ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΜΕΣΟΛΟΓΓΙΟΥ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ & ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Ο ΡΟΛΟΣ ΤΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΜΕΤΡΩΝ ΔΙΑΣΠΟΡΑΣ ΓΙΑ ΤΗΝ ΛΗΨΗ ΟΡΘΩΝ ΕΠΙΧΕΙΡΗΜΑΤΙΚΩΝ ΑΠΟΦΑΣΕΩΝ

ΚΑΡΡΑ ΓΕΩΡΓΙΑ

ΚΙΟΣΕ ΜΑΡΙΝΕΛΑ

ΕΙΣΗΓΗΤΡΙΑ
Κα. ΜΠΟΤΑ ΒΙΚΤΩΡΙΑ

ΜΕΣΟΛΟΓΓΙ 2011

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΜΕΣΟΛΟΓΓΙΟΥ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ & ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Ο ΡΟΛΟΣ ΤΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΜΕΤΡΩΝ ΔΙΑΣΠΟΡΑΣ ΓΙΑ ΤΗΝ ΛΗΨΗ ΟΡΘΩΝ ΕΠΙΧΕΙΡΗΜΑΤΙΚΩΝ ΑΠΟΦΑΣΕΩΝ

ΚΑΡΡΑ ΓΕΩΡΓΙΑ (Α.Μ. 12592)
ΚΙΟΣΕ ΜΑΡΙΝΕΛΑ (Α.Μ. 12354)

ΕΙΣΗΓΗΤΡΙΑ
Κα. ΜΠΟΤΑ ΒΙΚΤΩΡΙΑ

ΜΕΣΟΛΟΓΓΙ 2011

ΠΕΡΙΕΧΟΜΕΝΑ

1ο ΚΕΦΑΛΑΙΟ

1.1 ΕΙΣΑΓΩΓΗ

Η στατιστική είναι ένας κλάδος των εφαρμοσμένων μαθηματικών, ο οποίος βασίζεται σε ένα σύνολο αρχών και μεθοδολογιών για τον σχεδιασμό της διαδικασίας συλλογής δεδομένων (data), τη συνοπτική και αποτελεσματική παρουσίαση των δεδομένων και την ανάλυση και εξαγωγή συμπερασμάτων από τα δεδομένα. Πρόκειται για έναν κλάδο ο οποίος προσπαθεί από τις ιδιότητες του μέρους να εξάγει συμπεράσματα για το όλον.

Πιο συγκεκριμένα στην έρευνα αυτή προσπαθούμε να ερευνήσουμε την σημασία των μέτρων διασποράς στη λήψη επιχειρηματικών αποφάσεων. Εξετάζουμε ποια είναι τα μέτρα αυτά και το πώς χρησιμοποιούνται, εφαρμόζονται έτσι ώστε να παράσχουν την κατάλληλη πληροφόρηση στα άτομα που θα λάβουν σημαντικές επιχειρηματικές αποφάσεις πάνω σε ένα ζήτημα. Η ανάλυση που διεξάγεται εδώ ξεκινάει με την παράθεση των μέτρων θέσης, βλέπουμε ξεχωριστά ποια είναι τα βασικότερα από αυτά και έπειτα προχωράμε στα μέτρα διασποράς. Τα μέτρα διασποράς βασίζονται, δημιουργούνται μέσω των μέτρων θέσης και επομένως η αναφορά στα μέτρα θέσεως τα καθιστούσε αναγκαία στη παρούσα έρευνα. Έπειτα αναφερόμαστε στη σημασία των επιχειρηματικών αποφάσεων και εν συνεχεία παραθέτουμε εφαρμογές των μέτρων διασποράς σε διάφορα προβλήματα.

1.2 ΣΤΟΧΟΣ ΤΗΣ ΕΡΕΥΝΑΣ

Ο στόχος της έρευνας όπως ήδη έχουμε αναφέρει είναι να αναδείξει την σημασία των μέτρων διασποράς στη λήψη σωστών επιχειρηματικών αποφάσεων, δηλαδή στη συγκέντρωση της κατάλληλης, ορθής πληροφόρησης στο άτομο ή στα άτομα που είναι υπεύθυνα για τη λήψη μιας επιχειρηματικής απόφασης.

2^{o} ΚΕΦΑΛΑΙΟ: ΜΕΤΡΑ ΘΕΣΕΩΣ

2.1 Αριθμητικός Μέσος

Έστω οι παρατηρήσεις ενός τυχαίου δείγματος από έναν πληθυσμό. Τότε ο αριθμητικός μέσος, των παρατηρήσεων αυτών δίνεται από τη σχέση

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_\nu}{\nu} = \frac{1}{\nu} \cdot \sum_{i=1}^{\nu} x_i$$

Παράδειγμα 1

Παρακάτω δίνονται οι τελικές βαθμολογίες (με άριστα το 100) ενός τυχαίου δείγματος 15 φοιτητών της αρχαιολογικής σχολής σε ένα μάθημα τους κατά το ακαδ. έτος 2002-03.

56, 67, 34, 89, 67, 89, 78, 57, 48, 47, 89, 80, 59, 89, 94

Ο αριθμητικός μέσος της βαθμολογίας του τυχαίου δείγματος των 15 φοιτητών του ΕΑΠ είναι:

$$\bar{x} = \frac{1}{\nu} \cdot \sum_{i=1}^{\nu} x_i = \frac{1043}{15} = 69{,}53$$

Παράδειγμα 2

Παρακάτω δίνονται τα ημερομίσθια για 80 εργαζομένους για το έτος 2007.

Ημερομίσθια *(σε δεκάδες ευρώ)*	*Αριθμός Εργαζομένων* v_i	**Κεντρική τιμή** **m_i**
[5-10)	20	**7,5**
[10-15)	30	**12,5**
[15-20)	17	**17,5**
[20-25)	13	**22,5**
Σύνολο	**80**	

Μέσο ημερομίσθιο

$$\bar{X} = \frac{\sum_{i=1}^{4} v_i m_i}{\sum_{i=1}^{4} v_i} = \frac{1115}{80} = 13,94$$

Άρα το μέσο ημερομίσθιο είναι 13,94 δεκάδες ευρώ

Επισήμανση

Εύκολα αποδεικνύεται ότι εάν στις τιμές της μεταβλητής X προσθέσουμε έναν αριθμό α τότε και η μέση τιμή μεταβάλλεται κατά α.

Δηλαδή $\overline{X+a} = \bar{X} + a$ με $a \in \Re$

Επίσης εάν οι τιμές της X πολλαπλασιασθούν επί $k \in \Re$ τότε και η μέση τιμή πολλαπλασιάζεται επί κ.

Δηλαδή $\overline{\kappa \cdot X} = \kappa \cdot \overline{X}$

2.2 Διάμεσος

Η **διάμεσος, *M*,** είναι η τιμή που χωρίζει ένα σύνολο δεδομένων περίπου στη μέση εφόσον τα δεδομένα τοποθετηθούν σε αύξουσα σειρά. Ουσιαστικά η διάμεσος ενός συνόλου παρατηρήσεων είναι η τιμή που έχει την ιδιότητα ότι το πολύ 50% των μετρήσεων είναι μικρότερες από την τιμή αυτή και το πολύ το 50% των μετρήσεων είναι μεγαλύτερες από την τιμή αυτή.

Στην περίπτωση του παραδείγματος των φοιτητών της αρχαιολογίας η διάμεσος βρίσκεται αφού τοποθετήσουμε τις βαθμολογίες σε αύξουσα σειρά 34, 47, 48, 56, 57, 59, 67, 67, 78, 80, 89, 89, 89, 89, 94 Το πλήθος των παρατηρήσεων είναι 15, δηλαδή περιττός αριθμός, άρα η διάμεσος τιμή των βαθμολογιών θα είναι η τιμή της παρατήρησης που βρίσκεται στη θέση $\frac{n+1}{2}$ δηλαδή στη συγκεκριμένη περίπτωση η διάμεσος θα είναι η τιμή της παρατήρησης που βρίσκεται στη θέση $\frac{15+1}{2}=8$. Αρά η διάμεσος βαθμολογία είναι το 67. Στην περίπτωση που θα είχαμε 14 παρατηρήσεις η διάμεσος τιμή θα ήταν ο μέσος της 7ης και της 8ης παρατήρησης. Σε αυτήν την περίπτωση δηλαδή, η διάμεσος δίνεται από τον τύπο

$$M=\frac{1}{2}(X_{\frac{n}{2}+1}+X_{\frac{n}{2}})$$

Ο παραπάνω τύπος είναι για μη ομαδοποιημένα δεδομένα. Για ομαδοποιημένα ο τύπος είναι

$$M=X_{\lambda}+\frac{\delta}{v_i}(\frac{N}{2}-\Phi_i)$$

Όπου

X_{λ} : το κάτω μέρος της κλάσης στην οποία πέφτει η διάμεσος

δ : το πλάτος της κλάσης

ν_i : το πλήθος της κλάσης που πέφτει η διάμεσος

Ν: το πλήθος του δείγματος

Φ_i : η αθροιστική συχνότητα μέχρι την κλάση στην οποία πέφτει η διάμεσος, δεν συμπεριλαμβάνει την κλάση αυτή

Συγκρίνοντας τον αριθμητικό μέσο και την διάμεσο παρατηρούμε τα εξής:

Ο αριθμητικός μέσος:

- επηρεάζεται από την ύπαρξη ακραίων τιμών
- είναι χρήσιμος για συμπερασματολογία που αναφέρεται στο άθροισμα των τιμών του πληθυσμού
- είναι ευκολότερο να εργασθούμε με αυτόν θεωρητικά

Η διάμεσος

- Δεν επηρεάζεται από την ύπαρξη ακραίων τιμών
- Δεν είναι χρήσιμη για συμπερασματολογία που αναφέρεται στο άθροισμα των τιμών του πληθυσμού
- Είναι δύσκολο να εργασθούμε με αυτήν θεωρητικά

2.3 Επικρατούσα Τιμή

Ως **επικρατούσα τιμή,** ενός συνόλου δεδομένων χαρακτηρίζουμε εκείνη με τη μεγαλύτερη συχνότητα εμφάνισης.

Στην περίπτωση του παραδείγματος με τους φοιτητές η επικρατούσα τιμή της βαθμολογίας είναι το 89 που εμφανίζεται 4 φορές.

Στην περίπτωση που υπάρχουν δύο ή περισσότερες τιμές με την ίδια συχνότητα εμφάνισης τότε λέμε ότι τα δεδομένα έχουν δύο ή περισσότερες επικρατούσες τιμές.

Η επικρατούσα τιμή για τα ομαδοποιημένα δεδομένα δίνεται από τον τύπο :

$$M_o = X_\lambda + \delta \frac{\Delta_1}{\Delta_1 + \Delta_2}$$

Όπου:

$_T X_\lambda$ **: το** αριστερό άκρο της επικρατούσας κλάσης, δηλαδή αυτής με τη μεγαλύτερη συχνότητα

δ: Το πλάτος της κλάσης

Δ1 : η διαφορά συχνότητα της επικρατούσας κλάσης – τη συχνότητα της προηγούμενης κλάσης

Δ2: η διαφορά συχνότητα της επικρατούσας κλάσης – τη συχνότητα της επόμενης κλάσης

2.4 Τεταρτημόρια

Σε πολλές περιπτώσεις τα μέτρα κεντρικής τάσης των δεδομένων δεν εξασφαλίζουν ικανοποιητική περιγραφή τους. Στις περιπτώσεις αυτές προσδιορίζουμε κάποια άλλα μέτρα που αναφέρονται στη σχετική θέση των δεδομένων. Αυτά είναι τα λεγόμενα ποσοστιαία σημεία. Τα πιο συχνά χρησιμοποιούμενα ποσοστιαία σημεία είναι τα **τεταρτημόρια (Q1, Q2, Q3)** τα οποία χωρίζουν ένα σύνολο παρατηρήσεων σε τέταρτα. Για παράδειγμα το *Q1* είναι η τιμή η οποία χωρίζει το σύνολο των παρατηρήσεων σε δύο μέρη έτσι ώστε το πολύ 25% να είναι μικρότερες και το πολύ 75% να είναι μεγαλύτερες από την τιμή αυτή.

Όπως στην περίπτωση της διαμέσου, ο προσδιορισμός των τεταρτημορίων προϋποθέτει διάταξη των παρατηρήσεων κατά αύξουσα σειρά και στη συνέχεια εντοπισμό της θέσης τους. Ειδικότερα:

Το i=1, 2 ,...,3 τεταρτητομόριο *(Qi)* βρίσκεται στην $\left[\frac{i(n+1)}{2}\right]$ θέση. -Η τιμή του i=1, 2 ,3 τεταρτημορίου (Qi) είναι: $Q_i = X_{A_Q} + \Delta_Q[X_{A_Q+1} - X_{A_Q}]$

όπου A_Q = ακέραιο μέρος του πηλίκου $\left[\frac{i(n+1)}{4}\right]$.

και Δ_Q = δεκαδικό μέρος του πηλίκου $\left[\frac{i(n+1)}{4}\right]$

Για τα δεδομένα του παραδείγματος της βαθμολογίας των φοιτητών, τα οποία έχουν τοποθετηθεί κατά αύξουσα σειρά μεγέθους έχουμε:

Το 1ο τεταρτημόριο είναι η τιμή που βρίσκεται στη

$$\left[\frac{i(n+1)}{4}\right] = \frac{16}{4} = 4$$

.

Στη συγκεκριμένη περίπτωση $A_Q = 4$ και $\Delta_Q = 0$

$Q_1 = X_4 = 56$

Τα παραπάνω αφορούν αταξινόμητα δεδομένα. Ο τύπος για τα ταξινομημένα δεδομένα είναι

$$Q_K = X_\lambda + \frac{\delta}{v_i}(\frac{\kappa N}{4} - \Phi_i)$$

Όπου

Κ: το πρώτο ή το τρίτο τεταρτημόριο, λαμβάνει τιμές 1 και 3

X_λ : το κάτω μέρος της κλάσης στην οποία πέφτει το τεταρτημόριο

δ : το πλάτος της κλάσης

v_i : το πλήθος της κλάσης που πέφτει το τεταρτημόριο

Ν: το πλήθος του δείγματος

Φ_i : η αθροιστική συχνότητα μέχρι την κλάση στην οποία πέφτει το τεταρτημόριο, δεν συμπεριλαμβάνει την κλάση αυτή

Παράδειγμα

Ημερομίσθια (σε δεκάδες ευρώ)	***Αριθμός Εργαζομένων v_i***	**Κεντρική τιμή x_i**	**Αθροιστική συχνότητα Φ_i**
[5-10)	20	**7,5**	**20**
[10-15)	30	**12,5**	**50**
[15-20)	17	**17,5**	**67**
[20-25)	13	**22,5**	**80**

Υπολογίζουμε την τιμή του τρίτου τεταρτημορίου, άρα θα πρέπει να εντοπίσουμε τη θέση του

$$\frac{3n}{4}=\frac{3*80}{4}=60$$

Επομένως το Q_3 ανήκει στη τρίτη τάξη, δηλαδή στο διάστημα [10, 15). Υπολογίζουμε το Q_3 πλέον:

$$Q_3=X_\lambda+\frac{\delta}{v_i}(\frac{3N}{4}-\Phi_i)=15+5\frac{(60-50)}{17}=15+5*0.588=17.94$$

Υπολογίζουμε την τιμή του πρώτου τεταρτημορίου, άρα θα πρέπει να εντοπίσουμε τη θέση του

$$\frac{1n}{4}=\frac{1*80}{4}=20$$

Επομένως το Q_1 ανήκει στη τρίτη τάξη, δηλαδή στο διάστημα [5, 10). Υπολογίζουμε το Q_1 πλέον:

$$Q_1=X_\lambda+\frac{\delta}{v_i}(\frac{1N}{4}-\Phi_i)=5+5\frac{(20-0)}{20}=10$$

Τέλος προσθέτουμε τους τύπους για τα εκατοστημόρια και τα δεκατημόρια. Αυτοί είναι αντίστοιχα οι εξής:

Δεκατημόρια

$$D_K=X_\lambda+\frac{\delta}{v_i}(\frac{\kappa N}{10}-\Phi_i)$$

Όπου

K: 1,2,…,9

X_λ : το κάτω μέρος της κλάσης στην οποία πέφτει το δεκατημόριο

δ : το πλάτος της κλάσης

v_i : το πλήθος της κλάσης που πέφτει το δεκατημόριο

Ν: το πλήθος του δείγματος

Φ_i : η αθροιστική συχνότητα μέχρι την κλάση στην οποία πέφτει το δεκατημόριο δεν συμπεριλαμβάνει την κλάση αυτή

Εκατοστημόρια

$$C_K = X_\lambda + \frac{\delta}{v_i}(\frac{\kappa N}{100} - \Phi_i)$$

Όπου

Κ: 1,2,3, …,99, 100

X_λ : το κάτω μέρος της κλάσης στην οποία πέφτει το εκατοστημόριο

δ : το πλάτος της κλάσης

v_i : το πλήθος της κλάσης που πέφτει το εκατοστημόριο

Ν: το πλήθος του δείγματος

Φ_i : η αθροιστική συχνότητα μέχρι την κλάση στην οποία πέφτει το εκατοστημόριο δεν συμπεριλαμβάνει την κλάση αυτή

3^{O} ΚΕΦΑΛΑΙΟ: ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ

3.1 Εύρος

Το εύρος *(R)* ενός συνόλου παρατηρήσεων ορίζεται ως η διαφορά μεταξύ της μεγίστης και της ελαχίστης τιμής του συνόλου των παρατηρήσεων. $R = Xmax - Xmin$

Για τα δεδομένα του παραδείγματος της βαθμολογίας των 15 φοιτητών έχουμε ότι: $R = Xmax - Xmin = 94 - 34 = 60$

3.2 Ενδοτεταρτημοριακό Εύρος

Το ενδοτεταρτημοριακό εύρος *(IR)*, το οποίο ορίζεται ως, $IR = Q3 - Q1$ και περιλαμβάνει το 50% των παρατηρήσεων που βρίσκονται γύρω από τη διάμεσο. Για τα δεδομένα του παραδείγματος της βαθμολογίας των 15 φοιτητών έχουμε ότι:

$$IR = Q3 - Q1 = 89 - 56 = 33$$

Στη περίπτωση των ημερομισθίων το **ενδοτεταρτημοριακό εύρος** *(IR) είναι*

$$IR = Q3 - Q1 = 17{,}94 - 10 = 7{,}94$$

3.3 Διακύμανση

Η **Διακύμανση** (*S*) είναι ένα ακόμη μέτρο διασποράς το οποίο βασίζεται στην έννοια της απόστασης μιας παρατήρησης από το μέσο αριθμητικό των παρατηρήσεων. Ξεπερνά το πρόβλημα του μηδενικού αθροίσματος αποκλίσεων χρησιμοποιώντας όχι απόλυτες τιμές αλλά τα τετράγωνα των αποκλίσεων τα οποία έχουν πάντοτε μη αρνητικές τιμές. Στην περίπτωση δείγματος *n* παρατηρήσεων με δειγματικό μέσο $\bar{X}$ η δειγματική διακύμανση ορίζεται ως η μέση τιμή των τετραγώνων των αποκλίσεων των *n* τιμών του δείγματος από το δειγματικό μέσο, $s^2 = \frac{1}{\nu - 1}\sum_{i=1}^{\nu}(x_i - \bar{x})^2$

ο τύπος αυτός μπορεί να μετασχηματισθεί κάνοντας τις πράξεις και ως εξής:

$$s^2=\frac{1}{\nu-1}\left[\sum_{i=1}^{\nu}x_i^2-\frac{1}{\nu-1}\cdot(\sum_{i=1}^{\nu}x_i)^2\right]$$

$$\text{ή}\ \ s^2=\frac{1}{\nu-1}\sum_{i=1}^{\nu}x_i^2-\bar{x}^2$$

εάν οι παρατηρήσεις έχουν συχνότητες ν_i οι τύποι γράφονται ως εξής:

$$s^2=\frac{1}{\nu-1}\sum_{i=1}^{\mu}(x_i-\bar{x})^2\cdot\nu_i$$

$$s^2=\frac{1}{\nu-1}\left[\sum_{i=1}^{\nu}x_i^2\cdot\nu_i-\frac{1}{\nu-1}\cdot(\sum_{i=1}^{\nu}x_i\cdot\nu_i)^2\right]$$

$$s^2=\frac{1}{\nu-1}\sum_{i=1}^{\nu}x_i^2\cdot\nu_i-\bar{x}^2$$

Για τα δεδομένα του παραδείγματος της βαθμολογίας των 15 φοιτητών έχουμε ότι:

$$s^2=\frac{1}{\nu-1}\sum_{i=1}^{\nu}(x_i-\bar{x})^2=\frac{4993,7}{14}=356,69$$

Παράδειγμα

Παραθέτουμε παρακάτω την κατανομή των ημερομισθίων για 100 εργάτες και θα βρούμε την διασπορά των ημερομισθίων και με τους δύο τύπους:

Τάξεις ημερομισθίων σε δραχμές	X_i	v_i	$v_i X_i$	$v_i X_i^2$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$v_i(x_i - \bar{x})^2$
1500-2500	2000	5	10000	20000000	-2770	7672900	38364500
2500-3500	3000	13	39000	117000000	-1770	3132900	40727700
3500-4500	4000	20	80000	320000000	-770	592900	11858000
4500-5500	5000	35	175000	875000000	230	52900	1851500
5500-6500	6000	18	108000	648000000	1230	1512900	27232200
6500-7500	7000	7	49000	343000000	2230	4972900	34810300
7500-8500	8000	2	16000	128000000	3230	10432900	20865800
Σύνολο		100	477000	2451000000			175710000

$$N = 100, \sum v_i x_i = 477000, \bar{x} = 4770, \sum v_i x_i^2 = 2451000000, \sum v_i (x_i - \bar{x}) = 1757100$$

Αντικαθιστώντας στο τύπο

$$s^2 = \frac{1}{\nu - 1} \sum_{i=1}^{\mu} (x_i - \bar{x})^2 \cdot \nu_i = \frac{175710000}{99} = 1774848,48 => s = 1332,23$$

Αντικαθιστώντας και στο τύπο

$$s^2 = \frac{1}{\nu - 1} \sum_{i=1}^{\nu} x_i^2 \cdot \nu_i - \bar{x}^2 = \frac{2451000000}{99} - (\frac{477000}{99})^2 = 24757575,75 - 23214876,03 = 1542699,72$$
$$=> s = 1242,05$$

Έχουμε μια απόκλιση στους υπολογισμούς που οφείλεται σε στρογγυλοποιήσεις που γίνονται από το ηλεκτρονικό υπολογιστή. Η τυπική απόκλιση 1332,23 σημαίνει ότι το ημερομίσθιο κάθε εργάτη διαφέρει από μέσο όρο (4770) κατά μέσο όρο κατά 1332,23 δραχμές.

3.4. Τυπική Απόκλιση

Η διακύμανση είναι αναμφισβήτητα ένα πολύ χρήσιμο μέτρο διασποράς, Παρολαυτά, είναι δύσκολο να ερμηνευτεί δεδομένου ότι εκφράζεται στα τετράγωνα των μονάδων των παρατηρήσεων.

Η **τυπική απόκλιση** (S) ορίζεται ως η θετική τετραγωνική ρίζα της διακύμανσης, δηλαδή, $S = +\sqrt{S^2}$

Για τα δεδομένα του παραδείγματός $S = +\sqrt{S^2} = \sqrt{356,69} = 19,89$

3.5 Συντελεστής Μεταβλητότητας

Έστω ότι αντικείμενο έρευνας είναι η μεταβλητότητα που εμφανίζει η επίδοση των φοιτητών σε δύο μαθήματα: τα Μαθηματικά και τη Στατιστική. Ας θεωρήσουμε το παράδειγμα της βαθμολογίας των 100 φοιτητών και το οποίο έδωσε τα εξής αποτελέσματα:

Μαθηματικά: $50=X_M$ και $S_M = 4$

Στατιστική: $70=X_\Sigma$ και $S_\Sigma = 5$

Με βάση τις τιμές των $S_M = 4$ και $S_\Sigma = 5$ η διασπορά στην επίδοση των φοιτητών στη Στατιστική είναι μεγαλύτερη σε σύγκριση με εκείνη των Μαθηματικών. Το συμπέρασμα όμως αυτό είναι λανθασμένο δεδομένου ότι τα δύο σύνολα παρατηρήσεων έχουν διαφορετικούς μέσους και άρα δεν επιτρέπεται η σύγκριση της διασποράς με την τυπική απόκλιση. Το πρόβλημα της σύγκρισης που ανακύπτει στην περίπτωση αυτή αντιμετωπίζεται με τη βοήθεια του **συντελεστή μεταβλητότητας**, ο οποίος μετράει τη διασπορά των τιμών της κατανομής σε **σχετικούς όρους** και υπολογίζεται από τον τύπο:

$$CV = \frac{s}{\bar{x}}$$

Για την επίδοση των 100 φοιτητών στα Μαθηματικά και τη Στατιστική, σύμφωνα με τα δεδομένα του παραδείγματος, έχουμε:

Μαθηματικά: $CV_{\mathrm{M}} = \frac{s_{\mathrm{M}}}{\bar{x}_{\mathrm{M}}} = \frac{4}{50} = 0,08$

Στατιστική: $CV_{\Sigma} = \frac{s_{\Sigma}}{\bar{x}_{\Sigma}} = \frac{5}{70} = 0,07$

Συμπερασματικά μπορούμε να πούμε ότι η σχετική ανομοιογένεια (μεταβλητότητα) στην επίδοση των φοιτητών στη Στατιστική (7%) είναι μικρότερη από εκείνη στα Μαθηματικά (8%).

3.6 Μέση διαφορά Gini

Στη περίπτωση που ενδιαφερόμαστε για τις αποκλίσεις, διαφορές ανάμεσα στις τιμές μιας μεταβλητής X και όχι για τις αποκλίσεις από το μέσο αριθμητικό μέσο χρησιμοποιούμε την μέση διαφορά κατά Gini. Για παράδειγμα ένας υπάλληλος μπορεί να ενδιαφέρεται να μάθει κατά πόσοι διαφέρει ο μισθός του από τους μισθούς των άλλων υπαλλήλων και όχι από τον μέσο μισθό. Η μέση διαφορά που θα δούμε παρακάτω συμβολίζεται με d_i.

Παράδειγμα (αταξινόμητα δεδομένα)

Υποθέτουμε ότι τα δεδομένα 4, 5, 7, 8, 10 και 12 αφορούν εβδομαδιαίους μισθούς έξι υπαλλήλων. Θα υπολογίσουμε την μέση διαφορά. Για τον σκοπό υατό όλες τις διαφορές για όλους τους συνδυασμούς.

$|4-4|=0, |4-5|=1, |4-7|=3, \ldots, |4-12|=8 \Rightarrow 22$

$|5-4|=1, |5-5|=0, |5-7|=2, \ldots, |5-12|=7 \Rightarrow 18$

....

$|12-4|=8, |12-5|=7, |12-7|=5, \ldots, |12-12|=0 \Rightarrow 26$

Οι αριθμοί στο τέλος των εξισώσεων αφορούν τα αθροίσματα. Το άθροισμα όλων των διαφορών είναι ίσο με 112. Έχουμε επομένως:

$$d_I = \frac{\sum |x_i - x_j|}{n^2} = \frac{112}{6^2} = 3{,}111$$

Όπου $\sum |x_i - x_j|$ το συνολικό άθροισμα των διαφορών και n το πλήθος των τιμών στο δείγμα. Η τιμή που βρήκαμε παραπάνω σημαίνει ότι ο μισθό κάθε υπαλλήλου διαφέρει κατά μέσο από τους μισθούς των άλλων υπαλλήλων κατά 3.111 δραχμές.

Συνεχίζουμε με το να δούμε πως υπολογίζεται η μέση διαφορά σε ταξινομημένα δεδομένα. Σε αυτή την περίπτωση υπάρχουν δύο περιπτώσεις:

- Τα διαστήματα των τάξεων είναι άνισα
- Τα διαστήματα των τάξεων είναι ίσα

Στη πρώτη περίπτωση η μέση διαφορά υπολογίζεται με τον τύπο:

$$d_I = \frac{2}{N^2} \sum (N - \Phi_i) \cdot \Phi_i \cdot \delta_i$$

N: σύνολο συχνοτήτων της κατανομής

Φ_i: δεξιόστροφη αθροιστική σειρά συχνοτήτων

δ_i: διαστήματα τάξεων

Παράδειγμα

Ας υποθέσουμε ότι έχουμε την κατανομή 120 υπαλλήλων ανάλογα με το ύψος του μισθού τους. Θα υπολογίσουμε την μέση διαφορά κατά Gini.

Τάξεις μισθών	Αριθμός υπαλλήλων ν_i	Φ_i	N- Φ_i	(N- Φ_i) Φ_i	δ_i	(N- Φ_i) Φ_i δ_i
6-8	48	48	72	3456	2	6912
8-10	38	86	34	2924	2	5848
10-15	21	107	13	1391	5	6955
15-25	9	116	4	464	10	4640
25-50	4	120	0	0	25	0
Σύνολο	120					24355

N=120, (N- Φ_i) Φ_i δ_i =24355

Αντικαθιστούμε στο τύπο και έχουμε

$$d_I = \frac{2}{N^2}\sum(N-\Phi_i)\cdot\Phi_i\cdot\delta_i = \frac{2}{120^2}24355 = 3{,}383 \text{ χιλιάδες δραχμές}$$

Η παραπάνω τιμή σημαίνει ότι ο μισθός κάθε υπαλλήλου διαφέρει κατά μέσο όρο από τους μισθούς των άλλων υπαλλήλων κατά 3111 δραχμές.

Στη δεύτερη περίπτωση η μέση διαφορά υπολογίζεται με τον τύπο:

$$d_I = \frac{2\delta}{N^2}\sum(N-\Phi_i)\cdot\Phi_i$$

N: σύνολο συχνοτήτων της κατανομής

Φ_i: δεξιόστροφη αθροιστική σειρά συχνοτήτων

δ: το διάστημα των τάξεων

Προφανώς η πρώτη περίπτωση περιλαμβάνει και την δεύτερη καθώς είναι πιο γενική. Έχουμε κοινό πλάτος πλέον για τα διαστήματα. Οι υπολογισμοί είναι παρόμοιοι.

4^{O} ΚΕΦΑΛΑΙΟ: Επιχειρηματικές Αποφάσεις

Το ζητούμενο του θέματος δίνεται συναρτήσει της ρήσης του Herbert Simon, ότι η λήψη αποφάσεων είναι συνώνυμη του μάνατζμεντ. Αυτή η φράση του ουσιαστικά κλίνει μέσα της όλη την έννοια της απόφασης στη διοίκηση των οργανισμών. Σύμφωνα με τον ίδιο και προς κατανόηση του άνωθεν ορισμού, αναφέρει ότι η λήψη αποφάσεων συνίσταται στην εξεύρεση ευκαιριών για μια λήψη απόφασης, στην ανάπτυξη εναλλακτικών προσεγγίσεων και τέλος στην εξεύρεση πιθανών τρόπων δράσης (Χριστοδουλόπουλου Δ.,2006).

Το Μάνατζμεντ στο σύνολο του αναφέρεται στις δράσεις ενός οργανισμού, οι οποίες δράσεις έχουν να κάνουν με αποφάσεις, με προσεγγίσεις αυτών των αποφάσεων, με επιλογή των καλύτερων δυνατών δράσεων για τον οργανισμό. Με βάση αυτό ο νομπελίστας θεωρητικός συνδέει τις αποφάσεις με το μάνατζμεντ, μάλλον ταυτίζει τις δυο έννοιες (Facione, P. and Facione, N., 2007),

Η σύνδεση της λήψης αποφάσεων με το μάνατζμεντ, αναφέρεται και στη διαδικασία λήψης ανταγωνιστικών αποφάσεων. Συγκεκριμένα το μάνατζμεντ, βοηθά στην ανταγωνιστικότητα ενός οργανισμού, βάση του μάνατζμεντ, ένας οργανισμός μπορεί να εστιάσει στη στρατηγική λήψη αποφάσεων λαμβάνοντας υπ'όψιν τις ενδεχόμενες αποφάσεις των ανταγωνιστών (Plous, S.,2003).

Ο ρόλος της διοίκησης αποφάσεων είναι, μέσα σε μια επιχείρηση, να προγραμματίζει τις δραστηριότητες και τις ευθύνες των χρηματοοικονομικών διευθυντών. Ουσιαστικά η χρηματοοικονομική λειτουργία της επιχείρησης επικεντρώνει το ενδιαφέρον της σε τρεις βασικές κατηγορίες αποφάσεων: α) την απόφαση για το επενδυτικό έργο της επιχείρησης, β) την απόφαση για τον τρόπο με τον οποίο θα χρηματοδοτηθούν οι επιχειρηματικές δραστηριότητες, γ) την απόφαση για το μέρισμα που η εταιρία θα διανείμει.

Η σύγχρονη θεωρία της διοίκησης αποφάσεων λειτουργεί με βάση την υπόθεση ότι ο αντικειμενικός σκοπός κάθε επιχείρησης είναι η μεγιστοποίηση του πλούτου των μετόχων της. Για να επιτευχθεί η μεγιστοποίηση του πλούτου των μετόχων θα πρέπει να μεγιστοποιηθεί η αξία του μεριδίου τους στην επιχείρηση, δηλαδή να μεγιστοποιηθεί η αξία των μετοχών που αυτοί κατέχουν.

Η αξία της μετοχής βέβαια αντικατοπτρίζεται κάθε στιγμή από την αγοραία τιμή της, η οποία με την σειρά της αντανακλά τις επενδυτικές, χρηματοδοτικές και μερισματικές επιλογές που κάνει η επιχείρηση. Οι επιλογές αυτές καθορίζουν το μέγεθος και τον βαθμό κινδύνου των μελλοντικών ταμειακών ροών και η παρούσα αξία αυτών των ταμειακών ροών προσδιορίζει την συνολική αξία της επιχείρησης. Κατά συνέπεια, κάθε επιχειρηματική απόφαση πρέπει να αξιολογείται με βάση την επίδραση που αυτή θα έχει στις μελλοντικές ταμειακές ροές και γενικότερα στην συνολική αξία της επιχείρησης, η οποία μεταφράζεται σε αξία για τους μετόχους.

Στην πράξη όμως αυτό που συμβαίνει είναι, με βάση τις παραδοσιακές τεχνικές και τα στοιχεία που βρίσκονται δημοσιευμένα στις οικονομικές καταστάσεις, αποτελούν την βάση για τις περισσότερες επιχειρηματικές αποφάσεις. Με βάση τα δημοσιευμένα κέρδη οι επιχειρήσεις αποφασίζουν το ποσοστό που θα διαθέσουν για μερίσματα, οι πιστωτές αποφασίζουν αν θα συνεχίσουν να παρέχουν πίστωση ή θα την διακόψουν, οι επενδυτές αποφασίζουν να αγοράσουν ή να πωλήσουν τις μετοχές που κατέχουν και οι μέτοχοι αξιολογούν το ΔΣ της επιχείρησης, με την στήριξή τους ή την εκλογή νέων αντιπροσώπων.

Η διοίκηση αποφάσεων είναι ένα όργανο της διοίκησης, είναι η διαδικασία μετρήσεως, αναλύσεως, υπολογισμού και παρουσιάσεως του κόστους των προϊόντων ή των υπηρεσιών μιας επιχείρησης, καθώς και της αποδοτικότητας και της καλής λειτουργίας της. η λογιστική ασχολείται με τις παρακάτω δραστηριότητες:

1. την μέτρηση ή τον υπολογισμό και την εκτίμηση του κόστους των παραχθέντων προϊόντων ή υπηρεσιών. Σήμερα το κόστος προϋπολογίζεται με βάση τις τεχνικές του προδιαγραφές.

2.την ανάλυση του κόστους και τον προσδιορισμό των σχέσεων μεταξύ του κόστους και των διαφόρων παραγόντων που επιδρούν σ' αυτό. Η έννοια του κόστους είναι έννοια συσσωρευτική.

3.την καταχώρηση, την ταξινόμηση και την κατανομή του κόστους στα διάφορα κέντρα κόστους

4.την παρουσίαση του κόστους περιληπτικά ή λεπτομερειακά στην διοίκηση για την λήψη αποφάσεων

Γενικότερα ασχολείται και με τα παρακάτω:

1.σχεδιάζει την λειτουργία του συστήματος και των διαδικασιών της κοστολόγησης, ώστε να ικανοποιούνται οι ανάγκες της συγκεκριμένης επιχείρησης

2.ελέγχει το κόστος, δηλαδή ερευνά το κατά πόσο το πραγματικό κόστος ανταποκρίνεται στον στόχο που είχε τεθεί από την αρχή

3.αναλύει το κόστος. Αυτό μπορεί να γίνει είτε κατά προϊόν, είτε κατά περιοχές πωλήσεων, είτε κατά πελάτη, είτε κατά τμήμα της επιχείρησης.

4.σύγκριση του κόστους με το κόστος των εναλλακτικών προϊόντων, δραστηριοτήτων, μεθόδων

Βασικός σκοπός της διοίκησης αποφάσεων είναι:

- ο καθορισμός αναλυτικών αποτελεσμάτων και η αποτίμηση των αποθεμάτων
- η υποβοήθηση της διοικήσεως μιας επιχείρησης στον τομέα του προγραμματισμού
- ο έλεγχος της αποτελεσματικότητας της επιχειρήσεως

Βασικά η απόφαση αποβλέπει μαζί με τον σχεδιασμό και τον απολογιστικό έλεγχο, στην σωστή τιμολογιακή πολιτική.

Με τον όρο σχεδιασμό εννοούμε την μακροχρόνια πρόβλεψη και παράθεση στόχων εφικτών και τον καθορισμό της πολιτικής και στρατηγικής της επιτεύξεως των στόχων αυτών. Επιπλέον την λήψη όλων των απαραίτητων αποφάσεων, με την επιλογή της καλύτερης εναλλακτικής λύσης, που θα συντελέσουν στην επίτευξη των στόχων που τέθηκαν. Μακροχρόνιοι στόχοι των επιχειρήσεων, είναι κατά κανόνα, οι παρακάτω:

i. η μεγιστοποίηση των αναμενόμενων κερδών
ii. η μεγιστοποίηση των πωλήσεων και του μεριδίου της αγοράς
iii. ελαχιστοποίηση των εξόδων γενικά
iv. ελαχιστοποίηση του κατά μονάδα κόστους παραγωγής
v. μεγιστοποίηση των περιουσιακών στοιχείων
vi. μεγιστοποίηση της διανομής των κερδών και των μερισμάτων
vii. ελαχιστοποίηση του χρόνου παραγωγής, απωλειών χρήματος και διαφόρων άλλων απωλειών

viii.διάφοροι άλλοι εξειδικευμένοι στόχοι

Σκοπός του προσδιορισμού των αποφάσεων είναι η επιτυχία ενός καλού σχεδιασμού και μακροχρόνιου προγραμματισμού των δραστηριοτήτων της επιχειρήσεως. Ο σχεδιασμός διευκολύνεται με την θέσπιση αντικειμενικών σκοπών και με τον καθορισμό ειδικής στρατηγικής και πολιτικής της επιχείρησης.

Ο απολογιστικός έλεγχος διευκολύνεται με την παροχή των στοιχείων της λογιστικής κόστους και αποβλέπει στην ορθή εφαρμογή και επίτευξη των σχεδιασμένων στόχων, στην μέτρηση του αποτελέσματος και στην διαπίστωση τυχόν αποκλίσεων. Σκοπός της λογιστικής κόστους είναι και ο εσωτερικός έλεγχος. Έτσι αποβλέπει στον καταλογισμό των ευθυνών για τις αποκλίσεις και την λήψη διορθωτικών μέτρων για το μέλλον ή πολλές και στην αλλαγή στόχων, με την αυτόματη επαναπληροφόρηση. Η λογιστική κόστους διευκολύνει τον απολογιστικό έλεγχο, κάνοντας σύγκριση των αποτελεσμάτων που επιτεύχθηκαν με τα αναμενόμενα αποτελέσματα.

5^{O} ΚΕΦΑΛΑΙΟ: ΕΦΑΡΜΟΓΕΣ

Στο παρόν κεφάλαιο παρουσιάζουμε διάφορα προβλήματα που σχετίζονται με την λήψη επιχειρηματικών αποφάσεων. Τα περισσότερα από αυτά περιλαμβάνουν τα μέτρα διασποράς ως εργαλεία λήψης αποφάσεων. Μερικά προβλήματα χρησιμοποιούν τα μέτρα θέσης για την λήψη αποφάσεων. Ας δούμε όμως τα προβλήματα όπως παρουσιάζονται παρακάτω.

Εφαρμογή 1: Η αστυνομία χρησιμοποιώντας ρανταρ έλεγξε τις ταχύτητες 30 μοτοσυκλετιστών σε μια περιοχή. Τα δεδομένα παρατίθενται παρακάτω:

44 38 41 50 36 36 43 42 49 48

35 40 37 41 43 50 45 45 39 38

50 41 47 36 35 40 42 43 48 33

Να βρεθεί η μέση τιμή, η διάμεσος και η επικρατούσα τιμή

Απάντηση

Χρησιμοποιώντας το excel έχουμε τον παρακάτω πίνακα

Μέση τιμή	41.83333333
Διάμεσος	41.5
Επικρατούσα τιμή	36
Τυπική απόκλιση	5.004021371
Διασπορά	25.04022989
Εύρος	17
Ελάχιστη	33
Μέγιστη	50
Άθροισμα	1255

Άρα η μέση ταχύτητα είναι 41,83, το 50% τρέχει με ταχύτητα κάτω από 41,5 ενώ η επικρατούσα τιμή είναι 36.

Εφαρμογή 2: Ένας μετεωρολόγος καταγράφει τις υψηλές θερμοκρασίες σε μια περιοχή

για τις πρώτες 7 μέρες κάθε έτους. Οι θερμοκρασίες είναι (fahrenheit) 51, 49, 41, 53, 54,

68, 48. Ποια είναι η μέση τιμή και η διάμεσος;

Απάντηση

Η μέση τιμή υπολογίζεται παρακάτω

$$\text{Μέση τιμή} = \Sigma x / n = (51 + 49 + 53 + 41 + 68 + 48 + 54) / 7 = 52$$

Επιπλέον καθώς έχουμε περιττό αριθμό παρατηρήσεων η διάμεσος βρίσκεται στη μέση ακριβώς. Εάν διατάξουμε τα δεδομένα κατά αύξουσα σειρά παίρνουμε την ακόλουθη εικόνα: 41, 48, 49, 51, 53, 54, 68. Επομένως η μεσαία παρατήρηση είναι η 51 και έτσι η διάμεσος ισούται με 51.

Εφαρμογή 3: Τέσσερις φίλοι πραγματοποιούν ένα τεστ IQ. Οι τιμές του σκορ είναι 96, 100, 106, 114. Ποια είναι η μέση τιμή και η διάμεσος;

\

Απάντηση

Η μέση τιμή υπολογίζεται παρακάτω

Μέση τιμή = Σx / n = (96 + 100 + 106 + 114) / 4 = 104

Επιπλέον καθώς έχουμε άρτιο αριθμό παρατηρήσεων η διάμεσος είναι ο μέσος όρος των δύο μεσαίων παρατηρήσεων. Επομένως (100 + 106) / 2 = 103.

Εφαρμογή 4: Ένα δείγμα αποτελείται από 4 παρατηρήσεις {1, 3, 5, 7}. Ποια είναι η τυπική απόκλιση;

Απάντηση

Πρέπει να υπολογίσουμε πρώτα τον μέσο όρο

$$x = (1 + 3 + 5 + 7) / 4 = 4$$

Έπειτα υπολογίζουμε την τυπική απόκλιση με βάσει τον παρακάτω τύπο:

$$s = \text{τετραγωνική ρίζα } [\Sigma (x_i - x)^2 / (n - 1)]$$

$$s = \text{τετραγωνική ρίζα } \{ [(1 - 4)^2 + (3 - 4)^2 + (5 - 4)^2 + (7 - 4)^2] / (4 - 1) \}$$

$$s = \text{τετραγωνική ρίζα } \{ [(-3)^2 + (-1)^2 + (1)^2 + (3)^2] / 3 \}$$

$$s = \text{τετραγωνική ρίζα } \{ [9 + 1 + 1 + 9] / 3 \} = \text{τετραγωνική ρίζα } (20 / 3) =$$

$$\text{τετραγωνική ρίζα } (6.67) = 2.58$$

Εφαρμογή 5: Να υπολογιστεί ο αριθμητικός μέσος με βάσει τον παρακάτω πίνακα

Κλάση	0-50	50-100	100-150	150-200	200-250
Συχνότητα	8	12	20	25	15

Απάντηση

Κλάση	Συχνότητα φ	Κεντρική τιμή x	φx
0-50	8	25	200
50-100	12	75	900
100-150	20	125	2500
150-200	25	175	4375
200-250	15	225	3375
	N =80		11,35

Μέσος όρος = 11350/80 = 141.875

Εφαρμογή 6: Να βρεθεί η επικρατούσα τιμή της βαθμολογίας 16 μαθητών: 0, 0, 2, 2, 3,

3, 4, 5, 5, 5, 5, 6, 6, 7, 8.

Απάντηση

Επειδή η τιμή 5 επαναλαμβάνεται τις περισσότερες φορές η επικρατούσα τιμή είναι ίση

με 5

Εφαρμογή 7: Τα παρακάτω δεδομένα αφορούν ημερομίσθια μαθητευόμενων εργατών σε δύο βιομηχανίες:

Βιομηχανία Α: 100, 440, 450, 460, 470, 480, 400, 500, 500, 520, 540

Βιομηχανία Β: 70, 140, 150, 230, 380, 480, 500, 500, 750, 850, 900

Επειδή τα μέσα ημερομίσθια είναι ίδια και στις δύο βιομηχανίες (450), το ίδιο ισχύει και για την διάμεσο στη κάθε βιομηχανία (480), μπορεί να υποστηριχθεί ότι οι συνθήκες αμοιβής είναι ίδιες και στις δύο βιομηχανίες;

Απάντηση

Για να απαντήσουμε το ερώτημα αυτό μπορούμε να εξετάσουμε το εύρος των αμοιβών για κάθε βιομηχανία. Στη περίπτωση που δεν έχουν οι αμοιβές των δύο βιομηχανιών το ίδιο εύρος τότε η απάντηση στο ερώτημα θα είναι ότι οι συνθήκες αμοιβής δεν είναι ίδιες και στις δύο βιομηχανίες. Στη περίπτωση που οι αμοιβές έχουν το ίδιο εύρος τότε θα πρέπει να χρησιμοποιήσουμε κάποιο άλλο μέτρο διασποράς όπως τη τυπική απόκλιση ή το συντελεστή μεταβολής. Έχουμε ότι:

$$R_A = 540 - 100 = 100$$

$$R_B = 900 - 70 = 830$$

Παρατηρούμε ότι το εύρος των αμοιβών στη βιομηχανία Β να είναι μεγαλύτερο από το εύρος των αμοιβών στη βιομηχανία Α. Αυτό σημαίνει ότι οι αμοιβές στη βιομηχανία Α είναι περισσότερο συγκεντρωμένες γύρω από τη μέση αμοιβή ενώ στη βιομηχανία Β αποκλίνουν από την μέση αμοιβή. Άρα οι συνθήκες αμοιβής δεν είναι ίδιες και στις δύο βιομηχανίες.

Εφαρμογή 8: Για τις ανάγκες της διερεύνησης της κατανομής εισοδημάτων κάποιας περιοχής, ο διευθυντής της ΔΟΥ της περιοχής αυτής αποφάσισε δειγματοληπτικά να μελετήσει το φορολογητέο εισόδημα των κατοίκων της περιοχής. Έτσι επέλεξε ένα δείγμα 100 δηλώσεων του τελευταίου έτους. Από το δείγμα αυτό προέκυψαν τα κάτωθι στοιχεία.

Φορολογητέο εισόδημα (σε χιλιάδες ευρώ)	**Αριθμός δηλώσεων**
[10–20)	14
[20–30)	24
[30–40)	24
[40-50)	15
[50–60)	13
[60–70]	10
Σύνολο	100

Α) Να υπολογισθεί ο αριθμητικός μέσος, η διάμεσος και η διακύμανση των εισοδημάτων αυτών.

Στη συνέχεια ο διευθυντής της ΔΟΥ θέλει να διερευνήσει το ύψος των εσόδων αν σε αυτούς τους φορολογούμενους, ανεξάρτητα από τον φόρο που πληρώνουν, έκανε και μία συμπληρωματική κράτηση ύψους 2% επί του φορολογητέου εισοδήματος της κάθε δήλωσης.

Β) Πόσο θα ήταν το συνολικό ποσό που θα εισέπραττε η υπηρεσία από την συμπληρωματική κράτηση;

Απάντηση

Φορολογητέο εισόδημα (σε χιλιάδες ευρώ)	**Αριθμός δηλώσεων φ_i**	χ_i	$\varphi_i\chi_i$	Φ_i
[10–20)	14	15	210	14
[20–30)	24	25	600	38
[30–40)	24	35	840	62
[40-50)	15	45	675	77
[50–60)	13	55	715	90
[60–70]	10	65	650	100
Σύνολο	100		3690	

i) Μέσος

$$\bar{x} = \frac{\sum_{i=1}^{k} \nu_i \chi_i}{\sum_{i=1}^{k} \nu_i} = \frac{3690}{100} = 36,9$$

Επομένως το μέσο φορολογητέο εισόδημα είναι 36.900 ευρώ.

Διάμεσος

Εντοπίζουμε τη θέση της διαμέσου, 100/2 = 50. Επομένως η διάμεσος βρίσκεται στη τρίτη τάξη μεταξύ 30 και 40 χιλιάδων ευρώ.

$$M = X_\lambda + \frac{\delta}{v_i}\left(\frac{N}{2} - \Phi_i\right) = 30 + 10\frac{50-38}{24} = 35$$

Άρα η διάμεσος είναι ίση με 35000 ευρώ.

Διακύμανση

$$S^2 = \frac{\sum_{i=1}^{k} \phi_i (\chi_i - \bar{x})^2}{\sum_{i=1}^{n} v_i - 1} = \frac{\sum_{i=1}^{k} \phi_i (\chi_i - 36,9)^2}{99} = \frac{23339}{99} = 235,74$$

ii) Το συμπληρωματικό ποσό που θα εισέπραττε η ΔΥΟ από την συμπληρωματική κράτηση θα ήταν:

$$0,02 * \sum_{i=1}^{k} \phi_i \chi_i = 0,02 * 3690 = 73,8$$

73,8 χιλιάδες ευρώ

Εφαρμογή 9: Ο παρακάτω πίνακας δίνει την κατανομή συχνότητας των ηλικιών για ένα τυχαίο δείγμα 50 εργαζομένων σε μια μεγάλη εταιρεία.

Ηλικία	**Αριθμός Εργαζομένων**
20-<28	4
28-<36	6
36-<44	8
44-<52	12
52-<60	16
60-<68	4

Επιπλέον δίνεται ότι η τυπική απόκλιση των ηλικιών των εργαζομένων είναι 10,25.

Με βάση τα στοιχεία αυτά:

Να υπολογισθούν ο αριθμητικός μέσος και η τεταρτημοριακή απόκλιση των ηλικιών.

Απάντηση

Υπολογισμός βοηθητικών στοιχείων

Ηλικία (τάξεις)	*Κεντρική Τιμή Τάξης χ_i*	*Συχνότητα φ_i*	*Αθροιστική Συχνότητα Φ_i*	$\varphi_i * \chi_i$
20-<28	24	4	4	96
28-<36	32	6	10	192
36-<44	40	8	18	320
44-<52	48	12	30	576
52-<60	56	16	46	896
60-<68	64	4	50	256
Σύνολο		***50***		***2336***

Ερώτημα (α):

Αριθμητικός Μέσος

$$\overline{X} = \frac{\sum \nu_i \chi_i}{\sum \nu_i} = \frac{2336}{50} = 46,72 \Rightarrow \overline{X} = 46,72$$

Τεταρτημοριακή Απόκλιση

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

Θα πρέπει να υπολογιστούν τα Q_1, Q_3.

Εντοπισμός της θέσης του Q_1 :

$$\frac{n*i}{4}=\frac{50*1}{4}=\frac{50}{4}=12,5$$

Άρα το Q_1 ανήκει στην 3η τάξη (διάστημα 36 - < 44)

Υπολογισμός της τιμής του Q_1:

$$Q_1 = X_\lambda + \frac{\delta}{v_i}(\frac{N}{4} - \Phi_i) = 36 + 8\frac{(12,5-10)}{8} = 36 + 8\frac{2,5}{8} = 36 + 2,5$$
$$\Rightarrow Q_1 = 38,5$$

Εντοπισμός της θέσης του Q_3:

$$\frac{n*i}{4}=\frac{50*3}{4}=\frac{150}{4}=37,5$$

Άρα το Q_3 ανήκει στην 5η τάξη (διάστημα 52 - < 60)

Υπολογισμός της τιμής του Q_3:

$$Q_1 = X_\lambda + \frac{\delta}{v_i}(\frac{N}{4} - \Phi_i) = 52 + 8\frac{(37,5-30)}{16} = 52 + 8\frac{7,5}{16} = 52 + 3,75$$
$$\Rightarrow Q_3 = 55,75$$

Συνεπώς η Τεταρτημοριακή Απόκλιση είναι:

$$Q = \frac{Q_3 - Q_1}{2}$$
$$= \frac{55,75-38,5}{2} = \frac{17,25}{2} = 8,625$$

Εφαρμογή 10: Δίνονται το μήκος (σε μικροχιλιοστά) και το βάρος (σε γραμμάρια) ενός τυχαίου δείγματος 12 μικροεπεξεργαστών από την ημερήσια παραγωγή ενός εργοστασίου.

Μήκος (x)	10	12	11	13	12	16	14	20	19	21	23	21
Βάρος (y)	18	21	18	23	22	25	24	26	25	27	31	28

Με βάση τα στοιχεία αυτά:

(α) Να υπολογισθούν η διάμεσος, το πρώτο και το τρίτο τεταρτημόριο του μήκους των μικροεπεξεργαστών.

(β) Να διερευνηθεί ποια από τις δύο μεταβλητές (μήκος ή βάρος) παρουσιάζει τη μεγαλύτερη μεταβλητότητα, δεδομένου ότι η μέση τιμή και η διακύμανση του βάρους είναι 24 και 15,091, αντίστοιχα.

Απάντηση

(α) *Διατάσσουμε τα δεδομένα σε αύξουσα τάξη:*

Μήκος (x)	*10*	*11*	*12*	*12*	*13*	*14*	*16*	*19*	*20*	*21*	*21*	*23*

$$M = X_{\frac{n+1}{2}} = X_{\frac{12+1}{2}} = X_{\frac{13}{2}} = X_{6,5} = X_6 + 0,5*(X_7 - X_6) = 14 + 0,5*(16-14) = 15$$

$$Q_1 = X_{\frac{n+1}{4}} = X_{\frac{12+1}{4}} = X_{\frac{13}{4}} = X_{3.25} = X_3 + 0,25*(X_4 - X_3) = 12 + 0,25*(12-12) = 12$$

$$Q_1 = X_{\frac{3(n+1)}{4}} = X_{\frac{3(12+1)}{4}} = X_{\frac{39}{4}} = X_{9,75} = X_9 + 0,75*(X_{10} - X_9) = 20 + 0,75*(21-20) = 20,75$$

X_i	Y_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$	$Y_i - \overline{Y}$	$(Y_i - \overline{Y})^2$	X_i^2	Y_i^2
10	*18*	*–6*	*36*	*–6*	*36*	*100*	*324*
11	*18*	*–5*	*25*	*–6*	*36*	*121*	*324*
12	*21*	*–4*	*16*	*–3*	*9*	*144*	*441*
12	*22*	*–4*	*16*	*–2*	*4*	*144*	*484*
13	*23*	*–3*	*9*	*–1*	*1*	*169*	*529*
14	*24*	*–2*	*4*	*0*	*0*	*196*	*576*
16	*25*	*0*	*0*	*1*	*1*	*256*	*625*
19	*25*	*3*	*9*	*1*	*1*	*361*	*625*
20	*26*	*4*	*16*	*2*	*4*	*400*	*676*
21	*27*	*5*	*25*	*3*	*9*	*441*	*729*
21	*28*	*5*	*25*	*4*	*16*	*441*	*784*
23	*31*	*7*	*49*	*7*	*49*	*529*	*961*
192	***288***		***230***		***166***	***3302***	***7078***

(β) *Ο **Συντελεστής Μεταβλητότητας** δίνεται από τον τύπο:*

$$CV = \frac{s}{\overline{X}} * 100$$

Επομένως πρέπει να υπολογισθεί πρώτα ο αριθμητικός μέσος και η τυπική απόκλιση των τυχαίων μεταβλητών μήκος και βάρος των μικροεπεξεργαστών.

Μήκος των μικροεπεξεργαστών

Αριθμητικός Μέσος

$$\overline{X}=\frac{\sum_{i=1}^{n}X_i}{n}=\frac{192}{12}=16$$

Τυπική Απόκλιση

$$s^2=\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}=\frac{230}{11}=20,909$$

ή εναλλακτικά

$$s^2=\frac{\sum_{i=1}^{n}X_i^2-n\overline{X}^2}{n-1}=\frac{3302-12*16^2}{12-1}=\frac{3302-12*256}{11}=\frac{3302-3072}{11}=20,909$$

Άρα $s=+\sqrt{s^2}\Rightarrow s=+\sqrt{20,909}=4,573$

Αντικαθιστώντας τις τιμές των $\overline{X}, s$ στην (1) υπολογίζεται ο συντελεστής μεταβλητότητας για το <u>μήκος</u> των μικροεπεξεργαστών:

$$CV=\frac{s}{\overline{X}}*100=\frac{4,573}{16}*100=0,28581\cdot 100=28,581\ \mathbf{(1)}$$

Βάρος των μικροεπεξεργαστών

<u>Σημείωση: Ο μέσος όρος και η τυπική απόκλιση για το βάρος είναι δεδομένα, αλλά, εφόσον έχουμε τα y_i, μπορούν να υπολογιστούν ως εξής:</u>

Αριθμητικός Μέσος

$$\overline{Y}=\frac{\sum_{i=1}^{n}Y_i}{n}=\frac{288}{12}=24$$

Τυπική Απόκλιση

$$s^2=\frac{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}{n-1}=\frac{166}{11}=15,091$$

ή εναλλακτικά

$$s^2 = \frac{\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2}{n-1} = \frac{7078 - 12*24^2}{12-1} = \frac{7078-12*576}{11} = \frac{7078-6412}{11} = 15,091$$

Άρα $s = +\sqrt{s^2} \Rightarrow s = +\sqrt{15,091} = 3,885$

Αντικαθιστώντας τις τιμές των $\overline{X}, s$ στην (1) υπολογίζεται ο συντελεστής μεταβλητότητας για το <u>βάρος</u> των μικροεπεξεργαστών:

$$CV = \frac{s}{\overline{Y}} * 100 = \frac{3,885}{24} * 100 = 16,187 \cong 16,19$$

Τελικά, η μεγαλύτερη μεταβλητότητα παρουσιάζεται στο μήκος των μικροεπεξεργαστών.

Εφαρμογή 11: Τα παρακάτω δεδομένα εκφράζουν τη μέση θερμοκρασία του Ιουνίου 2004 σε βαθμούς Κελσίου (Χ) όπως καταγράφηκε σε 20 μεγάλες πόλεις της χώρας.

26	20	25	23	26
29	27	24	21	29
20	25	25	21	23
25	29	23	18	15

Δίνεται επίσης ότι η διακύμανση των θερμοκρασιών είναι 13,90 (βαθμοί Κελσίου).

α) Να υπολογισθούν ο αριθμητικός μέσος, η επικρατούσα τιμή, το εύρος και το ενδοτεταρτημοριακό εύρος των θερμοκρασιών σε βαθμούς Κελσίου.

β) Ένας ερευνητής που πρόκειται να χρησιμοποιήσει τις θερμοκρασίες αυτές για την υποβολή μιας μελέτης σε διεθνές συνέδριο είναι υποχρεωμένος να τις μετατρέψει σε βαθμούς Φαρενάιτ (Υ) με βάση τον τύπο:

$$Y = (9/5)*X + 32$$

Να υπολογισθούν ο αριθμητικός μέσος, η επικρατούσα τιμή, η τυπική απόκλιση και το εύρος των θερμοκρασιών σε βαθμούς Φαρενάιτ.

Απάντηση

Διατάσσουμε τα δεδομένα σε αύξουσα τάξη:

15 18 20 20 21 21 23 23 23 24 25 25 25 25 26 26 27 29 29 29

α)

Αριθμητικός Μέσος

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{474}{20} = 23,7$$

Επικρατούσα Τιμή

Ορίζεται ως η τιμή με τη μεγαλύτερη συχνότητα εμφάνισης. Στην προκειμένη περίπτωση,

$M_0 = 25$

Εύρος

$R = X_{\max} - X_{\min} = 29\text{-}15 \Rightarrow R = 14$

Ενδοτεταρτημοριακό εύρος

$IR = Q_3 - Q_1$

Θα πρέπει να υπολογιστούν τα Q_1 και Q_3

Παρατηρώ ότι για $i=1$ και $n=20$

$\frac{i(n+1)}{4}=\frac{(20+1)}{4}=5,25$. Άρα $A_Q=5$ και $\Delta_Q=0,25$.

Επομένως,

$$Q_1 = X_{\frac{(n+1)}{4}} = X_{A_Q} + \Delta_Q\left(X_{A_Q+1} - X_{A_Q}\right)$$
$$= X_5 + \Delta_Q\left(X_6 - X_5\right)$$

$$= 21 + 0,25(21-21) = 21$$

Παρατηρώ ότι για $i=3$ και $n=20$

$\frac{i(n+1)}{4}=\frac{3(20+1)}{4}=15,75$. Άρα $A_Q=15$ $\Delta_Q=0,75$.

Επομένως,

$$Q_3 = X_{\frac{3(n+1)}{4}}$$
$$= X_{A_Q} + \Delta_Q\left(X_{A_Q+1} - X_{A_Q}\right)$$
$$= X_{15} + \Delta_Q\left(X_{16} - X_{15}\right)$$
$$= 26 + 0,75(26-26) = 26$$

Άρα $IR = Q_3 - Q_1 = 26 - 21 = 5$

β) Χρησιμοποιώντας τις ιδιότητες των στατιστικών μέτρων και τα αποτελέσματα του πρώτου ερωτήματος καταλήγουμε στα εξής:

Αριθμητικός Μέσος

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{\sum[(9/5)X_i + 32]}{n} = \frac{\sum(9/5)X_i + 32n}{n} = \frac{(9/5)\sum X_i}{n} + \frac{32n}{n} = (9/5)\bar{X} + 32$$
$$\Rightarrow \bar{Y} = (9/5)\bar{X} + 32$$

$$\bar{Y} = (9/5)\bar{X} + 32 = (9/5)*23,7 + 32 = 42,66 + 32 = 74,66 \Rightarrow \bar{Y} = 74,66$$

Επικρατούσα Τιμή

$$T_{0Y} = (9/5)T_{0X} + 32 = (9/5)*25+32 = 77 \Rightarrow T_{0Y} = 77$$

Τυπική απόκλιση

$$S_Y^2 = \frac{\sum(Y_i-\bar{Y})^2}{n-1} = \frac{\sum\{[(9/5)X_i+32]-[(9/5)\bar{X}+32]\}^2}{n-1}$$

$$= \frac{\sum[(9/5)X_i-(9/5)\bar{X}]^2}{n-1} = \frac{\sum(9/5)^2(X_i-\bar{X})^2}{n-1}$$

$$= \frac{(9/5)^2\sum(X_i-\bar{X})^2}{n-1} = (9/5)^2 S_X^2$$

Επομένως $S_Y = \sqrt{(9/5)^2 S_X^2} = (9/5)S_X \Rightarrow S_Y = (9/5)S_X$

$$S_X = \sqrt{13,90} = 3,73$$

$$S_Y = (9/5)*3,73 \Rightarrow S_Y = 6,714$$

Εύρος

$$R_Y = Y_{max} - Y_{min} = [(9/5)X_{max}+32]-[(9/5)X_{min}+32] = (9/5)(X_{max}-X_{min})+32-32 = (9/5)R_X$$
$$\Rightarrow R_Y = (9/5)R_X$$

$$R_Y = (9/5)R_X = (9/5)(29-15) = (9/5)*14 = 25,2 \Rightarrow R_Y = 25,2$$

Εφαρμογή 12: Μία ασφαλιστική εταιρεία ενδιαφέρεται να μελετήσει την κατανομή του αριθμού των συμβολαίων στον κλάδο ασφάλισης αυτοκινήτων. Για το λόγο αυτό επέλεξε τυχαίο δείγμα 24 ασφαλιστών αυτοκινήτων και κατέγραψε τον αριθμό των συμβολαίων που έχει συνάψει κάθε ένας από αυτούς κατά τη διάρκεια ενός μήνα. Τα δεδομένα συνοψίζονται στον παρακάτω πίνακα.

14	10	19	21	9	8
12	15	18	17	16	13
10	11	12	14	16	17
13	14	18	14	15	13

(i) Να υπολογιστούν ο αριθμητικός μέσος, η επικρατούσα τιμή, το εύρος καθώς και το ενδοτεταρτημοριακό εύρος. Δίνεται ότι η τυπική απόκλιση του δείγματος είναι 3.30 συμβόλαια.

(ii) Ο προϊστάμενος των ασφαλιστών θεωρεί ότι τα έσοδα κάθε ασφαλιστή i, $i=1,\ldots,24$ από τα συμβόλαια που έχει συνάψει το συγκεκριμένο μήνα υπολογίζονται ως εξής:

$$I_i = 250X_i - 300$$

όπου X_i είναι ο αριθμός των συμβολαίων που έχει συνάψει ο ασφαλιστής i και I_i είναι τα έσοδα του. Να υπολογίσετε τον αριθμητικό μέσο, την τυπική απόκλιση και το εύρος των εσόδων των ασφαλιστών.

Απάντηση

Αριθμητικός Μέσος

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{339}{24} = 14.125$$

Επικρατούσα Τιμή

Ορίζεται ως η τιμή με τη μεγαλύτερη συχνότητα εμφάνισης. Στην προκειμένη περίπτωση, $T_o = 14$

Εύρος

$$R = X_{\max} - X_{\min} = 21 - 8 = 13$$

Ενδοτεταρτημοριακό εύρος

$$IR = Q_3 - Q_1$$

Θα πρέπει να υπολογιστούν τα Q_1 και Q_3

Παρατηρούμε ότι για i =1 και n= 24

$\frac{i(n+1)}{4} = \frac{(24+1)}{4} = 6.25$. Άρα $A_Q = 6$ και $\Delta_Q = 0,25$.

Επομένως,

$$Q_1 = X_{\frac{(n+1)}{4}} = X_A + \Delta_Q(X_{A+1} - X_A)$$
$$= X_6 + \Delta_Q(X_7 - X_6)$$
$$= 12 + 0,25(12 - 12) = 12$$

Παρατηρώ ότι για $i = 3$ και $n = 24$

$\frac{i(n+1)}{4} = \frac{3(24+1)}{4} = 18,75$. Άρα $A_Q = 18$ και $\Delta_Q = 0,75$..

Επομένως,

$$Q_3 = X_{\frac{3(n+1)}{4}} = X_A + \Delta_Q(X_{A+1} - X_A)$$

$$= X_{18} + \Delta_Q(X_{19} - X_{18})$$

$$= 16 + 0,75(17-16) = 16,75$$

Άρα $IR = Q_3 - Q_1 = 16,75 - 12 = 4,75$

(β) Χρησιμοποιώντας τις ιδιότητες των στατιστικών μέτρων και τα αποτελέσματα του πρώτου ερωτήματος καταλήγουμε στα εξής:

Αριθμητικός Μέσος

$\overline{I} = 250\overline{X} - 300 = 250*14.125 - 300 = 3231.25$

Τυπική απόκλιση

$S_I = 250S_x = 250*3.30 = 825$

Εύρος

$R_I = 250R_x = 250*13 = 3250$

Εφαρμογή 13: Ο πίνακας που ακολουθεί παρουσιάζει την κατανομή συχνοτήτων των ημερομισθίων (σε δεκάδες ευρώ), ενός τυχαίου δείγματος 80 εργαζομένων μιας μεγάλης επιχείρησης.

Ημερομίσθια (σε δεκάδες ευρώ)	***Αριθμός Εργαζομένων***
[5-10)	20
[10-15)	30
[15-20)	17
[20-25)	13
Σύνολο	**80**

Δίνεται ότι η τυπική απόκλιση των ημερομισθίων είναι 51,1 ευρώ.

(i) Να υπολογισθεί ο μέσος και η διάμεσος των ημερομισθίων.

(ii) Να προσδιορισθεί το ποσό πάνω από το οποίο βρίσκονται τα ημερομίσθια του 25% των εργαζομένων.

(iii) Η διοίκηση της επιχείρησης αποφάσισε να αυξήσει το ημερομίσθιο του κάθε εργαζομένου κατά 2 ευρώ. Να επαναπροσδιορισθεί ο μέσος και η διάμεσος των νέων ημερομισθίων και να υπολογισθεί ο συντελεστής μεταβλητότητάς τους.

Απάντηση

Υπολογισμός βοηθητικών στοιχείων

Ημερομίσθια (σε δεκάδες ευρώ)	*Αριθμός Εργαζομένων* φ_i	**Κεντρική τιμή** χ_i	**Αθροιστική συχνότητα** Φ_i	$\varphi_i \chi_i$
[5-10)	20	7,5	**20**	**150,00**
[10-15)	30	**12,5**	**50**	**375,00**
[15-20)	17	**17,5**	**67**	**297,50**
[20-25)	13	**22,5**	**80**	**292,50**
Σύνολο	**80**			**1115,00**

i) **Μέσο ημερομίσθιο**

$$\bar{X} = \frac{\sum_{i=1}^{4} \phi_i \chi_i}{\sum_{i=1}^{4} \phi_i} = \frac{1115}{80} = 13,94$$

13,94 δεκάδες ευρώ

Διάμεσος των ημερομισθίων

Εντοπίζουμε τη θέση του Μ

$$\frac{n}{2} = \frac{80}{2} = 40$$

Επομένως το Μ ανήκει στη δεύτερη τάξη, δηλαδή στο διάστημα [10, -5)

Υπολογισμός της τιμή του Μ

$$M = X_\lambda + \frac{\delta}{v_i}(\frac{N}{2} - \Phi_i) = 10 + 5\frac{(40-20)}{30} = 10 + 5\frac{100}{30} = 13,33$$

ii) Εφόσον ζητείται η τιμή του τρίτου τεταρτημορίου θα πρέπει να εντοπίσουμε τη θέση του

$$\frac{3n}{4} = \frac{3*80}{4} = 60$$

Επομένως το Q_3 ανήκει στη τρίτη τάξη, δηλαδή στο διάστημα [10, 15). Υπολογίζουμε το Q_3 πλέον:

$$Q_3 = X_\lambda + \frac{\delta}{v_i}(\frac{3N}{4} - \Phi_i) = 15 + 5\frac{(60-50)}{17} = 15 + 5*0.588 = 17.94$$

Άρα το 25% των εργαζομένων έχει ημερομίσθιο άνω των 179,4 ευρώ

iii) Για τα νέα ημερομίσθια έχουμε

Αριθμητικός μέσος = 13,94 + 0,2 = 14,14 δεκάδες ευρώ

Διάμεσος = 13,33 +0,2 = 13,53 δεκάδες ευρώ

Τυπική απόκλιση = 5,11 δεκάδες ευρώ

Συντελεστής μεταβλητότητας = 5,11 δεκάδες ευρώ

Εφαρμογή 14: Ο διευθυντής ενός υποκαταστήματος της τράπεζας στην προσπάθειά του να ελέγξει το επίπεδο των καταθέσεων στο υποκατάστημα που διευθύνει συλλέγει ένα τυχαίο δείγμα από 60 πελάτες του υποκαταστήματος και καταγράφει το ποσό των καταθέσεών τους (σε ευρώ, στρογγυλοποιημένο στην πλησιέστερη δεκάδα). Ο παρακάτω πίνακας περιέχει τα δεδομένα αυτά.

2530	2420	2870	3520	3030	3490	3110	2970	2930	2820
2870	2940	2150	3380	3010	2200	3120	2780	2720	4040
2330	2460	2550	3170	3020	3290	2980	3310	3240	3290
3230	2940	2650	3070	2810	2850	3080	2740	3560	3310
2240	2610	2980	4240	2710	2980	3740	3720	4240	4160
3830	2970	3490	2910	3380	3990	4110	2490	3870	3790

Α) Να υπολογιστούν ο αριθμητικός μέσος, το εύρος, το ενδοτεταρτημοριακό εύρος καθώς και η τυπική απόκλιση των δεδομένων

Β) Να κατασκευασθεί το Θηκόγραμμα του ποσού των καταθέσεων και με βάση αυτό να εξαχθούν συμπεράσματα για τη μορφή της κατανομής του.

Γ) Στην περίπτωση που η πολιτεία προχωρήσει σε επιβολή φορολογίας των καταθέσεων η οποία ανέρχεται στο 10% του ύψους των καταθέσεων, καθώς επίσης και στην επιβολή πάγιου τέλους που ανέρχεται σε 50 ευρώ, να υπολογισθεί ο αριθμητικός μέσος, η διάμεσος, το εύρος και η τυπική απόκλιση του ύψους των καταθέσεων στο δείγμα μετά την επιβολή της φορολογίας και του πάγιου τέλους.

Απάντηση

Χρησιμοποιώντας το excel έχουμε

Μέσος	3120.5
Εύρος	2090
Τυπική απόκλιση	525.177
Ασυμμετρία	0.412859

Η μέση τιμή υπολογίζεται από τον τύπο $\bar{y} = \dfrac{\sum_{i=1}^{n} y_i}{n} = \dfrac{187230}{60} = 3120,5$

Για την εύρεση των τυπικών αποκλίσεων χρησιμοποιούμε τον τύπο

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{16272885}{59}} = 525.177$$

$R = X_{max} - X_{min} = 4240 - 2150 = 2090$

$$Q_i = X_{\left[i\left(\frac{n+1}{4}\right)\right]} + \Delta_Q \left[X_{\left[i\left(\frac{n+1}{4}\right)\right]+1} - X_{\left[i\left(\frac{n+1}{4}\right)\right]}\right]$$

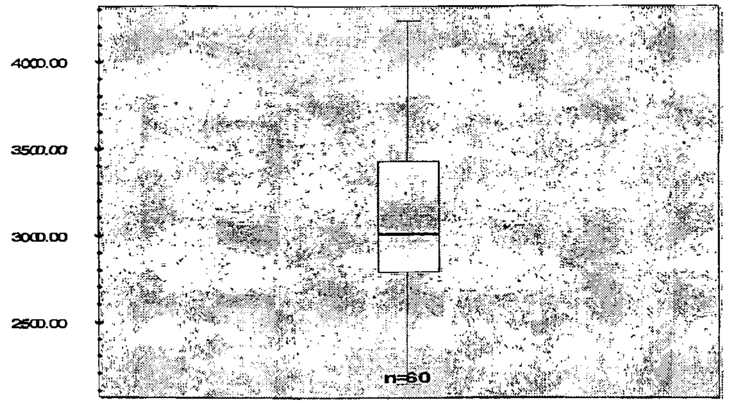
$Q_1 = 2780 + 0.25\,[2810 - 2780] = 2780 + 0.25*30 = 2780 + 7.5 = 2787.5$

$Q_3 = 3380 + 0.75\,[3490 - 3380] = 3380 + 0.75*110 = 3380 + 82.5 = 3462.5$

Άρα το ενδοτεταρτημοριακό εύρος είναι

$IR = Q_3 - Q_1 = 3462.5 - 2787.5 = 675$

Γ)



Το παραπάνω γράφημα επιβεβαιώνει την θετική ασυμμετρία καθώς βλέπουμε τα δεδομένα να είναι συγκεντρωμένα κάτω από την μαύρη γραμμή (διάμεσος). Τα όρια του ορθογωνίου αντιπροσωπεύουν το πρώτο και το τρίτο τεταρτημόριο. Τα άκρα της γραμμής αντιπροσωπεύουν την μέγιστη και ελάχιστη τιμή των δεδομένων.

Δ) Αν X είναι η αρχική τιμή τότε η τελική τιμή είναι 0.9X – 50

$E(0.9X - 50) = 0.9\,E(X) - 50 = 0.9*3120.5 - 50 = 2758.45$

$Var(0.9X - 50) = 0.9^2 Var(X) = 0.81 *525.77^2 \Rightarrow \sigma_X = 0.9*525.77 = 474.525$

$R(0.9X - 50) = 0.9R(X) = 0.9 *2090 = 1881$

E)

Κλάσεις	Κεντρική τιμή	Συχνότητα	vx	X^2	$v*x^2$
1750 -2250	2000	3	6000	4000000	12000000
2250 - 2750	2500	11	27500	6250000	68750000
2750 - 3250	3000	25	75000	9000000	225000000
3250 - 3750	3500	12	42000	12250000	147000000
3750 - 4250	4000	9	36000	16000000	144000000
		60	186500	47500000	596750000
			310.8333		
					288912.429
					537.505748

ΣΤ) Η μέση τιμή είναι ίση με 186500/60 = 310.833.

Η διασπορά κατόπιν υπολογισμών ισούται με 537.505748. Για την επικρατούσα τιμή και το πρώτο τεταρτημόριο κάνουμε τους παρακάτω υπολογισμούς

$$M_o = X_\lambda + \delta \frac{\Delta_1}{\Delta_1 + \Delta_2} = 2750 + \frac{25-11}{25-11+25-12} 500 =$$

$$= 2750 + \frac{14}{27} 500 = 2750 + 259.259 = 3009.259$$

$$Q_1 = X_\lambda + \frac{\delta}{v_i}\left(\frac{N}{4} - \Phi_i\right) = 2750 + \frac{\frac{60}{4} - 14}{25} 500 = 2750 + \frac{1}{25} 500 = 2770$$

Ζ) Προ φόρου

$$CV = \frac{\sigma}{\bar{x}} = \frac{525.77}{3120.5} = 0.168$$

Μετά τον φόρο

$$CV = \frac{\sigma}{\bar{x}} = \frac{474.525}{2758.45} = 0.172$$

Μετά τον φόρο παρατηρείται μεγαλύτερη μεταβλητότητα χωρίς να έχει μεγάλη διαφορά σε σχέση με την κατάσταση που είχαμε πριν το φόρο.

Εφαρμογή 15: Στον παρακάτω πίνακα δίνονται οι ώρες βλάβης των μηχανών ενός εργοστασίου κατά τη διάρκεια 13 συνεχών εβδομάδων και οι αντίστοιχες μονάδες παραγωγής της επιχείρησης για κάθε εβδομάδα.

Εβδομάδα	ώρες βλάβης X	μονάδες παραγωγής Y
1	1,0	24
2	1,5	22
3	2,0	17
4	0,5	27
5	1,0	25
6	2,2	16
7	1,7	21

8	1,3	24
9	0,3	28
10	1,8	20
11	2,1	18
12	0,7	28
13	0,3	31

I) Να υπολογιστούν οι μέσες τιμές για τις ώρες βλάβης και για τις μονάδες παραγωγής.

II) Να δημιουργηθεί διάγραμμα διασποράς στο οποίο θα απεικονίζονται στον άξονα των Χ οι ώρες βλάβης και στον άξονα των Υ οι μονάδες παραγωγής. Μπορούμε να θεωρήσουμε ότι οι τιμές Υ του διαγράμματος είναι συνάρτηση των αντίστοιχων τιμών Χ;

III) Ο υπεύθυνος παραγωγής θεωρεί ότι οι τιμές παραγωγής Υ σε μία εβδομάδα είναι γραμμική συνάρτηση των ωρών βλάβης Χ κατά την διάρκεια της εβδομάδας, σύμφωνα με τη σχέση $Y=\alpha-\beta X$. Από παλαιότερη εμπειρία του γνωρίζει ότι η παραγωγή, αν δεν υπάρχει βλάβη, θα είναι $\alpha=32$ μονάδες, ενώ για κάθε ώρα βλάβης η παραγωγή μειώνεται κατά $\beta=7$ μονάδες. Να υπολογιστούν οι υποθετικές μονάδες παραγωγής σύμφωνα με την υπόθεση γραμμικότητας του υπευθύνου της επιχείρησης.

IV) Να υπολογιστούν οι εβδομαδιαίες διαφορές πραγματικών μονάδων παραγωγής Υ και υποθετικών μονάδων παραγωγής. Επιπλέον να υπολογιστεί και ο μέσος όρος των διαφορών αυτών.

V) Με σκοπό την αύξηση της παραγωγής, η επιχείρηση σκέφτεται να προσλάβει μόνιμο τεχνίτη, ώστε οι ώρες βλάβης να μειωθούν κατά 10%. Να υπολογιστούν οι ώρες βλάβης Χ μειωμένες κατά 10% , και οι νέες υποτιθέμενες μονάδες παραγωγής με τον τύπο του III ερωτήματος. Υπολογίστε ακόμη τους αντίστοιχους μέσους όρους καθώς και την ποσοστιαία αύξηση του μέσου όρου παραγωγής με την πρόσληψη του τεχνίτη.

VI) Να υπολογιστεί με συνάρτηση του EXCEL πόσες φορές η παραγωγή είναι μικρότερη από 20 μονάδες μετά την μείωση των βλαβών κατά 10%.

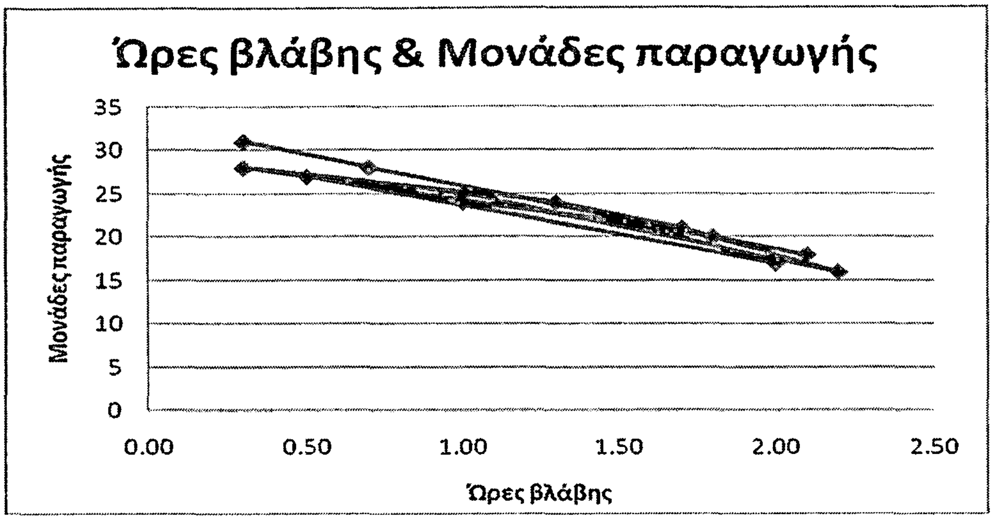
Απάντηση

I)

Εβδομάδα	ώρες βλάβης X	μονάδες παραγωγής Y
1	1.00	24
2	1.50	22
3	2.00	17
4	0.50	27
5	1.00	25
6	2.20	16
7	1.70	21
8	1.30	24
9	0.30	28
10	1.80	20
11	2.10	18
12	0.70	28
13	0.30	31
	1.261538	23.15384615

Ο μέσος χρόνος βλαβών είναι 1.26 ώρες και ο μέσος όρος παραγωγής είναι 23.15 μονάδες.

II)



Από το παραπάνω γράφημα παρατηρούμε ότι όσο αυξάνονται οι ώρες βλάβης τότε μειώνεται ο αριθμός των μονάδων παραγωγής. Επομένως μπορούμε να θερήσουμε ότι οι τιμές Υ του διαγράμματος είναι είναι συναρτήση των αντίστοιχων τιμών του Χ.

III)

Υποθετικές Μονάδες παραγωγής
25
21.5
18
28.5
25
16.6
20.1
22.9
29.9
19.4
17.3
27.1
29.9
23.16923077

Ο υποθετικός μέσος όρος μονάδων παραγωγής είναι 23.16.

IV)

Εβδομαδιαίες διαφορές
-1
0.5
-1
-1.5
0
-0.6
0.9
1.1
-1.9
0.6
0.7
0.9
1.1
-0.015384615

V)

Μείωση βλαβών κατά 10%	Υποθετικές Μονάδες παραγωγής	Ποσοστιαία αύξηση του μέσου όρου παραγωγής
0.9	25.7	3.88%
1.35	22.55	
1.8	19.4	
0.45	28.85	
0.9	25.7	
1.98	18.14	
1.53	21.29	
1.17	23.81	
0.27	30.11	
1.62	20.66	
1.89	18.77	
0.63	27.59	
0.27	30.11	
1.135385	24.05230769	

VI) Χρησιμοποιώντας την συνάρτηση CountIf στο excel υπολογίσαμε ότι η παραγωγή θα είναι μικρότερη από 20 μονάδες μόνο τρεις φορές.

Εφαρμογή 16: Τα παρακάτω στοιχεία προέρχονται από πρόχειρο διαγωνισμό ενός εστιατορίου για την προμήθεια ειδών και αφορούν στις ποσότητες και στις τιμές (σε ευρώ χωρίς Φ.Π.Α.) δύο προμηθευτών Α και Β.

				τιμή μονάδος	
α/α Είδους	***περιγραφή***	***ποσότητα (τεμ)***		**Προσφορά Α**	**Προσφορά Β**
1	καρέκλα	60		120,0	130,0
2	τραπέζι 70x70	18		180,0	150,0
3	τραπεζάκι 30x30	10		110,0	100,0
4	σκαμπό	15		70,0	85,0
5	μαξιλάρι καρέκλας	70		8,0	7,5
6	τραπεζομάντηλο	30		5,0	5,3
7	σταχτοδοχείο απλό	25		1,5	1,5
8	σταχτοδοχείο αντιανεμικό	15		2,3	2,2
9	χαρτοπετσετοθήκη	25		1,0	0,8
10	αλατοπίπερο επιτραπέζιο	20		2,7	2,9
11	λαδόξυδο επιτραπέζιο	20		3,3	3,6
12	ψωμιέρα ξύλινη	20		4,7	4,4
	ΣΥΝΟΛΟ				

Ι) Αν το εστιατόριο αποφασίσει να αγοράσει όλα τα είδη από τον προμηθευτή Α πόσο θα είναι το μερικό κόστος για κάθε είδος και πόσο το συνολικό κόστος για όλα τα είδη; Αν το εστιατόριο αποφασίσει να αγοράσει όλα τα είδη από τον προμηθευτή Β πόσο θα είναι το μερικό κόστος για κάθε είδος και πόσο το συνολικό κόστος για όλα τα είδη;

II) Αν το εστιατόριο αποφασίσει να αγοράσει κάποια είδη από τον προμηθευτή Α και κάποια από τον Β, ανάλογα ποιος προσφέρει μικρότερη τιμή, πόσο θα είναι το οικονομικότερο μερικό κόστος για κάθε είδος και πόσο το συνολικό οικονομικότερο κόστος για όλα τα είδη;

III) Να υπολογιστεί η τιμή επιβαρυμένη με ΦΠΑ 19% του οικονομικότερου συνολικού κόστους σε ένα κελί γράφοντας από πάνω «συνολικό κόστος με ΦΠΑ 19%».

IV) Στο ίδιο σύστημα ορθογωνίων αξόνων να κάνετε δύο διαγράμματα γραμμής όπου να εμφανίζονται α) το κόστος της προσφοράς Α ανά είδος για την προμήθεια όλων των μονάδων του κάθε είδους και β) το κόστος της προσφοράς Β ανά είδος για την προμήθεια όλων των μονάδων του κάθε είδους.

V) Τροποποιείστε τον πίνακα του ερωτήματος **II** παραπάνω ως εξής: Ενοποιήστε τα οκτώ είδη με το χαμηλότερο κόστος (τα 8 τελευταία είδη του πίνακα) σε μία κατηγορία με όνομα «ΔΙΑΦΟΡΑ ΕΙΔΗ». (Δηλ. ο νέος τροποποιημένος πίνακας να περιέχει συνολικά 5 στοιχεία. Τα 4 πρώτα, ήτοι: καρέκλα, τραπέζι 70χ70, τραπεζάκι 30χ30, σκαμπό και το «ΔΙΑΦΟΡΑ ΕΙΔΗ») Να κάνετε κυκλικό διάγραμμα που να εμφανίζει την ποσοστιαία κατανομή του οικονομικότερου κόστους σε σχέση με τα είδη που θέλει να προμηθευτεί το εστιατόριο.

Απάντηση

Ι)

περιγραφή	Μερικό κόστος προμηθευτή Α	Μερικό κόστος προμηθευτή Β
καρέκλα	7200	7800
τραπέζι 70x70	3240	2700
τραπεζάκι 30x30	1100	1000
σκαμπό	1050	1275
μαξιλάρι καρέκλας	560	525
τραπεζομάντηλο	150	159
σταχτοδοχείο απλό	37.5	37.5
σταχτοδοχείο αντιανεμικό	34.5	33
χαρτοπετσετοθήκη	25	20
αλατοπίπερο επιτραπέζιο	54	58
λαδόξυδο επιτραπέζιο	66	72
ψωμιέρα ξύλινη	94	88
ΣΥΝΟΛΟ	13611	13767.5

Το συνολικό κόστος από τον προμηθευτή Α θα είναι 13611 και από τον προμηθευτή Β 13767.5.

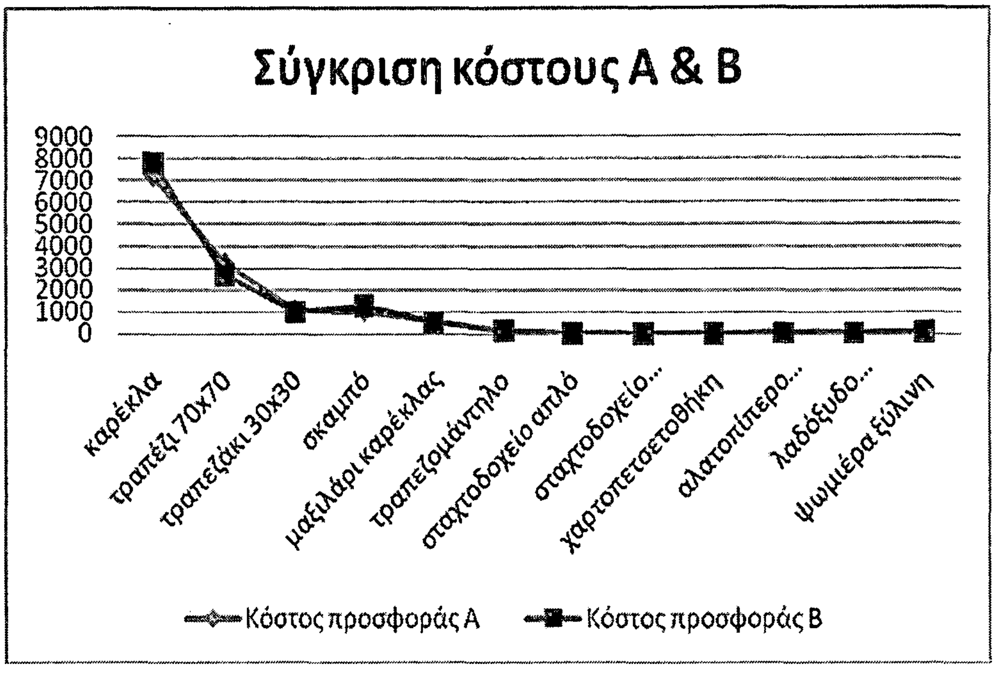
II)

Η οικονομικότερη λύση επιλέγοντας και από τους δύο
7200
2700
1000
1050
525
150
37.5
33
20
54
66
88
Σύνολο 12923.5

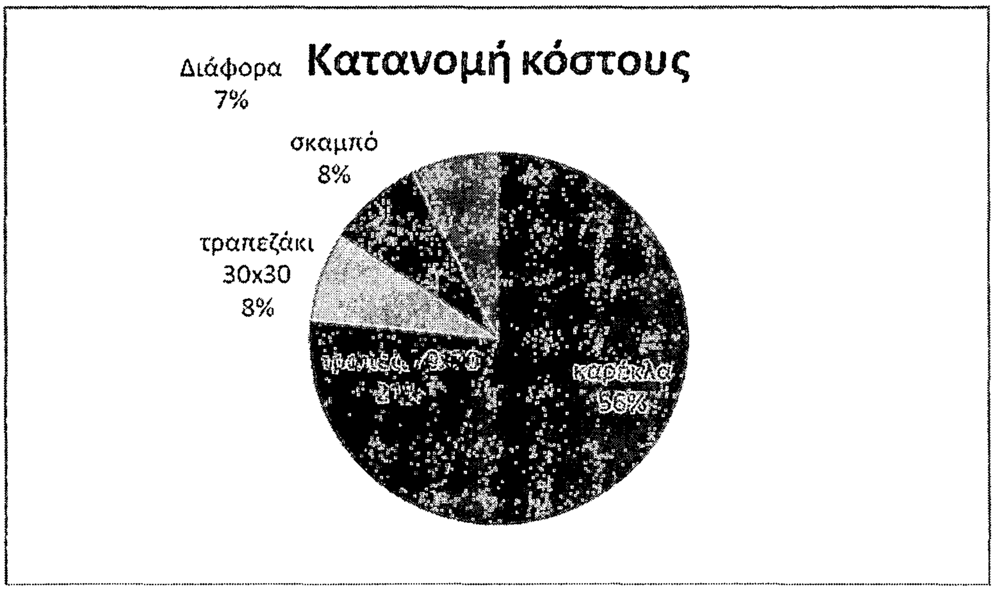
III)

Συνολικό κόστος με ΦΠΑ 19%
€15,378.97

IV)



V)



Κατανομή του οικονομικότερου κόστους για την προμήθεια των ειδών που χρειάζεται το εστιατόριο.

Εφαρμογή 17: Δίνεται η εβδομαδιαία κατανάλωση μπύρας (σε lt) σε δύο καταστήματα Α και Β με βάσει τον επόμενο πίνακα:

	Δ	Τ	Τ	Π	Π	Σ	Κ
Α:	845	1150	1280	963	1117	1426	1208
Β:	45	62	68	57	63	94	65

Ποιο από τα δύο καταστήματα παρουσιάζει περισσότερο ομοιογενή κατανάλωση προϊόντος;

Απάντηση

Η σύγκριση της ομοιογένειας θα γίνει μέσω των συντελεστών μεταβλητότητας για το σημείο πώλησης Α και Β.

Για το Α έχουμε $\sum x_i = 7989$, $\sum x_i^2 = 9342723$ και συνεπώς

$$\bar{X}_A = \frac{\sum_{i=1}^{k} x_i}{N} = \frac{7989}{7} = 1141,286$$

$$S_A^2 = \frac{1}{N-1}[\sum x_i^2 - \frac{(\sum x_i)^2}{N}] = \frac{1}{6}(9342723 - \frac{7989^2}{7}) = ... = 37498,57$$

Ενώ για το Β έχουμε $\sum x_i = 454$, $\sum x_i^2 = 30772$ και συνεπώς

$$\bar{X}_B = \frac{\sum_{i=1}^{k} x_i}{N} = \frac{454}{7} = 64,85714$$

$$S_B^2 = \frac{1}{N-1}[\sum x_i^2 - \frac{(\sum x_i)^2}{N}] = \frac{1}{6}(30772 - \frac{454^2}{7}) = ... = 221,1429$$

Κατά συνέπεια

$$CV_A = \frac{S_A}{\bar{X}_A} = \frac{\sqrt{37498,57}}{1141,286} = 0,169673$$

$$CV_B = \frac{S_B}{\bar{X}_B} = \frac{\sqrt{221,1429}}{64,85714} = 0,229287$$

Παρατηρούμε ότι αν και το σημείο πώλησης Α έχει μεγαλύτερη διακύμανση, παρουσιάζει πιο ομοιογενή κατανάλωση από το Β, αφού έχει μιρκότερο συντελεστή μεταβλητότητας από το Β. Η μεγαλύτερη διακύμανση εξηγείται από τη μεγαλύτερη τάξη μεγέθους της κατανάλωσης στο Α.

Εφαρμογή 18: Ο πίνακας που ακολουθεί παρουσιάζει την κατανομή συχνότητας των υπερωριών (σε ώρες), ενός τυχαίου δείγματος 60 εργαζομένων ενός μεγάλου πολυκαταστήματος για το μήνα Μάρτιο.

Υπερωρίες (σε ώρες)	*Αριθμός Εργαζομένων*
[0-5)	12
[5-10)	15
[10-15)	12
[15-20)	10
[20-25)	7
[25-30)	4
Σύνολο	**60**

(i) Να υπολογισθεί ο αριθμητικός μέσος, η διάμεσος, η επικρατούσα τιμή και το τρίτο τεταρτημόριο. Δίνεται ότι η τυπική απόκλιση του δείγματος είναι 7.62 ώρες.

(ii) Σε ένα άλλο πολυκατάστημα της ίδιας αλυσίδας, από ένα αντίστοιχο τυχαίο δείγμα 60 εργαζομένων υπολογίζεται ότι η μέση τιμή υπερωριών, για το μήνα Μάρτιο, είναι 10.25 και η τυπική απόκλιση είναι 7.4 ώρες. Με βάση τα στοιχεία αυτά να διερευνηθεί σε ποιό από τα δύο πολυκαταστήματα παρουσιάζεται μεγαλύτερη μεταβλητότητα όσον αφορά τις υπερωρίες των εργαζομένων τους.

Απάντηση

(i) Υπολογισμός βοηθητικών στοιχείων

Υπερωρίες	**Κεντρική Τιμή** x_i	**Αριθμός Εργαζομένων** φ_i	**Αθροιστική Συχνότητα** Φ_i	$\varphi_i x_i$
[0-5)	2.5	12	12	30
[5-10)	7.5	15	27	112.5
[10-15)	12.5	12	39	150
[15-20)	17.5	10	49	175
[20-25)	22.5	7	56	157.5
[25-30)	27.5	4	60	110
ΣΥΝΟΛΑ		**60**		**735**

Αριθμητικός Μέσος

$$\overline{X} = \frac{\sum_{i=1}^{6} v_i \chi_i}{\sum_{i=1}^{6} v_i} = \frac{735}{60} = 12.25 \Rightarrow \overline{X} = 12.25 \text{ ώρες}$$

Διάμεσος

Εντοπισμός της θέσης του M:

$$\frac{n}{2} = \frac{60}{2} = 30$$

Άρα το M ανήκει στην 3η τάξη (διάστημα 10 - <15)

Υπολογισμός της τιμής του M:

$$M = X_\lambda + \frac{\delta}{v_i}(\frac{N}{2} - \Phi_i) = 10 + 5\frac{(30-27)}{12} = 10 + 5\frac{3}{12} = 11.25 \text{ ώρες}$$

Τρίτο Τεταρτημόριο

Εντοπισμός της θέσης του Q_3:

$$\frac{n^*i}{4}=\frac{60^*3}{4}=\frac{180}{4}=45$$

Άρα το Q_3 ανήκει στην 4^η τάξη (διάστημα 15 – <20)

Υπολογισμός τιμής Q_3:

$$Q_K = X_\lambda + \frac{\delta}{v_i}(\frac{\kappa N}{4} - \Phi_i) = 15 + 5\frac{(45-39)}{10} = 15 + 5\frac{6}{10} = 18 \Rightarrow Q_1 = 18 \text{ ώρες}$$

Θα πρέπει να υπολογιστεί και το M_o

Εντοπισμός της θέσης του M_o:

Η τάξη με τη μεγαλύτερη συχνότητα είναι η 2^η (διάστημα 5 – <10). Άρα η επικρατούσα τιμή βρίσκεται στην τάξη αυτή.

Υπολογισμός της τιμής του M_o

$$M_o = X_\lambda + \delta\frac{\Delta_1}{\Delta_1+\Delta_2} = 5+5^*\frac{(15-12)}{(15-12)+(15-12)} = 5+5^*\frac{3}{6} \Rightarrow M_o = 7.5 \text{ ώρες}$$

(ii) Δεδομένου ότι τα δύο δείγματα έχουν διαφορετικούς μέσους η σύγκριση της μεταβλητότητας θα γίνει με βάση το Συντελεστή Μεταβλητότητας.

$$CV_A = \frac{S}{\overline{X}}*100 = \frac{7.62}{12.25}*100 = 62.20\%$$

και

$$CV_B = \frac{S}{\overline{X}}*100 = \frac{7.4}{10.25}*100 = 72.20\%$$

Επομένως το δεύτερο πολυκατάστημα έχει μεγαλύτερη μεταβλητότητα σε σχετικούς όρους, ενώ έχει μικρότερη τυπική απόκλιση.

Εφαρμογή 19: Η κατανομή 200 επιχειρήσεων ανάλογα με το ύψος των μηνιαίων πωλήσεων τους έχει ως εξής:

Αξία πωλήσεων (εκατομμύρια δρχ.)	Αριθμός επιχειρήσεων
4-6	10
6-8	20
8-10	30
10-12	80
12-14	30
14-16	20
16-18	10

Να υπολογιστούν και να ερμηνευτούν τα εξής στατιστικά μέτρα:

1) Ο μέσος αριθμητικός
2) Η διάμεσος
3) Η επικρατούσα τιμή

Απάντηση

Χρήσιμοι υπολογισμοί

Αξία πωλήσεων (εκατομμύρια δρχ.)	**Κεντρική Τιμή** χ_i	**Αριθμός επιχειρήσεων** φ_ι	**Αθροιστική Συχνότητα** Φ_ι	$\varphi_\iota \chi_\iota$
4-6	5	10	10	50
6-8	7	20	30	140
8-10	9	30	60	270
10-12	11	80	140	880
12-14	13	30	170	390
14-16	15	20	190	300
16-18	17	10	200	170
ΣΥΝΟΛΑ		**200**		2200

1) **Αριθμητικός Μέσος**

$$\overline{X} = \frac{\sum_{i=1}^{7} \nu_i \chi_i}{\sum_{i=1}^{7} \nu_i} = \frac{2200}{200} = 11 \Rightarrow \overline{X} = 11 \text{ αξία πωλήσεων}$$

Παρατηρούμε ότι η μέση αξία πωλήσεων είναι 11 εκατομμύρια δραχμές για τις 200 επιχειρήσεις

2) **Διάμεσος**

Εντοπισμός της θέσης του Μ:

$$\frac{n}{2} = \frac{200}{2} = 100$$

Άρα το Μ ανήκει στην 4^{η} τάξη (διάστημα 10 - <12)

Υπολογισμός της τιμής του Μ:

$$M = X_\lambda + \frac{\delta}{v_i}\left(\frac{N}{2} - \Phi_i\right) = 10 + 2\frac{(100-60)}{80} = 11$$

Παρατηρούμε το 50% των επιχειρήσεων έχει αξία πωλήσεων άνω των 11 εκατομμυρίων δραχμών.

3) Η επικρατούσα τιμή

Εντοπισμός της θέσης του M_0:

4) Η τάξη με τη μεγαλύτερη συχνότητα είναι η 4η (διάστημα 10- 12). Άρα η επικρατούσα τιμή βρίσκεται στην τάξη αυτή.

Υπολογισμός της τιμής του M_0:

$$M_o = X_\lambda + \delta\frac{\Delta_1}{\Delta_1 + \Delta_2} = 10 + 2 * \frac{(80-30)}{(80-30)+(80-30)} = 11 \Rightarrow M_o = 11$$

Εφαρμογή 20: Δίνονται οι εξής κατανομές συχνοτήτων:

Κατανομή (α)

Μηνιαίες αποδοχές (χιλ. δρχ.)	**Αριθμός υπαλλήλων**
45 - 50	**10**
50 – 55	**40**
55 - 60	**50**
60 – 65	**40**
65 - 70	**10**

Κατανομή (β)

Μηνιαίες αποδοχές (χιλ. δρχ.)	Αριθμός υπαλλήλων
30 - 40	20
40 – 50	40
50 – 60	50
60 – 70	40
70 – 80	20

Ερωτάται ένας εργαζόμενος από απόψεως αποδοχών, σε ποια κατανομή θα προτιμούσε να τοποθετηθεί; Στην α ή την β;

Απάντηση

Για να απαντήσουμε σε αυτό το ερώτημα θα υπολογίσουμε την μέση τιμή και τον συντελεστή μεταβλητότητας για τις δύο κατανομές. Από την παρατήρηση των αριθμών στους παραπάνω δύο πίνακες φαίνεται ότι θα υπάρξει μεγαλύτερη μεταβλητότητα στη δεύτερη κατανομή ενώ εκτιμούμε ότι οι μέσες αποδοχές θα είναι σχετικά παρόμοιες (πολύ κοντά) και στις δύο κατανομές. Ας δούμε όμως από τους υπολογισμούς τι θα προκύψει.

Χρήσιμοι υπολογισμοί

Κατανομή (α)

Μηνιαίες αποδοχές (χιλ. δρχ.)	Αριθμός υπαλλήλων φ_i	Κεντρική τιμή χ_i	Αθροιστική συχνότητα Φ_i	$\varphi_i \chi_i$
45 - 50	10	47,5	10	475
50 – 55	40	52,5	50	2100
55 - 60	50	57,5	100	2875
60 – 65	40	62,5	140	2500
65 - 70	10	67,5	150	675
Σύνολο	150			8625

Κατανομή (β)

Μηνιαίες αποδοχές (χιλ. δρχ.)	Αριθμός υπαλλήλων φ_i	Κεντρική τιμή χ_i	Αθροιστική συχνότητα Φ_i	$\varphi_i \chi_i$
30 - 40	20	35	20	700
40 – 50	40	45	60	1800
50 – 60	50	55	110	2750
60 – 70	40	65	150	2600
70 – 80	20	75	170	1500
Σύνολο	170			9350

Υπολογίζουμε τις αποδοχές και για τις δύο κατανομές

Αριθμητικός Μέσος (α) κατανομής

$$\overline{X} = \frac{\sum_{i=1}^{6} \nu_i \chi_i}{\sum_{i=1}^{6} \nu_i} = \frac{8625}{150} = 57,5 \Rightarrow \overline{X} = 57,5$$

Αριθμητικός Μέσος (β) κατανομής

$$\overline{X} = \frac{\sum_{i=1}^{6} \nu_i \chi_i}{\sum_{i=1}^{6} \nu_i} = \frac{9350}{170} = 55 \Rightarrow \overline{X} = 55$$

Υπολογισμός τυπικών αποκλίσεων

Κατανομή (α)

Αριθμός υπαλλήλων φ_i	**Κεντρική τιμή χ_i**	$(X_i - \overline{X})^2$	$f_i * (X_i - \overline{X})^2$
10	**47,5**	**100**	**1000**
40	**52,5**	**25**	**1000**
50	**57,5**	**0**	**0**
40	**62,5**	**25**	**1000**
10	**67,5**	**100**	**1000**
		Σύνολο	**4000**

$$S^2 = \frac{\sum_{i=1}^{5} \phi_i (\chi_i - \bar{x})^2}{\sum_{i=1}^{5} \phi_i - 1} = \frac{4000}{149} = 26,84$$

$$S = 5,181$$

Κατανομή (β)

Αριθμός υπαλλήλων φ_i	Κεντρική τιμή χ_i	$(X_i - \overline{X})^2$	$f_i*(X_i - \overline{X})^2$
20	35	400	8000
40	45	100	4000
50	55	0	0
40	65	100	4000
20	75	400	8000
		Σύνολο	24000

$$S^2 = \frac{\sum_{i=1}^{5} \phi_i (\chi_i - \bar{x})^2}{\sum_{i=1}^{5} \phi_i - 1} = \frac{24000}{169} = 142,01$$

$$S = 11,91$$

$$CV_a = \frac{S}{\overline{\overline{X}}} * 100 = \frac{5,18}{57,5} * 100 = 9,0\%$$

και

$$CV_\beta = \frac{S}{\overline{\overline{X}}} * 100 = \frac{11,91}{55} * 100 = 21,65\%$$

Από τα παραπάνω φαίνεται ότι η πρώτη κατανομή όχι μόνο έχει σχετικά υψηλότερο μέσο μισθό αλλά ταυτόχρονα έχει μικρότερη μεταβλητότητα στους μισθούς σε σχέση με την β κατανομή, λιγότερη από την μισή μεταβλητότητα της β. Επομένως ο εργαζόμενος θα ήταν λογικό να επιλέξει την α κατανομή.

Εφαρμογή 21: Μια βιομηχανία παράγει δύο τύπους λαμπτήρων T.V., A και B. Οι λαμπτήρες έχουν αντίστοιχα μέση διάρκεια ζωής $\bar{x}_a = 1495$ και $\bar{x}_b = 1875$ ώρες και μέσες αποκλίσεις τετραγώνου $s_a = 280$ και $s_b = 310$ ώρες.

Ερωτάται: α) ποιος τύπος έχει τη μεγαλύτερη απόλυτη διασπορά και σχετική διασπορά, β)ποιου τύπου λαμπτήρες παρουσιάζουν μεγαλύτερη ομοιογένεια;

Απάντηση

Μεγαλύτερη απόλυτη διασπορά έχουν οι λαμπτήρες τύπου β διότι $s_a = 280$ < $s_b = 310$. ως προς την σχετική διασπορά θα πρέπει να υπολογίσουμε τους συντελεστές μεταβλητότητας των δύο λαμπτήρων.

$$CV_a = \frac{S}{\overline{X}} * 100 = \frac{280}{1495} * 100 = 18,72\%$$

και

$$CV_\beta = \frac{S}{\overline{X}} * 100 = \frac{310}{1875} * 100 = 16,5\%$$

Παρατηρούμε ότι η σχετική μεταβλητότητα στον λαμπτήρα β να είναι μικρότερη του λαμπτήρα α. Επομένως οι λαμπτήρες τύπου β έχουν μεγαλύτερη ομοιογένεια.

Εφαρμογή 22: Σε μια βιομηχανία κατασκευής μπαταριών υπάρχουν δύο γραμμές παραγωγής. Η πρώτη γραμμή παραγωγής κατασκευάζει την μπαταρία τύπου Α και η δεύτερη την μπαταρία τύπου Β. Είναι γνωστό ότι το κόστος παρασκευής της μπαταρίας τύπου Α είναι 38 ευρώ ανά τεμάχιο και της μπαταρίας τύπου Β 40 ανά τεμάχιο. Στο παρακάτω πίνακα δίνεται ο χρόνος ζωής για ένα δείγμα 5 μπαταριών για τον κάθε τύπο.

Μπαταρία τύπου Α	Μπαταρία τύπου Β
20	26
26	32
24	19
22	20
18	23

Να βρεθούν τα παρακάτω:

Α) ο μέσος χρόνος ζωής ανά τύπο μπαταρίας

Β) το μέσο κόστος παρασκευής του κάθε τύπου μπαταρίας για κάθε χίλιες ώρες ζωής

Γ) ποιο δείγμα μπαταριών έχει την μεγαλύτερη ομοιογένεια

Απάντηση

α.

$$\bar{x}_A = \frac{20+26+24+22+18}{5} = 22 \text{ χιλ. ώρες}$$

$$\bar{x}_B = \frac{26+32+19+20+23}{5} = 24 \text{ χιλ. ώρες.}$$

β. Το μέσο κόστος ανά χιλιάδα ωρών και ανά τύπο μπαταρίας είναι:

για τον τύπο Α: $\frac{38}{22} \approx 1{,}73$ €/χιλ. ώρες

για τον τύπο Β: $\frac{40}{24} \approx 1{,}66$ €/χιλ. ώρες

Άρα συμφέρει ο τύπος Β.

γ. $$S_A^2 = \frac{(20-22)^2+(26-22)^2+(24-22)^2+(22-22)^2+(18-22)^2}{5} = 8(\text{χιλ. ώρες})^2$$

Επομένως $S_A = 2\sqrt{2}(\text{χιλ. ώρες})$

$$S_B^2 = \frac{(26-24)^2+(32-24)^2+(19-24)^2+(20-24)^2+(23-24)^2}{5} = 22(\text{χιλ.ώρες})^2$$

Επομένως $S_B = \sqrt{22}(\text{χιλ.ώρες})$

δ. $$CV_A = \frac{2\sqrt{2}}{22} = \frac{\sqrt{2}}{11},\ CV_B = \frac{\sqrt{22}}{24},\ \text{άρα } \frac{CV_A}{CV_B} = \frac{\frac{\sqrt{2}}{11}}{\frac{\sqrt{2}\sqrt{11}}{24}} = \frac{24}{11\sqrt{11}} \approx \frac{24}{11 \cdot 3{,}3} < 1$$

Επομένως $CV_A < CV_B$, δηλαδή μεγαλύτερη ομοιογένεια έχει το δείγμα Α.

Εφαρμογή 23: Στον επόμενο πίνακα δίνονται οι τιμές x_i, i=1,2,3,4 μιας μεταβλητής X με αντίστοιχες συχνότητες ν_i, i=1,2,3,4. Η συχνότητα ν_2 που αντιστοιχεί στην τιμή x_2=3 είναι άγνωστη. Δίνεται ότι η μέση τιμή των παρατηρήσεων είναι ίση με =4.

x_i	ν_i
2	6
3	;
5	3
8	4

α. Να δείξετε ότι $v_2 = 7$.

β. Να αποδείξετε ότι η διακύμανση των παρατηρήσεων είναι ίση με 4,9.

γ. Να εξετάσετε αν το δείγμα των τιμών της μεταβλητής Χ είναι ομοιογενές.

Απάντηση

x_i	v_i	$x_i * v_i$	$(x_i-\bar{x})^2 v_i$
2	6	12	24
3	α	3α	α
5	3	15	3
8	4	32	64
Άθροισμα	13+α	59+3α	91+α

Α) $\bar{x} = \frac{\sum_{1}^{4} x_i v_i}{v} \Leftrightarrow 4 = \frac{59+3a}{13+a} \Leftrightarrow a = 7$

Β) $s^2 = \frac{\sum_{i=1}^{4} (x_i - \bar{x})^2 v_i}{v} = \frac{98}{20} = 4.9$

Γ) $CV = \frac{s}{\bar{x}} = \frac{2.2}{4} = 0.55 = 55\% > 10\%$ άρα το δείγμα **δεν** είναι ομοιογενές

Εφαρμογή 24: Οι 70 δημόσιοι υπάλληλοι δευτεροβάθμιας εκπαίδευσης μιας νομαρχίας έχουν μέσο μηνιαίο μισθό 800 Ευρώ, ενώ οι υπάλληλοι πανεπιστημιακής εκπαίδευσης έχουν μέσο μισθό 1.100 Ευρώ. Ο μέσος μισθός όλων των υπαλλήλων στη νομαρχία είναι 890 ευρω.

Α. α. Να βρείτε πόσοι είναι οι υπάλληλοι πανεπιστημιακής εκπαίδευσης.

β. Ποιο είναι το μηνιαίο οικονομικό κονδύλι που απαιτείται για την αποπληρωμή όλων

των εργαζομένων στην νομαρχία;

Β. Την 1η Ιανουαρίου του έτους 2005 δόθηκε αύξηση 30 Ευρώ μηνιαία σε κάθε υπάλληλο δευτεροβάθμιας εκπαίδευσης και 4% σε κάθε υπάλληλοι πανεπιστημιακής εκπαίδευσης. Να υπολογίσετε:

α. Το μέσο μισθό των υπαλλήλων δευτεροβάθμιας εκπαίδευσης και το μέσο μισθό των

υπάλληλοι πανεπιστημιακής εκπαίδευσης όπως θα διαμορφωθεί μετά την αύξηση.

β. Το νέο μισθό των εργαζομένων στην νομαρχία.

Απάντηση

Θεωρούμε ν_Δ και ν_Π τον αριθμό των υπαλλήλων της δευτεροβάθμιας και πανεπιστημιακής εκπαίδευσης αντίστοιχα που εργάζονται στην νομαρχία. Επίσης θεωρούμε $\bar{x}_\Delta$ και $\bar{x}_\Pi$ τους μέσους μηνιαίους μισθούς των υπαλλήλων της δευτεροβάθμιας και πανεπιστημιακής εκπαίδευσης αντίστοιχα και $\bar{x}$ ο μέσος μηνιαίος μισθός όλων των υπαλλήλων.

Α.α θα έχουμε

$$\bar{x} = 890 <=> \frac{\nu_\Delta \cdot \bar{x}_\Delta + \nu_\Pi \cdot \bar{x}_\Pi}{\nu_\Delta + \nu_\Pi} = 890 <=>$$

$$\frac{70 \cdot 800 + \nu_\Pi \cdot 1100}{70 + \nu_\Pi} = 89 <=> 56000 + 1100\nu_\Pi = 62300 + 890\nu_\Pi <=>$$

$$1100\nu_\Pi + -890\nu_\Pi = 62300 - 56000 <=> 210\nu_\Pi = 6300 <=> \nu_\Pi = 30$$

Α.β. Αφού οι υπάλληλοι είναι συνολικά 70 +30 =100 και ο μέσος μηνιαίος μισθός είναι 890 ευρώ το συνολικό ποσό που απαιτείται για την αποπληρωμή τους είναι 89000 ευρώ κάθε μήνα.

Β.α Μετά την πρώτη Ιανουαρίου οι νέοι μέσοι μισθοί διαμορφώνονται ως εξής: για τους υπαλλήλους της δευτεροβάθμιας εκπαίδευσης $\bar{x}'_{\Delta} = \bar{x}_{\Delta} + 30 = 800 + 30 = 830$ ευρώ ενώ για τους υπαλλήλους της πανεπιστημιακής εκπαίδευσης

$\bar{x}'_{\Pi} = \bar{x}_{\Pi} + \frac{4}{10}\bar{x}_{\Pi} = 1,04 * 1100 = 1144$ ευρώ.

Β. Ο νέος μέσος μηνιαίος μισθός όλων των υπαλλήλων μετά την 1η Ιανουαρίου είναι

$$\bar{x} = \frac{\nu_{\Delta} \cdot \bar{x}'_{\Delta} + \nu_{\Pi} \cdot \bar{x}'_{\Pi}}{\nu_{\Delta} + \nu_{\Pi}} = \frac{70 \cdot 830 + 30 \cdot 1144}{70 + \nu_{\Pi}} = \frac{92420}{100} = 924,2$$

Εφαρμογή 25: Το μέσο ύψος 70.000 μαθητών της Γ' λυκείου είναι $\bar{x} = 172$ και η τυπική απόκλιση είναι s= 7 cm.

α. Να αποδείξετε ότι το δείγμα των μαθητών της Γ' λυκείου έχει ομοιογένεια ως προς το

ύψος.

β. Αν κατά την μέτρηση του ύψους όλων των μαθητών είχε από λάθος μετρηθεί 2 cm περισσότερο από το πραγματικό να βρείτε πόσο είναι το «πραγματικό» μέσο ύψος.

γ. Λαμβάνοντας υπόψη τα «πραγματικά» στοιχεία του ύψους των μαθητών, το δείγμα σε αυτήν την περίπτωση είναι περισσότερο ή λιγότερο ομοιογενές από το δείγμα για το οποίο είχαμε λάβει υπόψη τα «πλασματικά» στοιχεία του ύψους;

Απάντηση

α. Υπολογίζουμε τον συντελεστή μεταβολής για το δείγμα. Είναι

$$CV = \frac{S}{\bar{X}} \cdot 100 = \frac{7}{172} \cdot 100 = 4.06\% < 10\%$$

Επομένως το δείγμα έχει ομοιογένεια ως προς το ύψος.

β. Το πραγματικό μέσο ύψος θα είναι $\bar{y} = \bar{x} - 2 = 172 - 2 = 170$

γ. Η τυπική απόκλιση λαμβάνοντας υπόψη τα πραγματικά στοιχεία του ύψους των μαθητών δεν μεταβάλλεται και έτσι $s_y = s_x = 7$ cm. Ο συντελεστής μεταβλητότητας έχοντας υπόψη την πραγματική μέση τιμή του ύψους είναι

$$CV_y = \frac{S_y}{\bar{y}} \cdot 100 = \frac{7}{170} \cdot 100 = 4.11\% < 10\%$$

Παρατηρούμε ότι λαμβάνοντας υπόψη τα πραγματικά στοιχεία του ύψους των μαθητών ο συντελεστής μεταβλητότητας του δείγματος αυξάνει σε σχέση με το συντελεστή μεταβλητότητας του δείγματος που υπολογίσαμε λαμβάνοντας υπόψη τα πλασματικά στοιχεία του ύψους, δηλαδή το δείγμα με τα πραγματικά στοιχεία είναι λιγότερο ομοιογενές συγκριτικά με αυτό που είχαμε αρχικά.

Εφαρμογή 26: Να υπολογιστεί η συχνότητα και η παρατήρηση που λείπει στον παρακάτω πίνακα συχνοτήτων αν γνωρίζουμε ότι (ζυγός αριθμός παρατηρήσεων) η μέση τιμή είναι ίση με 3,02

x_i	v_i
2	6
2,02	α
5	3
5,7725	8
Άθροισμα	15+α

Απάντηση

$$\bar{x} = \frac{\sum_{1}^{4} x_i v_i}{v} \Leftrightarrow 3.02 = \frac{27 + 2,02a + 8\beta}{9 + a} \Leftrightarrow 27,18 + 3,02a = 27 + 2,02a + 8\beta \Leftrightarrow$$

$$0,18 + \alpha = 8\beta$$

Επειδη καμία από τις παρατηρήσεις δεν ισούται με 5,545 προφανώς η διάμεσος υπολογίζεται σαν το ημιάθροισμα των δύο μεσαίων παρατηρήσεων οι οποίες είναι διαδοχικοί αριθμοί (το β παίρνει προφανώς τιμές από 6 και πάνω). Με βάσει την τιμή της διαμέσου πρέπει 5 + β =2* 5,545 => β=6,09. Επομένως

$$0,18 + \alpha = 8 \cdot 5,7725 \Leftrightarrow 0,18 + \alpha = 40,18 \Leftrightarrow \alpha = 40$$

Εφαρμογή 27: Η βαθμολογία στα 10 μαθήματα ενός μαθητή είναι: 10, 11, 19, 10, 10, 15, 15, 15, 11, 10. Να υπολογίσετε:

α. Τη μέση τιμή. β. Τη διακύμανση. γ. Την τυπική απόκλιση. δ. Το εύρος.

ε. Το συντελεστή μεταβολής.

Απάντηση

Κατασκευάζουμε πρώτα τον παρακάτω πίνακα:

x_i	v_i	$x_i * v_i$	x^2_i	$x_i * v^2_i$
10	4	40	100	400
11	2	22	121	242
15	3	45	225	675
19	1	19	361	361
Άθροισμα	10	126		1678

Α. $\bar{x} = \dfrac{\sum_{1}^{4} x_i v_i}{10} = \dfrac{126}{10} = 12,6$

Β. $s^2 = \dfrac{1}{\nu}\left\{\sum_{i=1}^{k} x_i^2 v_i - \dfrac{\left(\sum_{i=1}^{k} x_i v_i\right)^2}{\nu}\right\} = \dfrac{1}{10}(1678 - \dfrac{126^2}{10}) = 167,8 - 158,76 = 9,04$

Γ. $s = \sqrt{9.04} \approx 3$

Δ. R = xmax – xmin = 19 – 10 = 9

Ε. $CV = \dfrac{S}{\bar{x}} \cdot 100 = \dfrac{3}{12.6} \cdot 100 = 0.238 = 23.8\%$

Εφαρμογή 28: Οι τιμές πώλησης δέκα ηλεκτρικών συσκευών (σε ευρώ) είναι: 265, 280, 280, 310, 400, 310, 280, 350, 280 και 265.

α. Να υπολογίσετε τη μέση τιμή και τη διάμεσο.

β. Εάν τα είδη πωληθούν με 10% έκπτωση, ποια από τα παραπάνω μεγέθη θα αλλάξουν και ποιες θα είναι οι τιμές τους μετά την αλλαγή που θα υποστούν;

γ. Εάν οι πελάτες επιβαρυνθούν με 10 € επιπλέον στη τιμή αγοράς για έξοδα μεταφοράς, ποια από τα παραπάνω μεγέθη του ερωτήματος β. θα αλλάξουν και ποιες θα είναι οι τιμές τους μετά την αλλαγή που θα υποστούν;

Απάντηση

Διατάσσουμε τις τιμές κατά αύξουσα σειρά:

265, 265, 280, 280, 280, 280, 310, 310, 350, 400

A. $\bar{x} = \frac{\sum_{1}^{10} x_i v_i}{10} = \frac{3020}{10} = 302$ και $\delta = \frac{x_5 + x_6}{2} = \frac{280 + 280}{2} = 280$

B. Οι τιμές με την έκπτωση είναι οι: $y_i = 0{,}9x_i$, με $i = 1,\ldots, 10$. Επομένως έχουμε:

$$\bar{y} = 0.9\bar{x} = 271.8 \text{ και } \delta_y = \frac{y_5 + y_6}{2} = \frac{0{,}9x_5 + 0{,}9x_6}{2} = 0{,}9\frac{x_5 + x_6}{2} = 0{,}9\delta_x = 252$$

Γ. Οι τιμές με την επιβάρυνση είναι οι $w_i = y_i + 10$, $i = 1,\ldots,10$. Επομένως έχουμε:

$$\bar{w} = \bar{y} + 10 = 281.8 \text{ και } \delta_w = \frac{w_5 + w_6}{2} = \frac{y_5 + 10 + y_6 + 10}{2} = \frac{y_5 + y_6}{2} + 10 = \delta_y + 10 = 262$$

Εφαρμογή 29: Α. Να γράψετε στο τετράδιό σας τον πίνακα των τιμών της μεταβλητής Χ σωστά συμπληρωμένο.

Τιμές Μεταβλητής X_i	Συχνό-τητα v_i	Σχετική Συχνό-τητα φ_i	Σχετική Συχνό-τητα $\varphi_i\%$	Αθροιστική Συχνότητα Φ_i	$x_i v_i$	x_i^2	$x_i^2 v_i$
1	10				10	1	10
2				35		4	
3						9	
ΣΥΝΟΛΟ	v=50	1	100				

Β. Να υπολογίσετε τη μέση τιμή και τη διάμεσο.

Γ. Να υπολογιστεί η διακύμανση s^2

Απάντηση

Α.

Τιμές Μεταβλητής X_i	Συχνό-τητα v_i	Σχετική Συχνό-τητα φ_i	Σχετική Συχνό-τητα $\varphi_i\%$	Αθροιστική Συχνότητα Φ_i	$x_i v_i$	x_i^2	$x_i^2 v_i$
1	10	0.2	20	10	10	1	10
2	25	0.50	50	35	50	4	100
3	15	0.3	30	50	45	9	135
ΣΥΝΟΛΟ	v=50	1	100		105		245

Β. Μέση τιμή $\bar{x} = \frac{\sum_{1}^{3} x_i v_i}{50} = \frac{105}{50} = 2.1$

Διάμεσος

Εντοπίζουμε τη θέση του Μ

$$\frac{n}{2}=\frac{50}{2}=25$$

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

Γ. $$s^2=\frac{1}{\nu}\left\{\sum_{i=1}^{k}x_i^2\nu_i-\frac{\left(\sum_{i=1}^{k}x_i\nu_i\right)^2}{\nu}\right\}=\frac{1}{50}(245-\frac{105^2}{50})=\frac{1}{50}(245-220.5)=0.49$$

Εφαρμογή 30: Ένα προϊόν πωλείται σε 10 διαφορετικά καταστήματα στις παρακάτω τιμές, σε Ευρώ:

8, 10, 13, 13, 15, 16, 18, 14, 14, 9.

α. Να υπολογίσετε τη μέση τιμή, τη διάμεσο και την επικρατούσα τιμή.

β. Να υπολογίσετε το εύρος, την τυπική απόκλιση και τον συντελεστή μεταβολής.

γ. Αν οι τιμές του προϊόντος σε όλα τα καταστήματα υποστούν έκπτωση 10%, να εξετάσετε αν θα μεταβληθεί ο συντελεστής μεταβολής.

Απάντηση

Α. $$\bar{x}=\frac{\sum_{1}^{\nu}x_i}{\nu}=\frac{130}{10}=13$$

$$\delta=\frac{13+14}{2}=13,5$$

Όσον αφορά την επικρατούσα τιμή υπάρχουν δύο, η 13 και η 14.

Β. R = 18 – 8 = 10

$$S^2 = \frac{1}{\nu}(\sum_{i=1}^{\nu} x_i - \bar{x})^2 = \frac{1}{10}[(8-13)^2 + (9-13)^2 + ... + (8-13)^2] = \frac{1}{10} \cdot 90 = 9$$

Επομένως S =3 και $CV = \frac{S}{\bar{x}} = \frac{3}{13} = 23.08\%$

Γ. Αν $x_1, x_2, ..., x_{10}$ οι αρχικές τιμές και $y_1, y_2, ..., y_{10}$ οι τιμές που προκύπτουν μετά την έκπτωση των αντίστοιχων τιμών κατά 10%. Σε αυτή την περίπτωση νέα μέση τιμή $\bar{y}$ θα μειωθεί κατά 10% δηλαδή

$$\bar{y} = \bar{x} - 0.1\bar{x} = 0.9\bar{x}$$

Η νέα τυπική απόκλιση θα είναι $S_y = |0.9| \cdot S_x$

$$CV_y = \frac{S_y}{\bar{y}} = \frac{0.9S_x}{0.9\bar{x}} = \frac{S_x}{\bar{x}} = CV_x$$

Άρα ο συντελεστής μεταβλητότητας δεν μεταβάλλεται.

Εφαρμογή 31: Δίνονται οι ετήσιες δαπάνες (σε Ευρώ) για την αγορά βιβλίων γενικού ενδιαφέροντος, δύο φοιτητών ενός Πανεπιστημιακού Τμήματος κατά τα 5 τελευταία χρόνια:

Έτος 2005 2006 2007 2008 2009

Ετήσιες Δαπάνες Φοιτητή Α(σε Ευρώ) 250 200 300 350 300

Ετήσιες Δαπάνες Φοιτητή Β (σε Ευρώ) 5000 5150 5100 5050 5100

α. Να υπολογιστούν οι τυπικές αποκλίσεις των ετήσιων δαπανών των δύο φοιτητών. Μπορούμε να ισχυριστούμε ότι, εφόσον οι τυπικές αποκλίσεις είναι ίσες, οι ετήσιες δαπάνες των δύο φοιτητών παρουσιάζουν την ίδια μεταβλητότητα; Αιτιολογείστε την απάντησή σας.

β. Να συγκριθεί η μεταβλητότητα των ετήσιων δαπανών των δύο φοιτητών και να εξαχθούν τα σχετικά συμπεράσματα.

Απάντηση

A. $\bar{x}_A = \frac{\sum_1^{v} x_i}{v} = \frac{1400}{5} = 280$

$\bar{x}_B = \frac{\sum_1^{v} x_i}{v} = \frac{25400}{5} = 5080$

$S_A^2 = \frac{1}{v}(\sum_{i=1}^{v} x_i - \bar{x})^2 =$

$= \frac{1}{5}[(250-280)^2 + (200-280)^2 + (300-280)^2$

$+(350-280)^2 + (300-280)^2] = \frac{1}{5} \cdot 13000 = 2600$

$S_A = 50.99$

$S_B^2 = \frac{1}{v}(\sum_{i=1}^{v} x_i - \bar{x})^2 =$

$= \frac{1}{5}[(5000-5080)^2 + (5150-5080)^2 + (5100-5080)^2$

$+(5050-5080)^2 + (5100-5080)^2] = \frac{1}{5} \cdot 13000 = 2600$

$S_B = 50.99$

Παρατηρούμε ότι οι δύο φοιτητές έχουν τις ίδιες τυπικές αποκλίσεις αλλά διαφέρον στις μέσες τιμές τους. Δεν μπορούμε να ισχυριστούμε ότι εφόσον οι τυπικές αποκλίσεις είναι ίσες, οι ετήσιες δαπάνες των δύο φοιτητών παρουσιάζουν την ίδια μεταβλητότητα καθώς οι τυπικές αποκλίσεις των δαπανών αναφέρονται σε διαφορετικούς μέσους όρους δαπανών.

Β. $CV_A = \frac{S_A}{\bar{y}_A} = \frac{50.99}{280} = 18.21\%$

$CV_B = \frac{S_B}{\bar{y}_A} = \frac{50.99}{5080} = 1\%$

Είναι προφανές από τα παραπάνω ότι υπάρχει πολύ μεγάλη μεταβλητότητα στις δαπάνες του φοιτητή Α ενώ για τον φοιτητή Β η μεταβλητότητα των δαπανών του είναι εξαιρετικά μικρή.

Εφαρμογή 32: Ο υπεύθυνος ποιότητας ενός οργανισμού παροχής υπηρεσιών στο πλαίσιο της πολιτικής βελτίωσης της ποιότητας υπηρεσιών θέλει να ελέγξει το χρόνο που δαπανούν οι πελάτες για τη διεκπεραίωση των υποθέσεών τους. Σε μία έρευνα καταγράφηκε ο χρόνος εξυπηρέτησης (σε λεπτά) ενός δείγματος 50 πελατών από τη στιγμή που εισέρχονται στον οργανισμό μέχρι τη διεκπεραίωση της αίτησής τους. Ο παρακάτω πίνακας περιέχει τους χρόνους εξυπηρέτησης των 50 πελατών.

25 31 38 28 38 41 69 23 24 51
31 35 32 57 36 31 33 42 27 24
37 36 35 23 27 35 31 32 28 26
37 47 31 25 31 33 54 69 34 29
35 33 38 59 68 44 33 29 32 65

α. Να υπολογιστούν ο αριθμητικός μέσος, η επικρατούσα τιμή, η διάμεσος, το εύρος, το ενδοτεταρτημοριακό εύρος καθώς και η διασπορά του χρόνου διεκπεραίωσης

β. Να κατασκευασθεί το Θηκόγραμμα του χρόνου διεκπεραίωσης και με βάση αυτό να εξαχθούν συμπεράσματα για τη μορφή της κατανομής του.

γ. Να κατασκευασθεί Πίνακας Κατανομής Συχνοτήτων για το χρόνο διεκπεραίωσης χρησιμοποιώντας τάξεις εύρους 10, με κάτω όριο της πρώτης τάξης τα 20 λεπτά.

δ. Χρησιμοποιώντας τα ταξινομημένα δεδομένα να υπολογιστούν ο αριθμητικός μέσος, η επικρατούσα τιμή, το τρίτο τεταρτημόριο και η διακύμανση του χρόνου διεκπεραίωσης. **ε.** Χρησιμοποιώντας ένα μοντέλο βελτίωσης της ποιότητας των παρεχόμενων υπηρεσιών ο υπεύθυνος του οργανισμού πιστεύει ότι ο χρόνος εξυπηρέτησης του κάθε πελάτη μπορεί να μειωθεί κατά πέντε λεπτά μέσα σε χρονικό διάστημα έξι μηνών. Να υπολογισθεί ο αριθμητικός μέσος, η τυπική απόκλιση, το εύρος και η διακύμανση του χρόνου διεκπεραίωσης μετά από έξι μήνες.

στ. Σε ποια από τις δύο περιπτώσεις του ερωτήματος (ε) παρατηρείται μεγαλύτερη μεταβλητότητα στο χρόνο εξυπηρέτησης;

Απάντηση

α) Για τον υπολογισμό των μεγεθών που ζητούνται δημιουργήθηκε αρχικά ο παρακάτω πίνακας

Αύξων αριθμός	**Χρόνος Xi (σε αύξουσα σειρά)**	**X^2_i**	**$(X_i\text{-}X_\mu)^3$**
1	23	529	-2767.5873
2	23	529	-2767.5873
3	24	576	-2217.3425
4	24	576	-2217.3425
5	25	625	-1745.3377
6	25	625	-1745.3377
7	26	676	-1345.5729
8	27	729	-1012.0481
9	27	729	-1012.0481
10	28	784	-738.76326
11	28	784	-738.76326
12	29	841	-519.71846
13	29	841	-519.71846
14	31	961	-220.34886
15	31	961	-220.34886
16	31	961	-220.34886
17	31	961	-220.34886
18	31	961	-220.34886
19	31	961	-220.34886
20	32	1024	-128.02406
21	32	1024	-128.02406
22	32	1024	-128.02406
23	33	1089	-65.939264
24	33	1089	-65.939264
25	33	1089	-65.939264
26	33	1089	-65.939264
27	34	1156	-28.094464
28	35	1225	-8.489664
29	35	1225	-8.489664
30	35	1225	-8.489664
31	35	1225	-8.489664
32	36	1296	-1.124864
33	36	1296	-1.124864
34	37	1369	-6.4E-05

35	37	1369	-6.4E-05
36	38	1444	0.884736
37	38	1444	0.884736
38	38	1444	0.884736
39	41	1681	62.099136
40	42	1764	122.023936
41	44	1936	337.153536
42	47	2209	988.047936
43	51	2601	2720.54714
44	54	2916	4878.40154
45	57	3249	7952.09594
46	59	3481	10590.0255
47	65	4225	21858.0543
48	68	4624	29675.8287
49	69	4761	32645.2735
50	69	4761	32645.2735
Σύνολο	1852	75964	123096.09

Στη συνέχεια περιγράφονται αναλυτικά οι υπολογισμοί :

Αριθμητικός μέσος

$$\bar{x} = \frac{\sum_{1}^{50} x_i}{10} = \frac{1852}{50} = 37.04 \quad \text{min}$$

Επικρατούσα τιμή M_o

Είναι γνωστό ότι ορίζεται ως η τιμή με τη μεγαλύτερη συχνότητα εμφάνισης. Στην προκειμένη περίπτωση ισούται με την τιμή 31 η οποία εμφανίζεται 6 φορές, δηλαδή M_o =31 min.

Διάμεσος

Πλήθος Δεδομένων n = 50, άρτιος αριθμός. Η διάμεσος είναι η τιμή στη θέση $\frac{n+1}{2}=\frac{50+1}{2}=25.5$. Άρα

$$M=\frac{1}{2}(X_{\frac{n}{2}}+X_{\frac{n}{2}+1})=\frac{1}{2}(X_{\frac{50}{2}}+X_{\frac{50}{2}+1})=\frac{1}{2}(X_{25}+X_{26})=\frac{33+33}{2}=33\,\text{min}$$

Εύρος R

Το εύρος δίνεται από τη διαφορά της μικρότερης παρατήρησης από τη μεγαλύτερη, δηλαδή $R=X_{\max}-X_{\min}=69-23=46\,\text{min}$

Ενδοτεταρτημοριακό εύρος

Το ενδοτεταρτημοριακό εύρος δίνεται από τον τύπο $IR=Q_3-Q_1$

Όπου: $Q_{3,}Q_1$: το πρώτο και τρίτο αντίστοιχα τεταρτημόριο. Επομένως για τον υπολογισμός του ενδοτεταρτημοριακού εύρους απαιτείται ο υπολογισμός των δύο τεταρτημορίων.

Πρώτο τεταρτημόριο

Είναι η τιμή στη θέση $\frac{i(n+1)}{4}$ με i = 1, δηλαδή η $X_{\frac{1(n+1)}{2}}$ αλλιώς

$$Q_1=X_{\frac{1(n+1)}{2}}=X_{A_Q}+\Delta_Q(X_{A_Q+1}-X_{A_Q})$$

Όπου

AQ: Το ακέραιο μέρος του πηλίκου $\frac{1(n+1)}{2}$

ΔQ: Το δεκαδικό μέρος του πηλίκου $\frac{1(n+1)}{2}$

Επομένως για n = 50 προκύπτει: = 12.75. Άρα AQ = 12 και ΔQ = 0.75.

Συνεπώς

$$Q_1 = X_{\frac{1(n+1)}{2}} = X_{12} + 0.75(X_{13} - X_{12}) = 29 + 0.75(29-29) = 29\,\text{min}$$

Τρίτο τεταρτημόριο

Για το πρώτο τεταρτημόριο εφαρμόζουμε τους ίδιους τύπους για i = 3 επειδή επιπλέον n = 50 έχουμε: $\frac{i(n+1)}{4} = \frac{3(50+1)}{4} = \frac{51}{4} = 38.25$

Άρα AQ = 38 και ΔQ = 0.25

Συνεπώς

$$Q_3 = X_{\frac{3(n+1)}{2}} = X_{38} + 0.25(X_{39} - X_{38}) = 38 + 0.25(41-38) = 38.75\,\text{min}$$

Τελικά το ενδοτεταρτημοριακό εύρος υπολογίζεται

$IR = Q_3 - Q_1$ = 38,75-29 = 9,75 min

Διασπορά S^2

Η διασπορά ή αλλιώς διακύμανση υπολογίζεται από τον τύπο:

$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$ που είναι ισοδύναμος με τον $S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1}$, τον οποίο επιλέξαμε να χρησιμοποιήσουμε. Για τις ανάγκες των υπολογισμών δημιουργήθηκε η τρίτη στήλη που φαίνεται στον πίνακα που δόθηκε αρχικά. Αντικαθιστώντας τα αθροίσματα προκύπτει:

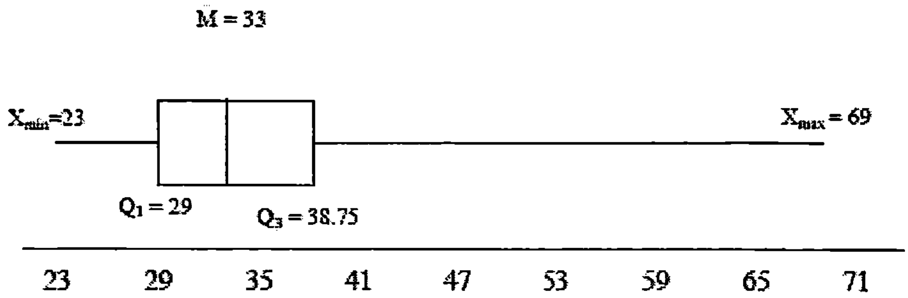
$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1} = \frac{75964 - 50 \cdot 37.04^2}{50-1} = 150.3249\,\text{min}^2$$

Στη συνέχεια παρατίθενται τα αποτελέσματα που βρέθηκαν συγκεντρωτικά:

	Με βάση τους τύπους	**Με βάση τις συναρτήσεις του Excel**
Αριθμητικός μέσος (X_μ)	37.04	37.04
Επικρατούσα τιμή (T_0)	31	31
Διάμεσος (M)	33	33
Εύρος (R)	46	46
Πρώτο τεταρτημόριο (Q_1)	29	29.5
Τρίτο τεταρτημόριο (Q_3)	38.75	38
Ενδοτεταρτημοριακό εύρος (IR)	9.75	8.5
Διασπορά (S^2)	150.324898	150.324898

β) Από το ερώτημα (α) έχουμε ήδη υπολογίσει όλα τα απαραίτητα για την κατασκευή του θηκογράμματος στοιχεία:

X_{min}=23, X_{max}=69, M=33, $Q_1 = 29$, $Q_3 = 38.75$.



Από το Θηκόγραμμα παρατηρούμε ότι:

14Q–= (δηλαδή το ενδιάμεσο διαχωριστικό τμήμα στο εσωτερικό του ορθογώνιου πλαισίου βρίσκεται πλησιέστερα προς το αριστερό άκρο του).

= (δηλαδή το μήκος της δεξιάς απόληξης είναι μεγαλύτερο από το αντίστοιχο της αριστερής)

Κατά συνέπεια συμπεραίνουμε ότι η κατανομή των δεδομένων παρουσιάζει συγκέντρωση στο αριστερό κομμάτι της

γ) Εφόσον το κάτω όριο της πρώτης κλάσης είναι το 20 και το πλάτος της 10 τότε αυτή είναι η [20 , 30). Για να συμπεριλάβουμε και την τελευταία παρατήρηση (69) χρειάζεται να δημιουργήσουμε συνολικά 5 ισοπλατείς κλάσεις. Οπότε έχουμε τις παρακάτω τάξεις και τον αντίστοιχο πίνακα κατανομής συχνοτήτων:

Τάξη	φ_i	χ_i	Φ_i	$\varphi_i\chi_i$	χ_i^2	$\varphi_i\chi_i^2$
20-<30	13	25	13	325	625	8125
30-<40	25	35	38	875	1225	30625
40-<50	4	45	42	180	2025	8100
50-<60	4	55	46	220	3025	12100
60-<70	4	65	50	260	4225	16900
Σύνολο	**50**	-	-	**1860**	**11125**	**75850**

Οι τρεις τελευταίες στήλες δημιουργήθηκαν για τις ανάγκες του ερωτήματος (δ)

δ) Από τα αποτελέσματα του προηγούμενου ερωτήματος υπολογίζουμε τα επόμενα μεγέθη:

Αριθμητικός μέσος $\bar{X}$

$$\bar{X} = \frac{\sum_{i=1}^{k} \phi_i \cdot \chi_i}{\sum_{i=1}^{k} \phi_i \cdot \chi_i} = \frac{\sum_{i=1}^{5} \phi_i \cdot \chi_i}{\sum_{i=1}^{5} \phi_i \cdot \chi_i} = \frac{1860}{50} = 37.2$$

Επικρατούσα τιμή M_0

Η κλάση με τη μεγαλύτερη συχνότητα είναι η δεύτερη επομένως

Το πλάτος της τάξης είναι δ = 10.

$\Delta_1 = 25 - 13 = 12$ και $\Delta_2 = 25 - 4 = 21$.

Αντικαθιστώντας τις τιμές βρίσκουμε:

$$M_o = X_\lambda + \delta \frac{\Delta_1}{\Delta_1 + \Delta_2} = 30 + 10 \frac{12}{12+21} = 33,636 \min$$

Τρίτο τεταρτημόριο Q_3

Το τρίτο τεταρτημόριο Q_3 βρίσκεται από τον τύπο:

$$Q_3 = X_\lambda + \frac{\delta}{v_i}(\frac{3N}{4} - \Phi_i)$$

Στη συγκεκριμένη περίπτωση έχουμε:

$$\frac{3n}{4} = \frac{3 \cdot 50}{4} = 37.5$$

Άρα το τρίτο τεταρτημόριο βρίσκεται στην $2^η$ τάξη. Οπότε:

$$Q_3 = X_\lambda + \frac{\delta}{v_i}(\frac{3N}{4} - \Phi_i)] = 30 + 10\frac{37.5 - 13}{25} = 135.878$$

Διακύμανση S^2

Η διακύμανση στην περίπτωση ομαδοποιημένων δεδομένων βρίσκεται από τον τύπο:

$$S^2 = \frac{\sum_{i=1}^{k} \phi_i (\chi_i - \bar{X})^2}{\sum_{i=1}^{k} \phi_i - 1}$$ ή από τον ισοδύναμο $$S^2 = \frac{\sum_{i=1}^{k} \phi_i \chi_i^2 - n\bar{X}^2}{\sum_{i=1}^{k} \phi_i - 1}$$

Εδώ ο υπολογισμός θα γίνει με βάση τον δεύτερο. Όλα τα απαραίτητα αθροίσματα έχουν ήδη υπολογιστεί. Επομένως:

$$S^2 = \frac{\sum_{i=1}^{k} f_i m_i^2 - n\bar{X}^2}{\sum_{i=1}^{k} f_i - 1} = \frac{75850 - 50 \cdot 37.2^2}{50 - 1} = 135,8776$$

Παρακάτω παραθέτουμε σε πίνακα τα στατιστικά μέτρα που υπολογίσαμε

Αριθμητικός Μέσος (Xμ)	37.2
Επικρατούσα Τιμή (M)	33.636
Τρίτο τεταρτημόριο (Q3)	39.800
Διασπορά (S2)	135.878

ε) Αν Xi είναι οι αρχικές μετρήσεις τότε οι μετασχηματισμένες μετρήσεις είναι:

Χρησιμοποιώντας τις ιδιότητες των στατιστικών μέτρων και τα αποτελέσματα του ερωτήματος (α) προκύπτουν:

Αριθμητικός μέσος $\bar{Y}$

$\bar{Y} = \bar{X} - 5 = 37.04 - 5 = 32.04\,\text{min}$

Τυπική απόκλιση S_Y

$S_Y = S_X = 12.2614\,\text{min}$

Εύρος R_Y

$R_Y = R_X = 46\,\text{min}$

Διακύμανση

$S_Y^2 = S_X^2 = 150.3249\,\text{min}^2$

στ) Προκειμένου να εξετάσουμε σε ποια από τις δύο περιπτώσεις του ερωτήματος (ε) παρατηρείται μεγαλύτερη μεταβλητότητα στο χρόνο εξυπηρέτησης θα υπολογίσουμε τους συντελεστές μεταβλητότητας για τις δύο περιπτώσεις.

$$CV_Y = \frac{S_Y}{\bar{Y}}100 = \frac{12.261}{32.04}100 = 38.27$$

$$CV_X = \frac{S_X}{\bar{X}}100 = \frac{12.261}{37.04}100 = 33.10$$

Όπως διαπιστώνεται στην δεύτερη περίπτωση ο συντελεστής μεταβλητότητας είναι μεγαλύτερος και κατά συνέπεια σημειώνεται μεγαλύτερη μεταβλητότητα.

Εφαρμογή 33: Μια βιομηχανία κάνει συσκευασία γάλακτος σε 4 μεγέθη κουτιών και σε

ποσοστά: 10%, 20%, 30% 40% με αντίστοιχο κόστος συσκευασίας 8, 6, 4, 2 ευρώ ανά κουτί.

(α΄) Να βρεθεί το μέσο κόστος συσκευασίας και η τυπική απόκλιση του κόστους αυτού.

(β΄) Αν το κόστος κάθε συσκευασίας αυξηθεί κατά 10% να βρεθεί εάν αλλάζει η τυπική απόκλιση του κόστους συσκευασίας.

Απάντηση

(α΄) Από τις συχνότητες επί τοις % βρίσκουμε ότι οι σχετικές συχνότητες

είναι 0,1, 0,2, 0,3, 0,4. Από την σχέση $\bar{x} = \sum_{i=1}^{k} x_i f_i$ βρίσκουμε ότι $\bar{x} = 4$

Για τον υπολογισμό της τυπικής απόκλισης χρησιμοποιούμε τον τύπο

$$s^2 = \frac{1}{v}\sum_{i=1}^{k}(x_i - \bar{x})^2 v_i$$

από τον οποίο έχουμε

$s^2 = 0.1(8-4)^2 + 0.2(6-4)^2 + 0.3(4-4)^2 + 0.4(2-4)^2 = 4$

$s = 2$

(β΄) Κάθε τιμή αυξάνεται κατά 10% επομένως πολλαπλασιάζεται επί 1,1.

Άρα και η τυπική απόκλιση θα πολλαπλασιασθεί επί 1,1 και θα γίνει 2,2.

Εφαρμογή 34: Θεωρούμε δύο δείγματα Α και Β με παρατηρήσεις:

Δείγμα Α: 12, 18, t_3, t_4, … , t_{25}

Δείγμα Β: 16, 14, t_3, t_4, … , t_{25}.

Δίνεται ότι $t_3+t_4+\ldots+t_{25}=345$.

α. Να αποδείξετε ότι οι μέσες τιμές των δύο δειγμάτων Α και Β αντίστοιχα είναι 15.

β. Έστω s_A^2 η διακύμανση του δείγματος Α και s_B^2 η διακύμανση του δείγματος Β, με $s_A^2 - s_B^2 = \frac{16}{25}$. Αν ο συντελεστής μεταβολής του δείγματος Α είναι ίσος με 1/15, να βρείτε τον συντελεστή μεταβολής του δείγματος Β.

Απάντηση

Α) $\bar{x}_A = \frac{12+18+t_3+\ldots+t_{25}}{25} = \frac{30+345}{25} = 15$

$$\bar{x}_B = \frac{16+14+t_3+\ldots+t_{25}}{25} = \frac{30+345}{25} = 15$$

Β) $s_A^2 = \frac{\sum_{i=1}^{25}(\bar{x}_A - A_i)^2}{25}$

$$s_B^2 = \frac{\sum_{i=1}^{25}(\bar{x}_B - B_i)^2}{25}$$

Όπου A_i, B_i οι παρατηρήσεις για το δείγμα Α και Β αντίστοιχα. Τότε

$$CV_A = \frac{s_A}{\bar{x}_A} \Rightarrow s_A = CV_A \bar{x}_A = 1$$

$$s_A^2 - s_B^2 = \frac{16}{25} \Rightarrow s_B^2 = 1 - \frac{16}{25} = \frac{9}{25}$$

$$s_B = \frac{3}{5}$$

$$CV_B = \frac{s_B}{\bar{x}_B} = \frac{\frac{3}{5}}{15} = \frac{1}{25}$$

Εφαρμογή 35: Ο παρακάτω πίνακας δίνει την κατανομή συχνότητας των μισθών τριάντα υπαλλήλων μιας δημόσιας υπηρεσίας:

Μισθός (€)	**Αριθμός Υπαλλήλων**
600 – 700	7
700 – 800	14
800 – 900	5
900 – 1000	3
1000 – 1100	1

Δίνεται επίσης ότι η τυπική απόκλιση των μισθών τους είναι 102,5€.

Με βάση τα δεδομένα αυτά να υπολογιστούν:

α. Ο αριθμητικός μέσος.

β. Το ενδοτεταρτημοριακό εύρος των μισθών.

Απάντηση

Υπολογισμός βοηθητικών στοιχείων

Μισθός (€)	**Κεντρική Τιμή** x_i	**Αριθμός Υπαλλήλων φι**	**Αθροιστική Συχνότητα Φι**	$\varphi_i x_i$
600 – 700	650	7	7	4550
700 – 800	750	14	21	10500
800 – 900	850	5	26	4250
900 – 1000	950	3	29	2850
1000 – 1100	1050	1	30	1050
ΣΥΝΟΛΑ		30		**23200**

α. Αριθμητικός Μέσος

$$\overline{X} = \frac{\sum_{i=1}^{k} \phi_i \chi_i}{n} = \frac{23200}{30} = 773,333 \rightarrow \overline{X} \cong 773,3€$$

β. Ενδοτεταρτημοριακό Εύρος

$I = Q_3 - Q_1$

Θα πρέπει να υπολογιστούν τα Q_1, Q_3.

Εντοπισμός της θέσης του Q_1:

$$\frac{n*i}{4} = \frac{30*1}{4} = \frac{30}{4} = 7,5$$

Άρα το Q_1 ανήκει στην $2^η$ τάξη (διάστημα 700 - < 800)

Υπολογισμός της τιμής του Q_1:

$$Q_1 = X_\lambda + \frac{\delta}{v_i}(\frac{N}{4} - \Phi_i) = 700 + 100\frac{(7,5-7)}{14} = 700 + 100\frac{0.5}{14} = 700 + 100 \cdot (0,036) =$$

$$= 700 + 3,6 \rightarrow Q_1 \cong 703,6€$$

Εντοπισμός της θέσης του Q_3:

$$\frac{n*i}{4} = \frac{30*3}{4} = \frac{90}{4} = 22,5$$

Άρα το Q_3 ανήκει στην $3^η$ τάξη (διάστημα 800 - < 900)

Υπολογισμός της τιμής του Q_3:

$$Q_3 = X_\lambda + \frac{\delta}{v_i}(\frac{3N}{4} - \Phi_i) = 800 + 100\frac{(22,5-21)}{5} = 800 + 100\frac{1,5}{5} = 800 + 100 \cdot (0,3) =$$

$$= 800 + 30 \rightarrow Q_1 = 830€$$

Συνεπώς το Ενδοτεταρτημοριακό Εύρος είναι:

$IR = Q_3 - Q_1 = 830 - 703,6 = 126,4 \Rightarrow I = 126,4$

Εφαρμογή 36: Οι αρχικές ετήσιες αποδοχές (σε χρηματικές μονάδες) των 12 πρώτων αποφοίτων του Τμήματος Διοίκησης Επιχειρήσεων ενός Πανεπιστημίου είναι οι ακόλουθες:

2050	2150	2250	2080	1955	1910	2090	2330	2140	2525	2120	2080

Να κατασκευαστεί το Θηκόγραμμα των παραπάνω δεδομένων και με βάση αυτό να εξαχθούν συμπεράσματα για τη μορφή της κατανομής τους.

Απάντηση

Κατατάσσω τα δεδομένα σε αύξουσα τάξη:

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
1910	1955	2050	2080	2080	2090	2120	2140	2150	2250	2330	2525

(α) Για να κατασκευαστεί το Θηκόγραμμα χρειάζεται να υπολογιστούν τα ακόλουθα στοιχεία:

X_{min}, X_{max}, M, Q_1, Q_3.

$X_{min} = 1910$

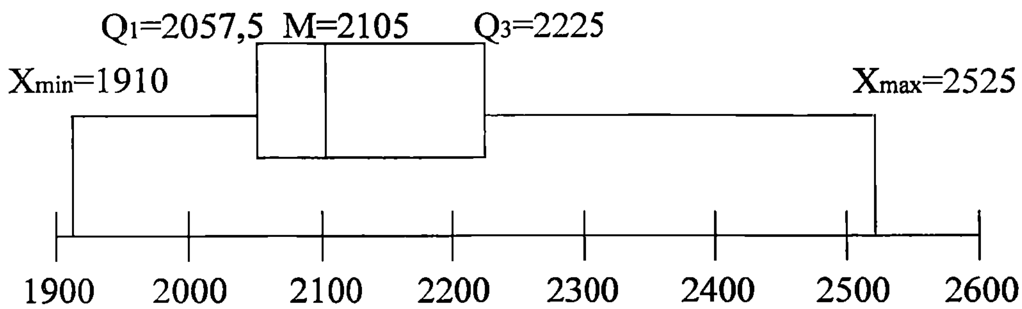
$X_{max} = 2525$

$$M = X_{\frac{n+1}{2}} = X_{\frac{12+1}{2}} = X_{\frac{13}{2}} = X_{6,5} = X_6 + 0,5*(X_7 - X_6) = 2090 + 0,5*(2120 - 2090) = 2105$$

$$Q_1 = X_{\frac{n+1}{4}} = X_{\frac{12+1}{4}} = X_{\frac{13}{4}} = X_{3,25} = X_3 + 0,25 * (X_4 - X_3) = 2050 + 0,25 * (2080 - 2050) = 2057,5$$

$$Q_3 = X_{\frac{3(n+1)}{4}} = X_{\frac{3*(12+1)}{4}} = X_{\frac{39}{4}} = X_{9,75} = X_9 + 0,75 * (X_{10} - X_9) = 2150 + 0,75 * (2250 - 2150) = 2225$$

Με βάση τα στοιχεία αυτά κατασκευάζουμε το Θηκόγραμμα των δεδομένων.



(β) Από το Θηκόγραμμα παρατηρούμε ότι

$Q_3 - M > M - Q_1$

(δηλαδή ότι το διαχωριστικό ενδιάμεσο τμήμα στο εσωτερικό του ορθογώνιου πλαισίου βρίσκεται πλησιέστερα προς το αριστερό άκρο του).

$X_{max} - Q_3 > Q_1 - X_{min}$

(δηλαδή το μήκος της δεξιάς απόληξης είναι μεγαλύτερο από το αντίστοιχο της αριστερής)

Κατά συνέπεια συμπεραίνουμε ότι η κατανομή των δεδομένων παρουσιάζει θετική ασυμμετρία.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Από την παραπάνω έρευνα προέκυψαν τα εξής:

- Η μέση τιμή είναι ένα μέτρο θέσης και χρησιμοποιείται για να μας δείξει που βρίσκεται κατά μέσο όρο μια ποσότητα
- Η διάμεσος, τα τεταρτημόρια, τα δεκατημόρια, τα εκατοστημόρια είναι μέτρα θέσης που χρησιμοποιούνται για να μας δείξουν που κινούνται συγκεκριμένα κομμάτια του δείγματος. Για παράδειγμα μια επιχείρηση μπορεί να θέλει να δει τι είδους προϊόντα αγοράζει το 10% των καλύτερων πελατών της. Θα χρησιμοποιήσει για αυτό τον λόγο τα δεκατημόρια.
- Η τυπική απόκλιση είναι ένα μέτρο μεταβλητότητας ως προς την μέση τιμή που χρησιμοποιείται για να δείξει αν ένα μέγεθος μεταβάλλεται πολύ ή λίγο. Μια χαλυβουργική εταιρεία που κατασκευάζει βίδες θα ήθελε η διάμετρος τους να είναι όσο το δυνατόν πιο σταθερή κατά την διάρκεια κατασκευής τους.
- Η διαφορά κατά Gini είναι και αυτό ένα μέτρο μεταβλητότητας όπως η τυπική απόκλιση αλλά δεν είναι ως προς την μέση τιμή αλλά ως προς όλες τις άλλες παρατηρήσεις.
- Ο συντελεστής μεταβλητότητας είναι και αυτό ένα μέτρο μεταβλητότητας ο οποίος δείχνει την ομοιογένεια ενός μεγέθους. Για παράδειγμα μια εταιρεία θέλει η παραγωγή της να είναι όσο το δυνατόν ομοιογενής.
- Το εύρος που είναι ένα μέτρο μεταβλητότητας δείχνει μεταξύ ποιου διαστήματος μπορούν να κινηθούν οι μετρήσεις ενός μεγέθους. Για παράδειγμα μια εταιρεία θέλει να ξέρει το εύρος αντοχής ενός υλικού που κατασκευάζει

- Το ενδοτεταρτημοριακό εύρος είναι ένα μέτρο μεταβλητότητας δείχνει μεταξύ ποιου διαστήματος μπορούν να κινηθούν οι μεσαίες 50% μετρήσεις ενός μεγέθους.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Θεοδώρου Η . Απολόπουλου, ¨Περιγραφική Στατιστική Επιχειρήσεων¨, Σύγχρονη Εκδοτική, Αθήνα 2003.

Χριστοδουλόπουλου Δ.,(2006), Λήψη Στρατηγικών επιχειρηματικών αποφάσεων και Συστήματα υποστήριξης Αποφάσεων, Οικονομικό Πανεπιστήμιο Αθηνών, σελ 1-8

Αποστολοπούλου, Θ. ¨Στατιστκή Επιχειρήσεων¨Ι, Αθήνα 1982

Κάκουλλου Θ.Ν, ¨Θεωρία Πιθανοτήτων¨, Αθήνα 1970

Αγγελοπούλου, Ε. ¨Φροντιστηριακές Σημειώσεις Στατιστικής¨, 6 τεύχη, Αθήνα 1977

Γιαννοπούλου, Θ. ¨Μαθηματικά ΙΙΙ¨(Μαθηματική Στατιστική), Εκδ. ΟΕΔΒ, Αθήνα, 1981

Θεοδωράκη, Γ.Α. ¨Μαθήματα Στατιστικής Ι¨ Αθήνα 1970

Κούρκουλου, Α.Θ. ¨Στατιστική Επιχειρήσεων¨, Λάρισα 1986

Facione, P. and Facione, N., (2007),Thinking and Reasoning in Human Decision Making, The California Academic Press / Insight Assessment,pp.34-38

Plous, S.,(2003), The Psychology of Judgment and Decision Making New York: McGraw-Hill