ΕΡΓΑΣΤΗΡΙΟ ΠΟΣΟΤΙΚΩΝ ΜΕΘΟΔΩΝ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΤΗΣ ΔΙΟΙΚΗΣΗΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
“ΜΕΤΑΠΤΥΧΙΑΚΟ ΔΙΠΛΩΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ – MBA”

ΣΤΑΤΙΣΤΙΚΗ
ΚΑΙ
ΘΕΩΡΙΑ ΑΠΟΦΑΣΕΩΝ

Επιλεγμένες Ασκήσεις με ενδεικτική λύση στην Περιγραφική Στατιστική

ΑΣΚΗΣΗ 1

Ο Πίνακας 1 δίνει την κατανομή συχνότητας των μισθών τριάντα υπαλλήλων μιας δημόσιας υπηρεσίας.

Πίνακας 1

Μισθός (χ. μ.)	**Αριθμός Υπαλλήλων**
60 – 70	7
70 – 80	14
80 – 90	5
90 – 100	3
100 – 110	1

Δίνεται επίσης ότι η τυπική απόκλιση των μισθών τους είναι 10,25 χ. μ.

i. Να υπολογισθούν ο αριθμητικός μέσος και η τεταρτημοριακή απόκλιση των μισθών.
ii. Αν υποθέσουμε ότι για την κατανομή των μισθών 50 υπαλλήλων μιας άλλης δημόσιας υπηρεσίας ο αριθμητικός μέσος και η τυπική απόκλιση είναι 109,56 χ. μ. και 13,17 χ. μ. αντίστοιχα να εξετασθεί σε ποια από τις δύο ομάδες υπαλλήλων παρατηρείται μεγαλύτερη διασπορά των μισθών.

ΛΥΣΗ

i.

Μισθός (χ. μ.)	**Κεντρική Τιμή** m_i	**Αριθμός Υπαλλήλων** f_i	**Αθροιστική Συχνότητα** F_i	$f_i m_i$
60 – 70	65	7	7	455
70 – 80	75	14	21	1050
80 – 90	85	5	26	425
90 – 100	95	3	29	285
100 – 110	105	1	30	105
ΣΥΝΟΛΑ		30		2320

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i m_i}{n} = \frac{2320}{30} = 77.3333 \rightarrow \bar{x} \cong 77.33$$

Εντοπισμός θέσης Q_1 : $\frac{n*i}{4} = \frac{30*1}{4} = \frac{30}{4} = 7.5$. Άρα το Q_1 ανήκει στην 2η τάξη.

(διάστημα 70 - < 80)

Εντοπισμός θέσης Q_3 : $\frac{n*i}{4} = \frac{30*3}{4} = \frac{90}{4} = 22.5$. Άρα το Q_3 ανήκει στην 3η τάξη.

(διάστημα 80 - < 90)

Υπολογισμός τιμής Q_1:

$$Q_1 = L_{Q_1} + \frac{\delta}{f_{Q_1}}\left(\frac{n*i}{4} - F_{Q_{1-1}}\right) = 70 + 10\frac{(7.5-7)}{14} = 70 + 10\frac{0.5}{14} = 70 + 10(0.036) =$$
$$= 70 + 0.36 \rightarrow Q_1 \cong 70.36$$

Υπολογισμός τιμής Q_3 :

$$Q_3 = L_{Q_3} + \frac{\delta}{f_{Q_3}}\left(\frac{n*i}{4} - F_{\Pi_{Q_{3-1}}}\right) = 80 + 10\frac{(22,5-21)}{5} = 80 + 10\frac{15}{5} = 80 + 10(0,3) =$$
$$= 80 + 3 \rightarrow Q_3 = 83$$

Υπολογισμός τεταρτημοριακής απόκλισης:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{83 - 70.36}{2} = \frac{12.64}{2} = 6.32 \rightarrow Q = 6.32$$

ii.

$$C = \frac{s}{\overline{x}}*100 = \frac{10.25}{77.33}*100 \rightarrow C \cong 13.25\%$$

$$C_1 = \frac{s_1}{\overline{x}_1}*100 = \frac{13.17}{109.56}*100 \rightarrow V_1 \cong 12.20\%$$

$C > C_1$: Άρα σε σχετικές τιμές η αρχική ομάδα παρουσιάζει μεγαλύτερη διασπορά.

ΑΣΚΗΣΗ 2

Τα ακόλουθα δεδομένα εκφράζουν τον αριθμό των αυτοκινήτων που φτάνουν σε ένα σταθμό διοδίων σε 10 τυχαία επιλεγμένα χρονικά διαστήματα των 10 λεπτών.

26 24 22 15 33 18 20 34 27 30

Να υπολογιστούν:

i. η διάμεσος
ii. το εύρος
iii. το τρίτο τεταρτημόριο των δεδομένων

ΛΥΣΗ

Κατατάσσω τα δεδομένα κατά αύξουσα τάξη:

15, 18, 20, 22, 24, 26, 27, 30, 33, 34

i. $$M = X_{\frac{n}{2}+\frac{1}{2}} = X_{\frac{n+1}{2}} = X_{\frac{n}{2}} + \frac{1}{2}\left(X_{\frac{n}{2}+1} - X_{\frac{n}{2}}\right) = X_5 + \frac{1}{2}\left(X_6 - X_5\right) =$$
$$= 24 + \frac{1}{2}(26-24) = 24 + \frac{1}{2}*1 \rightarrow M = 25$$

ii. $R = X_{\max} - X_{\min} = 34 - 15 \rightarrow R = 19$

iii. $$Q_3 = X_{\frac{3(n+1)}{4}} = X_{\frac{3*11}{4}} = X_{\frac{33}{4}} = X_{8.25} = X_8 + 0.25\left(X_9 - X_8\right) = 30 + 0.25(33-30) =$$
$$= 30 + 0.25*3 = 30 + 0.75 \rightarrow Q_3 = 30.75$$

ΑΣΚΗΣΗ 3

Μια αλυσίδα καταστημάτων έχει κατατάξει τα προϊόντα που διαθέτει σε 5 βασικές κατηγορίες. Ο παρακάτω πίνακας 2 παρουσιάζει την αξία των ετησίων πωλήσεων της εταιρείας, ανά κατηγορία προϊόντων, για το έτος 1997:

Πίνακας 2

ΚΑΤΗΓΟΡΙΑ	1	2	3	4	5
ΑΞΙΑ ΠΩΛΗΣΕΩΝ (σε χ.μ.)	250	50	75	25	100

Να υπολογιστεί ο βαθμός συγκέντρωσης και να σχολιαστεί.

ΛΥΣΗ

ΚΑΤΗΓΟΡΙΑ ΠΡΟΪΟΝΤΟΣ (i)	Xi	$\frac{Xi}{X}$	$\left(\frac{Xi}{X}\right)^2$
1	250	0.50	0.25
2	50	0.10	0.01
3	75	0.15	0.0285
4	25	0.05	0.0025
5	100	0.20	0.04
ΣΥΝΟΛΑ	500	1.00	0.3250

Θα χρησιμοποιήσω το συντελεστή Gini – Hirschman ο οποίος εκφράζει το βαθμό συγκέντρωσης για δεδομένα ταξινομημένα με βάση κάποιο ποιοτικό χαρακτηριστικό. Η γενική του μορφή είναι :

$$C = 100\sqrt{\sum_{i=1}^{n}\left(\frac{X_i}{X}\right)^2}$$

Στην προκειμένη περίπτωση:

$$C = 100\sqrt{0,325} \Rightarrow C = 100 * 0,57 \Rightarrow C = 57,0$$

Όπως είναι γνωστό ο συντελεστής C παίρνει γενικά τιμές στο διάστημα:

[$\frac{100}{\sqrt{n}}$: τέλεια ισοκατανομή, 100: απόλυτη συγκέντρωση].

Στη συγκεκριμένη περίπτωση ο συντελεστής C παίρνει τιμές στο διάστημα:

[$\frac{100}{\sqrt{5}} = 44,7$, 100].

Συνεπώς, συμπεραίνουμε ότι υπάρχει ελαφρά συγκέντρωση των ετησίων εξαγωγών της βιομηχανίας.

ΑΣΚΗΣΗ 4

Ο Πίνακας 3 δίνει την κατανομή συχνοτήτων των μηνιαίων υπερωριακών αποδοχών 100 εργαζομένων μιας επιχείρησης:

Πίνακας 3

Μηνιαίες Αποδοχές (δρχ.)	**Αριθμός Εργαζομένων**
20-<30	40
30-<40	25
40-<50	3
50-<60	1
60-<70	1
70-<80	30

Να υπολογιστούν:

i. Το Σωματείο των εργαζομένων της εταιρείας καταγγέλλει τη Διοίκηση για άνιση κατανομή των υπερωριών μεταξύ των εργαζομένων. Με βάση τα στοιχεία του Πίνακα 3 θα συμφωνούσατε ή όχι με τις καταγγελίες του Σωματείου; Αιτιολογείστε την απάντηση σας.

ii. Να προσδιοριστεί το διάστημα το οποίο δεν περιλαμβάνει το 30% των μικρότερων παρατηρήσεων και το 25% των μεγαλυτέρων.

ΛΥΣΗ

Μισθός (χ. μ.)	**Κεντρική Τιμή** m_i	**Αριθμός Υπαλλήλων** f_i	**Αθροιστική Συχνότητα** F_i	$f_i m_i$	$n-F_i$	$(n-F_i)F_i$
20 –< 30	25	40	40	1000	60	2400
30 –< 40	35	25	65	875	35	2275
40 –< 50	45	3	68	135	32	2176
50 –< 60	55	1	69	55	31	2139
60 –< 70	65	1	70	65	30	2100
70 –< 80	75	30	100	2250	0	0
ΣΥΝΟΛΑ		100		4380		11090

i.

Για να διαπιστωθεί αν ευσταθούν οι καταγγελίες του Σωματείου περί ανισοκατανομής των αποδοχών θα πρέπει να υπολογισθεί ο Συντελεστής Gini, για δεδομένα ταξινομημένα με βάση κάποιο ποσοτικό χαρακτηριστικό. Η γενική του μορφή είναι:

$$g = d / 2\bar{x}$$

όπου $d = \dfrac{2\delta}{n^2}\sum_{i=1}^{k}\left(n - F_i\right)F_i$ (d: μέση διαφορά Gini)

Στη προκειμένη περίπτωση:

$$\bar{X} = \frac{\sum_{i=1}^{k} f_i m_i}{n} = 43,80$$

$$d = \frac{2*10}{100^2}*11090 = \frac{2*10*11090}{100*100} = \frac{1109}{50} \rightarrow \quad d \cong 22,18$$

Άρα $g = \dfrac{22,18}{2*43,8} = \dfrac{22,18}{87,6} \rightarrow \quad g \cong 0,253$ (Γενικά: $0 \le g \le 1$)

Συνεπώς συμπεραίνουμε ότι η ανισοκατανομή των αποδοχών είναι περίπου το 25% της μέγιστης δυνατής ανισοκατανομής και κατά συνέπεια οι ισχυρισμοί του Σωματείου ευσταθούν σε μεγάλο βαθμό.

ii.

<u>Εντοπισμός θέσης D_3</u>: $\dfrac{n_i}{10} = \dfrac{100*3}{10} = 30$ Άρα το D_3 ανήκει στην 1η τάξη.

(διάστημα 20-<30)

<u>Υπολογισμός τιμής D_3</u>:

$$D_3 = L_{D_i} + \frac{\delta}{f_{D_i}}\left[\frac{n*i}{10} - F_{D_{3-1}}\right] = 20 + \frac{10}{40}\left(30 - 0\right) = 20 + \frac{30}{4} \rightarrow \quad D_3 = 27,5$$

<u>Εντοπισμός θέσης Q_3</u>:

$$\frac{n*i}{4} = \frac{100*3}{4} = \frac{300}{4} = 75$$

Άρα το Q_3 ανήκει στην 6η τάξη (διάστημα 60 - < 70)

<u>Υπολογισμός τιμής Q_3</u>:

$$Q_3 = L_{Q3} + \frac{\delta}{f_{Q3}}\left(\frac{n*i}{4} - F_{Q3-1}\right) = 70 + \frac{10}{30}\left(75 - 70\right) = \frac{2150}{30}$$

$$\rightarrow Q_3 \cong 71,67$$

Άρα το ζητούμενο διάστημα είναι (27.5, 71.67).

ΑΣΚΗΣΗ 5

Τα ποσοστά κέρδους των 4 δραστηριοτήτων ($Δ_1$-$Δ_4$) μιας εταιρείας για το 1998 ήταν 4.2%, 5.5%, 7.4%, και 10.1% επί των εσόδων αντίστοιχα. Να υπολογιστεί το μέσο ποσοστό κέρδους της εταιρείας αν γνωρίζουμε ότι τα έσοδα της από τις τέσσερις αυτές δραστηριότητες για το 1998 ήταν 300, 200, 50 και 30 εκ. δρχ. αντίστοιχα.

ΛΥΣΗ

Προϊόν	% Κέρδους (X_i)	Έσοδα Πωλήσεων (W_i)	W_iX_i
$Δ_1$	4.2	30	126
$Δ_2$	5.5	20	110
$Δ_3$	7.4	5	37
$Δ_4$	10.1	3	30.3
ΣΥΝΟΛΑ	**27.2**	**58**	**303.3**

Θα χρησιμοποιήσω σταθμικό αριθμητικό μέσο: $\mu = \dfrac{\sum_{i=1}^{N} W_i X_i}{\sum_{i=1}^{N} W_i}$

Άρα $\mu = \dfrac{303.3}{58} \cong 5.229 \rightarrow \quad \mu \cong 5.2\%$

Σημείωση: Ο (απλός) αριθμητικός μέσος δίνει $\mu = \dfrac{\sum_{i=1}^{N} X_i}{N} = \dfrac{27.2}{4} \rightarrow \quad \mu = 6.8\%$.

ΑΣΚΗΣΗ 6

Τα ακόλουθα δεδομένα εκφράζουν τα βάρη (σε κιλά) 16 τυχαία επιλεγμένων φοιτητών από αυτούς που είναι μέλη του αθλητικού τμήματος της Σχολής τους. Να κατασκευαστεί το Διάγραμμα Μίσχου–Φύλλου των βαρών και να σχολιαστεί.

80	82	85	64	58	73	70	72
86	85	68	65	57	88	90	82

ΛΥΣΗ

Κατατάσσω τα δεδομένα κατά αύξουσα τάξη:

57, 58, 64, 65, 68, 70, 72, 73, 80, 82, 82, 85, 85, 86, 88, 90.

5	7, 8
6	4, 5, 8
7	0, 2, 3
8	0, 2, 2, 5, 5, 6, 8
9	0

Το Διάγραμμα Μίσχου – Φύλλου μας δείχνει ότι οι περισσότεροι από τους φοιτητές του Αθλητικού Τμήματος έχουν βάρος μεταξύ 80 και 90 κιλά.

ΑΣΚΗΣΗ 7

Ο Πίνακας 4 δίνει την κατανομή συχνότητας του αριθμού εργατικών ατυχημάτων ανά εργατοώρες σε ένα τυχαίο δείγμα 50 επιχειρήσεων ενός βιομηχανικού κλάδου.

Πίνακας 4

Ατυχήματα ανά 1000 εργατοώρες	**Αριθμός Επιχειρήσεων**
1.45 -< 1.75	3
1.75 -< 2.05	12
2.05 -< 2.35	14
2.35 -< 2.65	9
2.65 -< 2.95	6
2.95 -< 3.25	5

Επιπλέον γνωρίζουμε ότι για τα δεδομένα του Πίνακα 4 η διάμεσος του αριθμού των εργατικών ατυχημάτων ανά 1000 εργατοώρες είναι 2,26.

Με βάση τα παραπάνω:

i. Να υπολογισθεί η τεταρτημοριακή απόκλιση του αριθμού των ατυχημάτων.
ii. Να διερευνηθεί η κατανομή ως προς τη συμμετρία της και να σχολιασθεί.

ΛΥΣΗ

Ατυχήματα Ανά 1000 εργατώρες	**Αριθμός επιχειρήσεων f_i**	**F_i**
1.45- <1.75	3	3
1.75-< 2.05	12	15
2.05-<2.35	14	29
2.35-<2.65	9	38
2.65-<2.95	7	45
2.95-<3.25	5	50
ΣΥΝΟΛΑ	50	

i. Τεταρτημοριακή απόκλιση:

$$Q = \frac{Q_3 - Q_1}{2}$$

<u>Εντοπισμός θέσης Q_1</u>:

$$\frac{n*i}{4} = \frac{50*1}{4} = 12.5$$

Άρα το Q_1 ανήκει στην $2^η$ τάξη (διάστημα 1.75 - < 2.05)

<u>Υπολογισμός τιμής Q_1</u> :

$$Q_1 = L_{Q_1} + \frac{\delta}{70}\left(\frac{n_i}{4} - F_{Q_{1-1}}\right) = 1.75 + \frac{0.3}{12}(12.5 - 3) \to Q_1 = 1.98$$

<u>Εντοπισμός θέσης Q_3</u>:

$$\frac{n*i}{4} = \frac{3*50}{4} = 37.5$$

Άρα το Q_3 ανήκει στην $4^η$ τάξη (διάστημα 2.35 - < 2.65)

<u>Υπολογισμός τιμής Q_3</u> :

$$Q_3 = L_{Q_3} + \frac{\delta}{f_{Q_3}}\left(\frac{n*i}{4} - \Phi_{\Pi Q_3}\right) = 2.35 + \frac{0.3}{9}(37.5 - 29) \to Q_3 = 2.63$$

Άρα:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{2.63 - 1.98}{2} \to Q = 0.325$$

ii. Συμμετρία:

$$S_B = \frac{(Q_3 - M) - (M - Q_1)}{Q_3 - Q_1} = \frac{(2.63 - 2.26) - (2.26 - 1.98)}{2.63 - 1.98} \to S_B = \frac{0.09}{0.65} = 0.14$$

Όπως είναι γνωστό ο συντελεστής S_B παίρνει γενικά τιμές στο διάστημα [-1: έντονη αρνητική ασυμμετρία, +1: έντονη θετική ασυμμετρία]. Παρατηρούμε ότι υπάρχει ελαφρά θετική ασυμμετρία και συμπεραίνουμε ότι υπάρχουν οριακά περισσότερες επιχειρήσεις με μικρό αριθμό εργατικών ατυχημάτων.