*) Αν μας δίνει f_i και όχι $f_i/\nu\%$ τότε εργαζόμαστε ανάλογα, βρίσκοντας τις τιμές της αθροιστικής συχνότητας αλλά σαν ενδιάμεση τιμή δεν λαμβάνουμε το 50% αλλά το μισό του πλήθους $\left(\frac{f}{2}\right)$

ii)

$F_i/\nu\%$
10
30
70
100

τότε



Τα τρίγωνα $\overset{\triangle}{AB\Gamma}$ και $\overset{\triangle}{A\Delta E}$ είναι όμοια

άρα: $\frac{AB}{A\Delta} = \frac{B\Gamma}{\Delta E} \Rightarrow \frac{50-30}{70-30} = \frac{x}{5-3} \Leftrightarrow \frac{20}{40} = \frac{x}{2} \Leftrightarrow x=1$

Άρα, $\delta = 5+1 = 6.$

Η βαθμολογία μιας ομάδας φοιτητών σε ένα μάθημα δίνεται από τον πίνακα:

ΒΑΘΜΟΛΟΓΙΑ (x_i)	ΦΟΙΤΗΤΕΣ (f_i)
5	4
6	5
7	10
8	1

Να υπολογίσετε
α. Το εύρος
β. Τη διασπορά
γ. Την τυπική απόκλιση του δείγματος

ΛΥΣΗ

α. $R = 8 - 5 = 3$

β. $\bar{x} = \dfrac{5.4 + 6.5 + 7.10 + 8.1}{4+5+10+1} = \dfrac{128}{20} = 6{,}4$

Άρα,

$$S'^2 = \frac{1}{f}\left\{\sum_{i=1}^{4} x_i^2 f_i - \frac{\left(\sum_{i=1}^{4} x_i f_i\right)^2}{f}\right\} \quad (1)$$

- όπου η στήλη $x_i v_i$ είναι:

$x_i f_i$:	20	30	70	8	ΣΥΝΟΛΟ: 128

- η στήλη $x_i^2 \cdot v_i$ είναι

$x_i^2 \cdot f_i$:	100	180	490	64	ΣΥΝΟΛΟ: 834

$$(1): \; S'^2 = \frac{1}{20} \cdot \left\{ 834 - \frac{128^2}{20} \right\} = \frac{834}{20} - \frac{16384}{400} = 0{,}74$$

Για την S^2 πιο συχνός τύπος είναι ο:

$$S^2 = \frac{1}{f-1}\left\{\sum_{i=1}^{4} x_i^2 f_i - \frac{\left(\sum_{i=1}^{4} x_i f_i\right)^2}{f}\right\} = \frac{834}{19} - \frac{16384}{380} =$$

$$= 43{,}89 - 43{,}11 = 0{,}78$$

Οι S'^2 και S^2 εκφράζουν την ίδια στατιστική έννοια (τη διασπορά)

γ. $S' = \sqrt{S'^2} = \sqrt{0{,}74}$ και $S = \sqrt{S^2} = \sqrt{0{,}78}$

Στον διπλανό πίνακα δίνονται οι ελάχιστες θερμοκρασίες (σε °C) στην Αθήνα και στη Θεσ/νίκη αντίστοιχα για 4 συνεχείς ημέρες. Να βρείτε ποιο από τα δύο δείγματα παρουσιάζει μεγαλύτερη ομοιογένεια (ή ακρίβεια). (Υπολογίστε την S')

Αθήνα	0	1	2	3
Θεσ/νίκη	1	2	4	5

ΛΥΣΗ

Έστω το δείγμα α για την Αθήνα και δείγμα β για τη Θεσ/νίκη.
Έτσι έχουμε ότι:

$\bar{x}_\alpha = \frac{3}{2}$ και $\bar{x}_\beta = 3$

Θα βρούμε την S'_α και S'_β
Για την S'_α, θα ισχύει:

$$S_\alpha'^2 = \frac{(0-\frac{3}{2})^2+(1-\frac{3}{2})^2+(2-\frac{3}{2})^2+(3-\frac{3}{2})^2}{4} = \frac{5}{4} \Rightarrow S'_\alpha = \frac{\sqrt{5}}{2}$$

Για την S'_β, θα ισχύει:

$$S_\beta'^2 = \frac{(1-3)^2+(2-3)^2+(4-3)^2+(5-3)^2}{4} = \frac{10}{4} \Rightarrow S'_\beta = \frac{\sqrt{10}}{2}$$

Έτσι, ο συντελεστής μεταβλητότητας των μετρήσεων για κάθε ένα από τα δείγματά μας είναι:

$$CV_\alpha = \frac{S'_\alpha}{\bar{x}_\alpha} = \frac{\sqrt{5}}{3} \quad \text{και} \quad CV_\beta = \frac{S'_\beta}{\bar{x}_\beta} = \frac{\sqrt{10}}{6}$$

Έστω $CV_\alpha > CV_\beta \sim \frac{\sqrt{5}}{3} > \frac{\sqrt{10}}{6} \Rightarrow 2\sqrt{5} > \sqrt{10} \Rightarrow \sqrt{20} > \sqrt{10}$ ισχύει

Άρα, $CV_\alpha > CV_\beta$

Δηλαδή το δείγμα β είναι περισσότερο ακριβής (ή έχει μεγαλύτερη ακρίβεια ή είναι πιο ομοιογενές) από το δείγμα α.