ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΜΕΣΟΛΟΓΓΙΟΥ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗ & ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ

Π Τ Υ Χ Ι Α Κ Η Ε Ρ Γ Α Σ Ι Α

ΘΕΜΑ: Στατιστική και ο ρόλος της στις επιχειρήσεις

ΜΠΑΛΑΦΑ ΒΑΣ. ΑΙΚΑΤΕΡΙΝΗ

ΠΙΤΣΑ ΒΑΣ. ΠΑΝΑΓΙΩΤΑ

ΕΙΣΗΓΗΤΗΣ

ΜΕΓΑΡΙΤΗΣ ΑΘΑΝΑΣΙΟΣ

Μ Ε Σ Ο Λ Ο Γ Γ Ι 2 0 1 3

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΜΕΣΟΛΟΓΓΙΟΥ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ & ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ

Π Τ Υ Χ Ι Α Κ Η Ε Ρ Γ Α Σ Ι Α

Στατιστική και ο ρόλος της στις επιχειρήσεις

ΜΠΑΛΑΦΑ ΒΑΣ. ΑΙΚΑΤΕΡΙΝΗ (Α.Μ. 15048)

aikabala@logistiki.teimes.gr

ΠΙΤΣΑ ΒΑΣ. ΠΑΝΑΓΙΩΤΑ (Α.Μ. 15094)

panapits@logistiki.teimes.gr

ΕΙΣΗΓΗΤΗΣ

ΜΕΓΑΡΙΤΗΣ ΑΘΑΝΑΣΙΟΣ

Μ Ε Σ Ο Λ Ο Γ Γ Ι 2 0 1 3

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Τ.Ε.Ι. ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ & ΟΙΚΟΝΟΜΙΑΣ (ΤΕΙΜ)
ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ & ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗΣ

ΕΙΣΗΓΗΤΙΚΗ ΕΚΘΕΣΗ ΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

υποβάλλεται ενσωματωμένη σε κάθε αντίτυπο της Πτυχιακής Εργασίας

ΤΙΤΛΟΣ ΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ *Ακριβής καταχώρηση του τίτλου του θέματος της Πτυχιακής*

ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ Ο ΡΟΛΟΣ ΤΗΣ ΣΤΙΣ ΕΠΙΧΕΙΡΗΣΕΙΣ

ΦΟΙΤΗΤΕΣ

ΕΠΩΝΥΜΟ	ΟΝΟΜΑ	ΑΡ. ΜΗΤΡ.
ΜΠΑΛΑΦΑ	ΑΙΚΑΤΕΡΙΝΗ	15048
ΠΙΤΣΑ	ΠΑΝΑΓΙΩΤΑ	15094

ΕΚΘΕΣΗ ΕΓΚΡΙΣΗΣ ΕΙΣΗΓΗΤΗ ΕΚΠΑΙΔΕΥΤΙΚΟΥ *καταχώρηση κειμένου από τον εισηγητή*

Η παρούσα πτυχιακή εργασία πραγματοποιήθηκε υπό την καθοδήγησή μου. Το θέμα της εργασίας είναι η Στατιστική και ο ρόλος της στις επιχειρήσεις. Δίνονται εφαρμογές των στατιστικών μέτρων και της απλής γραμμικής παλινδρόμησης στο χώρο των επιχειρήσεων. Κατά τη γνώμη μου, το θέμα αντιμετωπίστηκε επαρκώς, η δομή είναι ικανοποιητική και συνεπώς εισηγούμαι την έγκρισή της.

ΗΜΕΡΟΜΗΝΙΑ 13/9/2013

Απαραίτητη η υπογραφή του εισηγητή εκπαιδευτικού **ΥΠΟΓΡΑΦΗ** ....

Πίνακας περιεχομένων

Κεφάλαιο 2. Στατιστικά μέτρα

Κεφάλαιο 3. Παλινδρόμηση και συσχέτιση

Κεφάλαιο 4. Στατιστική και επιχειρήσεις

ΠΡΟΛΟΓΟΣ

Ο όρος “στατιστική” καθιερώθηκε απο τον καθηγητή του Πανεπιστημίου της Γοττίγγης Gottfried Achenwall (1719 – 1772) και παράγεται από τη λατινική λέξη “status” η οποία σημαίνει “κράτος”. Η σταστιστική θεωρείται μία επιστήμη η αποία ασχολείται με τη συλλογή,την επεξεργασία και την ανάλυση δεδομένων με στόχο τη λήψη ορθών αποφάσεων. Οι κυριότεροι μέθοδοι συλλογής στατιστικων δεδομένων είναι η απογραφή και η δειγματολοψία. Η απογραφή συνιστάται στη καταγραφή όλων των στοιχείων του πληθυσμου, ενώ η δειγματοληψία συνιστάται στη συλλογή στατιστικών δεδομένων μόνο από ένα τμήμα του στατιστικού πληθυσμού που θέλουμε να διερευνήσουμε. Για καλύτερη εξέταση των δεδομένων η στατιστική εφαρμόζει κάποιες παραμέτρους. Οι πάραμετροι αυτοί ονομάζονται,παράμετροι θέσεως, και παράμετροι διασποράς. Οι παράμετροι θέσεως είναι, η διάμεσος,τα τεταρτημόρια,τα δεκατημόρια,τα εκατοστημόρια και η επικρατούσα τιμή. Και οι παράμετροι διασποράς είναι το εύρος,το ημιενδοτεταρτημοριακό εύρος, η μέση απόκλιση, η τυπική απόκλιση, και ο συντελεστής μεταβλητικότητας. Σε περίπτωση που έχουμε διμετάβλητους ή πολυμετάβλητους στατιστικούς πληθυσμούς χρησιμοποιούμε τις μεθόδους παλινδρόμησης και συσχέτισης. Οι διαδικασίες αυτές ασχολούνται με τη μελέτη των συγκεκριμένων πληθυσμών και προσδιορίζουν τη σχέση εξάρτησης μεταξύ δύο μεταβλητών. Όσο αναφορά την εφαρμογή της στατιστικής στον επιχειρηματικό τομέα η στατιστική πραγματεύεται περισσότερο σε εμπορικές και βιομηχανικές επιχειρήσεις,και ο ρόλος της είναι να συγκεντρώνει το στατιστικό υλικό παρουσιάζοντας το σε πίνακες και σε διαγράμματα,και να υπολογίζει τους απαραίτητους στατιστικούς δείκτες. Οι κυριότερες στατιστικές δραστηριότητες μιας μεγάλης επιχείρησης είναι, το τμήμα Παραγωγής, το τμήμα Προσωπικού, το τμήμα Marketing, το τμήμα Οικονομικού ελέγχου, και το τμήμα Οικονομικών ή Στατιστικών Μελετών.

ΕΙΣΑΓΩΓΗ

Στατιστική είναι μία από τις σπουδαιότερες μεθόδους, η οποία καλείται να εξετάσει ένα μελετούμενο μέγεθος. Σε μια στατιστική έρευνα, διακρίνουμε τρία στάδια. Τη συλλογή του στατιστικού υλικού, την επεξεργασία και την ανάλυση του υλικού αυτού, και την εξαγωγή χρήσιμων συμπερασμάτων.

Στο πρώτο κεφάλαιο της εργασίας μας, αναλύουμε τον όρο “στατιστική”, τις βασικές έννοιες της στατιστικής περιγράφοντας λεπτομερώς για την έννοια του πληθυσμού, τους δείγματος, και της μεταβλητής. Στη συνέχεια ακολουθούν διάφοροι στατιστικοί πίνακες και γραφικές παραστάσεις, βάσει των οποίων καταλήγουμε στην εξακριβώση των αποτελέσματων μίας έρευνας.

Στο δεύτερο κεφάλαιο αναφέρουμε διάφορα στατιστικά μέτρα, τα οποία έχουν σκοπό τη συνοπτική παρουσίαση των δεδομένων για να διευκολυνθεί η μελέτη τους και να εξαχθούν χρήσιμα συμπεράσματα. Τα συμπεράσματα αυτά χρησιμεύουν για τη λήψη ορθών αποφάσεων. Τα μέτρα που θα περιγράψουμε είναι,τα μέτρα κεντρικής τάσης, τα μέτρα θέσεως και τα μέτρα διασποράς.

Ένα ξεχωριστό κομμάτι της εργασίας μας, το οποίο επεκτείνουμε στο τρίτο κεφάλαιο, είναι οι έννοιες της παλινδρόμησης και της συσχέτισης,οι οποίες θεωρούνται σχετικές μεταξύ τους, αφού και οι δύο διαδικασίες έχουν σκοπό τη μελέτη διμετάβλητων πληθυσμων. Ενώ η παλινδρόμηση προσδιορίζει τη σχέση μεταξύ εξάρτησης δύο μεταβλητών, ο συντελεστής γραμμικής συσχέτισης δίνει ένα μέτρο του μεγέθους της γραμμικής συσχέτισης μεταξύ δύο μεταβλητών.

Τέλος, στο τέταρτο κεφάλαιο της εργασίας μας σημειώνουμε το ρόλο της στατιστικής στις επιχειρήσεις, τονίζοντας πως η εφαρμογη στατιστικών μεθόδων αφορά περισσότερο βιομηχανικές και εμπορικές επιχειρήσεις. Για έναν επιχειρηματία ο οποίος γνωρίζει το προσωπικό του και έχει τη πλήρη επίβλεψη της επιχερήσης η εφαρμογή της στατιστικής δεν είναι απαραίτητη, όσο για ένα γενικό διευθυντή ο οποίος απασχολεί εκατοντάδες άτομα. Στο κλείσιμο του κεφαλαίου αυτού,εφαρμόζουμε διάφορα παραδείγματα με βάσει τις επιχειρήσεις αλλά πραγματοποιηούμε και πιο γενικές περιπτώσεις οσο αναφορά το κομμάτι της στατιστικής.

ΚΕΦΑΛΑΙΟ 1. ΣΤΑΤΙΣΤΙΚΗ

1.1 Ιστορική αναδρομή της στατιστικής

Ο όρος « στατιστική » προέρχεται από τη λατινική λέξη « status» που σημαίνει κράτος και δηλώνει αρχικά τη συλλογή στοιχείων για τις κρατικές ανάγκες (έκταση, πληθυσμός, παραγωγή κ.α.). Η αρχαιότερη συλλογή στατιστικών στοιχείων θεωρείται η απογραφή πληθυσμού που έγινε το 2238 Π.Χ. στην Κίνα από τον αυτοκράτορα Yao, τους Σίνες, τους Αιγύπτιους και τους Πέρσες. Επίσης ο όρος Στατιστική αναφέρεται και από το Σωκράτη (Ξενοφώντος ΄΄Απομνημονεύματα΄΄) και από τον Αριστοτέλη (΄΄Πολιτεία΄΄).

Η συγκέντρωση στατιστικών στοιχείων στην αρχαιότητα, είχε ως στόχο τον εντοπισμό των πολιτών να πληρώσουν φόρο ή να υπηρετήσουν ως πολεμιστές. Στην Ιταλία στις πόλεις Βενετία και Φλωρεντία κατά τη διάρκεια της Αναγέννησης άρχισε η συλλογή δεδομένων για τον πληθυσμό και την οικονομία, και γρήγορα επεκτάθηκε και σε άλλες χώρες της Δυτικής Ευρώπης. Ο μεγάλος ρυθμός θνησιμότητας στην Ευρώπη οφειλόταν στις επιδηματικές ασθένειες. Στις αρχικές καταγραφές των θανάτων από τη φοβερή ασθένεια την πανώλη, που εμφανίστηκε το 1348 και κράτησε πάνω από 400 χρόνια, προστέθηκαν στη συνέχεια και οι θάνατοι από άλλες αιτίες. Από δειγματοληπτική έρευνα που έκανε ο Άγγλος – Graunt το 1620 σε οικογένειες του Λονδίνου βρήκε ότι σε κάθε 88 άτομα υπήρχαν 3 θάνατοι. Χρησιμοποιώντας τους καταλόγους του Λονδίνου, που έδιναν 13.200 θανάτους το 1620, εκτίμησε τον πληθυσμό του Λονδίνου το έτος αυτό στα 387.200 άτομα.

Στην εποχή του Γουλιέλμου του Κατακτητή, έγινε μια σπουδαία απογραφή, στο τέλος του 11^{ου} αιώνα, αναφέρεται σε διάφορες μονάδες παραγωγής της Αγγλίας όπως ιχθυοτροφεία, μεταλλεία κ.α. Από το 16^{ο} έως το 19^{ο} αιώνα, η ραγδαία ανάπτυξη του εμπορίου ώθησε τις πολιτειακές αρχές στη μελέτη οικονομικών δεδομένων, όπως είναι το πλήθος και η δυναμικότητα βιομηχανιών κτλ.

Σε μια στατιστική έρευνα σήμερα μπορούμε να διακρίνουμε τρία στάδια: τη συλλογή στατιστικού υλικού, την επεξεργασία και παρουσίαση του και τέλος την ανάλυση αυτού του υλικού και την εξαγωγή χρήσιμων συμπερασμάτων. Τα τρία αυτά σχέδια επιτυγχάνονται με την εφαρμογή κατάλληλων για κάθε περίπτωση στατιστικών μεθόδων με την βοήθεια των υπολογιστών.

Συμπερασματικά λοιπόν μπορούμε να δώσουμε ως ορισμό της ΄΄Στατιστικής΄΄ το συνηθέστερο και πλέον γνωστό ορισμό του R. A. Fisher (1890- 1962), πατέρα της σύγχρονης στατιστικής.

Στατιστική είναι ένα σύνολο αρχών και μεθοδολογιών για:

- Σχεδιασμό της διαδικασίας συλλογής δεδομένων
- Τη συνοπτική και αποτελεσματική παρουσίασή τους
- Την ανάλυση και εξαγωγή αντίστοιχων συμπερασμάτων.

1.2 Βασικές έννοιες στατιστικής

1.2.1 Στατιστικη

Η στατιστική είναι η μέθοδος συλλογής, επεξεργασίας και εξαγωγής συμπερασμάτων για ένα μελετούμενο μέγεθος (π.χ. βαθμοί μαθητών σε μια τάξη). Πληροφόρηση για το μελετούμενο μέγεθος παίρνει κάποιος είτε συλλέγοντας στοιχεία, είτε εκτελώντας ένα πείραμα τύχης. Το σύνολο των άπειρου πλήθους αποτελεσμάτων ενός πειράματος τύχης ορίζει το λεγόμενο **δειγματοχώρο (S)**. Το σύνολο αυτό στη στατιστική καλείται **πληθυσμός**. Ο πληθυσμός χαρακτηρίζεται σαν **πεπερασμένος** αν είναι γνωστό το αριθμητικό του μέγεθος και **άπειρος** αν θεωρητικά μπορεί να γίνει άπειρος ο αριθμός πειραμάτων τύχης.

Σε πολλά πειράματα τύχης, όπως π.χ. μπορούν να θεωρηθούν οι ημερίσιες καιρικές συνθήκες ενός τόπου,ο δειγματοχώρος είναι δύσκολο να περιγραφεί γιατί περιλαμβάνει όλες τις ιδιότητες (π.χ. υγρασία, θερμοκρασία, ατμοσφαιρική πίεση κ.λπ.) της ατμόσφαιρας. Στην περίπτωση αυτή μπορούν να θεωρηθεί χωριστά κάθε ιδιότητα, π.χ. μέγιστη ημερίσια θερμοκρασία του αέρα ή ύψος βροχής 24ώρου κ.λπ., οπότε στο δειγματοχώρο ορίζονται υποσύνολα, το καθένα με ορισμένη φυσική ιδιότητα.

1.2.2 Πληθυσμός – Μεταβλητή – Δείγμα

Όπως προαναφέρθηκε και προηγουμένως, αυτό που μας ενδιαφέρει είναι να εξετάσουμε τα στοιχεία ενός συνόλου ως προς ένα ή περισσότερα χαρακτηριστικά τους. Αυτό συμβαίνει, για παράδειγμα, όταν ενδιαφερόμαστε για:

1) Τον αριθμό των υπαλλήλων μιας επιχείρησης.
2) Τις προτιμήσεις των ψηφοφόρων εν όψει των προσεχών εκλογών.
3) Το ύψος, το βάρος, την ομάδα αίματος, το φύλο κ.λπ.

Σε καθένα από τα παραδείγματα αυτά έχουμε ένα σύνολο και θέλουμε να εξετάσουμε τα στοιχεία του ως προς ένα ή περισσότερα χαρακτηριστικά τους. Ένα τέτοιο σύνολο λέγεται **πληθυσμός** (population). Τα στοιχεία του πληθυσμού συχνά αναφέρονται ως μονάδες ή άτομα πληθυσμού. Τα χαρακτηριστικά ως προς τα οποία εξετάζουμε ένα πληθυσμό λέγονται **μεταβλητές** (variables) και τις συμβολίζουμε με το κεφαλαίο γράμμα X,Y,Z,B,... κ.λπ. οι δυνατές τιμές που μπορεί να πάρει μια μεταβλητή λέγονται **τιμές της μεταβλητής**. Από τη διαδοχική εξέταση των ατόμων του πληθυσμού ως προς ένα χαρακτηριστικό τους προκύπτει μια σειρά από δεδομένα, που λέγονται στατιστικά δεδομένα ή παρατηρήσεις. Για παράδειγμα, αν εξετάζαμε το βάρος δέκα ατόμων, τα σατιστικά δεδομένα ή παρατηρήσεις που θα προκύψουν μπορεί να είναι: 70, 70, 90, 70, 80, 85, 85, 85, 80, 90. Οι δυνατές τιμές που μπορεί να πάρει η μεταβλητή 'βάρος' είναι: 70, 90, 80, 85.

Τις μεταβλητές τις διακρίνουμε:

1) Σε **ποιοτικές ή κατηγορικές** μεταβλητες των οποίων οι τιμές δεν είναι αριθμοί. Τέτοιες είναι, για παράδειγμα, το φύλο (με τιμές κορίτσι, αγόρι), η οικονομική κατάσταση και η υγεία των ανθρώπων (που μπορεί να χαρακτηριστεί ως κακή, μέτρια, καλή ή πολύ καλή).
2) Σε **ποσοτικές** μεταβλητες, των οποίων οι τιμές είναι αριθμοί και διακρίνονται σε:
   - Σε **συνεχείς** μεταβλητές, που είναι αυτές που μπορούν να πάρουν οποιαδήποτε τιμή μέσα στο πεδίο ορισμού τους (π.χ. η θερμοκρασία).
   - Σε **διακριτές ή ασυνεχείς** μεταβλητές που είναι αυτές που μπορούν να λάβουν τιμές μόνο από το σύνολο φυσικών αριθμών (π.χ. 1, 2, 3, ..., 6) κτλ.

Δείγμα ονομάζεται το υπούνολο του πληθυσμού το οποίο μπορούμε να καταγράψουμε υπό τους περιορισμούς (υλικούς και χρονικούς) της έρευνάς μας. Για παράδειγμα, αν θέλουμε να βρούμε τι ποσοστό ελλήνων έχει πράσινα μάτια, είναι πρακτικά αδύνατο να ρωτήσουμε όλους τους Έλληνες. Επομένως μελετάμε ένα δείγμα του πληθυσμού των Ελλήνων, προκειμένου να εκτιμήσουμε το ζητούμενο ποσοστό. Το πλήθος των στοιχείων του δείγματος λέγεται **μέγεθος του δείγματος**. Το δείγμα είναι ένα σύνολο από υποκείμενα που έχει επιλεγεί κατάλληλα ώστε να αντιπροσωπεύσει έναν ολόκληρο πληθυσμό. Προφανώς, η διαδικασία επιλογής δείγματος αποτελεί μια εξαιρετικά σημαντική διαδικασία, καθώς καθορίζει την εγκυρότητα των αποτελεσμάτων της μελέτης. Αν το δείγμα δεν είναι αντιπροσωπευτικό υπάρχει μεγάλη πιθανότητα η μελέτη να οδηγήσει σε εσφαλμένα συμπεράσματα.

Στην πράξη, ένα δείγμα είναι αντιπροσωπευτικό του πληθυσμού όταν έχει χρησιμοποιηθεί η διαδικασία της τυχαίας δειγματοληψίας (random sampling) για την απόκτησή του. Η τυχαία δειγματοληψία απαιτεί κάθε υποκείμενό του πληθυσμού να έχει την ίδια πιθανότητα να επιλέγει. Όπως και οι πληθυσμοί, ένα δείγμα μπορεί να είναι αντίστοιχα από πολύ μικρό εώς πολύ μεγάλο. Εύκολα γίνεται αντιληπτό ότι όσο μεγαλύτερο είναι το δείγμα, τόσο πιο αντιπροσωπευτικό θα είναι, καθώς αυξάνεται ο αριθμός των υποκείμενων που επιλέγονται από τον πληθυσμό.

Στη στατιστική όταν χρησιμοποιούμε δεδομενα κρίνεται αναγκαίο να προσδιορίσουμε εάν τα δεδομένα προέρχονται από έναν πληθυσμό ή από ένα δείγμα. Για να εξυπηρετηθεί αυτός ο διαχωρισμός, η στατιστική χρησιμοποιεί τον όρο παράμετροι (parameter) για να περιγράψει δεδομένα που αναφέρονται στον πληθυσμό και τον όρο στατιστικος δείκτης (statistic) για τα δεδομένα που συσχετίζονται με ένα δείγμα.

1.2.3 Συλλογή στατιστικών δεδομένων

Οι κυριότερες μέθοδοι συλλογής στατιστικών δεδομένων είναι η απογραφή και η δειγματοληψία.

Απογραφή είναι μια μέθοδος συλλογής στατιστικών δεδομένων, που ακολουθούμε για να πάρουμε όλες τις απαραίτητες πληροφορίες, για έναν πληθυσμό εξετάζοντας όλα τα άτομα του πληθυσμού ως προς τα χαρακτηριστικά που μας ενδιαφέρουν.

Δειγματοληψία ονομάζεται η διαδικασία καταγραφής ενός υποσυνόλου του πληθυσμού. Προχωρούμε σε δειγματοληψία γιατί η απογραφή είναι δύσκολη, οικονομικά και χρονικά ασύμφορη και πολλές φορές αδύνατη. Για αυτόν τον λόγο επιλέγουμε ένα υποσύνολο του πληθυσμού, το δείγμα. Συλλέγουμε τις παρατηρήσεις απ το δείγμα και στην συνέχεια γενικεύουμε τα συμπεράσματα για ολόκληρο τον πληθυσμό. Τα συμπεράσματα όμως, που θα προκύψουν από την μελέτη του δείγματος θα είναι αναξιόπιστα δηλαδή θα ισχύουν με ικανοποιητική προσέγγιση για ολόκληρο τον πληθυσμό, μόνο όταν η επιλογή του δείγματος θα έχει γίνει με τέτοιο τρόπο, ώστε το δείγμα να είναι αντιπροσωπευτικό.

Οι λόγοι για τον οποίο συμβαίνει μια δειγματοληψία είναι οι οικονομικοί και χρονικοί περιορισμοί που υπάρχουν αλλά και η περιορισμένη πρόσβαση στον πληθυσμό. Οι περιορισμοί αυτοί δεν μειώνουν την αξία της δειγματοληψίας καθώς μπορεί να δώσει ακριβή και αξιόπιστα αποτελέσματα ιδιαίτερα όταν ο πληθυσμός που μελετούμε είναι ομοιογενής ως προς το χαρακτηριστικό που μας ενδιαφέρει. Επίσης, η δειγματοληψία μπορεί να είναι περισσότερο αξιόπιστη από μία απογραφή, όταν η γνώση του ερωτώμενου πως πρόκειται για απογραφή αυξάνει τη μεροληψία της απόκρισης π.χ. οι αποκρίσεις των αλλοδαπών στην εθνική απογραφή, οι οποίες ίσως να είναι πιο ειλικρινείς σε δειγματοληψία ή ακόμα στην περίπτωση όπου δεν υπάρχουν αξιόπιστοι κατάλογοι του πληθυσμού όπως στις μη αναπτυγμένες χώρες. Τέλος η δειγματοληψία μειώνει το κόστος της έρευνας σε πραγματικούς πληθυσμούς, δηλαδή σε πληθυσμούς των οποίων το μέγεθος είναι γνωστό κάθε μία χρονική στιγμή.

Η επιλογή του αντιπροσωπευτικού δείγματος αποτελει πολύ σοβαρή και δύσκολη διαδικασία. Ο κακός σχεδιασμός και η εκτέλεση της στατιστικής έρευνας, η μη αντιπροσωπευτικότηρα του δείγματος, ο μη σωστός καθορισμός του μεγέθους του δείγματος αποτελούν μερικά βασικά μειονεκτήματα στη διαδικασία επιλογής ενός δείγματος.

Το σφάλμα μιας δειγματοληψίας χωρίζεται σε τυχαίο και συστηματικό. Τυχαίο σφάλμα δειγματοληψίας ονομάζεται η διαφορά μεταξύ των μετρήσεων του δείγματος και των πραγματικών μετρήσεων το οποίο θα υπάρχει στην έρευνα μας. Το τυχαίο σφάλμα προκύπτει με φυσικό τρόπο καθώς η μέση τιμή (ή άλλα στατιστικά) του υποσυνόλου του πληθυσμού που επιλέγουμε ως δείγμα είναι πρακτικά αδύνατο να είναι ίση με τη μέση τιμή του πληθυσμού, λόγω των τυχαίων σφαλμάτων της δειγματοληψίας. Αν η δειγματοληψία γίνει με κάποια πιθανοθεωρητική μέθοδο τότε το σφάλμα μπορεί να εκτιμηθεί ενώ αν γίνει με κάποια μη πιθανοθεωρητική μέθοδο (όπως συχνά συμβαίνει στην πράξη) τότε ο υπολογισμός του δεν είναι δυνατός.

Συστηματικό σφάλμα δειγματοληψίας ονομάζεται το σφάλμα που εμφανίζεται λόγω των σφαλμάτων σχεδίασης της δειγματοληψίας, όπως π.χ. αν μετράς την ευχαρίστηση από την απόκτηση ενός προϊόντος και έχεις δύο ομάδες που ρωτάνε, με τη μία να έχεις μία πολύ όμορφη γυναίκα ως συνεντευξιαστή και την άλλη έναν άνδρα. Το συστηματικό σφάλμα είναι διαφορετικό από το τυχαίο σφάλμα, καθώς οφείλεται αποκλειστικά στον σχεδιασμό της έρευνας.

1.3 Παρουσίαση στατιστικών δεδομένων

Μετά τη συλλογή στατιστικών δεδομένων ακολουθεί το στάδιο της παρουσίασης συνοπτικών πινάκων ή γραφικων παραστάσεων, ώστε να κατανοούνται ευκολότερα από τον αναγνώστη και να διευκολύνεται το έργο της στατιστικής αναλύσεως. Η παρουσίαση των στατιστικών δεδομένων γίνεται με τους εξής τρόπους:

1) Με μορφή στατιστικων πινάκων και
2) Με στατιστικά διαγράμματα.

1.4 Στατιστικοί πίνακες

1.4.1 Έννοια και κατηγορίες στατιστικών πινάκων

Η παρουσίαση των στατιστικων δεδομένων στους πίνακες γίνεται με τέτοιο τρόπο, ώστε να επιτυγχάνεται η συνοπτική εμφάνιση των αριθμιτικών δεδομένων, η οποία διευκολύνει τη σύγκριση των δεδομένων και ενημερώνει εύκολα τον αναγνώστη.

Οι στατιστικοί πίνακες διακρίνονται στους:

1) **Γενικούς πίνακες**, οι οποίοι περιέχουν όλες τις πληροφορίες που προκύπτουν από μια στατιστική έρευνα.
2) **Ειδικούς πίνακες**, οι οποίοι είναι συνοπτικοί κι σαφείς. Τα στοιχεία τους αποτελούνται συνήθως από τους γενικούς πίνακες.

1.4.2 Τεχνική κατασκευή στατιστικών πινάκων

Κάθε πίνακας που έχει κατασκευαστεί σωστά πρέπει να περιέχει:

1) Τον τίτλο, που γράφεται στο επάνω μέρος του πίνακα και δηλώνει με σαφήνεια και συνοπτικά το περιεχόμενο του πίνακα.
2) Τις επικεφαλίδες, των γραμμών και στηλών, που δείχνουν συνοπτικά την φύση και τις μονάδες μέτρησης των δεδομένων.
3) Το κύριο σώμα, περιέχει τα στατιστικά δεδομένα, τα οποία μπορεί να είναι ποσοτικής ή ποιοτικής φύσεως.
4) Την πηγή, σε κάθε στατιστικό πίνακα είναι απαραίτητο να γράφεται η πηγή από τα οποία έχουν ληφθεί τα δεδομένα που περιέχει ο πίνακας. Η πηγή γράφεται στο κάτω μέρος του πίνακα και περιέχει το όνομα του συγγραφέα και του εκδότη ή την υπηρεσία που εκδίδει το δημοσίευμα, τον τίτλο και τη χρονολογία έκδοσης του δημοσιεύματος.

Παρακάτω δίνονται μερικοί στατιστικοί πίνακες, που διευκρινίζουν την εφαρμογή προηγούμενων εννοιών.

Πίνακας 1

Ποσοστά ανεργίας στην Ελλάδα κατά ομάδες ηλικιών

Έτη 2012-2013

Ηλικία	**2012**	**2013**
15-24	51,5	57,5
25-34	30,4	36
35-44	19,7	23,4
45-54	17,2	21,3
55-64	12,9	16,3
Σύνολο	131,7	154,5

Πηγή: Ι.Κ.Α

Πίνακας 2

Σχολίκος πληθυσμός κατά φύλο και βαθμίδα εκπαιδεύσεως κατά στο σχολικό έτος 2012-2013

Βαθμίδες εκπαίδευσεως	**Αγόρια**	**Κορίτσια**	**Σύνολο**
Προσχολικη εκπαίδευση	65.183	48.384	113.567
Δημοτική εκπαίδευση	398.132	390.749	788.881
Μέση γενική εκπαίδευση	427.419	445.213	872.632
Μέση 'Τα και Ε' εκπαίδευση	72.848	56.947	129.795
Ανώτατη (Α.Ε.Ι) εκπαίδευση	57.246	60.283	117.529
Τεχνολογική (Τ.Ε.Ι)	44.654	29.182	73.836
Σύνολο	1.065.482	1.030.488	2.096.240

Πηγή: Ε.Σ.Υ.Ε « Στατιστική της Εκπαίδευσης 2012-2013»

1.5 Πίνακες κατανομής συχνοτήτων

Οι στατιστικοί πίνακες και γραφικές παραστάσεις αποτελούν χρήσιμα μέσα για να παρουσιάσουμε τα δεδομένα καθαρά, σύντομα και με σαφήνεια. Επίσης μπορούν να αποκαλύψουν σημαντικά χαρακτηριστικά των δεδομένων, όπως το εύρος τους, τη συμμετρικότητα τους ή την ύπαρξη ακραίων τιμών.

Έστω $x_1, x_2, \dots, x_k$ οι τιμές μιας μεταβλητής X, που αφορά τα στοιχεία ενός δείγματος μεγέθους ν, $k \leq \nu$. Στην τιμή x_i αντιστοιχίζεται η **συχνότητα** ν_i, δηλαδή ο φυσικός αριθμός που δείχνει πόσες φορές εμφανίζεται η τιμή x_i της μεταβλητής X στο σύνολο των παρατηρήσεων. Για τις συχνότητες ισχύουν:

- $0 \leq \nu_i \leq \nu, i = 1,2, \dots, k.$
- $\nu_1 + \nu_2 + \cdots + \nu_k = \nu.$

Εάν διαιρέσουμε τη συχνότητα ν_i με το μέγεθος ν του δείγματος, προκύπτει η **σχετική συχνότητα** f_i της τιμής x_i, δηλαδή

$$f_i = \frac{\nu_i}{\nu}, i = 1,2, \dots, k.$$

Για τις σχετικές συχνότητες ισχύουν:

- $0 \leq f_i \leq 1, i = 1,2, \dots, k$.
- $f_1 + f_2 + \cdots + f_k = 1$.

Συνήθως, τις σχετικές συχνότητες f_i τις εκφραζουμε επί τοις εκατό, οπότε συμβολίζονται με $f_i\%$.

Οι ποσότητες x_i, ν_i, f_i, $i = 1,2, \dots, k$ για ένα δείγμα συγκεντρώνονται σε ένα συνοπτικό πίνακα, που ονομάζεται **πίνακας συχνοτήτων**.

Για παράδειγμα, για την μεταβλήτη X : 'αριθμός αδελφών' του πίνακα 3 οι συχνότητες για τις τιμές x_1=0, x_2= 1, x_3 =2, x_4 =3 είναι, αντίστοιχα, ν_1 =8, ν_2 =22, ν_3=7, ν_4 =3.

Πίνακας 3

Κατανομή συχνοτήτων της μεταβλητής X: ' αριθμός αδελφών' των φοιτητών

Αριθμός αδελφών $\boldsymbol{x_i}$	**Συχνότητα** $\boldsymbol{\nu_i}$	**Σχετική Συχνότητα** $\boldsymbol{f_i}$	**Σχετική Συχνότητα** $\boldsymbol{f_i\%}$
0	8	0,200	20,0
1	22	0,550	55,0
2	7	0,175	17,5
3	3	0,075	7,5
Σύνολο	40	1,000	100,0

1.6 Αθροιστικές συχνότητες

Στην περίπτωση των ποσοτικών μεταβλητών εκτός από τις συχνότητες ν_i και f_i χρησιμοποιούνται οι **αθροιστικές συχνότητες** N_i και οι **αθροιστικές σχετικές συχνότητες** F_i, οι οποίες εκφράζουν το πλήθος και το ποσοστό αντίστοιχα των παρατηρήσεων που είναι μικρότερες ή ίσες της τιμής x_i. Εάν $x_1 < x_2 < \cdots < x_k$, τότε ισχύουν:

- $N_i = \nu_1 + \nu_2 + \cdots + \nu_i,\ i = 1,2, \dots, k$.
- $F_i = f_1 + f_2 + \cdots + f_i,\ i = 1,2, \dots, k$.

Συχνά οι F_i πολλαπλασιάζονται επί 100 εκφραζόμενες έτσι επί τοις εκατό, δηλαδή

$$F_i\% = 100\ F_i.$$

Πίνακας 4

Κατανομή συχνοτήτων και αθροιστικών συχνοτήτων της μεταβλητής 'αριθμός αδελφών' των φοιτητών

Αριθμος αδελφών x_i	**Συχνότητα ν_i**	**Σχετ. Συχνότητα f_i**	**Σχετ. Συχνότητα $f_i\%$**	**Αθροι. Συχνοτητα N_i**	**Αθροι. Σχετ. Συχνότητα F_i**	**Αθροι. Σχετ. Συχνότητα $F_i\%$**
0	8	0,200	20,0	8	0,200	20,0
1	22	0,550	55,0	30	0,750	75,0
2	7	0,175	17,5	37	0,925	92,5
3	3	0,075	7,5	40	1,000	100,0
Σύνολο	40	1,000	100,0	----	-----	-----

1.7 Γραφική παράσταση κατανομής συχνοτήτων

Τα στατιστικά δεδομένα παρουσιάζονται πολλές φορές και υπό μορφή γραφικών παραστάσεων ή διαγραμμάτων.

Οι γραφικές παραστάσεις παρέχουν πιο σαφή εικόνα του χαρακτηριστικού σε σχέση με τους πίνακες, είναι πολύ πιο ενδιαφέρουσες και ελκυστικές, χωρίς βέβαια να προσφέρουν περισσότερη πληροφορία από εκείνη που περιέχεται στους αντίστοιχους πίνακες συχνοτήτων. Επί πλέον με τα διαγράμματα διευκολύνεται η σύγκριση μεταξύ ομοειδών στοιχείων για το ίδιο ή για διαφορετικά χαρακτηριστικά.

Υπάρχουν διάφοροι τρόποι γραφικής παρουσίασης, ανάλογα με το είδος των δεδομένων που έχουμε. Όπως όμως οι στατιστικοί πίνακες έτσι και τα στατιστικά διαγράμματα πρέπει να συνοδεύονται από α) τον τίτλο, β) την κλίμακα με τις τιμές των μεγεθών που απεικονίζονται, γ) το υπόμνημα που επεξηγεί συνήθως τις τιμές της μεταβλητής και δ) την πηγή των δεδομένων.

Α) Ραβδόγραμμα

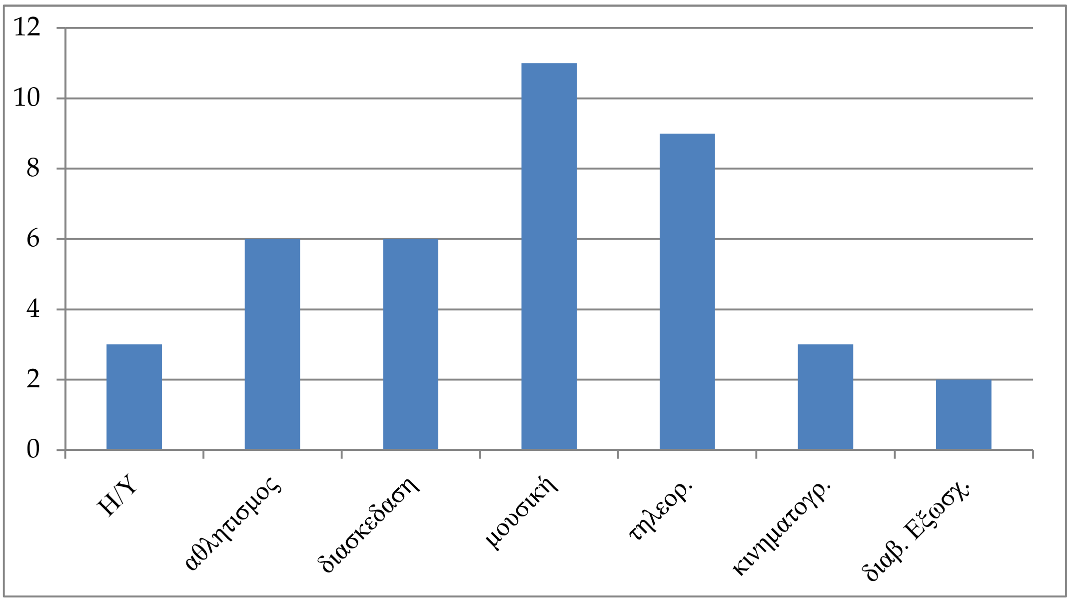
Το ραβδόγραμμα (barchart) χρησιμοποιήται για τη γραφική παράσταση των τιμών μιας ποιοτικής μεταβλητής. Το ραβδόγραμμα αποτελείται από ορθογώνιες στήλες που οι βάσεις τους βρίσκονται πάνω στον οριζόντιο ή τον κατακόρυφο άξονα. Σε κάθε τιμή της μεταβλητής Χ αντιστοιχεί σε μία ορθογώνια στήλη της οποίας το ύψος είναι ίσο με την αντίστοιχη συχνότητα ή σχετική συχνότητα. Έτσι έχουμε αντίστοιχα το ραβδόγραμμα συχνοτήτων και το ραβδόγραμμα σχετικών συχνοτήτων. Τόσο η απόσταση μεταξύ των στηλών όσο και το μήκος των βάσεών τους καθορίνται αυθαίρετα. Στον παρακάτω πίνακα 5 έχουμε την κατανομή συχνοτήτων της μεταβλητής Χ : '**απασχόληση στον ελεύθερο χρόνο**' και στα Σχήματα 1 (α) και (β) τα αντίστοιχα ραβδογράμματα **συχνοτήτων** και **σχετικών συχνοτήτων**.

Πίνακας 5

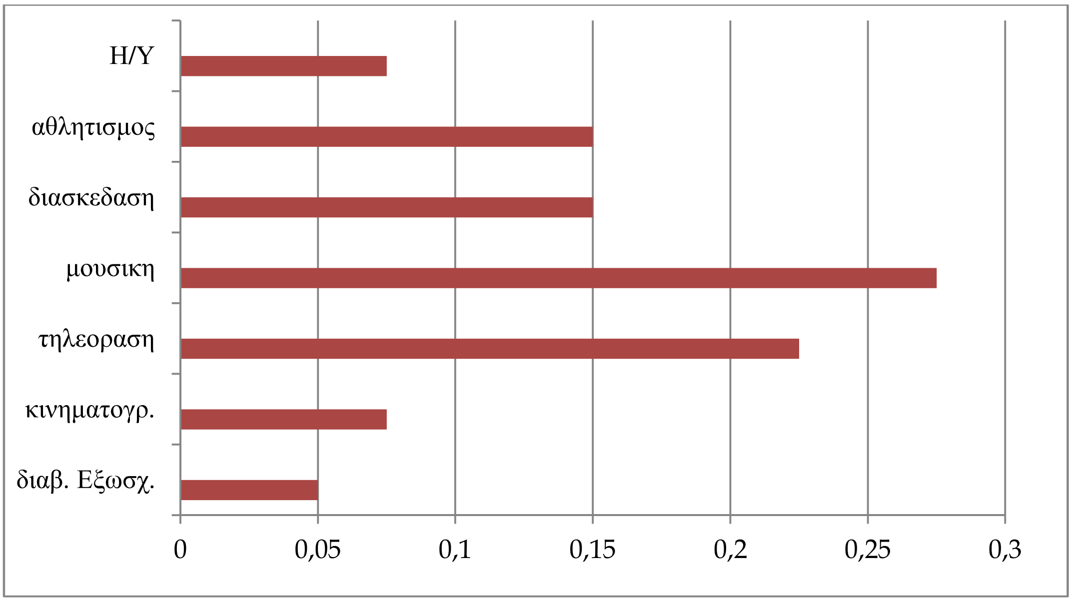
Κατανομή συχνοτήτων για την απασχόληση στον ελεύθερο χρόνο τους των φοιτητών

i	**Απασχόληση** x_i	**Συχνότητα** ν_i	**Σχετική συχνότητα** f_i	**Σχετική συχνότητα** f_i%
1	Υπολογιστές	3	0,075	7,5
2	Αθλητισμός	6	0,150	15,0
3	Διασκέδαση	6	0,150	15,0
4	Μουσική	11	0,275	27,5
5	Τηλεόραση	9	0,225	22,5
6	Κινηματογράφος	3	0,075	7,5
7	Διάβασμα εξωσχ. Βιβλίων και άλλα	2	0,050	5,0
Σύνολο		40	1,000	100,0

Μερικές φορές σε ένα ραβδόγραμμα συχνοτήτων ο ρόλος των δύο αξόνων είναι δυνατών να αντιστραφεί, όπως φαίνεται στο Σχήμα 1 (β), που παριστάνεται το ραβδόγραμμα σχετικών συχνοτήτων της ίδιας μεταβλητής.



Σχήμα 1 (α)



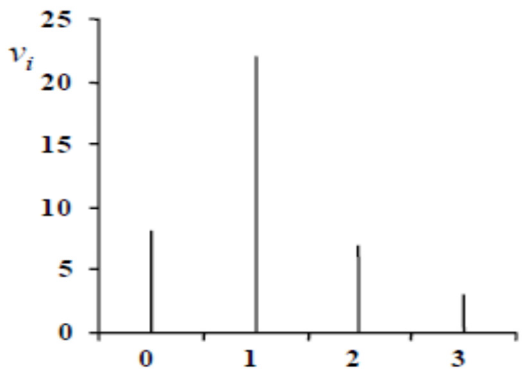
Σχήμα 1 (β)

Ραβδόγραμμα συχνοτήτων (α) και σχετικών συχνοτήτων (β) για την απασχόληση των φοιτητών.

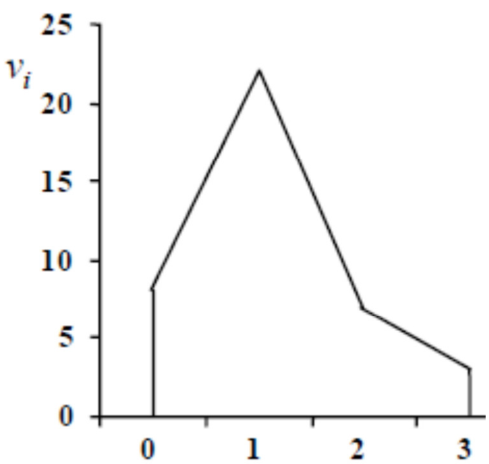
Β) Διάγραμμα συχνοτήτων

Όταν έχουμε μια ποσοτική μεταβλητή αντί του ραβδογράμματος χρησιμοποιείται το **διάγραμμα συχνοτήτων**. Αυτό μοιάζει με το ραβδόγραμμα με μόνη διαφορά ότι αντί να χρησιμοποιούμε συμπαγή ορθογώνια υψώνουμε σε κάθε x_i (υποθέτοντας ότι $x_1 < x_2 < \cdots < x_k$) μία κάθετη γραμμή με μήκος ίσο προς την αντίστοιχη συχνότητα όπως φαίνεται στο Σχήμα 2 (α).

Ενώνοντας τα σημεία (x_i, ν_i) ή (x_i, f_i) έχουμε το λεγόμενο **πολύγωνο συχνοτήτων** ή **πολύγωνο σχετικών συχνοτήτων**, αντίστοιχα, που μας δίνουν μια γενική ιδέα για την μεταβολή της συχνότητας ή της σχετικής συχνότητας όσο μεγαλώνει η τιμή της μεταβλητής που εξετάζουμε, όπως φαίνεται στο Σχήμα 2 (β).



Σχήμα 2 (α)



Σχήμα 2 (β)

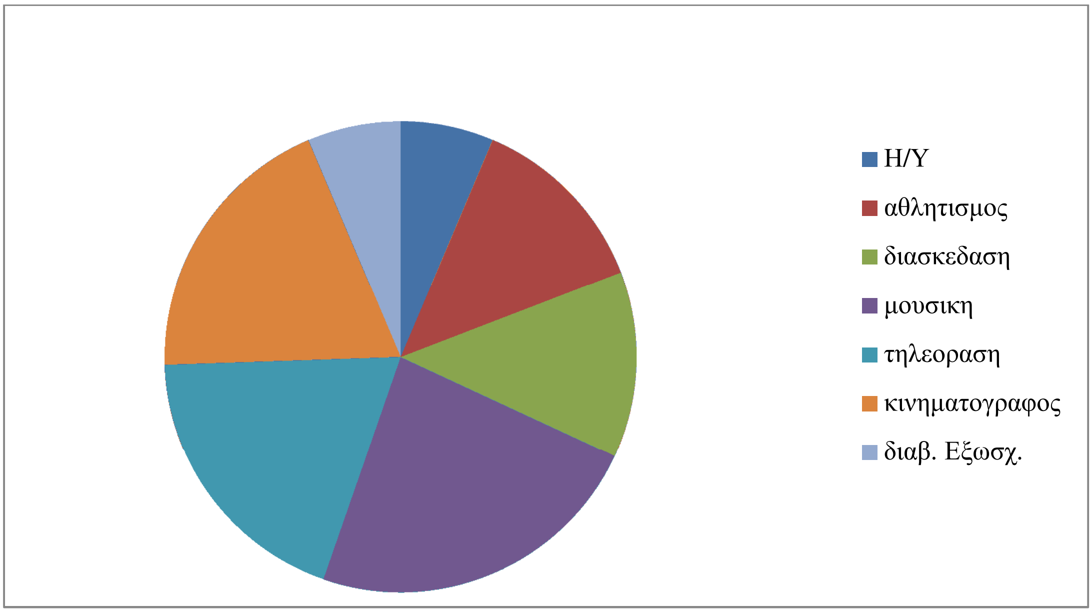
Γ) Κυκλικό διάγραμμα

Το **κυκλικό διάγραμμα** χρησιμοποιείται για τη γραφική παράσταση τόσο των ποιοτικών όσο και των ποσοτικών δεδομένων, όταν οι διαφορετικές τιμές της μεταβλητής είναι σχετικά λίγες. Για να κατασκευάσουμε ένα κυκλικό διάγραμμα χωρίζουμε τον κυκλικό δίσκο σε κυκλικούς τομείς των οποίων οι επίκεντρες γωνίες βαίνουν σε τόξα α_i ανάλογα με τις συχνότητες ν_i των τιμών της μεταβλητής.

Επειδή τα ποσά α_i και ν_i είναι ανάλογα, ισχύει:

$$\frac{\alpha_i}{360^\circ} = \frac{\nu_i}{\nu} \Leftrightarrow \alpha_i = \frac{\nu_i}{\nu} \cdot 360^\circ = f_i \cdot 360^\circ.$$

Στο Σχήμα 3 παριστάνεται το αντίστοιχο κυκλικό διάγραμμα σχετικών συχνοτήτων της ‘απασχόλησης των φοιτητών’.



Σχήμα 3

Κυκλικό διάγραμμα σχετικών συχνοτήτων της απασχόλησης φοιτητών

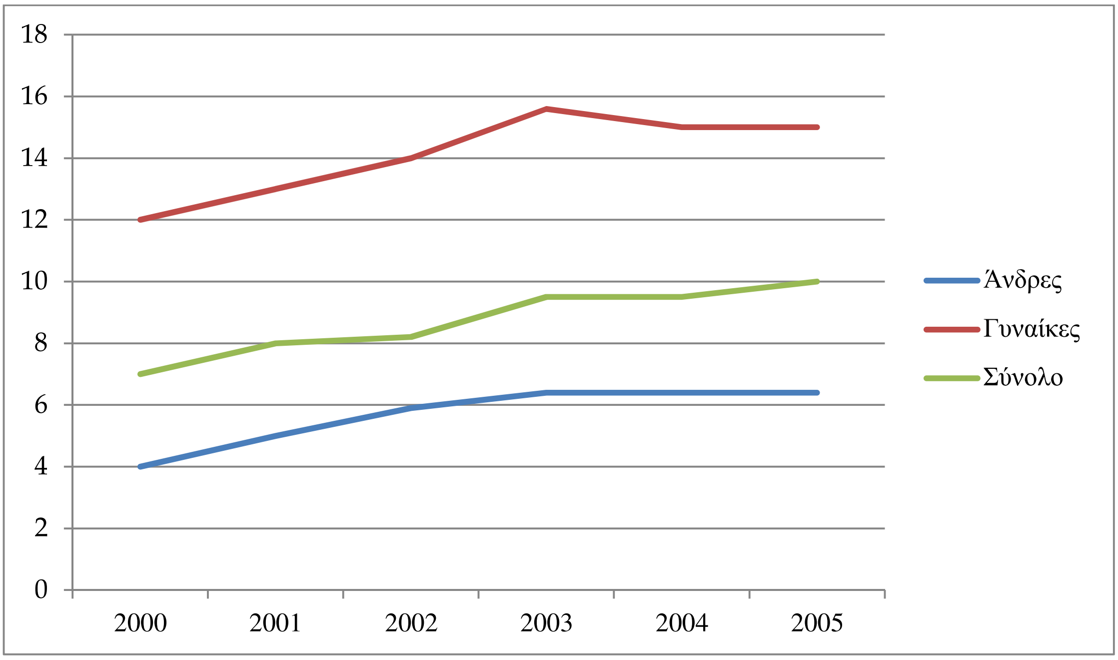
Δ) Σημειόγραμμα

Όταν έχουμε λίγες παρατηρήσεις, η κατανομή τους μπορεί να περιγραφεί με το **σημειόγραμμα**, στο οποίο οι τιμές παριστάνονται γραφικά σαν σημεία υπεράνω ενός οριζόντιου άξονα. Στο παρακάτω σχήμα 4 έχουμε το σημειόγραμμα των χρόνων (σε λεπτά) 4, 2, 3, 1, 5, 6, 4, 2, 3, 4, 7, 4, 8, 6, 3 που χρειάστηκαν 15 φοιτητές για να λύσουν ένα πρόβλημα.

Σχήμα 4

Ε) Χρονόγραμμα

Το **χρονόγραμμα** ή **χρονολογικό διάγραμμα** χρησιμοποιείται για τη γραφική απεικόνιση ενός οικονομικού, δημογραφικού ή άλλου μεγέθους. Ο οριζόντιος άξονας χρησιμοποιείται συνήθως ως άξονας μέτρησης του χρόνου και ο κάθετος ως άξονας μέτρησης της εξεταζόμενης μεταβλητής. Στο παρακάτω Σχήμα 5 έχουμε το χρονόγραμμα του ποσοστού ανεργίας στη χώρα μας από το 2000 εως το 2005.



Σχήμα 5

Ποσοστό ανεργίας στην Ελλάδα, πηγή Ε.Σ.Υ.Ε

Στο σχήμα παρατηρούμε ότι στο γυναικείο πληθυσμό υπάρχει συστηματικά ποσοστό ανεργίας, γύρω στις 8 εκατοστιαίες μονάδες. Στο διάστημα 2003-2005 το ποσοστό ανεργίας έχει σταθεροποιηθεί γύρω στο 6,5% για τους άνδρες και γύρω στο 15% για τις γυναίκες.

1.8 Ιστόγραμμα συχνοτήτων

Στην περίπτωση διακριτής μεταβλητής, όταν το πλήθος των τιμών της είναι μεγάλο, αλλά πολύ περισσότερο σε συνεχή μεταβλητή X, ταξινομούμε τα δεδομένα σε ίσα διαδοχικά διαστήματα της μορφής [,) (κλάσεις ή τάξεις) και καταγράφουμε τις συχνότητες των παρατηρήσεων που ανήκουν σε κάθε κλάση. Για την i κλάση $[a, b)$ η κεντρική τιμή x_i είναι:

$$x_i = \frac{a+b}{2}.$$

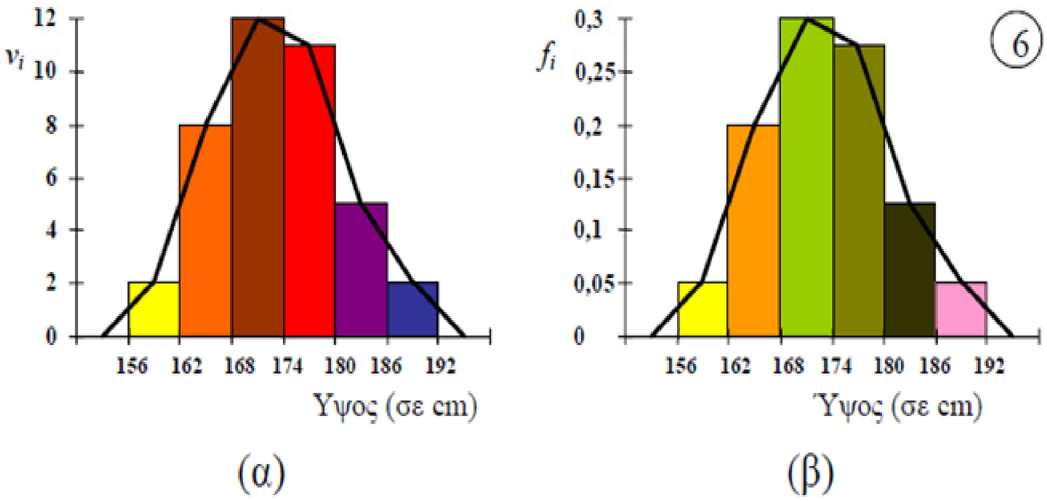
Ιστόγραμμα συχνοτήτων ονομάζεται η γραφική παράσταση ενός πίνακα συχνοτήτων με ομαδοποιημένα δεδομένα. Στον οριζόντιο άξονα ενός συστήματος ορθογωνίων αξόνων σημειώνουμε, με κατάλληλη κλίμακα, τα όρια των κλάσεων. Στη συνέχεια κατασκεάζουμε διαδοχικά ορθογώνια (ιστούς), από καθένα από τα οποία έχει βάση ίση με το πλάτος της κλάσης και ύψος τέτοιο, ώστε **το εμβαδόν τουορθογωνίου να ισούται με τη συχνότητα** της κλάσης αυτής.

Α) Κλάσεις ίσου πλάτους

Θεωρώντας το πλάτος c ως μονάδα μέτρησης στον οριζόντιο άξονα, το ύψος κάθε ορθογωνίου είναι ίσο προς τη συχνότητα της αντίστοιχης κλάσης, έτσι ώστε να ισχύει ότι το εμβαδόν των ορθογωνίων είναι ίσο με τις αντίστοιχες συχνότητες. Με ανάλογο τρόπο γίνεται και το **ιστόγραμμα σχετικών συχνοτήτων**.

Αν στα ιστογράμματα συχνοτήτων προσθέσουμε ακόμα δύο κλάσεις, στην αρχή και στο τέλος, με συχνότητα μηδέν και στην συνέχεια ενώσουμε τα μέσα των άνω βάσεων των ορθογωνίων, σχηματίζεται το λεγόμενο **πολύγωνο συχνοτήτων**. Το εμβαδόν του χωρίου που ορίζεται από το πολύγωνο συχνοτήτων και τον οριζόντιο άξονα είναι ίσο με το άθροισμα των συχνοτήτων, δηλαδή με το μέγεθος του δείγματος ν.

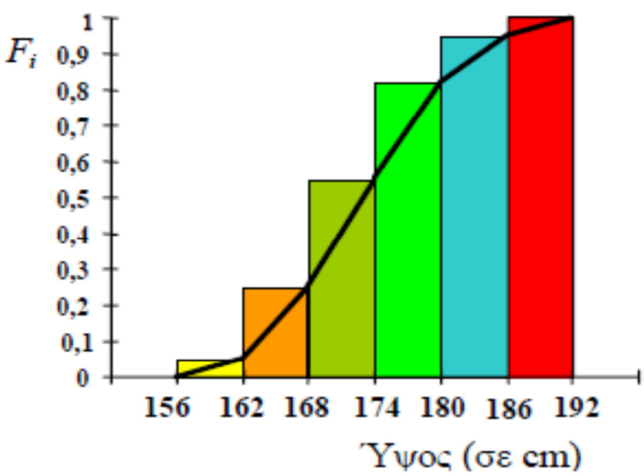
Στο παρακάτω Σχήμα 6 κατασκευάζεται από το ιστόγραμμα σχετικών συχνοτήτων και το πολύγωνο σχετικών συχνοτήτων με εμβαδόν ίσο με 1.



Σχήμα 6

Ιστόγραμμα και πολύγωνο (α) συχνοτήτων και (β) σχετικών συχνοτήτων.

Το **ιστόγραμμα αθροιστικών συχνοτήτων** και **σχετικών συχνοτήτων** κατασκευάζονται με τον ίδιο τρόπο. Αν ενώσουμε σε ένα ιστόγραμμα αθροιστικών συχνοτήτων τα δεξιά άκρα των άνω βάσεων των ορθογωνίων με ευθύγραμμα τμήματα βρίσκουμε το **πολύγωνο αθροιστικών συχνοτήτων της κατανομής**. Στο σχήμα 7 παριστάνεται το ιστόγραμμα και το πολύγωνο αθροιστικών σχετικών συχνοτήτων για το ύψος των φοιτητων.



Σχήμα 7

Β) Κλάσεις άνισου πλάτους

Συνήθως επιλέγουμε κλάσεις ίσου πλάτους, υπάρχουν όμως και περιπτώσεις που είναι απαραίτητο να έχουμε και κλάσεις διαφορετικού πλάτους όπως, για παράδειγμα στην περίπτωση όπου οι συχνότητες σε κάποιες κλάσεις να είναι πολύ μικρές τότε γίνεται συγχώνευση κελιών ή ακόμα και στην κατανάλωση νερού.

Για παράδειγμα, η διάρκεια (σε sec) $\nu = 80$ τηλεφωνημάτων που έγιναν τυχαία από ένα κινητό τηλέφωνο όπως δίνεται στον παρακάτω πίνακα.

Πίνακας 6

Διάρκεια τηλεφωνημάτων σε sec	Συχνότητα ν_i
0-20 20-25 25-30 30-40	20 20 24 16
Σύνολο	$\nu = 80$

Το αντίστοιχο ιστόγραμμα συχνοτήτων κατασκευάζεται πάλι, έτσι ώστε το εμβαδόν κάθε ορθογωνίου να ισούται με τη συχνότητα της αντίστοιχης κλάσης. Άρα, αν c_i είναι το πλάτος της κλάσης i με συχνότητα ν_i, το ύψος του ορθογωνίου θα είναι:

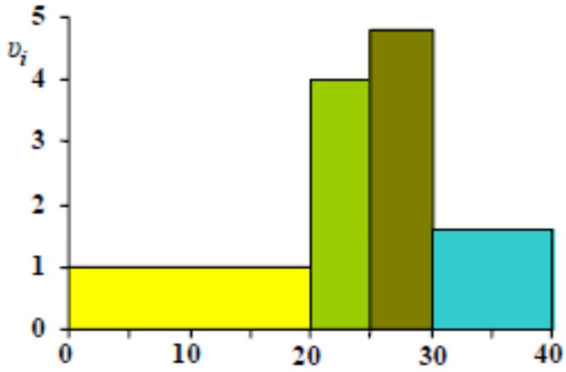
$$u_i = \frac{\nu_i}{c_i}, i = 1,2, \dots, k.$$

Για την κατασκευή του ιστογράμματος συχνοτήτων χρειαζόμαστε τα πλάτη των κλάσεων και τα ύψη των ορθογωνίων. Αυτά δίνονται στον επόμενο πίνακα.

Πίνακας 7

Διάρκεια τηλεφωνημάτων σε sec	**Πλάτος κλάσης c_i**	**Συχνότητα ν_i**	**Ύψος $u_i = \frac{\nu_i}{c_i}$**	**Ύψος $u_i = \frac{f_i\%}{c_i}$**
0-20 20-25 25-30 30-40	20 5 5 10	20 20 24 16	1,0 4,0 4,8 1,6	1,25 5,00 6,00 2,00

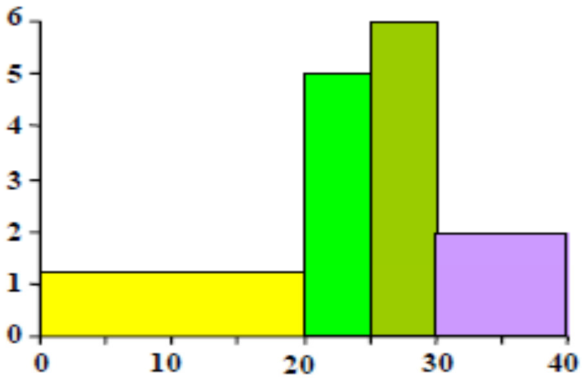
Το ιστόγραμμα συχνοτήτων δίνεται στο Σχήμα 8 (α). Παρατηρούμε ότι το άθροισμα των εμβαδών όλων των ορθογωνιών είναι ίσο με το συνολικό μέγεθος δείγματος ν, όπως δηλαδή συμβαίνει και στο ιστόγραμμα με κλάσεις ίσου πλάτους



Σχήμα 8 (α)

Ιστόγραμμα συχνοτήτων

Με ανάλογο τρόπο κατασκευάζεται και το ιστόγραμμα σχετικών συχνοτήτων Σχήμα 8 (β) αρκεί να χρησιμοποιήσουμε ως ύψος των ορθογωνίων το λόγο των σχετικών συχνοτήτων προς το πλάτος των κλάσεων

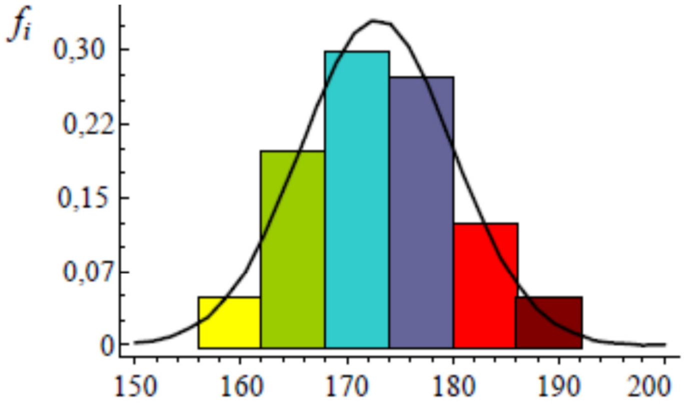


Σχήμα 8 (β)

Ιστόγραμμα σχετικών συχνοτήτων

1.9 Καμπύλες συχνοτήτων

Εάν ο αριθμός των κλάσεων είναι αρκετά μεγάλος δηλαδή τείνει στο άπειρο και το πλάτος των κλάσεων είναι πολύ μικρό δηλαδή τείνει στο μηδέν, τότε η πολυγωνική γραμμή συχνοτήτων τείνει να πάρει τη μορφή μιας ομαλής καμπύλης, η οποία ονομάζεται **καμπύλη συχνοτήτων**, όπως φαίνεται στο παρακάτω Σχήμα 9 όπου καμπύλη συχνοτήτων για το ύψος των φοιτητών.



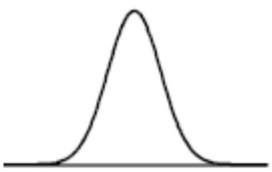
Σχήμα 9

Οι καμπύλες συχνοτήτων έχουν μεγάλη εφαρμογή στη στατιστική, οι ιδιότητες τους μπορούν να χρησιμοποιηθούν για την εξαγωγή χρήσιμων συμπερασμάτων.

Μερικές κατανομές συχνοτήτων:

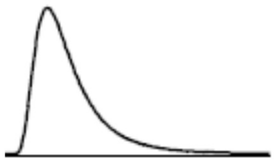
(α)

Στο σχήμα (α) οι παρατηρήσεις ‘κατανέμονται’ ομοιόμορφα σε ένα διάστημα [α,β], η κατανομή αυτή λέγεται **ομοιόμορφη**.



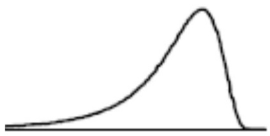
(β)

Στο σχήμα (β), η κατανομή με ‘ κωδωνοειδή’ μορφή λέγεται **κανονική κατανομή**.



(γ)

Όταν οι παρατηρήσεις δεν είναι συμμετρικά κατανεμημένες, η κατανομή λέγεται ασύμμετρη με θετική ασυμμετρία ως προς την κατανομή σχήμα (γ).



(δ)

Στο σχήμα (δ) βλέπουμε την κατανομή που λέγεται αρνητική ασυμμετρία.

ΚΕΦΑΛΑΙΟ 2: ΣΤΑΤΙΣΤΙΚΑ ΜΕΤΡΑ

2.1 Μέτρα κεντρικής τάσης

Ένα από τα σημαντικότερα μέτρα κεντρικής τάσης αλλά και γενικότερα της στατιστικής θεωρείται ο μέσος αριθμητικός (Arithmetic Mean or Average) ή αλλιώς μέση τιμή.

2.1.1 Μέσος αριθμητικός ή μέση τιμή

Ο **μέσος αριθμητικός** ή **μέση τιμή** είναι ένα από τα σπουδαιότερα μέτρα της στατιστικής. Συμβολίζεται με $\bar{x}$ όταν τα αριθμητικά δεδομένα αφορούν ένα δείγμα, και με “μ” όταν τα δεδομένα αφορούν ολόκληρο πληθυσμό. Όταν έχουμε ένα σύνολο ν παρατηρήσεων, ο αριθμητικός μέσος είναι το άθροισμα των παρατηρήσεων δια το πλήθος των παρατηρήσεων. Για τον υπολογισμό του αριθμητικου μέσου διακρίνουμε δύο περιπτώσεις:

1) Υπολογισμός αριθμητικού μέσου βάσει αταξινόμητων δεδομένων.

2) Υπολογισμός αριθμητικού μέσου βάσει ταξινομημέων δεδομένων.

Α. Υπολογισμός αριθμητικού μέσου βάσει αταξινόμητων δεδομένων

Σε περίπτωση αταξινόμητων δεδομένων ο μέσος αιρθμητικός διακρίνεται δύο τρόπους υπολογισμού. Στον απλό ή αστάθμητο μέσο αριθμητικό και στο σταθμικό μέσο αριθμητικό.

1) ΑΠΛΟΣ Η ΑΣΤΑΘΜΗΤΟΣ ΑΡΙΘΜΗΤΙΚΟΣ ΜΕΣΟΣ

Όταν έχουμε μια σειρά 10 τιμών: x_1, x_2, x_3, ..., x_{10} ο αστάθμητος αριθμητικός μέσος είναι το πηλίκο τους αθροίσματος των τιμών δια το πλήθος των τιμών, και υπολογίζεται με τον τύπο:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i.$$

ΠΑΡΑΔΕΙΓΜΑ: Έστω η μισθοδοσία 6 υπαλλήλων για το μήνα Δεκέμβριο: 1000€, 900€, 750€, 1100€, 800€, 700€.

Τότε ο αστάθμητος αριθμητικός μέσος είναι:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_6}{6} = \frac{1000 + 900 + 750 + 1100 + 800 + 700}{6} = \frac{5300}{6} = 883{,}33€.$$

Αυτό σημαίνει πως αν οι υπάλληλοι είχαν την ίδια μισθοδοσία, θα έπαιρναν και οι 6 υπάλληλοι 883,33 ο καθένας.

2) ΣΤΑΘΜΙΚΟΣ ΑΡΙΘΜΗΤΙΚΟΣ ΜΕΣΟΣ

Όταν έχουμε μια σειρά 10 τιμών: x_1, x_2, x_3, ..., x_{10} στις οποίες δίνεται διαφορετική σημασία,που εκφράζονται από κάποιους αριθμούς που λέγονται συντελεστές βαρύτητας (ή σταθμίσεως): w_1, w_2, w_3, ..., w_{10}, τότε χρησιμοποιούμε το **σταθμικό αριθμητικό μέσο** και υπολογίζεται με τον τύπο:

$$\bar{x}_w = \frac{x_1 w_1 + x_2 w_2 + \cdots + x_n w_n}{w_1 + w_2 + \cdots + w_n}.$$

ΠΑΡΑΔΕΙΓΜΑ: Εστώ οι βαθμολογίες ενός μαθητή στις πανελλαδικές εξετάσεις:

Μαθηματικά 16, Βιολογία 12, Φυσική 11, Έκθεση 12, Χημεία 15.

Και οι συντελεστές βαρύτητας των μαθημάτων, αντίστοιχα 2, 1.5, 1, 0.5, 1.

Τότε, έχουμε

$$\begin{aligned} x_1 w_1 + x_2 w_2 + x_3 w_3 + x_4 w_4 + x_5 w_5 &= 16 \cdot 2 \ + \ 12 \cdot 1{,}5 \ + \ 11 \cdot 1 \ + \ 12 \cdot 0{,}5 \ + \ 15 \cdot 1 \\ &= 32 \ + \ 18 \ + \ 11 \ + \ 6 \ + \ 15 \ = \ 82 \end{aligned}$$

και

$$w_1 + w_2 + w_3 + w_4 + w_5 = 2 \ + \ 1.5 \ + \ 1 \ + \ 0.5 \ + \ 1 \ = \ 6.$$

Άρα, ο σταθμικός αριθμητικός μέσος είναι:

$$\bar{x}_w = \frac{82}{6} = 13{,}6.$$

Β) ΥΠΟΛΟΓΙΣΜΟΣ ΑΡΙΘΜΗΤΙΚΟΥ ΜΕΣΟΥ ΒΑΣΕΙ ΤΑΞΙΝΟΜΗΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ ΣΕ ΚΑΤΑΝΟΜΗ ΣΥΧΝΟΤΗΤΩΝ ΜΕ ΑΝΙΣΑ ΔΙΑΣΤΗΜΑΤΑ ΤΑΞΕΩΝ.

Όταν τα δεδομένα που έχουμε είναι ταξινομημένα σε μια κατανομή συχνοτήτων, τότε λέμε ότι ο μέσος αριθμητικός είναι σταθμικός. Στη φάση αυτή ως τιμές της μεταβλητής (x_i) θεωρούνται οι κεντρικοί όροι των τάξεων και οι συντελεστές σταθμίσεως οι αντίστοιχες συχνότητες κάθε κατανομής (f_i). Για να υπολογίσουμε το μέσο αριθμητικό σε αυτή τη περίπτωση έχουμε 2 τρόπους υπολογισμού, τον άμεσο και τον έμμεσο.

ΑΜΕΣΟΣ ΤΡΟΠΟΣ: Έστω x_1, x_2, x_3, ..., x_{10} οι κεντρικοί όροι των τάξεων μιας κατανομής συχνοτήτων και f_1, f_2, f_3, ..., f_{10} οι αντίστοιχες συχνότητες της κατανομής. Για τον υπολογισμό, λοιπόν του μέσου αριθμητικού χρησιμοποιούμε τον τύπο:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \cdots + x_n f_n}{f_1 + f_2 + \cdots + f_n} = \frac{1}{N} \cdot \sum_{i=1}^{n} x_i f_i,$$

όπου

$$N = f_1 + f_2 + \cdots + f_n = \sum_{i=1}^{n} x_i f_i.$$

ΠΑΡΑΔΕΙΓΜΑ: Έστω το μέσο ημερομίσθιο μιάς κατανομής 100 εργατών.

Πίνακας 1

Τάξεις ημερομισθίων σε ευρώ (κλάσεις)	**Κεντρικοί όροι τάξεων x_i**	**Αριθμός εργατών f_i**	**Γινόμενα $x_i f_i$**
1500-2500	2000	5	10000
2500-3500	3000	13	39000
3500-4500	4000	20	80000
4500-5500	5000	35	175000
5500-6500	6000	18	108000
6500-7500	7000	7	49000
7500-8500	8000	2	16000
Σύνολο	------------	$N = 100$	477000

Το άθροισμα των γινομένων $x_i f_i$ δηλώνει το συνολικό ποσό της ημερήσιας μισθοδοσίας των 100 εργατών.

Με βάση τον τύπο έχουμε:

$$\bar{x} = \frac{1}{N} \cdot \sum_{i=1}^{7} x_i f_i = \frac{477000}{100} = 4700.$$

Ερμηνεία:
Αν όλοι οι εργάτες αμοίβονταν εξίσου, έπαιρναν δηλαδή το ίδιο ημερομίσθιο, τότε το ημερομίσθιο κάθε εργάτη θα ήταν 4700 ευρώ.

ΕΜΜΕΣΟΣ ΤΡΟΠΟΣ: Για τον υπολογισμό μέσου αριθμητικού με έμμεσο τρόπο ακολουθούμε την εξής διαδικασία:

α) Αντικαθιστούμε τη μεταβλητή x_i (δηλαδή τους κεντρικούς όρους των τάξεων) με μια νέα μεταβλητή ξ_i. Η διαφορά ανάμεσα στη μεταβλητή x_i με τη μεταβλητή ξ_i είναι πως η μεταβλητή x_i αποτελείται από μεγάλους αριθμούς, ενώ η μεταβλητή ξ_i αποτελείται από μονοψήφιους αριθμούς.

β) Επιλέγουμε μια τιμής της μεταβλητής x_i, και αυτή είναι συνήθως η τιμή που αντιστοιχεί στη μεγαλύτερη συχνότητα της κατανομής, την οποία συμβολίζουμε με το x_o.

γ) Στη συνέχεια, από κάθε τιμή της μεταβλητής x_i αφαιρούμε την τιμή x_o και τις διαφορές αυτών τις διαιρούμε με το διάστημα των τάξεων, το οποίο συμβολίζεται με δ. Έτσι, έχουμε τον τύπο:

$$\xi_i = \frac{x_i - x_0}{\delta}$$

δ) Έπειτα πολλαπλασιάζουμε κάθε τιμή της μεταβλητής ξ_i με τις αντίστοιχες συχνότητες ($=f_i$), και τα γινόμενα τα γράφουμε σε μια στήλη που συμβολίζεται με $\xi_i f_i$.

ε) Τέλος για να υπολογίσουμε το μέσο αριθμητικό εφαρμόζουμε τον τύπο:

$$\bar{x} = \frac{x_0 + \delta \cdot \sum \xi_i f_i}{\sum f_i}.$$

ΠΑΡΑΔΕΙΓΜΑ: Έστω το μέσο ημερομίσθιο μιάς κατανομής 100 εργατών.

Πίνακας 2

Τάξεις ημερομισθίων σε ευρώ (κλάσεις)	**Κεντρικοί όροι** x_i	**Συχνότητες** f_i	ξ_i	$\xi_i f_i$
1500-2500	2000	5	-3	-15
2500-3500	3000	13	-2	-26
3500-4500	4000	20	-1	-20
4500-5500	5000	35	0	0
5500-6500	6000	18	1	18
6500-7500	7000	7	2	14
7500-8500	8000	2	3	6
Σύνολο	---------	100	-------	-23

Με βάση την παραπάνω διαδικασία, στον πίνακα έχουμε:

$$x_0 = 5000, \qquad \delta = 1000, \qquad \sum f_i = 100, \qquad \sum \xi_i f_i = -23.$$

Σύμφωνα με τον τύπο του αριθμητικού μέσου αντικαθιστούμε τα δεδομένα και βρίσκουμε:

$$\bar{x} = \frac{x_0 + \delta \cdot \sum \xi_i f_i}{\sum f_i} = \frac{5000 + 1000 \cdot (-23)}{100} = 4770 \text{ ευρώ}.$$

Έτσι παρατηρούμε ότι και με τον άμεσο και με τον έμμεσο τρόπο έχουμε το ίδιο αποτέλεσμα.

Γ) ΥΠΟΛΟΓΙΣΜΟΣ ΑΡΙΘΜΗΤΙΚΟΥ ΜΕΣΟΥ ΒΑΣΕΙ ΤΑΞΙΝΟΜΗΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ ΣΕ ΚΑΤΑΝΟΜΗ ΣΥΧΝΟΤΗΤΩΝ ΜΕ ΑΝΙΣΑ ΔΙΑΣΤΗΜΑΤΑ ΤΑΞΕΩΝ

Σε περίπτωση που το διάστημα των ταξεων (δ) δεν είναι ίδιο για όλες τις τάξεις, τότε στον άμεσο τρόπο ακολουθούμε την ίδια διαδικασία, όπως και στην κατανομή συχνοτήτων με ίσα διαστήματα τάξεως, όταν θέλουμε να υπολογίσουμε το μέσο αριθμητικό. Στον έμμεσο τρόπο, για τον υπολογισμό του μέσου αριθμητικού ισχύουν τα παραπάνω με τη διαφορά ότι για το σχηματισμό της στήλης ξ_i, επιλέγουμε ένα διάστημα τάξεως δ, έτσι ώστε οι διαιρούμενες με το δ διαφορές $(x_i - x_0)$ να δίνουν ολιγόψηφα πηλίκα. Αλλιώς για τον υπολογισμό αριθμητικού μέσου χρησιμοποιούμε τον άμεσο τρόπο.

ΠΑΡΑΔΕΙΓΜΑ: Ο παρακάτω πίνακας μας δείχνει τη διαδικασία υπολογισμού του μέσου αριθμητικού σε περίπτωση που τα διαστήματα των τάξεων είναι άνισα. Έστω λοιπόν, το μέσο οικογενειακό εισόδημα 1000000 οικογενειών.

Πίνακας 3

Τάξεις οικογενειακού εισοδήματος σε ευρώ (κλάσεις)	**Κεντρικοί όροι** x_i	**Αριθμός οικογενειών** f_i	ξ_i	$\xi_i f_i$
0-50000	2500	7200	-1,25	-90000
50000-100000	7500	165200	-0,75	123000
100000-200000	150000	360800	0	0
200000-400000	300000	360000	1,5	540000
400000-1000000	700000	42000	5,5	231000
Σύνολο	-------	1000000	---	557100

Υπολογίζοντας, ξέρουμε ότι

$$x_0 = 150000, \qquad \delta = 100000, \qquad \sum f_i = 1000000, \qquad \sum \xi_i f_i = 557100.$$

Έτσι αντικαστώντας τα παραπάνω δεδομένα έχουμε δύο επιλογές:

1) Με τον άμεσο τρόπο το μέσο οικογενειακό εισόδημα είναι:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{205710000000}{1000000} = 205710 \text{ ευρώ}.$$

2) Με τον έμμεσο τρόπο το μέσο οικογενειακό εισόδημα είναι:

$$\bar{x} = \frac{x_0 + \delta \cdot \sum \xi_i f_i}{\sum f_i} = \frac{150000 + 100000 \cdot 557100}{1000000} = 205710 \text{ ευρώ}.$$

2.2 Μέτρα παράμετροι θέσεως

Με βάση τα προηγούμενα, είπαμε πως ο μέσος αριθμητικός είναι ένα από τα σπουδαιότερα στατιστικά μέτρα. Αυτό βέβαια συμβαίνει όταν έχουμε μια κατανομή συχνοτήτων της οποίας η καμπύλη συχνοτήτων είναι συμμετρική. Σε περίπτωση που η κατανομή συχνοτήτων είναι ασυμμετρική, τότε ο μέσος αριθμητικός δεν είναι αντιπροσωπευτικό στατιστικό μέτρο για τη λήψη ορθών αποφάσεων. Για την αντιμετώπιση της συγκεκριμένης περίπτωσης υπάρχουν άλλα στατιστικά μέτρα, τα οποία οδηγούν σε αξιόπιστες πληροφορίες για τον εξεταζόμενο στατιστικό πληθυσμό. Τα στατιστικά αυτά μέτρα ονομάζονται παράμετροι θέσεως και είναι:
α) η διάμεσος,
β) τα τερτατημόρια,
γ) τα δεκατημόρια,
δ) τα εκατοστημόρια και
ε) η επικρατούσα τιμή ή τύπος.

2.2.1 Διάμεσος

Όταν έχουμε μια σειρά 10 τιμών τις οποίες τις ταξινομούμε με αύξουσα σειρά, κατά τη φυσική τους κατάσταση, δηλαδή από τη μικρότερη προς τη μεγαλύτερη, τότε η **διάμεσος** ορίζεται ως η μεσαία τιμή όταν το 10 είναι περιττός αριθμός, ή ο μέσος όρος των δύο μεσσαίων τιμών όταν ο 10 είναι άρτιος αριθμός. Η διάμεσος συμβολίζεται με το σύμβολο "Μ" και θεωρείται το σπουδαιότερο στατιστικό μέτρο θέσεως. Για τον υπολογισμό της διαμέσου διακρίνουμε τις παρακάτω περιπτώσεις:

Α) ΥΠΟΛΟΓΙΣΜΟΣ ΔΙΑΜΕΣΟΥ ΑΤΑΞΙΝΟΜΗΤΩΝ ΔΕΔΟΜΕΝΩΝ

Σε περίπτωση που τα δεδομένα είναι αταξινόμητα και έχουν τοποθετηθεί κατά αύξουσα σειρά για τον υπολογισμό της διαμέσου διακρίνουμε 2 περιπτώσεις:
α) Όταν το πλήθος των τιμώς είναι μονός αριθμός, τότε η μεσαία τιμή που κατέχει την κεντρική θέση είναι η διάμεσος.

ΠΑΡΑΔΕΙΓΜΑ: Έστω τιμές:

121, 135, 127, 139, 114, 129, 148, 137, 161.

Αρχικά κατατάσσουμε τις τιμές κατά αύξουσα σειρά:

114, 121, 127, 129, 135, 137, 139, 148, 161.

Η κεντρική θέση των τιμών αυτών είναι

$$\frac{n+1}{2} = \frac{9+1}{2} = \frac{10}{2} = 5\,.$$

Αυτό σημαίνει ότι η διάμεσος βρίσκεται στη 5η θέση, και είναι η τιμή 135.

β) Όταν το πλήθος των τιμών είναι ζυγός αριθμός, τότε σε αυτή την περίπτωση υπάρχουν δύο τιμές που κατέχουν την κεντρική θέση και όχι μία. Ο μέσος όρος των τιμών αυτών που κάτεχουν την κεντρική θέση είναι η διάμεσος.

ΠΑΡΑΔΕΙΓΜΑ: Έστω οι παρακάτω τιμές,κατά αύξουσα σειρά:

114, 121, 127, 129, 135, 137, 139, 148.

Η κεντρική θέση των τιμών αυτών είναι:

$$\frac{n+1}{2} = \frac{8+1}{2} = 4{,}5.$$

Αυτό σημαίνει ότι η διάμεσος βρίσκεται ανάμεσα στη 4η και στη 5η θέση.

Επομένως, βάσει του τύπου, η διάμεσος είναι:

$$M = \frac{129+135}{2} = 132.$$

Β) ΥΠΟΛΟΓΙΣΜΟΣ ΔΙΑΜΕΣΟΥ ΤΑΞΙΝΟΜΗΜΕΩΝ ΔΕΔΟΜΕΝΩΝ ΣΕ ΚΑΤΑΝΟΜΗ ΣΥΧΝΟΤΗΤΩΝ

Στην περίπτωση που τα δεδομένα είναι ταξινομημένα σε N κλάσεις, ο υπολογισμός της διαμέσου γίνεται ως εξής:

1) Υπολογίζουμε τις αθροιστικές συχνότητες: $\Phi_1, \Phi_2, \dots, \Phi_N$.
2) Εντοπίζουμε μεταξύ ποιων αθροιστικών συχνοτήτων Φ_{i-1} και Φ_i βρίσκεται ο αριθμός $N/2$. Εάν ο αριθμός $N/2$ συμπίπτει με μία από τις αθροιστικές συχνότητες, τότε ως διάμεσος ορίζεται το άνω άκρο της αντίστοιχης κλάσης.
3) Χρησιμοποιούμε τον τύπο:

$$M = \alpha_{i-1} + \frac{\delta}{f_i}\left(\frac{N}{2} - \Phi_{i-1}\right).$$

όπου:

α_{i-1} : το κατώτερο όριο της τάξης i της διαμέσου,

f_i : η συχνότητα της τάξης i της διαμέσου,

Φ_{i-1} : η αθροιστική συχνότητα της τάξης $i-1$ η οποία προηγείται της τάξης i,

δ : το πλάτος της τάξης i της διαμέσου.

ΠΑΡΑΔΕΙΓΜΑ: Έστω μια κατανομή συχνοτήτων 100 εργατών.

Πίνακας 4

Τάξεις ημερομισθιών σε ευρώ (κλάσεις)	**Συχνότητες f_i**	**Φ_i**
2000-3000	6	6
3000-4000	14	20
4000-5000	21	41
5000-6000	36	74
6000-7000	18	92
7000-8000	5	97
8000-9000	3	100
Σύνολο	100	--------

Έχουμε:

- $N/2 = \frac{100}{2} = 50$
- $\alpha_{i-1} = 5000$
- $f_i = 36$
- $\Phi_{i-1} = 41$
- $\delta = 1000$

Οπότε:

$$M = \alpha_{i-1} + \frac{\delta}{f_i}\left(\frac{N}{2} - \Phi_{i-1}\right) = 5000 + \frac{1000}{36}(50-41) = 5250 \text{ ευρώ}.$$

Ερμηνεία:

Αυτό σημαίνει ότι το 50% των εργατών έχουν ημερομίσθιο μέχρι 5250 ευρώ, και το υπόλοιπο 50% έχει ημερομίσθιο άνω των 5250 ευρώ.

2.2.2 Τεταρτημόρια – Δεκατημόρια – Εκατοστημόρια

Παράμετροι θέσεως εκτός απο τη διάμεσο, είναι επίσης και τα **τεταρτημόρια** (Quartiles), τα **δεκατημόρια** (Deciles) και τα **εκατοστημόρια** (Centiles). Ο ρόλος των τεταρτημορίων είναι να υποδιαιρούν το σύνολο των τιμών σε τέσσερις ισοπληθείς ομάδες, των δεκατημορίων σε δέκα ισοπληθείς ομάδες, και των εκατοστημορίων σε εκατό ισοπληθείς ομάδες. Τα μέτρα αυτά χρησιμοποιούνται αρκετά συχνά για τη μελέτη ενός συνόλου δεδομένων.

Α) ΤΕΤΑΡΤΗΜΟΡΙΑ

Το πρώτο τεταρτημόρια συμβολίζεται με το Q_1, το δεύτερο τεταρτημόριο με Q_2 και το τρίτο τεταρτημόριο συμβολίζεται με το Q_3.

Για το Q_1, έχουμε αριστερά το πολύ 25% των παρατηρήσεων και δεξία το πολύ 75% των παρατηρήσεων. Ομοίως και για το Q_3, έχουμε αριστερά το πολύ 75% των παρατηρήσεων και δεξιά το πολύ 25% των παρατηρήσεων. Για το Q_2 το οποιό αντιστοιχεί στο 50% συμπεραίνουμε πως είναι και η διάμεσος.

Για τον υπολογισμό των τεταρτημορίων έχουμε τις εξής περιπτώσεις:

α) Όταν τα δεδομένα είναι αταξινόμητα το Q_1 το εντοπίζουμε σε ποια θέση είναι, με τον τύπο

$$\frac{n+1}{4},$$

ενώ το Q_3 το εντοπίζουμε σε ποια θέση είναι, με τον τύπο

$$\frac{3(10+1)}{4}.$$

β) Όταν τα δεδομένα είναι αταξινόμητα, τότε για το πρώτο τεταρτημόριο χρησιμοποιούμε τον τύπο

$$Q_1 = \alpha_{i-1} + \frac{\delta}{f_i}\left(\frac{N}{4} - \Phi_{i-1}\right)$$

και για το τρίτο τεταρτημόριο χρησιμοποιούμε τον τύπο

$$Q_3 = \alpha_{i-1} + \frac{\delta}{f_i}\left(\frac{3N}{4} - \Phi_{i-1}\right).$$

Σημείωση: Για τον υπολογιμό των τεταρτημορίων ακολουθούμε την ίδια διαδικασία που ακολουθήσαμε και για τη διάμεσο.

ΠΑΡΑΔΕΙΓΜΑ: Με βάση τα δεδομένα του προηγούμενου πίνακα έχουμε:

Για το πρώτο τεταρτημόριο:

- $N/4 = \frac{100}{4} = 25$
- $\alpha_{i-1} = 4000$
- $f_i = 21$
- $\Phi_{i-1} = 20$
- $\delta = 1000$

Βάσει του τύπου λοιπόν έχουμε:

$$Q_1 = \alpha_{i-1} + \frac{\delta}{f_i}\left(\frac{N}{4} - \Phi_{i-1}\right) = 4000 + \frac{1000}{21}(25 - 20) = 4238 \text{ ευρώ}.$$

Ερμηνεία:

Αυτό σημαίνει ότι τα πρώτα 25% των εργατών έχουν ημερομίσθιο μέχρι 4238 ευρώ.

Για το τρίτο τεταρτημόριο:

- $3N/4 = \frac{300}{4} = 75$
- $\alpha_{i-1} = 6000$
- $f_i = 18$
- $\Phi_{i-1} = 74$
- $\delta = 1000$

Βάσει του τύπου λοιπόν έχουμε:

$$Q_3 = \alpha_{i-1} + \frac{\delta}{f_i}\left(\frac{3N}{4} - \Phi_{i-1}\right) = 6000 + \frac{1000}{21}(75 - 74) = 6056 \text{ ευρώ.}$$

Ερμηνεία:

Αυτό σημαίνει ότι τα πρώτα 25% των εργατών παίρνουν ημερομίσθιο μέχρι 6056 ευρώ, ενώ τα υπόλοιπα 25% των εργατών παίνρουν πάνω από 6056 ευρώ.

B) ΔΕΚΑΤΗΜΟΡΙΑ ΚΑΙ ΕΚΑΤΟΣΤΗΜΟΡΙΑ

Έτσι αντίστοιχα και τα δεκατημόρια υπολογίζονται βάσει του τύπου:

$$D_k = \alpha_{i-1} + \frac{\delta}{f_i}\left(\frac{kN}{10} - \Phi_{i-1}\right), \kappa = 1,2, \dots, 9$$

και τα εκατοστημόρια βάσει του τύπου:

$$C_k = \alpha_{i-1} + \frac{\delta}{f_i}\left(\frac{kN}{100} - \Phi_{i-1}\right), \kappa = 1,2, \dots, 99$$

ΠΑΡΑΔΕΙΓΜΑ: Με βάσει πάλι τα δεδομένα του προηγούμενου πίνακα, για το 10^{o} εκατοστημόριο, έχουμε:

- $10N/100 = 10$
- $\alpha_{i-1} = 3000$
- $f_i = 14$
- $\Phi_{i-1} = 6$
- $\delta = 1000$

Βάσει του τύπου λοιπόν έχουμε:

$$C_{10} = \alpha_{i-1} + \frac{\delta}{f_i}\left(\frac{10N}{100} - \Phi_{i-1}\right) = 3000 + \frac{1000}{14}(10 - 6) = 3285 \text{ ευρώ.}$$

Ερμηνεία:

Αυτό σημαίνει ότι τα πρώτα 10% των εργατών έχουν ημερομίσθιο μέχρι 3285 ευρώ.

2.2.3 Επικρατούσα τιμή ή τύπος

Επικρατούσα τιμή ή **Τύπος** είναι η τιμή εκείνη της μεταβλητής που αντιστοιχεί στη μεγαλύτερη συχνότητα της κατανομής. Συμβολίζεται με M_o και λέγεται και αλλιώς **Σημείο Μεγαλύτερης Συχνότητας**. Για τον υπολογισμό της επικρατούσας τιμής διακρίνουμε δύο περιπτώσεις. Όταν τα δεδομένα είναι ταξινομημένα και όταν είναι αταξινόμητα.

Στην περίπτωση που τα δεδομένα είναι ταξινομημένα, τότε η επικρατούσα τιμή δίνεται από τον τύπο:

$$M_0 = \alpha_{i-1} + \delta \cdot \frac{\Delta_1}{\Delta_1 + \Delta_2}$$

όπου:

α_{i-1} : το κατώτερο όριο της τάξεως στην οποία αντιστοιχεί η μεγαλύτερη συχνότητα (f_i).

δ : το πλάτος της τάξεως με τη μεγαλύτερη συχνότητα.

$\Delta_1 = f_i - f_{i-1}$: διαφορά μεγαλύτερης συχνότητας μείον προηγούμενη.

$\Delta_2 = f_i - f_{i+1}$: διαφορά μεγαλύτερης συχνότητα μείον επόμενη.

ΠΑΡΑΔΕΙΓΜΑ: Με βάση τον προηγούμενο πίνακα έχουμε:

- $\alpha_{i-1} = 5000$
- $\delta = 1000$
- $\Delta_1 = f_i - f_{i-1} = 36 - 21 = 15$
- $\Delta_2 = f_i - f_{i+1} = 36 - 18 = 18$

Με βάσει το τύπο λοιπόν έχουμε:

$$M_0 = \alpha_{i-1} + \delta \cdot \frac{\Delta_1}{\Delta_1 + \Delta_2} = 5000 + 1000 \cdot \frac{15}{15 + 18} = 5450 \text{ ευρώ.}$$

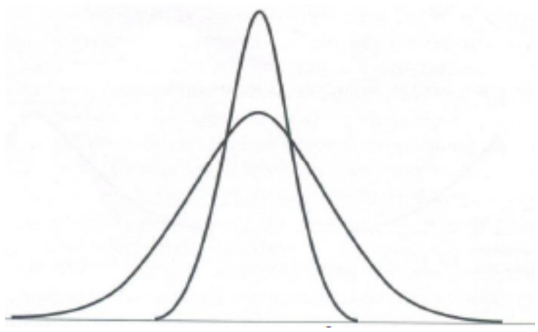
Ερμηνεία :

Αυτό σημαίνει ότι το συχνότερο ημερομίσθιο είναι περίπου 5450 ευρώ.

Στην περίπτωση που τα δεδομένα είναι αταξινόμητα τότε ο τύπος εντοπίζεται στη μεγαλύτερη συχνότητα των δεδομένων.

2.3 Μέτρα διασποράς

Τα μέτρα κεντρικής τάσης από μόνα τους δεν επαρκούν για την επαρκή περιγραφή ενός συνόλου αριθμητικών μετρήσεων. Η αντιπροσωπευτικότητά τους εξαρτάται σε μεγάλο βαθμό από την ετερογένεια που παρουσιάζουν οι μετρήσεις. Στο Σχήμα 1 εμφανίζονται δύο κατανομές εντελώς διαφορετικές η μία με την άλλη, οι οποίες όμως έχουν την ίδια μέση τιμή, διάμεσο και επικρατούσα τιμή. Είναι προφανές ότι η περιγραφή των δύο αυτών κατανομών, διά μέσου μόνο των μέτρων κεντρικής τάσης, δεν επαρκεί για την εξαγωγή ασφαλών συμπερασμάτων.



Σχήμα 1

Αυτό που διαφοροποιεί τις δύο κατανομές και δεν προσδιορίζεται από τα μέτρα κεντρικής τάσης, είναι η διασπορά των τιμών τους γύρω από το κέντρο τους. Ο προσδιορισμός αυτός γίνεται με τη βοήθεια των μέτρων διασποράς.

Για να γίνει κατανοητός ο λόγος χρήσης των μέτρων διασποράς, παραθέτουμε το ακόλουθο παράδειγμα.

ΠΑΡΑΔΕΙΓΜΑ:

Δυο όμιλοι επιχειρήσεων, που ο καθένας αποτελείται από 10 επιχειρήσεις, είχαν ετήσια έσοδα (σε χιλιάδες ευρώ) για το οικονομικό έτος 2013 τα ποσά που αναγράφονται στον παρακάτω πίνακα:

Πίνακας 5

Όμιλος Α	502	500	496	503	499	500	504	498	501	497
Όμιλος Β	500	498	510	495	500	492	501	497	503	504

Οι μέσες τιμές των εσόδων είναι:

$$\bar{x}_A = \frac{1}{10}\sum_{i=1}^{10} x_i = \frac{502 + 500 + 496 + 503 + 499 + 500 + 504 + 498 + 501 + 497}{10} = 500$$

και

$$\bar{x}_B = \frac{1}{10}\sum_{i=1}^{10} x_i = \frac{500 + 498 + 510 + 495 + 500 + 492 + 501 + 497 + 503 + 504}{10} = 500$$

Όπως προκύπτει από τις τιμές των δειγμάτων, και οι δύο όμιλοι έχουν μέση τιμή εσόδων 500000 ευρώ. Όμως τα έσοδα των επιχειρήσεων του όμιλου Α είναι από 496000 έως 504000, ενώ του ομίλου Β από 492000 μέχρι 510000. Από τα διαστήματα αυτά προκύπτει ότι τα έσοδα των επιχειρήσεων του ομίλου Α βρίσκονται κοντά στη μέση τιμή, ενώ το αντίστοιχο διάστημα για τον όμιλο Β απλώνεται σε τιμές που βρίσκονται πιο μακριά από τη μέση τιμή. Επομένως είναι φανερό ότι δεν είναι αρκετή η γνώση της μέσης τιμής.

Τα μέτρα διασποράς στοχεύουν στον προσδιορισμό της μεταβλητότητας που παρουσιάζει ένα σύνολο μετρήσεων. Τα μέτρα αυτά χρησιμοποιούνται σε συνδυασμό με τα μέτρα θέσης και από κοινού περιγράφουν τις κατανομές δεδομένων με τρόπο συμπληρωματικό. Τα μέτρα διασποράς είναι:

1) Το εύρος
2) Το ενδοτετρτημοριακό εύρος
3) Μέση απόκλιση
4) Η τυπική απόκλιση
5) Ο συντελεστής μεταβλητότητας

2.3.1 Εύρος

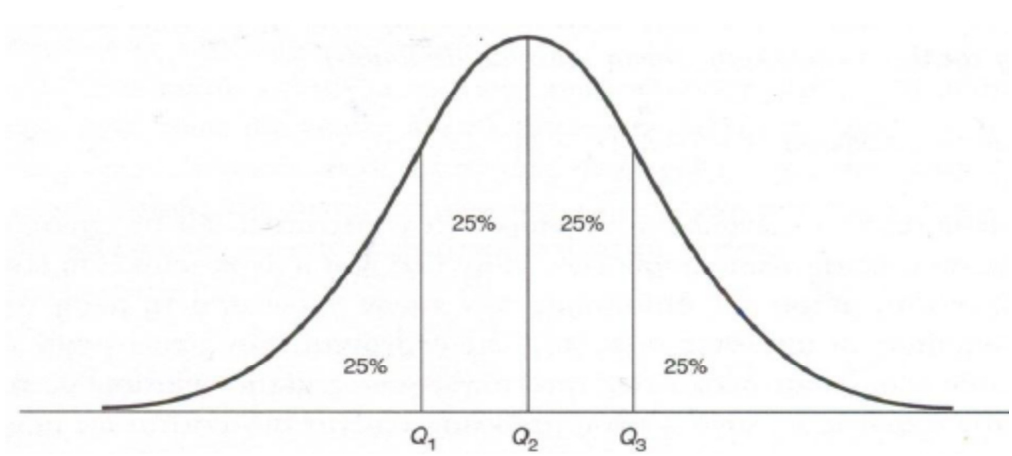
Το **εύρος** είναι το απλούστερο από όλα τα μέτρα διασποράς και ορίζεται ως η διαφορά μεταξύ μέγιστης και ελάχιστης τιμής ενός συνόλου μετρήσεων.

Εύρος R = max – min

Αν και το εύρος είναι εύκολο στον προσδιορισμό του, η χρηστικότητά του είναι εξεραιτικά περιορισμένη. Και αυτό διότι στον υπολογισμό του υπεισέρχονται δύο μόνο τιμές, οι πλέον ακραίες. Εξαιτίας αυτού του γεγονότος είναι εξαιρετικά ευαίσθητο στην ύπαρξη τιμών που διαφοροποιούνται πολύ των υπολοίπων, επομένως, ο προσδιορισμός της διασποράς των μετρήσεων δία μέσου του εύρους μπορεί να οδηγήσει σε παραπλανητικά συμπεράσματα.

2.3.2 Ενδοτεταρτημοριακό εύρος

Η διαφορά τρίτου και πρώτου τεταρτημορίου $H = Q_3 - Q_1$ ονομάζεται ενδοτεταρτημοριακό εύρος. Το ενδοτεταρτημοριακό εύρος, όπως γίνεται άμεσα αντιληπτό από τον ορισμό του, δεν επηρεάζεται από πιθανές ακραίες τιμές που μπορεί να υπάρχουν στο σύνολο των μετρήσεων (Σχήμα 2).



Σχήμα 2

2.3.3 Μέση απόκλιση

Τα μέτρα διασποράς που μέχρι στιγμής αναφέρθηκαν δεν δίνουν καμία πληροφορία για την διασπορά των δεδομένων γύρω από το μέσο της κατανομής τους. Επιπλέον, κανένα από αυτά δεν λαμβάνει υπόψη κατά τον υπολογισμό του το σύνολο των τιμών της κατανομής.

Έστω $x_1, x_2, \ldots, x_k$ οι τιμές μιας μεταβλητής X, που αφορά τα στοιχεία ενός δείγματος μεγέθους ν, $k \leq \nu$ και $\bar{x}$ η μέση τιμή. Η Η **μέση απόκλιση** ή **μέση απόλυτη απόκλιση** δίνεται από τον τύπο:

$$MAD = \frac{\nu_1 \cdot |x_1 - \bar{x}| + \nu_2 \cdot |x_2 - \bar{x}| + \cdots + \nu_k \cdot |x_k - \bar{x}|}{\nu - 1} = \frac{1}{\nu - 1}\sum_{i=1}^{k} \nu_i \cdot |x_i - \bar{x}|.$$

2.3.4 Διακύμανση και τυπική απόκλιση

Ένας εναλλακτικός τρόπος για να αποφύγουμε το πρόβλημα των προσήμων στον υπολογισμό των αποκλίσεων $x_i - \bar{x}$, και επιπλέον για να μην εμπλακούμε με τη χρήση των απόλυτων τιμών, είναι να χρησιμοποιήσουμε ως αποστάσεις των τιμών από τη μέση τιμή τους τα τετράγωνα των αποκλίσεων. Η μέση τιμή των τετραγώνων των αποκλίσεων ονομάζεται **διακύμανση.**

Η διακύμανση όταν υπολογίζεται στο σύνολο των στοιχείων ενός πληθυσμού, δίνεται από τον τύπο:

$$\sigma^2 = \frac{\nu_1 \cdot (x_1 - \mu)^2 + \nu_2 \cdot (x_1 - \mu)^2 + \cdots + \nu_k \cdot (x_1 - \mu)^2}{N} = \frac{1}{N}\sum_{i=1}^{k} \nu_i (x_i - \mu)^2.$$

όπου $x_1, x_2, \dots, x_k$ οι πληθυσμιακές τιμές, μ η πληθυσμιακή μέση τιμή και N το πλήθος των στοιχείων του πληθυσμού.

Σε περίπτωση δειγματικών δεδομένων, ο υπολογισμός της διακύμανσης γίνεται με την βοήθεια του τύπου:

$$s^2 = \frac{\nu_1 \cdot (x_1 - \bar{x})^2 + \nu_2 \cdot (x_1 - \bar{x})^2 + \cdots + \nu_k \cdot (x_1 - \bar{x})^2}{\nu - 1} = \frac{1}{\nu - 1}\sum_{i=1}^{k} \nu_i (x_i - \bar{x})^2.$$

όπου $x_1, x_2, \dots, x_k$ οι δειγματικές τιμές, $\bar{x}$ η δειγματική μέση τιμή και ν το μέγεθος του δείγματος.

Ο λόγος για τον οποίο, στον τύπο της δειγματικής διακύμανσης αντί της διαίρεσης του αθροίσματος των τετραγώνων των αποκλίσεων διά του ν χρησιμοποιείται το $\nu - 1$, είναι η υποεκτίμηση που προκύπτει για τη διακύμανση του πληθυσμού, όταν αυτή εκτιμηθεί από τα δειγματικά δεδομένα. Η διακύμανση εκφράζει τον ίδιο τύπο πληροφορίας με τη μέση απόκλιση, έχει όμως ορισμένες πολύ σημαντικές ιδιότητες, όπως π.χ. η δυνατότητά της να αναλύεται σε επιμέρους συνιστώσεις. Ο υπολογισμός της απλουστεύεται όταν χρησιμοποιείται ο τύπος:

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

Ο τύπος αυτός προκύπτει από τον αρχικό τύπο της διακύμανσης ως εξής:

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

$$= \frac{1}{\nu - 1}\left[\sum_{i=1}^{k} \nu_i x_i^2 - 2\bar{x}\sum_{i=1}^{k} \nu_i x_i + \sum_{i=1}^{k} \nu_i \bar{x}^2\right] = \frac{1}{\nu - 1}\left[\sum_{i=1}^{k} \nu_i x_i^2 - 2\nu\bar{x}^2 + \nu\bar{x}^2\right]$$

$$= \frac{1}{\nu - 1}\left[\sum_{i=1}^{k} \nu_i x_i^2 - \nu\bar{x}^2\right] = \frac{1}{\nu - 1}\left[\sum_{i=1}^{k} \nu_i x_i^2 - \frac{1}{\nu}\left(\sum_{i=1}^{k} \nu_i x_i\right)^2\right].$$

Η διακύμανση εκφράζεται σε μονάδες, οι οποίες είναι τα τετράγωνα των αρχικών μονάδων της εξεταζόμενης μεταβλητής. Για να έχουμε ένα μέτρο το οποίο μετράει τη μεταβλητότητα μιας μεταβλητής και να εκφράζεται στις ίδιες μονάδες που εκφράζεται η μεταβλητή, παίρνουμε την τετραγωνική ρίζα της διακύμανσης. Η τετραγωνική ρίζα της διακύμανσης ονομάζεται **τυπική απόκλιση** και στην πράξη είναι το ευρύτερα χρησιμοποιούμενο μέτρο διασποράς. Η τυπική απόκλιση χρησιμοποιούμενη από κοινού με τη μέση τιμή, μπορούν να περιγράψουν αποτελεσματικά την κατανομή ενός συνόλου μετρήσεων. Η μέση τιμή ορίζει το σημείο γύρω από το οποίο τείνουν να συσσωρεύονται οι τιμές μιας κατανομής, ενώ η τυπική απόκλιση προσδιορίζει το βαθμό της διασποράς των τιμών γύρω από το σημείο αυτό.

Ο τύπος της τυπικής απόκλισης είναι:

$$s = \sqrt{s^2}.$$

Όσο μεγαλύτερη είναι η τιμή της τυπικής απόκλισης, τόσο μεγαλύτερη είναι η μεταβλητικότητα των παρατηρήσεων από το μέσο αριθμητικό, δηλαδή οι παρατηρήσεις είναι ευρέως διασκορπισμένες περί το μέσο αριθμητικό, αντιστρόφως, μικρή τυπική απόκλιση δηλώνει ότι οι τιμές βρίσκονται κοντά στο μέσο όρο.

2.3.5 Συντελεστής μεταβλητότητας

Επειδή, η τυπική απόκλιση έχει μονάδες μέτρησης οι οποίες είναι ίδιες με τις μονάδες του μεγέθους στο οποίο αναφέρεται, στερείται νοήματος η σύγκριση τυπικών αποκλίσεων μεταβλητών που μετρώνται σε διαφορετικές μονάδες. Δεν έχει νόημα για παράδειγμα να συγκρίνουμε την τυπική απόκλιση μετρήσεων βάρους και μετρήσεων θερμοκρασίας. Για τη σύγκριση της μεταβλητότητας, χρησιμοποιείται ο **συντελεστής μεταβλητότητας**. Ο συντελεστής μεταβλητότητας είναι ένα σχετικό μέτρο διασποράς και εκφράζει την τυπική απόκλιση ενός συνόλου μετρήσεων ως ποσοστό (%) επί της μέσης τιμής και δίνεται από τον τύπο:

$$CV = \frac{s}{\bar{x}}, \quad \bar{x} \neq 0.$$

Συμπερασματικά μπορούμε να πούμε ότι όσο πιο μικρός είναι ο συντελεστής μεταβλητότητας, τόσο μεγαλύτερη ομοιογένεια υπάρχει στις τιμές της μεταβλητής. Γενικά, δεχόμαστε ότι ένα δείγμα θα είναι ομοιογενές, εάν ο συντελεστής μεταβλητότητας δεν είναι μεγαλύτερος του 10%.

ΚΕΦΑΛΑΙΟ 3: ΠΑΛΙΝΔΡΟΜΗΣΗ ΚΑΙ ΣΥΣΧΕΤΙΣΗ

3.1 Διαγράμματα διασποράς

Έστω X και Y δύο μεταβλητές με τιμές $x_1, x_2, \ldots, x_n$ και $y_1, y_2, \ldots, y_n$, αντίστοιχα. **Διάγραμμα διασποράς** ονομάζουμε την παράσταση σε ορθογώνιο σύστημα αξόνων όλων των σημείων με συντεταγμένες (x_i, y_i). Τα σημεία αυτά σχηματίζουν ένα 'νέφος' ή 'σμήνος' σημείων, από την προσεκτική παρατήρηση του οποίου μπορούμε να πάρουμε πληροφορίες για τη σχέση εξάρτησης που ενδεχομένως υπάρχει μεταξύ των μεταβλητών X και Y.

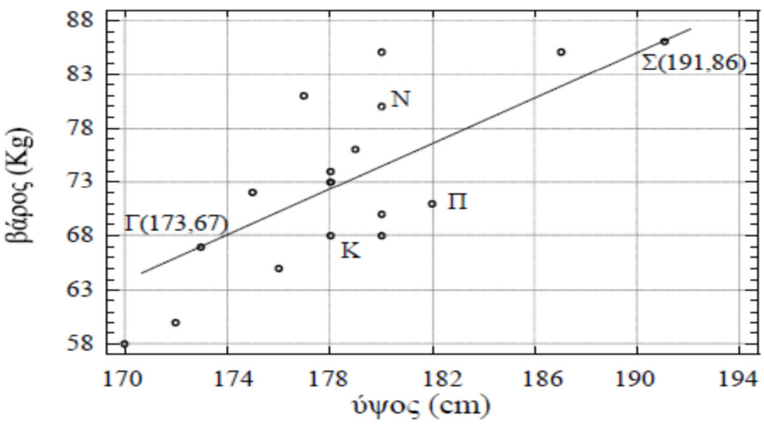
Ο παρακάτω πίνακας δίνει τα ύψη X (σε cm) και τα βάρη Y (σε kg) των 18 αγοριών της Γ' Λυκείου.

Πίνακας 1

Λίστα υψών (σε cm) και βάρος (σε kg) των 18 αγοριών

Μαθητής	Ύψος X	Βάρος Y	Μαθητής	Ύψος X	Βάρος Y
Α	170	58	Κ	178	68
Β	172	60	Λ	179	76
Γ	173	67	Μ	180	68
Δ	175	72	Ν	180	80
Ε	176	65	Ξ	180	70
Ζ	177	81	Ο	180	85
Η	178	73	Π	182	71
Θ	178	74	Ρ	187	85
Ι	178	73	Σ	191	86

Στο παράδειγμα αυτό έχουμε την περίπτωση όπου σε κάθε άτομο (μαθητή) γίνονται δύο μετρήσεις. Δηλαδή το δείγμα αποτελείται από τα ζεύγη τιμών των συνεχών μεταβλητών X (ύψος) και Y (βάρος).



Σχήμα 1

Διάγραμμα διασποράς

Στο παράδειγμα αυτό παρατηρούμε ότι οι ψηλοί μαθητές είναι συνήθως και πιο βαρείς. Η προσεκτική παρατήρηση ενός διαγράμματος διασποράς μπορεί να μας δώσει σημαντικές πληροφορίες για τη σχέση εξάρτησης που ενδεχομένως υπάρχει μεταξύ των μεταβλητών τις οποίες εξετάζουμε.

3.2 Απλή γραμμική παλινδρόμηση

Η απλή γραμμική παλινδρόμηση ποσοτικοποιεί τη σχέση δύο συνεχών τυχαίων μεταβλητων Χ και Υ, υπό τη μορφή ενός γραμμικού υποδείγματος, από το οποίο οι τιμές της μιας μεταβλητηής προβλέπονται (ή εκτιμώνται) από τις τιμές της άλλης. Αν οι τιμές της μεταβλητής Υ εκτιμώνται από τις τιμές της μεταβλητής Χ, τότε η Υ ονομάζεται **εξαρτημένη μεταβλητή** και η Χ **ανεξάρτητη μεταβλητή**.

3.2.1 Μέθοδος ελαχίστων τετραγώνων

Από το παραπάνω διάγραμμα διασποράς φαίνεται καθαρά ότι υπάρχει μία σχέση ανάμεσα στο βάρος και στο ύψος των 18 μαθητών. Τα σημεία (x_i, y_i) είναι συγκεντρωμένα περίπου γύρω από μια ευθεία, δηλαδή η σχέση μεταξύ των Χ και Υ είναι κατά προσέγγιση γραμμική. Η ευθεία που θα προσαρμόζεται καλύτερα στα σημεία αυτά καλείται **ευθεία παλινδρόμησης** της Υ πάνω στη Χ. Όπως γνωρίζουμε, η εξίσωση μιας ευθείας δίνεται από τη σχέση:

$$y = a + bx$$

όπου a και b είναι παράμετροι τις οποίες θέλουμε να υπολογίσουμε ή όπως λέμε, να ΄΄εκτιμήσουμε΄΄, έτσι ώστε η ευθεία που θα προκύψει να μας δίνει όσο το δυνατόν την καλύτερη περιγραφή της σχέσης (εξάρτησης) που υπάρχει μεταξύ των μεταβλητών X και Y.

Η πιο απλή διαδικασία προσαρμογής μιας ευθείας γραμμής σε ένα διάγραμμα διασποράς είναι με το μάτι. Αυτή όμως έχει πολλά μειονεκτήματα παρά την απλότητα της. Το κυριότερο είναι η έλλειψη αντικειμενικότητας, αφού διάφορα άτομα μπορούν να χαράξουν διαφορετικές μεταξύ τους ευθείες. Χρειαζόμαστε λοιπόν μια ακριβέστερη μέθοδο για την προσαρμογή μιας ευθείας γραμμής σε τέτοιου είδους δεδομένα. Μια μέθοδος που χρησιμοποιείται για την εκτίμηση των παραμέτρων a και b, άρα και για την εύρεση της εξίσωσης της καλύτερης ευθείας που προσαρμόζεται στα δεδομένα, είναι η ΄΄ μέθοδος ελαχίστων τετραγώνων ΄΄.

Η μέθοδος των ελαχίστων τετραγώνων συνίσταται στον προσδιορισμό των παραμέτρων a και b, έτσι ώστε να ελαχιστοποιείται το άθροισμα των τετραγώνων των κατακόρυφων αποστάσεων των σημείων $(x_i, y_i), i = 1,2, \ldots, n$ από την ευθεία $y = a + bx$, δηλαδή το

$$\sum_{i=1}^{n}(y_i - a - bx_i)^2 \qquad (1)$$

να γίνεται ελάχιστο.

Οι τιμές των παραμέτρων a και b, που ελαχιστοποιούν την (1), καλούνται **εκτιμήτριες ελαχίστων τετραγώνων**, συμβολίζονται με $\hat{a}$ και $\hat{b}$, αντιστοίχως και αποδεικνύεται ότι δίνονται από τις σχέσεις:

$$\hat{b} = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x}.$$

Η ευθεία

$$\hat{y} = \hat{a} + \hat{b}x$$

καλείται **ευθεία ελαχίστων τετραγώνων** ή **ευθεία παλινδρόμησης** της Y πάνω στη X.

3.2.2 Ερμηνεία των εκτιμητριών ελαχίστων τετραγώνων

Στην εξίσωση ελαχίστων τετραγώνων $\hat{y} = \hat{a} + \hat{b}x$ η τιμή της εκτιμήτριας $\hat{a}$ της παραμέτρου a παριστάνει την τεταγμένη του σημείου στο οποίο η ευθεία τέμνει τον άξονα y΄y, δηλαδή την τιμή της εξαρτημένης μεταβλητής Y όταν $x = 0$.

Έστω τώρα δύο τιμές x_1 και $x_2 = x_1 + 1$ της ανεξάρτητης μεταβλητής. Τότε, λαμβάνοντας τη διαφορά των αντίστοιχων προβλεπόμενων τιμών της εξαρτημένης μεταβλητής βρίσκουμε:

$$\hat{y}_2 - \hat{y}_1 = \hat{a} + \hat{b}x_2 - \left(\hat{a} + \hat{b}x_1\right) = \hat{a} + \hat{b}(x_1 + 1) - \left(\hat{a} + \hat{b}x_1\right) = \hat{b}.$$

Δηλαδή:

$$\hat{y}_2 = \hat{y}_1 + \hat{b}.$$

Συνεπώς, ο συντελεστής διεύθυνσης $\hat{b}$ της ευθείας $\hat{y} = \hat{a} + \hat{b}x$ παριστάνει τη μεταβολή της εξαρτημένης μεταβλητής Y όταν το X μεταβληθεί κατά μία μονάδα. Έτσι, όταν το X αυξηθεί κατά μία μονάδα, τότε το $\hat{y}$ αυξάνεται κατά $\hat{b}$ μονάδες όταν $\hat{b} > 0$ ή ελαττώνεται κατά $\hat{b}$ μονάδες όταν $\hat{b} < 0$.

3.2.3 Συντελεστή γραμμικής συσχέτισης του Pearson

Ο **συντελεστής γραμμικής συσχέτισης** δύο μεταβλητών X και Y ορίζεται με βάση ένα δείγμα n ζευγών παρατηρήσεων $(x_i, y_i), i = 1,2,\dots,n$, συμβολίζεται με $r(\mathrm{X},\mathrm{Y})$ ή απλά με r και δίνεται από τον τύπο:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

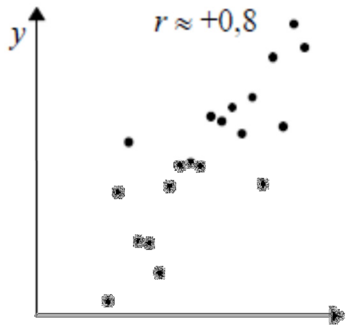
Από τον ορισμό του r παρατηρούμε ότι για μεγάλες τιμές x_i της X και y_i της Y (μεγαλύτερες από τη μέση τιμή τους) οι διαφορές $x_i - \bar{x}$ και $y_i - \bar{y}$ είναι θετικές, οπότε το γινόμενό τους είναι θετικό. Όμοια και για τις μικρές τιμές x_i και y_i οι διαφορές $x_i - \bar{x}$ και $y_i - \bar{y}$ είναι αρνητικές, οπότε το γινόμενό τους είναι πάλι θετικό. Επομένως, όταν σε μεγάλες τιμές της μεταβλητής Χ αντιστοιχούν και μεγάλες τιμές της Υ, ή σε μικρές τιμές της Χ αντιστοιχούν μικρές τιμές της Υ, τότε ο συντελεστής συσχέτισης είναι θετικός και λέμε ότι Χ και Υ είναι θετικά συσχετισμένες.

Ο συντελεστής συσχέτισης είναι καθαρός αριθμός, δηλαδή δεν εκφράζεται σε συγκεκριμένες μονάδες μέτρησης, επομένως είναι ανεξάρτητος των χρησιμοποιούμενων μονάδων μέτρησης των μεταβλητών Χ και Υ. Επιπλέον ισχύει πάντοτε ότι:

$$-1 \leq r \leq 1.$$

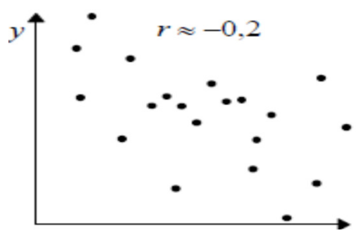
Πιο συγκεκριμένα όταν:

- $0 < r \leq 1$, τότε οι Χ και Υ είναι **θετικά** γραμμικά συσχετισμένες (Σχήμα 2 (α)).



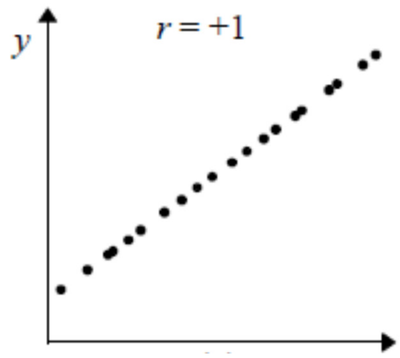
Σχήμα 2 (α)

- $-1 \leq r < 0$, τότε οι Χ και Υ είναι **αρνητικά** γραμμικά συσχετισμένες (Σχήμα 2 (β)).



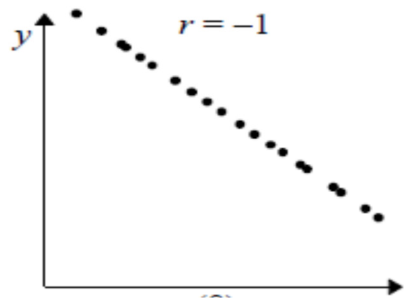
Σχήμα 2 (β)

- $r = 1$, τότε έχουμε **τέλεια θετική γραμμική συσχέτιση** και όλα τα σημεία βρίσκονται πάνω σε μία ευθεία με θετική κλίση (Σχήμα 2 (γ)).



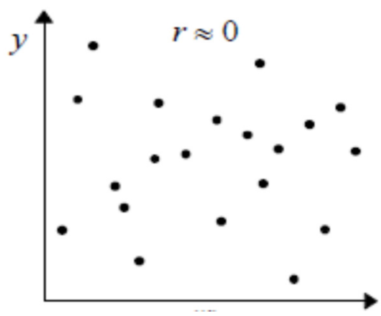
Σχήμα 2 (γ)

- $r = -1$, τότε έχουμε **τέλεια αρνητική γραμμική συσχέτιση** και όλα τα σημεία βρίσκονται πάνω σε μία ευθεία με αρνητική κλίση (Σχήμα 2 (δ)).



Σχήμα 2 (δ)

- $r = 0$, τότε δεν υπάρχει γραμμική συσχέτιση μεταξύ των μεταβλητών. Οι μεταβλητές X και Y είναι γραμμικά ασύσχετιστες (Σχήμα 2 (ε)).



Σχήμα 2 (ε)

Αποδεικνύεται ότι ο συντελεστής γραμμικής συσχέτισης r δίνεται ισοδύναμα και από τον παρακάτω τύπο, η χρήση του οποίου διευκολύνει συχνά τους υπολογισμούς κυρίως στην περίπτωση που οι $\bar{x}, \bar{y}$ δεν είναι ακέραιοι:

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}}.$$

Όταν ο συντελεστής συσχέτισης είναι κοντά στο -1 ή 1, η γραμμική συσχέτιση των μεταβλητών X και Y είναι ισχυρή (συνήθως χαρακτηρίζουμε ισχυρές τις συσχετίσεις όταν $|r| > 0.9$) ενώ όταν είναι κοντά στο 0 οι μεταβλητές X και Y είναι πρακτικά ασυσχέτιστες.

3.2.4 Συντελεστής προσδιορισμού

Στην προηγούμενη παράγραφο μιλήσαμε για το συντελεστή συσχέτισης (= r), ο οποίος μετράει την ένταση (το βαθμό) της εξαρτήσεως μεταξύ των μεταβλητών X και Y, όταν η σχέση εξαρτήσεως είναι γραμμικής μορφής. Θέλουμε τώρα να καθορίσουμε ένα στατιστικό μέτρο, το οποίο θα μας δείξει όχι μόνο το συνδετικό κρίκο ματαξύ παλινδρομήσεως και συσχετίσεως, αλλά θα δώσει μια κατάλληλη ερμηνεία της τιμής των συντελεστών συσχετίσεως. Το στατιστικό αυτό μέτρο ονομάζεται **συντελεστής προσδιορισμού**.

Ένας τρόπος για να αξιολογήσουμε την προσαρμογή της ευθείας των ελάχιστων τετραγώνων είναι να υπολογίσουμε το συντελεστή προσδιορισμού. Ο συντελεστής προσδιορισμού της δειγματικής ευθείας της παλινδρόμησης, συμβολίζεται με R^2, ορίζεται ως το τετράγωνο του δειγματικού συντελεστή συσχέτισης, δηλαδή:

$$R^2 = r^2.$$

Επειδή ο δειγματικός συντελεστής συσχέτισης παίρνει τιμές στο διάστημα $[-1,1]$, ο συντελεστής προσδιορισμού παίρνει τιμές στο διάστημα $[0,1]$. Όταν $R^2 = 1$, όλα τα σημεία που αναπαριστούν τις δειγματικές τιμές των X και Y βρίσκονται τοποθετημένα επί της ευθείας των ελαχίστων τετραγώνων. Όταν $R^2 = 0$, δεν υπάρχει γραμμική σχέση μεταξύ των δειγματικών τιμών των X και Y.

Αποδεικνύεται ότι ισχύει ο τύπος:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}.$$

Πράγματι, έχουμε:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{\sum_{i=1}^{n}(\hat{a} + \hat{b}x_i - \hat{a} - \hat{b}\bar{x})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} =$$

$$b^2 \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]^2 \left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}\right] =$$

$$\frac{\left[\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right]^2}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]} = r^2.$$

Ο συντελεστής προσδιορισμού, επομένως, μπορεί να ερμηνευθεί ως το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής το οποίο εξηγείται από την ανεξάρτητη μεταβλητή.

ΚΕΦΑΛΑΙΟ 4: ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΕΠΙΧΕΙΡΗΣΕΙΣ

4.1 Ο ρόλος της στατιστικής στις επιχειρήσεις

Η πρώτη εφαρμογή στατιστικών μεθόδων για την αντιμετώπιση και τη λύση προβλημάτων του επιχειρηματικού τομέα έγινε στις αρχές του 20 αιώνα Η.Π.Α. Με τον τρόπο αυτό είχε ως συνέπεια, τη συγκέντρωση προσωπικού και υποκαταστημάτων σε όλη την έκταση αυτής της χώρας. Τα τελευταία χρόνια υπάρχει μεγάλη ανάπτυξη βιομηχανικών και εμπορικών επιχειρήσεων στις Η.Π.Α. Με στατιστικές μεθόδους αυτό οφείλεται στην καλύτερη οργάνωση της παραγωγής και της διανομής των εμπορευμάτων. Γενικότερα η εφαρμογή των στατιστικών μεθόδων αφορά κυρίως εμπορικές και βιομηχανικές επιχειρήσεις.

Η εφαρμογή της στατιστικής δεν είναι ιδιαίτερα απαραίτητη για έναν επιχειρηματία ο οποίος έχει τον έλεγχο της επιχείρησης, γνωρίζει όλο το μηχανισμό της επιχείρησης και τους συνεργάτες του προσωπικά, και γενικότερα φροντίζει ο ίδιος για τη διαχείριση της. Αντίθετα για έναν γενικό διευθυντή ο οποίος απασχολεί εκατοντάδες άτομα, τα οποία βρίσκονται συνήθως μακριά από το κέντρο της επιχείρησης, η στατιστική θεωρείται χρησιμότατο πληροφοριακό εργαλείο για τη λήψη ορθών επιχειρηματικών αποφάσεων και την άσκηση κοινωνικής,οικονομικής και τιμολογιακής πολιτικής.

Ο ρόλος του στατιστικού μιας μεγάλης επιχείρησης είναι να συγκεντρώνει το στατιστικό υλικό, να το παρουσιάζει σε πίνακες και διαγράμματα, και να υπολογίζει τους απαραίτητους στατιστικούς δείκτες. Τα στοιχεία λοιπόν αυτά, ο γενικός διευθυντής, έχει την υποχρέωση να τα επεξεργαστεί και να τα χρησιμοποιήσει για τη λήψη ορθών επιχειρηματικών αποφάσεων.

Οι κυριότερες στατιστικές δραστηριότητες μιας μεγάλης επιχείρησης είναι:

α) Το τμήμα Παραγωγής,

β) Το τμήμα Προσωπικού,

γ) Το τμήμα Marketing,

δ) Το τμήμα Οικονομικού ελέγχου και

ε) Το τμήμα Οικονομικών ή Στατιστικών μελετών.

Το Τμήμα Παραγωγής έχει στόχο το Στατιστικό έλεγχο ποιότητας των παραγόμενων προϊόντων. Η τεχνική της μαζικής παραγωγής απαιτεί την ικανοποίηση ορισμένων προδιαγραφών, που σημαίνει ότι τα παραγόμενα προϊόντα πρέπει να έχουν ορισμένες διαστάσεις, χημικές συνθέσεις κτλ, οι οποίες δεν πρέπει να διαφέρουν απο το προκαθορισμένο πρότυπο προϊόν. Εφαρμόζοντας τις μεθόδους του στατιστικού ελέγχου ποιότητας, έχουμε τη δυνατότητα να εντοπίσουμε με ακρίβεια τις αποκλίσεις ανάμεσα στο παραγόμενο προϊόν και στο πρότυπο κόστος. Οι αποκλίσεις αυτές δημιουργούνται κατά τη διάρκεια της παραγωγικής διαδικασίας. Έτσι με την εφαρμογή των μεθόδων ρυθμίζουμε τις συνθήκες της παραγωγής, με σκοπό τον περιορισμό των αποκλίσεων σε λογικά πλαίσια, και με το τρόπο αυτό μειώνουμε το ποσοντό των ελαττωματικών προϊόντων, μειώνουμε το κόστος παραγωγής, αυξάνουμε τη παραγωγικότητα, φροντίζουμε για τη διατήρηση της ποιότητας των καλών παραγόμενων προϊόντων, αυξάνουμε την κατανάλωση, και τέλος καταλήγουμε στην αύξηση κερδών της επιχείρησης.

Το τμήμα Προσωπικού ασχολείται γενικότερα με την παρακολούθηση του προσωπικού. Δηλαδή, με τις αμοιβές του προσωπικού, τους μισθούς, τα ημερομίσθια, τα επιδόματα κτλ. Με το χρόνο εργασίας, ο οποίος και καθορίζει τους μισθούς κτλ. Με τις ικανότητες και τη δύναμη του προσωπικού,ανάλογα με τη κατηγορία που εντάσσεται ο κάθε εργαζόμενος (εργάτες, μαθητευόμενοι, υπάλληλοι κτλ.)

Το τμήμα Marketing είναι μία από τις πιό απαραίτητες λειτουργίες των επιχειρήσεων και έχει σκοπό την έρευνα των προτιμήσεων των καταναλωτών. Οι επιχειρήσεις προκειμένου να εντοπίσουν τις προτιμήσεις και τις ανάγκες των καταναλωτών σε περίπτωση που σκοπεύουν στην παραγωγή νέων προϊόντων, διενεργούν σε δείγματοληπτικές έρευνες.

Το τμήμα Οικονομικού ελέγχου τηρεί ένα σύστημα πρότυπου κόστους για την άσκηση τιμολογιακής πολιτικής. Συνδιάζει λογιστικές και στατιστικές μεθόδους για τη κατάρτιση του προϋπολογισμού του επόμενου οικονομικού έτους, και περιλαμβάνει: πρώτες ύλες, εργατικό δυναμικό, πωλήσεις, καθαρά κέρδη κα.

Και τέλος, το τμήμα Οικονομικών ή Στατιστικών μελετών αναλύει γενικές επιχειρηματικές προβλέψεις των επιχειρηματικών δραστηριοτήτων για διάφορους οικονομικούς παράγοντες, όπως για παράδειγμα για τις τιμές των προϊόντων. Συντονίζει τη στατιστική εργασία άλλων τμημάτων της επιχείρησης και για τη σωστή λειτουργία του τμήματος απαιτεί οικονομολόγο ή στατιστικό.

4.2 Εφαρμογές της στατιστικής στο χώρο των επιχειρήσεων

Εφαρμογή 1

Το πολύγωνο συχνοτήτων της κατανομής Χ των ετήσιων μισθών (σε εκατοντάδες Ευρώ) ενός δείγματος εργαζομένων, ομαδοποιημένης σε κλάσεις ίσου πλάτους, έχει κορυφές τα σημεία:

Α(20,0), Β(40,5), Γ(60,10), Δ(80,20), Ε(100,30), Ζ(120, ν_5), Η(140,10), Θ(160,0).

Η κατακόρυφη γραμμή με εξίσωση x = 100 διαιρεί το χωρίο που ορίζεται από το πολύγωνο συχνοτήτων και τον οριζόντιο άξονα σε δύο ισεμβαδικά χωρία.

Ε1. Να αποδείξετε ότι $\nu_5 = 25$.

Ε2. Να κατασκευάσετε το ιστόγραμμα συχνοτήτων της κατανομής.

Ε3. Να υπολογίσετε τις τιμές των μέτρων θέσης της κατανομής.

Ε4. Αν σαν «όριο φτώχιας» θεωρήσουμε τον μισθό των 7.200 ευρώ, να εκτιμήσετε το ποσοστό επί τοις % των φτωχών του δείγματος.

Λύση

Ε1. Έστω α το άκρο της πρώτης κλάσης και c το πλάτος κάθε κλάσης. Γνωρίζουμε ότι στο πολύγωνο συχνοτήτων θεωρούμε αριστερά και δεξιά δύο υποτιθέμενες κλάσεις με μηδενικές συχνότητες και συνδέουμε τα μέσα των άνω βάσεων του ορθογωνίου. Επομένως, έχουμε:

$$\alpha - 20 = \frac{c}{2}$$

και

$$\alpha + 6c + \frac{c}{2} = 160.$$

Λύνουμε το σύστημα και έχουμε:

$$\begin{cases} \alpha + \dfrac{13c}{2} = 160 \\ \alpha - \dfrac{c}{2} = 20 \end{cases} \Leftrightarrow \begin{cases} c = 20 \\ \alpha = 30 \end{cases}$$

Έτσι οι κλάσεις είναι:

[30,50), [50,70), [70,90), [90,110), [110,130), [130,150).

Γνωρίζουμε ότι το εμβαδόν του χωρίου που περικλίεται από τον οριζόντιο άξονα και το πολύγωνο συχνοτήτων ισούται με το μέγεθος του δείγματος.

Έχουμε: $\nu_1 = 5$, $\nu_2 = 10$, $\nu_3 = 20$, $\nu_4 = 30$, $\nu_6 = 10$.

Επειδή η ευθεία x = 100 χωρίζει το χωρίο σε δύο ισοδύναμα χωρία έχουμε ότι:

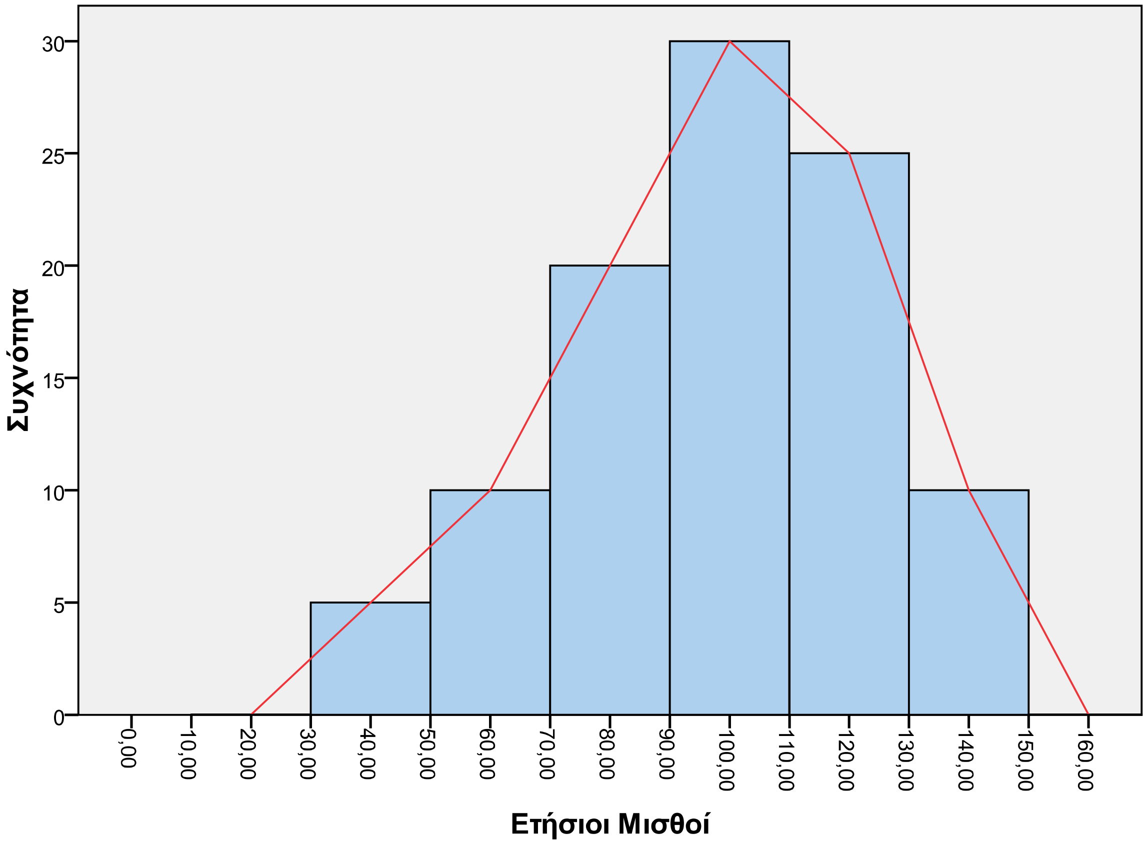
$$\nu_1 + \nu_2 + \nu_3 + \frac{\nu_4}{2} = \frac{\nu_4}{2} + \nu_5 + \nu_6 \Leftrightarrow 5 + 10 + 20 = \nu_5 + 10 \Leftrightarrow \nu_5 = 25.$$

Ε2. Φτιάχνουμε τον πίνακα και στη συνέχεια σχηματίζουμε το ακόλουθο πολύγωνο συχνοτήτων.

Πίνακας 1

Κλάσεις	Κέντρα κλάσεων x_i	ν_i	$x_i\nu_i$	N_i
[30,50)	40	5	200	5
[50,70)	60	10	600	15
[70,90)	80	20	1600	35
[90,110)	100	30	3000	65
[110,130)	120	25	3000	90
[130,150)	140	10	1400	100
Σύνολο		100	9800	

Διάγραμμα 1



Ε3. Αφού η κατακόρυφη γραμμή με εξίσωση x = 100 διαιρεί το χωρίο που ορίζεται από το πολύγωνο συχνοτήτων και τον οριζόντιο άξονα σε δύο ισεμβαδικά χωρία, έχουμε ότι η διάμεσος είναι το 100.

Η μέση τιμή είναι:

$$\bar{x} = \frac{1}{100} \cdot \sum_{i=1}^{6} x_i \nu_i = \frac{9800}{100} = 98.$$

Ε4. Ψάχνουμε να βρούμε το ποσοστό των ατόμων, που παίρνουν κάτω από 7.200€. Στο διάστημα [70,90) βρίσκεται το 20%, οπότε θεωρώντας ότι κατανέμονται ομοιόμορφα στο διάστημα [70,72) θα βρίσκεται το 2%. Επομένως, το συνολικό ποσοστό που ζεί κάτω από το όριο της φτώχειας είναι:

5% + 10% + 2% = 17%.

Εφαρμογή 2

Τα κέρδη σε ευρώ μιας αλυσίδας καταστημάτων ειδών διατροφής ακολουθούν περίπου την κανονική κατανομή. Γνωρίζουμε ότι το 84% των καταστημάτων έχουν κέρδη λιγότερα από 1200 ευρώ, ενώ το 97,5% των καταστημάτων έχουν κέρδη πάνω από 600 ευρώ.

Ε1. Να υπολογίσετε τη μέση τιμή, την τυπική απόκλιση και τη διάμεσο των κερδών.

Ε2. Να υπολογίσετε τη διακύμανση και να προσεγγίσετε το εύρος των κερδών.

Ε3. Μπορεί το σύνολο των καταστημάτων της αλυσίδας να θεωρηθεί ομοιογενές ως προς τα κέρδη; Αν το δείγμα δεν είναι ομοιογενές, κατά ποια σταθερή ποσότητα πρέπει να αυξηθούν τα κέρδη των καταστημάτων για να γίνει το δείγμα ομοιογενές;

Λύση

Ε1. Επειδή το 84% των καταστημάτων έχουν κέρδη λιγότερα από 1200 ευρώ, έχουμε:

$$\bar{x} + s = 1200.$$

Ενώ επειδή το 97,5% των καταστημάτων έχουν κέρδη πάνω από 600 ευρώ, έχουμε:

$$\bar{x} - 2s = 600.$$

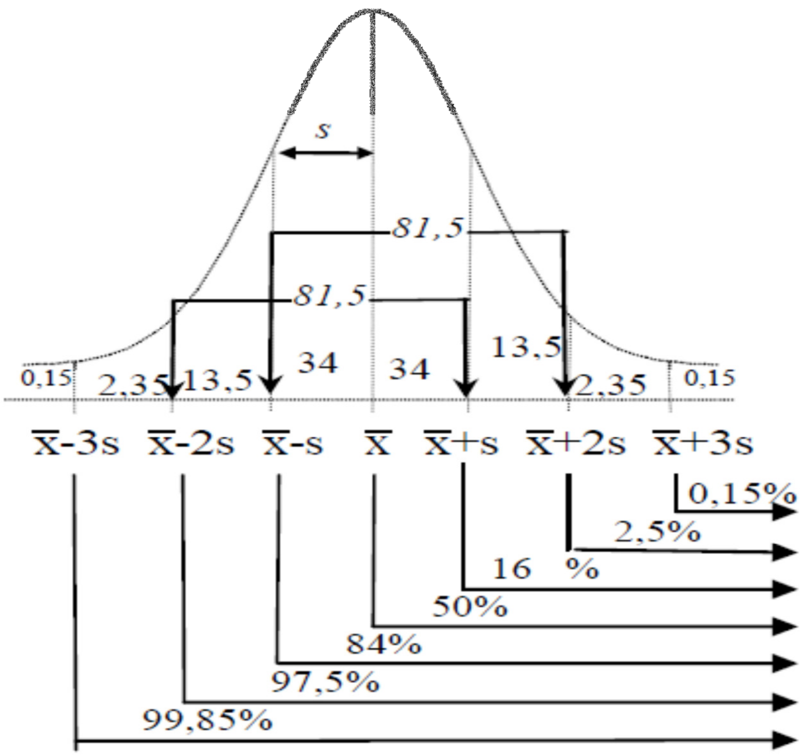
Λύνουμε το σύστημα και έχουμε:

$$\begin{cases} \bar{x} - 2s = 600 \\ \bar{x} + s = 1200 \end{cases} \Leftrightarrow \begin{cases} s = 200 \\ \bar{x} = 1000 \end{cases}$$

Επειδή έχουμε κανονική κατανομή,

$$M = \bar{x} = 1000.$$

Διάγραμμα 2



Ε2. Έχουμε:

$$s^2 = 40000$$

και

$$R = 6s = 1200.$$

Ε3. Ισχύει

$$CV = \frac{s}{\bar{x}} = \frac{200}{1000} = 0{,}2 > 0{,}1.$$

Άρα, το σύνολο των καταστημάτων της αλυσίδας δεν είναι ομοιογενές.

Έστω ότι κατά $c > 0$ πρέπει να αυξηθούν τα κέρδη των καταστημάτων, οπότε τα νέα κέρδη θα είναι

$$y_i = x_i + c.$$

Τότε, έχουμε:

$$\bar{y} = \bar{x} + c = 1000 + c$$

και

$$s_Y = s_X = 200.$$

Ο νέος συντελεστής μεταβλητότητας για να είναι το δείγμα ομοιογενές πρέπει να είναι:

$$CV \leq 0{,}1 \Leftrightarrow \frac{s_Y}{\bar{y}} \leq 0{,}1 \Leftrightarrow \frac{200}{1000 + c} \leq 0{,}1$$

$$\Leftrightarrow 100 + 0{,}1c \geq 200 \Leftrightarrow 0{,}1c \geq 200 \Leftrightarrow c \geq 1000.$$

Επομένως, για να συμβαίνει αυτό η μικρότερη τιμή της σταθεράς είναι η $c = 1000$.

Εφαρμογή 3

Τα ψύγεια μιας εταιρείας συντήρησης τροφίμων είναι κατανεμημένα σε 4 κλάσεις σύμφωνα με τη θερμοκρασία τους X (σε ^oC) η οποία κυμαίνεται από -4^oC έως 4^oC. Αν η δεύτερη κλάση έχει τριπλάσιο αριθμό ψυγείων από την πρώτη και η τέταρτη πενταπλάσιο της πρώτης τότε:

Ε1. Να παρασταθούν τα δεδομένα σε πίνακα συχνοτήτων και να δειχθεί ότι η μέση της θερμοκρασίας των ψυγείων είναι $\bar{x} = 1^oC$.

Ε2. Έστω ότι η τρίτη κλάση έχει ίδιο αριθμό ψυγείων με την πρώτη.

α) Να υπολογίσετε τη διάμεσο θερμοκρασία.

β) Αν γνωρίζουμε ότι η θερμοκρασία 34 ψυγείων είναι μικρότερη των 0.5^oC, να βρεθεί ο αριθμός των ψυγείων που κατέχει η εταιρεία.

Λύση

Ε1. Έστω c το πλάτος κάθε κλάσης, τότε

$$c = \frac{R}{k} = \frac{4 - (-4)}{4} = 2.$$

Έχουμε ότι η δεύτερη κλάση έχει τριπλάσιο αριθμό ψυγείων από την πρώτη και η τέταρτη πενταπλάσιο της πρώτης, οπότε: $\nu_2 = 3\nu_1$ και $\nu_4 = 5\nu_1$.

Έτσι σχηματίζουμε τον ακόλουθο πίνακα:

Πίνακας 2

Κλάσεις	Κέντρο Κλάσης x_i	ν_i	$x_i\nu_i$
[-4,-2)	-3	ν_1	$-3\nu_1$
[-2,0)	-1	$3\nu_1$	$-3\nu_1$
[0,2)	1	ν_3	ν_3
[2,4)	3	$5\nu_1$	$15\nu_1$
Σύνολα		$9\nu_1 + \nu_3$	$9\nu_1 + \nu_3$

Έχουμε:

$$\bar{x} = \frac{9\nu1 + \nu3}{9\nu1 + \nu3} = 1^{o}C.$$

Ε2.

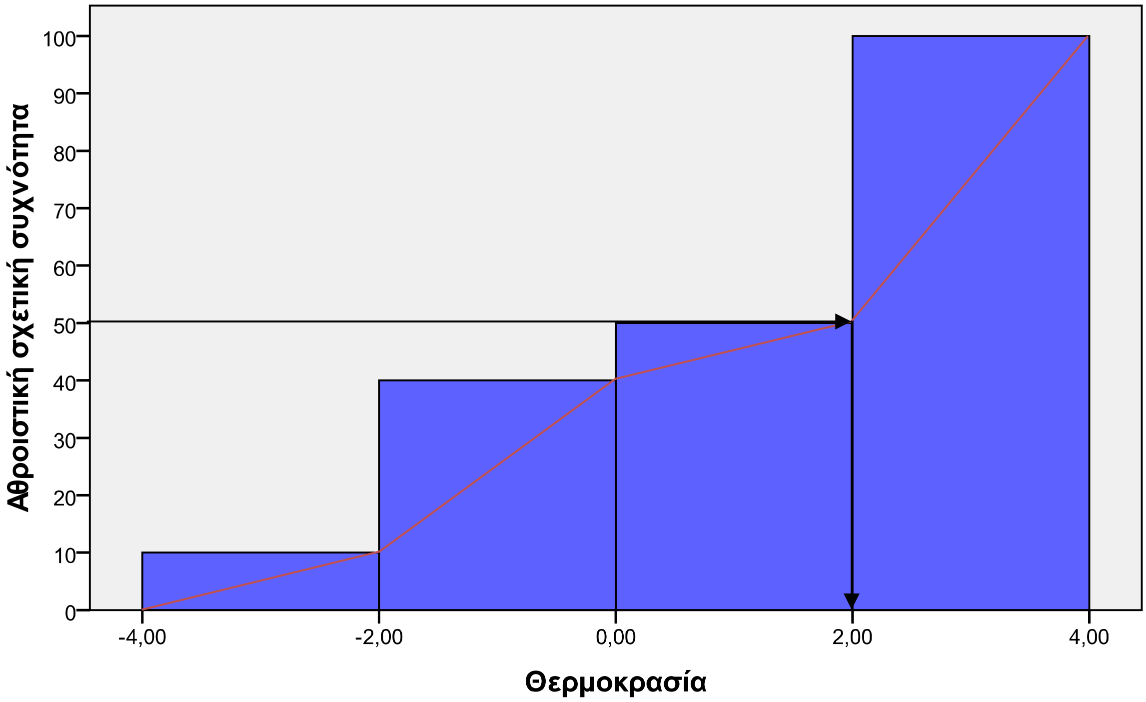
α) Αν $\nu_1 = \nu_3$ ο πίνακας γίνεται:

Πίνακας 3

Κλάσεις	Κέντρο Κλάσης x_i	ν_i	$x_i\nu_i$	f_i	$F_i\%$
[-4,-2)	-3	ν_1	$-3\nu_1$	0,1	10
[-2,0)	-1	$3\nu_1$	$-3\nu_1$	0,3	40
[0,2)	1	ν_1	ν_1	0,1	50
[2,4)	3	$5\nu_1$	$15\nu_1$	0,5	100
Σύνολα		$10\nu_1$	$10\nu_1$		

Όταν έχουμε ομαδοποιημένες τιμές η διάμεσος βρίσκεται γραφικά. Στο επόμενο σχήμα είναι το ιστόγραμμα και το πολύγωνο αθροιστικών σχετικών συχνοτήτων.

Διάγραμμα 3



Στον κάθετο άξονα φέρνουμε από το $F_i\% = 50\%$ παράλληλη ευθεία στον οριζόντιο άξονα. Στο σημείο που τέμνει το πολύγωνο φέρνουμε κάθετη ευθεία στον άξονα των τιμών x_i. Στο σημείο που τέμνει τον άξονα της μεταβλητής είναι η τιμή της διαμέσου.

Άρα, έχουμε $M = 2$.

β) Στο [0,2) βρίσκεται το 10% των ψυγείων, οπότε θεωρώντας ότι τα ψυγεία είναι ομοιόμορφα κατανεμημένα στις κλάσεις, στο [0,0.5) θα βρίσκεται το 2,5% των ψυγείων. Επομένως στο [-4,0.5) βρίσκεται το 42,5% των ψυγείων. Έχουμε ότι 42,5% των ψυγείων είναι 34. Οπότε, συνολικά ψυγεία είναι:

$$34\frac{100}{42,5} = 80.$$

Εφαρμογή 4

Σε 10 καταστήματα στην επαρχία συναντήσαμε τις παρακάτω τιμές πώλησης ενός προϊόντος (σε λεπτά) 74, 78, 76, 70, 80, 74, 76, 78, 72, 72.

Ε1. Να βρείτε τα παρακάτω μέτρα για το παραπάνω δείγμα:

α) Μέση τιμή

β) Διάμεσο

γ) Εύρος

δ) Διακύμανση και

ε) Να κρίνετε αν το παραπάνω δείγμα είναι ομοιογενές.

Αν για τα ίδια προϊόντα, από έρευνα σε 15 καταστήματα της Αθήνας, οι τιμές πώλησης (σε λεπτά) βρέθηκε ότι έχουν μέση τιμή 70. Να βρείτε

Ε2. Τη μέση τιμή πώλησης του προϊόντος για όλα τα καταστήματα της Αθήνας και της επαρχίας.

Ε3. Ποια πρέπει να είναι η μεγαλύτερη τιμή της τυπικής απόκλισης για την τιμή πώλησης του προϊόντος σε όλα τα καταστήματα, ώστε το συνολικό δείγμα να παραμένει ομοιογενές;

Λύση

Ε1. α) Έχουμε:

$$\bar{x} = \frac{74 + 78 + 76 + 70 + 80 + 74 + 76 + 78 + 72 + 72}{10} = \frac{750}{10} = 75.$$

β) Σε αύξουσα σειρά οι παρατηρήσεις είναι:

70, 72, 72, 74, 74, 76, 76, 78, 78, 80

Το πλήθος τους είναι άρτιο και συνεπώς η διάμεσός τους είναι το ημιάθροισμα των δύο μεσαίων παρατηρήσεων, δηλαδή:

$$M = \frac{74 + 76}{2} = 75.$$

γ) Το εύρος είναι:

$$R = 80 - 70 = 10.$$

δ) Σύμφωνα με τα δεδομένα έχουμε τον παρακάτω πίνακα:

Πίνακας 4

x_i	ν_i	$\nu_i x_i$	$(x_i - \bar{x})^2$	$\nu_i(x_i - \bar{x})^2$
70	1	70	25	25
72	2	144	9	18
74	2	144	1	2
76	2	152	1	2
78	2	156	9	18
80	1	80	25	25
Σύνολα	10	750		90

Οπότε:

$$s^2 = \frac{1}{10-1}\sum_{i=1}^{6} \nu_i (x_i - \bar{x})^2 = \frac{90}{9} = 9.$$

ε) Έχουμε:

$$CV = \frac{s}{\bar{x}} = \frac{3}{75} = 0{,}04 \leq 0{,}1$$

και συνεπώς το δείγμα είναι ομοιογενές.

Ε2. Αν x_i οι τιμές στα προϊόντα της επαρχίας και y_i οι τιμές στα προϊόντα της Αθήνας, τότε η μέση τιμή της πώλησεις είναι:

$$\bar{z} = \frac{10\bar{x} \ \ + 15\bar{y}}{10+15} = \frac{10 \cdot 75 \ \ + 15 \cdot 70}{10+15} = 72.$$

Ε3. Για να είναι ομοιογενές ένα δείγμα πρέπει:

$$CV \leq 0{,}1 \Leftrightarrow \frac{s_Z}{\bar{z}} \leq 0{,}1 \Leftrightarrow \frac{s_Z}{\overline{72}} \leq 0{,}1 \Leftrightarrow s_Z \leq 7{,}2.$$

Άρα, η μεγαλύτερη τιμή της τυπικής απόκλισης για την τιμή πώλησης του προϊόντος σε όλα τα καταστήματα, ώστε το συνολικό δείγμα να παραμένει ομοιογενές είναι 7,2.

Εφαρμογή 5

Τα επενδυμένα κεφάλαια Χ και τα πραγματοποιηθέντα κέρδη Υ δέκα επιχειρήσεων κατά το έτος 2013 είχαν ως εξής (σε χιλιάδες ευρώ):

x_i	30	40	120	250	50	150	220	60	100	200
y_i	5	7	15	28	10	20	27	8	14	24

Ε1. Να κατασκευαστεί το διάγραμμα διασποράς των δύο μεταβλητών.

Ε2. Να προσδιορισθεί η ευθεία γραμμικής παλινδρόμησης.

Ε3. Να βρεθεί ο συντελεστής γραμμικής συσχέτισης r και να σχολιασθεί το αποτέλεσμα.

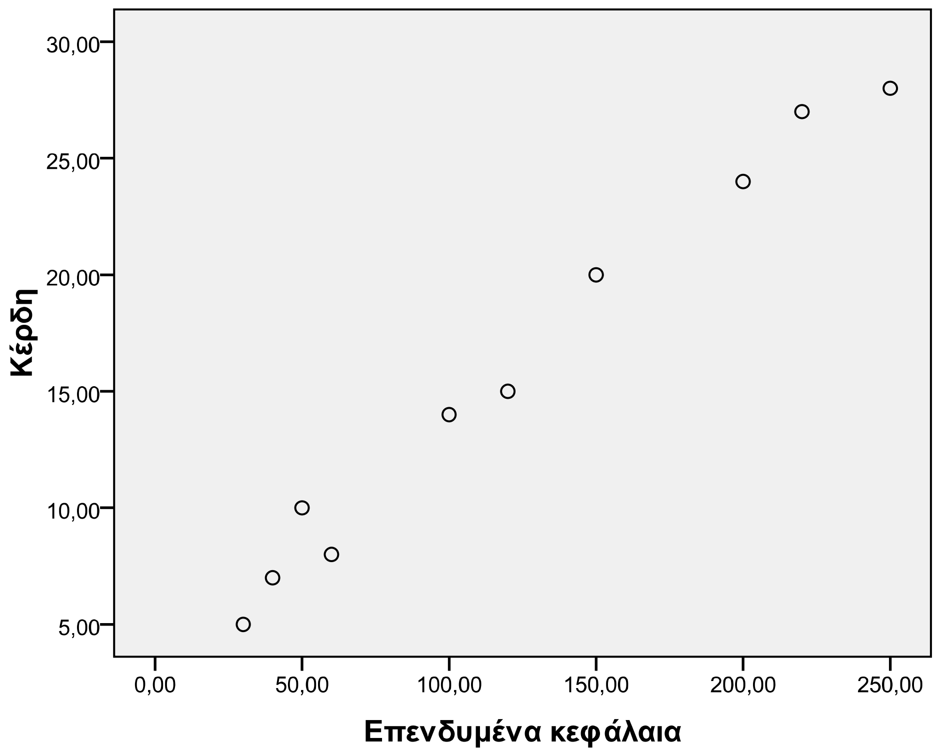
Ε4. Να βρεθούν οι διαφορές μεταξύ των πραγματικών τιμών της εξαρτημένης μεταβλητής Υ και των αντίστοιχων εκτιμώμενων τιμών.

Ε5. Ποιο είναι το κέρδος που περιμένουμε να έχει μια επιχείρηση, εάν είναι γνωστό ότι έχει επενδύσει 150 χιλιάδες ευρώ;

Λύση

Ε1. Το διάγραμμα διασποράς δίνεται παρακάτω:

Διάγραμμα 5



Ε2. Φτιάχνουμε τον παρακάτω πίνακα:

Πίνακας 5

ΕΠΙΧΕΙΡΗΣΗ	x_i	y_i	$x_i y_i$	x_i^2	y_i^2	$\hat{y}$
1	30	5	150	900	25	1,6105
2	40	7	280	1600	49	0,6485
3	120	15	1800	14400	225	-7,0475
4	250	28	7000	62500	784	-19,5535
5	5	10	50	25	100	4,0155
6	150	20	3000	22500	400	-9,9335
7	220	27	5940	48400	729	-16,6675
8	60	8	480	3600	64	-1,2755
9	100	14	1400	10000	196	-5,1235
10	200	24	4800	40000	576	-14,7435
$n = 10$	**1175**	**158**	**24900**	**203925**	**3148**	**-**

Από τον παραπάνω πίνακα έχουμε:

$$\sum_{i=1}^{10} x_i = 1175, \sum_{i=1}^{10} y_i = 158, \sum_{i=1}^{10} x_i y_i = 24900, \sum_{i=1}^{10} x_i^2 = 203925, \sum_{i=1}^{10} y_i^2 = 3148.$$

Επίσης, έχουμε:

$$\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = \frac{1175}{10} = 117{,}5$$

και

$$\bar{y} = \frac{1}{10}\sum_{i=1}^{10} y_i = \frac{158}{10} = 15{,}8.$$

Η ευθεία γραμμικής παλινδρόμησης είναι:

$$\hat{y} = \hat{a} + \hat{b}x,$$

όπου

$$\hat{b} = \frac{10\sum_{i=1}^{10} x_i y_i - \left(\sum_{i=1}^{10} x_i\right)\left(\sum_{i=1}^{10} y_i\right)}{10\sum_{i=1}^{10} x_i^2 - \left(\sum_{i=1}^{10} x_i\right)^2}$$

$$= \frac{10 \cdot 24900 - 1175 \cdot 158}{10 \cdot 203925 - 1175^2} = 0{,}0962$$

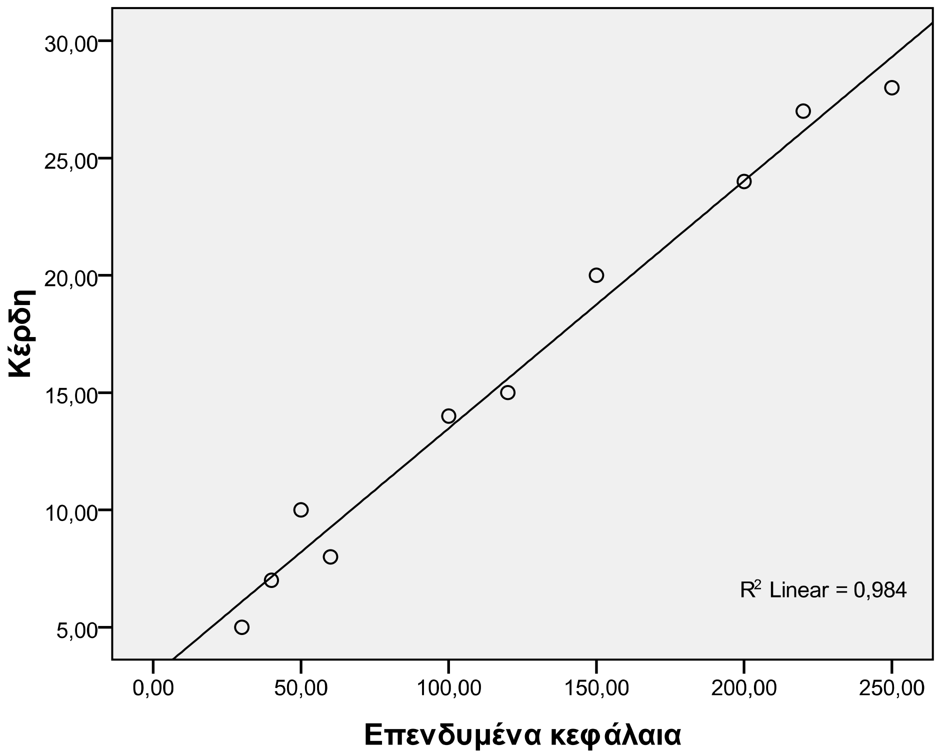
και

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 15{,}8 - 0{,}0962 \cdot 117{,}5 = 4{,}4965.$$

Επομένως, η ζητούμενη εξίσωση παλινδρόμησης είναι:

$$\hat{y} = 4{,}4965 + 0{,}0962x.$$

Διάγραμμα 6



- Η σταθερά $\hat{a} = 4{,}4965$ προσδιορίζει (θεωρητικά) τα πραγματοποιηθέντα κέρδη που αντιστοιχούν στην τιμή $\mathrm{X} = 0$.
- Η σταθερά $\hat{b} = 0{,}0962$ δείχνει ότι όταν τα επενδυμένα κεφάλαια αυξηθούν κατά 1000 ευρώ, τότε τα πραγματοποιηθέντα κέρδη θα αυξηθούν κατά 96.2 ευρώ.

Ε3. Ο συντελεστής γραμμικής συσχέτισης είναι:

$$r = \frac{10\sum_{i=1}^{10} x_i y_i - \left(\sum_{i=1}^{10} x_i\right)\left(\sum_{i=1}^{10} y_i\right)}{\sqrt{10\sum_{i=1}^{10} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{10\sum_{i=1}^{10} y_i^2 - \left(\sum_{i=1}^{10} y_i\right)^2}}$$

$$= \frac{10 \cdot 24900 - 1175 \cdot 158}{\sqrt{10 \cdot 203925 - 1175^2}\sqrt{10 \cdot 3148 - 158^2}} = 0{,}992.$$

Επειδή η τιμή του συντελεστή συσχέτισης είναι πολύ κοντά στη μονάδα, υπάρχει ισχυρή θετική συσχέτιση μεταξύ των μεταβλητών Χ και Υ.

Ε4. Οι διαφορές $\hat{u}_i = y_i - \hat{y}_i$ μεταξύ των πραγματικών τιμών της εξαρτημένης μεταβλητής Υ και των αντίστοιχων εκτιμώμενων τιμών δίνονται στον παρακάτω πίνακα:

Πίνακας 6

ΕΠΙΧΕΙΡΗΣΗ	x_i	y_i	$y_i - \hat{y}_i$
1	30	5	3,3895
2	40	7	6,3515
3	120	15	22,0475
4	250	28	47,5535
5	5	10	5,9845
6	150	20	29,9335
7	220	27	43,6675
8	60	8	9,2755
9	100	14	19,1235
10	200	24	38,7435
$\boldsymbol{n = 10}$	**1175**	**158**	**-**

Ε5. Το κέρδος που περιμένουμε να έχει μια επιχείρηση, εάν είναι γνωστό ότι έχει επενδύσει 150 χιλιάδες ευρώ θα βρεθεί από την εξίσωση παλινδρόμησης για $x = 150$. Έχουμε:

$$\hat{y} = 4{,}4965 + 0{,}0962 \cdot 150 = 18{,}96265.$$

Άρα, το κέρδος που περιμένουμε να έχει μια επιχείρηση, εάν είναι γνωστό ότι έχει επενδύσει 150 χιλιάδες ευρώ είναι 18962,65 ευρώ.

Εφαρμογή 6

Μια εταιρεία έχει εγκαταστημένα 20 υποκαταστήματα (σημεία πώλησης των προϊόντων της) σε επιλεγμένες περιοχές, κατά το δυνατόν ισοδύναμες μεταξύ τους από πλευράς ευκαιριών και δυνατοτήτων ως προς το ετήσιο ύψος του κύκλου εργασιών τους (ετήσιου τζίρου).

Η εταιρεία προσέλαβε τη φετινή περίοδο ένα νέο γενικό διευθυντή, ο οποίος αναδιοργάνωσε την επιχείρηση και εφάρμοσε νέες επιστημονικές μεθόδους, στην οργάνωση και διοίκηση (management) καθώς και στον τομέα των πωλήσεων της εταιρείας. Μετά συμπλήρωση της νέας οικονομικής χρήσης ο γενικός διευθυντής αποφασίζει να μελετήσει τα έσοδα πωλήσεων των 20 υποκαταστημάτων. Έτσι θα αντλήσει συμπεράσματα, που θα συγκρίνει με αυτά της περσινής περιόδου, για να μπορέσει να εξάγει κάποια τελικά συμπεράσματα σχετικά με το τι προέκυψε μετά την αναδιοργάνωση της επιχείρησης.

Οι πωλήσεις (σε εκατομμύρια ευρώ) των 20 υποκαταστημάτων της φετινής χρήσης ήταν:

105	112	115	118	123	123	124	125	127	128
132	133	134	136	138	138	142	145	149	156

Ως προς τη περσινή χρήση γνωρίζουμε ότι η μέση τιμή των πωλήσεων ήταν $\bar{x}_1 = 90000000$ ευρώ και η τυπική απόκλιση $s_1 = 10000000$ ευρώ. Να εξετασθεί εάν η επιχείρηση βελτίωσε την οικονομική της θέση.

Λύση

Δημιουργούμε αρχικά τον παρακάτω πίνακα και υπολογίζουμε τη νέα μέση τιμή και τη νέα τυπική απόκλιση:

Πίνακας 7

Κλάσεις Πωλήσεων	Αριθμός Υποκ/των ν_i	x_i	x_i^2	$x_i\nu_i$	$x_i^2\nu_i$
[100,110)	1	105	11025	105	11025
[110,120)	3	115	13225	345	39675
[120,130)	6	125	15625	750	93750
[130,140)	6	135	18225	814	109350
[140,150)	3	145	21025	435	63075
[150,160)	1	155	24025	155	24025
Σύνολο	**20**			**2600**	**340900**

Η νέα μέση τιμή των πωλήσεων της εταιρείας θα είναι:

$$\bar{x}_2 = \frac{1}{20}\sum_{i=1}^{6} x_i\nu_i = \frac{2600}{20} = 130 \text{ εκατομμύρια ευρώ.}$$

Η διακύμανση των πωλήσεων θα είναι:

$$s_2^2 = \frac{1}{19}\left[\sum_{i=1}^{6} \nu_i x_i^2 - 20\bar{x}^2\right] = \frac{1}{19}\left(340900 - 20 \cdot 130^2\right) = 152{,}63.$$

Συνεπώς, η νέα τυπική απόκλιση είναι:

$$s_2 = \sqrt{s^2} = \sqrt{152{,}63} \cong 12{,}35 \text{ εκατομμύρια ευρώ.}$$

Εδώ έχουμε ένα σημείο που χρειάζεται προσοχή. Δεν πρέπει να αφεθούμε στην πρώτη εικόνα των τυπικών αποκλίσεων και να πούμε ότι επειδή η φετινή τυπική απόκλιση $s_2 = 12{,}35$ είναι μεγαλύτερη από της περσινής $s_1 = 10$, η διασπορά των εσόδων φέτος παρουσιάζει χειρότερη εικόνα. Για να είμαστε σίγουροι προχωράμε στον υπολογισμό και των αντίστοιχων συντελεστών μεταβλητότητας CV_1 και CV_2 και έχουμε:

$$CV_1 = \frac{s_1}{\bar{x}_1} = \frac{10}{90} = 0{,}111$$

και

$$CV_2 = \frac{s_2}{\bar{x}_2} = \frac{12{,}35}{130} = 0{,}095.$$

Από τη μέχρι τώρα μελέτη μπορούμε να πούμε ότι η επιχείρηση βελτίωσε την οικονομική της θέση, διότι:

I. Έχουμε σημαντική αύξηση της μέσης τιμής των πωλήσεων από 90000000 ευρώ πέρυσι 130000000 ευρώ φέτος.

II. Η διασπορά των πωλήσεων φέτος παρουσιάζει καλύτερη εικόνα, δηλαδή έχουμε καλύτερη ομοιογένεια ως προς τα έσοδα, αφού:

$$CV_1 = 0{,}111 < 0{,}095 = CV_2.$$

Πίνακας 8

Στατιστικά Μέτρα	**Περσινή χρήση (1)**	**Φετινή χρήση (2)**	**Μεταβολή των (1) και (2)**
Μέση Τιμή	90	130	Αύξηση
Τυπική Απόκλιση	10	12,35	Αύξηση
Διακύμανση	100	152,63	Αύξηση
Συντελ. Μεταβλητότητας	0,111	0,095	Μείωση

Εφαρμογή 7

Είναι γνωστό ότι ένα τμήμα των εμπορευμάτων που πωλούνται από τις επιχειρήσεις επιστρέφεται σε μερικές περιπτώσεις, από τους αγοραστές στον πωλητή για διάφορους λόγους. Για παράδειγμα, επειδή δεν τηρήθηκαν οι προδιαγραφές που είχαν συμφωνηθεί για το εμπόρευμα, επειδή καθυστέρησε πολύ η παράδοση του εμπορεύματος, ή επειδή υπάρχουν ελαττωματικά εμπορεύματα κ.τ.λ. Μια εμπορική επιχείρηση Α είχε την περασμένη χρονιά πωλήσεις εμπορευμάτων αξίας 500 εκατομμύρια ευρώ και επιστροφές εμπορευμάτων αξίας 50 εκατομμύρια ευρώ. Ο νέος γενικός διευθυντής θέλοντας να εξακριβώσει εάν οι επιστροφές των εμπορευμάτων γίνονται σε όρια επιτρεπτά, αναθέτει στο διευθυντή πωλήσεων τη διερεύνηση του θέματος.

Η διεύθυνση πωλήσεων, ανάμεσα στα άλλα στοιχεία που συνέλεξε για μελέτη και διερεύνηση του θέματος, συγκέντρωσε και τις περσινές ετήσιες πωλήσεις εμπορευμάτων με τις αντίστοιχες επιστροφές για τις 10 πρώτες σε πωλήσεις επιχειρήσεις του ίδιου κλάδου εμπορίας. Τα στοιχεία παρουσιάζονται στον παρακάτω πίνακα (σε εκατομμύρια ευρώ).

Αξία πωλήσεων x_i	20	30	40	40	50	50	60	70	80	90
Αξία επιστροφών y_i	1	3	3	4	5	6	6	8	9	10

Ο διευθυντής πωλήσεων θέλησε να διαπιστώσει:

Ε1. Το βαθμό συσχέτισης που υπάρχει μεταξύ της αξίας των εμπορευμάτων που πωλούνται ετησίως και της αξίας αυτών που επιστρέφονται στην ομάδα των 10 επιχειρήσεων.

Ε2. Την εξίσωση παλινδρόμησης που θα μπορούσε να εκφράσει τη σχέση που υπάρχει μεταξύ της αξίας των πωλήσεων και της αξίας των επιστροφών κατά έτος.

Ε3. Ποια θα ήταν η αναμενόμενη αξία των εμπορευμάτων που επιστρέφονται, αν οι πωλήσεις αυξάνονταν κατά μια μονάδα, δηλαδή κατά 10 εκατομμύρια ευρώ.

Ε4. Ποια θα ήταν η αξία των επιστροφών που θα είχε μια επιχείρηση, αν οι ετήσιες πωλήσεις της είχαν ύψος 35 εκατομμύρια ευρώ.

Λύση

Ε1. Για να υπολογίσουμε το συντελεστή γραμμικής συσχέτισης των μεταβλητών Χ και Υ συμπληρώνουμε τον παρακάτω πίνακα:

Πίνακας 9

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
20	1	20	400	1
30	3	90	900	9
40	3	120	1600	9
40	4	160	1600	16
50	5	250	2500	25
50	6	300	2500	36
60	6	360	3600	36
70	8	560	4900	64
80	9	720	6400	81
90	10	900	8100	100
530	**55**	**3480**	**32500**	**377**

Από τον παραπάνω πίνακα έχουμε:

$$\sum_{i=1}^{10} x_i = 530, \sum_{i=1}^{10} y_i = 55\,, \sum_{i=1}^{10} x_i y_i = 3480, \sum_{i=1}^{10} x_i^2 = 32500, \sum_{i=1}^{10} y_i^2 = 377.$$

Επίσης, έχουμε:

$$\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = \frac{530}{10} = 53$$

και

$$\bar{y} = \frac{1}{10}\sum_{i=1}^{10} y_i = \frac{55}{10} = 5{,}5.$$

Συνεπώς, ο συντελεστής γραμμικής συσχέτισης είναι:

$$r = \frac{10\sum_{i=1}^{10} x_i y_i - (\sum_{i=1}^{10} x_i)(\sum_{i=1}^{10} y_i)}{\sqrt{10\sum_{i=1}^{10} x_i^2 - (\sum_{i=1}^{n} x_i)^2}\sqrt{10\sum_{i=1}^{10} y_i^2 - \left(\sum_{i=1}^{10} y_i\right)^2}}$$

$$= \frac{10 \cdot 3480 - 530 \cdot 55}{\sqrt{10 \cdot 32500 - 530^2}\sqrt{10 \cdot 377 - 55^2}}$$

$$= 0{,}986.$$

Επειδή η τιμή του συντελεστή συσχέτισης είναι πολύ κοντά στη μονάδα, υπάρχει ισχυρή θετική συσχέτιση μεταξύ των μεταβλητών X και Y.

E2. Η ευθεία γραμμικής παλινδρόμησης είναι:

$$\hat{y} = \hat{a} + \hat{b}x,$$

όπου

$$\hat{b} = \frac{10\sum_{i=1}^{10} x_i y_i - (\sum_{i=1}^{10} x_i)(\sum_{i=1}^{10} y_i)}{10\sum_{i=1}^{10} x_i^2 - \left(\sum_{i=1}^{10} x_i\right)^2}$$

$$= \frac{10 \cdot 3480 - 530 \cdot 55}{10 \cdot 32500 - 530^2}$$

$$= 0{,}128$$

και

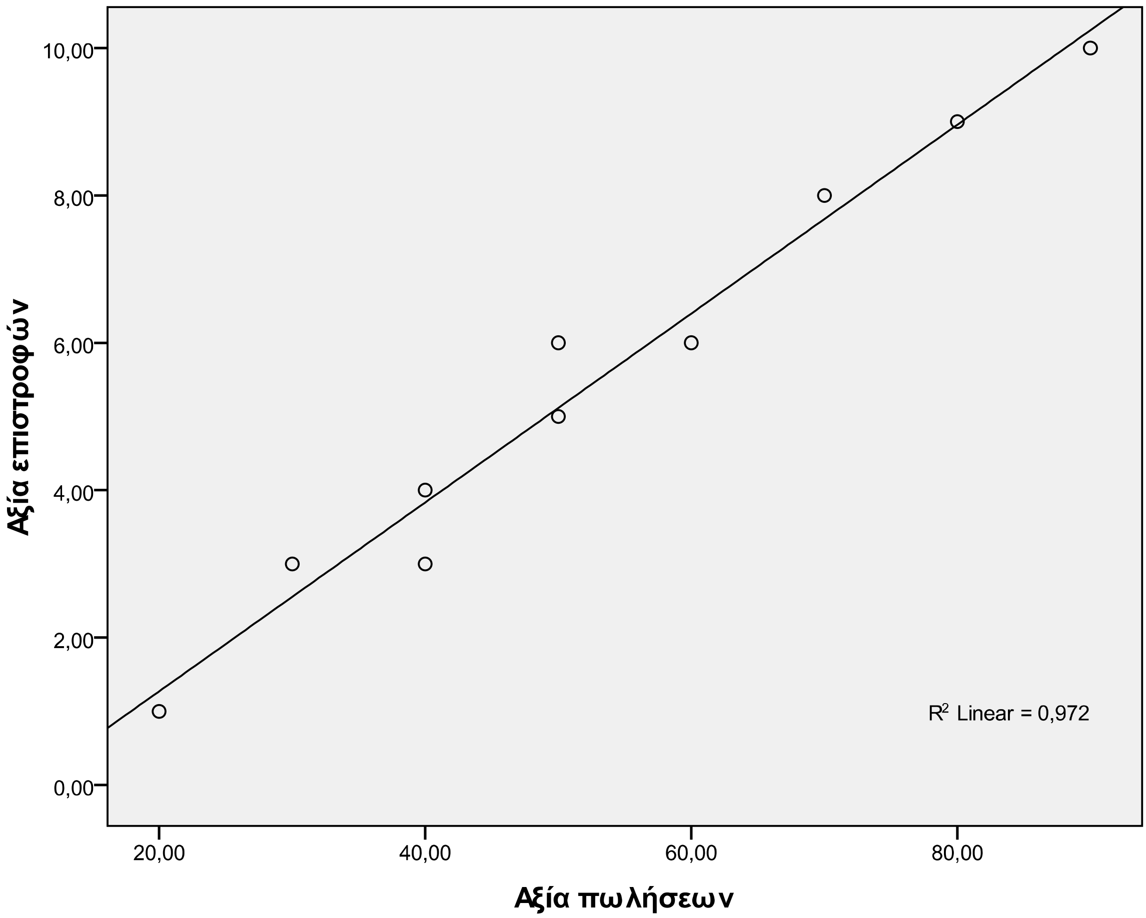
$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 5{,}5 - 0{,}128 \cdot 53 = -1{,}29.$$

Επομένως, η ζητούμενη εξίσωση παλινδρόμησης είναι:

$$\hat{y} = -1{,}29 + 0{,}128x.$$

Η ευθεία παλινδρόμησης και το διάγραμμα διασποράς δίνονται παρακάτω:

Διάγραμμα 7



Ε3. Εάν οι πωλήσεις αυξάνονταν κατά μια μονάδα, τότε:

$$\hat{y} = -1{,}29 + 0{,}128(x + 1) = -1{,}29 + 0{,}128x + 0{,}128 = \hat{y} + 0{,}128.$$

Επομένως, εάν οι πωλήσεις αυξάνονταν κατά μια μονάδα δηλαδή κατά 1000000 ευρώ, τότε το ύψος των επιστρεφόμενων εμπορευμάτων θα αυξανόταν κατά 128000 ευρώ, όσο ακριβώς και η τιμή του συντελεστή $\hat{b}$.

Ε4. Εάν οι ετήσιες πωλήσεις μιας επιχείρησης είχαν ύψος 35 εκατομμύρια ευρώ, τότε η αξία των εμπορευμάτων που επιστράφηκαν θα ήταν:

$$\hat{y} = -1{,}29 + 0{,}128 \cdot 35 = 3{,}19 \text{ ή } 3190000 \text{ ευρώ}.$$

Εφαρμογή 8

Ο παρακάτω πίνακας δίνει τα ύψη των ετησίων διαφημιστικών εξόδων σε χιλιάδες ευρώ που έγιναν από τις 30 πρώτες σε πωλήσεις επιχειρήσεις, ενός εμπορικού κλάδου.

Διαφημιστικά έξοδα	**Αριθμός επιχειρήσεων**
[0, 20)	2
[20, 40)	3
[40, 60)	8
[60, 80)	10
[80, 100)	4
[100, 120)	2
[120, 140)	1
Σύνολο	**30**

Ε1. Ζητείται να υπολογισθούν και να ερμηνευθούν:

Ε1.1. Τα τεταρτημόρια.

Ε1.2. Το ενδοτεταρτημοριακό εύρος.

Ε1.3. Η μέση τιμή.

Ε1.4. Η διακύμανση.

Ε1.5. Η τυπική απόκλιση.

Ε2. Να κατασκευαστεί το ιστόγραμμα συχνοτήτων και να εξαχθούν συμπεράσματα για τη μορφή της κατανομής τους.

Λύση

Ε1.1. Για να υπολογίσουμε τα τεταρτημόρια πρέπει πρώτα να βρούμε τις αθροιστικές συχνότητες.

Ο πίνακας αθροιστικών συχνοτήτων δίνεται παρακάτω:

Πίνακας 10

Κλάσεις	**Συχνότητα** ν_i	**Αθροιστική Συχνότητα** Φ_i
[0, 20)	2	2
[20, 40)	3	5
[40, 60)	8	13
[60, 80)	10	23
[80, 100)	4	27
[100, 120)	2	29
[120, 140)	1	30
Σύνολο	**30**	

Υπολογισμός πρώτου τεταρτημορίου.

Ο αριθμός

$$\frac{N}{4} = \frac{30}{4} = 7{,}5$$

βρίσκεται μεταξύ των αθροιστικών συχνοτήτων $\Phi_2 = 5$ και $\Phi_3 = 13$.

Έχουμε:

$$\alpha_2 = 40, f_3 = 8, \delta = 20.$$

Επομένως:

$$Q_1 = \alpha_2 + \frac{\delta}{f_3}\left(\frac{N}{4} - \Phi_2\right) = 40 + \frac{20}{8}\left(\frac{30}{4} - 5\right) = 46{,}25.$$

Υπολογισμός δεύτερου τεταρτημορίου.

Ο αριθμός

$$\frac{N}{2} = \frac{30}{2} = 15$$

βρίσκεται μεταξύ των αθροιστικών συχνοτήτων $\Phi_3 = 13$ και $\Phi_4 = 23$.

Έχουμε:

$$\alpha_3 = 60, f_4 = 10, \delta = 20.$$

Επομένως:

$$M = Q_2 = \alpha_3 + \frac{\delta}{f_4}\left(\frac{N}{2} - \Phi_3\right) = 60 + \frac{20}{10}\left(\frac{30}{2} - 13\right) = 64.$$

Υπολογισμός τρίτου τεταρτημορίου.

Ο αριθμός

$$\frac{3N}{4} = \frac{90}{4} = 22.5$$

βρίσκεται μεταξύ των αθροιστικών συχνοτήτων $\Phi_3 = 13$ και $\Phi_4 = 23$.

Έχουμε:

$$\alpha_3 = 60, f_4 = 10, \delta = 20.$$

Επομένως:

$$Q_3 = \alpha_3 + \frac{\delta}{f_4}\left(\frac{3N}{4} - \Phi_3\right) = 60 + \frac{20}{10}\left(\frac{90}{4} - 13\right) = 79.$$

ΕΡΜΗΝΕΙΑ: Εάν ταξινομηθούν τα διαφημιστικά έξοδα των επιχειρήσεων κατά αύξουσα σειρά, τότε τα τεταρτημόρια χωρίζουν το πλήθος τους σε τέσσερις ίσες ομάδες. Για παράδειγμα, έχουμε ότι το πρώτο 25% των διαφημιστικών εξόδων του δείγματος έχει τιμή μέχρι και 46250 ευρώ ενώ το υπόλοιπο 75% του δείγματος έχει τιμή πάνω από 46250 ευρώ.

Ε1.2. Το ενδοτεταρτημοριακό έυρος είναι:

$$H = Q_3 - Q_1 = 79 - 46{,}25 = 32{,}75.$$

ΕΡΜΗΝΕΙΑ: Το ενδοτεταρτημοριακό εύρος μας δίνει το εύρος του 50% των «κεντρικών» διαφημιστικών εξόδων του δείγματος, απαλλαγμένο από το 25% των μικρότερων και το 25% των μεγαλύτερων πωλήσεων.

Ε1.3. Για να υπολογίσουμε τη μέση τιμή κατασκευάζουμε τον παρακάτω πίνακα:

Πίνακας 11

Κλάσεις	x_i	ν_i	$\nu_i x_i$
[0, 20)	10	2	20
[20, 40)	30	3	90
[40, 60)	50	8	400
[60, 80)	70	10	700
[80, 100)	90	4	360
[100, 120)	110	2	220
[120, 140)	130	1	130
Σύνολο		**30**	**1920**

Η μέση τιμή είναι:

$$\bar{x} = \frac{1}{30} \cdot \sum_{i=1}^{7} \nu_i x_i = \frac{1920}{30} = 64.$$

ΕΡΜΗΝΕΙΑ: Οι 30 επιχειρήσεις στη χρονιά που εξετάσαμε δαπάνησαν κατά μέσο όρο 64000 ευρώ για διαφημιστικά έξοδα.

Ε1.4. Για να υπολογίσουμε τη διακύμανση κατασκευάζουμε τον παρακάτω πίνακα:

Πίνακας 12

Κλάσεις	$\boldsymbol{x_i}$	$\boldsymbol{\nu_i}$	$\boldsymbol{x_i^2}$	$\boldsymbol{\nu_i x_i^2}$
[0, 20)	10	2	100	200
[20, 40)	30	3	900	2700
[40, 60)	50	8	2500	20000
[60, 80)	70	10	4900	49000
[80, 100)	90	4	8100	32400
[100, 120)	110	2	12100	24200
[120, 140)	130	1	16900	16900
Σύνολο		**30**		**145400**

Η διακύμανση είναι:

$$s^2 = \frac{1}{30-1}\left[\sum_{i=1}^{7} \nu_i x_i^2 - 30\bar{x}^2\right] = \frac{1}{29}(145400 - 30 \cdot 64^2) = 776{,}55.$$

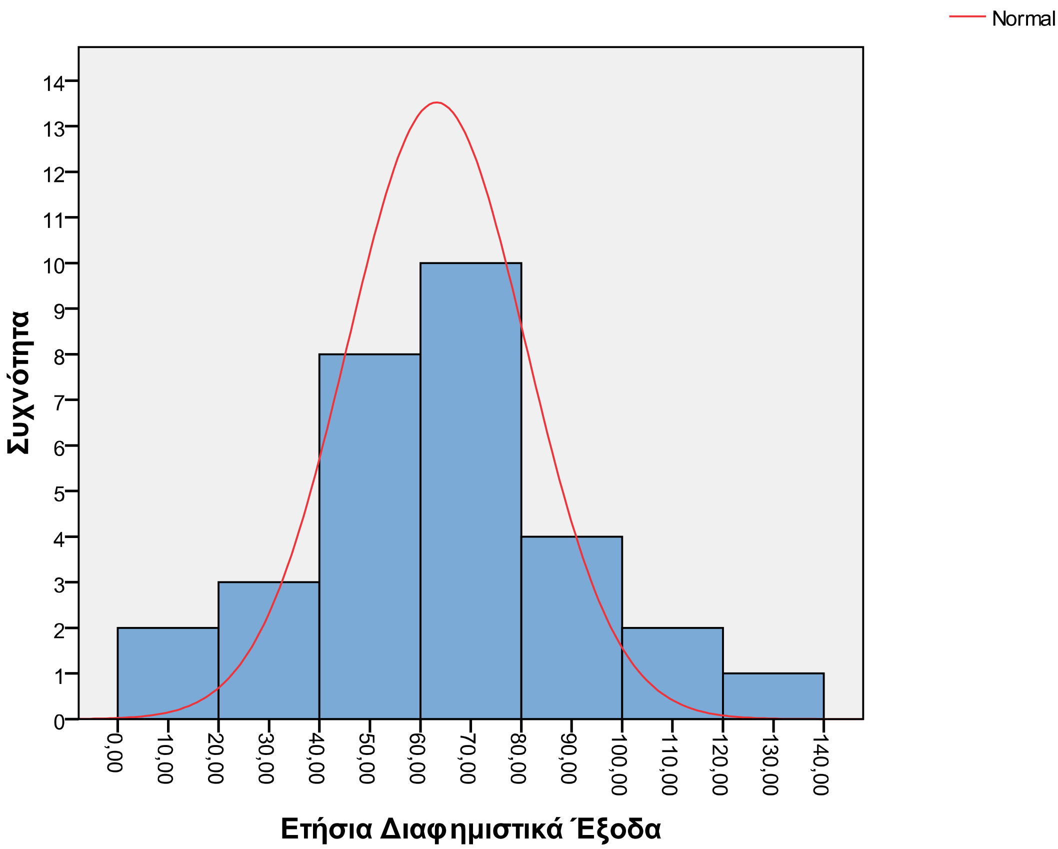
Ε1.5. Η τυπική απόκλιση είναι:

$$s = \sqrt{s^2} = \sqrt{776{,}55} \approx 27{,}87.$$

ΕΡΜΗΝΕΙΑ: Όπως γνωρίζουμε, όσο μικρότερες είναι οι τιμές της s^2 και της s τόσο πιο συγκεντρωμένες γύρω από τη μέση τιμή βρίσκονται οι παρατηρήσεις μας. Επίσης η τετραγωνική ρίζα του μέσου αριθμητικού των τετραγώνων των αποκλίσεων των δεδομένων μας από τη μέση ετήσια δαπάνη διαφήμισης (64000 ευρώ) ισούται με 27870 ευρώ περίπου.

E2. Το ιστόγραμμα συχνοτήτων δίνεται παρακάτω:

Διάγραμμα 8



Από το ιστόγραμμα βλέπουμε ότι η κατανομή των ετησίων διαφημιστικών εξόδων φαίνεται να είναι περίπου κανονική (είναι σχεδόν συμμετρική και δεν έχει μακριές ουρές).

ΣΥΜΠΕΡΑΣΜΑΤΑ

Η στατιστική είναι ένα ισχυρό εργαλείο στην υπηρεσία οποιασδήποτε επιστήμης και παρέχει πολλές δυνατότητες, όσον αφορά στον προσδιορισμό της μεταβλητότητας, στην αντιμετώπιση, στην πρόβλεψη, στον σχεδιασμό και λήψη αποφάσεων ενώ ταυτόχρονα μας εξασφαλίζει κέρδος, χρόνου και χρήματος. Αυτό είναι σημαντικό σε μια επιχείρηση.

Καταλήγουμε στο συμπέρασμα πως η εφαρμογή των στατιστικών μεθόδων στις επιχειρήσεις έδωσε λύσεις σε προβλήματα που αφορούσαν τον επιχειρηματικό τομέα, με συνέπεια την ορθή ανάπτυξη βιομηχανικών και εμπορικών επιχειρήσεων. Η εφαρμογή της στατιστικής όμως δεν είναι απαραίτητη σε μια μικρή επιχείρηση ενώ για μια μεγάλη επιχείρηση η στατιστική είναι χρήσιμη ως πληροφοριακό εργαλείο που καθιστά πιο εύκολη τη λήψη αποφάσεων σε όλους τους τομείς της επιχείρησης (π.χ επιχειρηματικές αποφάσεις, άσκηση κοινωνικής, οικονομικής και τιμολογιακής πολιτικής).

Σε μια μεγάλη επιχείρηση ο στατιστικός συγκεντρώνει τα στατιστικά στοιχεία και έπειτα απ' τον απαραίτητο υπολογισμό των στατιστικών δεικτών τα παρουσιάζει σε πίνακες ή διαγράμματα με σκοπό την επεξεργασία τους και την χρησιμοποίησή τους στην λήψη ορθών αποφάσεων. Κατά συνέπεια, η εφαρμογή της στατιστικης οδηγεί στην βελτίωση της επιχείρησης σε οποιοδήποτε τομέα του οποίου έγινε συγκέντρωση των στατιστικών στοιχείων.

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] Αδαμάπουλος Λ., Δαμιανού Χ. και Σβέρκος Α., *Μαθηματικά και στοιχεία στατιστικής*, Οργανισμός εκδόσεως διδακτικών βιβλίων, 2011.

[2] Αποστολόπουλος Θ., *Περιγραφική στατιστική επιχειρήσεων*, Σύγχρονη Εκδοτική, 2003.

[3] Γναρδέλης Χ., *Εφαρμοσμένη στατιστική*, Εκδόσεις Παπαζήση, 2003.

[4] Καραγεώργου Δ., Κόκλα Α., Παπακωσταντίνου Ε., *Στατιστική για τις επιχειρήσεις*, Οργανισμός έκδοσης διδακτικών βιβλίων, Αθήνα, 1999.

[5] Κιόχος Π., Κιόχος Α., *Στατιστική για τις επιχειρήσεις και την οικονομία*, εκδ. Ελένη Κιόχου, 2010.

[6] Λιώκη-Λειβαδά Η. και Ασημακόπουλος Δ.Ν., *Εισαγωγή στην εφαρμοσμένη στατιστική*, *Τεύχος 1 Μεθοδολογίες*, Συμμετρία, 2007.

[7] Λιώκη-Λειβαδά Η. και Ασημακόπουλος Δ.Ν., *Εισαγωγή στην εφαρμοσμένη στατιστική*, *Τεύχος 2 Ασκήσεις*, Συμμετρία, 2007.

[8] Μπάρλας Α.Χ., *Μαθηματικά Γ' Λυκείου - Γενικής Παιδείας*, Ελληνοεκδοτική, 2009.