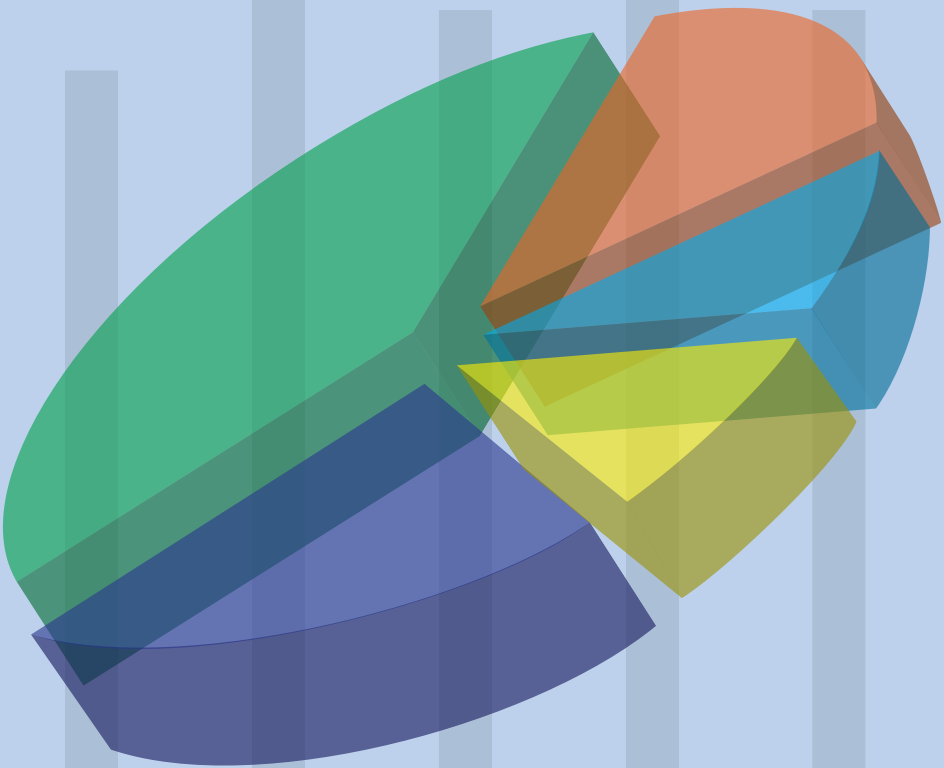
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΙΝΣΤΙΤΟΥΤΟ ΕΚΠΑΙΔΕΥΤΙΚΗΣ ΠΟΛΙΤΙΚΗΣ

ΣΤΑΤΙΣΤΙΚΗ ΕΠΙΧΕΙΡΗΣΕΩΝ



Β΄ ΕΠΑ.Λ.

ΤΟΜΕΑΣ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΙΝΣΤΙΤΟΥΤΟ ΤΕΧΝΟΛΟΓΙΑΣ ΥΠΟΛΟΓΙΣΤΩΝ ΚΑΙ ΕΚΔΟΣΕΩΝ «ΔΙΟΦΑΝΤΟΣ»

ΣΤΑΤΙΣΤΙΚΗ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΣΤΟΙΧΕΙΑ ΑΡΧΙΚΗΣ ΕΚΔΟΣΗΣ

Συγγραφείς:

- Καραγεώργος Δημήτριος, *Λέκτορας Πανεπιστημίου Αθηνών, τ. Σύμβουλος Π.Ι.*
- Κόκλα Άννα – Μαρία, *Καθηγήτρια Β/θμιας Εκπαίδευσης.*
- Παπακωνσταντίνου Ευάγγελος, *Σχολικός Σύμβουλος Β/θμιας Εκπαίδευσης.*

Εποπτεία στα πλαίσια του Π.Ι.:

Ηλιάδης Νικόλαος, *Σύμβουλος Π.Ι.*

Κριτές:

Γιαλαμάς Βασίλειος, *Επίκουρος καθηγητής Πανεπιστημίου Αθηνών.*

Δαλιεράκη Ελισάβετ, *Καθηγήτρια Β/θμιας Εκπαίδευσης.*

Κανελλόπουλος Σωτήριος, *Καθηγήτης Β/θμιας Εκπαίδευσης.*

Γλωσσική Επιμέλεια:

Βαρδαλάχου Αικατερίνη, *Φιλόλογος.*

ΣΤΟΙΧΕΙΑ ΕΠΑΝΕΚΔΟΣΗΣ

Η επανέκδοση του παρόντος βιβλίου πραγματοποιήθηκε από το Ινστιτούτο Τεχνολογίας Υπολογιστών & Εκδόσεων «Διόφαντος» μέσω ψηφιακής μακέτας.

ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΙΝΣΤΙΤΟΥΤΟ ΕΚΠΑΙΔΕΥΤΙΚΗΣ ΠΟΛΙΤΙΚΗΣ

Καραγεώργος Δημήτριος **Κόκλα Άννα – Μαρία**

Παπακωνσταντίνου Ευάγγελος

Η συγγραφή και η επιστημονική επιμέλεια του βιβλίου πραγματοποιήθηκε
υπό την αιγίδα του Παιδαγωγικού Ινστιτούτου

ΣΤΑΤΙΣΤΙΚΗ ΕΠΙΧΕΙΡΗΣΕΩΝ

Β΄ ΕΠΑ.Λ.

ΤΟΜΕΑΣ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΙΝΣΤΙΤΟΥΤΟ ΤΕΧΝΟΛΟΓΙΑΣ ΥΠΟΛΟΓΙΣΤΩΝ ΚΑΙ ΕΚΔΟΣΕΩΝ «ΔΙΟΦΑΝΤΟΣ»

ΠΡΟΛΟΓΟΣ

Το βιβλίο αυτό καλύπτει την ύλη του μαθήματος της «Στατιστικής επιχειρήσεων» και απευθύνεται στους μαθητές της Α΄ τάξης του Α΄ κύκλου του τομέα οικονομίας και Διοίκησης των Τεχνολογικών Επαγγελματικών Εκπαιδευτηρίων (Τ.Ε.Ε.). Γράφτηκε στα πλαίσια της νέας μεταρρύθμισης (Ν.2525/97), που καθιέρωσε τα Τ.Ε.Ε, με σκοπό τον εκσυγχρονισμό των προγραμμάτων σπουδών και των αντίστοιχων διδακτικών εγχειριδίων της Τεχνικο-επαγγελματικής Εκπαίδευσης.

Για τη συγγραφή του βιβλίου οι συγγραφείς έλαβαν υπόψη την ελληνική και ξένη εμπειρία για παρόμοια εγχειρίδια και τις γενικότερες επιστημονικές απόψεις για τη διδασκαλία της Στατιστικής στη Γενική και Επαγγελματική Εκπαίδευση. Η ανάπτυξη της ύλης έγινε με τέτοιο τρόπο, ώστε: α) Να ανταποκρίνεται στις νοητικές δυνατότητες, στις ανάγκες και στα ενδιαφέροντα των μαθητών, για τους οποίους προορίζεται, β) να βοηθήσει στην ανάπτυξη της κριτικής ικανότητας και δημιουργικότητας των μαθητών και γ) να δείξει τη χρησιμότητα της Στατιστικής σε διάφορες άλλες επιστήμες, καθώς και στην αντιμετώπιση πραγματικών προβλημάτων.

Το περιεχόμενο του βιβλίου χωρίζεται σε επτά κεφάλαια. Στο πρώτο κεφάλαιο αναπτύσσονται με συντομία ο σκοπός και η μεθοδολογία που χρησιμοποιεί η στατιστική και δίνονται οι ορισμοί των βασικών όρων που αναφέρονται στη συνέχεια.

Στο δεύτερο κεφάλαιο παρουσιάζεται η συμβολή της στατιστικής στη σύγχρονη επιχείρηση, ώστε όλοι οι μαθητές να κατανοήσουν τη χρησιμότητα του γνωστικού αντικειμένου που διδάσκονται.

Στο τρίτο κεφάλαιο αναπτύσσονται οι μέθοδοι συλλογής, ταξινόμησης και παρουσίασης στοιχείων (στατιστικών δεδομένων), καθώς και η χρησιμοποίησή τους στις επιχειρήσεις.

Στο τέταρτο κεφάλαιο παρουσιάζονται τα μέτρα θέσης μιας κατανομής, ο τρόπος υπολογισμού της και η σπουδαιότητά της στους διάφορους τομείς της καθημερινής ζωής και κυρίως στις επιχειρήσεις.

Στο πέμπτο κεφάλαιο παρουσιάζονται τα κυριότερα μέτρα διασποράς μίας κατανομής, ο τρόπος υπολογισμού τους και η χρησιμότητά τους στην ερμηνεία της συμπεριφοράς του φαινομένου που μελετάμε.

Στο έκτο κεφάλαιο αναπτύσσεται η μεθοδολογία για τον προσδιορισμό του βαθμού αλληλεξάρτησης δύο μεταβλητών και δίνεται ο τρόπος υπολογισμού του συντελεστή γραμμικής συσχέτισης που μετράει το βαθμό συσχέτισης των μεταβλητών αυτών.

Στο έβδομο κεφάλαιο αναπτύσσονται οι χρονολογικές σειρές, που έχουν μεγάλη χρησιμότητα στους διάφορους τομείς της οικονομίας.

Σε όλο το βιβλίο χρησιμοποιούνται παραδείγματα και εφαρμογές από την καθημερινή ζωή. Οι ασκήσεις και τα προβλήματα είναι κατάλληλα επιλεγμένα να προκαλούν το ενδιαφέρον των μαθητών και να αντιμετωπίζονται με ευκολία.

Ευελπιστούμε ότι η γνώση που θα αποκτήσουν οι μαθητές και από το βιβλίο αυτό θα τους κάνει ικανότερους να αντιμετωπίσουν με επιτυχία τις δυσκολίες της ζωής.

Οι συγγραφείς, ως μαχόμενοι εκπαιδευτικοί, γνωρίζουν καλά ότι οι στόχοι του βιβλίου αυτού δεν θα ευοδωθούν ποτέ, αν εκείνοι που θα το διδάξουν δεν το μελετήσουν με προσοχή και δεν κοπιάσουν στην τάξη. Προσπαθήσαμε το βιβλίο αυτό να είναι διδακτικά όσο γίνεται καλύτερο και πλήρες από άποψη υλικού, ώστε οι διδάσκοντες να βοηθηθούν στο δύσκολο έργο τους.

Οι συγγραφείς αισθάνονται την ανάγκη να ευχαριστήσουν:

α) Το σύμβουλο του Παιδαγωγικού Ινστιτούτου (Π.Ι.) Νίκο Ηλιάδη, ο οποίος εισηγήθηκε την ανάθεση της συγγραφής του βιβλίου στο τμήμα της Τ.Ε.Ε και τα μέλη του τμήματος που αποδέχτηκαν την εισήγηση.

β) Τη συνάδελφο Δέσποινα Μοσχολέα για τη διοικητική διευκόλυνση στα πλαίσια του Π.Ι.

γ) Τους κριτές: Βασίλη Γιαλαμά, Επίκουρο Καθηγητή Πανεπιστημίου Αθηνών, την Ελισάβετ Δαλιεράκη, Καθηγήτρια Β/θμιας Εκπαίδευσης, τον Σωτήριο Κανελλόπουλο Καθηγητή Β/θμιας Εκπαίδευσης και

δ) Τη φιλόλογο Βαρδαλάχου Αικατερίνη για τη γλωσσική επιμέλεια.

Τέλος, το βιβλίο αυτό, σαν ανθρώπινο δημιούργημα, ασφαλώς δεν είναι τέλειο. Ο μόνος τρόπος για να γίνει καλύτερο είναι ο νηφάλιος και ελεύθερος διάλογος, που έχει καθιερώσει η επιστημονική παράδοση. Οποιοσδήποτε μαθητής ή καθηγητής ή οποιοσδήποτε ενδιαφέρεται για την Παιδεία στον τόπο μας και θέλει να κάνει οποιαδήποτε σχόλια, παρατηρήσεις ή κρίσεις για το βιβλίο αυτό, είναι ευπρόσδεκτος να τις αποτείνει στο Παιδαγωγικό Ινστιτούτο, Μεσογείων 396, 15310 Αγία Παρασκευή (υπόψη Τμήματος Τ.Ε.Ε.), τηλ. 6014207 – 6016364, FAX: 6016388, 6003230 από όπου και θα ενημερώνονται οι συγγραφείς.

Ιούλιος 1999 Οι συγγραφείς

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΙΣΑΓΩΓΗ

1.1 Γενικά
1.2 Αντικείμενο και σκοπός της Στατιστικής
1.3 Η Στατιστική των επιχειρήσεων
1.4 Πληθυσμός – Μεταβλητές
1.5 Περιγραφική Στατιστική – Έννοια και περιεχόμενο
1.6 Επαγωγική Στατιστική – Έννοια και περιεχόμενο

❖ Ο **William Petty** ονομάζει τη στατιστική «*πολιτική αριθμητική*» (1683)

❖ «*Η στατιστική είναι η επιστήμη που μελετά τα γεγονότα κατά μάζας, δηλαδή κατά πολυαρίθμους ομοειδείς μονάδας*», W. Lexis (1837-1914), Γερμανός στατιστικός.

❖ Ο **Sir R. Fischer** είναι η σημαντικότερη επιστημονική παρουσία του εικοστού αιώνα στο χώρο της στατιστικής και θεωρείται από τους περισσότερους ως ο πατέρας της σύγχρονης στατιστικής επιστήμης.

ΚΕΦΑΛΑΙΟ 1ο
ΕΙΣΑΓΩΓΗ

1.1 Γενικά

Η λέξη στατιστική, κατά μία άποψη, προέρχεται από τη λατινική λέξη «status», που σημαίνει κράτος-κατάσταση και δήλωνε, από τότε που πρωτοχρησιμοποιήθηκε, συλλογή στοιχείων για τις κρατικές ανάγκες. Οι ασχολούμενοι με τη συλλογή και ανάλυση αυτών των στοιχείων ονομάζονταν «κρατικοί» ή «στατιστικοί». Άλλη μία εκδοχή της προέλευσης της λέξης στατιστική, σύμφωνα με τις απόψεις που διατυπώνονται στα περισσότερα ετυμολογικά λεξικά, είναι ότι αυτή έλκει την καταγωγή της από το ελληνικό ρήμα στατίζω, που σημαίνει τοποθετώ, διαπιστώνω, προσδιορίζω [στατιστικός –ή –ό (γαλλ. statistique< (στατίζω)].

Κάνοντας μία σύντομη **ιστορική αναδρομή** αξίζει να αναφέρουμε ότι οι πρώτες προσπάθειες συλλογής στατιστικών στοιχείων ανάγονται στους αρχαιότατους χρόνους. Από την ιστορία πληροφορούμαστε ότι σε απογραφές πληθυσμού, γης κτλ. προέβαιναν και λαοί της αρχαιότητας (Έλληνες, Κινέζοι, Αιγύπτιοι κ.λ.π).

Προχωρώντας στους αιώνες, διαπιστώνουμε ότι το πρώτο βιβλίο στατιστικού περιεχομένου γράφτηκε το 1583 από τον *Fr. Sansovino.* Αργότερα, στο τέλος του Μεσαίωνα, στην Ευρώπη, διάφορες κυβερνήσεις άρχισαν να συγκεντρώνουν συστηματικά στατιστικά στοιχεία γεννήσεων, θανάτων, γάμων κτλ. Η Στατιστική ως αυτοτελής επιστήμη εφαρμόζεται από τον 17ο αιώνα. Τότε, άρχισε να διαμορφώνεται ένας νέος κλάδος, που προήλθε από τη μελέτη των τυχερών παιχνιδιών, η θεωρία των πιθανοτήτων, η οποία προάγεται και συμπληρώνεται κυρίως από τους *Bernoulli, Gauss, Laplace*. Η στατιστική έχει πλέον εισαχθεί (18ος αιώνας) στις ακαδημαϊκές σπουδές και η συστηματική οργάνωση των κρατικών στατιστικών υπηρεσιών αλλά και επιστημονικών εταιρειών είναι πλέον γεγονός (19ος αιώνας).

Αξίζει να αναφερθεί ότι ο πρώτος Κυβερνήτης της Ελλάδας, Ι. Καποδίστριας, είχε ενδιαφερθεί σοβαρότατα για τη δημιουργία στατιστικής υπηρεσίας στην Ελλάδα, διαβλέποντας τον κρίσιμο ρόλο που θα έπαιζε στη δημιουργία του νέου κράτους.

Έτσι αρκετά νωρίς για την ελληνική και διεθνή πραγματικότητα, το 1833, ιδρύθηκε η Υπηρεσία Γενικής Στατιστικής του Κράτους, η οποία τέθηκε στη δικαιοδοσία του Υπουργείου των Εσωτερικών. Με το Κανονιστικό Διάταγμα (Κ.Δ.) της 29/4 του 1834, πάλι στο Υπουργείο Εσωτερικών, ιδρύθηκε το «Γραφείο Δημοσίας Οικονομίας», το οποίο συμπεριέλαβε και την Υπηρεσία Γενικής Στατιστικής του Κράτους.

Περί τα τέλη του 19ου αιώνα η Στατιστική έχει το κατάλληλο επιστημονικό υπόβαθρο για την ανάπτυξη στατιστικών μεθόδων.

Θα πρέπει εδώ να σημειώσουμε τη συμβολή των Άγγλων *Karl Pearson* (1857-1936) και *R.A. Fisher* (1890-1962) στη θεμελίωση της σύγχρονης Στατιστικής, όπως επίσης και του Βέλγου στατιστικολόγου *A. Quetelet* ο οποίος ήταν και ο κύριος διοργανωτής του πρώτου διεθνούς συνεδρίου Στατιστικής το 1853.

Στις αρχές του 20ου αιώνα, χρησιμοποιήθηκε η στατιστική μέθοδος ελέγχου και ποιότητας των βιομηχανικών προϊόντων και θεωρείται ως η πρώτη σε ευρεία εφαρμογή χρήση στατιστικών μεθόδων στην παραγωγή.

1.2 Αντικείμενο και σκοπός της Στατιστικής

Σήμερα, ακούμε από τα Μέσα Μαζικής Ενημέρωσης ή διαβάζουμε στα διάφορα έντυπα για «στατιστικές». Αυτές είναι πληροφορίες που αναφέρονται σε συγκεκριμένο αντικείμενο έρευνας ή γεγονός και δίνονται με μορφή πινάκων ή διαγραμμάτων. Ανάλογα με το αντικείμενο έρευνας ή το γεγονός, η Στατιστική παίρνει και ιδιαίτερη ονομασία. Για παράδειγμα, μιλάμε για «Στατιστική Επιχειρήσεων», για «Γεωργική Στατιστική», για «Στατιστική Τουρισμού» κτλ., όταν τα αριθμητικά δεδομένα αναφέρονται αντίστοιχα στις επιχειρήσεις, στη γεωργία ή στον τουρισμό κ.ο.κ.

Εκτός από τη σημασία αυτή, η «Στατιστική» ως κλάδος των εφαρμοσμένων επιστημών όχι μόνο συγκεντρώνει και παρουσιάζει πληροφορίες, αλλά ταυτόχρονα μελετά και αναλύει τις παρατηρήσεις ή μετρήσεις που αναφέρονται σε ένα συγκεκριμένο πεδίο έρευνας. Έτσι η Στατιστική περιλαμβάνει τόσο τις μεθόδους συλλογής και επεξεργασίας στοιχείων όσο και τις μεθόδους ανάλυσης και μελέτης τους. Από τη μελέτη αυτή προκύπτουν οι σχέσεις που υπάρχουν στα διάφορα φαινόμενα και διατυπώνονται συμπεράσματα που είναι χρήσιμα για τη λήψη ορθών αποφάσεων.

Μπορούμε λοιπόν να πούμε ότι:

Στατιστική είναι η εφαρμοσμένη επιστήμη που ασχολείται με τις επιστημονικές μεθόδους σχεδιασμού μιας μελέτης, συλλογής, επεξεργασίας, παρουσίασης και ανάλυσης στοιχείων με σκοπό την εξαγωγή συμπερασμάτων για τη λήψη ορθών αποφάσεων.

Σύμφωνα με τον ορισμό αυτό της Στατιστικής μπορούμε να απαριθμήσουμε τα βασικά στάδια που ακολουθούμε για τη μελέτη ενός φαινομένου.

Αυτά είναι:
- **Ο σχεδιασμός της μελέτης του φαινομένου.**
- **Η συγκέντρωση των απαραίτητων πληροφοριών (στατιστικών στοιχείων)**
- **Η επεξεργασία και παρουσίαση των στοιχείων αυτών.**
- **Η ανάλυση των στοιχείων με επιστημονικές μεθόδους.**
- **Η εξαγωγή χρήσιμων συμπερασμάτων.**

Η Στατιστική ως επιστήμη έχει τους δικούς της συμβολισμούς, ορολογία, θεωρήματα και τεχνικές. Χρησιμοποιείται σε όλες σχεδόν τις άλλες επιστήμες και στους περισσότερους τομείς της ανθρώπινης δραστηριότητας (διοίκηση, δημογραφία, βιομηχανία, εμπόριο, πολιτική, ιατρική κτλ.).

Η μεθοδολογία της Στατιστικής και η χρησιμοποίησή της έχουν παρεξηγηθεί από πολλούς ανθρώπους. Έχουμε ακούσει ότι ο στατιστικός είναι ο άνθρωπος που φτιάχνει ακριβή διαγράμματα από παράλογες υποθέσεις ή που στηρίζει επιστημονικά τα ψέματά του κτλ.

Η στατιστική ως επιστήμη δεν ευθύνεται βέβαια για όλα αυτά. Μπορούμε όμως να αναζητήσουμε την ευθύνη για τη δημιουργία τέτοιων εντυπώσεων και παρεξηγήσεων είτε σε εκείνους που χρησιμοποιούν τη Στατιστική λανθασμένα χωρίς να εξακριβώνουν τη γνησιότητα των στοιχείων τους είτε σε εκείνους που τη χρησιμοποιούν σκόπιμα για κερδοσκοπικούς, δημαγωγικούς ή άλλους λόγους που έχουν σκοπό να παραπλανήσουν τους αποδέκτες των στατιστικών μελετών.

Η επίδραση της Στατιστικής στη ζωή μας είναι σήμερα πολύ μεγάλη. Εκτός από την εφαρμογή της στην απογραφή, στατιστικές μέθοδοι χρησιμοποιούνται για παράδειγμα στη μελέτη της γεννητικότητας, της θνησιμότητας, της μετανάστευσης κτλ. προκειμένου να ληφθούν ορθές αποφάσεις. Χρησιμοποιούνται επίσης στον τομέα της εκπαίδευσης για την εξαγωγή χρήσιμων συμπερασμάτων, όσον αφορά το εκπαιδευτικό σύστημα μιας χώρας, και στη μετεωρολογία, η οποία δεν μπορεί να κάνει καμία πρόβλεψη χωρίς τη στατιστική ανάλυση των διαφόρων στοιχείων που επηρεάζουν το φαινόμενο που μελετάει κάθε φορά.

Η μεθοδολογία της Στατιστικής χρησιμοποιείται επίσης σε γεωργικές μελέτες, στην ιατρική και φαρμακευτική έρευνα κτλ. Η Στατιστική και η θεωρία των πιθανοτήτων χρησιμοποιούνται για να ληφθούν σοβαρές αποφάσεις. Για παράδειγμα, με στατιστικές μεθόδους ελέγχεται η ποιότητα των προϊόντων που παράγονται από μια βιομηχανία και στη συνέχεια αποφασίζεται η διάθεσή τους στην αγορά.

Όλα τα αναπτυγμένα κράτη χρησιμοποιούν τη Στατιστική για τη μελέτη των οικονομικών δραστηριοτήτων (παραγωγή, εισαγωγές, εξαγωγές, τιμές αγαθών κτλ.) και των διοικητικών και κοινωνικών θεμάτων (διοίκηση, κοινωνικές ασφαλίσεις κτλ.).

Ως εδώ αναφέρθηκαν μερικοί μόνο τομείς στους οποίους προσφέρει τις υπηρεσίες της η Στατιστική, αλλά είναι σκόπιμο να γίνει ιδιαίτερη αναφορά στο ρόλο της Στατιστικής στο χώρο των επιχειρήσεων.

1.3 Η Στατιστική των Επιχειρήσεων

Η μέχρι τώρα διεθνής εμπειρία από τις εφαρμογές της Στατιστικής στον επιχειρηματικό τομέα έχει καταστήσει τις στατιστικές μεθόδους απαραίτητο εργαλείο στην άσκηση της σύγχρονης επιχειρηματικής δραστηριότητας.

Τα στατιστικά στοιχεία που αφορούν την περιουσιακή κατάσταση και εξέλιξη των επιχειρήσεων, όπως αυτά που αναφέρονται στα αποθέματα, τον κεφαλαιουχικό εξοπλισμό, τις αγορές και τις πωλήσεις, τα αποτελέσματα των διαφημιστικών δαπανών κτλ., μπορούν, αφού μελετηθούν κατάλληλα, να βοηθήσουν σημαντικά στη λήψη ορθών αποφάσεων.

Στις γνωστές έρευνες αγοράς, που διεξάγουν οι σύγχρονες επιχειρήσεις, ο ρόλος της Στατιστικής είναι καθοριστικός και τα συμπεράσματα είναι πολύτιμα για το μέλλον της επιχείρησης.

Με τη βοήθεια της Στατιστικής οι επιχειρήσεις μπορούν να κάνουν προβλέψεις για την εξέλιξη των πωλήσεών τους, να εξασφαλίζουν τον έλεγχο της ποιότητας των παραγόμενων προϊόντων, να ελέγχουν το κόστος παραγωγής κτλ.

Η ανάπτυξη ολόκληρων επιστημονικών κλάδων, όπως είναι η Οικονομετρία, η Επιχειρησιακή Έρευνα, η Διαφήμιση κ.ά., οφείλεται σε μεγάλο βαθμό στη Στατιστική, της οποίας η χρησιμότητα γίνεται κάθε μέρα και περισσότερο εμφανής.

Συνεπώς, η Στατιστική είναι απαραίτητη στη σύγχρονη Διοίκηση Επιχειρήσεων, γιατί η λήψη επιχειρηματικών αποφάσεων πρέπει να βασίζεται σε επιστημονικές μεθόδους και όχι μόνο στη διαίσθηση του επιχειρηματία ή στην τύχη.

1.4 Πληθυσμός – Μεταβλητές

Για την εξαγωγή χρήσιμων συμπερασμάτων είναι απαραίτητο σε κάθε στατιστική έρευνα να εξετάζονται τα στοιχεία ενός συνόλου ως προς ένα ή περισσότερα χαρακτηριστικά του. Η αρχή αυτή ακολουθείται όταν, για παράδειγμα, ενδιαφερόμαστε να βγάλουμε συμπεράσματα για:

- Τον αριθμό των επισκεπτών των μουσείων κατά φύλο και ηλικία.
- Τα ποσοστά της ανεργίας κατά ομάδα ηλικιών, φύλο και επίπεδο εκπαίδευσης.
- Την τουριστική κίνηση στα ξενοδοχειακά καταλύματα.
- Τις συνέπειες του καπνίσματος στην υγεία των καπνιστών κτλ.

Σε κάθε ένα από τα παραπάνω παραδείγματα έχουμε **ένα σύνολο και θέλουμε να εξετάσουμε τα στοιχεία του ως προς ένα ή περισσότερα χαρακτηριστικά τους. Ένα τέτοιο σύνολο ονομάζεται πληθυσμός (population).** Με την έννοια αυτή θα χρησιμοποιείται ο όρος πληθυσμός στα παρακάτω. Για την πληρότητα του βιβλίου αναφέρουμε ότι στη Στατιστική χρησιμοποιείται και ο όρος «στατιστικός πληθυσμός» με την ακόλουθη σημασία:

Ως στατιστικό πληθυσμό ορίζουμε τις μετρήσεις ή παρατηρήσεις που αναφέρονται σε κάποιο χαρακτηριστικό ή σε κάποια ιδιότητα των μονάδων του πληθυσμού.

Τα στοιχεία του πληθυσμού καθώς και του στατιστικού πληθυσμού τα λέμε **άτομα** ή **στοιχεία** ή **στατιστικές μονάδες του πληθυσμού.**

Για να πάρουμε μια απόφαση για έναν πληθυσμό (για παράδειγμα αν είναι αποδεκτό το ποσοστό των ελαττωματικών λαμπτήρων που φτιάχνει ένα εργοστάσιο) συνήθως εξετάζουμε ένα μόνο μέρος (δηλ. μερικούς μόνο λαμπτήρες) από το σύνολο αυτό. Επιλέγουμε δηλαδή μια μικρή ομάδα ή ένα υποσύνολο του πληθυσμού, το οποίο καλείται **δείγμα**, απ' όπου αντλούνται πληροφορίες για τα χαρακτηριστικά του συγκεκριμένου πληθυσμού. Δηλαδή:

Δείγμα είναι κάθε τμήμα του πληθυσμού.

Η επιλογή των στοιχείων αυτού του δείγματος πρέπει να γίνεται με «**τυχαία διαδικασία**». Αυτό σημαίνει ότι το δείγμα λαμβάνεται με κάποιο τρόπο που είναι έξω από τον έλεγχο εκείνου που κάνει το πείραμα ή τη μελέτη. Υπάρχουν πολλοί μαθηματικοί ορισμοί για το «τυχαίο δείγμα», αλλά εμείς, όπου αναφέρουμε τον όρο, **θα εννοούμε ότι κάθε στοιχείο του πληθυσμού έχει την ίδια ευκαιρία να επιλεγεί και η επιλογή του δεν επηρεάζει την επιλογή κανενός άλλου στοιχείου.** Βασισμένοι τώρα στο δείγμα αυτό, οδηγούμαστε σε ορισμένα συμπεράσματα για ολόκληρο τον πληθυσμό.

Ας πάρουμε ένα απλό παράδειγμα: Μια φαρμακευτική εταιρεία θέλοντας να εισάγει ένα νέο φάρμακο (Α) στην αγορά κάνει μια έρευνα για να διαπιστώσει, αν πραγματικά το φάρμακο αυτό είναι αποτελεσματικότερο από το φάρμακο Β, που ήδη κυκλοφορεί. Από τους ασθενείς που έχουν ανάγκη αυτά τα φάρμακα παίρνουμε τυχαία δύο ισοπληθείς ομάδες, για παράδειγμα, 30 ασθενείς σε κάθε ομάδα. Στη μία ομάδα δίνουμε το φάρμακο Α και στην άλλη το Β για 10 ημέρες. Την ημέρα που αρχίζουμε το πείραμα εξετάζουμε όλα τα άτομα και καταγράφουμε για το καθένα την κατάστασή του για τα ίδια στοιχεία (αρτηριακή πίεση, αριθμός αιμοσφαιρίων, ζάχαρο κτλ.). Την 11η ημέρα κάνουμε πάλι εξετάσεις και καταγράφουμε τα νέα αποτελέσματα. Αναλύουμε με στατιστικές μεθόδους τα αποτελέσματα για κάθε ομάδα, συγκρίνουμε τα αποτελέσματα μεταξύ των δύο ομάδων και τελικά βγάζουμε το συμπέρασμα ποιο από τα δύο φάρμακα είναι καλύτερο.

Τα άτομα ενός πληθυσμού εξετάζονται ως προς μια ή περισσότερες χαρακτηριστικές ιδιότητές τους. Στο παράδειγμα με τη φαρμακευτική εταιρεία οι ασθενείς αποτελούν τον πληθυσμό και ο καθένας είναι στοιχείο του πληθυσμού. Οι αριθμοί 14, 19, 18, 16,…, 17, που δείχνουν τις τιμές της αρτηριακής πίεσης των ατόμων που εξετάζονται, αποτελούν τις παρατηρήσεις μας (τα στατιστικά δεδομένα), ενώ η «αρτηριακή πίεση» είναι το χαρακτηριστικό του πληθυσμού ως προς το οποίο εξετάζουμε τα άτομα και ονομάζεται **μεταβλητή**. Οι αριθμοί ή άλλες συμβολικές εκφράσεις που μετρούν ή εκφράζουν τις διάφορες καταστάσεις μιας μεταβλητής ονομάζονται **τιμές της μεταβλητής.** Μια μεταβλητή συμβολίζεται με τα γράμματα Χ, Ψ, Ζ,…, ενώ οι τιμές της μεταβλητής με τα αντίστοιχα μικρά (για παράδειγμα, οι τιμές της μεταβλητής Χ συμβολίζονται με x_1, x_2, x_3,…)

Παραθέτουμε τον παρακάτω πίνακα παραδειγμάτων για καλύτερη κατανόηση των μεταβλητών και των τιμών τους:

Πίνακας 1.1

Μεταβλητές	**Τιμές ή παρατηρήσεις**
-Αριθμός υπαλλήλων ενός λογιστηρίου	1, 2, 3, 4, 5, 6, 7…
-Οικογενειακή κατάσταση των υπαλλήλων μιας επιχείρησης	Άγαμος/η, έγγαμος/η, χήρος/α, διαζευγμένος/η, σε διάσταση
-Φύλο ανθρώπων	Αρσενικό, θηλυκό
-Εβδομαδιαίες πωλήσεις μιας αντιπροσωπείας αυτοκινήτων	0, 1, 2, 3, 4, 5,…
-Ανάστημα μαθητών (σε cm) ενός σχολείου	158, 160, 161, 171, 180,…

Οι μεταβλητές διακρίνονται σε δύο κατηγορίες. Στις **ποσοτικές** και στις **ποιοτικές:**

1) Ποσοτικές μεταβλητές (quantitative variables) είναι εκείνες που επιδέχονται μέτρηση και οι τιμές τους είναι πραγματικοί αριθμοί. Τέτοιες μεταβλητές είναι το βάρος ενός μαθητή, η θερμοκρασία, η υγρασία, ο αριθμός των κοριτσιών μιας οικογένειας κτλ.

2) Ποιοτικές μεταβλητές (qualitative variables) είναι εκείνες που δεν επιδέχονται μέτρηση και οι τιμές τους δεν εκφράζονται με αριθμούς, αλλά με λέξεις. Για παράδειγμα, το είδος των βιβλίων (λογοτεχνικό, ιστορικό), το χρώμα των αυτοκινήτων, η ένδειξη της ρίψης ενός νομίσματος κτλ.

Οι ποσοτικές μεταβλητές, τώρα, διακρίνονται σε **συνεχείς** και **διακριτές (ασυνεχείς):**

α) Συνεχής μεταβλητή (continuous variable) είναι εκείνη που μπορεί να πάρει όλες τις τιμές ενός διαστήματος της ευθείας των πραγματικών αριθμών. Για παράδειγμα, το βάρος ενός μαθητή, το ανάστημα, η ηλικία κτλ. είναι συνεχείς μεταβλητές.

β) Διακριτή ή ασυνεχής μεταβλητή (discrete variable) είναι εκείνη που παίρνει μόνο μεμονωμένες τιμές (πεπερασμένο ή αριθμήσιμο πλήθος τιμών). Για παράδειγμα, η ένδειξη ενός ζαριού, ο αριθμός των παιδιών μιας οικογένειας είναι ασυνεχείς μεταβλητές.

1.5 Περιγραφική Στατιστική – Έννοια και περιεχόμενο

Ανάλογα με τον τρόπο που χρησιμοποιούμε τη Στατιστική διακρίνουμε σήμερα δύο μεγάλους τομείς της: την **Περιγραφική Στατιστική** και την **Επαγωγική Στατιστική ή Στατιστική Συμπερασματολογία,** καθώς και ένα σχετικά πρόσφατο τομέα της, που ονομάζεται **σχεδιασμός πειραμάτων.**

Ο πρώτος τομέας, η **Περιγραφική Στατιστική**, με τον οποίο θα ασχοληθούμε παρακάτω, περιλαμβάνει τις μεθόδους για τη συλλογή, ταξινόμηση, περιγραφή και παρουσίαση ενός συνόλου δεδομένων. Συνήθως τα δεδομένα που έχουμε στη διάθεσή μας σε μία στατιστική μελέτη είναι πολλά και σε ακατάστατη μορφή, με αποτέλεσμα να μην μπορούμε να διακρίνουμε την πληροφορία που περιλαμβάνουν. Χρησιμοποιώντας τις μεθόδους και τις τεχνικές της Περιγραφικής Στατιστικής παρουσιάζουμε τα δεδομένα αυτά σε εύχρηστο μέγεθος και εύκολα επεξεργάσιμη μορφή. **Τα στατιστικά αποτελέσματα (συμπεράσματα)** που προκύπτουν με βάση τις μεθόδους αυτές **έχουν καθαρά περιγραφικό χαρακτήρα και αναφέρονται αποκλειστικά στο σύνολο των δεδομένων που εξετάζουμε (δείγμα), ενώ δεν μπορούν να χρησιμεύσουν για να προχωρήσουμε σε γενικεύσεις που αφορούν ένα ευρύτερο σύνολο δεδομένων (πληθυσμός).**

1.6 Επαγωγική Στατιστική – Έννοια και περιεχόμενο

Ο δεύτερος τομέας, η **Επαγωγική Στατιστική ή Στατιστική Συμπερασματολογία**, περιλαμβάνει μεθόδους που μας βοηθούν να καταλήξουμε σε συμπεράσματα από το μέρος (δείγμα)-το οποίο μπορούμε να μελετήσουμε με τη βοήθεια των μεθόδων της Περιγραφικής Στατιστικής-στο σύνολο (πληθυσμός), και συνεπώς με αυτή την έννοια τα αποτελέσματα έχουν επαγωγικό χαρακτήρα. Η διαδικασία αυτή πραγματοποιείται με τη βοήθεια της θεωρίας των πιθανοτήτων. Τα συμπεράσματα αυτά οδηγούν στη λήψη ορισμένων αποφάσεων, όπως στο παράδειγμα της φαρμακευτικής εταιρείας που προαναφέρθηκε, δηλαδή από τη μελέτη της συμπεριφοράς ενός δείγματος (ομάδα Α και Β ασθενών) βγάζουμε το συμπέρασμα ποιο από τα δύο φάρμακα είναι το καλύτερο για να χρησιμοποιηθεί στο σύνολο των ασθενών. Γι' αυτό πολλές φορές λέμε ότι η **Στατιστική δημιουργεί τις αποφάσεις.**

Η διαδικασία αυτή μπορεί να αποδοθεί σχηματικά ως εξής:

Συλλογή πληροφοριών από ένα δείγμα → **Εξαγωγή συμπερασμάτων**

Ο σχεδιασμός πειραμάτων βασίζεται σε ένα συνδυασμό θεωρίας πιθανοτήτων και μιας στοιχειώδους αλλά ασυνήθιστης λογικής.

Στο βιβλίο αυτό θα μελετήσουμε τις τεχνικές και τις μεθόδους της Περιγραφικής Στατιστικής και θα χρησιμοποιήσουμε παραδείγματα από το χώρο των επιχειρήσεων, ώστε να γνωρίσει ο μαθητής πώς εφαρμόζονται αυτές οι τεχνικές στον επιχειρηματικό τομέα.

ΕΡΩΤΗΣΕΙΣ ΚΑΤΑΝΟΗΣΗΣ

1. Η Στατιστική είναι η επιστήμη που ασχολείται ....

   ....

   με σκοπό...

2. Σε μια στατιστική έρευνα μπορούμε να διακρίνουμε τρία στάδια:

   α)…………………………………………

   β)…………………………………………

   γ)…………………………………………

3. Πόσο απαραίτητη είναι η στατιστική στη σύγχρονη επιχειρηματική δράση; Απαντήστε συνοπτικά.

4. Στη Στατιστική, ως πληθυσμό εννοούμε ....

   ....

5. Υποθέτουμε ότι ο δήμαρχος της πόλης όπου ζείτε θέλει να μάθει τη γνώμη των δημοτών του για την κατασκευή ενός γηπέδου καλαθοσφαίρισης (basketball). Αν ρωτήσει 100 άτομα που μένουν στη γειτονιά σας, νομίζετε ότι θα μπορούσε να αποκτήσει μια ιδέα για τη γνώμη των δημοτών; Απαντήστε χρησιμοποιώντας τους όρους «πληθυσμό» και «δείγμα».

6. Για να εκτιμήσουμε τις αποδοχές των υπαλλήλων μιας μεγάλης εταιρείας (αριθμός εργαζομένων: 5000 άτομα), καταγράφουμε τις αποδοχές των διευθυντικών στελεχών της εταιρείας. Κατά την άποψή σας θα βγάλουμε σωστά συμπεράσματα για τις αποδοχές των υπαλλήλων της εταιρείας; Δικαιολογήστε τον ισχυρισμό σας.

7. Μεταβλητή ονομάζουμε ....και τιμές

   της μεταβλητής λέγονται ....

8. Οι μεταβλητές διακρίνονται σε δύο κατηγορίες:

   α) ……………………………………… β) …………………………………

9. Οι ποσοτικές μεταβλητές διακρίνονται σε:

   α) …………………………………… β) ………………………………..

10. Αντιστοιχίστε κάθε μεταβλητή της στήλης Α με τον αντίστοιχο χαρακτηρισμό της στήλης Β:

Στήλη Α	**Στήλη Β**
1) Το επάγγελμα 2) Αριθμός τροχαίων ατυχημάτων (σε ένα Σαββατοκύριακο) 3) Μονάδες camping 4) Αφίξεις αλλοδαπών στα σύνορα κατά φύλο 5) Εισπράξεις μουσείων 6) Αριθμός ανέργων αρρένων κατά ομάδα ηλικιών 7) Κατάσταση υγείας	Α. Ποιοτική μεταβλητή Ποσοτική μεταβλητή Β. Διακριτή Γ. Συνεχής

11. α) Ας υποθέσουμε ότι ένας κατασκευαστής διεξάγει μια έρευνα, για να καθορίσει τη λιανική τιμή του προϊόντος του σε μια συγκεκριμένη περιοχή. Αυτή η μεταβλητή είναι διακριτή ή συνεχής;
    β) Σε συνδυασμό με την προηγούμενη μελέτη ο κατασκευαστής θέλει επίσης να προσδιορίσει τον αριθμό των μονάδων του προϊόντος που πουλήθηκαν στην περιοχή αυτή σε μία εβδομάδα, κατά τη διάρκεια της οποίας είχε γίνει διαφήμιση του προϊόντος. Αυτή η μεταβλητή είναι συνεχής ή διακριτή;

12. Οι δύο μεγάλοι τομείς της Στατιστικής είναι:

    α) …………………………………….. και β) ....

    και πρόσφατα έχει αναπτυχθεί ένας τομέας της Στατιστικής

    που λέγεται ....

    Να αναφέρετε συνοπτικά ό,τι γνωρίζετε για τους δύο μεγάλους τομείς της Στατιστικής.

ΕΡΓΑΣΙΑ ΣΤΗΝ ΤΑΞΗ

Η εκμάθηση των βασικών εννοιών της Στατιστικής γίνεται πιο ενδιαφέρουσα, όταν μπορούμε να εργαστούμε με πληροφορίες που συλλέγονται από εμάς και αποθηκεύονται σε έναν υπολογιστή. Η απλούστερη συλλογή δεδομένων μπορεί να αφορά πληροφορίες για τους συμμαθητές σας. Η έρευνα που ακολουθεί μπορεί να δώσει αρκετά δεδομένα με τα οποία να απαντήσετε σε μερικές ενδιαφέρουσες ερωτήσεις. Αποθηκεύστε τα δεδομένα που θα συλλέξετε!

(Σημειώστε με το σύμβολο Χ αν δεν γνωρίζετε την απάντηση)

1. Φύλο
2. Ηλικία
3. Ύψος (σε cm)
4. Ύψος του πατέρα σας
5. Ύψος της μητέρας σας
6. Χρώμα μαλλιών
7. Χρώμα ματιών
8. Είστε δεξιόχειρας, αριστερόχειρας ή αμφίχειρας;
9. Τον βαθμό σας στα μαθηματικά
10. Τον αριθμό της ταυτότητάς σας
11. Τα ενδιαφέροντά σας στον ελεύθερο χρόνο σας
12. Το επάγγελμα που προτιμάτε
13. Προσθέστε μία ή περισσότερες ερωτήσεις που σας ενδιαφέρουν

Στη συνέχεια παρουσιάζεται ένας εύκολος τρόπος συλλογής των δεδομένων, ώστε κάθε μαθητής να έχει από ένα αντίγραφο.

- Μοιράστε ένα έντυπο ερωτηματολογίου, όπου κάθε μαθητής θα γράφει τις απαντήσεις του. Στο τέλος θα πρέπει να υπάρχει κάτι παρόμοιο με το υπόδειγμα που ακολουθεί.
- Κάθε μαθητής να απαντήσει τις ερωτήσεις σε μία σελίδα.
- Να δοθεί ένα αντίγραφο σε κάθε μαθητή.

ΥΠΟΔΕΙΓΜΑ									
α/α	**1 Φύλο**	**2 Ηλικία**	**3 Ύψος**	**4 Ύψος πατέρα**	**5 Ύψος μητέρας**	**6 Μαλλιά**	**7 Μάτια**	**8 Χέρι**	**9 Βαθμός μαθηματικών**
1	Θ	16	167	178	158	Κ	Μ	Α	15
2	Θ	15	160	175	163	Κ	Κ	Δ	09
3	Α	17	170	184	168	Μ	Μ	Δ	18
4	Θ	15	173	175	170	Κ	Μ	Δ	11
5	Θ	16	158	182	168	Ξ	Κ	Α	19

- Αποθηκεύστε τα δεδομένα που συγκεντρώσατε σε Η/Υ για περαιτέρω ανάλυση.
- Προσδιορίστε το είδος κάθε μεταβλητής.

Στο τέλος του πέμπτου κεφαλαίου κάθε μαθητής θα έχει κάνει τη δική του μικρή στατιστική έρευνα.

ΚΕΦΑΛΑΙΟ 2ο

Η ΣΥΜΒΟΛΗ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ ΣΤΗ ΣΥΓΧΡΟΝΗ ΕΠΙΧΕΙΡΗΣΗ

❖ Οι ρίζες της διοίκησης των επιχειρήσεων βρίσκονται στις αρχές του 20ού αιώνα στη σχολή της **«Επιστημονικής Διοίκησης» των Αμερικανών F.W. Taylor, H. Cannt** και **F. Gilbeth** και ξεκίνησε από τη βελτίωση των μεθόδων παραγωγής στη βιομηχανία.

❖ Με τον όρο **«Τεϋλορισμός»** εννοούμε: ένα σύστημα οργάνωσης εργασίας και ελέγχου του χρόνου, που οφείλεται κυρίως στον Αμερικάνο **F. W. Taylor**. Κλασσικό του έργο παραμένει το **«Scientific management».**

ΚΕΦΑΛΑΙΟ 2ο

Η ΣΥΜΒΟΛΗ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ ΣΤΗ ΣΥΓΧΡΟΝΗ ΕΠΙΧΕΙΡΗΣΗ

2.1 Η σύγχρονη διοικητική των επιχειρήσεων (management) – Γενικά

Η οργάνωση και η διοίκηση των επιχειρήσεων (διοικητική των επιχειρήσεων) γίνεται σήμερα με επιστημονικές μεθόδους. Ο ταχύτατος ρυθμός ανάπτυξης της τεχνολογίας, ιδιαίτερα στον τομέα των υπολογιστών και των τηλεπικοινωνιών, η αύξηση του κεφαλαίου των επιχειρήσεων, παράλληλα με τον πολυεθνικό χαρακτήρα που έχουν πολλές από αυτές και η παγκοσμιοποίηση της οικονομίας που συντελείται έχουν επηρεάσει σε μεγάλο βαθμό τη σύγχρονη διοίκηση των επιχειρήσεων.

Όλα τα παραπάνω οδηγούν την επιστήμη της διοίκησης επιχειρήσεων σε αναζήτηση νέων προτύπων και μεθόδων της οργανωτικής δομής και λειτουργίας των επιχειρήσεων και σε επαναπροσδιορισμό των θέσεών της. Σήμερα, θα μπορούσαμε να θεωρήσουμε **τη σύγχρονη διοικητική των επιχειρήσεων ως ένα σύστημα δράσης, που μέσα από τον προγραμματισμό, το συντονισμό και τον έλεγχο στην όλη λειτουργία της επιχείρησης, στοχεύει στη σωστή διαχείριση των πόρων της για την επίτευξη του σκοπού της, που είναι το κέρδος.**

Τελευταία χρησιμοποιείται πολύ και ο όρος **management** για να αποδώσει την παραπάνω έννοια, ο οποίος προήλθε από το γαλλικό «*manager*», που σημαίνει «*διαχειρίζομαι τα του οίκου*».

Από την εποχή του *F. Taylor* και μέχρι σήμερα η επιστήμη της διοίκησης επιχειρήσεων, όπως και η τεχνολογία, εξελίχθηκε τόσο πολύ, ώστε η κάποτε «κραταιά» θεωρία του τεϋλορισμού σήμερα πλέον εγκαταλείπεται.

2.2 Έρευνες αγορών – Διαφήμιση

Ένα από τα πλέον σημαντικά και ζωτικής σημασίας πεδία δραστηριότητας των σύγχρονων επιχειρήσεων είναι αυτά της έρευνας αγορών και της διαφήμισης. Εδώ η συνδρομή και η εφαρμογή της Στατιστικής είναι σοβαρότατη. Οι έρευνες αγορών πολύ απλά στοχεύουν στην ακριβή χαρτογράφηση των αναγκών της αγοράς σε αγαθά και, φυσικά, στον προσδιορισμό του βαθμού της ζήτησής τους.

Η πρώτη στον κόσμο Διαφήμιση Αυτοκινήτου.

Πρόκειται για το μοντέλο Patent-Motor-Waygen της Benz του 1889.
Το διαφημιστικό μήνυμα ήταν:
«Γλιτώνετε από τα έξοδα αμαξά και συντήρησης των αλόγων!»

Για μια επιτυχή έρευνα αγοράς απαιτείται κατηγοριοποίηση των υποψηφίων αγοραστών-καταναλωτών, ώστε να προσδιορίσουμε ευκολότερα σε ποια ή ποιες κατηγορίες αγοραστών θα απευθύνουμε τα διαφημιστικά μηνύματά μας και στη συνέχεια τα προϊόντα μας.

Κριτήρια κατηγοριοποίησης είναι: η ηλικία των υποψήφιων αγοραστών, το φύλο, το επάγγελμα, το μορφωτικό επίπεδο, τα ενδιαφέροντά τους (hobbies) κ.ά.

Συνήθως στις έρευνές μας κάνουμε χρήση ερωτηματολογίων για να διερευνήσουμε τις προθέσεις του αγοραστικού κοινού. Τα βασικά χαρακτηριστικά των ερωτηματολογίων είναι:

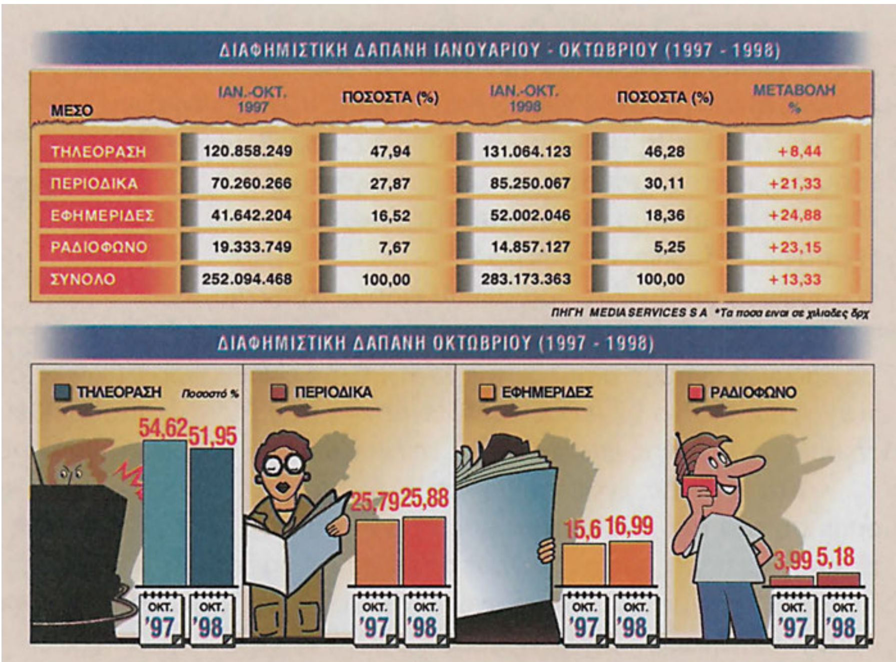
1. **Η σαφήνεια και απλότητα των ερωτήσεων.**
2. **Η διακριτικότητα στον τρόπο διατύπωσης των ερωτήσεων, ώστε να μη θίγεται ή να μη φοβάται να απαντήσει ο ερωτώμενος.**
3. **Η συντομία στις απαντήσεις των ερωτήσεων.**
4. **Η αποδοχή της σπουδαιότητας της έρευνας.**
5. **Η αποφυγή διπλών ερωτήσεων.**
6. **Η αποφυγή υπόδειξης ορισμένου είδους απάντησης.**

Στις έρευνες αγοράς ο ρόλος της Στατιστικής είναι καθοριστικός. Πρέπει να απαντήσει για το μέγεθος του δείγματος που θα χρησιμοποιηθεί, να προσδιορίσει τον τρόπο επιλογής του, την κοινωνική διαστρωμάτωση που θα έχει κτλ. Κατόπιν πρέπει να γίνει η κατάλληλη στατιστική επεξεργασία και η παρουσίαση των αποτελεσμάτων που προέκυψαν κ.ά., ώστε η διοίκηση της επιχείρησης να πάρει ασφαλείς αποφάσεις σε καίρια ζητήματα, όπως για παράδειγμα, την πραγματοποίηση ή όχι επενδύσεων, την επέκταση της εμπορικής δραστηριότητας σε άλλες αγορές κτλ.

Ένα άλλο σημαντικό βήμα είναι η διαπίστωση των τάσεων της αγοράς, ως προς την τιμή, τη διακίνηση και την προώθηση του προϊόντος που μας ενδιαφέρει, καθώς και ο τρόπος που θα φτάσει το μήνυμα παραγωγής του νέου προϊόντος σε όλους τους καταναλωτές ώστε να το δεχθεί και ο πλέον δύσπιστος αγοραστής. Στο σημείο αυτό είναι αναγκαία **η χρησιμοποίηση της διαφήμισης,** για να φτάσει το μήνυμα σε όλους τους αγοραστές, να σπάσει η λεγόμενη αδράνεια της αγοράς και να δοκιμάσει ο καταναλωτής ένα νέο προϊόν.

Φυσικά υπάρχουν οι έρευνες αγορών και οι διαφημίσεις, που γίνονται, για να διατηρηθεί ένα παλαιό προϊόν στην αγορά ή να καλυτερεύσει τη θέση του στις προτιμήσεις των αγοραστών. Σε κάθε περίπτωση οι στατιστικές έρευνες είναι αναγκαίες για να διαπιστώνουμε τα αποτελέσματα μιας διαφήμισης, να επισημαίνουμε τυχόν τρωτά σημεία στην όλη διακίνηση του προϊόντος που μας είναι άγνωστα, με αποτέλεσμα η κάθε επιχείρηση να παρεμβαίνει έγκαιρα, για να αποφεύγονται δυσάρεστες εξελίξεις.

Ο παρακάτω πίνακας από τον ημερήσιο τύπο δείχνει χαρακτηριστικά τη σημασία που έχει αποκτήσει η Στατιστική σε συνδυασμό με τη διαφήμιση στην καθημερινή ζωή.



ΔΙΑΦΗΜΙΣΤΙΚΗ ΔΑΠΑΝΗ ΙΑΝΟΥΑΡΙΟΥ - ΟΚΤΩΒΡΙΟΥ (1997 - 1998)

ΜΕΣΟ	ΙΑΝ.-ΟΚΤ. 1997	ΠΟΣΟΣΤΑ (%)	ΙΑΝ.-ΟΚΤ. 1998	ΠΟΣΟΣΤΑ (%)	ΜΕΤΑΒΟΛΗ %
ΤΗΛΕΟΡΑΣΗ	120.858.249	47,94	131.064.123	46,28	+8,44
ΠΕΡΙΟΔΙΚΑ	70.260.266	27,87	85.250.067	30,11	+21,33
ΕΦΗΜΕΡΙΔΕΣ	41.642.204	16,52	52.002.046	18,36	+24,88
ΡΑΔΙΟΦΩΝΟ	19.333.749	7,67	14.857.127	5,25	+23,15
ΣΥΝΟΛΟ	252.094.468	100,00	283.173.363	100,00	+13,33

ΠΗΓΗ MEDIA SERVICES S A *Τα ποσα ειναι σε χιλιαδες δρχ

Πηγή: *Κυριακάτικη Ελευθεροτυπία 1998*

2.3 Λήψη επιχειρηματικών αποφάσεων

Η συμβολή της Στατιστικής στη διαδικασία λήψης επιχειρηματικών αποφάσεων είναι ουσιώδης, ιδιαίτερα μάλιστα όταν αναφερόμαστε σε συλλογικό επίπεδο υψηλόβαθμων στελεχών και όχι τόσο στις ατομικές καθημερινές και συχνές αποφάσεις που λαμβάνονται από τους υπαλλήλους-στελέχη της επιχείρησης.

Το τμήμα ή **η Διεύθυνση Στατιστικής** είναι το κατεξοχήν αρμόδιο για την επεξεργασία, παρουσίαση και ανάλυση των εσωτερικών δεδομένων της επιχείρησης, που λαμβάνονται κυρίως από το Λογιστήριο, τη Διεύθυνση Πωλήσεων, τη Διεύθυνση Προσωπικού κτλ. και τα οποία αφορούν κυρίως τον έλεγχο της ομαλής εσωτερικής λειτουργίας της επιχείρησης, της πορείας των πωλήσεων κτλ. Η Διεύθυνση Στατιστικής είναι επίσης υπεύθυνη για τη συλλογή πληροφοριών που αφορούν το ύψος των συνολικών πωλήσεων του κλάδου που δραστηριοποιείται η επιχείρηση, το ύψος των διαφημιστικών δαπανών και των προγραμματισμένων επενδύσεων των άλλων ομοειδών επιχειρήσεων κτλ. Όλα αυτά τα δεδομένα τα αναλύει με σκοπό την εξαγωγή χρήσιμων συμπερασμάτων που θα βοηθήσουν ουσιαστικά στη λήψη επιχειρηματικών αποφάσεων από τη διοίκηση της επιχείρησης.

Επίσης η Στατιστική αναλύοντας σειρά δεδομένων προηγούμενων χρονικών περιόδων της επιχείρησης και του κλάδου στον οποίο ανήκει, όπως για παράδειγμα, τις ετήσιες δαπάνες των τελευταίων 15 ετών ή τα ετήσια έσοδα της τελευταίας 15ετίας, έχει τη δυνατότητα με ικανοποιητική προσέγγιση να προβλέψει και να πληροφορήσει τη διοίκηση της επιχείρησης για τις πωλήσεις του επόμενου έτους ή της επόμενης πενταετίας. Έτσι, για παράδειγμα, η επιχείρηση οδηγείται σε μια ορθή λήψη αποφάσεων ως προς τον πενταετή προγραμματισμό των κεφαλαιουχικών επενδύσεών της. Αργότερα στο κεφάλαιο των χρονολογικών σειρών θα δούμε αναλυτικά αυτή την περίπτωση των προβλέψεων.

2.4 Ποιότητα διαχείρισης – Αποτελεσματική διοίκηση

Προαπαιτούμενο προσόν για την παροχή υπηρεσιών ή αγαθών υψηλής ποιότητας από μέρους της επιχείρησης, ώστε να επιβιώσει και να αναπτυχθεί, είναι να διακρίνεται και η ίδια για την ποιότητα της διαχείρισής της και την αποτελεσματική διοίκηση.

Κατ' αρχήν ας καθορίσουμε πότε ένα προϊόν είναι ποιοτικά καλό. Ένα προϊόν **θεωρείται ποιοτικά καλό, όταν ανταποκρίνεται στις προδιαγραφές και στις επιθυμίες του καταναλωτή,** τις οποίες οι επιστημονικά και ποιοτικά οργανωμένες επιχειρήσεις έχουν καταγράψει από προηγούμενες έρευνες αγορών. Είναι γνωστό ότι καμία διαφημιστική εταιρεία δεν θα αναλάβει να παρουσιάσει ένα κακό προϊόν ως καλό, αλλά θα προσπαθήσει ένα καλό προϊόν να το προβάλλει σαν καλύτερο.

Για να επιτευχθούν τα παραπάνω, η επιχείρηση θα πρέπει να έχει μία καλή σχέση κόστους-απόδοσης. Αυτό σημαίνει ότι η επιχείρηση έχει μια διαχείριση ποιότητας, η οποία τη βοηθά να διακριθεί από τις υπόλοιπες του κλάδου της. Η επιτυχία των στόχων της (αποτελεσματικότητα) είναι συνάρτηση των ποιοτικών χαρακτηριστικών της. Ενδεικτικά αναφέρουμε μερικά ποιοτικά χαρακτηριστικά μιας επιχείρησης:

- Ορθή διαχείριση των ανθρώπινων πόρων. Αυτό σημαίνει ενεργοποίηση και ουσιαστική συμμετοχή των εργαζομένων σε κάθε εκδήλωση της επιχειρηματικής δραστηριότητας (ενημέρωση για τους στόχους της επιχείρησης, συνεχής επιμόρφωση, ενθάρρυνση στην ανάπτυξη πρωτοβουλιών κτλ.).
- Υψηλού επιπέδου χρηματοοικονομική διοίκηση, διακρίνοντας τους ευνοϊκούς για την επιχείρηση τρόπους χρηματοδότησης, κάνοντας σωστή αξιολόγηση επενδύσεων και διαχείριση οικονομικών κινδύνων κτλ. Τομέας ιδιαίτερα δύσκολος που απαιτεί λεπτούς χειρισμούς και πολύ εξειδικευμένα στελέχη.
- Δυνατότητες συνεχούς ελέγχου όλων των φάσεων της παραγωγής του προϊόντος και υψηλές προδιαγραφές (standards) ποιοτικού ελέγχου των προϊόντων.
- Τακτικές έρευνες των προθέσεων του καταναλωτικού κοινού, καλή επικοινωνιακή σχέση με την πελατεία και τους συνεργάτες.
- Ταχεία ενσωμάτωση νέων τεχνολογιών, που θα βελτιώσουν ποιοτικά το προϊόν, και νέων μεθόδων που θα βελτιώσουν τη λειτουργία της επιχείρησης.
- Προσπάθειες ώστε η επιχείρηση να είναι καινοτόμος στην όλη λειτουργία και πολιτική της, για να γίνει από τις επώνυμες του κλάδου της.
- Εμμονή στην τήρηση των αρχών λειτουργίας που έχει θέσει η επιχείρηση και στην τήρηση προδιαγραφών ποιότητας, ώστε να μπορέσει να δημιουργήσει η επιχείρηση αυτό που λέμε επώνυμο προϊόν.

Η συμβολή της Στατιστικής, για να επιτύχει μια επιχείρηση την ποιοτική διάσταση τόσο στην παραγωγή του προϊόντος της όσο και στη λειτουργία της, είναι σημαντική. Με τη μελέτη απλών «Ειδικών Περιπτώσεων» (case studies) που θα παρουσιάσουμε μετά το τέλος κάθε κεφαλαίου θα γίνουν περισσότερο κατανοητά τα όσα αναφέραμε.

2.5 Η θέση της Στατιστικής στον ενοποιημένο οικονομικό χώρο

Για όλους τους διεθνείς οργανισμούς, οικονομικούς ή όχι, για παράδειγμα, Ουνέσκο, Ο.Η.Ε, Διεθνής Τράπεζα κτλ., η Στατιστική αποτελεί ένα από τα σοβαρότερα στηρίγματά τους και τα πιο σημαντικά εργαλεία επίτευξης των σκοπών τους. Η Στατιστική θα τους δώσει στοιχεία επεξεργασμένα για να αποκτήσουν μια σωστή εικόνα για ό,τι θέλουν να μελετήσουν ή για να πιστοποιήσουν την επιτυχία ή όχι των μέτρων και των προγραμμάτων που εφάρμοσαν.

Για την Ελλάδα είναι σημαντικό το γεγονός ότι, λόγω της συμμετοχής της στην Ευρωπαϊκή Ένωση (Ε.Ε.), έχει τη δυνατότητα της εύκολης πρόσβασης στα στοιχεία και τις πληροφορίες που αφορούν τις άλλες χώρες της Ε.Ε., μέσω της **Στατιστικής Υπηρεσίας της Ευρώπης- Eurostat.** Έτσι, είναι πιθανό στοιχεία που δεν υπάρχουν στην Εθνική Στατιστική Υπηρεσία της Ελλάδας (Ε.Σ.Υ.Ε.) και τις άλλες εθνικές στατιστικές υπηρεσίες των άλλων κρατών, να υπάρχουν στην Eurostat. Με τη βοήθεια της Eurostat επιχειρείται η ύπαρξη ενιαίας ονοματολογίας και στατιστικής ορολογίας, ώστε να αποφεύγονται πιθανές παρανοήσεις ακόμη και σε απλά θέματα. **Η Eurostat λειτουργεί και ως ο γενικός συντονιστής των στατιστικών υπηρεσιών των 15 χωρών της Ε.Ε.** Αυτό βοηθά στην ακόμη μεγαλύτερη προαγωγή της Στατιστικής και στην επένδυση της μεγάλης σημασίας που έχει για την επιβίωση των επιχειρήσεων σήμερα, γεγονός ιδιαίτερα σημαντικό και για τις ελληνικές επιχειρήσεις, που χρειάζονται πληροφορίες για να στηρίξουν την ανάπτυξή τους.

ΕΡΩΤΗΣΕΙΣ ΚΑΤΑΝΟΗΣΗΣ

1. Η σύγχρονη διοικητική των επιχειρήσων είναι ένα σύστημα δράσης που μέσα από τον .... , το ...., και τον ...., στην όλη λειτουργία της επιχείρησης, στοχεύει ....

2. Ο όρος «management» χρησιμοποιείται .... , προήλθε από .... και σημαίνει ....

3. Οι ρίζες της διοίκησης των επιχειρήσεων χρονολογούνται από ....

4. Οι έρευνες αγορών στοχεύουν..

5. Κριτήρια κατηγοριοποίησης των υποψηφίων αγοραστών είναι: ....

6. Τα βασικά χαρακτηριστικά ενός ερωτηματολογίου είναι:

α) ……………………………….. δ) ………………………………..

β) ……………………………….. ε) ………………………………..

γ) ……………………………….. στ) ………………………………..

7. Να αναφέρετε συνοπτικά την αναγκαιότητα της Στατιστικής στις έρευνες αγορών.

8. Να αναφέρετε συνοπτικά την άποψή σας για τη συμβολή της Στατιστικής στη λήψη επιχειρηματικών αποφάσεων.

9. Το κατεξοχήν αρμόδιο τμήμα για την επεξεργασία, παρουσίαση και ανάλυση στατιστικών δεδομένων είναι ....

10. Είναι ένα προϊόν θεωρείται ποιοτικά καλό, όταν ……………………………

………..…………………………………………………………………………

11. Η διαφήμιση έχει κύριο σκοπό να παρουσιάσει στον καταναλωτή ένα σκάρτο προϊόν ως καλό.

Σ ☐ Λ ☐

12. Να αναφέρετε και να αναπτύξετε πέντε βασικά ποιοτικά χαρακτηριστικά που θα πρέπει να έχει μια άρτια λειτουργούσα επιχείρηση.

13. Ποιος είναι ο ρόλος της Στατιστικής στη λειτουργία των διεθνών οργανισμών;

14. Να αναφέρετε συνοπτικά τη σπουδαιότητα της Ευρωπαϊκής Στατιστικής υπηρεσίας για την Ελλάδα αλλά και για τις άλλες χώρες της Ευρωπαϊκής Ένωσης.

ΚΕΦΑΛΑΙΟ 3ο

ΣΥΛΛΟΓΗ ΚΑΙ ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΤΟ ΧΩΡΟ ΤΩΝ ΕΠΙΧΕΙΡΗΣΕΩΝ

❖ **George Gallup:** Αμερικανός δημοσιογράφος που ίδρυσε το 1935 ινστιτούτο σφυγμομέτρησης της κοινής γνώμης με δειγματοληψία. Ο σχεδιασμός και η εκτέλεση της δειγματοληψίας χρειάζονται ιδιαίτερη προσοχή, για να προκύψουν αξιόπιστα συμπεράσματα, τα οποία θα ισχύουν με ικανοποιητική ακρίβεια για ολόκληρο τον πληθυσμό.

❖ Απογραφή πληθυσμού έγινε στην Κίνα από τον αυτοκράτορα Υάο το 2.238 π.Χ. Στοιχειώδεις απογραφές φαίνεται να έχουν πραγματοποιηθεί από τους Σίνες, τους Αιγυπτίους, τους Πέρσες κ.ά. Στους Ρωμαίους η πρώτη απογραφή πληθυσμού έγινε επί Ρωμύλου (753-715 π.Χ.).

ΚΕΦΑΛΑΙΟ 3ο

ΣΥΛΛΟΓΗ ΚΑΙ ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΤΟ ΧΩΡΟ ΤΩΝ ΕΠΙΧΕΙΡΗΣΕΩΝ

3.1 Γενικά

Η συλλογή των στατιστικών δεδομένων αποτελεί σημαντικό στάδιο κάθε Στατιστικής έρευνας. Απαιτεί ιδιαίτερη προσοχή, διότι, αν τα στοιχεία που θα συγκεντρώσουμε είναι ανακριβή και αναξιόπιστα, τότε θα οδηγηθούμε σε παραπλανητικά συμπεράσματα και λανθασμένες αποφάσεις.

Στη χώρα μας υπάρχει η Εθνική Στατιστική Υπηρεσία Ελλάδας (Ε.Σ.Υ.Ε.), που συγκεντρώνει στατιστικά στοιχεία, τα οποία και δημοσιεύει στην ετήσια στατιστική επετηρίδα ή στο μηνιαίο στατιστικό δελτίο και έτσι μπορεί κάθε ενδιαφερόμενος να έχει τα απαραίτητα στοιχεία για το αντικείμενο έρευνάς του. Εκτός από την Ε.Σ.Υ.Ε. υπάρχουν και άλλοι φορείς που συγκεντρώνουν στατιστικά στοιχεία. Τέτοιοι φορείς είναι τα ινστιτούτα ερευνών, δημόσιοι και ιδιωτικοί οργανισμοί, τράπεζες κτλ. Για τη συλλογή και παρουσίαση στατιστικών στοιχείων εφαρμόζονται διάφορες μέθοδοι, από τις οποίες θα αναφέρουμε τις σπουδαιότερες.

3.2 Συλλογή στατιστικών δεδομένων

α) Η απογραφή (census), με την οποία συγκεντρώνονται πληροφορίες για όλα τα άτομα του πληθυσμού και όχι μόνο για ένα δείγμα του. Ιδιαίτερη σημασία έχει η απογραφή του πληθυσμού μιας χώρας, διότι αποτελεί την κύρια πηγή πληροφοριών ως προς τα δημογραφικά, οικονομικά και κοινωνικά δεδομένα κάθε χώρας. Γι' αυτό η απογραφή αυτή αποτελεί σημαντικό γεγονός και γίνεται με μεγάλη κρατική φροντίδα, ώστε τα στοιχεία που θα συγκεντρωθούν να δίνουν την πραγματική εικόνα της χώρας. Στη χώρα μας απογραφή του πληθυσμού γίνεται κάθε δέκα χρόνια από την Ε.Σ.Υ.Ε., ενώ απογραφή των βιομηχανικών και εμπορικών επιχειρήσεων κάθε πέντε χρόνια. Απογραφές γεωργίας, κτηνοτροφίας κτλ. δεν γίνονται σε καθορισμένο χρόνο.

β) Η δειγματοληψία (sampling) είναι η μέθοδος κατά την οποία συγκεντρώνονται παρατηρήσεις μόνο για ένα μέρος πληθυσμού, **το δείγμα**. Η μέθοδος αυτή εφαρμόζεται, όταν ο πληθυσμός που εξετάζουμε έχει μεγάλο πλήθος στοιχείων ή όταν η απογραφή είναι πρακτικά αδύνατη. Για παράδειγμα, αν θέλουμε να εξετάσουμε τα ελαστικά που κατασκευάζει μια βιομηχανία ως προς την αντοχή τους, δεν μπορούμε να πάρουμε όλα τα ελαστικά, αφού η εξέταση ενός ελαστικού συνεπάγεται την καταστροφή του. Γι' αυτό εξετάζουμε έναν ορισμένο αριθμό ελαστικών (δείγμα) και τα συμπεράσματα που προκύπτουν τα γενικεύουμε για όλα τα ελαστικά (πληθυσμό). Όλοι σήμερα γνωρίζουμε τις δημοσκοπήσεις (Gallup), που δημοσιεύουν τα διάφορα έντυπα για τη λαϊκή υποστήριξη που έχουν τα πολιτικά κόμματα ή οι επικρατέστερες πολιτικές προσωπικότητες μιας χώρας.

γ) Η συνεχής εγγραφή των ατόμων ενός πληθυσμού είναι η μέθοδος κατά την οποία καταχωρούνται σε ειδικά βιβλία πληροφορίες για τα άτομα. Για παράδειγμα, οι καταχωρήσεις που γίνονται στα ληξιαρχεία (γεννήσεων, θανάτων, γάμων) και στις διάφορες κρατικές υπηρεσίες (των αδειών κυκλοφορίας των αυτοκινήτων κτλ.).

Ακολουθώντας είτε τη μέθοδο της απογραφής είτε της δειγματοληψίας συλλέγουμε τις πληροφορίες και τις καταγράφουμε σε ειδικά στατιστικά έντυπα.

Μετά τη συγκέντρωση των στατιστικών στοιχείων έχουμε το στάδιο της επεξεργασίας τους. Κατά την επεξεργασία ελέγχουμε αν συμπληρώθηκαν σωστά οι απαντήσεις των ερωτηματολογίων και αν αντιγράφηκαν σωστά τα διάφορα στοιχεία που ζητήσαμε να συγκεντρωθούν. Στη συνέχεια καταχωρούμε τα στοιχεία σε ομάδες με τέτοιο τρόπο, ώστε όλα τα στοιχεία κάθε ομάδας να έχουν μια ιδιαίτερη χαρακτηριστική ιδιότητα. Σήμερα όλη αυτή η διαδικασία γίνεται με μηχανογράφηση.

3.3 Ταξινόμηση και παρουσίαση στατιστικών δεδομένων

Τα στατιστικά δεδομένα πρέπει να παρουσιάζονται με τρόπο απλό, συνοπτικό και σαφή, ώστε να είναι εύκολη η κατανόησή τους από τον κάθε ενδιαφερόμενο. Η παρουσίαση των στατιστικών στοιχείων μπορεί να γίνει με τρεις τρόπους:

- **με συνοπτικές εκθέσεις ή αναφορές**
- **με στατιστικούς πίνακες**
- **με διαγράμματα**

3.3.1 Εκθέσεις ή αναφορές

Με τον τρόπο αυτό **ενσωματώνουμε τα στατιστικά στοιχεία στο κείμενο των εκθέσεων**, δεν έχουμε όμως το πλεονέκτημα της συνοπτικής παρουσίασης των δεδομένων που επιτυγχάνεται με τους άλλους δύο τρόπους. Έτσι πολλές φορές ο αναγνώστης κουράζεται διαβάζοντας το κείμενο της έκθεσης, ενώ επιπλέον χρειάζεται να έχει καλή μνήμη για να συγκρατήσει τα πιο σημαντικά στοιχεία της. Οι εκθέσεις ή οι αναφορές χρησιμεύουν συνήθως ως ερμηνευτικά σημειώματα των αποτελεσμάτων μιας έρευνας και γι' αυτό χρησιμοποιούνται συνήθως από κοινού με τους πίνακες και τα διαγράμματα. Στη συνέχεια θα αναφερθούμε αναλυτικά στους στατιστικούς πίνακες και στα διαγράμματα, που είναι απαραίτητα στη Στατιστική.

3.3.2 Στατιστικοί πίνακες

Με τους στατιστικούς πίνακες, **οι οποίοι αποτελούν συστηματικές κατατάξεις αριθμητικών δεδομένων σε γραμμές και στήλες,** επιτυγχάνουμε τη συνοπτική παρουσίαση των δεδομένων, ώστε αφενός να διευκολυνόμαστε σε τυχόν συγκρίσεις και αφετέρου να ενημερώνεται ο κάθε ενδιαφερόμενος εύκολα και γρήγορα.

Οι πίνακες μπορούν να διακριθούν σε δύο κατηγορίες:

α) Σε γενικούς πίνακες, οι οποίοι περιέχουν κάθε πληροφορία που μας έχει δώσει μια μεγάλη στατιστική έρευνα. Είναι μεγάλου μεγέθους, περιλαμβάνουν πολλά λεπτομερειακά στοιχεία και αποτελούν πηγές στατιστικών πληροφοριών.

β) Σε ειδικούς πίνακες, οι οποίοι συνήθως αντλούν τα στατιστικά στοιχεία από τους γενικούς πίνακες. Είναι μικρού μεγέθους, συνοπτικοί και απλού περιεχομένου.

Σε κάθε πίνακα που έχει συνταχθεί σωστά, εκτός από το **κύριο σώμα** (κορμό), που περιέχει διαχωρισμένα μέσα στις γραμμές και στις στήλες τα στατιστικά δεδομένα, παρατηρούμε και τα εξής ειδικότερα στοιχεία:

α) Τον **τίτλο**, που γράφεται στο επάνω μέρος, και πρέπει με σαφήνεια να δηλώνει το περιεχόμενο του πίνακα και να είναι περιληπτικός.
β) Τις **επικεφαλίδες των στηλών (και γραμμών),** που δείχνουν συνοπτικά τη φύση και τη μονάδα μέτρησης των δεδομένων.
γ) Την **πηγή,** που γράφεται στο κάτω μέρος του πίνακα και δείχνει την προέλευση των στατιστικών δεδομένων.
δ) Τις **υποσημειώσεις,** που γράφονται στο κάτω μέρος του πίνακα και πριν από την πηγή, αν θεωρείται απαραίτητο να δοθούν κάποιες επεξηγήσεις σχετικά με τις επικεφαλίδες των στηλών ή τις μονάδες μέτρησης των δεδομένων.

Στη συνέχεια θα ασχοληθούμε με τη μελέτη και την παρουσίαση των ειδικών πινάκων, οι οποίοι διακρίνονται σε δύο κατηγορίες, όπως άλλωστε και οι γενικοί πίνακες:

α) **Πίνακες απλής εισόδου**. Όταν οι μονάδες του πληθυσμού που εξετάζουμε ερευνώνται ως προς ένα ποιοτικό ή ποσοτικό χαρακτηριστικό τους, τότε παρουσιάζονται με τους **πίνακες απλής εισόδου**. Τέτοιος πίνακας είναι ο ακόλουθος:

Πίνακας 3.1

Άνεργοι ανά κατηγορία κατά τα έτη 1996-1997 *(σε χιλιάδες)*

Κατηγορία	1996	1997
Σύνολο ανέργων	446,4	440,4
Ήταν σε διαθεσιμότητα	0,9	1,1
Έχουν βρει μια εργασία και περιμένουν να αναλάβουν	5,1	4,5
Άνεργοι κάτω από ένα μήνα	15,8	15,0
Άνεργοι 1-2 μήνες	32,2	30,0
Άνεργοι 3-5 μήνες	54,0	49,8
Άνεργοι 6-11 μήνες	77,1	87,8
Άνεργοι 12 μήνες και άνω (άνεργοι μακράς διάρκειας)	260,0	251,3
Δεν άρχισαν να ζητούν εργασία	1,3	0,8
«ΝΕΟΙ» άνεργοι	228,9	218,3
Ποσοστά % ανέργων μακράς διάρκειας	58,2	57,1
Ποσοστά % «ΝΕΩΝ» ανέργων	51,3	49,6

Πηγή: *Ε.Σ.Υ.Ε*

Άλλα παραδείγματα πινάκων απλής εισόδου αποτελούν οι πίνακες 3.2 και 3.3 της επόμενης σελίδας

Πίνακας 3.2

Η ηλικία της Ευρωβουλής

ΚΛΑΣΗ ΗΛΙΚΙΑΣ	ΑΡΙΘΜΟΣ	ΠΟΣΟΣΤΟ
80	2	0,32%
75-79	2	0,32%
70-74	9	1,44%
65-69	38	6,08%
60-64	49	7,82%
55-59	96	15,33%
50-54	120	19,17%
45-49	97	15,49%
40-44	74	11,82%
35-39	45	7,9%
30-34	21	3,35%
25-30	2	0,32%
Αγνωστη	42	6,71%

Πηγή: *Κυριακάτικη Ελευθεροτυπία – Ευρωεκλογές 1999*

Πίνακας 3.3

Το επαγγελματικό προφίλ των ευρωβουλευτών

	ΑΡΙΘΜΟΣ	ΠΟΣΟΣΤΟ
Δικηγόροι-νομικοί	74	11,82%
Καθηγητές Πανεπιστημίων	65	10,38%
Δημόσιοι υπάλληλοι	42	6,71%
Δημοσιογράφοι-εκδότες	41	6,55%
Καθηγητές Λυκείων-Γυμνασίων	58	9,26%
Επιχειρηματίες-στελέχη επιχειρήσεων	57	9,10%
Μηχανολόγοι-φυσικοί -χημικοί-αρχιτέκτονες	31	4,95%
Γεωλόγοι-γεωπόνοι	25	3,99%
Οικονομολόγοι-τραπεζίτες	33	5,27%
Πολιτικοί	47	7,51%
Δημοδιδάσκαλοι	7	1,12%
Εργάτες-συνδικαλιστές	26	4,15%
Ιδιωτικοί υπάλληλοι	11	1,76%
Ιατροί	27	4,31%
Διπλωμάτες	8	1,28%
Τεχνίτες-μηχανικοί	17	2,71%
Πρώην στρατιωτικοί	5	0,80%
Δικαστές-εισαγγελείς	5	0,80%
Ιερείς	4	0,63%
Μεταφραστές-διερμηνείς	7	1,12%
Νοσοκόμοι/ες	3	0,48%
Οικοκυρές	1	0,16%
Καλλιτέχνες-συγγραφείς	10	1,60%
Διάφορα	3	0,48%
Αγνωστο	19	3,03%

Πηγή: *Κυριακάτικη Ελευθεροτυπία – Ευρωεκλογές 1999*

β) Πίνακες διπλής εισόδου. Όταν οι μονάδες του πληθυσμού που εξετάζουμε μελετώνται συγχρόνως ως προς δύο ποιοτικά ή ποσοτικά χαρακτηριστικά, τότε χρησιμοποιούμε τους **πίνακες διπλής εισόδου**. Τέτοιοι πίνακες είναι οι ακόλουθοι:

Πίνακας 3.4

Άνεργοι κατά ομάδες ηλικιών (έως 29 ετών), φύλο και επίπεδο εκπαίδευσης, στο σύνολο της χώρας κατά τα έτη 1995-1997 *(σε χιλιάδες)*

Επίπεδο εκπαίδευσης	Έτη	Σύνολο		Έως 19 ετών		20-24 ετών		25-29 ετών	
		Α	Θ	Α	Θ	Α	Θ	Α	Θ
Τριτοβάθμια	1995	32,5	60,5	-	-	5,1	20,7	10,9	23,6
	1996	34,9	67,2	-	-	8,0	22,2	13,1	25,5
	1997	34,6	66,7	-	-	6,8	21,8	14,4	24,0
Δευτεροβάθμια	1995	90,1	131,5	9,5	24,7	32,7	46,4	20,3	26,2
	1996	86,6	147,3	11,9	28,2	29,3	48,9	17,9	27,8
	1997	92,2	141,4	11,9	22,2	32,5	49,3	18,4	26,2
Πρωτοβάθμια/ή	1995	53,5	56,6	4,5	3,9	4,8	4,5	5,2	5,2
Δεν πήγαν	1996	45,6	64,8	4,9	5,0	3,9	5,3	4,3	6,4
Σχολείο	1997	46,2	59,2	4,6	4,3	3,6	4,6	4,4	4,0

Α = Άρρενες Θ = Θήλεις ***Πηγή:*** *Ε.Σ.Υ.Ε*

Πίνακας 3.5

Δείκτης αξίας Λιανικών Πωλήσεων Απριλίου 1998									
Ανάλυση του Δείκτη Αξίας Λιανικών Πωλήσεων κατά περιοχές και κατηγορίες									
Ιανουάριος (1988 = 100,0)									
Κατηγορίες	**Σύνολο Χώρας**			**Περιφέρεια Πρωτεύουσας**			**Λοιπή χώρα**		
1-4	**1998**	**1997**	**Μετα-βολές %**	**1998**	**1997**	**Μετα-βολές %**	**1998**	**1997**	**Μετα-βολές %**
Γενικός	321,8	301,0	6,9	353,3	309,0	8,5	297,7	278,8	6,8
Διατροφής	359,7	335,4	7,3	388,8	365,7	6,3	321,3	295,4	8,8
Ένδυση-υπόδηση	401,6	366,4	9,6	459,8	415,9	10,6	361,5	321,2	12,5
Επίπλων-οικιακού εξοπλισμού &	269,5	255,1	5,7	287,6	281,2	2,2	235,3	217,7	8,1
Λοιπών ειδών	301,1	289,4	4,0	288,3	267,4	7,8	301,2	297,6	1,2

- ***Πίνακες κατανομής συχνοτήτων***

Μία πολύ συνηθισμένη μορφή πινάκων, που χρησιμοποιούμε στη Στατιστική για τη μελέτη των στατιστικών στοιχείων, είναι οι πίνακες κατανομής συχνοτήτων. Οι πίνακες αυτοί συντάσσονται με κατάλληλη κατάταξη και συστηματική ομαδοποίηση των τιμών της μεταβλητής που εξετάζουμε.

Εδώ θα παρουσιάσουμε πώς γίνεται η σύνταξη αυτών των πινάκων, όταν έχουμε να μελετήσουμε: α) μια διακριτή μεταβλητή και β) μια συνεχή μεταβλητή.

Α. Διακριτή μεταβλητή

Τα παρακάτω δεδομένα (παρατηρήσεις) αναφέρονται στον αριθμό των υπαλλήλων των λογιστηρίων 30 εμπορικών επιχειρήσεων:

5, 6, 7, 2, 4, 5, 3, 2, 5, 4, 1, 2, 3, 4, 5, 5, 6, 6, 4, 3, 3, 4, 4, 5, 6, 5, 4, 7, 7, 4.

Εδώ μας ενδιαφέρει να έχουμε μια σύντομη και συνοπτική εικόνα του πλήθους των υπαλλήλων του λογιστηρίου που απασχολούνται στις 30 εμπορικές επιχειρήσεις. Γι' αυτό διευκολύνεται η μελέτη μας, αν παρουσιάσουμε τα στοιχεία από το μικρότερο προς το μεγαλύτερο.

Κατ' αρχήν λοιπόν τοποθετούμε τις παρατηρήσεις κατά αύξουσα* σειρά (από τη μικρότερη προς τη μεγαλύτερη τιμή), δηλαδή:

1, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 6, 6, 6, 6, 7, 7, 7.

Μια τέτοια παρουσίαση στατιστικών στοιχείων λέγεται **«διάταξη των στοιχείων»**. Όταν όμως το πλήθος των στοιχείων είναι μεγάλο η διάταξή τους είναι επίπονη και γι' αυτό είναι απαραίτητη η χρήση Η/Υ. Στη συνέχεια εφαρμόζουμε τη μέθοδο της διαλογής.

Αν συμβολίσουμε τη μεταβλητή «αριθμός υπαλλήλων λογιστηρίου» με Χ, θα παρατηρήσουμε ότι παίρνει τιμές (x_i) από 1 μέχρι και 7.

Θα μπορούσε κάποιος να ρωτήσει: «πόσες επιχειρήσεις απασχολούν τρία άτομα στο λογιστήριό τους;» ή «πόσα άτομα απασχολούν συνήθως οι περισσότερες επιχειρήσεις;». Τέτοιες ερωτήσεις μπορούν να απαντηθούν εύκολα, αν παρουσιάσουμε τα παραπάνω στοιχεία σ' έναν πίνακα κατανομής συχνοτήτων.

* Η φθίνουσα, δηλαδή από τη μεγαλύτερη προς τη μικρότερη τιμή.

Πίνακας 3.6

Αριθμός υπαλλήλων λογιστηρίου 30 εμπορικών επιχειρήσεων

Αριθμός υπαλλήλων λογιστηρίου (x_i)	Διαλογή	Αριθμός Παρατηρήσεων (συχνότητα) v_i
1	I	1
2	III	3
3	IIII	4
4	IIII IIII	8
5	IIII III	7
6	IIII	4
7	III	3
Σύνολο		30

Έτσι παρατηρούμε ότι 4 επιχειρήσεις απασχολούν τρία άτομα, ενώ ο μεγαλύτερος αριθμός των επιχειρήσεων (δηλαδή 8) απασχολεί 4 άτομα στο λογιστήριο. Ο αριθμός 4 λέγεται συχνότητα της τιμής 3 και ο αριθμός 8 λέγεται συχνότητα της τιμής 4 της μεταβλητής Χ.

Μπορούμε τώρα να διατυπώσουμε τον ορισμό:

Συχνότητα μιας τιμής x_i της μεταβλητής Χ λέγεται ο φυσικός αριθμός v_i που φανερώνει πόσες φορές παρουσιάζεται στο δείγμα (ή σ' ολόκληρο τον πληθυσμό, αν αναφερόμαστε σ' αυτόν) η τιμή αυτή.

Είναι φανερό ότι το άθροισμα των συχνοτήτων ισούται με το σύνολο των παρατηρήσεών μας. Στο παράδειγμά μας είναι: 1 + 3 + 4 + 8 + 7 + 4 + 3 = 30.

Γενικά, αν v_1, v_2, v_3, ... v_κ είναι οι συχνότητες των x_1, x_2, ...x_κ τιμών μιας μεταβλητής Χ, τότε

$$v_1 + v_2 + v_3 + \ldots + v_\kappa = v,$$

όπου v το πλήθος των παρατηρήσεων.

Χάρη ευκολίας θα χρησιμοποιούμε το απλό σύμβολο Σ (άθροισμα) που μας βοηθάει να γράφουμε με συντομία μεγάλα αθροίσματα, δηλαδή:

$$\nu_1 + \nu_2 + \nu_3 + \ldots + \nu_\kappa = \sum \nu_i = \sum \nu = \nu, \quad 1 \le i \le \kappa$$

Σε δείγματα μεγάλου μεγέθους εξυπηρετεί περισσότερο η χρήση ποσοστού αντί της συχνότητας της τιμής x_i της μεταβλητής. Για παράδειγμα, πιο εύκολα συγκρίνουμε τα μεγέθη, αν ρωτήσουμε τι ποσοστό των επιχειρήσεων απασχολεί 4 υπαλλήλους στο λογιστήριο. Η απάντηση θα ήταν: $\frac{8}{30} = \mathbf{0{,}27} \cong \mathbf{27\%}$ των επιχειρήσεων. Ο αριθμός $\frac{8}{30}$ λέγεται **σχετική συχνότητα** και ο αριθμός 27% λέγεται **επί τοις εκατό σχετική συχνότητα.**

Έτσι διατυπώνουμε τον ορισμό:

Σχετική συχνότητα (f) μίας τιμής x_i της μεταβλητής X λέγεται το πηλίκο της συχνότητας ν_i προς το πλήθος ν των παρατηρήσεων, δηλ. $f_i = \frac{\nu_i}{\nu}$. Η σχετική συχνότητα εκφράζεται συνήθως σε ποσοστό επί τοις εκατό, δηλ. $f_i\% = \frac{\nu_i}{\nu} \cdot 100$

Συμπληρώνουμε λοιπόν τον πίνακα:

Πίνακας 3.7

Σχετικές συχνότητες της μεταβλητής: «αριθμός υπαλλήλων λογιστηρίου»

Αριθμός υπαλλήλων	Συχνότητα	Σχετική Συχνότητα	
x_i	ν_i	f_i	f_i %
1	1	0,033	3,3 %
2	3	0,1	10 %
3	4	0,133	13,3 %
4	8	0,267	26,7 %
5	7	0,234	23,4 %
6	4	0,133	13,3 %
7	3	0,1	10 %
Σύνολο 28	30	1	100 %

Θα μπορούσαμε επίσης να ρωτήσουμε πόσες επιχειρήσεις απασχολούν το πολύ 4 υπαλλήλους στο λογιστήριό τους. Από τον πίνακα 3.6 παρατηρούμε ότι, αν προσθέσουμε τις συχνότητες των τιμών 1, 2, 3, 4, θα βρούμε 1 + 3 + 4 + 8 = 16 επιχειρήσεις. Μπορούμε λοιπόν να ορίσουμε ως:

Αθροιστική συχνότητα N_i μιας τιμής x_i λέγεται το άθροισμα των συχνοτήτων των τιμών της μεταβλητής που είναι μικρότερες ή ίσες με την τιμή αυτή.

Η αθροιστική συχνότητα υπολογίζεται ως εξής: Αν οι τιμές x_1, x_2, …, x_i, …, $x_κ$ μιας ποσοτικής μεταβλητής είναι σε αύξουσα τάξη, τότε η αθροιστική συχνότητα της τιμής x_i είναι:

$$N_i = v_1 + v_2 + v_3 + \ldots + v_i \text{ για } i = 1, 2, \ldots, κ.$$

Είναι ακόμη πολύ πιθανό να ακούσουμε την ακόλουθη ερώτηση: «Τι ποσοστό των επιχειρήσεων απασχολεί το πολύ 4 υπαλλήλους στο λογιστήριο;»

Είναι φανερό ότι η απάντηση στο ερώτημα είναι: $\frac{16}{30} = 0{,}533 = 53{,}3\%$ των επιχειρήσεων, δηλαδή 3,3% + 10% + 13,3% + 26,7% = 53,3%. Ο αριθμός 53,3% λέγεται σχετική αθροιστική συχνότητα της τιμής $x_4 = 4$, δηλαδή:

Σχετική αθροιστική συχνότητα F_i ή επί τοις εκατό σχετική αθροιστική συχνότητα $F_i\%$ της τιμής x_i λέγεται το άθροισμα συχνοτήτων f_i ή $f_i\%$ των τιμών που είναι μικρότερες ή ίσες προς αυτή.

Γενικά η σχετική αθροιστική συχνότητα είναι:

$$F_i = f_1 + f_2 + f_3 + \ldots + f_i, \text{ για } i = 1, 2, \ldots, κ.,$$

και η επί τοις εκατό σχετική αθροιστική συχνότητα: $F_i\% = 100 \cdot F_i$.

Ο επόμενος πίνακας συμπληρώνει τους πίνακες 3.6 και 3.7:

Πίνακας 3.8

Πίνακας συχνοτήτων και αθροιστικών συχνοτήτων της μεταβλητής «αριθμός υπαλλήλων λογιστηρίου»

Αριθμός υπαλλήλων λογιστηρίου x_i	**Συχνότητα** v_i	**Σχετική συχνότητα** $f_i\%$	**Αθροιστική συχνότητα** N_i	**Σχετική αθροιστική συχνότητα** $F_i\%$
1	1	3,3	1	3,3
2	3	10	1 + 3 = 4	13,3
3	4	13,3	1 + 3 + 4 = 8	26,6
4	8	26,7	1 + 3 + 4 + 8 = 16	53,3
5	7	23,4	1 + 3 + 4 + 8 + 7 = 23	76,7
6	4	13,3	1 + 3 + 4 + 8 + 7 + 4 = 27	90
7	3	10	1 + 3 + 4 + 8 + 7 + 4 + 3 = 30	100
Σύνολο	30	100		

Β. Συνεχής μεταβλητή

Όταν σε κάποιο πρόβλημα η μεταβλητή είναι συνεχής, δηλ. μπορεί να πάρει κάθε τιμή μεταξύ δύο αριθμών α και β με α < β ή αν το πλήθος των τιμών που παίρνει η μεταβλητή είναι πολύ μεγάλο, τότε η προηγούμενη παρουσίαση των παρατηρήσεων με πίνακες και διαγράμματα είναι πρακτικά δύσκολη και γι' αυτό καταφεύγουμε σε ομαδοποίηση των παρατηρήσεων*.

Έτσι αν έχουμε 50 μαθητές και τους μελετάμε ως προς το ύψος τους, τότε δεν είναι εύκολο να φτιάξουμε πίνακα γράφοντας όλες τις τιμές της μεταβλητής «ύψος», γιατί αυτό μπορεί να πάρει κάθε τιμή, για παράδειγμα από 155cm μέχρι 189cm.

Για ευκολία τα δεδομένα παρουσιάζονται κατά αύξουσα σειρά:
155, 156, 159, 160, 162, 162, 164, 166, 166, 167, 167, 167, 167, 167, 168, 168, 169, 171, 171, 171, 171, 172, 172, 172, 172, 172, 172, 172, 173, 174, 176, 176, 177, 177, 177, 178, 178, 178, 179, 179, 181, 181, 181, 182, 182, 183, 184, 185, 186, **189.**

* Το ίδιο συμβαίνει, όταν η μεταβλητή είναι διακριτή και το πλήθος πολύ μεγάλο.

Η διαφορά της μικρότερης τιμής m = 155 από τη μεγαλύτερη τιμή M = 189 ονομάζεται **εύρος** (range) και συμβολίζεται με **R**, δηλαδή:

R = (τιμή της μεγαλύτερης παρατήρησης – τιμή της μικρότερης παρατήρησης)

R = M - m

Στο συγκεκριμένο παράδειγμα επειδή το εύρος του δείγματος R = 189 - 155 = 34 είναι σχετικά μικρό, χωρίζουμε το διάστημα [155, 190) σε 7 υποδιαστήματα που ονομάζονται **κλάσεις ή τάξεις** και το καθένα έχει πλάτος c = 5 cm. Τιμές ίσες με τα ανώτερα όρια των κλάσεων συμπεριλαμβάνονται στις αμέσως επόμενες κλάσεις. Έτσι η τιμή 160 συμπεριλαμβάνεται στη 2η κλάση ενώ η 165 συμπεριλαμβάνεται στην 3η κλάση, δηλαδή η 2η κλάση είναι το ανοικτό δεξιά διάστημα [160, 165) κ.ο.κ. Σύμφωνα με τα παραπάνω κατασκευάζουμε τον πίνακα συχνοτήτων της μεταβλητής «ύψος»:

Πίνακας 3.9

Κλάση αναστημάτων [,) cm	**Κεντρική τιμή x'_i / cm**	**Συχνότητα ν_i**	**Σχετική συχνότητα f_i %**
155-160	157,5	3	6
160-165	162,5	4	8
165-170	167,5	10	20
170-175	172,5	13	26
175-180	177,5	10	20
180-185	182,5	7	14
185-190	187,5	3	6
Σύνολο		50	100

Η κεντρική τιμή της 1ης κλάσης είναι ο αριθμός:

$$x_1' = \frac{155+160}{2} = 157{,}5$$

Ομοίως:

η κεντρική τιμή της 2ης κλάσης είναι ο αριθμός: $x_2' = \frac{160+165}{2} = 162{,}5$ κ.ο.κ.
Το πλάτος κάθε κλάσης ισούται με 5 cm.

Τονίζουμε ότι οι συχνότητες αναφέρονται στην κλάση και όχι στις μεμονωμένες τιμές της μεταβλητής. Για παράδειγμα, η κλάση [165, 170) έχει συχνότητα 10, σημαίνει ότι 10 τιμές της μεταβλητής βρίσκονται σ' αυτή την κλάση. Παρακάτω για να διευκολύνουμε τους υπολογισμούς θα λέμε ότι **η κεντρική τιμή της κλάσης**, δηλ. το 167,5 cm έχει τη συχνότητα 10 της κλάσης.

Γενικεύοντας λοιπόν:

Για να ομαδοποιήσουμε τις παρατηρήσεις χωρίζουμε το διάστημα μεταβολής ($α_ο$, $α_κ$) της μεταβλητής X σε υποδιαστήματα της μορφής [$α_{i-1}$, $α_i$) που ονομάζονται κλάσεις, όπως φαίνεται στον παρακάτω πίνακα:

Πίνακας 3.10

Κλάση [,)	**Κεντρική τιμή (xi′)**	**Συχνότητα ($ν_i$)**
$α_0 - α_1$	x_1'	$ν_1$
$α_1 - α_2$	x_2'	$ν_2$
.	.	.
.	.	.
.	.	.
$α_{i-1} - α_i$	x_i'	$ν_i$
.	.	.
.	.	.
.	.	.
$α_{κ-1} - α_κ$	$x_κ'$	$ν_κ$
		$Σν_i = ν$

Τα άκρα των κλάσεων λέγονται αντίστοιχα, το μεν $α_{i-1}$ **κατώτερο όριο** (lower limit) το δε $α_i$ **ανώτερο όριο** (upper limit) της κλάσης. Η διαφορά $\boldsymbol{α_i - α_{i-1}}$ ονομάζεται **πλάτος κλάσης** (class width) ή **διάστημα κλάσης** (class interval) και συμβολίζεται με c. Επειδή οι παρατηρήσεις σε κάθε κλάση θεωρείται ότι κατανέμονται ομοιόμορφα, μπορούν να αντιπροσωπευθούν από την κεντρική τιμή ή κεντρικό όρο της κλάσης, που υπολογίζεται από τη σχέση:

$$x_i' = \frac{\alpha_{i-1} + \alpha_i}{2} \qquad (1)$$

δηλαδή από το ημιάθροισμα του κατώτερου και ανώτερου ορίου κάθε κλάσης.

Οι συχνότητες ν_i μας δίνουν τον αριθμό των παρατηρήσεων που περιέχονται στις αντίστοιχες κλάσεις. Ακόμα με **M** και **m** συμβολίζουμε τη μεγαλύτερη και τη μικρότερη τιμή αντίστοιχα που έχει πάρει η μεταβλητή X.

Ένα σημαντικό ερώτημα είναι, σε πόσα υποδιαστήματα –κλάσεις– θα χωρίσουμε το διάστημα μεταβολής της μεταβλητής X. Υπάρχουν κάποιοι εμπειρικοί τύποι που δίνουν απάντηση σε ορισμένες περιπτώσεις, γενικά όμως δεχόμαστε ότι πρέπει να επιτυγχάνεται η ομοιογένεια των δεδομένων και η απλότητα της παρουσίασής τους. Στις πρακτικές εφαρμογές, ο αριθμός των κλάσεων κυμαίνεται κατά μέσον όρο στις 8 με 10. Επίσης, στην πλειονότητα των πρακτικών εφαρμογών έχουμε κλάσεις ίσου πλάτους. Φυσικά όμως υπάρχουν και περιπτώσεις όπου αναγκαζόμαστε να έχουμε κλάσεις άνισου πλάτους, όπως για παράδειγμα στις κατανομές δαπανών, ημερών ανεργίας κτλ.

3.3.3 Στατιστικά διαγράμματα

Οι στατιστικοί πίνακες κατανομής συχνοτήτων των μεταβλητών είναι ένας χρήσιμος τρόπος παρουσίασης των δεδομένων και διευκολύνει την επεξεργασία τους. Έχει παρατηρηθεί όμως ότι η χρήση γραφημάτων στην παρουσίαση των δεδομένων μας δίνει τη δυνατότητα να σχηματίσουμε μια γρήγορη και σαφή εικόνα των δεδομένων. Η πληροφορία που παίρνουμε από το σχήμα της κατανομής είναι σημαντική για την κατανόηση του χαρακτηριστικού που μελετάμε.

Τα στατιστικά διαγράμματα (diagrams) ή γραφικές απεικονίσεις αποτελούν έναν τρόπο παρουσίασης των στατιστικών μας δεδομένων με τη χρησιμοποίηση γεωμετρικών σχημάτων ή άλλων συμβόλων.

Όπως και στην περίπτωση των πινάκων έτσι και εδώ ένα στατιστικό διάγραμμα πρέπει να έχει τα εξής βασικά στοιχεία:

α) **Τον τίτλο.**
β) **Την κλίμακα με τις τιμές μεγεθών** που απεικονίζονται.
γ) **Το υπόμνημα,** που επεξηγεί συνήθως τις τιμές της μεταβλητής, και
δ) **Την πηγή των δεδομένων.**

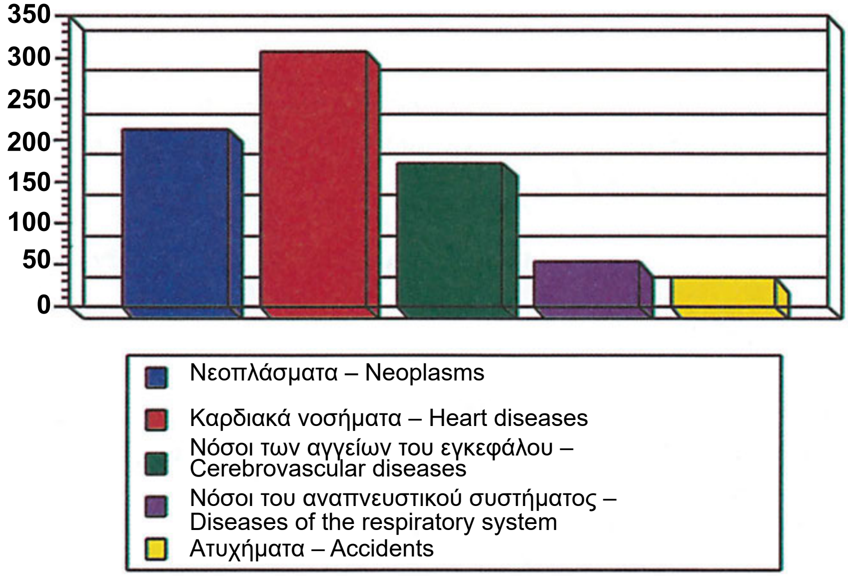
Παρακάτω παραθέτουμε μερικά από τα κυριότερα είδη διαγραμμάτων:

- **Ραβδόγραμμα** (barchart)

Το ραβδόγραμμα χρησιμοποιείται για τη γραφική παράσταση των τιμών μιας **ποιοτικής μεταβλητής** και αποτελείται από κατακόρυφα ή οριζόντια ορθογώνια ίσου πλάτους (με κενά μεταξύ τους) που λέγονται ράβδοι. Σε κάθε τιμή της μεταβλητής X αντιστοιχεί μία ορθογώνια στήλη, της οποίας το ύψος είναι ίσο με συχνότητα v_i ή τη σχετική συχνότητα f_i της τιμής αυτής. Όταν στον κατακόρυφο άξονα βάλουμε τις συχνότητες v_i, τότε το γράφημα ονομάζεται **ραβδόγραμμα συχνοτήτων**, ενώ όταν βάλουμε τις σχετικές συχνότητες έχουμε το **ραβδόγραμμα σχετικών συχνοτήτων.**

Στο ραβδόγραμμα 3.1 παρουσιάζονται οι κυριότερες αιτίες θανάτου στην πατρίδα μας το έτος 1996 με βάση τα στοιχεία της απογραφής του 1991.

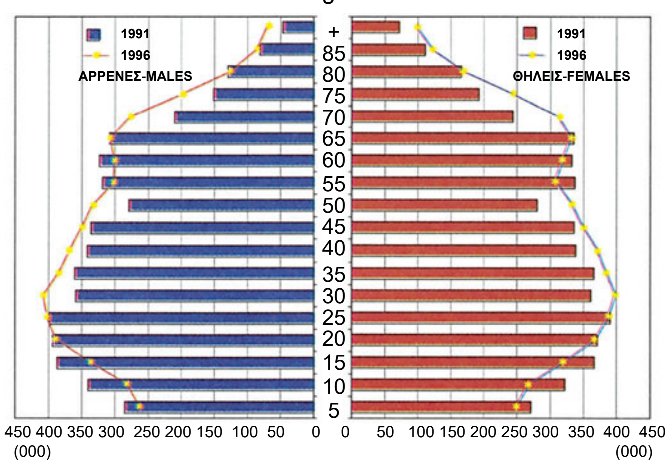
ΚΥΡΙΟΤΕΡΕΣ ΑΙΤΙΕΣ ΘΑΝΑΤΟΥ – 1996 – MAIN CAUSES OF DEATH



(Σχ. 3.1) Ραβδόγραμμα **Πηγή:** *Ε.Σ.Υ.Ε*

Το ραβδόγραμμα 3.2 δείχνει την κατανομή ανά ομάδες ηλικιών του πληθυσμού της Ελλάδας σύμφωνα με την απογραφή του 1991

ΠΛΗΘΥΣΜΟΣ – POPULATION
(Απογραφή 1991 – Εκτίμηση 30.6.1996)
(1991 Census – Estimation on 30.6.1996)
Ηλικία
Age



(Σχ.3.2) Ραβδόγραμμα **Πηγή:** *Ε.Σ.Υ.Ε*

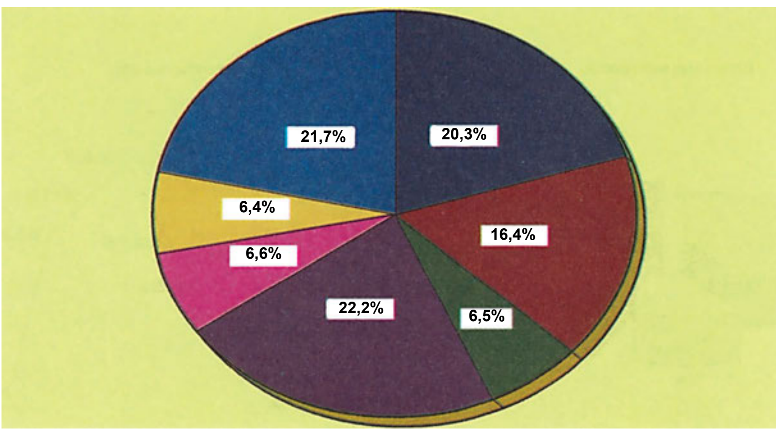
- **Κυκλικό διάγραμμα**

Τα κυκλικά διαγράμματα χρησιμοποιούνται κυρίως για τη γραφική απεικόνιση των σχετικών συχνοτήτων τόσο των ποιοτικών όσο και των ποσοτικών μεταβλητών. Τις σχετικές συχνότητες τις παριστάνουμε με κυκλικούς τομείς ενός ολόκληρου κύκλου. Έτσι ολόκληρος ο κύκλος παριστάνει τη σχετική συχνότητα 100%. Το σχήμα που προκύπτει με τον τρόπο αυτό λέγεται **κυκλικό διάγραμμα**. Για να κατασκευάσουμε, για παράδειγμα, ένα τέτοιο κυκλικό διάγραμμα χωρίζουμε τον κυκλικό δίσκο σε κυκλικούς τομείς των οποίων οι επίκεντρες γωνίες βαίνουν σε τόξα ανάλογα με τις συχνότητες των τιμών της μεταβλητής. Η γωνία 360° (πλήρης κύκλος) αντιστοιχεί στο σύνολο ν των παρατηρήσεων. Το τόξο $α_i$ αντιστοιχεί στη συχνότητα $ν_i$. Τα ποσά είναι ανάλογα, οπότε:

$$\frac{α_i}{360^\circ} = \frac{ν_i}{ν} \quad α_i = \frac{ν_i}{ν} \cdot 360^\circ \ , \ i = f_i \cdot 360^\circ \ , \ i = 1, 2, \ldots, κ$$

Το κυκλικό διάγραμμα του σχήματος 3.3 παρουσιάζει τα ποσοστά των ατόμων που απασχολούνταν κατά κλάδο οικονομικής δραστηριότητας στην Ελλάδα το 1996.

Απασχολούμενοι κατά κλάδο οικονομικής δραστηριότητας
(Έρευνα Εργατικού Δυναμικού 1996)
Person employed by branch of economic activity
(Labour Force Survey, 1996)



(Σχ. 3.3) **Πηγή:** Ε.Σ.Υ.Ε

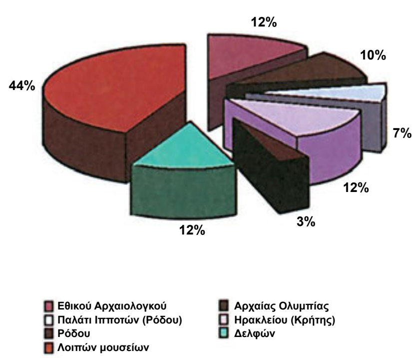
- Γεωργία, Κτηνοτροφία, Δάση, Κυνήγι, Αλιεία
- Ορυχεία, Μεταποίηση, Ηλεκτρισμός, Φωταέριο
- Οικοδομήσεις, Δημόσια έργα
- Εμπόριο, Εστιατόρια, Ξενοδοχεία
- Τράπεζες, Ασφάλειες
- Λοιπές Υπηρεσίες
- Μεταφορές, Επικοινωνίες

Ένα ακόμα παράδειγμα κυκλικού διαγράμματος δίνεται στο σχήμα 3.4

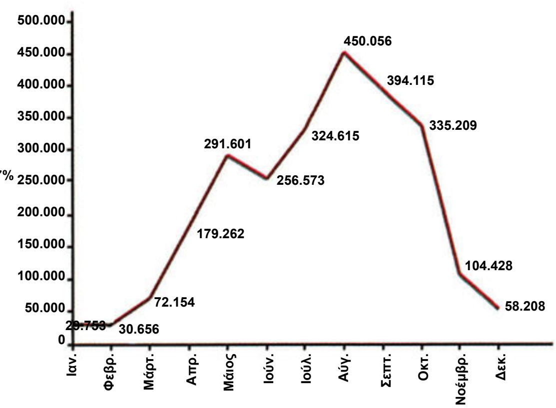
Επισκέπτες μουσείων: 1997

Μουσεία	**Ιαν.**	**Φεβρ.**	**Μάρτ.**	**Απρ.**	**Μάιος**	**Ιούν.**	**Ιούλ.**	**Αύγ**	**Σεπτ.**	**Οκτ.**	**Νοέμβρ.**	**Δεκ.**	**Σύνολο**
Εθνικού Αρχαιολογικού	6.452	10.374	13.552	25.171	36.116	32.238	36.343	44.763	41.764	35.184	8.329	4.301	294.587
Αρχαίας Ολυμπίας	1.947	1.268	6.184	18.701	35.624	27.915	35.248	42.910	42.834	28.863	5.955	2.503	249.952
Παλατίου Ιπποτών (Ρόδου)	1.300	820	2.400	12.880	260	230	30.200	43.500	32.300	28.000	3.400	2.703	157.990
Ηρακλείου (Κρήτης)	1.500	1.500	5.900	27.800	10.900	39.400	50.100	62.300	51.500	42.300	3.200	2.000	298.400
Ρόδου	700	600	990	500	10.900	9.100	11.000	16.000	11.600	11.700	1.700	900	75.690
Δελφών	5.100	4.300	12.200	29.500	41.500	33.800	33.500	40.300	46.200	34.600	8.300	4.900	294.200
Λοιπών μουσείων	12.754	11.794	31.018	60.210	96.061	91.120	128.224	200.283	167.917	154.562	73.544	40.904	1.068.391
Σύνολο	29.753	30.656	72.154	179.262	291.601	256.573	324.615	450.056	394.115	335.209	104.428	58.208	2.526.630

Επισκέπτες ανά μουσείο



Επισκέπτες ανά μήνα



(Σχ. 3.4)

Πηγή: *Ε.Σ.Υ.Ε*

- **Διάγραμμα συχνοτήτων** (frequency diagram)

Στην περίπτωση που έχουμε μια ποσοτική διακριτή μεταβλητή, αντί του ραβδογράμματος χρησιμοποιούμε το διάγραμμα συχνοτήτων ή το διάγραμμα σχετικών συχνοτήτων.

Η κατανομή συχνοτήτων σε μη ομαδοποιημένα διακριτά δεδομένα μπορεί να παρασταθεί με ένα γράφημα κάθετων γραμμών, στο οποίο **το ύψος κάθε γραμμής αντιπροσωπεύει τη συχνότητα της τιμής της μεταβλητής.** Οι διακριτές γραμμές θέλουν να δηλώσουν τη φύση της διακριτής μεταβλητής.

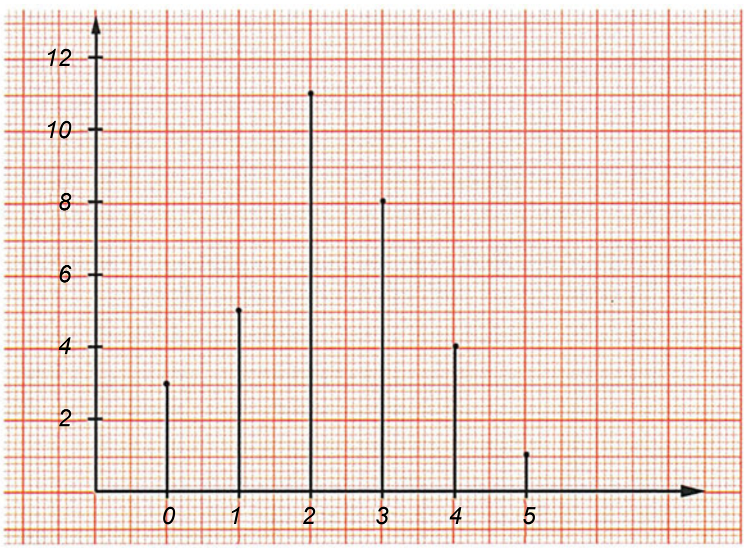
Η όλη εργασία φαίνεται καλύτερα στο ακόλουθο παράδειγμα:

Πίνακας 3.11

Αριθμός παιδιών κατά οικογένεια	0	1	2	3	4	5	
Συχνότητα (ν)	3	5	11	8	2	1	Σύνολο 30

Είναι φανερό ότι εδώ έχουμε μία συνάρτηση με ανεξάρτητη μεταβλητή X, που παίρνει τιμές $x_1, x_2, \ldots, x_\kappa$ και μία εξαρτημένη μεταβλητή ν, που παίρνει τιμές $\nu_1, \nu_2, \ldots, \nu_\kappa$ αντίστοιχα. Μπορούμε λοιπόν να πάρουμε ένα ορθογώνιο σύστημα αξόνων και στον οριζόντιο άξονα να πάρουμε τις τιμές x_i ενώ στον κατακόρυφο τις τιμές ν_i, οπότε έχουμε αμέσως τα σημεία Α, Β, Γ, …, Ν που είναι εικόνες των ζευγών (x_1, ν_1), (x_2, ν_2), … (x_κ, ν_κ).

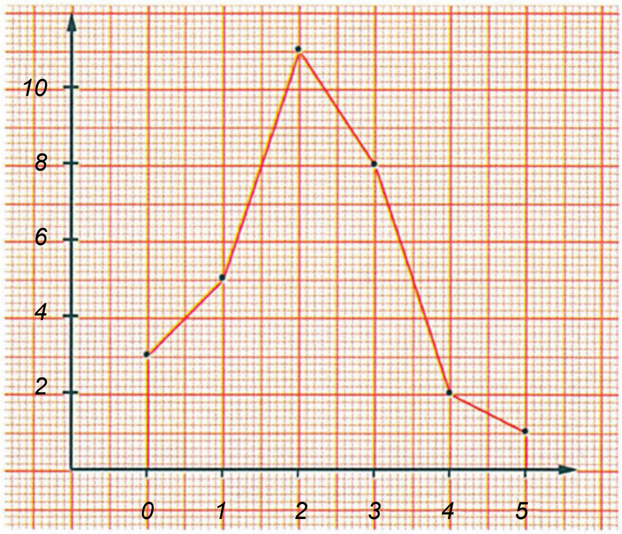
Το σύνολο των σημείων αυτών αποτελεί το **διάγραμμα συχνοτήτων**.



(Σχ. 3.5) Διάγραμμα Συχνοτήτων

Συνήθως ως διάγραμμα συχνοτήτων δεν παίρνουμε τα σημεία Α, Β, Γ, …, Ν, αλλά τα ευθύγραμμα τμήματα που παριστάνουν τις τεταγμένες αυτών των σημείων.

Ενώνοντας τα σημεία (x_i, v_i) ή (x_i, f_i) έχουμε το λεγόμενο **πολύγωνο συχνοτήτων ή πολύγωνο σχετικών συχνοτήτων** αντίστοιχα *(Σχ. 3.6)* που μας δίνει μία εικόνα για τη μεταβολή της συχνότητας σε σχέση με τη μεταβολή των τιμών της μεταβλητής που εξετάζουμε.

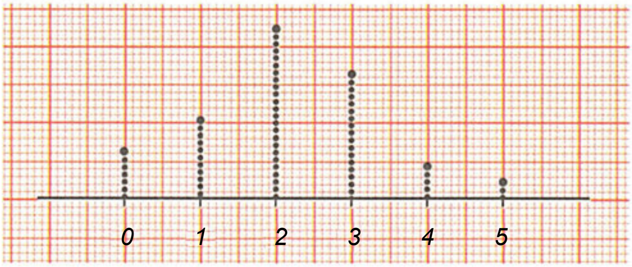


(Σχ.3.6) Πολύγωνο συχνοτήτων

- **Σημειόγραμμα** (dot diagram)

Η κατανομή του παραπάνω παραδείγματος θα μπορούσε επίσης να παρασταθεί με ένα σημειόγραμμα, στο οποίο οι τιμές της μεταβλητής απεικονίζονται πάνω σε έναν οριζόντιο άξονα και οι συχνότητες των αντίστοιχων τιμών παριστάνονται γραφικά σαν σημεία υπεράνω αυτού του άξονα.

Στο σημειόγραμμα που ακολουθεί βλέπουμε την κατανομή συχνοτήτων της μεταβλητής «αριθμός παιδιών» του παραδείγματος που ήδη αναφέραμε:



(Σχ.3.7) Σημειόγραμμα

- **Ιστόγραμμα Συχνοτήτων**

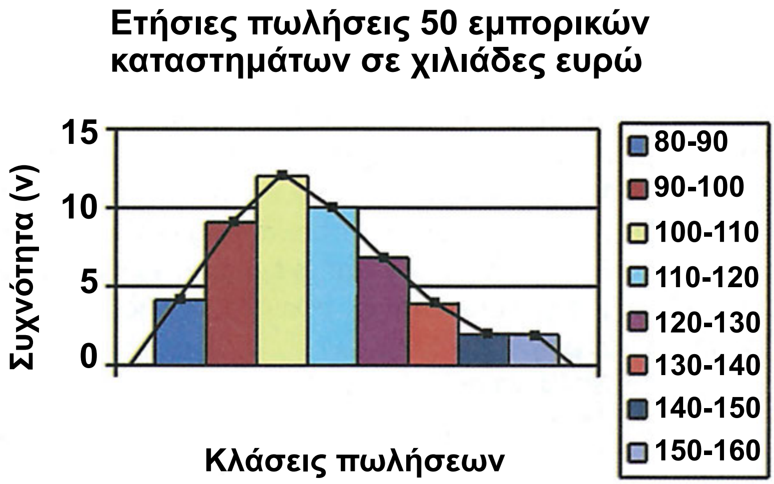
Για τη γραφική απεικόνιση ομαδοποιημένων δεδομένων χρησιμοποιούμε τα ιστογράμματα (histograms). Για την κατασκευή των ιστογραμμάτων χρησιμοποιούμε διαδοχικά ορθογώνια που έχουν βάσεις ίσες με το πλάτος c - το οποίο παίρνουμε ως μονάδα μέτρησης - των διαστημάτων των κλάσεων ($α_{i-1}$, $α_i$) και είναι τοποθετημένα επάνω στον οριζόντιο άξονα. Σε κάθε ιστόγραμμα το εμβαδόν του κάθε ορθογωνίου είναι ανάλογο της συχνότητας. Όταν οι κλάσεις είναι ίσου πλάτους, τότε το ύψος κάθε ορθογωνίου είναι ίσο προς τη συχνότητα της αντίστοιχης κλάσης, **έτσι ώστε το εμβαδόν των ορθογωνίων να είναι ίσο με τις αντίστοιχες συχνότητες,** αφού το πλάτος του ορθογωνίου λαμβάνεται ως μονάδα μέτρησης.

Ο πίνακας 3.12 δίνει τις ετήσιες πωλήσεις 50 ομοειδών εμπορικών καταστημάτων σε χιλιάδες ευρώ.

Πίνακας 3.12

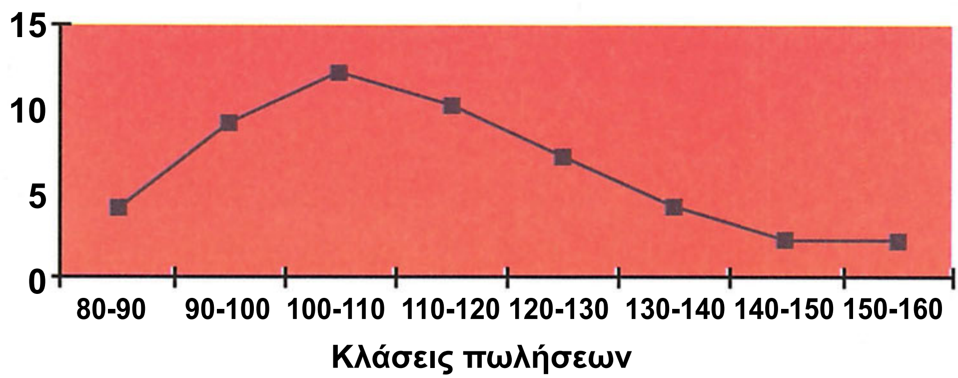
Κλάσεις πωλήσεων σε χιλιάδες ευρώ	**Συχνότητα (v_i)**	**Σχετική συχνότητα (f_i%)**	**Αθροιστική Συχνότητα (N_i)**
[80-90)	4	8	4
[90-100)	9	18	13
[100-110)	12	24	25
[110-120)	10	20	35
[120-130)	7	14	42
[130-140)	4	8	46
[140-150)	2	4	48
[150-160)	2	4	50
	Σv = 50	Σf =100	

Με βάση τον πίνακα αυτό θα κατασκευάσουμε το ακόλουθο ιστόγραμμα συχνοτήτων της μεταβλητής «ετήσιες πωλήσεις 50 εμπορικών καταστημάτων»:

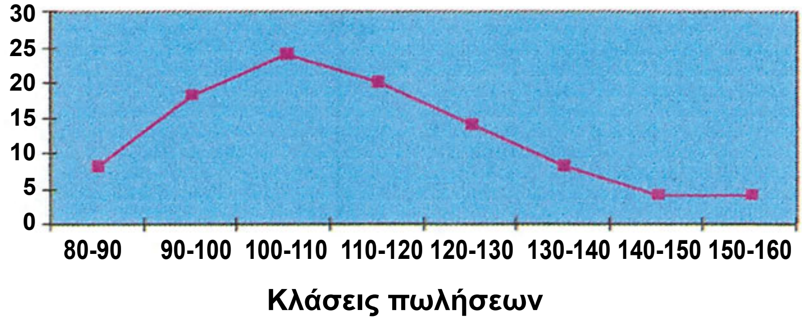


(Σχ. 3.8) Ιστόγραμμα συχνοτήτων.

Αν ενώσουμε τα μέσα των επάνω βάσεων των ορθογωνίων ενός ιστογράμματος, σχηματίζεται μια τεθλασμένη γραμμή, που λέγεται **πολύγωνο συχνοτήτων (frequency polygon)** (Σχ. 3.9, 3.10). Επίσης θα μπορούσαμε, αφού έχουμε χωρίσει τον οριζόντιο άξονα σε διαστήματα που αντιστοιχούν στις ίσου πλάτους κλάσεις, να ορίσουμε σημείο επάνω από το μέσο κάθε διαστήματος και σε ύψος ίσο με την αντίστοιχη συχνότητα κάθε κλάσης. Στη συνέχεια, ενώνοντας αυτά τα σημεία με ευθύγραμμα τμήματα σχηματίζουμε το πολύγωνο συχνοτήτων. Συνήθως προεκτείνουμε την πολυγωνική γραμμή από τα δύο άκρα της μέχρι να κλείσει με τον x-άξονα, οπότε σχηματίζεται ένα πολύγωνο.

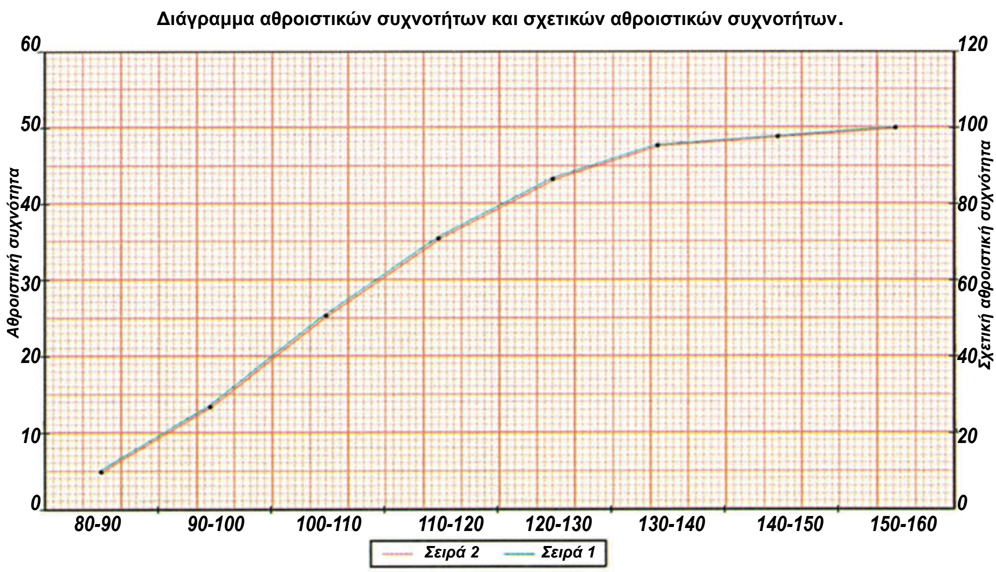


(Σχ. 3.9) Πολύγωνο συχνοτήτων.



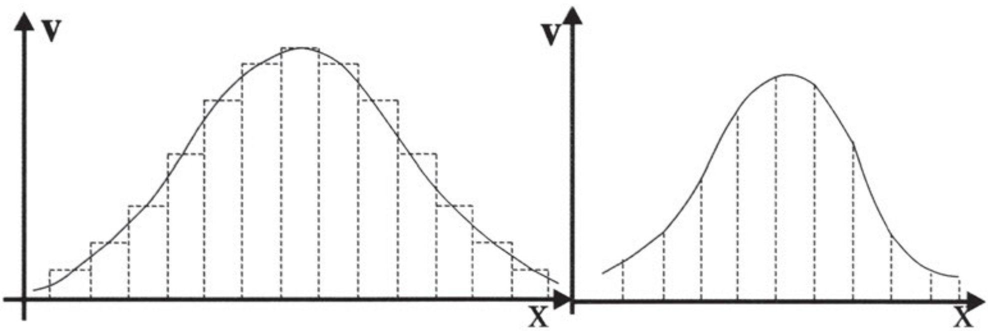
(Σχ. 3.10) Πολύγωνο σχετικών συχνοτήτων.

Με παρόμοιο τρόπο κατασκευάζουμε και τα ιστογράμματα αθροιστικών συχνοτήτων και σχετικών αθροιστικών συχνοτήτων (Σχ. 3.11). Ενώνουμε τα δεξιά άκρα των άνω βάσεων των ορθογωνίων και βρίσκουμε το διάγραμμα αθροιστικών συχνοτήτων της κατανομής.



(Σχ. 3.11) Διάγραμμα αθροιστικών συχνοτήτων και σχετικών αθροιστικών συχνοτήτων.

Όταν το πλάτος των κλάσεων είναι πολύ μικρό, τότε οι πλευρές του πολυγώνου συχνοτήτων έχουν μικρό μήκος. Στην περίπτωση αυτή το πολύγωνο συχνοτήτων μοιάζει με καμπύλη γραμμή. Είναι φανερό ότι **τα πολύγωνα συχνοτήτων των συνεχών μεταβλητών γίνονται καμπύλες γραμμές,** όσο αυξάνει ο αριθμός των κλάσεων, όπως φαίνεται στο σχήμα (3.12).



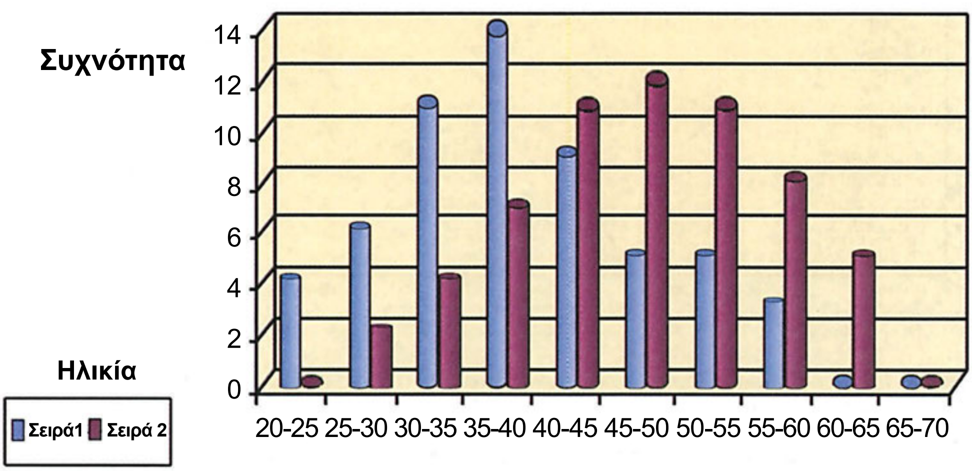
(Σχ.3.12) Πολύγωνο συχνοτήτων συνεχούς μεταβλητής.

Συγκριτικά πολύγωνα συχνοτήτων

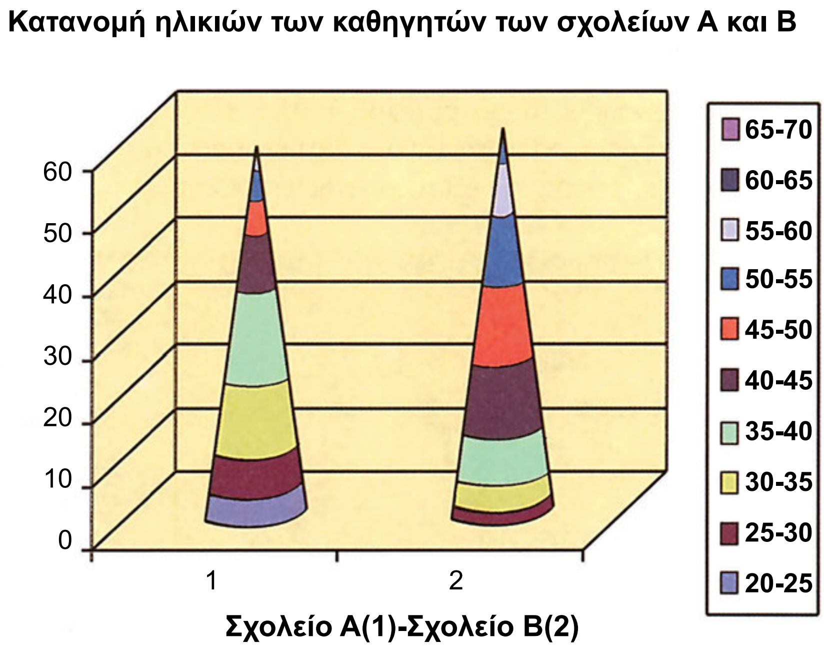
Τα πολύγωνα συχνοτήτων είναι πολύ χρήσιμα, όταν θέλουμε να συγκρίνουμε δύο ή περισσότερες ομάδες δεδομένων (Σχ. 3.15).

Στο παρακάτω διάγραμμα (Σχ. 3.13) παρουσιάζονται δύο διαγράμματα συχνοτήτων που συγκρίνουν την κατανομή ηλικιών των καθηγητών στο σχολείο Α και στο σχολείο Β.

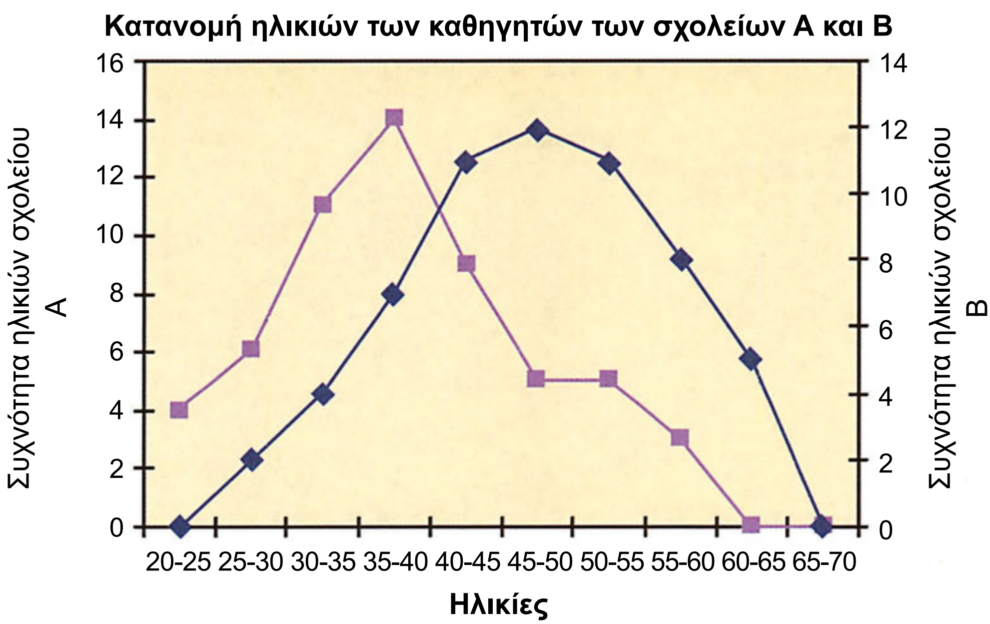
Διαγράμματα που χρησιμοποιούνται για να συγκρίνουμε δύο ή περισσότερες ομάδες δεδομένων είναι επίσης τα κωνοειδή και τα κυλινδρικά (Σχ. 3.14)



(Σχ. 3.13) Συγκριτικό διάγραμμα συχνοτήτων με κυλίνδρους.



(Σχ. 3.14) Συγκριτικό διάγραμμα συχνοτήτων με κώνους.



(Σχ. 3.15) Συγκριτικά πολύγωνα συχνοτήτων.

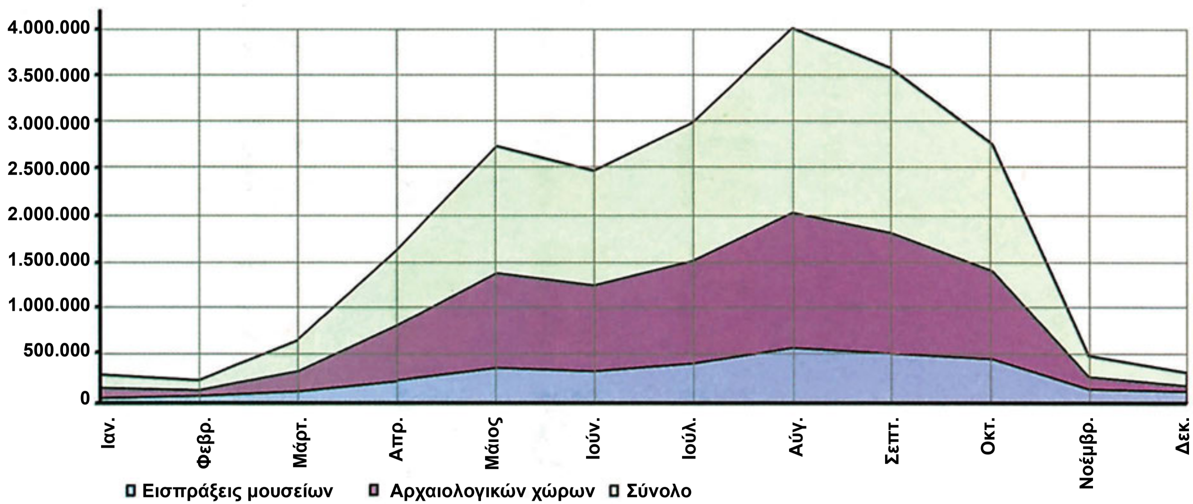
- **Χρονολογικά διαγράμματα ή χρονοδιαγράμματα** (time charts)

Χρησιμοποιούνται για τη γραφική απεικόνιση της διαχρονικής εξέλιξης ενός οικονομικού μεγέθους, δημογραφικού ή άλλου είδους (Σχ. 3.16, 3.17). Ο οριζόντιος άξονας χρησιμοποιείται ως άξονας μέτρησης του χρόνου και ο κάθετος ως άξονας μέτρησης της εξεταζόμενης μεταβλητής.

(Σχ. 3.16) Συγκριτικό χρονολογικό διάγραμμα. ***Πηγή:*** *Κυριακάτικη Ελευθεροτυπία*

Εισπράξεις μουσείων και αρχαιολογικών χώρων: 1997

Εισπράξεις	Ιαν.	Φεβρ.	Μάρτ.	Απρ.	Μάιος	Ιούν.	Ιούλ.	Αύγ.	Σεπτ.	Οκτ.	Νοέμβρ.	Δεκ.	Σύνολο
Μουσείων	34.322	38.238	83.253	207.077	336.500	301.837	379.700	540.525	479.847	412.029	104.428	58.208	2.975.964
Αρχαιολογικών χώρων	103.928	68.658	231.658	589.783	1.013.100	920.159	1.095.087	1.447.670	1.288.820	958.725	125.459	73.674	7.916.721



(Σχ.3.17) Χρονολογικό διάγραμμα των εισπράξεων διαφόρων μουσείων και αρχαιολογικών χώρων. **Πηγή:** *Ε.Σ.Υ.Ε*

- **Ειδογράμματα** (pickograms)

Αυτού του είδους τα διαγράμματα αποτελούνται από εικόνες ομοειδών προσώπων ή πραγμάτων και παρουσιάζουν συνήθως τη διαχρονική εξέλιξη του μεγέθους που απεικονίζεται. Χρησιμοποιούνται κυρίως στη διαφήμιση, αλλά και για συγκρίσεις μεταξύ δύο ή περισσοτέρων μεγεθών.

Εξαγωγές πετρελαίου μιας πετρελαιοπαραγωγού χώρας κατά τα έτη 1995-1998

Εξαγωγή ποσότητας πετρελαίου **2.000.000** τόνων

1995

1996

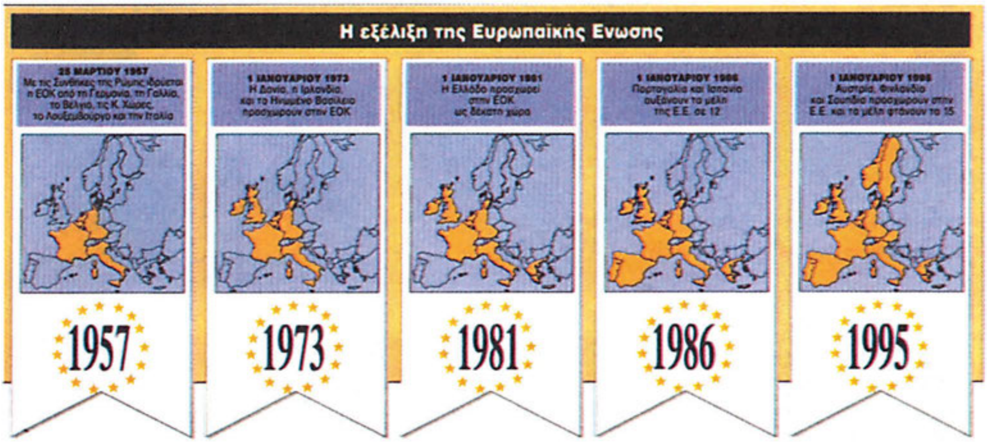
1997

1998

(Σχ.3.18) Ειδόγραμμα με βαρέλια πετρελαίου (Υποθετικά δεδομένα)

- **Χαρτοδιαγράμματα** (map charts)

Στα διαγράμματα αυτά η γραφική παράσταση των δεδομένων γίνεται πάνω σε γεωγραφικούς χάρτες και τα στατιστικά δεδομένα απεικονίζονται στις αντίστοιχες γεωγραφικές περιοχές με διάφορα χρώματα ή διαγραμμίσεις που δείχνουν το μέγεθος της εξεταζόμενης μεταβλητής. Στο υπόμνημα εξηγείται η αντιστοιχία χρωμάτων (ή διαγραμμίσεων) και τιμών της μεταβλητής (Σχ. 3.19).



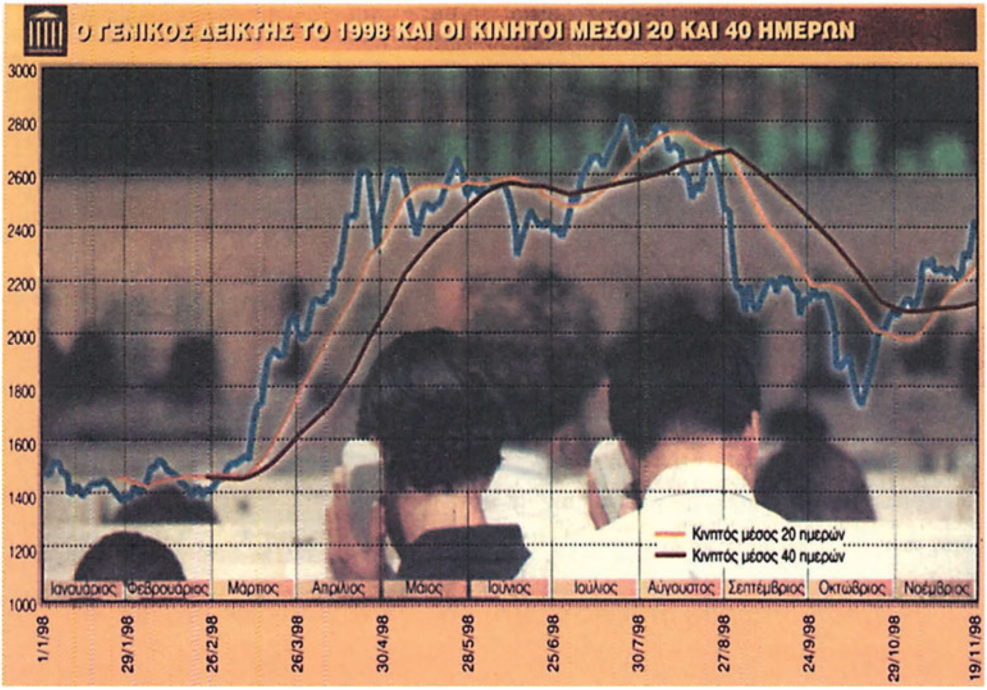
(Σχ. 3.19) Χαρτοδιάγραμμα με τις χώρες-μέλη της Ευρωπαϊκής Ένωσης.

3.4 Η χρησιμοποίηση από τις επιχειρήσεις στατιστικών πινάκων και διαγραμμάτων

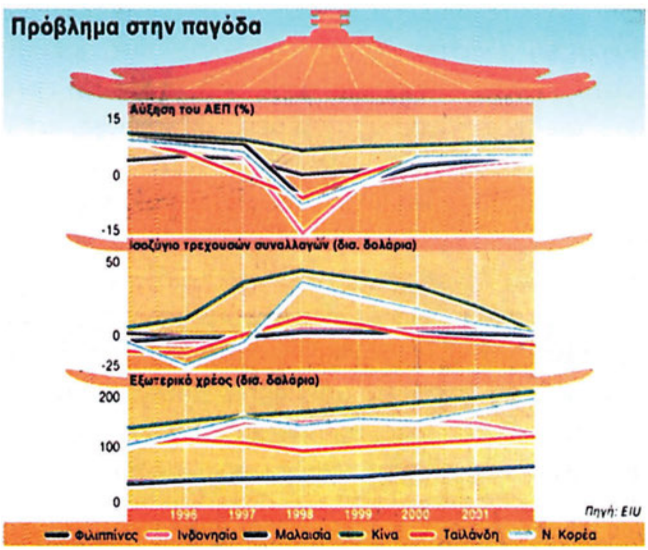
Σήμερα γίνεται ευρύτατη χρήση των πινάκων και των διαγραμμάτων από τις ιδιωτικές και δημόσιες επιχειρήσεις και από τους οργανισμούς, με κυριότερο σκοπό τη δημιουργία μιας θετικής εικόνας για την επιχείρηση. Οι επιχειρήσεις που λειτουργούν με επιτυχία προβάλλουν μέσω των πινάκων και των διαγραμμάτων κυρίως: α) Τη θετική οικονομική τους κατάσταση, β) τον ικανοποιητικό βαθμό επίτευξης των στόχων που είχαν θέσει στο παρελθόν και γ) τις ευνοϊκές προϋποθέσεις για μια επιτυχή και ασφαλή πορεία της επιχείρησης.

Η χρήση πινάκων και διαγραμμάτων από τις επιχειρήσεις γίνεται και για εσωτερικούς, ερευνητικούς και επιστημονικούς λόγους που αφορούν την ίδια την επιχείρηση. Με τους πίνακες και τα διαγράμματα γίνεται πιο εύκολη και γρήγορη η εσωτερική ενημέρωση των στελεχών, των εργαζομένων της επιχείρησης και των πολύ στενών συνεργατών της, ιδιαίτερα στις τακτικές και έκτακτες ενημερωτικές συναντήσεις (meetings) που γίνονται.

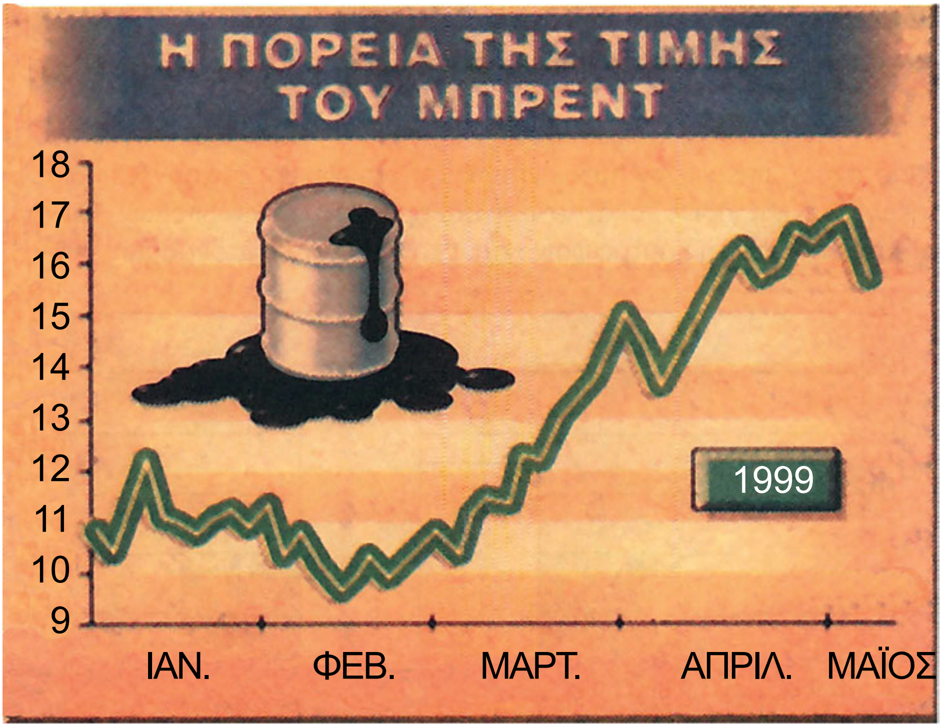
Ειδικότερα για τα διαγράμματα, θα πρέπει να γνωρίζουμε ότι λόγω των τεράστιων δυνατοτήτων των ηλεκτρονικών υπολογιστών μπορούμε να κατασκευάζουμε εξαιρετικής ακρίβειας διαγράμματα με συνθέσεις χρωμάτων και πρωτοποριακά σχήματα που ελκύουν την προσοχή του κοινού. Μερικά από αυτά που διακρίνονται για την πρωτοτυπία τους παραθέτουμε στις αμέσως επόμενες σελίδες:



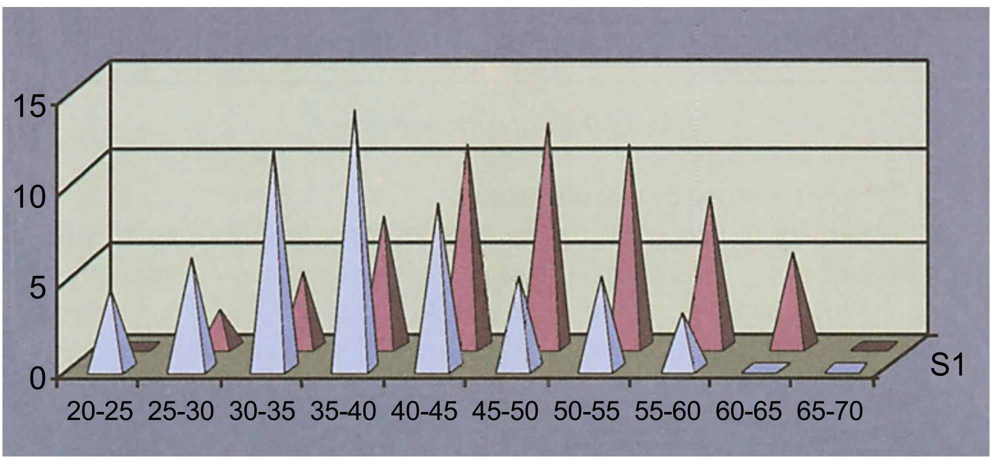
(Σχ. 3.20) Χρονολογικό διάγραμμα ***Πηγή:*** *Κυριακάτικη Ελευθεροτυπία 1999*



(Σχ. 3.21) Συγκριτικό διάγραμμα σε σχήμα παγόδας. ***Πηγή:*** *FIU*



(Σχ. 3.22) Χρονολογικό διάγραμμα. ***Πηγή:*** *Κυριακάτικη Ελευθεροτυπία 1999*

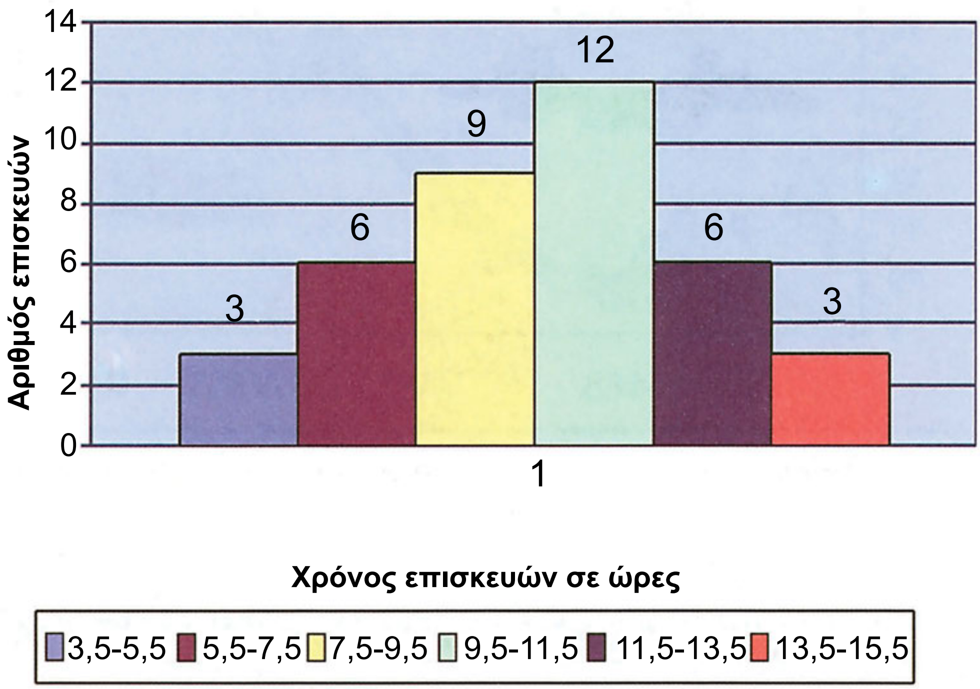


(Σχ. 3.23) Πυραμίδες. Συγκριτικό διάγραμμα της κατανομής των ηλικιών των εργαζομένων δύο επιχειρήσεων.

ΕΦΑΡΜΟΓΕΣ

- **Εφαρμογή 1η**

Σε μία σχολή μηχανικών αυτοκινήτων καταγράφηκαν οι χρόνοι που χρειάστηκαν για την πλήρη αποκατάσταση σοβαρών μηχανικών βλαβών. Τα αποτελέσματα της μελέτης αυτής φαίνονται στο παρακάτω ιστόγραμμα:



(Σχ. 3.24) Ιστόγραμμα συχνοτήτων.

α) Πόσες επισκευές έγιναν συνολικά;

β) Πόσες επισκευές χρειάστηκαν περισσότερες από 11,5 ώρες για την ολοκλήρωσή τους; Τι ποσοστό του συνόλου των επισκευών αποτελούν οι επισκευές που έγιναν σε περισσότερο από 11,5 ώρες;

γ) Τι ποσοστό του συνόλου των επισκευών είχε διάρκεια από 7,5 ως 11,5 ώρες;

Λύση

Παρατηρώντας το ιστόγραμμα μπορούμε να κατασκευάσουμε έναν πίνακα συχνοτήτων και αθροιστικών συχνοτήτων:

Πίνακας 3.13

α/α	Χρόνος επισκευών (ώρες)	Συχνότητα v_i	Σχετική συχνότητα f_i	Αθροιστική συχνότητα N_i	Σχετική αθροιστική συχνότητα $F_i\%$
1	[3,5 - 5,5)	3	0,077	3	7,7
2	[5,5 - 7,5)	6	0,154	3 + 6 = 9	7,7 + 15,4 = 23,1
3	[7,5 - 9,5)	9	0,230	$\underbrace{(3+6)}_{9}+9=18$	23,1 + 23 = 46,1
4	[9,5 -11,5)	12	0,308	$\underbrace{(3+6+9)}_{18}+12=30$	46,1 + 30,8 = 76,9
5	[11,5 - 13,5)	6	0,154	$\underbrace{(3+6+9+12)}_{30}+6=36$	76,9 + 15,4 = 92,3
6	[13,5 - 15,5)	3	0,077	$\underbrace{(3+6+9+12+6)}_{36}+3=39$	92,3 + 7,7 = 100
	Σύνολο	39	1		

α) Συνολικά επισκευάστηκαν 39 μηχανικές βλάβες.

β) Από την αθροιστική συχνότητα παρατηρούμε ότι η διάρκεια επισκευής των μηχανικών βλαβών, που ήταν μεγαλύτερη από 11,5 ώρες, βρίσκεται στην 5η και 6η κλάση. Άρα σύμφωνα με την αθροιστική συχνότητα ο αριθμός των βλαβών που επισκευάστηκαν βρίσκεται ως εξής: $N_6 – N_4 = 39 – 30 = 9$ μηχανικές βλάβες.

Αντίστοιχα από τη σχετική αθροιστική συχνότητα θα έχουμε:
$F_6 – F_4 = 100 – 76,9 = 23,1$. Δηλαδή 23,1% του συνόλου των μηχανικών βλαβών.

γ) Η διάρκεια επισκευών από 7,5 ως 11,5 ώρες αντιστοιχεί στην 3η και 4η κλάση. Άρα το ποσοστό των μηχανικών βλαβών που επισκευάστηκαν θα είναι: 23% + 30,8% = 53,8%

Θα μπορούσαμε επίσης να το υπολογίσουμε χρησιμοποιώντας τη σχετική αθροιστική συχνότητα ως εξής: $F_4 - F_2 = 76{,}9 - 23{,}1 = 53{,}8\%$ ή

$$\frac{9+12}{39} = \frac{21}{39} = 0{,}538 \text{ ή } 53{,}8\%$$

- **Εφαρμογή 2η**

Η κατανομή συχνοτήτων που δίνεται στον πίνακα 3.14 αναφέρεται στα ύψη 50 ανδρών – μετρημένα σε εκατοστά και στρογγυλοποιημένα στην πλησιέστερη δεκάδα.

Πίνακας 3.14

Ύψος (εκ.)	Συχνότητα
140	1
150	6
160	8
170	21
180	10
190	4

α) Ποιο είναι το μικρότερο δυνατό ύψος που θα μπορούσε να έχει το άτομο που καταγράφηκε με ύψος 140 εκ.;

β) Με τα δεδομένα του παραπάνω πίνακα συχνοτήτων να κατασκευάσετε έναν πίνακα συχνοτήτων ομαδοποιημένων δεδομένων με κλάσεις ίσου πλάτους και στη συνέχεια να σχηματίσετε το αντίστοιχο κυκλικό διάγραμμα.

γ) Να κατασκευάσετε ένα ιστόγραμμα συχνοτήτων και το αντίστοιχο πολύγωνο συχνοτήτων, το οποίο να απεικονίζει τα δεδομένα του πίνακα. Εξηγήστε την κατασκευή του ιστογράμματος.

δ) Να κατασκευάσετε έναν πίνακα αθροιστικών και σχετικών αθροιστικών συχνοτήτων και τα αντίστοιχα διαγράμματα.

i) Υπολογίστε από το διάγραμμα πόσα άτομα έχουν ύψος μικρότερο από 170 εκ.

ii) Υπολογίστε πόσα άτομα έχουν ύψος από 165 εκ. έως 185 εκ.

Λύση

α) Επειδή οι μετρήσεις έγιναν με προσέγγιση δεκάδας το άτομο που καταγράφηκε με ύψος 140 εκ. θα μπορούσε να έχει πραγματικό ύψος που κυμαίνεται από 135-144 εκ.

β) Κατασκευάζουμε τον πίνακα συχνοτήτων και σχετικών συχνοτήτων:

Πίνακας 3.15

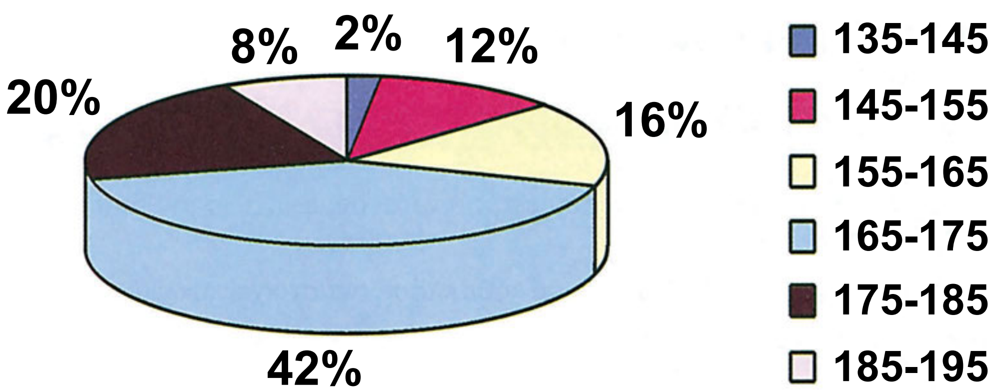
Ύψος (εκ.)	Συχνότητα	Σχετική συχνότητα f_i%
135-145	1	2
145-155	6	12
155-165	8	16
165-175	21	42
175-185	10	20
185-195	4	8
	50	100

Η κατασκευή κυκλικού διαγράμματος γίνεται σύμφωνα με τον τύπο:

$\frac{v_i}{v} \cdot 360° = f_i\% \cdot 360°$ δηλαδή, σε κάθε κλάση αντιστοιχεί κυκλικός τομέας αντίστοιχου τόξου, όπως αναγράφεται στον επόμενο πίνακα:

Πίνακας 3.16

Ύψος (εκ.) [,)		Τόξο κυκλικού τομέα
135-145	⟶	$2\% \cdot 360^\circ = 7{,}2^\circ$
145-155	⟶	$12\% \cdot 360^\circ = 43{,}2^\circ$
155-165	⟶	$16\% \cdot 360^\circ = 57{,}6^\circ$
165-175		$42\% \cdot 360^\circ = 151{,}2^\circ$
175-185		$20\% \cdot 360^\circ = 72^\circ$
185-195		$8\% \cdot 360^\circ = 28{,}8^\circ$



(Σχ.3.25) Κυκλικό διάγραμμα σχετικών συχνοτήτων.

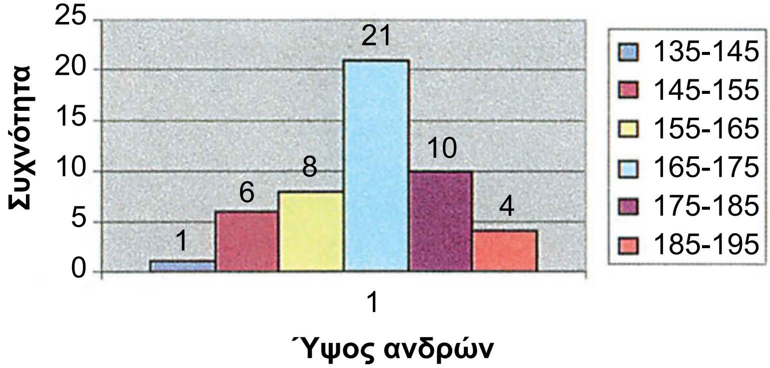
γ) Συμπληρώνουμε τον πίνακα 3.17 με τις αθροιστικές και σχετικές αθροιστικές συχνότητες και προκύπτει ο παρακάτω πίνακας:

Πίνακας 3.17

Ύψος (εκ.)	Κέντρο κλάσης x_i	Συχνότητα v_i	Αθροιστική συχνότητα N_i	Σχετική συχνότητα f_i%	Σχετική αθροιστική συχνότητα F_i%
135-145	140	1	1	2	2
145-155	150	6	7	12	14
155-165	160	8	15	16	30
165-175	170	21	36	42	72
175-185	180	10	46	20	92
185-195	190	4	50	8	100
Σύνολο		50		100	

Για να κατασκευάσουμε το αντίστοιχο ιστόγραμμα βρίσκουμε κατ' αρχήν το εύρος μεταβολής των τιμών της μεταβλητής δηλαδή, R = 135 – 195 = 60εκ. και το χωρίζουμε σε έξι ίσα διαστήματα 60 : 6 = 10. Άρα η κάθε κλάση θα έχει πλάτος 10, δηλαδή θα απεικονίζει μεταβολή ύψους 10εκ. Στη συνέχεια θα κατασκευάσουμε ορθογώνια με ίσες βάσεις –οι οποίες απεικονίζουν τις κλάσεις της μεταβλητής «ύψος»– και με ύψος που θα αντιστοιχεί στη συχνότητα κάθε κλάσης.

Αν ενώσουμε τα μέσα των επάνω βάσεων αυτών των ορθογωνίων θα προκύψει μια τεθλασμένη γραμμή, που ονομάζεται πολύγωνο συχνοτήτων, και όταν προεκταθεί έτσι, ώστε να διέρχεται και από το μέσο της πρώτης κάθετης πλευράς του πρώτου ορθογωνίου καθώς επίσης και από το μέσο της τελευταίας κάθετης πλευράς του τελευταίου ορθογωνίου, σχηματίζει με τον άξονα x΄x ένα πολύγωνο όπως φαίνεται στα παρακάτω σχήματα 3.27 και 3.28.

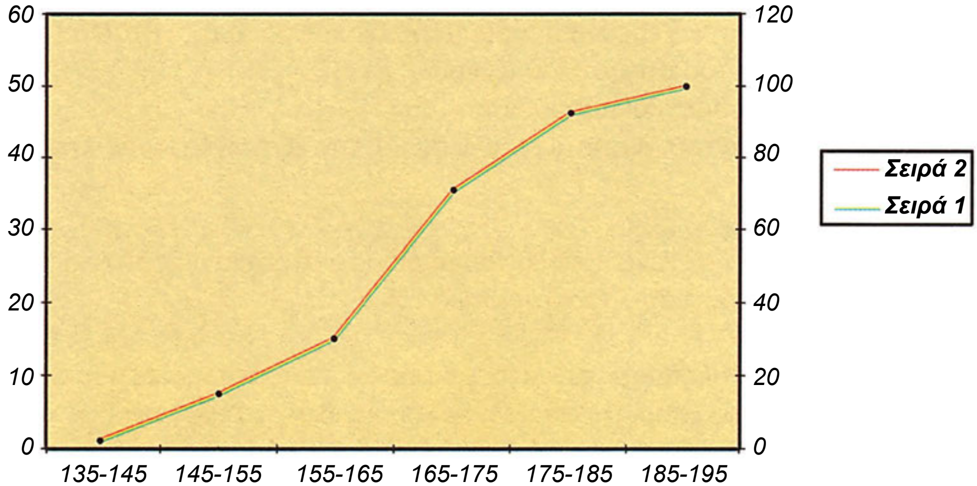


(Σχ. 3.26) Ιστόγραμμα συχνοτήτων.

(Σχ. 3.27) Πολύγωνο συχνοτήτων.

δ) Ο πίνακας 3.17 έχει συμπληρωθεί με τις αθροιστικές συχνότητες και με τις σχετικές αθροιστικές συχνότητες. **i)** Μπορούμε κατ' αρχήν να παρατηρήσουμε από το διάγραμμα ότι **36 άτομα έχουν ύψος μικρότερο από 170 εκ.** **ii)** Από τον πίνακα συχνοτήτων μπορούμε να υπολογίσουμε ότι ύψος που κυμαίνεται από 165-185 εκ. έχουν $\mathbf{\nu_4 + \nu_5 = 21 + 10 = 31}$ **άτομα ή αλλιώς**

$\mathbf{N_5 - N_3 = 46 - 15 = 31}$ **άτομα.** Έτσι έχουμε το παρακάτω διάγραμμα:



(Σχ. 3.28) Διάγραμμα αθροιστικών συχνοτήτων και σχετικών αθροιστικών συχνοτήτων.

ΑΝΑΚΕΦΑΛΑΙΩΣΗ

1. **Στατιστική** είναι η επιστήμη που ασχολείται με τη συλλογή, επεξεργασία, παρουσίαση και ανάλυση στοιχείων, με σκοπό την εξαγωγή συμπερασμάτων για τη λήψη ορθών αποφάσεων.

2. Σε μια στατιστική έρευνα μπορούμε να διακρίνουμε τρία στάδια:
α) Τη συλλογή του στατιστικού υλικού.
β) Την επεξεργασία και παρουσίασή του.
γ) Την ανάλυση του στατιστικού υλικού και την εξαγωγή συμπερασμάτων.

3. **Πληθυσμός** είναι ένα σύνολο του οποίου τα στοιχεία εξετάζονται ως προς ένα ή περισσότερα χαρακτηριστικά τους.

4. **Στατιστικός πληθυσμός** είναι οι μετρήσεις ή παρατηρήσεις που αναφέρονται σε κάποιο χαρακτηριστικό ή σε κάποια ιδιότητα των μονάδων του πληθυσμού.

5. **Δείγμα** είναι ένα τμήμα του πληθυσμού.

6. Τα χαρακτηριστικά ή οι ιδιότητες των στατιστικών μονάδων ενός πληθυσμού τα ονομάζουμε **μεταβλητές** και τις συμβολίζουμε με τα κεφαλαία γράμματα X, Y, Z, …. Οι αριθμητικές ή άλλες συμβολικές εκφράσεις τους λέγονται τιμές της μεταβλητής και συμβολίζονται με τα αντίστοιχα μικρά γράμματα.

7. Οι μεταβλητές διακρίνονται σε **ποιοτικές και ποσοτικές μεταβλητές.** Οι ποσοτικές διακρίνονται σε συνεχείς και διακριτές (ασυνεχείς) μεταβλητές.

8. Οι σπουδαιότερες μέθοδοι συλλογής στατιστικών στοιχείων είναι:
α) **Η απογραφή**, με την οποία γίνεται καταγραφή όλων των στοιχείων του πληθυσμού
β) **Η δειγματοληψία**, με την οποία γίνεται συλλογή δεδομένων μόνο από ένα μέρος του πληθυσμού, το δείγμα.
γ) **Η συνεχής εγγραφή των ατόμων ενός πληθυσμού**, με την οποία γίνεται καταχώρηση πληροφοριών για τα άτομα σε ειδικά βιβλία.

9. Η παρουσίαση των στατιστικών στοιχείων μπορεί να γίνει με τρεις τρόπους:
α) με συνοπτικές εκθέσεις ή αναφορές
β) με στατιστικούς πίνακες και
γ) με διαγράμματα

10. Σε κάθε πίνακα διακρίνουμε: το κύριο σώμα ή κορμό του πίνακα, τον τίτλο, τις επικεφαλίδες των στηλών και γραμμών, τις υποσημειώσεις και την πηγή.

11. Όταν οι μονάδες του πληθυσμού ερευνώνται ως προς ένα χαρακτηριστικό τους, χρησιμοποιούμε τους πίνακες **απλής εισόδου** και όταν ερευνώνται συγχρόνως ως προς δύο χαρακτηριστικά τους, χρησιμοποιούμε τους πίνακες **διπλής εισόδου**.

12. **Συχνότητα** μιας τιμής x_i της μεταβλητής Χ λέγεται ο φυσικός αριθμός ν_i που φανερώνει πόσες φορές παρουσιάζεται στο δείγμα (ή σε ολόκληρο τον πληθυσμό, αν αναφερόμαστε σ’ αυτόν) η τιμή αυτή.

13. **Τα στατιστικά διαγράμματα** αποτελούν έναν τρόπο παρουσίασης των στατιστικών δεδομένων με τη χρησιμοποίηση γεωμετρικών σχημάτων και συμβόλων. Σε ένα στατιστικό διάγραμμα μπορούμε να διακρίνουμε: τον τίτλο, την κλίμακα με τις τιμές μεγεθών, το υπόμνημα και την πηγή.

14. Μερικά από τα κυριότερα είδη διαγραμμάτων είναι: τα ραβδογράμματα, τα κυκλικά διαγράμματα, τα διαγράμματα συχνοτήτων, τα σημειογράμματα, τα ιστογράμματα συχνοτήτων, τα χρονολογικά διαγράμματα ή χρονοδιαγράμματα, τα ειδογράμματα, τα χαρτοδιαγράμματα.

ΕΡΩΤΗΣΕΙΣ ΚΑΤΑΝΟΗΣΗΣ

1. Οι σπουδαιότερες μέθοδοι συλλογής στατιστικών στοιχείων είναι:

α) .... β) .... γ) ....

Να αναφέρετε περιληπτικά τι γνωρίζετε για την κάθε μέθοδο.

2. Η παρουσίαση των στατιστικών στοιχείων μπορεί να γίνει με τους εξής τρόπους:

α) .... β) .... γ) ....

3. Τι επιτυγχάνουμε με τη βοήθεια των στατιστικών πινάκων;

4. Τα κυριότερα στοιχεία που μπορούμε να διακρίνουμε σε έναν πίνακα είναι:

α) .... β) .... γ) ....

δ) .... ε) ....

5. Όταν μελετάμε τις μονάδες του πληθυσμού ως προς δύο χαρακτηριστικά συγχρόνως, χρησιμοποιούμε τους πίνακες..........................εισόδου, ενώ όταν μελετάμε τις μονάδες του πληθυσμού ως προς ένα χαρακτηριστικό, χρησιμοποιούμε τους πίνακες ....εισόδου.

6. Συχνότητα μιας τιμής x_i της μεταβλητής Χ λέγεται ………………………………...

……………………………………………………………………………………………

7. Αθροιστική συχνότητα N_i μιας τιμής x_i λέγεται …………………………………

……………………………………………………………………………………………

και σχετική αθροιστική συχνότητα F_i μιας τιμής x_i λέγεται ………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

8. Η σχετική αθροιστική συχνότητα ισούται: $F_i = f_1 + f_2 + \ldots + f_i$, για $i = 1, 2, \ldots, \kappa$.

Σ ☐ Λ ☐

και η επί τοις εκατό σχετική αθροιστική συχνότητα $F_i\% =$ …………………

....

9. α) Πώς καταφεύγουμε σε ομαδοποίηση των παρατηρήσεων;

β) Εύρος του δείγματος ονομάζεται ……………………………………………

....

10. Σε ένα στατιστικό πίνακα μπορούμε να διακρίνουμε:

α) …………………………… β) …………………………………

γ) …………………………… δ) …………………………………

11. Να αναφέρετε μερικά από τα κυριότερα είδη διαγραμμάτων:

α) …………………………… β) …………………………………

γ) …………………………… δ) …………………………………

ΑΣΚΗΣΕΙΣ

1. Δίνεται η ποσοστιαία σύνθεση (%) του προσωπικού μιας επιχείρησης ως προς το μορφωτικό τους επίπεδο.

Επίπεδο μόρφωσης	**Ποσοστό (%)**
Πτυχιούχοι Τριτ/θμιας Εκπαίδευσης	25
Απόφοιτοι Λυκείου	50
Απόφοιτοι Γυμνασίου	25

i) Αν ο αριθμός των υπαλλήλων της επιχείρησης είναι 40 άτομα, κατασκευάστε τον πίνακα συχνοτήτων και σχετικών συχνοτήτων.
ii) Να απεικονίσετε τα παραπάνω δεδομένα με ραβδόγραμμα και με κυκλικό διάγραμμα.

2. Οι βαθμοί των 32 μαθητών της Α΄ τάξης των Τ.Ε.Ε στην Ιστορία είναι: 15, 9, 18, 20, 12, 18, 12, 11, 10, 14, 16, 10, 18, 20, 19, 16, 14, 15, 13, 15, 8, 17, 19, 11, 13, 15, 14, 11, 15, 16, 12, 17.
i) Να κατασκευάσετε πίνακα συχνοτήτων και σχετικών συχνοτήτων.
ii) Να κατασκευάσετε πίνακα αθροιστικών συχνοτήτων και σχετικών αθροιστικών συχνοτήτων.
iii) Πόσοι μαθητές έχουν βαθμό μικρότερο ή ίσο του 14 και πόσοι μαθητές έχουν βαθμό μεγαλύτερο του 15;

3. Για τα δεδομένα της άσκησης 2 να κατασκευάσετε το διάγραμμα συχνοτήτων, [το πολύγωνο συχνοτήτων, καθώς και το διάγραμμα αθροιστικών συχνοτήτων και σχετικών αθροιστικών συχνοτήτων].
Από το διάγραμμα ή με κατάλληλους υπολογισμούς να υπολογίσετε:
i) Το ποσοστό των μαθητών που απέτυχαν (βαθμός<10).
ii) Το ποσοστό των μαθητών που οι βαθμοί τους κυμαίνονται από 13 ως και 16.
iii) Το ποσοστό των μαθητών που αρίστευσαν, δηλαδή που ο βαθμός τους είναι μεγαλύτερος του 18. Τι παρατηρείτε για τη βαθμολογία της συγκεκριμένης τάξης;

4. Οι 50 εργάτες ενός εργοστασίου έχουν τις παρακάτω ηλικίες:

21	43	50	25	55	30	28	40	31	51
18	47	52	34	47	32	27	41	35	54
30	48	36	43	38	33	27	39	41	43
32	22	46	52	29	32	34	34	42	36
35	28	57	56	20	38	35	27	40	35

i) Να ομαδοποιήσετε τις ηλικίες σε 8 κλάσεις

ii) Να κατασκευάσετε πίνακα συχνοτήτων, αθροιστικών συχνοτήτων, σχετικών συχνοτήτων και σχετικών αθροιστικών συχνοτήτων.

iii) Να κατασκευάσετε το ιστόγραμμα συχνοτήτων και το πολύγωνο συχνοτήτων.

5. Ο παρακάτω πίνακας δείχνει τον αριθμό των τροχαίων ατυχημάτων κατά τα έτη [1988 – 1997:]

Έτος	Σύνολο ατυχημάτων
1988	20.753
1989	20.299
1990	19.609
1991	20.764
1992	22.006
1993	22.165
1994	22.222
1995	22.798
1996*	23.606
1997*	24.319

Πηγή: *Ε.Σ.Υ.Ε*

Να παρασταθεί ο αριθμός των οδικών ατυχημάτων κατά τα έτη 1988 – 1997 με χρονολογικό διάγραμμα και να απεικονισθεί η αυξητική εξέλιξη του αριθμού αυτού με χρονολογικό διάγραμμα.

6. Ο παρακάτω πίνακας δείχνει το σύνολο των εξαγωγών μιας επιχείρησης σε χιλιάδες ευρώ κατά τα έτη 1996-1997.

Έτος	Αφρική	Αμερική	Ασία	Ευρώπη
1996	80,4	120,2	150,6	230,8
1997	50,5	60,7	130,2	190,6

i) Κατασκευάστε δύο κυκλικά διαγράμματα που να απεικονίζουν τις εξαγωγές της επιχείρησης κατά τα έτη 1996 και 1997. Με τη βοήθεια των κυκλικών διαγραμμάτων να συγκρίνετε τις ετήσιες εξαγωγές της επιχείρησης.

ii) Πόσο τοις εκατό αυξήθηκαν ή μειώθηκαν οι εξαγωγές της επιχείρησης τα έτη 1996-1997 σε κάθε ήπειρο;

7. Η στατιστική υπηρεσία της τροχαίας θέλοντας να μελετήσει τους μήνες όπου εμφανίζεται αύξηση των σοβαρών τραυματισμών σε οδικά ατυχήματα, καθώς και τη μεταβολή του αριθμού των σοβαρά τραυματισμένων κατά τα έτη 1996-1997, κατασκεύασε τον παρακάτω πίνακα:

Μήνες	Σοβαρά ατυχήματα	
	1997 ν_i	1996 ν_i^*
Ιανουάριος	238	190
Φεβρουάριος	201	211
Μάρτιος	278	189
Απρίλιος	326	249
Μάιος	330	294
Ιούνιος	454	319
Ιούλιος	538	344
Αύγουστος	553	433
Σεπτέμβριος	383	317
Οκτώβριος	372	284
Νοέμβριος	319	231
Δεκέμβριος	295	215
Σύνολο	4.287	3.276

i) Συμπληρώστε τον πίνακα συχνοτήτων με μία στήλη, όπου να φαίνεται η μεταβολή (Δ) του αριθμού των βαριά τραυματιών κατά τα έτη 1996, 1997 (δηλαδή Δ = αριθμός τραυματιών 1997 – αριθμός τραυματιών 1996) στους αντίστοιχους μήνες ενός έτους.

ii) Υπολογίστε το ποσοστό της μεταβολής αυτής ως προς τον αριθμό των τραυματιών του 1996, δηλαδή $\left(\frac{\Delta}{\nu_i^*}\cdot 100\right)$ [με προσέγγιση δέκατου και συμπληρώστε τον πίνακα].

iii) Κατασκευάστε συγκριτικά μηνιαία ραβδογράμματα για τον αριθμό των σοβαρά τραυματισμένων σε τροχαία ατυχήματα κατά τα έτη 1996-1997.

iv) Κατασκευάστε ένα χρονολογικό διάγραμμα της εξέλιξης της ποσοστιαίας μεταβολής του ερωτήματος (β) για τα έτη 1996, 1997 σύμφωνα με τον πίνακα της άσκησης.

8. Τα παρακάτω δεδομένα αφορούν την οικογενειακή κατάσταση των υπαλλήλων μιας επιχείρησης.

Άνδρες ανύπαντροι: 22

Γυναίκες ανύπαντρες: 8

Άνδρες παντρεμένοι: 50

Γυναίκες παντρεμένες: 15

Άνδρες διαζευγμένοι: 3

Γυναίκες διαζευγμένες: 2

i) Να παρουσιαστεί η οικογενειακή κατάσταση των υπαλλήλων της επιχείρησης με τη βοήθεια πίνακα διπλής εισόδου.

ii) Να απεικονισθούν τα παραπάνω δεδομένα της οικογενειακής κατάστασης των υπαλλήλων με ραβδόγραμμα.

ΤΑ ΒΑΣΙΚΑ ΜΕΤΡΑ ΘΕΣΗΣ ΜΙΑΣ ΚΑΤΑΝΟΜΗΣ

4.1 Έννοια και σημασία των μέτρων θέσης
4.2 Μέσο εύρος
4.3 Μέση τιμή ή μέσος αριθμητικός
4.4 Διάμεσος
4.5 Επικρατούσα τιμή
4.6 Τεταρτημόρια
4.7 Η σπουδαιότητα των μέτρων θέσης στην επιχειρηματική δραστηριότητα
4.8 Εφαρμογές στις επιχειρήσεις

❖ *«Ο στατιστικός μέσος είναι μια περιγραφή, μια αντιπροσωπευτική ποσότητα που τίθεται για όλη την ομάδα, ο άριστος αντιπρόσωπος της ομάδας που εξετάζουμε»*
A. Quetelet (1796-1874).

❖ Ο **Quetelet** είναι από τους πρώτους που μελέτησαν συστηματικά και σε έκταση το πρόβλημα των μέσων τιμών, έχοντας παραδεχτεί ότι η έννοια των μέσων τιμών χρησιμοποιήθηκε ενστικτωδώς στην προεπιστημονική εποχή.

❖ Πολλοί σοφοί της αρχαιότητας, όπως ο **Αρχιμήδης**, ο **Μέτων** κ.ά., χρησιμοποίησαν τις μέσες τιμές στις έρευνές τους. Υποστηρίζεται ότι ο Πυθαγόρας βρήκε και χρησιμοποίησε πρώτος τον μέσο αρμονικό και τον μέσο γεωμετρικό.

ΚΕΦΑΛΑΙΟ 4ο

ΤΑ ΒΑΣΙΚΑ ΜΕΤΡΑ ΘΕΣΗΣ ΜΙΑΣ ΚΑΤΑΝΟΜΗΣ

4.1 Έννοια και σημασία των μέτρων θέσης

Στα προηγούμενα κεφάλαια ασχοληθήκαμε με τη συλλογή και την ταξινόμηση στατιστικών δεδομένων. Σκοπός μας ήταν να αντλήσουμε πληροφορίες από αυτά για τα χαρακτηριστικά του πληθυσμού που εξετάζουμε. Η συνοπτική παρουσίαση των δεδομένων με εποπτικές εικόνες και με πίνακες είδαμε ότι διευκόλυνε πολύ αυτή τη διαδικασία. Όμως η χρήση μόνο των στατιστικών πινάκων δεν είναι αρκετή σε πολλές περιπτώσεις. Μια από αυτές είναι ότι δεν μπορούμε με τους πίνακες να συγκρίνουμε εύκολα ομοειδείς έρευνες σε διαφορετικούς πληθυσμούς. Για παράδειγμα, αν έχουμε δύο πίνακες που παρουσιάζουν την επίδοση των μαθητών δύο τμημάτων της Β΄ Τ.Ε.Ε. στα Μαθηματικά, τότε δεν μπορούμε μόνο με τους πίνακες να συμπεράνουμε ποιο τμήμα έχει την καλύτερη επίδοση ή, ακόμα δυσκολότερο, να θυμόμαστε απέξω ολόκληρους πίνακες δεδομένων. Γι' αυτό είναι απαραίτητο, τις περισσότερες φορές, να αντικαταστήσουμε το σύνολο των αριθμητικών δεδομένων με ορισμένους αντιπροσωπευτικούς αριθμούς που ονομάζονται **παράμετροι ή μέτρα ή στατιστικά περιγραφικά μέτρα της κατανομής.**

Ονομάζουμε στατιστικό περιγραφικό μέτρο της κατανομής τον αριθμό που συνοψίζει βασικά χαρακτηριστικά των παρατηρήσεων του συνόλου των δεδομένων που εξετάζουμε.

Στην περίπτωση μονομεταβλητών δεδομένων (δεδομένων που προέρχονται από μετρήσεις ενός μεμονωμένου χαρακτηριστικού, π.χ. ύψος) έχουμε δύο τύπους περιληπτικών στατιστικών μέτρων:

- Τα στατιστικά μέτρα θέσης ή μέτρα θέσης και
- Τα στατιστικά μέτρα διασποράς ή μέτρα διασποράς.

Τα μέτρα θέσης είναι στατιστικές παράμετροι οι οποίες καθορίζουν μια τιμή της μεταβλητής, γύρω από την οποία τείνουν να συγκεντρωθούν οι παρατηρήσεις μας.

Στο κεφάλαιο αυτό θα μελετήσουμε τα κυριότερα μέτρα θέσης που χρησιμοποιούνται για την περιγραφή της γενικής τάσης των τιμών της μεταβλητής που εξετάζουμε. Αυτά είναι:

- Το μέσο εύρος.
- Η μέση τιμή ή μέσος αριθμητικός.
- Η διάμεσος.
- Η επικρατούσα τιμή και
- Τα τεταρτημόρια

4.2 Μέσο εύρος

Το πρώτο μέτρο θέσης που θα εξετάσουμε είναι το μέσο εύρος που ορίζεται ως εξής:

$$\textbf{Μέσο εύρος} = \frac{\textbf{Μεγαλύτερη τιμή + Μικρότερη τιμή}}{\textbf{2}}$$

(4.1)

Αν, για παράδειγμα, οι τιμές (σε ευρώ) της μετοχής μιας εταιρείας σε 5 συνεχείς συνεδρίες του Χρηματιστηρίου Αξιών Αθηνών (Χ.Α.Α) ήταν:

22, 20, 23, 16, 18

τότε το μέσο εύρος της τιμής της ήταν $\frac{16+23}{2} = 19{,}5$ Ευρώ

4.3 Μέση τιμή ή μέσος αριθμητικός

Η μέση τιμή ή μέσος αριθμητικός (*arithmetic mean*) είναι το δεύτερο σημαντικό μέτρο θέσης που θα μας απασχολήσει.

Έστω ότι μια μεταβλητή Χ παίρνει τις τιμές x_1, x_2 ,…$x_ν$.

Ορίζουμε ως **μέση τιμή** ή **μέσο αριθμητικό** των τιμών x_1, x_2 ,…,$x_ν$ της μεταβλητής Χ το πηλίκο του αθροίσματος των τιμών αυτών προς το πλήθος τους ν.

Δηλαδή: $$\text{μέση τιμή} = \frac{x_1 + x_2 + \dots + x_ν}{ν}$$

Αν εξετάζαμε όλο τον πληθυσμό, η μέση τιμή συμβολίζεται διεθνώς με το ελληνικό γράμμα **μ**, ενώ αν αναφερόμαστε σε δείγμα του πληθυσμού, συμβολίζεται με $\bar{\mathbf{x}}$

Στα παρακάτω θεωρούμε ότι μελετούμε δείγματα του πληθυσμού και γι' αυτό θα χρησιμοποιούμε το σύμβολο $\bar{x}$ για τη μέση. Επομένως:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_ν}{ν} = \frac{\sum_{i=1}^{ν} x_i}{ν} = \frac{\sum x}{ν}$$

(4.2)

Παράδειγμα 4.1:

Η διοίκηση επτά εμπορικών επιχειρήσεων, E_1, E_2, E_3, E_4, E_5, E_6, E_7, θέλει να υπολογίσει τη μέση ετήσια αξία των πωλήσεων των επιχειρήσεων. Το ύψος των ετήσιων πωλήσεων κάθε επιχείρησης (σε χιλιάδες ευρώ) παρουσιάζεται στον παρακάτω πίνακα:

Πίνακας 4.1

E_1	**E_2**	**E_3**	**E_4**	**E_5**	**E_6**	**E_7**
35	30	25	20	20	35	40

Ο υπεύθυνος πωλήσεων υπολογίζει τη **μέση τιμή ή μέσο αριθμητικό** των πωλήσεων των επτά επιχειρήσεων ως εξής:

$$\text{Μέση τιμή} = \frac{35+30+25+20+20+35+40}{7} = \frac{205}{7} = 29{,}286$$

χιλιάδες ευρώ και καταλήγει στο συμπέρασμα ότι η μέση τιμή της αξίας των πωλήσεων των επιχειρήσεών του ετησίως ήταν 29.286 ευρώ. Παρατηρεί επίσης ότι η επιχείρηση E_2 είχε ετήσιο ύψος πωλήσεων 30.000 ευρώ, ποσό που βρίσκεται πολύ κοντά στη μέση τιμή. Αντίστοιχα οι E_4 και E_5 είχαν σημαντικά μειωμένο ποσό πωλήσεων σε σχέση με τη μέση τιμή, ενώ η E_7 είχε σημαντικά μεγαλύτερο ποσό πωλήσεων.

4.3.1 Υπολογισμός της μέσης τιμής

Στη περίπτωση των αταξινόμητων δεδομένων (δηλαδή αυτών που δεν περιέχονται σε πίνακα συχνοτήτων), η μέση τιμή υπολογίζεται με τη βοήθεια της σχέσης (4.2).

Σε πολλά προβλήματα όμως, το πλήθος των δεδομένων είναι αρκετά μεγάλο.

- Στην περίπτωση αυτή ο **τρόπος υπολογισμού της μέσης τιμής γίνεται ευκολότερος, αν κατασκευάσουμε τον πίνακα συχνοτήτων**. Ένα τέτοιο πρόβλημα παρουσιάζεται στο ακόλουθο παράδειγμα και αφορά διακριτή μεταβλητή:

Παράδειγμα 4.2

Δίνεται ο αριθμός των υπαλλήλων που εργάζονται σε 25 εμπορικά καταστήματα:

6,1,3,5,4,2,6,3,6,6,4,3,2,5,6,7,4,5,4,7,7,8,7,8,6

Να υπολογιστεί η μέση τιμή του αριθμού των υπαλλήλων που εργάζονται στις επιχειρήσεις αυτές.

Λύση

Η μεταβλητή που εξετάζουμε είναι ο «αριθμός των υπαλλήλων που απασχολεί μια επιχείρηση». Κατασκευάζουμε τον πίνακα διαλογής και τον πίνακα κατανομής συχνοτήτων αυτής της μεταβλητής (πίνακας 4.2).

Πίνακας 4.2

Αριθμός υπαλλήλων (x_i)	Διαλογή	Συχνότητα (v_i)
1	I	1
2	II	2
3	III	3
4	IIII	4
5	III	3
6	IIII II	6
7	IIII	4
8	II	2
Σύνολο		25

Από τον πίνακα διαλογής παρατηρούμε ότι υπάρχουν επιχειρήσεις που απασχολούν τον ίδιο αριθμό υπαλλήλων. Στην περίπτωση αυτή η μέση τιμή υπολογίζεται ευκολότερα ως εξής:

$$\bar{x} = \frac{1\cdot 1 + 2\cdot 2 + 3\cdot 3 + 4\cdot 4 + 3\cdot 5 + 6\cdot 6 + 4\cdot 7 + 2\cdot 8}{25} = 5$$

Επομένως η μέση τιμή του αριθμού των υπαλλήλων ανά επιχείρηση είναι 5.

Από τα παραπάνω προκύπτει ότι **σε διακριτά δεδομένα η μέση τιμή υπολογίζεται με την ακόλουθη διαδικασία:**

- **Τα δεδομένα ταξινομούνται σε πίνακα συχνοτήτων**
- **Η μέση τιμή υπολογίζεται από τη σχέση:**

$$\bar{x} = \frac{v_1x_1 + v_2x_2 + \ldots + v_Kx_K}{v_1 + v_2 + \ldots + v_K} = \frac{v_1x_1 + v_2x_2 + \ldots + v_Kx_K}{v} = \frac{\sum_{i=1}^{K} v_ix_i}{v}$$

ή συντομότερα $\bar{x} = \frac{\sum vx}{v}$

(4.3)

Στην περίπτωση που τα ταξινομημένα δεδομένα είναι τιμές διακριτής μεταβλητής η μέση τιμή υπολογίζεται με τη βοήθεια της σχέσης (4.3).

- Όταν όμως τα δεδομένα είναι **τιμές συνεχούς μεταβλητής, τότε αυτά ομαδοποιούνται** και η μέση τιμή υπολογίζεται με τη διαδικασία που περιγράφεται στο επόμενο παράδειγμα.

Παράδειγμα 4.3

Τα έσοδα (σε δεκάδες χιλιάδες ευρώ – 10.000 €) 100 εμπορικών επιχειρήσεων ενός ομίλου, κατά τη διάρκεια μιας ορισμένης χρονικής περιόδου εμφανίζονται ομαδοποιημένα στον πίνακα 4.3:

Πίνακας 4.3

Έσοδα (σε δεκ. χιλ. ευρώ)	Επιχειρήσεις
0-10	4
10-20	11
20-30	12
30-40	13
40-50	25
50-60	15
60-70	8
70-80	7
80-90	3
90-100	2

Να βρεθεί η μέση τιμή των εσόδων των επιχειρήσεων του ομίλου.

Λύση

Υπολογίζουμε την κεντρική τιμή κάθε κλάσης και προσθέτουμε στον πίνακα συχνοτήτων δύο στήλες, τη στήλη των κεντρικών τιμών (x'_i) και τη στήλη των γινομένων των συχνοτήτων με την αντίστοιχη κεντρική τιμή ($\nu_i x'_i$)

Πίνακας 4.4

Έσοδα (σε δεκ. χιλ. ευρώ) (x_i)	**Κεντρική τιμή κλάσης (x_i')**	**Αριθμός επιχειρήσεων (ν_i)**	**$\nu_i x_i'$**
[0-10)	5	4	20
[10-20)	15	11	165
[20-30)	25	12	300
[30-40)	35	13	455
[40-50)	45	25	1125
[50-60)	55	15	825
[60-70)	65	8	520
[70-80)	75	7	525
[80-90)	85	3	255
[90-100)	95	2	190
Σύνολο		ν =100	$\sum \nu x'$ 4380

Η μέση τιμή υπολογίζεται από τη σχέση:

$$\bar{x} = \frac{\nu_1 x_1' + \nu_2 x_2' + \ldots + \nu_K x_K'}{\nu_1 + \nu_2 + \ldots + \nu_K}$$

Δηλαδή,

$$\bar{x} = \frac{20+165+300+455+1125+825+520+525+255+190}{100} \text{ δεκ. χιλ. ευρώ}$$

$$\bar{x} = \frac{4380}{100} = 43{,}8 \text{ δεκ. χιλ. ευρώ}$$

Άρα η μέση τιμή των εσόδων είναι 438.000 ευρώ.

Για τον υπολογισμό λοιπόν του μέσου αριθμητικού σε ομαδοποιημένα δεδομένα παίρνουμε ως τιμές x_i της μεταβλητής X **τις κεντρικές τιμές x_i' των κλάσεων**. Στη συνέχεια η μέση τιμή υπολογίζεται εύκολα από τη σχέση:

$$\bar{x} = \frac{\nu_1 x_1' + \nu_2 x_2' + \ldots + \nu_\kappa x_\kappa'}{\nu_1 + \nu_2 + \ldots + \nu_\kappa} = \frac{\sum_{i=1}^{\kappa} \nu_i x_i'}{\nu} = \frac{\sum \nu x'}{\nu}$$

(4.4)

4.3.2 Ιδιότητες της μέσης τιμής.

Μερικές χαρακτηριστικές ιδιότητες της μέσης τιμής είναι οι ακόλουθες:

1) Αν σε όλες τις τιμές μιας μεταβλητής Χ προσθέσουμε (ή αφαιρέσουμε) μια σταθερή ποσότητα κ, τότε η μέση τιμή αυξάνεται (ή ελαττώνεται) κατά τη σταθερή αυτή ποσότητα.

Παράδειγμα 4.4

Ο ιδιοκτήτης των επιχειρήσεων του παραδείγματος 4.1 θέλοντας να αυξήσει τα έσοδά του έκανε μια μεγάλη διαφημιστική εκστρατεία. Μετά το τέλος της διαφημιστικής αυτής καμπάνιας παρατηρήθηκε ότι το ετήσιο ύψος των πωλήσεων κάθε επιχείρησης αυξήθηκε κατά 2 χιλιάδες ευρώ. Επομένως τα έσοδα των επιχειρήσεων είναι (σε χιλιάδες ευρώ):

E_1	E_2	E_3	E_4	E_5	E_6	E_7
37	32	27	22	22	37	42

Η νέα μέση τιμή είναι:

$$\bar{x}^* = \frac{37+32+27+22+22+37+42}{7} = 31{,}286 \text{ χιλιάδες ευρώ}$$

δηλαδή, είναι αυξημένη κατά 2 χιλ. ευρώ σε σχέση με την προηγούμενη.

Γενικά: $\bar{x}^* = \bar{x} \pm \kappa$

όπου $\bar{x}^*$ η μέση τιμή που προκύπτει από τα δεδομένα, τα οποία έχουν μεταβληθεί κατά μια σταθερή ποσότητα κ.

Με αντίστοιχα παραδείγματα μπορούμε να δείξουμε και τις επόμενες ιδιότητες του μέσου αριθμητικού:

2) Αν οι τιμές μιας μεταβλητής Χ είναι ίσες μεταξύ τους, τότε η μέση τιμή των τιμών της είναι μια από αυτές τις τιμές, δηλαδή

$$\boxed{\text{Αν } x_1 = x_2 = \ldots = x_\nu = k, \text{ τότε } \bar{x} = k}$$

3) Αν όλες οι τιμές της μεταβλητής Χ πολλαπλασιαστούν επί μια σταθερή ποσότητα α, τότε και η μέση τιμή πολλαπλασιάζεται επί α.

Δηλαδή: $\boxed{\bar{x}^* = \alpha\bar{x}}$

4) Αν από όλες τις τιμές μιας μεταβλητής Χ αφαιρέσουμε τη μέση τιμή, τότε το άθροισμα των διαφορών $(x_i - \bar{x})$ είναι ίσο με το μηδέν

Δηλαδή: $\boxed{\sum(x_i - \bar{x}) = 0.}$

όπου i = 1, 2, …, ν.

5) Αν ένας πληθυσμός χωριστεί σε κ υποπληθυσμούς, που ο καθένας τους έχει N_1, N_2,…N_κ μονάδες και αντίστοιχους μέσους μ_1, μ_2,…μ_κ, τότε ο μέσος αριθμητικός όλου του πληθυσμού ισούται με:

$$\boxed{\mu = \frac{N_1\mu_1 + N_2\mu_2 + \ldots + N_\kappa\mu_\kappa}{N_1 + N_2 + \ldots + N_\kappa}}$$

4.4 Διάμεσος

Το επόμενο μέτρο θέσης που θα μελετήσουμε είναι η **διάμεσος** (median) **ή διάμεσος τιμή.**

Αν οι τιμές μια μεταβλητής έχουν διαταχθεί από τη μικρότερη μέχρι τη μεγαλύτερη σε αύξουσα τάξη, τότε **η μεσαία παρατήρηση είναι η διάμεσος (δ).**

Η διάμεσος ορίζεται ως:

- **Η τιμή της μεταβλητής που αντιστοιχεί στη μεσαία παρατήρηση, αν το πλήθος των παρατηρήσεων είναι περιττό.**

- **Το ημιάθροισμα των τιμών της μεταβλητής που αντιστοιχούν στις δύο μεσαίες παρατηρήσεις, αν το πλήθος τους είναι άρτιο.**

Παράδειγμα 4.5:

α) Τα ημερομίσθια 5 εργατών (σε χιλιάδες ευρώ) είναι τα ακόλουθα:
38, 40, 35, 38, 39
Αν τοποθετηθούν σε αύξουσα τάξη είναι:
35, 38, **38**, 39, 40
Στην περίπτωση αυτή η διάμεσος είναι **δ = 38**, δηλαδή το διάμεσο ημερομίσθιο είναι 38 ευρώ.

β) Αν τα ημερομίσθια οκτώ (8) εργατών είναι:
30, 28, 31, 28, 35, 36, 40, 37 (ευρώ)
τότε τοποθετώντας τις τιμές σε αύξουσα τάξη έχουμε:
28, 28, 30, 31, 35, 36, 37, 40

Παρατηρούμε ότι έχουμε **δύο μεσαίες τιμές, τις 31 και 35. Η διάμεσος, λοιπόν, είναι το ημιάθροισμα των δύο μεσαίων τιμών:**

$\delta = \frac{31+35}{2} = \frac{66}{2} = 33$, δηλαδή το διάμεσο ημερομίσθιο ισούται με 33 ευρώ.

Γενικά:

Διάμεσος ή διάμεση τιμή (δ) *λέγεται η τιμή της μεταβλητής που χωρίζει το σύνολο των τιμών σε δύο ισοπληθείς ομάδες, όταν οι τιμές που παίρνει η μεταβλητή τοποθετηθούν σε αύξουσα τάξη.*

Είναι φανερό ότι η διάμεσος είναι η τιμή της μεταβλητής που αντιστοιχεί στη $\frac{\nu+1}{2}$ θέση, όπου ν είναι το πλήθος των τιμών της μεταβλητής Χ τοποθετημένων σε αύξουσα τάξη.

Η διάμεσος στο παράδειγμα 4.1 είναι η τιμή της μεταβλητής που βρίσκεται στην $\frac{7+1}{2} = 4$η θέση, εφόσον οι τιμές τοποθετηθούν σε αύξουσα τάξη (20, 20, 25, 30, 35, 35, 40). Η διάμεσος τιμή των ετησίων πωλήσεων ισούται με 30.000 ευρώ.

4.4.1 Υπολογισμός της διαμέσου

- Όταν **οι τιμές της μεταβλητής δίνονται σε πίνακα συχνοτήτων**, για να βρούμε τη διάμεσο συμπληρώνουμε τον πίνακα με τη στήλη των αθροιστικών συχνοτήτων και εργαζόμαστε όπως στο παράδειγμα 4.6 που ακολουθεί:

Παράδειγμα 4.6

Ο πίνακας 4.6 παρουσιάζει τη βαθμολογία μιας ομάδας σπουδαστών στις εξετάσεις του μαθήματος της Στατιστικής:

Πίνακας 4.6

Βαθμός (x_i)	1	2	3	4	5	6	7	8	9	10
Αριθμός σπουδαστών (v_i)	5	3	5	2	20	20	25	10	5	5

Να βρεθεί η διάμεσος των τιμών της μεταβλητής «βαθμός στο μάθημα της Στατιστικής»

Λύση

Κατασκευάζουμε τον πίνακα (4.7) των αθροιστικών συχνοτήτων:

Πίνακας 4.7

Βαθμός (x_i)	**Αριθμός σπουδαστών Συχνότητα (v_i)**	**Αθροιστική συχνότητα (N_i)**
1	5	5
2	3	8
3	5	13
4	2	15
5	20	35
6	20	55
7	25	80
8	10	90
9	5	95
10	5	100
Σύνολο	100	

Θα μπορούσαμε να προσδιορίσουμε τη θέση της διαμέσου των τιμών της μεταβλητής Χ: «βαθμός στο μάθημα της Στατιστικής» λέγοντας ότι η διάμεσος θα αφήνει το 50% των παρατηρήσεων κάτω από αυτή και 50% πάνω από αυτή. Θα είναι λοιπόν η τιμή της μεταβλητής που αντιστοιχεί στη $\frac{\nu+1}{2}$ θέση, δηλαδή στην 50,5η θέση. Στη συγκεκριμένη περίπτωση εννοούμε ότι η διάμεσος θα βρίσκεται ανάμεσα στην 50η και στην 51η παρατήρηση, επειδή το πλήθος των τιμών είναι άρτιο και δεν υπάρχει μεσαία παρατήρηση. Από την αθροιστική συχνότητα βρίσκουμε ότι υπάρχουν 35 σπουδαστές με βαθμό Στατιστικής μέχρι και 5 και 55 σπουδαστές με βαθμό μέχρι και 6, οπότε ο βαθμός τόσο του 50ου όσο και του 51ου σπουδαστή θα είναι 6. Άρα η διάμεσος είναι η τιμή της μεταβλητής Χ που έχει αθροιστική συχνότητα N_6 = 55 και ισούται με δ = $\frac{6+6}{2}=6$.

Συμπεραίνουμε λοιπόν ότι για να υπολογίσουμε τη διάμεσο σε **ταξινομημένα διακριτά δεδομένα** ακολουθούμε την παρακάτω διαδικασία:

- Συμπληρώνουμε τον πίνακα συχνοτήτων με τη στήλη των αθροιστικών συχνοτήτων.

- Προσδιορίζουμε τις αθροιστικές συχνότητες N_{i-1} και N_i μεταξύ των οποίων βρίσκεται ο αριθμός $\frac{\nu+1}{2}$ που αντιστοιχεί στη θέση της διαμέσου.

- Η τιμή (x_i) της μεταβλητής Χ που βρίσκεται στη $\frac{\nu+1}{2}$ θέση είναι η διάμεσος

- Όταν **οι τιμές της μεταβλητής είναι ομαδοποιημένες,** για να υπολογίσουμε τη διάμεσο εργαζόμαστε όπως στο παράδειγμα (4.8)

Παράδειγμα 4.8

Ο παρακάτω πίνακας δείχνει τα ετήσια τηλεφωνικά έξοδα (σε ευρώ) των 1200 νοικοκυριών μιας κωμόπολης:

Πίνακας 4.8

Τηλεφωνικά έξοδα (σε ευρώ)	Αριθμός νοικοκυριών
0 - 50	50
50 - 100	200
100 - 150	250
150 - 200	250
200 - 250	200
250 - 300	150
300 - 350	60
350 - 400	20
400 - 450	10
450 - 500	10

Να υπολογιστεί η διάμεσος της μεταβλητής «Τηλεφωνικά έξοδα».

Λύση

Συμπληρώνουμε τον πίνακα συχνοτήτων με τη στήλη της αθροιστικής συχνότητας.

Πίνακας 4.9

α/α	Τηλεφωνικά έξοδα (x_i σε ευρώ)	Αριθμός ν_i	Αθροιστική συχνότητα N_i	
1	[0, 50)	50	50	N_{i-1}
2	[50, 100)	200	250	
3	[100, 150)	250	500	
4	[150, 200)	250	750	$\frac{\nu}{2}=600$ (θέση διαμέσου)
5	[200, 250)	200	950	
6	[250, 300)	150	1100	
7	[300, 350)	60	1160	
8	[350, 400)	20	1180	N_i
9	[400, 450)	10	1190	
10	[450, 500)	10	1200	
	Σύνολο	1200		

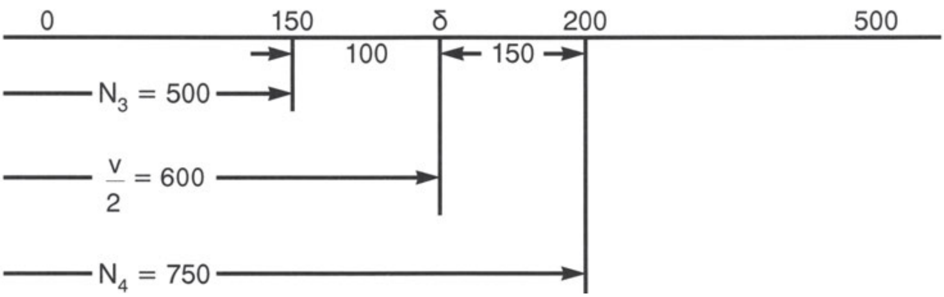
Θα προσδιορίσουμε και πάλι τη θέση της διαμέσου λέγοντας ότι το πολύ το 50% των παρατηρήσεων (50% · 1200 = 600) βρίσκονται κάτω από αυτή.

Άρα η διάμεσος θα βρίσκεται ανάμεσα στην 600 και 601 παρατήρηση γιατί το πλήθος των δεδομένων είναι άρτιο.

Γενικεύοντας, σε ομαδοποιημένα δεδομένα θα μπορούσαμε να θεωρήσουμε ως θέση της διαμέσου τη $\frac{\nu}{2}$ θέση, όπου ν το πλήθος των τιμών της μεταβλητής.

Από τη στήλη των αθροιστικών συχνοτήτων βλέπουμε ότι ο αριθμός 600 βρίσκεται ανάμεσα στις αθροιστικές συχνότητες $N_3 = 500$ και $N_4 = 750$. Άρα η διάμεσος περιέχεται στην κλάση $[\alpha_{i-1}, \alpha_i) = [150, 200)$, στην οποία αντιστοιχεί η αθροιστική συχνότητα $N_4 = 750$.

Υποθέτοντας ότι οι παρατηρήσεις σε κάθε κλάση κατανέμονται ομοιόμορφα θα υπολογίσουμε τη διάμεσο με την απλή μέθοδο των τριών.



(Σχ.4.1)

Στη συγκεκριμένη κλάση περιέχονται $\nu_4 = 250$ παρατηρήσεις οι οποίες κατανέμονται ομοιόμορφα σε διάστημα πλάτους c = 50, άρα η κάθε παρατήρηση καταλαμβάνει πλάτος $\frac{c}{\nu_i} = \frac{50}{250} = \frac{1}{5}$.

Επομένως οι 100 παρατηρήσεις ($\frac{\nu}{2} - N_{i-1} = 600 - 500$) που περιέχονται στην ίδια κλάση αρχίζοντας από 501$^{\eta}$ θα κατανέμονται ομοιόμορφα σε διάστημα πλάτους: $100 \cdot \frac{1}{5} = 20$

Το πλάτος όμως αυτό ισούται με $\delta - \alpha_{i-1} = \delta - 150$.

Άρα $\delta - 150 = 100 \cdot \frac{1}{5}$.

δ = 150 + 20 δηλαδή η διάμεσος των τηλεφωνικών εξόδων είναι 170 ευρώ.

Γενικεύοντας, αντικαθιστούμε στη σχέση (1) και έχουμε ότι:

$$\delta - \alpha_{i-1} = \left(\frac{\nu}{2} - N_{i-1}\right) \cdot \frac{c}{\nu_i}$$

$$\boxed{\delta = \alpha_{i-1} + \left(\frac{\nu}{2} - N_{i-1}\right) \cdot \frac{c}{\nu_i}}$$

(4.5)

όπου:

α_{i-1} = το κατώτερο όριο της κλάσης που περιέχει τη διάμεσο.
ν_i = η συχνότητα της κλάσης.
c = το πλάτος της κλάσης (αναφερόμαστε σε κλάσεις ίσου πλάτους).
N_{i-1} = η αθροιστική συχνότητα της προηγούμενης κλάσης.
ν = το πλήθος των παρατηρήσεων.

4.5 Επικρατούσα τιμή

Ένα άλλο μέτρο θέσης είναι η **επικρατούσα τιμή** (Ε.Τ).

Ως επικρατούσα τιμή (mode) **ορίζεται η τιμή της μεταβλητής που έχει τη μεγαλύτερη συχνότητα**, για παράδειγμα, αν οι τιμές της μεταβλητής είναι:

12, 19, **16**, **16**, 17, 13, **16**, 20

τότε η επικρατούσα τιμή είναι **16,** που έχει συχνότητα 3

Αν όλες οι τιμές έχουν την ίδια συχνότητα, τότε δεν ορίζεται επικρατούσα τιμή.

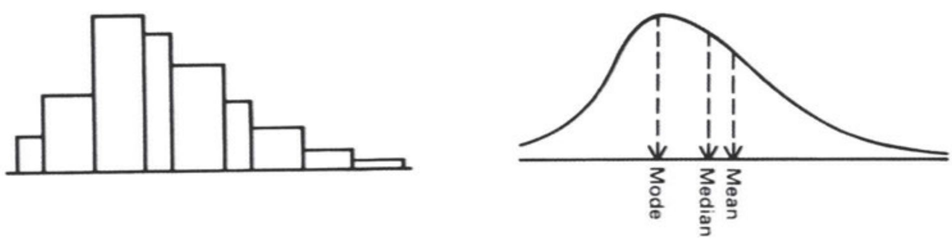
➢ Τα τέσσερα μέτρα θέσης που μελετήσαμε χρησιμοποιούνται για την περιγραφή της γενικής τάσης των τιμών της μεταβλητής Χ. Το ερώτημα όμως που γεννάται είναι ποιο είναι το καταλληλότερο μέτρο για να συνοψίσει βασικά χαρακτηριστικά των δεδομένων μας.

- Το μέσο εύρος είναι εύκολο να υπολογιστεί, αλλά για τον υπολογισμό του χρησιμοποιούνται μόνο οι δύο ακραίες τιμές της κατανομής.
- Η μέση τιμή λαμβάνει υπόψη όλες τις τιμές της μεταβλητής.
- Η διάμεσος δεν επηρεάζεται από τις ακραίες τιμές και για τον υπολογισμό της λαμβάνονται υπόψη όλες οι τιμές.
- Η επικρατούσα τιμή είναι ικανοποιητικό μέτρο θέσης, αν **η κατανομή είναι συμμετρική, δηλαδή αν η μέση τιμή ($\bar{x}$), η διάμεσος (δ) και η επικρατούσα τιμή (Ε.Τ) συμπίπτουν** όπως φαίνεται στο σχήμα 4.2.

Η κατανομή αυτή ονομάζεται **κανονική κατανομή**.

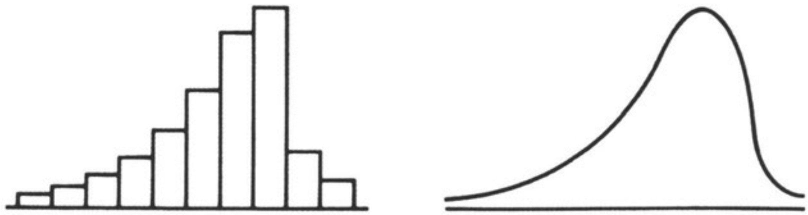
(Σχ.4.2)

Κατανομή που δεν είναι συμμετρική ονομάζεται **ασύμμετρη**. Μια ασύμμετρη κατανομή παρουσιάζει **θετική ασυμμετρία** όταν η μέση τιμή είναι μεγαλύτερη από τη διάμεσο και την επικρατούσα τιμή (Σχ.4.3).



(Σχ.4.3)

Μια ασύμμετρη κατανομή παρουσιάζει **αρνητική ασυμμετρία** όταν η μέση τιμή (Μ.Τ) είναι μικρότερη από τη διάμεσο (δ) και την επικρατούσα τιμή (Ε.Τ)



(Σχ.4.4)

Συμπερασματικά, τα πιο εύχρηστα μέτρα είναι το μέσο εύρος και η μέση τιμή, αλλά η τιμή και των δύο επηρεάζεται σημαντικά από τις ακραίες τιμές της κατανομής. Η επικρατούσα τιμή δεν είναι πάντα αντιπροσωπευτικό μέτρο της κατανομής, ιδιαίτερα αν η κατανομή δεν είναι συμμετρική. Έτσι θα μπορούσαμε να πούμε ότι η διάμεσος είναι το καλύτερο μέτρο θέσης μιας κατανομής, δεν είναι όμως πάντα εύκολος ο τρόπος υπολογισμού της.

4.6 Τεταρτημόρια

Σε συνέχεια του ορισμού της διαμέσου ορίζουμε τα τεταρτημόρια (quartiles)

Τεταρτημόρια λέγονται οι τιμές της μεταβλητής που χωρίζουν το σύνολο των τιμών της σε 4 ισοπληθείς ομάδες, όταν οι τιμές της μεταβλητής τοποθετηθούν σε αύξουσα τάξη.

✓ Δηλαδή, σε ένα σύνολο διατεταγμένων παρατηρήσεων σε αύξουσα τάξη ορίζουμε:

Το **πρώτο τεταρτημόριο Q_1** ως την τιμή της μεταβλητής, κάτω της οποίας βρίσκεται το πολύ το 25% του συνολικού αριθμού των παρατηρήσεων.

Το **δεύτερο τεταρτημόριο Q_2** συμπίπτει με τη διάμεσο, η οποία έχει ήδη ορισθεί, δηλαδή $Q_2 = δ$.

Τέλος **το τρίτο τεταρτημόριο Q_3** ορίζεται αντίστοιχα ως η τιμή της μεταβλητής, κάτω της οποίας βρίσκεται το πολύ το 75% του συνολικού αριθμού των παρατηρήσεων.

Σύμφωνα με τα παραπάνω, αν ν είναι το πλήθος των διατεταγμένων παρατηρήσεων, τότε η θέση που κατέχει το πρώτο τεταρτημόριο Q_1 είναι $\frac{\nu+1}{4}$ και η θέση που κατέχει το τρίτο τεταρτημόριο Q_3 είναι $\frac{3(\nu+1)}{4}$.

Ο υπολογισμός των τεταρτημορίων σε **ομαδοποιημένα δεδομένα** γίνεται όπως ο αντίστοιχος της διαμέσου ή γραφικά με τη βοήθεια του διαγράμματος των αθροιστικών συχνοτήτων. Η θέση των τεταρτημορίων Q_1 και Q_3 σε ομαδοποιημένα δεδομένα δεχόμαστε ότι βρίσκεται στη $\frac{\nu}{4}$ και $\frac{3\nu}{4}$ θέση αντίστοιχα.

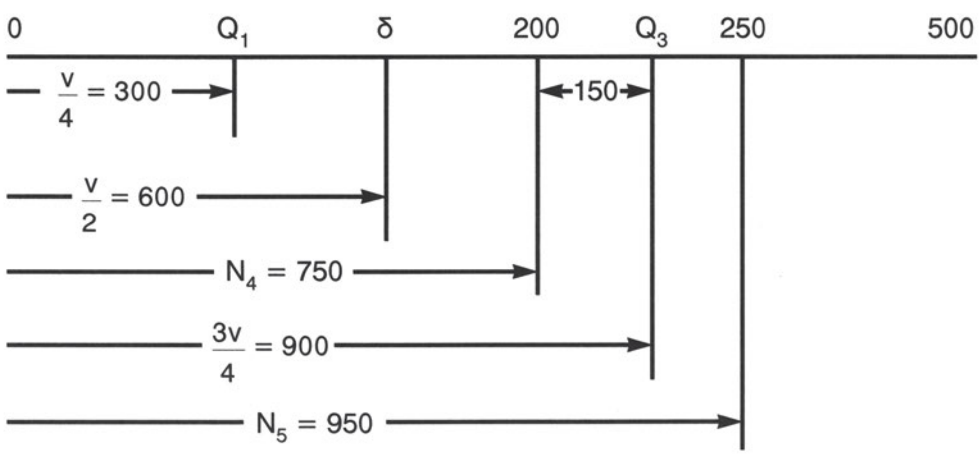
Παράδειγμα 4.9:

Στο παράδειγμα 4.8 υπολογίσαμε τη διάμεσο των τιμών της μεταβλητής «Τηλεφωνικά έξοδα». Ζητάμε να υπολογίσουμε το 1ο και το 3ο τεταρτημόριο.

Λύση

Από τον πίνακα (4.9) των αθροιστικών συχνοτήτων εντοπίζουμε τη θέση των Q_1 και Q_3 στον $\frac{\nu}{4} = 300$ και $\frac{3\nu}{4} = 900$ αριθμό αντίστοιχα.

Για τον υπολογισμό του Q_1 και Q_3 ακολουθούμε την εξής διαδικασία, όπως φαίνεται στο σχήμα (4.5):



(Σχ. 4.5)
Υπολογισμός Q_3

Υποθέτουμε ότι οι παρατηρήσεις είναι ομοιόμορφα κατανεμημένες σε κάθε κλάση. Παρατηρούμε ότι ο αριθμός 900 βρίσκεται μεταξύ των αθροιστικών συχνοτήτων $N_4 = 750$ και $N_5 = 950$, άρα το Q_3 θα είναι η τιμή της 900ής παρατήρησης, η οποία περιέχεται στην κλάση [200, 250). Η κλάση αυτή έχει συχνότητα $\nu_5 = 200$ και αθροιστική συχνότητα $N_5 = 950$, άρα η 900ή παρατήρηση θα βρίσκεται στα

$$\frac{\frac{3\nu}{4} - N_{i-1}}{\nu_i} = \frac{900 - 750}{200} = \frac{150}{200} = \frac{3}{4}$$ του συνολικού πλάτους ($c = 50$) της κλάσης και

θα ισούται με: $Q_3 = 200 + \frac{3}{4} \cdot 50 = 237{,}5$

Ομοίως $Q_1 = 100 + \frac{50}{250} \cdot 50 = 100 + 10 = 110$

Θα μπορούσαμε να ερμηνεύσουμε τις τιμές του Q_1 και Q_3 λέγοντας ότι το 25% των νοικοκυριών δαπάνησε για τηλεφωνικά έξοδα μέχρι και 110 ευρώ, ενώ το 75% των νοικοκυριών δαπάνησε ετησίως για τηλεφωνικά έξοδα μέχρι και 237,5 ευρώ.

Χάρη συντομίας και όπως προκύπτει από ανάλογη διαδικασία με εκείνη της διαμέσου, μπορούμε να διατυπώσουμε τύπους υπολογισμού των τεταρτημορίων Q_1 και Q_3 αντίστοιχα:

$$Q_1 = \alpha_{i-1} + \frac{\frac{\nu}{4} - N_{i-1}}{\nu_i} \cdot c$$

$$Q_3 = \alpha_{i-1} + \frac{\frac{3\nu}{4} - N_{i-1}}{\nu_i} \cdot c$$

(4.6)

Ο υπολογισμός των τεταρτημορίων καθώς και της διαμέσου μπορεί να γίνει και γραφικά από το διάγραμμα αθροιστικών συχνοτήτων.

4.7 Η σπουδαιότητα των μέτρων θέσης στην επιχειρηματική δραστηριότητα

Τα μέτρα θέσης που μελετήσαμε, και ιδιαίτερα η μέση τιμή και η διάμεσος, είναι από τα πιο συχνά χρησιμοποιούμενα στατιστικά μέτρα και αποτελούν πολύτιμο πληροφοριακό υλικό για την άσκηση οποιασδήποτε επιχειρηματικής πολιτικής. Ακόμη και από την κρατική πλευρά υπάρχει μεγάλο ενδιαφέρον για αυτά, αφού η εφαρμογή τους βοηθάει πολύ στην άσκηση ορθής κοινωνικής, οικονομικής κτλ. πολιτικής.

Παρόλο όμως που τα μέτρα θέσης είναι πολύ σημαντικά και απαραίτητα, δεν μας δίνουν πλήρη εικόνα του φαινομένου που εξετάζουμε.

Έτσι τις περισσότερες φορές δεν προχωρούμε σε εξαγωγή συμπερασμάτων και σε λήψη επιχειρηματικών αποφάσεων, αν δεν λάβουμε υπόψη μας και τις πρόσθετες πληροφορίες που παρέχουν τα μέτρα διασποράς, για τα οποία θα μιλήσουμε στο επόμενο κεφάλαιο.

Στην επόμενη παράγραφο παρατίθεται ένα απλό παράδειγμα από τον επιχειρηματικό χώρο, στο οποίο θα εφαρμόσουμε όλα τα στατιστικά μέτρα που μάθαμε μέχρι τώρα, για να τα κατανοήσουμε καλύτερα.

4.8 Εφαρμογές στις επιχειρήσεις

Περίπτωση μελέτης οφειλών σε προμηθευτές:

Από τους ισολογισμούς 40 επιχειρήσεων πήραμε τα παρακάτω στοιχεία, που είναι οι οφειλές (υποχρεώσεις) αυτών των επιχειρήσεων στους προμηθευτές τους, σε χιλιάδες ευρώ: 2, 4, 3, 7, 18, 16, 13, 62, 66, 68, 68, 74, 76, 76, 77, 51, 52, 54, 54, 56, 58, 22, 26, 26, 28, 32, 32, 33, 34, 37, 38, 38, 42, 44, 46, 48, 48, 48, 48, 49.

Ζητούνται:

α) Να βρεθεί η επικρατούσα τιμή, β) να ομαδοποιηθούν τα παραπάνω δεδομένα σε κλάσεις ίσου πλάτους, γ) να υπολογισθεί η μέση τιμή και δ) να υπολογιστούν η διάμεσος, το πρώτο και το τρίτο τεταρτημόριο.

Λύση

α) Η επικρατούσα τιμή είναι 48, η οποία έχει εμφανιστεί περισσότερες φορές από κάθε άλλη τιμή στην ομάδα των στατιστικών στοιχείων που εξετάζουμε. Δηλαδή, 4 επιχειρήσεις οφείλουν από 48.000 ευρώ η κάθε μια.

β) Ομαδοποιούμε τα δεδομένα σε οκτώ κλάσεις ίσου εύρους c = 10 και σχηματίζουμε τις στήλες (1) και (2) του πίνακα 4.10.

γ) Για τον υπολογισμό της μέσης τιμής κατασκευάζουμε τις στήλες (3) (των κεντρικών τιμών) και (4) στον πίνακα 4.10, από τον οποίο βρίσκουμε ν = 40 και $\Sigma\nu x' = 1660$. Αντικαθιστώντας τα δύο αυτά αθροίσματα στον τύπο εύρεσης της μέσης τιμής έχουμε:

$$\bar{x} = \frac{\sum \nu x'}{\sum \nu} = \frac{1660}{40} = 41{,}5$$

Δηλαδή $\bar{x}$ = 41.500 ευρώ.

Πίνακας 4.10

(1) **Κλάσεις οφειλών**	(2) **Αριθμός προμηθευτών** ν_i	(3) **Κεντρικοί όροι κλάσεων** x_i'	(4) $\nu_i x_i'$
[0-10)	4	5	20
[10-20)	3	15	45
[20-30)	4	25	100
[30-40)	7	35	245
[40-50)	8	45	360
[50-60)	6	55	330
[60-70)	4	65	260
[70-80)	4	75	300
Σύνολο	ν = 40		$\Sigma\nu x' = 1660$

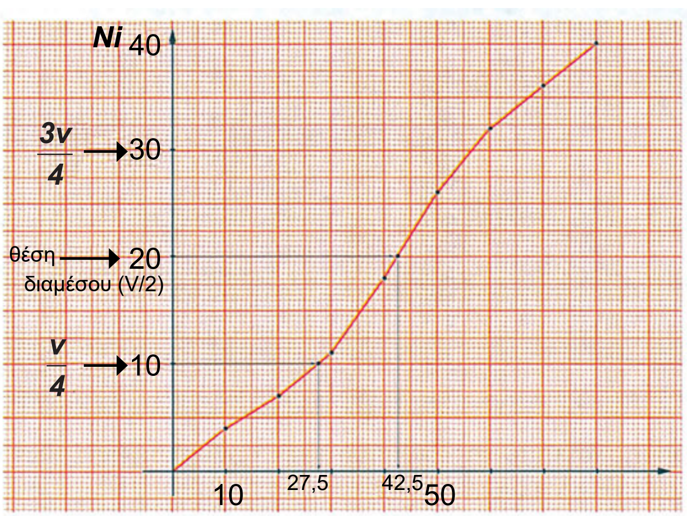
δ) Για τον υπολογισμό της διαμέσου και των τεταρτημορίων κατασκευάζουμε τον αριθμητικό πίνακα 4.11:

Πίνακας 4.11

(1) Τάξεις οφειλών	(2) Αριθμός προμηθευτών ν_i	(3) Αθροιστικές συχνότητες N_i	
[0-10)	4	4	
[10-20)	3	7	$\frac{\nu}{4} = 10$
[20-30)	4	11	
[30-40)	7	18	$\frac{\nu}{2} = 20$
[40-50)	8	26	$\frac{3\nu}{4} = 30$
[50-60)	6	32	
[60-70)	4	36	
[70-80)	4	40	
Σύνολο	ν = 40		

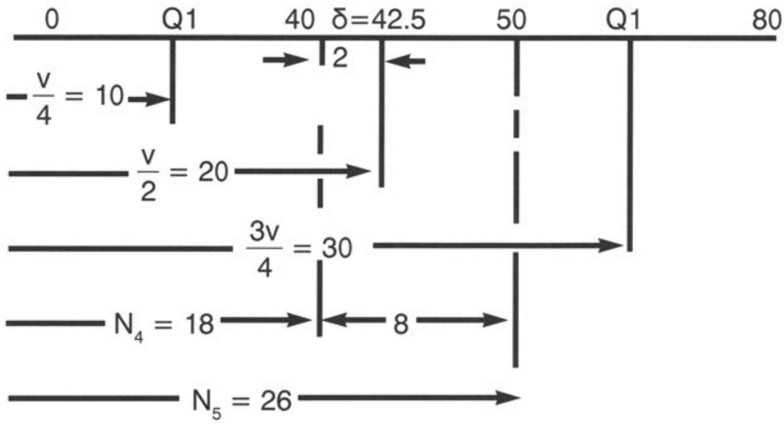
Από τη στήλη (3) των αθροιστικών συχνοτήτων βρίσκουμε ότι ο αριθμός $\frac{\nu}{2} = \frac{40}{2} = 20$ βρίσκεται μεταξύ των αθροιστικών συχνοτήτων 18 και 26 και συνεπώς η διάμεσος περιέχεται στην τάξη [40-50).

Γραφικά ο υπολογισμός της διαμέσου γίνεται ως εξής:
Κατασκευάζουμε το διάγραμμα αθροιστικών συχνοτήτων.



(Σχ.4.6)

Η διάμεσος θα είναι η τιμή της 20ής παρατήρησης. Στο διάγραμμα των αθροιστικών συχνοτήτων βρίσκουμε το σημείο με συντεταγμένες (δ, 20) και παρατηρούμε ότι η τετμημένη του σημείου, που θα είναι και η τιμή της διαμέσου, είναι προσεγγιστικά $\delta \cong 42{,}5$ χιλιάδες ευρώ.



(Σχ. 4.7)

Η διάμεσος θα μπορούσε να υπολογιστεί επίσης με την απλή μέθοδο των τριών.

Σε πλάτος c = 10 κατανέμονται ομοιόμορφα 8 παρατηρήσεις, οι οποίες περιέχονται στην κλάση [40-50), άρα η κάθε μια παρατήρηση καταλαμβάνει πλάτος ίσο με $\frac{10}{8}$ του συνολικού πλάτους της κλάσης.

Η διάμεσος που είναι η 20η παρατήρηση της συνολικής κατανομής, αλλά η (20η – 18η =) 2η παρατήρηση της συγκεκριμένης κλάσης, θα βρίσκεται σε πλάτος ίσο με $2 \cdot \frac{10}{8}$ αρχίζοντας από τη 19η παρατήρηση, η οποία έχει την τιμή 40.

Άρα $\delta - 40 = 2 \cdot \frac{10}{8}$

$\delta = 40 + 2{,}5$

$\delta = 42{,}5$ χιλιάδες ευρώ.

Ο υπολογισμός μπορεί επίσης να γίνει με χρήση του τύπου της διαμέσου:

$$\delta = \alpha_{i-1} + \frac{\frac{v}{2} - N_{i-1}}{v_i} \cdot c$$

όπου $\alpha_{i-1} = 40$

$\frac{v}{2} = 20$

$N_{i-1} = 18$

$N_i = 8$

$c = 10$

αντικαθιστώντας τον τύπο προκύπτει ότι:

$$\delta = 40 + \frac{20 - 18}{8} \cdot 10$$

$$\delta = 40 + \frac{2}{8} \cdot 10$$

$\delta = 40 + 2{,}5$

$\delta = 42{,}5$ χιλιάδες ευρώ.

- Υπολογισμός τεταρτημορίων.

Εργαζόμαστε όπως και προηγουμένως για τον υπολογισμό της διαμέσου. Από το διάγραμμα των αθροιστικών συχνοτήτων βρίσκουμε τις συντεταγμένες των σημείων:

$\left(Q_1, \frac{v}{4}\right)$ και $\left(Q_3, \frac{3v}{4}\right)$ δηλαδή (Q_1, 10) και (Q_3, 30), οπότε οι τιμές του Q_1 και του Q_3 είναι: $Q_1 = 27{,}5$ και $Q_3 = 56{,}7$ με προσέγγιση δεκάτου.

Μπορούμε να υπολογίσουμε τα τεταρτημόρια χρησιμοποιώντας τους τύπους:

$$Q_1 = \alpha_{i-1} + \frac{\frac{v}{4} - N_{i-1}}{v_i} \cdot c \qquad \text{και} \qquad Q_3 = \alpha_{i-1} + \frac{\frac{3v}{4} - N_{i-1}}{v_i} \cdot c$$

όπου $\alpha_{i-1} = 20$ και $\alpha_{i-1} = 50$

$\frac{v}{4}=10$ | $\frac{3v}{4}=30$

$N_{i-1} = 7$ | $N_{i-1} = 26$

$v_i = 4$ | $v_i = 6$

και $c = 10$

άρα

$$Q_1 = 20 + \frac{10-7}{4}\cdot 10 = 20 + \frac{3}{4}\cdot 10 = 20 + 7{,}5 = 27{,}5$$

και

$$Q_3 = 50 + \frac{30-26}{6}\cdot 10 = 50 + \frac{4}{6}\cdot 10 \cong 56{,}7$$

Από τον υπολογισμό των παραπάνω μέτρων θέσης παρατηρούμε ότι η μέση τιμή των οφειλών των 40 επιχειρήσεων στους προμηθευτές τους είναι 41.500 ευρώ. Το 50% των επιχειρήσεων οφείλει μέχρι και 42.500 ευρώ στους προμηθευτές. Το 25% των επιχειρήσεων οφείλει μέχρι και 27.500 ευρώ ενώ υπάρχει και ένα 25% των επιχειρήσεων που οφείλει περισσότερο από 56.700 ευρώ περίπου.

ΑΝΑΚΕΦΑΛΑΙΩΣΗ

1. Τα **μέτρα θέσης** είναι στατιστικές παράμετροι, που καθορίζουν μια τιμή της μεταβλητής που εξετάζουμε, γύρω από την οποία τείνουν να συγκεντρωθούν οι παρατηρήσεις μας.

2. Η **μέση τιμή** είναι το πιο σημαντικό και συχνά χρησιμοποιούμενο στην πράξη στατιστικό μέτρο και ορίζεται ως το πηλίκο του αθροίσματος των τιμών x_1, x_2, …, x_ν που παίρνει μια μεταβλητή Χ προς το πλήθος ν των τιμών της.

3. Το **μέσο εύρος** ορίζεται ως εξής:

$$\text{Μέσο εύρος} = \frac{\textit{Μεγαλύτερη τιμή} + \textit{Μικρότερη τιμή}}{2}$$

4. Στην περίπτωση που έχουμε αταξινόμητα δεδομένα, η μέση τιμή υπολογίζεται από τον τύπο:

$$\bar{x} = \frac{\sum x}{\nu}$$

5. Για τον υπολογισμό της μέσης τιμής από πίνακα συχνοτήτων, χρησιμοποιούμε τους τύπους:

$$\bar{x} = \frac{\sum \nu x}{\nu}$$

ή

$$\bar{x} = \frac{\sum \nu x'}{\nu}$$

όπου x΄ η κεντρική τιμή κάθε κλάσης σε ομαδοποιημένα δεδομένα.

6. Όταν οι τιμές που παίρνει η μεταβλητή τοποθετηθούν κατά αύξουσα τάξη, δηλαδή από τη μικρότερη προς τη μεγαλύτερη, τότε η τιμή της μεταβλητής που χωρίζει το σύνολο των τιμών σε δύο ισοπληθείς ομάδες, ονομάζεται **διάμεσος τιμή ή απλά διάμεσος (δ).**

7. Ο υπολογισμός της διαμέσου μπορεί να γίνει:
α) Γραφικά από το διάγραμμα αθροιστικών συχνοτήτων.
β) Εμπειρικά (μέθοδος των τριών).
γ) Με τον τύπο

$$\delta = \alpha_{i-1} + \left(\frac{\nu}{2} - N_{i-1}\right) \cdot \frac{c}{\nu_i}$$

8. Οι τιμές εκείνες της μεταβλητής που χωρίζουν το συνολικό αριθμό των τιμών της σε 4 ισοπληθείς ομάδες λέγονται τεταρτημόρια. Το δεύτερο τεταρτημόριο Q_2 είναι η διάμεσος, δηλαδή $Q_2 = \delta$.

9. Ο υπολογισμός των τεταρτημορίων από πίνακα συχνοτήτων γίνεται όπως και στη διάμεσο, δηλαδή γραφικά, εμπειρικά (μέθοδος των τριών) και με τους τύπους:

$$Q_1 = \alpha_{i-1} + \frac{\frac{\nu}{4} - N_{i-1}}{\nu_i} \cdot c$$

$$Q_3 = \alpha_{i-1} + \frac{\frac{3\nu}{4} - N_{i-1}}{\nu_i} \cdot c$$

10. Η τιμή της μεταβλητής η οποία αντιστοιχεί στη μεγαλύτερη συχνότητα της εξεταζόμενης κατανομής, ονομάζεται επικρατούσα τιμή.

ΕΡΩΤΗΣΕΙΣ ΚΑΤΑΝΟΗΣΗΣ

1. Τα μέτρα θέσης είναι στατιστικές παράμετροι, οι οποίες καθορίζουν ……

....

2. Ως μέση τιμή ορίζεται ....

....

.... και υπολογίζεται

α) σε αταξινόμητα δεδομένα ως $\overline{x}$ …………………………………………

β) σε ταξινομημένα δεδομένα ως $\overline{x}$ …………………………………………

3. α) Αν σε όλες τις τιμές μίας μεταβλητής αφαιρέσουμε τον αριθμό 2, τότε η τιμή της μέσης τιμής

☐ αυξάνεται κατά 2 μονάδες

☐ μειώνεται κατά 2 μονάδες

☐ παραμένει η ίδια.

β) Αν όλες οι τιμές μίας μεταβλητής πολλαπλασιασθούν επί 2, τότε η τιμή της μέσης τιμής:

☐ αυξάνεται κατά 2 μονάδες.

☐ διπλασιάζεται.

☐ παραμένει η ίδια.

4) Αναφέρατε τις ιδιότητες της μέσης τιμής.

5) Η διάμεσος ορίζεται ως ....

....

Αν το πλήθος των παρατηρήσεων είναι περιττό, τότε η τιμή της διαμέσου ....

ή αν το πλήθος των παρατηρήσεων είναι άρτιο ....
....

6) Η θέση της διαμέσου σε ταξινομημένα διακριτά δεδομένα είναι ο αριθμός ...., ενώ σε ομαδοποιημένα δεδομένα βρίσκεται στην ....

7) Η επικρατούσα τιμή ορίζεται ως ....
....

8) Τεταρτημόρια λέγονται οι τιμές της μεταβλητής ....
....

9) Η μέση τιμή ν αριθμών ισούται με 5. Αν προσθέσουμε τον αριθμό 13 στο άθροισμα των ν αριθμών, η νέα μέση τιμή ισούται με 6. Να βρεθεί το πλήθος των αριθμών.

10) Να βρεθεί για κάθε μία από τις παρακάτω κατανομές, ποιο είναι το καταλληλότερο μέτρο κεντρικής τάσης:

Κατανομές	**Μέτρο θέσης**
(α) Μισθοί όλων των καθηγητών του σχολείου σας	(1) Η Μέση τιμή
(β) Οι χρόνοι, στους οποίους 10 άλογα διανύουν μία δεδομένη απόσταση.	(2) Διάμεσος
(γ) Την ώρα, με ακρίβεια λεπτού, όπως τη δείχνουν 10 ρολόγια.	
(δ) Οι βαθμοί του τμήματός σας στο πρώτο διαγώνισμα Στατιστικής.	(3) Επικρατούσα τιμή

Να εξηγήσετε την επιλογή σας.

11. Αν η μέση τιμή των αριθμών 12, 18, 21, x, 13 είναι 17, τότε η τιμή του x είναι:

Α. 24 ☐ Β. 21 ☐ Γ. 17 ☐ Δ. 15 ☐

ΑΣΚΗΣΕΙΣ

Ομάδα Α.

1. Οι ηλικίες των υπαλλήλων ενός εμπορικού καταστήματος είναι:

25, 22, 20, 20, 23, 25, 25, 30, 35, 40.

Να βρεθούν:

α) Η μέση τιμή

β) η διάμεσος ηλικία και

γ) το πρώτο και τρίτο τεταρτημόριο.

2. Οι μηνιαίες αποδοχές των υπαλλήλων μιας επιχείρησης (σε δεκάδες ευρώ) είναι:

210, 220, 200, 180, 150, 160, 180, 200, 250, 260.

Να βρεθεί:

α) Η μέση τιμή

β) η διάμεσος του μηνιαίου μισθού

γ) το πρώτο και τρίτο τεταρτημόριο.

3. Η μέση τιμή του μηνιαίου μισθού των υπαλλήλων μιας επιχείρησης είναι 1100 ευρώ. α) Αν ο μισθός κάθε υπαλλήλου αυξηθεί κατά 100 ευρώ, ποια μεταβολή θα επέλθει στη μέση τιμή του; β) Αν αυξηθούν οι μισθοί όλων των εργαζομένων κατά 10%, ποια μεταβολή θα επέλθει στη μέση τιμή του;

4. Ο παρακάτω πίνακας περιέχει τον αριθμό ωρών υπερωριακής απασχόλησης που πραγματοποιήθηκε κατά τη διάρκεια μιας ορισμένης χρονικής περιόδου από τους εργάτες μιας βιομηχανίας:

Ώρες(x_i)	0	1	2	3	4	5	6	7	8	9	10
Εργάτες(v_i)	5	5	10	15	15	20	15	10	10	10	5

Να βρεθούν και να ερμηνευθούν: α) Η μέση τιμή, β) η διάμεσος, γ) το πρώτο τεταρτημόριο.

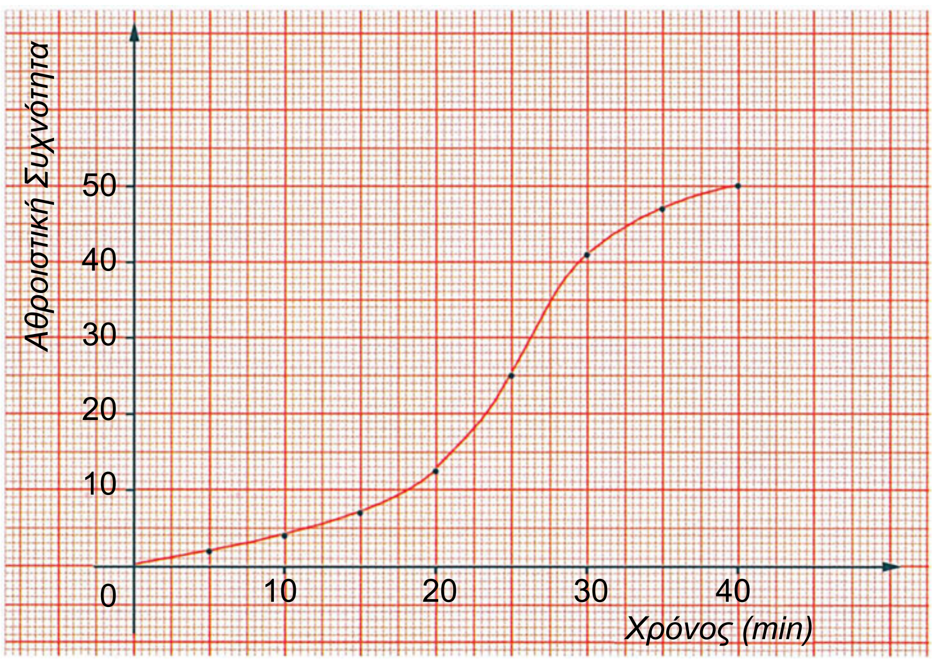
5. Ο παρακάτω πίνακας δείχνει τα ετήσια έξοδα πληρωμής λογαριασμών κατανάλωσης ηλεκτρικού ρεύματος (σε ευρώ) 100 νοικοκυριών.

Έξοδα	Αριθμός νοικοκυριών
40-60	5
60-80	10
80-100	10
100-120	15
120-140	20
140-160	20
160-180	10
180-200	5
200-220	5

Να υπολογιστούν και να ερμηνευθούν: α) Η μέση τιμή και β) η διάμεσος.

Ομάδα Β

1. Από το παρακάτω διάγραμμα αθροιστικής συχνότητας παρατηρούμε το χρόνο που πέρασαν 50 άτομα μέσα σε ένα πολυκατάστημα:



α) Από το διάγραμμα να βρείτε πόσα άτομα έμειναν στο πολυκατάστημα από 17 έως και 27 λεπτά.

β) Το 25% των ατόμων έμεινε λιγότερο ή ίσο με t λεπτά. Να βρείτε το t.

γ) Το 50% των ατόμων πέρασε περισσότερο από s λεπτά. Να βρείτε το s.

δ) Υπολογίστε τη διάμεσο. Τι παρατηρείτε;

2. Ο παρακάτω πίνακας δείχνει τους βαθμούς 30 μαθητών στο μάθημα της Στατιστικής της Α΄ τάξης των Τ.Ε.Ε.

Βαθμοί	Αριθμός μαθητών
7	1
8	2
11	3
12	1
13	2
14	5
15	7
16	5
17	3
18	1

α) Να βρείτε τη μέση τιμή, τη διάμεσο, το Q_1 και το Q_3. Τι παρατηρείτε για τη βαθμολογία του τμήματος;

β) Αν η βαθμολογία όλων των μαθητών αυξανόταν κατά δύο μονάδες, πόση θα γινόταν η μέση τιμή;

3. Από τους 10 παίκτες μιας ομάδας μπάσκετ οι 4 είναι Έλληνες και η μέση τιμή του ύψους τους είναι 200 εκ., οι 4 είναι κοινοτικοί και η μέση τιμή του ύψους τους είναι 208 εκ. και οι άλλοι 2 είναι ξένοι και η μέση τιμή του ύψους τους είναι 204 εκ.

α) Ποια είναι η μέση τιμή του ύψους των παικτών της ομάδας;

β) Κατά τη διάρκεια της αγωνιστικής περιόδου ο ένας ξένος παίκτης που είχε ύψος 200 εκ. πήρε μεταγραφή σε άλλη ομάδα και ο προπονητής ήθελε στη θέση του να πάρει έναν Έλληνα. Πόσο πρέπει να είναι το ύψος του Έλληνα παίκτη, ώστε η μέση τιμή του ύψους της ομάδας να είναι 205 εκ.;

4. Δώδεκα μαθητές μιας τάξης εκτίμησαν το ύψος ενός καμπαναριού μιας εκκλησίας. Οι εκτιμήσεις τους σε μέτρα είναι:

47, 52, 52, 54, 52, 50, 51, 50, 48, 53, 54, 49.

α) Υπολογίστε τη διάμεσο των εκτιμήσεων.

β) Υπολογίστε την επικρατούσα τιμή.

γ) Υπολογίστε τη μέση τιμή ($\bar{x}$), των 12 εκτιμήσεων.

δ) Ένας μαθητής από τους δώδεκα αναθεώρησε την εκτίμησή του και η νέα μέση τιμή των 12 εκτιμήσεων είναι $\bar{x} + 0{,}5$. Βρείτε κατά πόσο αυξήθηκε η εκτίμηση αυτού του μαθητή.

ε) Ο καθηγητής της τάξης έκανε και αυτός μία εκτίμηση για το ύψος του καμπαναριού και όταν η εκτίμησή του προστέθηκε στις αρχικές 12, η μέση τιμή των 13 πια εκτιμήσεων είναι $\bar{x} + 0{,}5$. Βρείτε την εκτίμηση του καθηγητή.

5. Οι μηνιαίες αποδοχές 100 υπαλλήλων μιας επιχείρησης κυμαίνονται μεταξύ 1400 και 2000 ευρώ. Ακόμη γνωρίζουμε ότι:

10 υπάλληλοι αμείβονται με 1500 ευρώ και κάτω.

30 υπάλληλοι αμείβονται με 1600 ευρώ και κάτω.

40 υπάλληλοι έχουν μισθούς μεγαλύτερους από 1700 ευρώ και

10 υπάλληλοι έχουν μισθούς μεγαλύτερους από 1800 ευρώ.

Ζητείται: α) Να υπολογιστεί η μέση τιμή, β) να υπολογιστεί η διάμεσος και γ) να υπολογιστεί το πρώτο και το τρίτο τεταρτημόριο.

ΚΕΦΑΛΑΙΟ 5ο

ΤΑ ΒΑΣΙΚΑ ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ ΜΙΑΣ ΚΑΤΑΝΟΜΗΣ

❖ *«...Η έρευνα των αιτιωδών σχέσεων μεταξύ των οικονομικών φαινομένων παρουσιάζει αρκετά προβλήματα ιδιαίτερης δυσκολίας, αφού ο στατιστικός μπορεί σπάνια ή ποτέ να κάνει πειράματα μόνος του. Αυτός πρέπει να δεχθεί τα δεδομένα της καθημερινής εμπειρίας …»* ***G. U. Yule*** *(1871-1951)*

❖ «…Η πολιτική οικονομία στο μέλλον πρέπει να αποβεί ποσοτική επιστήμη, χρησιμοποιώντας ευρύτατα την τεχνική της Στατιστικής, για να μελετήσει τα ποσοτικά δεδομένα και να αναζητήσει τις σχέσεις και τις αναλογίες που υπάρχουν μεταξύ τους….» **A. Aftalion**

ΚΕΦΑΛΑΙΟ 5ο

ΤΑ ΒΑΣΙΚΑ ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ ΜΙΑΣ ΚΑΤΑΝΟΜΗΣ

5.1 Γενικά

Για να γίνει κατανοητή η αναγκαιότητα της χρήσης των μέτρων διασποράς, ας μελετήσουμε το παράδειγμα που ακολουθεί.

Παράδειγμα

Ο ιδιοκτήτης ενός εργοστασίου συσκευασίας πλαστικών σφαιριδίων για B-B gun (είδος αεροβόλου πιστολιού) σκέφτεται να αντικαταστήσει μια από τις μηχανές συσκευασίας (Α) που θεωρείται παλαιού τύπου με μια μηχανή (Β) νέας τεχνολογίας. Πριν αγοράσει την καινούργια μηχανή θέλει να ελέγξει την αποτελεσματικότητά της. Για το λόγο αυτό διαλέχτηκαν 10 πακέτα σφαιριδίων από την κάθε μηχανή και μετρήθηκαν τα αντίστοιχα βάρη (σε p)*:

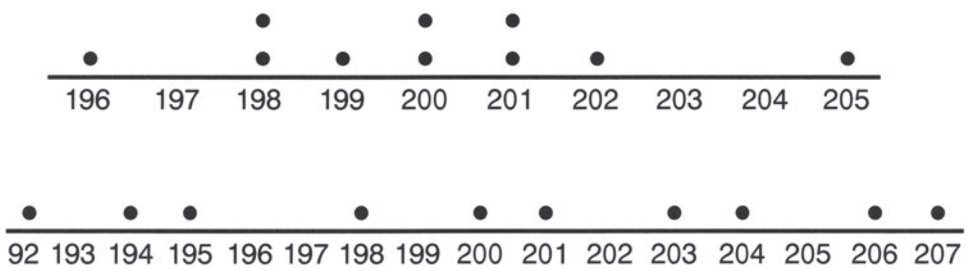
Πίνακας 5.1

Μηχανή Α	196	198	198	199	200	200	201	201	202	205
Μηχανή Β	192	194	195	198	200	201	203	204	206	207

* Pond = μονάδα μέτρησης βάρους του Τεχνικού Συστήματος, που έχει πλέον αντικατασταθεί από το 1N (1 Newton): 1N > 9810p.

Θα απεικονίσουμε τα δεδομένα με ένα σημειόγραμμα:

Βάρη (σε p)*



(Σχ. 5.1)

Αν υπολογίσουμε τη μέση τιμή των δύο ομάδων:

$$\bar{x}_A = \frac{\sum \nu x}{\nu} = \frac{1\cdot 196 + 2\cdot 198 + 1\cdot 199 + 2\cdot 200 + 2\cdot 201 + 1\cdot 202 + 1\cdot 205}{10} = 200$$

και

$$\bar{x}_B = \frac{\sum \nu y}{\nu} = \frac{1\cdot 192 + 194 + 195 + 198 + 200 + 201 + 203 + 204 + 206 + 207}{10} = 200$$

παρατηρούμε ότι και στις δυο ομάδες η μέση τιμή είναι ίδια. Από το σημειόγραμμα όμως βλέπουμε ότι τα βάρη των πακέτων των πλαστικών σφαιριδίων που συσκεύασε η πρώτη μηχανή κυμαίνονται από 196 έως 205 p, είναι δηλαδή συγκεντρωμένα γύρω από τη μέση τιμή, ενώ τα βάρη της Β μηχανής τα οποία κυμαίνονται από 192 έως 208 p, είναι περισσότερο διασπαρμένα.

Είναι λοιπόν φανερό ότι δεν μπορούμε να καταλήξουμε σε αξιόπιστα συμπεράσματα για τα δεδομένα μας και να τα συγκρίνουμε με άλλα ομοειδή δεδομένα γνωρίζοντας μόνο τη μέση τιμή, αλλά είναι απαραίτητο να χρησιμοποιήσουμε και άλλες παραμέτρους που να φανερώνουν τον τρόπο διασποράς των τιμών της μεταβλητής. **Αυτές οι παράμετροι ονομάζονται μέτρα διασποράς (measures of dispersion) και δείχνουν το πόσο είναι «απλωμένα» τα δεδομένα γύρω από κάποιο μέτρο θέσης.**

Τα μέτρα διασποράς που θα μελετήσουμε είναι:

- Το εύρος μεταβολής
- Η μέση απόλυτη απόκλιση
- Η διακύμανση
- Η τυπική απόκλιση

5.2 Εύρος μεταβολής

Το εύρος μεταβολής (range) είναι το απλούστερο μέτρο διασποράς. Είναι η διαφορά της μικρότερης τιμής από τη μεγαλύτερη τιμή της μεταβλητής Χ, δηλαδή:

R = μεγαλύτερη τιμή της Χ – μικρότερη τιμή της Χ

Το εύρος μεταβολής των τιμών της μεταβλητής υπολογίζεται ευκολότερα, όταν οι τιμές έχουν διαταχθεί σε αύξουσα τάξη. Στο παράδειγμα της (§5.1) το εύρος της μεταβλητής «βάρος πακέτου» της μηχανής Α είναι $R_A = 205 - 196 = 9$ ενώ της μηχανής Β είναι $R_B = 208 - 192 = 16$.

Το μειονέκτημα του εύρους μεταβολής είναι ότι για τον υπολογισμό του λαμβάνονται υπόψη μόνο οι δυο ακραίες τιμές της μεταβλητής. Παρόλα αυτά χρησιμοποιείται από τους οικονομολόγους σαν μέτρο διασποράς στο χρηματιστήριο για τις τιμές των μετοχών, στη μελέτη της μεταβολής των μισθών κ.λπ.

5.3 Μέση απόλυτη απόκλιση

Στο παράδειγμα των μηχανών συσκευασίας είδαμε ότι το εύρος μεταβολής των τιμών της μεταβλητής «βάρος πακέτου» στη μηχανή νέας τεχνολογίας είναι 16, ενώ το εύρος μεταβολής των αντίστοιχων τιμών στη μηχανή παλαιού τύπου είναι μόνο 9.

Ας δούμε τώρα πόσο διασπαρμένες είναι οι τιμές της μεταβλητής «Βάρος πακέτου» γύρω από τη μέση τιμή στην περίπτωση της μηχανής Β.

192 – 200 = - 8	*200 – 200 = 0*	*201 – 200 = 1*
194 – 200 = - 6		*203 – 200 = 3*
195 – 200 = - 5		*204 – 200 = 4*
198 – 200 = - 2		*206 – 200 = 6*
		207 – 200 = 7

Αν υπολογίσουμε το μέσο αριθμητικό των αποκλίσεων των τιμών της μεταβλητής από τη μέση τιμή $\overline{x}$ θα βρούμε:

$$\frac{-8-6-5-2+0+1+3+4+6+7}{10}=\frac{-20+20}{10}=0$$

Επειδή λοιπόν πάντα το άθροισμα $\sum\left(x_i-\overline{x}\right)=0$ χρησιμοποιούμε τις απόλυτες τιμές των αποκλίσεων για να καταλήξουμε σε έναν αντιπροσωπευτικό αριθμό που ονομάζεται μέση απόλυτη απόκλιση (Μ.Α.Α.) και ισούται με:

$$\text{Μ.Α.Α.}=\frac{\sum\left|x_i-\overline{x}\right|}{\nu}$$

Για τη μηχανή Β η μέση τιμή των απόλυτων τιμών των αποκλίσεων είναι:

$$\frac{8+6+5+2+0+1+3+4+6+7}{10}=\frac{42}{10}=4,2$$

αριθμός που μας δείχνει πόσο, κατά μέσο όρο, απέχουν οι παρατηρήσεις μας από τη μέση τιμή.

Είναι φανερό ότι όσο μικρότερος είναι ο αριθμός αυτός τόσο περισσότερο είναι συγκεντρωμένες οι παρατηρήσεις γύρω από τη μέση τιμή που είναι πλέον μια αντιπροσωπευτική και αξιόπιστη παράμετρος των τιμών της κατανομής.

Θα μπορούσαμε λοιπόν να θεωρήσουμε τη μέση απόλυτη απόκλιση ή μέση απόκλιση (mean deviation) ως ένα μέτρο διασποράς το οποίο ορίζεται ως εξής:

Αν μια μεταβλητή Χ παίρνει ν τιμές: $x_1, x_2, x_3, \ldots x_\nu$, που έχουν μέση τιμή $\bar{x}$, και σχηματίσουμε τις διαφορές $x_1 - \bar{x}$, $x_2 - \bar{x}$,…$x_\nu - \bar{x}$,τότε ως **μέση απόλυτη απόκλιση της μεταβλητής Χ** ορίζεται η μέση τιμή των απολύτων τιμών των διαφορών αυτών, δηλαδή:

$$\text{M.A.A.} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \ldots + |x_\nu - \bar{x}|}{\nu} = \frac{\sum_{i=1}^{\nu} |x_i - \bar{x}|}{\nu} = \frac{\sum |x - \bar{x}|}{\nu}$$

(5.2)

Υπολογίζοντας τη μέση απόλυτη απόκλιση των τιμών της μεταβλητής «βάρος πακέτου» για τη μηχανή Α βρίσκουμε ότι:

$$\text{M.A.A.} = \frac{1\cdot|196-200| + 2\cdot|198-200| + 1\cdot|199-200| + 2\cdot|200-200| + 2\cdot|201-200| + 1\cdot|202-200| + 1\cdot|205-200|}{10}$$

$$= \frac{1\cdot 4 + 2\cdot 2 + 1\cdot 1 + 2\cdot 0 + 2\cdot 1 + 1\cdot 2 + 1\cdot 5}{10} = \frac{4+4+1+2+2+5}{10} = \frac{18}{10} = 1{,}8$$

Άρα συμπεραίνουμε ότι η μηχανή Α είναι πιο αξιόπιστη από τη μηχανή νέας τεχνολογίας , γιατί 4,2 > 1,8.

- Στις ομαδοποιημένες παρατηρήσεις ακολουθούμε παρόμοια διαδικασία μόνο που ως x_i παίρνουμε τις κεντρικές τιμές των κλάσεων.

Επειδή η μέση απόλυτη απόκλιση παρουσιάζει κάποια δυσκολία στους αλγεβρικούς υπολογισμούς λόγω των απόλυτων τιμών, χρησιμοποιούμε συνήθως άλλα μέτρα διασποράς με τα οποία θα ασχοληθούμε στις επόμενες παραγράφους.

5.4 Διακύμανση και τυπική απόκλιση

Τα μέτρα διασποράς που χρησιμοποιούμε συνήθως είναι η διακύμανση και η τυπική απόκλιση (variance and standard deviation). Για κάθε τιμή της μεταβλητής Χ υπολογίζεται η απόκλιση από τη μέση τιμή, όπως κάναμε στο προηγούμενο παράδειγμα. Ο βαθμός συγκέντρωσης όμως των τιμών x_1, x_2, x_3, …. x_ν μιας μεταβλητής Χ γύρω από τη μέση τιμή $\bar{x}$ προσδιορίζεται ευκολότερα με τη βοήθεια των θετικών αριθμών $(x_i - \bar{x})^2$.

Διακύμανση των τιμών της μεταβλητής Χ ορίζουμε τη μέση τιμή των τετραγώνων των αποκλίσεων των τιμών της μεταβλητής από το μέσο αριθμητικό της ($\bar{x}$). Τη συμβολίζουμε με s^2 και δίνεται από τον τύπο:

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_\nu - \bar{x})^2}{\nu} = \frac{\sum_{i=1}^{\nu}(x_i - \bar{x})^2}{\nu} = \frac{\sum(x - \bar{x})^2}{\nu}$$

(5.3)

Επειδή οι μονάδες στις οποίες εκφράζεται η διακύμανση είναι τα τετράγωνα των μονάδων στις οποίες εκφράζονται οι τιμές της μεταβλητής, χρησιμοποιούμε αντί της διακύμανσης τη θετική τετραγωνική της ρίζα που λέγεται τυπική απόκλιση και έχει τις ίδιες μονάδες μέτρησης με τις τιμές της μεταβλητής.

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{\nu}}$$

- Ο παραπάνω τύπος χρησιμοποιείται, όταν

1. Οι τιμές x_1, x_2, ... x_ν αποτελούν συνολικό πληθυσμό.
2. Οι τιμές x_1, x_2, ... x_ν αποτελούν δείγμα από πληθυσμό και ζητάμε να υπολογίσουμε τη διασπορά μέσα στο ίδιο το δείγμα.

- Όσο μικρότερες είναι οι τιμές της διακύμανσης και της τυπικής απόκλισης τόσο πιο συγκεντρωμένες γύρω από τη μέση τιμή βρίσκονται οι τιμές της μεταβλητής Χ, με αποτέλεσμα η μέση τιμή ($\bar{x}$) να αποτελεί αντιπροσωπευτικό στατιστικό μέτρο για την κατανομή της μεταβλητής.

5.5 Υπολογισμός της διακύμανσης από αταξινόμητα δεδομένα

Ο υπολογισμός της διακύμανσης και της τυπικής απόκλισης από αταξινόμητα δεδομένα γίνεται με τη βοήθεια της σχέσης:

$$s^2 = \frac{\sum(x-\bar{x})^2}{\nu} \quad (5.4)$$

Θα υπολογίσουμε τη διακύμανση και την τυπική απόκλιση στο παράδειγμα των μηχανών συσκευασίας. Έχουμε ήδη υπολογίσει ότι $\bar{x}$ = 200 και στις δυο περιπτώσεις και συμπληρώνουμε τον πίνακα 5.1 με τις στήλες των αποκλίσεων των τιμών της μεταβλητής από τη μέση τιμή $\bar{x}$, καθώς και με τις στήλες των τετραγώνων των αποκλίσεων αυτών.

Πίνακας 5.2

ΜΗΧΑΝΗ Α			**ΜΗΧΑΝΗ Β**		
Βάρος x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	Βάρος x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
196	–4	16	192	–8	64
198	–2	4	194	–6	36
198	–2	4	195	–5	25
199	–1	1	198	–2	4
200	0	0	250	0	0
200	0	0	201	1	1
201	1	1	203	3	9
201	1	1	204	4	16
202	2	4	206	6	36
205	5	25	207	7	49
Σύνολο		56			240

Άρα $s_A^2 = \dfrac{\sum(x-\overline{x})^2}{v} = \dfrac{56}{10} = 5{,}6$

$$s_B^2 = \frac{\sum(x-\overline{x})^2}{v} = \frac{240}{10} = 24$$

και αντίστοιχα

$$s_A = \sqrt{5{,}6} \cong 2{,}4$$

$$s_B = \sqrt{24} \cong 4{,}9$$

Παρατηρούμε ότι $s_B > s_A$ γεγονός που ενισχύει τον ισχυρισμό μας ότι η μηχανή Α είναι πιο αξιόπιστη.

Αξίζει να αναφέρουμε ότι για τον υπολογισμό της διακύμανσης μπορούμε να χρησιμοποιήσουμε έναν πιο εύχρηστο τύπο, δηλαδή:

$$\boxed{s^2 = \frac{\sum x^2}{v} - \overline{x}^2}$$

(5.5)

όπου ν το πλήθος των τιμών της μεταβλητής, που μας λέει ότι **«η διακύμανση ισούται με τη μέση τιμή των τετραγώνων των τιμών της μεταβλητής μείον το τετράγωνο της μέσης τιμής».**

5.6 Υπολογισμός της διακύμανσης από ταξινομημένα δεδομένα

Σε περίπτωση που τα δεδομένα είναι ταξινομημένα σε πίνακα κατανομής συχνοτήτων, υπολογίζουμε τη διακύμανση ακολουθώντας τη διαδικασία του παρακάτω παραδείγματος:

Παράδειγμα 5.2

Ο πίνακας 5.3 παρουσιάζει τις επιδόσεις που πέτυχαν οι 50 αγωνιζόμενοι σε αγώνες τοξοβολίας παίδων (μέγιστη επίδοση 100). Να βρεθεί η μέση τιμή των επιδόσεων που πέτυχαν σε ένα γύρο και η τυπική απόκλιση.

Πίνακας 5.3

Επιδόσεις	Συχνότητα
66	2
67	5
68	10
69	12
70	9
71	6
72	4
73	2
Σύνολο	50

Λύση

Συμπληρώνουμε τον πίνακα 5.3 με τις στήλες που λείπουν.

Πίνακας 5.4

Επιδόσεις x_i	**Συχνότητα** v_i	$v_i x_i$	x_i^2	$v_i x_i^2$
66	2	132	4356	8712
67	5	335	4489	22445
68	10	680	4624	46240
69	12	828	4761	57132
70	9	630	4950	44100
71	6	426	5041	30246
72	4	288	5184	20736
73	2	146	5329	10658
Σύνολο	$v = 50$	$\Sigma vx = 3465$		$\Sigma vx^2 = 240269$

Άρα $\bar{x} = \frac{\sum vx}{v} = \frac{3465}{50} = 69{,}3$

$$s^2 = \frac{\sum vx^2}{v} - \bar{x}^2 = \frac{240269}{50} - (69{,}3)^2 = 4805{,}38 - 4802{,}49 = 2{,}89$$

$s = 1{,}7$

Άρα η μέση τιμή των επιδόσεων στους αγώνες τοξοβολίας είναι 69,3 και η τυπική απόκλιση 1,7.

Καταλήγουμε λοιπόν στο συμπέρασμα ότι: αν x_1, x_2, ... x_κ είναι οι τιμές της μεταβλητής X και v_1, v_2, ... v_κ είναι οι αντίστοιχες συχνότητες, τότε η διακύμανση των δεδομένων δίνεται από τον παρακάτω τύπο:

$$s^2 = \frac{\sum_{i=1}^{\kappa} v_i x_i^2}{\sum_{i=1}^{\kappa} v_i} - \bar{x}^2 = \frac{\sum vx^2}{\sum v} - \bar{x}^2 \quad \text{για } i = 1, 2, \ldots, \kappa$$

(5.6)

- Αν τα δεδομένα είναι ομαδοποιημένα σε κλάσεις ίσου πλάτους, τότε ο υπολογισμός της διακύμανσης γίνεται από τον τύπο

$$s^2 = \frac{\sum_{i=1}^{\kappa'} v_i x'^2_i}{\sum_{i=1}^{\kappa'} v_i} - \bar{x}^2 = \frac{\sum v x'^2}{\sum v} - \bar{x}^2$$

(5.7)

όπου x'_i , είναι η κεντρική τιμή της κάθε κλάσης που αντιπροσωπεύει τις τιμές που περιέχονται στην κλάση.

Σχόλιο: Οι συμβολισμοί s^2 και s, για τη διακύμανση και την τυπική απόκλιση αντίστοιχα, αναφέρονται σε δείγμα του πληθυσμού και όχι σε όλο τον πληθυσμό. Αν αναφερόμαστε σε όλο τον πληθυσμό μεγέθους N, τότε τη διακύμανση τη συμβολίζουμε με σ^2 , την τυπική απόκλιση με σ και έχουμε:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N} \quad \text{με} \quad \mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

Παράδειγμα 5.3:

Ο παρακάτω πίνακας δίνει τον αριθμό ημερών παραμονής 1100 πελατών μιας ξενοδοχειακής Α.Ε. κατά τη διάρκεια μιας ορισμένης χρονικής περιόδου.

Πίνακας 5.5

Ημέρες παραμονής	Αριθμός πελατών
0 – 4	100
4 – 8	150
8 – 12	300
12 – 16	200
16 – 20	150
20 – 24	50
24 – 28	30
28 –32	20
Σύνολο	1100

Να υπολογισθεί η διακύμανση και η τυπική απόκλιση.

Λύση

Συμπληρώνουμε τον πίνακα (5.5) με τις στήλες που λείπουν.

Πίνακας 5.6

Ημέρες Παραμονής	Αριθμός πελατών ν_i	Κέντρο κλάσης x'_i	$\nu_i x'_i$	x'^2_i	$\nu_i x'^2_i$
[0 , 4)	100	2	200	4	400
[4 , 8)	150	6	900	36	5.400
[8 , 12)	300	10	3000	100	30.000
[12 , 16)	200	14	2800	196	39.200
[16 , 20)	150	18	2700	324	48.600
[20 , 24)	50	22	1100	484	24.200
[24 , 28)	30	26	780	676	20.280
[28 , 32)	20	30	600	900	18.000
Σύνολο	ν = 1.000		$\Sigma\nu x'$ = 12.080		$\Sigma\nu x'^2$ = 186.080

Άρα $\bar{x} = \frac{\sum \nu x'}{\nu} = \frac{12.080}{1000} = 12{,}08$

Δηλαδή η μέση χρονική διάρκεια παραμονής των πελατών είναι περίπου 13 ημέρες.

Η διακύμανση θα ισούται με:

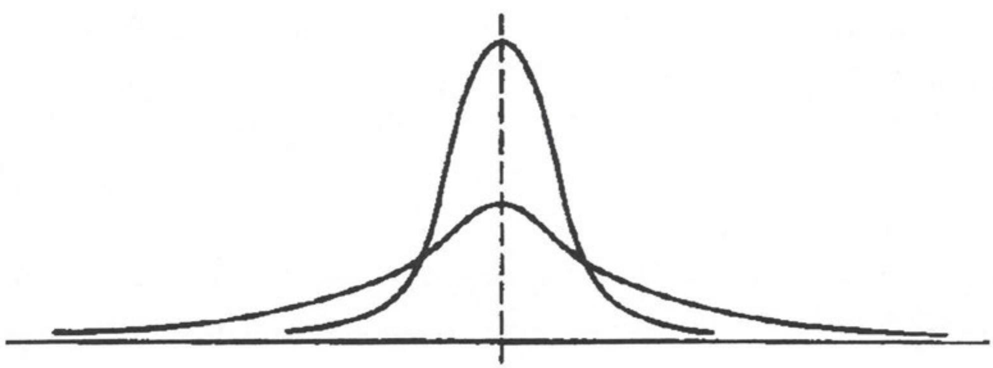
$$s^2 = \frac{\sum \nu x'^2}{\Sigma \nu} - \bar{x}^2 = \frac{186.080}{1000} - (12{,}08)^2 = 40{,}15$$

και η τυπική απόκλιση, που ως γνωστόν είναι η θετική τετραγωνική ρίζα της διακύμανσης, θα ισούται με $s = \sqrt{40{,}15} \cong 6{,}3$

Ας δούμε τη χρησιμότητα της μέσης τιμής και της τυπικής απόκλισης στην πιο γνωστή κατανομή συχνοτήτων, την κανονική κατανομή (βλ. παράγραφο 4.5).

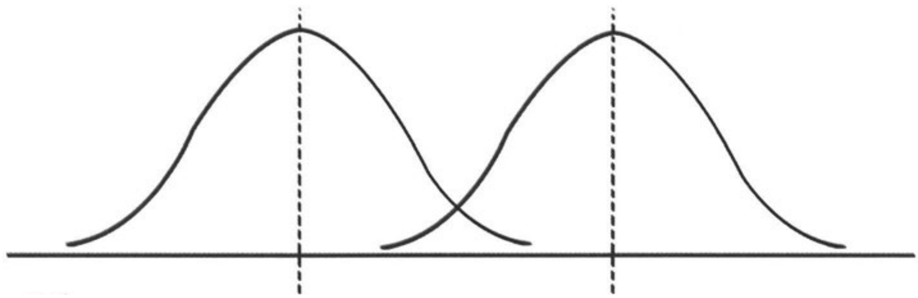
Ως γνωστόν η τυπική απόκλιση μετρά τη διασπορά των τιμών της μεταβλητής γύρω από τη μέση τιμή. Όσο μεγαλύτερη είναι η διασπορά των τιμών τόσο μεγαλύτερη είναι η τυπική απόκλιση. Ας παρατηρήσουμε τα παρακάτω διαγράμματα:

α) Δυο κανονικές κατανομές με την ίδια μέση τιμή, αλλά διαφορετική απόκλιση.



(Σχ. 5.2)

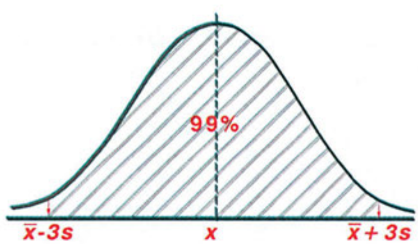
β) Δυο κανονικές κατανομές με διαφορετική μέση τιμή αλλά την ίδια απόκλιση.



(Σχ. 5.3)

Το 68% των μεσαίων τιμών βρίσκονται στο διάστημα $(\overline{x} - s,\ \overline{x} + s)$

(Σχ. 5.4)

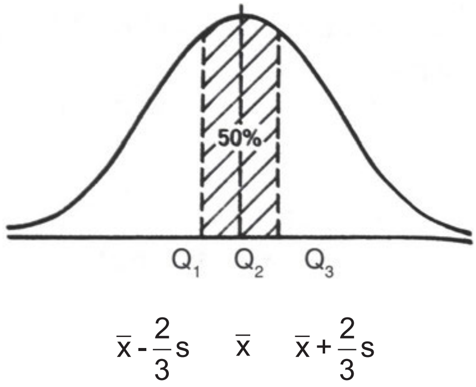


Το 99,7% των μεσαίων τιμών βρίσκονται στο διάστημα $(\overline{x} - 3s,\ \overline{x} + 3s)$

(Σχ. 5.5)

Αξίζει να σημειώσουμε ότι τα τεταρτημόρια Q_1 και Q_3 είναι οι τιμές της μεταβλητής που βρίσκονται σε απόσταση $\frac{2}{3} \cdot s$, αριστερά και δεξιά της μέσης τιμής.

(Σχ. 5.6)



Παράδειγμα 5.4:

Παίρνουμε όλους τους μηνιαίους μισθούς των εργαζομένων σε ένα εργοστάσιο και υπολογίζουμε ότι η μέση τιμή των μισθών είναι $\bar{x}$ =1400 ευρώ, η τυπική απόκλιση s = 240 ευρώ και η κατανομή των μισθών είναι κανονική.

Τότε, σύμφωνα με τα προηγούμενα, μπορούμε να ισχυριστούμε ότι:

i) Το 68,3% των εργαζομένων παίρνει μισθούς 1400 ± 1·240, δηλ. μεταξύ 1160 ευρώ και 1640 ευρώ.

ii) Το 95% των εργαζομένων παίρνει μισθούς 1400 ± 2·240, δηλ. μεταξύ 920 ευρώ και 1880 ευρώ.

iii) Το 99,7% των εργαζομένων παίρνει μισθούς 1400 ± 3·240, δηλ. μεταξύ 680 ευρώ και 2120 ευρώ και

iv) Το 50% των εργαζομένων παίρνει μισθούς 1400 ± 2/3·240, δηλ. μεταξύ 1240 ευρώ και 1560 ευρώ.

5.7 Ιδιότητες της διακύμανσης

Ας θεωρήσουμε ένα σύνολο αριθμών x_1, x_2, …, x_ν με μέση $\bar{x}$ τιμή και τυπική απόκλιση s.

- ***Αν σε κάθε τιμή της μεταβλητής προσθέσουμε ή αφαιρέσουμε ένα σταθερό αριθμό κ, τότε η διακύμανση παραμένει αμετάβλητη.***

$$s^2_y = s^2_x$$

- ***Αν κάθε τιμή της μεταβλητής Χ πολλαπλασιασθεί (ή διαιρεθεί) με ένα σταθερό αριθμό (α), τότε η διακύμανση τιμών της μεταβλητής Υ = αΧ***

($Y = \frac{X}{\alpha}, \alpha \neq 0$) πολλαπλασιάζεται (ή διαιρείται) με το τετράγωνο του αριθμού αυτού, δηλαδή:

$$s^2_y = \alpha^2 s^2_x$$

$$(s^2_y = \frac{s^2_x}{\alpha^2})$$

- ***Αν οι τιμές της μεταβλητής X είναι ίσες μεταξύ τους δηλαδή:***
  $x_1 = x_2 = \ldots = x_\nu = \kappa$, τότε $s^2_x = 0$

Παράδειγμα 5.5

Έστω $x_1, x_2, \ldots, x_\nu$ οι τιμές μιας μεταβλητής X με μέσο $\bar{x} = 6$ και διακύμανση $s^2_x = 4$ και $y_1, y_2, \ldots, y_\nu$ οι τιμές μιας μεταβλητής Y για την οποία ισχύει $y = 4x - 3$. Ζητάμε τη μέση τιμή και τη διακύμανση των τιμών της μεταβλητής Y.

Λύση

Υπενθυμίζουμε συνοπτικά τις ιδιότητες του μέσου που είναι χρήσιμες σ' αυτήν την περίπτωση.

Έστω ότι x_i ($i = 1, 2, \ldots, \nu$) είναι οι τιμές της μεταβλητής X με μέσο $\bar{x}$ τότε οι τιμές $\alpha \cdot x_i$, έχουν μέσο $\alpha \cdot \bar{x}$ και οι τιμές $y_i = \alpha x_i + \beta$ έχουν μέσο $\bar{y} = \alpha \bar{x} + \beta$. Στη συγκεκριμένη περίπτωση όπου $y = 4x - 3$ θα είναι $\bar{y} = 4\bar{x} - 3$ δηλαδή $\bar{y} = 4 \cdot 6 - 3 = 21$

Αν x_i, $i = 1, 2, \ldots, \nu$, είναι οι τιμές της μεταβλητής X με διακύμανση s^2_x, τότε οι τιμές $\alpha \cdot x_i$, $i = 1, 2, \ldots, \nu$ θα έχουν διακύμανση $\alpha^2 s^2_x$ και οι τιμές $y_i = \alpha x_i + \beta$ θα έχουν διακύμανση επίσης $\alpha^2 s^2_x$ δηλαδή:

$$s^2_y = \alpha^2 s^2_x$$
$$s^2_y = 16\, s^2_x$$
$$s^2_y = 16 \cdot 4$$
$$s^2_y = 64$$
$$s_y = 8$$

5.8 Συντελεστής μεταβλητότητας

Δυο όμιλοι επιχειρήσεων, που ο καθένας αποτελείται από 5 επιχειρήσεις, είχαν ετήσιες δαπάνες (σε χιλιάδες ευρώ) για το οικονομικό έτος 1999 τα ποσά που αναγράφονται στον παρακάτω πίνακα:

Πίνακας 5.7

Όμιλος Α	200	250	300	300	350
Όμιλος Β	5000	5050	5100	5100	5150

Παρατηρούμε ότι το εύρος μεταβολής των δαπανών για τους δυο ομίλους είναι:

$R_A = 350 – 200 = 150$ (σε χιλιάδες ευρώ)

$R_B = 5150 – 5000 = 150$ (σε χιλιάδες ευρώ)

Οι μέσες τιμές των δαπανών είναι:

$$\bar{x}_A = \frac{200+250+300+300+350}{5} = 280\ ,\ \bar{x}_B = \frac{5000+5050+5100+5100+5150}{5} = 5080$$

και οι διακυμάνσεις των δαπανών σε κάθε όμιλο είναι:

$$s_A^2 = \frac{(200-280)^2+(250-280)^2+(300-280)^2+(300-280)^2+(350-280)^2}{5} = 2600$$

$$s_B^2 = \frac{(5000-5080)^2+(5050-5080)^2+(5100-5080)^2+(5100-5080)^2+(5150-5080)^2}{5} = 2600$$

Παρατηρούμε ότι η διακύμανση $s^2_A = s^2_B = 2600$ οπότε $s_A = s_B \cong 51.000$ ευρώ.

Βλέπουμε ότι η διασπορά των δαπανών είναι ίδια και στους δυο ομίλους. Μπορούμε άραγε να ισχυριστούμε ότι οι 51.000 ευρώ της τυπικής απόκλισης για τον όμιλο Α, που έχει μέση τιμή ετησίων δαπανών $\bar{x}_A$ = 280.000 ευρώ και οι 51.000 ευρώ για τον όμιλο Β, που έχει μέση τιμή ετησίων δαπανών $\bar{x}_B$ = 5.080.000 ευρώ, έχουν την ίδια βαρύτητα;

Ένας τέτοιος ισχυρισμός θα ήταν λανθασμένος, γιατί η σχέση των τιμών της μεταβλητής «ετήσιες δαπάνες» με την τιμή της απόκλισης, είναι διαφορετική για τις επιχειρήσεις των δυο ομίλων.

Ένα μέτρο που μας βοηθάει να αντιμετωπίσουμε τέτοια προβλήματα και μας δίνει τη δυνατότητα συγκρίσεων είναι ο συντελεστής μεταβλητότητας **(CV). Ο συντελεστής μεταβλητότητας είναι καθαρός αριθμός και συσχετίζει την τυπική απόκλιση με τη μέση τιμή των τιμών της μεταβλητής.**

Ορίζεται ως ο λόγος της τυπικής απόκλισης s προς τη μέση τιμή και συνήθως εκφράζεται ως ποσοστό.

Δηλαδή: $CV = \dfrac{\text{τυπική απόκλιση}}{\textit{μέση τιμή}} \cdot 100 = \dfrac{s}{\bar{x}} \cdot 100$

Στο παράδειγμά μας ο συντελεστής μεταβλητότητας για τις δυο επιχειρήσεις θα είναι αντίστοιχα: $CV_A = \dfrac{51}{280} \cdot 100 = 18{,}21$ δηλαδή, $CV_A \% = 18{,}21\%$

και $CV_B = \dfrac{51}{5080} \cdot 100 = 1{,}004$ δηλαδή, $CV_B \% \cong 1{,}00\%$.

Κατά συνέπεια όσο μικρότερο είναι το ποσοστό αυτό τόσο περισσότερη ομοιογένεια υπάρχει στις τιμές της μεταβλητής μας. Στο συγκεκριμένο παράδειγμα ο συντελεστής μεταβλητότητας (coefficient of variation)* των τιμών της μεταβλητής «ετήσιες δαπάνες» του ομίλου Α είναι πολύ μεγαλύτερος από αυτόν του ομίλου Β, άρα η διασπορά των τιμών στον όμιλο Α σε σχέση με το μέσο όρο είναι πολύ μεγαλύτερη από την αντίστοιχη διασπορά στον όμιλο Β.

* Ο συντελεστής μεταβλητότητας προτάθηκε από τον Karl Pearson.

5.9 Η σπουδαιότητα των μέτρων διασποράς για τις επιχειρήσεις

Είναι πλέον προφανές, μετά τη μελέτη που κάναμε πάνω στα μέτρα θέσης και στα μέτρα διασποράς, ότι, αν θέλουμε να έχουμε μία πληρέστερη και ασφαλέστερη εικόνα σχετικά με ένα πρόβλημα ή με ένα θέμα που εξετάζουμε, επιβάλλεται η χρησιμοποίηση των στατιστικών μέτρων θέσης μαζί με αυτά της διασποράς.

Για παράδειγμα, αν ένας οικονομικός ερευνητής, μετά τη μελέτη των οικονομικών δεδομένων που έκανε για λογαριασμό ενός οργανισμού ή μιας επιχείρησης, διαπιστώσει την ύπαρξη υψηλού βαθμού διακύμανσης, αυτό θα πρέπει να τον προβληματίσει και να τον οδηγήσει σε περαιτέρω έρευνα και μελέτη των δεδομένων του. Τα μέτρα διασποράς, ως στατιστικά μέτρα που μας πληροφορούν πόσο διάσπαρτες (απλωμένες) γύρω από τα μέτρα θέσης είναι οι παρατηρήσεις μας, αποτελούν ένα επιπλέον εργαλείο στα χέρια των αναλυτών για να αντιληφθούν το βαθμό αντιπροσωπευτικότητας και αξιοπιστίας των μέτρων θέσης.

Στις επιχειρήσεις και στον οικονομικό χώρο γίνεται ευρύτατη χρησιμοποίηση των μέτρων διασποράς. Για παράδειγμα, ακούμε στην αγορά να γίνεται λόγος για διακύμανση των ποσοτήτων που πωλούνται από τις επιχειρήσεις. Τελευταία πολύ συχνά ακούμε για το ύψος των διακυμάνσεων που παρουσιάζεται στις τιμές των μετοχών των εταιρειών που είναι στο Χρηματιστήριο και για τη διακύμανση στις συναλλαγματικές ισοτιμίες του Ευρώ σε σχέση με άλλα νομίσματα.

Στην επόμενη παράγραφο παραθέτουμε μια απλή περίπτωση μελέτης (case study), ώστε να δούμε την εφαρμογή των όσων αναφέραμε στον τομέα των επιχειρηματικών δραστηριοτήτων.

5.10 Εφαρμογές στις επιχειρήσεις

Περίπτωση μελέτης αποτελέσματος μετά από αναδιοργάνωση επιχείρησης.

Η Εταιρεία G and B Hellas – Neon 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 τη συμπλήρωση της νέας οικονομικής χρήσης, ο γενικός διευθυντής αποφασίζει να μελετήσει τα έσοδα πωλήσεων των 20 υποκαταστημάτων. Έτσι θα αντλήσει συμπεράσματα, που θα τα συγκρίνει με αυτά της περσινής περιόδου, για να μπορέσει να βγάλει κάποια τελικά συμπεράσματα σχετικά με το τι προέκυψε μετά την αναδιοργάνωση της επιχείρησης.

Οι πωλήσεις (σε δεκάδες χιλιάδες ευρώ – 10.000 €) 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 την περσινή χρήση γνωρίζουμε ότι η μέση τιμή των πωλήσεων ήταν $\overline{x}$ = 900.000 ευρώ και η διασπορά των πωλήσεων για τα 20 καταστήματα ήταν s_1 = 100.000 ευρώ. Ακόμη οι χαμηλότερες πωλήσεις υποκαταστήματος ήταν 700.000 ευρώ και οι υψηλότερες 1.400.000 ευρώ.

Ο γενικός διευθυντής συντάσσει αρχικά τον παρακάτω πίνακα και υπολογίζει τη νέα μέση τιμή και τη νέα τυπική απόκλιση:

Πίνακας 5.8

Κλάσεις πωλήσεων	Αριθμός υποκ/των	x'	x^2	νx	νx^2
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
Αθροίσματα	20			2600	340900

Έχουμε λοιπόν ότι η νέα μέση τιμή των πωλήσεων της εταιρείας θα είναι:

$\bar{x}_2 = \frac{\Sigma \nu x}{\nu} = \frac{2600}{20} = 130$ δεκάδες χιλιάδες ευρώ (1.300.000 ευρώ)

Η νέα τυπική απόκλιση s_2 θα είναι:

$$s_2 = \sqrt{\frac{\Sigma \nu x^2}{\nu} - \bar{x}^2}$$

$$s_2 = \sqrt{\frac{340900}{20} - 130^2}$$

$s_2 = 12{,}04 \cong 12$
$s_2 = 12$ δεκάδες χιλιάδες ευρώ (120.000 ευρώ)

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

$CV_1 \frac{s_1}{\bar{x}_1} = \frac{10}{90} = 0{,}111$ ή $CV_1 = 11{,}1\%$

ή $CV_2 = \frac{s_2}{\bar{x}_2} = \frac{12}{130} = 0{,}092$ ή $CV_2 = 9{,}2\%$

Αποτέλεσμα: Όλες οι μεταβολές των στατιστικών μέτρων είναι ευνοϊκές.

Πίνακας 5.9

Τιμές των στατιστικών μέτρων των 2 διαχειριστικών χρήσεων.

Στατιστικά μέτρα	**Προηγούμενη διαχειριστική χρήση (1)**	**Φετινή διαχειριστική χρήση (2)**	**Μεταβολή των χρήσεων 1 και 2**
I. Μέση τιμή - $\bar{x}$ -	$\bar{x}_1 = 90$	$\bar{x}_2 = 130$	αύξηση (+)
II.Διακύμανση - s^2-	$s^2_1 = 100$	$s^2_2 = 144$	αύξηση (+)
III.Τυπική απόκλιση (s)	$s_1 = 10$	$s_2 = 12$	αύξηση (+)
IV.Συντελεστής μεταβλητότητας - CV -	$CV_1 = 11{,}1\%$	$CV_2 = 9{,}2\%$	μείωση (-)
V.Εύρος μεταβολής - Ε.Μ.-	$Ε.Μ._1 = 70$	$Ε.Μ._2 = 51$	μείωση (-)

Συμπέρασμα:

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

α) έχουμε σημαντική αύξηση της μέσης τιμής των πωλήσεων από 900.000 ευρώ πέρυσι σε 1.300.000 ευρώ φέτος (130 – 90 = 40), ποσοστό αύξησης περίπου 44,44%.

β) Η διασπορά των πωλήσεων φέτος παρουσιάζει καλύτερη εικόνα, δηλαδή έχουμε καλύτερη ομοιογένεια ως προς τα έσοδα, αφού ο φετινός $CV_2 = 9{,}2\% < CV_1 = 11{,}1\%$

Ακόμη, έχουμε φέτος ένα εύρος μεταβολής ίσο με: 156 – 105 = 51 δεκάδες χιλιάδες ευρώ, το οποίο είναι αρκετά μικρότερο από το περσινό, που ήταν 140 – 70 = 70 δεκάδες χιλ. ευρώ. Φυσικά η μελέτη δε σταματά εδώ, θα υπάρξει συνέχεια με την επεξεργασία και άλλων στοιχείων και με τη χρήση και άλλων στατιστικών μέτρων, όμως και τα μέχρι τώρα αποτελέσματα που βρέθηκαν είναι σοβαρά και σημαντικά.

ΑΝΑΚΕΦΑΛΑΙΩΣΗ

1. Οι παράμετροι που φανερώνουν το βαθμό διασποράς των δεδομένων γύρω από κάποιο μέτρο θέσης ονομάζονται **μέτρα διασποράς**. Τα μέτρα διασποράς που εξετάσαμε στο κεφάλαιο αυτό είναι: το εύρος μεταβολής, η μέση απόλυτη απόκλιση, η διακύμανση και η τυπική απόκλιση.

2. **Το εύρος μεταβολής** ορίζεται ως η διαφορά της μικρότερης από τη μεγαλύτερη τιμή της μεταβλητής δηλαδή:

R = μεγαλύτερη τιμή της X – μικρότερη τιμή της X.

Το εύρος μεταβολής επηρεάζεται από τις ακραίες τιμές της μεταβλητής.

3. **Η μέση απόλυτη απόκλιση** των τιμών της μεταβλητής X από τη μέση τιμή $\bar{x}$ ορίζεται ως ο μέσος αριθμητικός των απολύτων τιμών των διαφορών $x_i - \bar{x}$ για i = 1, 2, … , ν δηλαδή,

$$\text{M.A.A.} = \frac{\sum |x_i - \bar{x}|}{\nu}$$

4. **Η διακύμανση** ορίζεται ως ο μέσος αριθμητικός των τετραγώνων των διαφορών των τιμών μιας μεταβλητής από τη μέση τιμή της, δηλαδή:

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_\nu - \bar{x})^2}{\nu} = \frac{\sum (x - \bar{x})^2}{\nu}$$

5. **Η τυπική απόκλιση** είναι η θετική τετραγωνική ρίζα της διακύμανσης και έχει τις ίδιες μονάδες με τις τιμές της μεταβλητής X. Όσο μικρότερες είναι οι τιμές της διακύμανσης και της τυπικής απόκλισης τόσο πιο συγκεντρωμένες γύρω από το μέσο βρίσκονται οι τιμές της μεταβλητής X.

6. Για τον υπολογισμό της διακύμανσης σε αταξινόμητα δεδομένα χρησιμοποιούμε τους τύπους:

$$s^2 = \frac{\sum(x - \bar{x})^2}{\nu} \quad ή \quad s^2 = \frac{\sum x^2}{\nu} - \bar{x}^2$$

7. Για τον υπολογισμό της διακύμανσης σε ταξινομημένα δεδομένα χρησιμοποιούμε τους τύπους:

α) από πίνακα συχνοτήτων $s^2 = \frac{\sum \nu x^2}{\nu} - \bar{x}^2$

β) σε ομαδοποιημένα δεδομένα: $s^2 = \frac{\sum \nu x'^2}{\nu} - \bar{x}^2$

όπου x' η κεντρική τιμή της κλάσης.

8. Οι ιδιότητες της διακύμανσης είναι:

α) αν $x_1 = x_2 = \ldots = x_\nu = \kappa$ τότε $s^2 = 0$

β) αν σε όλες τις τιμές της X προσθέσουμε ή αφαιρέσουμε μια σταθερή ποσότητα κ, τότε η διακύμανση s^2 παραμένει αμετάβλητη.

γ) αν όλες οι τιμές της X πολλαπλασιασθούν ή διαιρεθούν με μια σταθερή ποσότητα α, τότε η διακύμανση s^2 πολλαπλασιάζεται ή διαιρείται με α^2.

9. **Ο συντελεστής μεταβλητότητας** είναι καθαρός αριθμός και συσχετίζει την τυπική απόκλιση με το μέγεθος της μέσης τιμής των τιμών της μεταβλητής. Ορίζεται ως ο λόγος της τυπικής απόκλισης s προς τη μέση τιμή των τιμών της μεταβλητής και συνήθως εκφράζεται με ποσοστό δηλ. $CV = \frac{s}{\bar{x}} \cdot 100$.

10. Τα μέτρα διασποράς αποτελούν σημαντικό συμπλήρωμα των μελετών που κάνουμε για να γνωρίζουμε το βαθμό αξιοπιστίας και αντιπροσωπευτικότητας των μέτρων θέσης.

ΕΡΩΤΗΣΕΙΣ ΚΑΤΑΝΟΗΣΗΣ

1. Παράμετροι ή μέτρα διασποράς λέγονται οι αριθμοί: ....

....

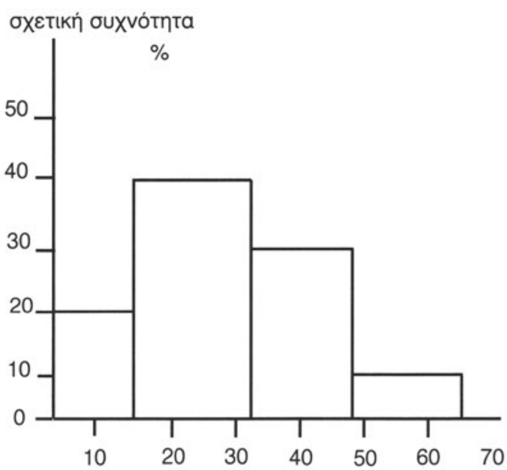
....

2. Το εύρος (R) είναι η διαφορά ....

...., και ισούται με:

R =.. . Το εύρος δε χρησιμοποιείται συνήθως ως αντιπροσωπευτικό μέτρο διασποράς, γιατί για τον υπολογισμό του χρησιμοποιούνται μόνο ....

3.



Το εύρος των παρατηρήσεων που παριστάνονται στο παραπάνω διάγραμμα είναι:

Α: 60 Β: 14 Γ: 64 Δ: 16 Ε: 28

4. Ως μέση απόλυτη απόκλιση των τιμών της μεταβλητής X ορίζεται........

....

Διατυπώστε ανάλογο τύπο αν τα δεδομένα περιέχονται σε πίνακα συχνοτήτων.

5. Ως διακύμανση των τιμών της μεταβλητής X ορίζουμε …………………… και τη συμβολίζουμε με s^2 = ... Η τυπική απόκλιση ορίζεται .... Οι μονάδες στις οποίες εκφράζεται η διακύμανση είναι .... , γι’ αυτό αντί της διακύμανσης χρησιμοποιούμε την τυπική απόκλιση, η οποία έχει ....

....

6. Η μέση τιμή των αριθμών 3, 6, 7, α, 14 είναι 8. Η τυπική απόκλιση είναι:

Α: 6,5 Β: 1,5 Γ: 3,74 Δ: 2

7. Αν για τις 9 τιμές μιας μεταβλητής ισχύει ότι: $\Sigma\,(x - \bar{x})^2 = 234$, τότε η τυπική απόκλιση είναι:

Α: 26 Β: 5,1 Γ: 9 Δ: 2

8. Η διακύμανση σε αταξινόμητα δεδομένα δίνεται από τον πιο εύχρηστο τύπο:

$s^2 =$

9. Η διακύμανση σε δεδομένα που περιέχονται σε πίνακα συχνοτήτων δίνεται από:

$s^2 =$

και σε ομαδοποιημένα δεδομένα δίνεται από:

$s^2 =$

10. Κανονική κατανομή λέγεται η κατανομή συχνοτήτων:..........................

....

11. Ο Συντελεστής μεταβλητότητας εκφράζει ....

και ισούται με CV =..

12. Αν η μέση τιμή είναι 2 και ο συντελεστής μεταβλητότητας είναι 50%, τότε η τυπική απόκλιση είναι:

Α: 0,5 Β: 50 Γ: 1 Δ: 5

13. Τα μέτρα διασποράς είναι σημαντικά γιατί: ....

....

ΑΣΚΗΣΕΙΣ

Ομάδα Α΄

1. Να υπολογίσετε τη μέση τιμή, τη μέση απόλυτη απόκλιση και την τυπική απόκλιση των αριθμών: 2, 3, 6, 9.

Δυο αριθμοί α και β προστίθενται στο σύνολο των τεσσάρων αριθμών έτσι, ώστε η μέση τιμή αυξάνεται κατά 1 και η διακύμανση κατά 2,5. Να βρεθούν οι α και β.

2. Οι μηνιαίοι μισθοί 5 υπαλλήλων μιας επιχείρησης σε ευρώ είναι 800, 1000, 1000, 1200, 1400.

Να βρεθούν και να ερμηνευθούν: α) η μέση τιμή, β) η μέση απόλυτη απόκλιση και γ) η τυπική απόκλιση.

3. Να δείξετε ότι η τυπική απόκλιση των αριθμών 1, 2, 3, 4, 5, 6, 7 είναι 2. Χρησιμοποιώντας το αποτέλεσμα που βρήκατε υπολογίστε την τυπική απόκλιση των αριθμών: i) 11, 12, 13, 14, 15, 16, 17

ii) 100, 200, 300, 400, 500, 600, 700

4. Οι μηνιαίες αποδοχές δύο ομάδων εργαζομένων σε ευρώ δίνονται παρακάτω:

Α′ ομάδα: 1000 1200 1250 1000 1200 1300

Β′ ομάδα: 1000 1100 1200 1000 1250 1400

Να βρεθούν:

i) Η μέση τιμή των αποδοχών της κάθε ομάδας εργαζομένων.

ii) Η τυπική απόκλιση της κάθε ομάδας.

iii) Η μέση απόλυτη απόκλιση κάθε ομάδας.

iv) Να συγκριθούν και να σχολιαστούν τα αποτελέσματα.

5. Το ύψος των μηνιαίων εξόδων ενοικίου που καταβάλλουν 25 εμπορικές επιχειρήσεις δίνεται από τον ακόλουθο πίνακα σε χιλιάδες ευρώ.

Ενοίκια	2	3	5	6	7	8	9	10
Επιχειρήσεις	1	2	6	7	4	2	2	1

Να βρεθούν και να ερμηνευθούν:
i) Η μέση τιμή του ενοικίου.
ii) Η διάμεση τιμή του ενοικίου.
iii) Το εύρος μεταβολής.
iv) Η διακύμανση και η τυπική απόκλιση.

6. Ο παρακάτω πίνακας δίνει τα ύψη των ετήσιων 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

Ζητείται να υπολογισθούν και να ερμηνευθούν:
i) Η μέση τιμή.
ii) Η διακύμανση.
iii) Η τυπική απόκλιση.

Ομάδα Β΄

1. Σε εργοστάσιο συσκευασίας αλατιού, ο υπεύθυνος ισχυρίζεται ότι σε κάθε σάκο αλάτι περιέχονται 25 kg. Επιλέχθηκαν 80 σάκοι και μετρήθηκε η μάζα (x σε kg) του κάθε σάκου. Τα αποτελέσματα ήταν τα εξής:

$$\Sigma\,(x-25) = 27{,}2 \text{ και } \Sigma\,(x-25)^2 = 85{,}1$$

Να βρείτε τη μέση τιμή και την τυπική απόκλιση των μαζών.

2. Το ύψος των ετήσιων κερδών μίας ομάδας 20 επιχειρήσεων σε χιλιάδες ευρώ είναι:

Κέρδη	25	30	35	40	45	50
Επιχειρήσεις	3	3	6	5	2	1

Ζητείται να υπολογισθούν και να ερμηνευθούν:
i) Η μέση τιμή.
ii) Η διάμεσος.
iii) Το εύρος μεταβολής.
iv) Η τυπική απόκλιση.
v) Ο συντελεστής μεταβλητότητας.

3. Ο παρακάτω πίνακας δείχνει τις ετήσιες ακαθάριστες αποδοχές (σε χιλιάδες ευρώ) μιας ομάδας 30 υπαλλήλων σε μια επιχείρηση.

Μηνιαίες αποδοχές	**Αριθμός υπαλλήλων**
20 – 25	5
25 – 30	15
30 – 35	5
35 – 40	3
40 – 45	1
45 –50	1
Σύνολο	30

Ζητείται:
i) Πόσος είναι ο μέσος ετήσιος ακαθάριστος μισθός;
ii) Ποιος είναι ο ετήσιος ακαθάριστος μισθός κάτω από τον οποίο βρίσκεται το 50% των αποδοχών των εργαζομένων.
iii) Να υπολογισθεί ο συντελεστής μεταβλητότητας.

4. Σε μια βιομηχανία αυτοκινήτων ο αριθμός των εργαζομένων το έτος 1995, κατανεμημένος σε ομάδες ηλικιών, φαίνεται στον παρακάτω πίνακα. Υπολογίστε τη μέση τιμή, τη διάμεσο, τη διακύμανση και την τυπική απόκλιση των ηλικιών των εργαζομένων σ' αυτή τη βιομηχανία.

Εκτιμήστε το ποσοστό των υπαλλήλων των οποίων οι ηλικίες βρίσκονται σε απόσταση μιας τυπικής απόκλισης γύρω από το μέσο (δηλαδή σε διάστημα $\bar{x}$ - s, $\bar{x}$ + s).

Τι παρατηρείτε; Εξηγήστε τον ισχυρισμό σας με κατάλληλο διάγραμμα.

Ηλικία	Αριθμός εργαζομένων
15 – 20	6600
20 – 25	6500
25 – 30	5600
30 – 35	5000
35 – 40	4200
40 – 45	3700
45 – 50	3500
50 – 55	3000
55 – 60	2400
60 – 65	2200
Σύνολο	42700

5. Οι ετήσιες πωλήσεις (X) των 15 καταστημάτων-παραρτημάτων (αλυσίδα καταστημάτων), που έχει μεγάλη εμπορική επιχείρηση, δίνονται σε δεκάδες χιλιάδες ευρώ, από τον παρακάτω πίνακα:

x_i	50	60	70	80	90	100	110
v_i	1	2	4	4	2	1	1

Ζητείται να υπολογισθούν:

i) Το μέσο ετήσιο ύψος πωλήσεων των 15 καταστημάτων.

ii) Το εύρος μεταβολής.

iii) Η διακύμανση και η τυπική απόκλιση.

iv) Ο συντελεστής μεταβλητότητας.

Μια άλλη επιχείρηση του ίδιου εμπορικού κλάδου παρουσίασε μέσο ετήσιο ποσό πωλήσεων 700.000 ευρώ και τυπική απόκλιση 170.000 ευρώ. Να συγκρίνετε και να σχολιάσετε τη δραστηριότητα των δύο επιχειρήσεων.

ΦΥΛΛΟ ΑΥΤΟΕΛΕΓΧΟΥ ΣΤΑ ΜΕΤΡΑ ΘΕΣΕΩΣ ΚΑΙ ΔΙΑΣΠΟΡΑΣ

1. Η τιμή της μεταβλητής, η οποία χωρίζει το σύνολο των παρατηρήσεών μας σε δυο ισοπληθείς ομάδες, ονομάζεται:

α) Μέση τιμή ❒
β) Διάμεσος ❒
γ) Πρώτο τεταρτημόριο ❒
δ) Επικρατούσα τιμή ❒

2. Η τιμή που παρουσιάζεται περισσότερες φορές σε μια ομάδα στατιστικών στοιχείων ονομάζεται:

α) Μέση τιμή ❒
β) Διάμεσος ❒
γ) Επικρατούσα τιμή ❒
δ) Πρώτο τεταρτημόριο ❒

3. Το δεύτερο τεταρτημόριο είναι:

α) Η μέση τιμή ❒
β) Η διάμεσος ❒
γ) Η επικρατούσα τιμή ❒
δ) Τίποτα από τα παραπάνω ❒

4. Η μέση τιμή και η διάμεσος:

α) Έχουν πάντα την ίδια τιμή ❒
β) Ποτέ δεν έχουν την ίδια τιμή ❒
γ) Μπορεί να έχουν είτε την ίδια είτε διαφορετική τιμή ❒

5. Το πηλίκο του αθροίσματος των τιμών μιας μεταβλητής Χ προς το πλήθος των τιμών της ονομάζεται:

α) Επικρατούσα τιμή ❒
β) Διάμεσος ❒
γ) Μέση τιμή ❒
δ) Πρώτο τεταρτημόριο ❒

6. Στις ίδιες μονάδες που εκφράζονται οι τιμές της μεταβλητής Χ εκφράζεται και:

α) Η τυπική απόκλιση ❒
β) Η διακύμανση ❒

γ) Ο συντελεστής μεταβλητότητας ❒

7. Δεν προσφέρεται για αλγεβρικούς υπολογισμούς:
α) Η διακύμανση ❒
β) Η τυπική απόκλιση ❒
γ) Η μέση απόλυτη απόκλιση ❒
δ) Ο συντελεστής μεταβλητότητας ❒

8. Στα μέτρα θέσης ανήκουν:
α) Η διάμεσος και η μέση τιμή ❒
β) Η διάμεσος και η διακύμανση ❒
γ) Η διακύμανση και η τυπική απόκλιση ❒
δ) Το εύρος μεταβολής και η επικρατούσα τιμή ❒

9. Αν όλες οι τιμές μιας μεταβλητής πολλαπλασιασθούν επί μια σταθερή ποσότητα α τότε:
α) Η διακύμανση πολλαπλασιάζεται επί α ❒
β) Η διακύμανση πολλαπλασιάζεται επί $α^2$ ❒
γ) Η διακύμανση παραμένει αμετάβλητη ❒

10. Η τυπική απόκλιση είναι η τετραγωνική ρίζα της:
α) Διαμέσου ❒
β) Διακύμανσης ❒
γ) Μέσης τιμής ❒

11. Η διαφορά της μικρότερης από τη μεγαλύτερη τιμή των δεδομένων μας, δίνεται από:
α) Την επικρατούσα τιμή ❒
β) Το εύρος μεταβολής ❒
γ) Το συντελεστή μεταβλητότητας ❒
δ) Την τυπική απόκλιση ❒

12. Ο συντελεστής μεταβλητότητας δίνεται από το πηλίκο της:
α) Διακύμανσης προς τη μέση τιμή ❒
β) Διακύμανσης προς την τυπική απόκλιση ❒
γ) Τυπικής απόκλισης προς τη μέση τιμή ❒
δ) Μέσης τιμής προς την τυπική απόκλιση ❒

13. Από τα παρακάτω μεγέθη καθαρός αριθμός είναι:

α) Η διακύμανση ❒
β) Η τυπική απόκλιση ❒
γ) Η μέση τιμή ❒
δ) Ο συντελεστής μεταβλητότητας ❒

ΚΕΦΑΛΑΙΟ 6ο

ΠΑΛΙΝΔΡΟΜΗΣΗ – ΣΥΣΧΕΤΙΣΗ

❖ *«Σπάνια πλέον πληροφορούμαι οτιδήποτε τόσο έξυπνο, ώστε να εντυπωσιάσει τη φαντασία, όπως η υπέροχη μορφή της κοσμικής τάξης που εκφράζεται με το "Νόμο της Συχνότητας Σφάλματος". Ο Νόμος θα είχε προσωποποιηθεί από τους Έλληνες και λατρευτεί σαν Θεός, εάν τον εγνώριζαν»* **Sir Francis Galton** (1822-1911).

❖ Η **μέθοδος των ελαχίστων τετραγώνων** δημοσιεύτηκε ολοκληρωμένη για πρώτη φορά το 1806, σε μια εργασία του Γάλλου **Legendre Adrien – Marie** (1752-1833) με τον τίτλο "Nouvelles methodes pour la determination des orbites comètes". Νωρίτερα από τον Legendre, ο **Gauss**, σε ηλικία μόλις 17 ετών, είχε καταλήξει στη διατύπωση των ελαχίστων τετραγώνων, αλλά δεν προέβαλλε αυτό το εύρημά του, προφανώς επειδή δεν το θεώρησε ως κάτι πολύ σημαντικό.

ΚΕΦΑΛΑΙΟ 6ο

ΠΑΛΙΝΔΡΟΜΗΣΗ* – ΣΥΣΧΕΤΙΣΗ

6.1 Γενικά

Στα προηγούμενα κεφάλαια μελετήσαμε σε κάθε πληθυσμό ένα χαρακτηριστικό του. Εξετάσαμε, για παράδειγμα, τις πωλήσεις ενός κλάδου επιχειρήσεων, τους μισθούς των υπαλλήλων μιας επιχείρησης, καθώς και τη βαθμολογία των μαθητών μιας τάξης ή το ύψος τους κτλ. Σε κάθε περίπτωση κατασκευάσαμε τον πίνακα και το διάγραμμα συχνοτήτων. Τέτοιες κατανομές, που προκύπτουν από την εξέταση ενός πληθυσμού ως προς μια μόνο μεταβλητή ή ιδιότητα, ονομάζονται **μονοπαραμετρικές κατανομές**.

Συχνά όμως οι τιμές μιας μεταβλητής, για παράδειγμα, πωλήσεις ενός προϊόντος, επηρεάζονται άμεσα από τις τιμές μιας άλλης μεταβλητής, όπως από τις διαφημιστικές δαπάνες ή από το βαθμό ζήτησης ενός αγαθού κ.ο.κ. Κατά συνέπεια, είναι αναγκαία η ταυτόχρονη μελέτη δύο ή και περισσότερων μεταβλητών, με σκοπό να αναζητήσουμε μια σχέση που να περιγράφει όσο το δυνατό καλύτερα, την εξάρτηση που υπάρχει μεταξύ τους. Οι κατανομές που προκύπτουν από τη μελέτη ενός πληθυσμού ως προς δύο μεταβλητές ονομάζονται **διπαραμετρικές**.

Είναι γνωστό ότι σε πολλούς κλάδους των Μαθηματικών η σχέση που υπάρχει μεταξύ των μεταβλητών είναι τέτοια, ώστε, αν γνωρίζουμε τη μεταβολή της μιας μεταβλητής, μπορούμε να προβλέψουμε ακριβώς τη μεταβολή και της άλλης. Για παράδειγμα, αν διπλασιαστεί το μήκος της πλευράς ενός

* Ο όρος **παλινδρόμηση** (regression) χρησιμοποιήθηκε για πρώτη φορά το 1877 από τον Άγγλο Sir Francis Galton (1822-1911), ο οποίος μελετώντας τη σχέση μεταξύ των αναστημάτων των γονέων και των αναστημάτων των παιδιών τους, παρατήρησε ένα είδος επαναφοράς (παλινδρόμησης) των αναστημάτων των παιδιών στα αναστήματα των γονέων τους.

τετραγώνου γνωρίζουμε ότι θα τετραπλασιασθεί το εμβαδόν του. Στην οικονομία όμως συνήθως αυτό δεν ισχύει, δηλαδή δεν μπορούμε να ισχυρισθούμε ότι, αν αυξηθούν οι διαφημιστικές δαπάνες μιας επιχείρησης, ανάλογα θα αυξηθούν και τα έσοδά της.

Στο κεφάλαιο αυτό, λοιπόν, θα αναζητήσουμε μια σχέση εξάρτησης μεταξύ της μεταβλητής X, την οποία θεωρούμε ανεξάρτητη και συνηθίζεται να λέγεται **ερμηνευτική**, και της μεταβλητής Y, την οποία θεωρούμε εξαρτημένη από τη X. Θα προσπαθήσουμε να εκτιμήσουμε τις τιμές της μεταβλητής Y με βάση τις τιμές της μεταβλητής X. Επίσης θα μελετήσουμε το πώς μπορεί να μετρηθεί ο βαθμός εξάρτησης, που τυχόν υπάρχει, μεταξύ των δύο μεταβλητών.

6.2 Διάγραμμα διασποράς

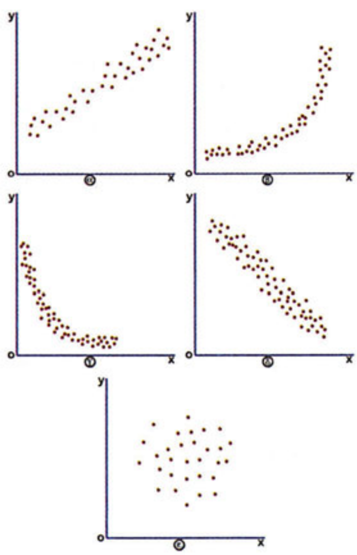
Ένας απλός τρόπος για να μελετήσουμε από κοινού τις δύο μεταβλητές είναι να κατασκευάσουμε το λεγόμενο διάγραμμα διασποράς, του οποίου η κατασκευή είναι πολύ απλή.

Θεωρούμε ένα ορθοκανονικό σύστημα αξόνων, όπου στον οριζόντιο άξονα τοποθετούνται οι τιμές της ανεξάρτητης μεταβλητής X και στον κάθετο οι τιμές της εξαρτημένης μεταβλητής Y. Αφού πρόκειται για την ταυτόχρονη μελέτη δύο μεταβλητών X και Y οι παρατηρήσεις μας θα προσδιορίζονται από τα ζεύγη τιμών:

$$(x_1, y_1), (x_2, y_2), \ldots , (x_\nu, y_\nu)$$

και γνωρίζουμε ότι σε κάθε ζεύγος αντιστοιχεί ένα σημείο του επιπέδου με τετμημένη x_i και τεταγμένη y_i.

Όπως βλέπουμε στο διάγραμμα 6.1, οι παρατηρήσεις μας σχηματίζουν ένα πλήθος σημείων που λέγεται **νέφος σημείων ή διάγραμμα διασποράς ή στικτό διάγραμμα** (scatter diagram).



(Σχ.6.1)

Τίθεται λοιπόν το ερώτημα, αν θα μπορούσε να χαραχθεί στα διαγράμματα διασποράς μια ευθεία ή καμπύλη γραμμή, ανάλογα με την περίπτωση, η οποία να αναπαριστά (να περιγράφει), όσο το δυνατόν καλύτερα, τη σχέση που υπάρχει μεταξύ των δύο μεταβλητών.

Θα αναζητήσουμε την απάντηση στο ερώτημα αυτό μέσα από μερικά παραδείγματα.

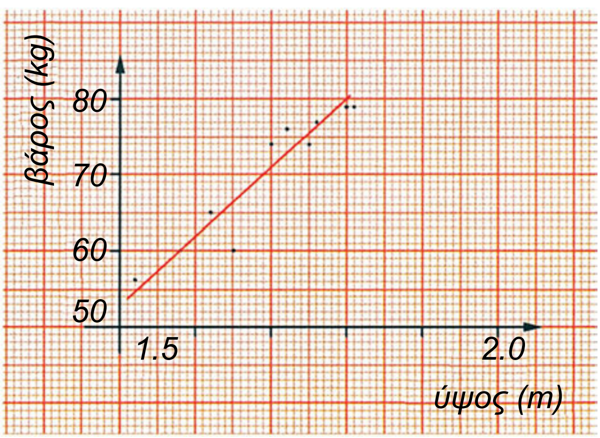
Παράδειγμα 6.1:

Ο παρακάτω πίνακας δίνει τα ύψη και τα αντίστοιχα βάρη 10 μαθητών:

Πίνακας 6.1

Μαθητής	1ος	2ος	3ος	4ος	5ος	6ος	7ος	8ος	9ος	10ος
Χ: ύψος (m)	1.52	1.58	1.62	1.65	1.70	1.72	1.75	1.76	1.80	1.81
Υ:βάρος (kp)	56	60	65	60	74	76	74	77	79	79

Από μια απλή παρατήρηση βλέπουμε ότι πρέπει να υπάρχει κάποια σχέση μεταξύ των δύο μεταβλητών ύψος και βάρος. Έτσι φαίνεται ότι ο ψηλότερος μαθητής είναι και βαρύτερος. Παρατηρούμε, όμως, ότι από αυτή τη γενική διαπίστωση, ο 4ος και ο 7ος μαθητής αποτελούν εξαίρεση. Σε ορθογώνιο σύστημα αξόνων δίνουμε και το διάγραμμα των παρατηρήσεών μας. Το διάγραμμα αυτό το ονομάζουμε διάγραμμα διασποράς.



(Σχ. 6.2)

Παράδειγμα 6.2:

Αν δ είναι το μήκος της διαμέτρου ενός κύκλου και L το αντίστοιχο μήκος του κύκλου, τότε έχουμε τον επόμενο πίνακα και το αντίστοιχο διάγραμμα διασποράς.

Πίνακας 6.2

δ σε cm.	1	2	3	4	5
L σε cm.	3,14	6,28	9,42	12,56	15,70

Έτσι η σχέση μεταξύ του μήκους της διαμέτρου και του αντίστοιχου μήκους του κύκλου εκφράζεται:

α) Αριθμητικά στον παραπάνω πίνακα,
β) γραφικά από μια ευθεία γραμμή (αφού η δ θεωρητικά μπορεί να πάρει κάθε θετική τιμή) και
γ) αλγεβρικά από το γνωστό μας τύπο: $L = 2\pi\rho = \pi{\cdot}\delta$, όπου $\pi \cong 3{,}14$

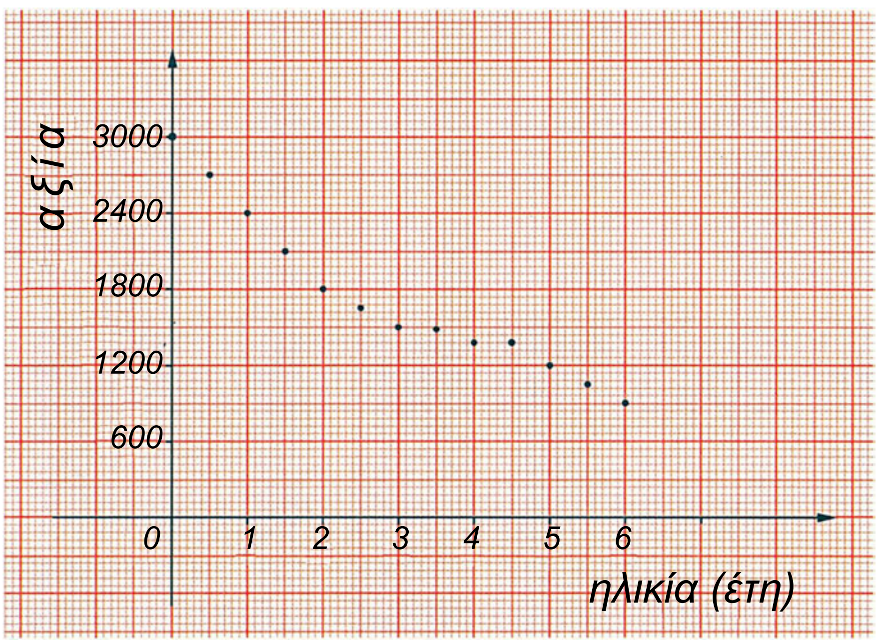
(Σχ.6.3)

Παράδειγμα 6.3:

Ο πίνακας 6.3 και το αντίστοιχο διάγραμμα δείχνουν τη σχέση μεταξύ της ηλικίας ενός δικύκλου και της αντίστοιχης αξίας του σε ευρώ.

Πίνακας 6.3

Ηλικία (έτη)	0	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6
Αξία	3000	2700	2400	2100	1800	1650	1500	1500	1350	1350	1200	1050	900



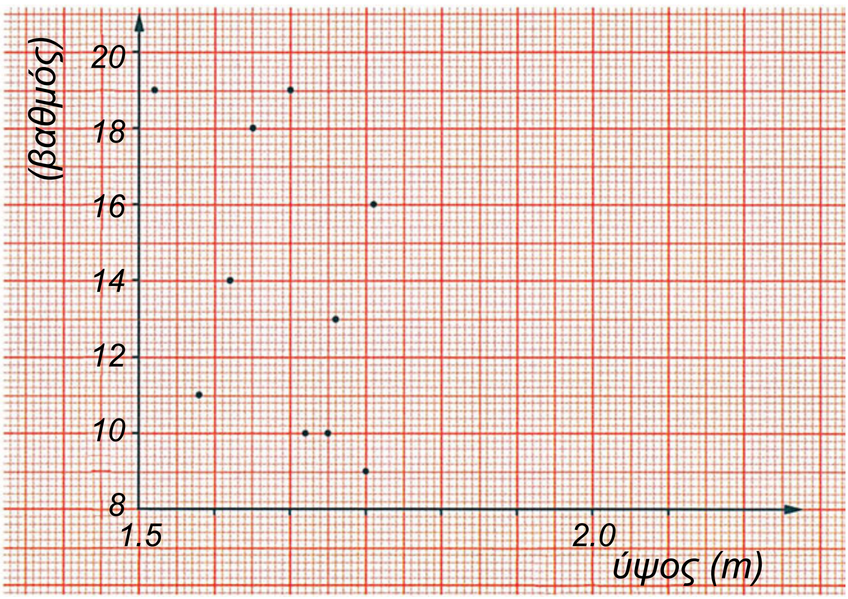
(Σχ.6.4)

Παράδειγμα 6.4:

Για τους 10 μαθητές του πρώτου παραδείγματος ενδιαφερθήκαμε να δούμε αν υπάρχει σχέση μεταξύ του ύψους τους και της επίδοσής τους στα μαθηματικά. Ο πίνακας 6.4 και το αντίστοιχο διάγραμμα διασποράς μας βοηθούν σ' αυτό.

Πίνακας 6.4

Μαθητής	1ος	2ος	3ος	4ος	5ος	6ος	7ος	8ος	9ος	10ος
X: ύψος	1,52	1,58	1,62	1,65	1,70	1,72	1,75	1,76	1,80	1,81
Y: βαθμός	19	11	14	18	19	10	10	13	9	16



(Σχ.6.5)

Όσο και να παρατηρήσουμε τον πίνακα και το διάγραμμα δεν είναι δυνατό να καταλήξουμε σε συμπέρασμα ως προς τη σχέση ύψους και επίδοσης στα μαθηματικά. Στο διάγραμμα παρατηρούμε ότι η διασπορά των σημείων είναι τέτοια, που δε μας θυμίζει καμία γνωστή μας καμπύλη.

Στα προηγούμενα τέσσερα παραδείγματα παρατηρούμε τα εξής:

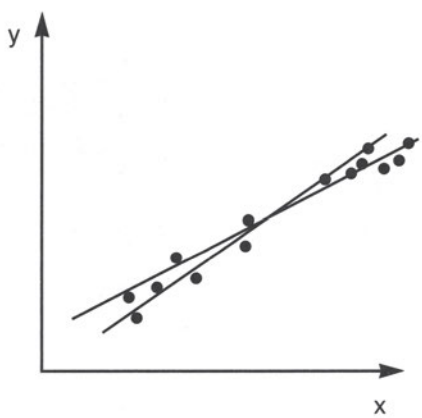
Στο πρώτο παράδειγμα τα σημεία (x, y) του διαγράμματος διασποράς φαίνεται να βρίσκονται κοντά σε μια ευθεία γραμμή, η οποία και περιέχει μερικά από αυτά. Στο δεύτερο όλα τα σημεία βρίσκονται πάνω στην ευθεία που έχει εξίσωση: $L = \pi \cdot \delta$. Στο τρίτο τα σημεία φαίνεται να βρίσκονται κοντά σε μια ευθεία, αλλά μπορεί κάποιος να πει ότι είναι κοντά σε κάποια καμπύλη από τις γνωστές μας υπερβολές $y = \frac{\alpha}{x}$, $x \neq 0$, ενώ στο τέταρτο η διασπορά δε μας θυμίζει καμία γνωστή γραμμή.

Έτσι στις τρεις περιπτώσεις διαπιστώνουμε κάποια σχέση (συμμεταβολή ή συγγένεια) μεταξύ των δύο μεταβλητών X και Y. Στην 4η περίπτωση δεν παρατηρούμε καμία σχέση μεταξύ τους.

6.3 Γραμμική παλινδρόμηση

Είναι φανερό από τα παραδείγματα που αναφέραμε στην προηγούμενη παράγραφο ότι μια προσεγγιστικά «ευθύγραμμη» μορφή νέφους των σημείων του διαγράμματος διασποράς δύο μεταβλητών X και Y, μας οδηγεί στην αναζήτηση μιας ευθείας $y = \alpha + \beta x$ **που να δίνει την καλύτερη δυνατή περιγραφή της σχέσης που υπάρχει μεταξύ των μεταβλητών X και Y.**

Ας παρατηρήσουμε το διάγραμμα διασποράς (Σχ.6.6):

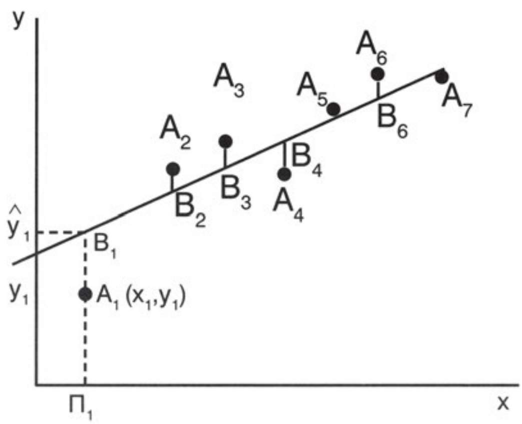


(Σχ.6.6)

Βλέπουμε ότι είναι πολύ πιθανόν δύο μελετητές με το «μάτι» να φέρουν δύο ευθείες. Το ερώτημα που τίθεται όμως είναι: «ποια είναι εκείνη η ευθεία που προσεγγίζει καλύτερα τα σημεία του διαγράμματος διασποράς, δηλαδή τις εικόνες των ζευγών (x_i,y_i);».

Σήμερα η μέθοδος που χρησιμοποιείται σχεδόν αποκλειστικά για τον προσδιορισμό της ευθείας, που περνά όσο το δυνατόν πιο κοντά από τα σημεία των παρατηρήσεών μας – δηλαδή της **ευθείας παλινδρόμησης – είναι γνωστή ως μέθοδος των ελαχίστων τετραγώνων** (method of least squares).

Έστω ότι η εξίσωση της γραμμής είναι η ευθεία y = α + βx. Κατασκευάζουμε το διάγραμμα διασποράς των σημείων A_i (x_i, y_i), όπου i = 1, 2, 3, …, ν (Σχ. 6.7).



(Σχ. 6.7)

Η ευθεία παλινδρόμησης θα είναι η ευθεία, για την οποία το άθροισμα των τετραγώνων των διαφορών των προβλεπόμενων τιμών $(\hat{y}_i)$ της Y από τις αντίστοιχες πραγματικές τιμές (y_i) γίνεται ελάχιστο.

Από το παραπάνω διάγραμμα παρατηρούμε ότι για το σημείο A_1 (x_1, y_1) που δεν ανήκει στην ευθεία (ε) θα ισχύει :

$$\hat{y}_1 - y_1 = Π_1B_1 - Π_1A_1 = (α + \beta\, x_1) - y_1 = ε_1$$

$$\text{Δηλαδή } \hat{y}_1 - y_1 = ε_1$$

ενώ για το σημείο A_5 (x_5, y_5) που βρίσκεται στην ευθεία (ε) θα ισχύει:

$$\hat{y}_5 - y_5 = ε_5 = 0$$

Παρατηρούμε επίσης ότι η τετμημένη του B_i συμπίπτει με την τετμημένη του A_i, ενώ η τεταγμένη του B_i είναι $\hat{y}_i = α + \beta\, x_i$ γιατί το B_i ανήκει στην (ε):

$$y = α + βx$$

Οι διαφορές $\hat{y}_i$ - y_i συμβολίζονται διεθνώς με το ελληνικό γράμμα ε_i, όταν αναφερόμαστε στη μελέτη πληθυσμού και με e_i, όταν εξετάζουμε δείγματα.

Η ευθεία για την οποία το άθροισμα:
$\Sigma\varepsilon_i^2 = (A_1B_1)^2 + (A_2B_2)^2 + (A_3B_3)^2 + (A_4B_4)^2 + ...$
Δηλ. $\Sigma = (\hat{y}_1 - y_1)^2 + (\hat{y}_2 - y_2)^2 + (\hat{y}_3 - y_3)^2 + ...$
ή $\Sigma = (\alpha + \beta x_1 - y_1)^2 + (\alpha + \beta x_2 - y_2)^2 + (\alpha + \beta x_3 - y_3)^2 + ...$

γίνεται ελάχιστο* ορίζεται ως ευθεία παλινδρόμησης και περιγράφεται από την εξίσωση:

$$\hat{y} = \hat{\alpha} + \hat{\beta}x$$

Μπορούμε να υπολογίσουμε τις παραμέτρους $\hat{\alpha}$ και $\hat{\beta}$ της ευθείας, έτσι ώστε το $\Sigma\varepsilon_i$ να γίνεται ελάχιστο λύνοντας το σύστημα των **κανονικών εξισώσεων** (normal equations):

$$\boxed{\begin{aligned} \sum \hat{y} &= \nu\hat{\alpha} + \hat{\beta}\sum x \\ \sum xy &= \hat{\alpha}\sum x + \hat{\beta}\sum x^2 \end{aligned}} \quad (6.1)$$

Όπου $\sum y = \sum_{i=1}^{\nu} y_i$, $\sum x = \sum_{i=1}^{\nu} x_i$, $\sum xy = \sum_{i=1}^{\nu} x_i y_i$, $\sum x^2 = \sum_{i=1}^{\nu} x_i^2$ και ν το σύνολο των πραγματικών τιμών της Y, δηλαδή το πλήθος των ζευγών των παρατηρήσεών μας.

Από την πρώτη εξίσωση των κανονικών εξισώσεων προκύπτει ότι:

$$\frac{\sum y}{\nu} = \frac{\nu\hat{\alpha}}{\nu} + \frac{\beta\sum x}{\nu} \Leftrightarrow \overline{y} = \hat{\alpha} + \hat{\beta}\overline{x} \Leftrightarrow \hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$$

ή

$$\boxed{\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}}$$

* Λαμβάνουμε το άθροισμα $\sum(y - \hat{y})^2$ και όχι το $\sum(y - \hat{y})$, επειδή οι θετικές και αρνητικές διαφορές $(y - \hat{y})$ αθροιζόμενες αλληλοαναιρούνται, δηλαδή $\sum(y - \hat{y})=0$.

Η πλήρης και ακριβής διατύπωση της εξίσωσης που αντιπροσωπεύει το μοντέλο γραμμικής παλιδρόμησης είναι $y = \alpha + \beta x + \varepsilon$, όπου ε είναι ο όρος τυχαίο σφάλμα (random error term).

από τη λύση του συστήματος των κανονικών εξισώσεων προκύπτει η παράμετρος $\hat{\beta}$:

$$\hat{\beta}=\frac{\sum xy-\frac{\sum x\sum y}{v}}{\sum x^2-\frac{(\sum x)^2}{v}}=\frac{v\sum xy-\sum x\sum y}{v\sum x^2-(\sum x)^2}$$

(6.2)

ή διαιρώντας αριθμητή και παρονομαστή με το πλήθος των παρατηρήσεων στο τετράγωνο έχουμε ότι:

$$\hat{\beta}=\frac{\frac{\sum xy}{v}-\frac{\sum x\sum y}{v^2}}{\frac{\sum x^2}{v}-\frac{(\sum x)^2}{v^2}}\Leftrightarrow\hat{\beta}=\frac{\frac{\sum xy}{v}-\left(\frac{\sum x}{v}\right)\left(\frac{\sum y}{v}\right)}{\frac{\sum x^2}{v}-\left(\frac{\sum x}{v}\right)^2}$$ απ' όπου προκύπτει ένας

εύχρηστος τύπος υπολογισμού της παραμέτρου $\hat{\beta}$:

$$\hat{\beta}=\frac{\frac{\sum xy}{v}-\bar{x}\cdot\bar{y}}{\frac{\sum x^2}{v}-\bar{x}^2}$$

(6.3)

Οι αριθμοί $\hat{\alpha}$ και $\hat{\beta}$ ονομάζονται εκτιμήτριες ελαχίστων τετραγώνων

Παρατηρούμε ότι η ευθεία παλινδρόμησης $\hat{y}=\hat{\alpha}+\hat{\beta}$ διέρχεται από το σημείο $M(\bar{x},\bar{y})$ γιατί $\bar{y}=\hat{\alpha}+\hat{\beta}\bar{x}$. Έτσι μπορούμε να γράψουμε την εξίσωση της ευθείας παλινδρόμησης στη μορφή:

$$\hat{y}-\bar{y}=\hat{\beta}(x-\bar{x}) \text{ και } \hat{\beta}=\frac{\hat{y}-\bar{y}}{x-\bar{x}}=\text{εφω},$$

όπου ω η γωνία κλίσης της ευθείας.

Η παράμετρος $\hat{\beta}$, γνωστή ως συντελεστής διεύθυνσης της ευθείας ή γωνιακός συντελεστής της ευθείας παλινδρόμησης ή απλά συντελεστής παλινδρόμησης, $\hat{\beta}$ προσδιορίζει τη μεταβολή της εξαρτημένης μεταβλητής Y, όταν η ερμηνευτική μεταβλητή X μεταβληθεί κατά μια μονάδα.

Αν $\hat{\beta} > 0$, τότε οι τιμές της Y αυξάνονται, όταν αυξάνονται αντίστοιχα οι τιμές της ερμηνευτικής μεταβλητής X.

Αν $\hat{\beta} < 0$, τότε οι τιμές της Y μειώνονται, όταν αυξάνονται αντίστοιχα οι τιμές της ερμηνευτικής μεταβλητής X.

Παράδειγμα 6.5:

Το αεροδρόμιο μιας επαρχίας μπορεί να εξυπηρετεί ικανοποιητικά ορισμένο αριθμό δρομολογίων (έστω z_o) την ημέρα. Τους θερινούς μήνες όμως, εξαιτίας της αυξημένης τουριστικής κίνησης, παρατηρούνται καθυστερήσεις στα δρομολόγια των αεροσκαφών. Η διοίκηση του αεροδρομίου έδωσε εντολή να καταγραφούν οι χρόνοι καθυστέρησης (y_i) των πτήσεων καθώς και ο επιπλέον του z_o αριθμός δρομολογίων (x_i) σε διάρκεια 24 ωρών. Ο πίνακας 6.5 παρουσιάζει τα (υποθετικά) στοιχεία της έρευνας αυτής.

Να προσδιορίσετε την εξίσωση της ευθείας παλινδρόμησης της εξαρτημένης μεταβλητής Y (χρόνος καθυστέρησης σε λεπτά) ως προς την ερμηνευτική μεταβλητή X (αριθμός επιπλέον πτήσεων) και να ερμηνευθούν οι παράμετροι $\hat{\alpha}$ και $\hat{\beta}$ της ευθείας παλινδρόμησης.

Πίνακας 6.5

Αριθμός επιπλέον δρομολογίων x_i	1	2	3	4	5	6	7
Χρόνος καθυστέρησης (σε λεπτά) y_i	3	5	7	9	11	13	15

(Υποθετικά δεδομένα)

Λύση

Κατασκευάζουμε τον πίνακα (6.6):

Πίνακας 6.6

α/α	x_i	y_i	x^2_i	y^2_i	x_iy_i
1	1	3	1	9	3
2	2	5	4	25	10
3	3	7	9	49	21
4	4	9	16	81	36
5	5	11	25	121	55
6	6	13	36	169	78
7	7	15	49	225	105
Σύνολο	$\sum x = 28$	$\sum y = 63$	$\sum x^2 = 140$	$\sum y^2 = 679$	$\sum xy = 308$

Αντικαθιστώντας στο σύστημα των κανονικών εξισώσεων:

$$\sum y = \nu\hat{\alpha} + \hat{\beta}\sum x$$

$$\sum xy = \hat{\alpha}\sum x + \hat{\beta}\sum x^2 \quad \text{έχουμε}$$

$$\begin{cases} 63 = 7\hat{\alpha} + 28\hat{\beta} \\ 308 = 28\hat{\alpha} + 140\hat{\beta} \end{cases} \Leftrightarrow \begin{cases} (-4)63 = (-4)7\hat{\alpha} + (-4)28\hat{\beta} \\ 308 = 28\hat{\alpha} + 140\hat{\beta} \end{cases} \Leftrightarrow$$

$$\begin{cases} -252 = -28\hat{\alpha} - 112\hat{\beta} \\ 308 = 28\hat{\alpha} + 140\hat{\beta} \end{cases} \Leftrightarrow \begin{cases} 63 = 7\hat{\alpha} + 28\hat{\beta} \\ 56 = 28\hat{\beta} \end{cases}$$

Άρα $\hat{\beta}$ = 2 και $\hat{\alpha}$ = 1, οπότε η ευθεία παλινδρόμησης είναι $y' = 1 + 2x$. Θα μπορούσαμε επίσης να υπολογίσουμε τις παραμέτρους $\hat{\alpha}$ και αντικαθιστώντας στους τύπους: $\hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x}$ και $\hat{\beta} = \dfrac{\dfrac{\sum xy}{\nu} - \bar{x}\,\bar{y}}{\dfrac{\sum x^2}{\nu} - \bar{x}^2}$

Υπολογίζουμε: $\bar{x} = \frac{\sum x}{\nu} = \frac{28}{7} = 4$ και $\bar{y} = \frac{\sum y}{\nu} = \frac{63}{7} = 9$

και έχουμε:

$$\hat{\beta} = \frac{\frac{308}{7} - 4 \cdot 9}{\frac{140}{7} - 16} = \frac{44 - 36}{20 - 16} = \frac{8}{4} = 2 \quad \text{και} \quad \hat{\alpha} = 9 - 2 \cdot 4 = 1$$

επομένως, όπως και πριν

$$\hat{y} = 2x + 1$$

- **Ερμηνεία των $\hat{\alpha}$ και $\hat{\beta}$.**

Από την εξίσωση της $y' = 2x + 1$ προκύπτει ότι, όταν $x = 0$, τότε $\hat{y} = 1 = \hat{\alpha}$.

Ως γνωστό το σημείο Β (0,1) βρίσκεται στον άξονα Υ και είναι το σημείο τομής της ευθείας παλινδρόμησης με τον άξονα Υ. Η τιμή $y' = 1 = \hat{\alpha}$ παριστάνει την προβλεπόμενη τιμή της μεταβλητής Υ, όταν η τιμή της Χ είναι $x = 0$.

Στο συγκεκριμένο παράδειγμα η τιμή $\beta' = 2$ σημαίνει ότι η εξάρτηση των μεταβλητών είναι θετική, δηλαδή, όταν αυξάνεται η τιμή της Χ κατά μια μονάδα (ένα επιπλέον δρομολόγιο), η τιμή της μεταβλητής Υ αυξάνεται κατά δύο μονάδες (δύο λεπτά).

Πράγματι για $x = 1$ $\hat{y} = 2 \cdot 1 + 1 = 3$
και για $x = 2$ $\hat{y} = 2 \cdot 2 + 1 = 5$

Δηλαδή, όταν ο αριθμός των δρομολογίων ξεπερνά την τιμή εύρυθμης λειτουργίας του αεροδρομίου κατά μια μονάδα ($x = 1$), τότε ο χρόνος καθυστέρησης των πτήσεων αυξάνεται κατά δύο λεπτά ($\beta = 2$). Η μη μηδενική τιμή της παραμέτρου, σημαίνει ότι, όταν το αεροδρόμιο λειτουργεί σε κανονικές συνθήκες, οπότε ο αριθμός των επιπλέον δρομολογίων είναι μηδέν ($x = 0$), τότε πάλι έχουμε καθυστέρηση ενός λεπτού $\hat{y} = 2 \cdot 0 + 1$ και $\hat{y} = 1$. Αυτό σημαίνει ότι, λόγω της μη καλής οργάνωσης και διοίκησης του αεροδρομίου, δεν υπάρχει πλήρης εκμετάλλευση των δυνατοτήτων του, ώστε να αποφεύγεται η καθυστέρηση του ενός λεπτού, όταν ο αερολιμένας λειτουργεί κανονικά. (Υπενθύμιση: τα δεδομένα ήταν υποθετικά).

Αξίζει να σημειωθεί ότι εκτός από την ταυτόχρονη μελέτη δύο μεταβλητών, η οποία αποτελεί την περίπτωση της **απλής παλινδρόμησης**, έχουμε και την περίπτωση της **πολλαπλής παλινδρόμησης** (multiple regression), όπου μελετώνται συγχρόνως περισσότερες από δύο μεταβλητές.

6.4 Συντελεστής συσχέτισης

Ας ξαναδούμε τα δεδομένα του παραδείγματος (6.5), δηλαδή το χρόνο καθυστέρησης των πτήσεων (y_i) και τον αριθμό των επιπλέον δρομολογίων (x_i). Στο νέφος αυτών των σημείων (x_i y_i) προσαρμόσαμε την ευθεία παλινδρόμησης

$$y = 2x + 1.$$

Η ίδια έρευνα λόγω της υπεραυξημένης εναέριας κυκλοφορίας επαναλήφθηκε και τον επόμενο μήνα. Τα αποτελέσματα περιέχονται στον πίνακα που ακολουθεί:

Πίνακας 6.7

Αριθμός επιπλέον δρομολογίων (x_i)	2	3	3	4	7	7	9
Χρόνος καθυστέρησης (σε λεπτά) (y_i)	3,5	5	11	15	18,5	-2,25	26,25

(Υποθετικά δεδομένα)

Θα προσδιορίσουμε και πάλι την εξίσωση της ευθείας παλινδρόμησης της Υ επάνω στη Χ και θα χαράξουμε τις αντίστοιχες ευθείες παλινδρόμησης των δεδομένων των δύο συνόλων.

Συμπληρώνουμε τον πίνακα (6.7) για να βρούμε τις τιμές των $\hat{\alpha}$ και $\hat{\beta}$ ώστε να προσδιορίσουμε την $\hat{y} = \hat{\alpha} + \hat{\beta} \cdot x$.

Πίνακας 6.8

α/α	x_i	y_i	x^2_i	$x_i y_i$
1	2	3,5	4	17
2	3	5	9	15
3	3	11	9	33
4	4	15	16	60
5	7	18,5	49	129,5
6	7	-2,25	64	-15,75
7	9	26,25	64	236,25
Σύνολο	35	77	215	465

Το ζεύγος (7, -2,25), που φαίνεται στον πίνακα, δηλώνει ότι 7 αεροπλάνα της 6[ης] χρονικής ζώνης (προφανώς της πλέον ήρεμης, για παράδειγμα, νυχτερινής ή μεταμεσονύκτιας χρονικής ζώνης) προσγειώθηκαν κατά 2,25 λεπτά νωρίτερα από την καθορισμένη ώρα άφιξης.

Από τους τύπους υπολογισμού του $\hat{\beta}$, επιλέγουμε εκείνον που είναι πιο εύχρηστος σύμφωνα με τις τιμές των δεδομένων που έχουμε. Άρα:

$$\hat{\beta} = \frac{\nu\sum xy - \sum x \sum y}{\nu \sum x^2 - (\sum x)^2} = \frac{7 \cdot 465 - 35 \cdot 77}{7 \cdot 215 - (35)^2} = \frac{3.255 - 2.695}{1505 - 1225} = \frac{560}{280} = 2$$

όπως έχουμε ήδη αναφέρει στην προηγούμενη παράγραφο $\hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x}$

άρα υπολογίζουμε: $\bar{y} = \frac{\sum y}{\nu} = \frac{77}{7} = 11$ και $\bar{x} = \frac{35}{7} = 5$, οπότε $\hat{\alpha} = 11 - 2 \cdot 5 = 1$.

Άρα η εξίσωση της ευθείας παλινδρόμησης είναι $y = 2x + 1$.

Παρατηρούμε ότι για δύο διαφορετικά σύνολα δεδομένων –πίνακας 6.5 και πίνακας 6.7– η ευθεία που περιγράφει καλύτερα τη σχέση μεταξύ των μεταβλητών X και Y είναι η $y = 2x + 1$. Κατασκευάζουμε λοιπόν τα διαγράμματα διασποράς για τα δύο σύνολα ζευγών των παρατηρήσεών μας, με σκοπό να αντλήσουμε περισσότερες πληροφορίες για τη σχέση μεταξύ των δύο μεταβλητών.

Αριθμός επιπλέον πτήσεων

(Σχ. 6.8)

Είναι φανερό από το διάγραμμα ότι στην πρώτη περίπτωση οι συντεταγμένες των σημείων (x_i, y_i) είναι σημεία της ευθείας $y = 2x + 1$, ενώ στη δεύτερη περίπτωση τα περισσότερα σημεία βρίσκονται γύρω από την ευθεία $y = 2x + 1$.

Σύμφωνα με αυτά καταλήγουμε στο συμπέρασμα ότι πρέπει να αναζητήσουμε ένα μέτρο που να μας δείχνει πόσο ισχυρή είναι η συγγένεια ή η συμμεταβολή μεταξύ των μεταβλητών X και Y.

Αυτή η σχέση μεταξύ των δύο μεταβλητών ονομάζεται **συσχέτιση** (correlation). Ο βαθμός γραμμικής συσχέτισης μετριέται από έναν καθαρό αριθμό που ονομάζεται **συντελεστής συσχέτισης*** (correlation coefficient) **και συμβολίζεται με r.**

Ο συντελεστής γραμμικής συσχέτισης δίνεται από τον τύπο:

$$r = \frac{v\sum xy - (\sum x)(\sum y)}{\sqrt{v\sum x^2 - (\sum x)^2} \bullet \sqrt{v\sum y^2 - (\sum y)^2}}$$

(6.4)

Ο συντελεστής συσχέτισης r μεταβάλλεται από –1 έως 1, δηλαδή $-1 \leq r \leq 1$

και έχει τις εξής ιδιότητες:

*Ο συντελεστής συσχέτισης ονομάζεται και συντελεστής Pearson.

- Το πρόσημό του φανερώνει, αν η συσχέτιση είναι θετική ή αρνητική και είναι το ίδιο με του συντελεστή παλινδρόμησης. Οι δύο μεταβλητές X και Y θα λέμε ότι είναι θετικά συσχετισμένες, όταν η αύξηση των τιμών της μιας έχει ως συνέπεια και την αύξηση των τιμών της άλλης, ενώ θα λέμε ότι είναι αρνητικά συσχετισμένες, όταν η αύξηση των τιμών της μιας έχει συνέπεια τη μείωση των τιμών της άλλης.
- Το μέτρο του φανερώνει πόσο ισχυρά συσχετισμένες είναι οι μεταβλητές.

Ισχυρή συσχέτιση εμφανίζουν οι μεταβλητές, όταν ο συντελεστής συσχέτισης πλησιάζει τις τιμές + 1 ή – 1.

Αν $r = 1$, τότε έχουμε τέλεια θετική συσχέτιση. Όλες οι παρατηρήσεις βρίσκονται επάνω στην ευθεία παλινδρόμησης, η οποία έχει θετική κλίση ($\hat{\beta} > 0$)

Αν $r = -1$, τότε έχουμε τέλεια αρνητική συσχέτιση. Όλες οι παρατηρήσεις βρίσκονται επάνω στην ευθεία παλινδρόμησης, η οποία έχει αρνητική κλίση ($\hat{\beta} < 0$)

Αν $r = 0$, τότε οι μεταβλητές δεν εμφανίζουν γραμμική συσχέτιση και λέμε ότι οι μεταβλητές είναι γραμμικά ασυσχέτιστες.

Αναλυτικότερα επισημαίνουμε τα εξής:

α) Όταν $0,8 \leq r < 1$ ή $-1 < r \leq -0,8$ λέμε ότι έχουμε πολύ ισχυρή συσχέτιση.
β) Όταν $0,7 \leq r < 0,8$ ή $-0,8 < r \leq -0,7$ λέμε ότι έχουμε ισχυρή συσχέτιση.
γ) Όταν $0,5 \leq r < 0,7$ ή $-0,7 < r \leq -0,5$ λέμε ότι έχουμε μέση συσχέτιση.
δ) Όταν $0,3 \leq r < 0,5$ ή $-0,5 < r \leq -0,3$ λέμε ότι έχουμε ασθενή συσχέτιση.
ε) Όταν $-0,3 \leq r < 0,3$ λέμε ότι δεν έχουμε συσχέτιση και
στ) Όταν $r = \pm 1$ λέμε ότι έχουμε τέλεια συσχέτιση.

Στο παράδειγμα 6.5 ο συντελεστής συσχέτισης είναι ίσος με:

$$r = \frac{7 \cdot 308 - 28 \cdot 63}{\sqrt{7 \cdot 140 - 28^2} \cdot \sqrt{7 \cdot 679 - 63^2}} = 1,$$

δηλαδή οι μεταβλητές X και Y έχουν τέλεια θετική συσχέτιση. Άρα όλα τα σημεία με συντεταγμένες (x_i, y_i) βρίσκονται πάνω στην ευθεία παλινδρόμησης $y = 2x + 1$, όπως φαίνεται και στο διάγραμμα διασποράς.

Για το δεύτερο σύνολο ζευγών του πίνακα 6.7 ο συντελεστής συσχέτισης είναι:

$$r = \frac{560}{\sqrt{280} \cdot \sqrt{4008,4}} \cong 0,53$$

Άρα τα δεδομένα μας εμφανίζουν μέση συσχέτιση.

6.5 Η σημασία της παλινδρόμησης και της συσχέτισης στη σύγχρονη επιχείρηση

Η παλινδρόμηση και η συσχέτιση είναι δύο πολύ χρήσιμες μέθοδοι για τη σύγχρονη επιχείρηση που συμβάλλουν σημαντικά στην επιχειρηματική πρόβλεψη, στον προγραμματισμό της επιχειρηματικής δράσης και στη λήψη ορθών επιχειρηματικών αποφάσεων. Χρησιμοποιούνται για την ποσοτική διερεύνηση των σχέσεων, οι οποίες υπάρχουν μεταξύ των διαφόρων οικονομικών μεγεθών και βοηθούν σημαντικά στη χάραξη της οικονομικής πολιτικής που ασκείται είτε από το κράτος είτε από ιδιωτικές επιχειρήσεις.

Η γραμμική παλινδρόμηση, όπως είδαμε, μελετά και προσδιορίζει μια γραμμική σχέση –εφόσον υπάρχει– μεταξύ των μεταβλητών των διμεταβλητών πληθυσμών. Με βάση τις τιμές της ανεξάρτητης μεταβλητής και τη σχέση $\hat{y} = \hat{\alpha} + \hat{\beta} \cdot x$ εκτιμούμε-προβλέπουμε με προσέγγιση τις αντίστοιχες τιμές της εξαρτημένης μεταβλητής. Η συσχέτιση είναι ένα μέτρο που φανερώνει, αν η εξάρτηση των μεταβλητών που εξετάζουμε είναι έντονη, μέτρια, ασθενής ή και μηδενική, καθώς και αν πρόκειται για μια θετική ή αρνητική συσχέτιση.

Στην πράξη, συνήθως προσδιορίζουμε την εξίσωση της ευθείας παλινδρόμησης και υπολογίζουμε τη συσχέτιση των δύο μεταβλητών, γιατί έτσι η εικόνα που προκύπτει από την από κοινού εξέταση των μεταβλητών αυτών είναι πιο ολοκληρωμένη.

Η παλινδρόμηση και η συσχέτιση είναι, ίσως, οι μέθοδοι εκτίμησης που εφαρμόζονται περισσότερο στις οικονομικές σχέσεις. Αυτό φαίνεται σε διάφορους κλάδους της οικονομίας, όπως στην οικονομετρία, στις τεχνικές έρευνας αγοράς κ.λπ. Καταλαβαίνουμε λοιπόν ότι αποτελούν ένα πολύ σημαντικό όργανο εργασίας και έρευνας για το σημερινό οικονομολόγο, όπως και για άλλους ερευνητές και επιστήμονες.

6.6 Εφαρμογές στις επιχειρήσεις

Εφαρμογή I.

Μελέτη προβλήματος πωλήσεων.
Περίπτωση μελέτης της Διεύθυνσης Πωλήσεων της Α.Ε. «Αστήρ – 3Ε».

Είναι γνωστό ότι ένα τμήμα των εμπορευμάτων που πωλούνται από τις επιχειρήσεις επιστρέφεται, σε μερικές περιπτώσεις, από τους αγοραστές στον πωλητή (επιστροφές εμπορευμάτων) 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 κτλ.

Η εμπορική επιχείρηση «Αστήρ – 3Ε» Α.Ε. είχε την περασμένη χρονιά πωλήσεις εμπορευμάτων αξίας 5.000.000 ευρώ και επιστροφές εμπορευμάτων αξίας 500.000 ευρώ. Ο νέος γενικός διευθυντής θέλοντας να εξακριβώσει αν 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 πίνακα 6.9:

Πίνακας 6.9

Αξία πωλήσεων (x)	20	30	40	40	50	50	60	70	80	90
Αξία επιστροφών (y)	1	3	3	4	5	6	6	8	9	10

(δεδομένα σε εκατοντάδες χιλιάδες ευρώ – 100.000 €).

Ο διευθυντής πωλήσεων θέλησε να διαπιστώσει:

1. Το βαθμό συσχέτισης που υπάρχει μεταξύ της αξίας των εμπορευμάτων που πωλούνται ετησίως και της αξίας αυτών που επιστρέφονται στην ομάδα των 10 επιχειρήσεων.
2. Την εξίσωση παλινδρόμησης που θα μπορούσε να εκφράσει τη σχέση που υπάρχει μεταξύ της αξίας των πωλήσεων και της αξίας των επιστροφών κατά έτος.
3. Ποια θα ήταν η αναμενόμενη αξία των εμπορευμάτων που επιστρέφο-

νται, αν οι πωλήσεις αυξάνονταν κατά μια μονάδα, δηλαδή 100.000 ευρώ;

4. Ποια θα ήταν η αξία των επιστροφών που θα είχε μια επιχείρηση, αν οι ετήσιες πωλήσεις της είχαν ύψος 35 εκατοντάδες χιλιάδες ευρώ;

5. Ποια η διάμεση αξία των πωλήσεων κατά έτος αυτής της ομάδας επιχειρήσεων;

6. Ποια η διάμεση αξία των επιστροφών κατά έτος;

7. Όπου κρίνεται απαραίτητο να γίνει το σχετικό διάγραμμα.

Λύση

1. Για να υπολογίσουμε τον συντελεστή γραμμικής συσχέτισης των μεταβλητών «αξία πωλήσεων» και «αξία επιστροφών» συμπληρώνουμε τον πίνακα των δεδομένων:

Πίνακας 6.10

Αξία πωλήσεων x	**Αξία επιστροφών y**	**x.y**	**x^2**	**y^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
$\sum x = 530$	$\sum y = 55$	$\sum xy = 3480$	$\sum x^2 = 32500$	$\sum y^2 = 377$

(σε εκατοντάδες χιλιάδες ευρώ)

αντικαθιστούμε στον τύπο υπολογισμού του συντελεστή συσχέτισης και έχουμε:

$$r = \frac{\nu\sum xy - \sum x\sum y}{\sqrt{\nu\sum x^2 - (\sum x)^2}\sqrt{\nu\sum y^2 - (\sum y)^2}} = \frac{10 \cdot 3480 - 530 \cdot 55}{\sqrt{10 \cdot 32500 - (530)^2}\sqrt{10 \cdot 377 - (55)^2}} =$$

$$= \frac{5650}{\sqrt{44100}\sqrt{745}} = \frac{5650}{210 \cdot 27{,}295} = \frac{5650}{5731{,}95} = 0{,}986$$

Παρατηρούμε ότι υπάρχει ισχυρή συσχέτιση μεταξύ των μεταβλητών (X) και (Y).

2. Θα υπολογίσουμε τις εκτιμήτριες ελαχίστων τετραγώνων $\hat{\alpha}$, $\hat{\beta}$ και θα προσδιορίσουμε την εξίσωση της ευθείας παλινδρόμησης.

Από τα στοιχεία του πίνακα υπολογίζουμε τον συντελεστή διεύθυνσης της ευθείας παλινδρόμησης:

$$\hat{\beta} = \frac{\nu\sum y - \sum xy}{\nu\sum x^2 - (\sum x)^2} = \frac{5650}{44100} = 0{,}128$$

επίσης υπολογίζουμε

$$\bar{x} = \frac{\sum x}{\nu} = \frac{530}{10} = 53 \quad \text{και} \quad \bar{y} = \frac{\sum y}{\nu} = \frac{55}{10} = 5{,}5$$

οπότε από $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$ προκύπτει ότι:

$\hat{\alpha} = 5{,}5 - 0{,}128 \cdot 53 \Leftrightarrow \hat{\alpha} = -1{,}29$

Άρα η ευθεία που προσαρμόζεται καλύτερα στα δεδομένα είναι:

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

Παρατηρούμε ότι, αν $\hat{y} = 0$, τότε $x \cong 10$ (γιατί αν $\hat{y} = 0$ τότε $-1{,}29 + 0{,}128x = 0 \Leftrightarrow 0{,}128x = 1{,}29 \Leftrightarrow x \cong 10$) που σημαίνει ότι για πωλήσεις εμπορευμάτων αξίας περίπου μέχρι και 1.000.000 ευρώ (10×100.000) δεν υπάρχουν επιστροφές εμπορευμάτων (υπενθυμίζουμε ότι τα δεδομένα είναι σε εκατοντάδες χιλιάδες ευρώ).

3. Αν οι πωλήσεις αυξάνονταν κατά μια μονάδα, τότε

$$\hat{y}^* = -1,29 + 0,128\,(x+1) = \underbrace{-1,29 + 0,128}_{\hat{y}} + 0,128 = \hat{y} + 0,128$$

Συμπεραίνουμε ότι, αν οι πωλήσεις αυξάνονταν κατά μια μονάδα, δηλαδή κατά 100.000 ευρώ, τότε το ύψος των επιστρεφόμενων εμπορευμάτων θα αυξανόταν κατά 0,128 εκατοντάδες χιλιάδες ευρώ, δηλαδή κατά 12.800 ευρώ όσο ακριβώς και η τιμή του συντελεστή $\hat{\beta}$, ($\hat{\beta}$ = 0,128).

4. Αν οι ετήσιες πωλήσεις μιας επιχείρησης ήταν της τάξης των 35 εκατοντάδων χιλιάδων ευρώ, τότε το ύψος των επιστρεφόμενων εμπορευμάτων θα υπολογιζόταν από τον τύπο:

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

$$\hat{y} = -1,29 + 0,128 \cdot 35 = 3,19$$

Δηλαδή, η αξία των εμπορευμάτων που επιστράφηκαν θα ήταν 319.000 ευρώ.

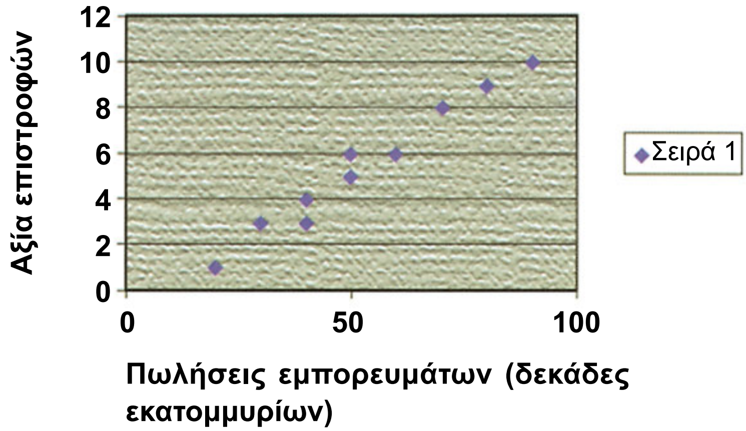
5. Η μέση αξία των εμπορευμάτων που πουλήθηκαν κατ' έτος από αυτή την ομάδα των επιχειρήσεων είναι 5.300.000 ευρώ ($\bar{x}$ = 53) και αντίστοιχα η μέση αξία των επιστροφών είναι 550.000 ($\bar{y}$ = 5,5). Οι συντεταγμένες του σημείου (53, 5,5) επαληθεύουν την ευθεία παλινδρόμησης.

6. Η διάμεση αξία των προϊόντων που επιστράφηκαν κατ' έτος υπολογίζεται ως εξής:

Αξία επιστροφών : 1, 3, 3, 4, **5**, **6**, 6, 8, 9, 10.

Ο αριθμός των επιχειρήσεων είναι 10 (άρτιος) άρα η διάμεσος θα βρίσκεται στη $\frac{10+1}{2} = 5,5$ θέση, δηλαδή ανάμεσα στην 5η και 6η παρατήρηση και θα ισούται με $\frac{5+6}{2} = 5,5$, δηλαδή η διάμεσος αξία των επιστροφών θα είναι 550.000 ευρώ.

Παρατηρούμε ότι η μέση τιμή της αξίας των εμπορευμάτων των 10 επιχειρήσεων που επιστρέφονται κατ' έτος είναι 550.000 ευρώ, αριθμός που συμπίπτει με τη διάμεσο της αξίας των εμπορευμάτων που επιστρέφονται. Παρόλα αυτά δεν μπορούμε να ισχυριστούμε ότι η κατανομή είναι συμμετρική, γιατί έχουμε δύο παρατηρήσεις (y = 3 και y = 6), που εμφανίζουν συχνότητα ίση με 2, η οποία είναι και η μεγαλύτερη, δηλαδή δεν έχουμε επικρατούσα τιμή.



ΑΝΑΚΕΦΑΛΑΙΩΣΗ

1. Σκοπός της παλινδρόμησης είναι η καλύτερη δυνατή περιγραφή της σχέσης που τυχόν υπάρχει μεταξύ δύο ή περισσότερων μεταβλητών, ενώ ο σκοπός της συσχέτισης είναι να μετρήσει το βαθμό εξάρτησης των μεταβλητών.

2. Ασχοληθήκαμε με τη μελέτη δύο μεταβλητών θεωρώντας τη **Χ ανεξάρτητη** και την **Υ εξαρτημένη μεταβλητή** από τη Χ και πιο συγκεκριμένα εξετάσαμε την **ευθύγραμμη παλινδρόμηση**.

3. Αν απεικονίσουμε τις τιμές των μεταβλητών πάνω σε ένα ορθοκανονικό σύστημα αξόνων, τότε τα ζεύγη των τιμών (x_i, y_i) των παρατηρήσεών μας δημιουργούν ένα νέφος σημείων, το οποίο ονομάζεται **διάγραμμα διασποράς ή στικτό διάγραμμα.** Το διάγραμμα διασποράς μας δίνει μια πρώτη εικόνα της σχέσης που υπάρχει μεταξύ των δύο μεταβλητών, ως προς το είδος και το βαθμό της εξάρτησης των μεταβλητών.

4. Με τη βοήθεια της μεθόδου των **ελαχίστων τετραγώνων** μπορούμε να προσαρμόσουμε σ' ένα διάγραμμα μια ευθεία ή καμπύλη γραμμή, η οποία θα περιγράφει κατά τον καλύτερο δυνατό τρόπο τη σχέση μεταξύ των Υ και Χ. Στο συγκεκριμένο κεφάλαιο ασχοληθήκαμε με την προσαρμογή μιας ευθείας γραμμής, δηλαδή της ευθείας παλινδρόμησης.

5. Οι τιμές των παραμέτρων $\hat{\alpha}$ και $\hat{\beta}$ σύμφωνα με τη μέθοδο των ελαχίστων τετραγώνων, προκύπτουν από τη λύση του παρακάτω συστήματος εξισώσεων, γνωστού ως συστήματος κανονικών εξισώσεων:

$$\sum y = \nu\hat{\alpha} + \hat{\beta}\sum x$$

$$\sum xy = \hat{\alpha}\sum x + \hat{\beta}\sum x^2$$

6. Στην εξίσωση **γραμμικής παλινδρόμησης** $\hat{y} = \hat{\alpha} + \hat{\beta} \cdot x$ η παράμετρος $\hat{\alpha}$ μας δείχνει το σημείο τομής του κάθετου άξονα y από την ευθεία παλινδρόμησης, δηλαδή η $\hat{\alpha}$ μας δίνει την τιμή της y για x = 0.

7. Η παράμετρος $\hat{\beta}$ προσδιορίζει τη μεταβολή που επέρχεται στην εξαρτημένη μεταβλητή Υ, όταν η Χ μεταβληθεί κατά μια μονάδα. Αν το $\hat{\beta}$ είναι θετικό ($\hat{\beta} > 0$), τότε η εξάρτηση που υπάρχει μεταξύ των δύο μεταβλητών είναι θετική, ενώ αν το $\hat{\beta}$ είναι αρνητικό ($\hat{\beta} < 0$), τότε η εξάρτηση είναι αρνητική.

8. **Οι τύποι υπολογισμού** των παραμέτρων $\hat{\alpha}$ και $\hat{\beta}$ της εξίσωσης παλινδρόμησης $\hat{y} = \hat{\alpha} + \hat{\beta} \cdot x$ είναι:

και $$\hat{\beta} = \frac{\nu \sum xy - \left(\sum x\right)\left(\sum y\right)}{\nu \sum x^2 - \left(\sum x\right)^2}$$

ή $$\hat{\beta} = \frac{\dfrac{\sum xy}{\nu} - \bar{x} \cdot \bar{y}}{\dfrac{\sum x^2}{\nu} - \bar{x}^2}$$

9. **Οι τιμές της εξαρτημένης μεταβλητής** που υπολογίζονται με τη βοήθεια της εξίσωσης παλινδρόμησης συμβολίζονται με $\hat{y}$ για να διακρίνονται από τις αντίστοιχες πραγματικές τιμές y, ενώ οι διαφορές $\hat{y}$ - y συμβολίζονται με $\varepsilon = \hat{y} - y$

10. Με το **συντελεστή γραμμικής συσχέτισης r** μετράμε το βαθμό εξάρτησης μεταξύ δύο εξεταζόμενων μεταβλητών, με την προϋπόθεση φυσικά, ότι η σχέση εξάρτησης έχει γραμμική μορφή.

11. Όσο πιο κοντά είναι οι παρατηρήσεις μας στην ευθεία παλινδρόμησης, τόσο πιο **έντονη είναι η συσχέτιση** που υπάρχει μεταξύ των δύο μεταβλητών.

12. **Ο συντελεστής συσχέτισης είναι ανεξάρτητος** από τις μονάδες μέτρησης που χρησιμοποιούμε για τις εξεταζόμενες μεταβλητές και είναι **καθαρός αριθμός,** δηλαδή δεν εκφράζεται σε συγκεκριμένες μονάδες μέτρησης.

13. Ο συντελεστής συσχέτισης r και ο συντελεστής παλινδρόμησης $\hat{\beta}$ **έχουν πάντοτε το ίδιο πρόσημο (+ ή -).** Η τιμή του συντελεστή συσχέτισης κυμαίνεται μεταξύ –1 και +1. Δηλαδή: $-1 \le r \le +1$

14. Αν r = +1, έχουμε **τέλεια θετική συσχέτιση.**
Αν r = -1, έχουμε **τέλεια αρνητική συσχέτιση.**
Αν r = 0, οι μεταβλητές είναι **γραμμικά ασυσχέτιστες.**

15. Αν οι μεταβλητές είναι μεταξύ τους ανεξάρτητες, τότε ο συντελεστής γραμμικής συσχέτισης r = 0. Αν όμως έχουμε r = 0 δεν σημαίνει κατ' ανάγκη ότι οι μεταβλητές είναι ανεξάρτητες. Απλά δεν υπάρχει μεταξύ τους γραμμική συσχέτιση.

16. **Ο τύπος υπολογισμού του συντελεστή συσχέτισης r** είναι:

$$r = \frac{\nu\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\nu\sum x^2 - \left(\sum x\right)^2} \cdot \sqrt{\nu\sum y^2 - \left(\sum y\right)^2}}$$

ΕΡΩΤΗΣΕΙΣ ΚΑΤΑΝΟΗΣΗΣ

1. Κατανομές που προκύπτουν από τη μελέτη ενός πληθυσμού ως προς δύο μεταβλητές ονομάζονται……………………………………………………

2. Αν απεικονίσουμε τις τιμές των μεταβλητών πάνω σε ένα κανονικό σύστημα αξόνων, τότε τα ζεύγη τιμών (x_i, y_i) των παρατηρήσεων δημιουργούν

.... , το οποίο ονομάζεται

διάγραμμα ....

3. Η ευθεία παλινδρόμησης δίνει την καλύτερη δυνατή περιγραφή ....
....

4. Η ευθεία παλινδρόμησης, η οποία προσδιορίζεται από τη μέθοδο των

ελαχίστων τετραγώνων, θα είναι η ευθεία για την οποία ....

....

5. Οι παράμετροι $\hat{\alpha}$ και $\hat{\beta}$ προσδιορίζονται από την επίλυση του συστήματος

των κανονικών εξισώσεων, που είναι οι εξής: ....

....

6. Να γράψετε τον τύπο που χρησιμοποιούμε για τον υπολογισμό της εκτιμήτριας $\hat{\alpha}$ της ευθείας παλινδρόμησης $\hat{y} = \hat{\alpha} + \hat{\beta} \cdot x$, καθώς και τους τύπους υπολογισμού της εκτιμήτριας $\hat{\beta}$.

7. Έστω $\hat{y} = \hat{\alpha} + \hat{\beta} \cdot x$ η εξίσωση της ευθείας παλινδρόμησης της μεταβλητής Y πάνω στη X.

- Αν $\hat{\beta} > 0$, τότε αυτό φανερώνει ότι, όταν το x αυξάνεται (ή μειώνεται), το

y...

- Αν $\hat{\beta} < 0$, τότε αυτό φανερώνει ότι, όταν το x αυξάνεται (ή μειώνεται), το

y...

8. Ο συντελεστής συσχέτισης φανερώνει ....

....

9. Ο τύπος υπολογισμού του συντελεστή συσχέτισης είναι:

r =...

10. Ο συντελεστής συσχέτισης παίρνει τιμές: ... ≤ r ≤ ...

11. Αν ο συντελεστής συσχέτισης μεταξύ των μεταβλητών X και Y είναι ίσος με 1, τότε λέμε ότι έχουν .... συσχέτιση, ενώ, αν είναι ίσος με –1, τότε λέμε ότι έχουν .... συσχέτιση.

Αν ο συντελεστής συσχέτισης είναι ίσος με 0, τότε οι μεταβλητές είναι

....

12. Αν η εκτιμήτρια ελαχίστων τετραγώνων $\hat{\beta}$ είναι:

- $\hat{\beta} > 0$, τότε ο συντελεστής συσχέτισης r

είναι ....

- $\hat{\beta} < 0$, τότε ο συντελεστής συσχέτισης r

είναι ....

ΑΣΚΗΣΕΙΣ

Ομάδα Α΄

1. Η αξία των επενδυμένων κεφαλαίων (x) και τα αντίστοιχα κέρδη (y) ομάδας 10 ομοειδών βιομηχανικών επιχειρήσεων περιέχονται στον παρακάτω πίνακα (τα δεδομένα εκφράζουν δεκάδες χιλιάδες ευρώ).

Κεφάλαιο (x)	80	90	100	120	150	180	200	250	300	350
Κέρδη (y)	15	20	25	25	30	30	40	40	60	80

Ζητείται:

i) Να βρεθεί ο συντελεστής γραμμικής συσχέτισης r, να σχολιασθεί το αποτέλεσμα και

ii) Να ερμηνευθεί η παράμετρος $\hat{\beta}$ της $\hat{y} = \hat{\alpha} + \hat{\beta} \cdot x$.

2. Η από κοινού μελέτη των μεταβλητών X και Y μας έδωσε, μετά από υπολογισμούς, τα παρακάτω αθροίσματα:

$\sum x = 30$ $\quad \sum y = 33$ $\quad \sum x^2 = 196$

$\sum xy = 210$ $\quad \sum y^2 = 227$ $\quad \nu = 6$

Ζητείται:

i) Να βρεθεί ο συντελεστής γραμμικής συσχέτισης r.

ii) Να βρεθεί η εξίσωση γραμμικής παλινδρόμησης $\hat{y} = \hat{\alpha} + \hat{\beta} \cdot x$.

3. Ο παρακάτω πίνακας δεδομένων δείχνει τα ετήσια εισοδήματα x σε εκατοντάδες ευρώ και τα αντίστοιχα έξοδα τηλεφωνικών λογαριασμών y σε εκατοντάδες ευρώ μιας ομάδας 7 νοικοκυριών:

x	300	350	400	450	500	550	600
y	9	8	10	13	13	16	15

Ζητείται:

i) Να βρεθεί ο συντελεστής γραμμικής συσχέτισης r.
ii) Να υπολογισθούν οι παράμετροι $\hat{\alpha}$ και $\hat{\beta}$ της εξίσωσης παλινδρόμησης

4. Ο παρακάτω πίνακας αναφέρεται σε 10 ομοειδείς εμπορικές επιχειρήσεις και δείχνει τις ετήσιες πωλήσεις τους (y) και τις αντίστοιχες ετήσιες διαφημιστικές τους δαπάνες (x), (τα δεδομένα εκφράζονται σε χιλιάδες ευρώ).

x	20	22	25	30	32	35	40	50	60	70
y	250	270	260	350	330	300	370	480	450	500

Ζητείται:

i) Να βρεθεί ο συντελεστής γραμμικής συσχέτισης r και να σχολιασθεί το αποτέλεσμα.
ii) Να προσδιορισθεί η εξίσωση γραμμικής παλινδρόμησης.

5. Ο παρακάτω πίνακας μας δείχνει τη μέση ετήσια κατά κεφαλή κατανάλωση y, σε κιλά, ενός αγαθού, σε σχέση με το κατά κεφαλή εισόδημα x σε εκατοντάδες ευρώ μιας ομάδας ατόμων:

x	65	68	70	73	75	78	85	90
y	40	40	41	43	45	48	50	50

Ζητείται:

i) Να βρεθεί και να σχολιασθεί η τιμή του συντελεστή γραμμικής συσχέτισης.
ii) Να προσδιορισθεί η εξίσωση γραμμικής παλινδρόμησης $\hat{y} = \hat{\alpha} + \hat{\beta} \cdot x$.
iii) Να ερμηνευθούν οι τιμές των παραμέτρων $\hat{\alpha}$ και $\hat{\beta}$ της εξίσωσης παλινδρόμησης.
iv) Να βρεθούν οι διαφορές μεταξύ των πραγματικών τιμών της εξαρτημένης μεταβλητής y και των αντίστοιχων τιμών που μας δίνει η $\hat{y} = \hat{\alpha} + \hat{\beta} \cdot x$.

6. Πέντε άτομα μετέχουν σε πρόγραμμα ελάττωσης βάρους. Ο αριθμός των εβδομάδων (x) που κάθε άτομο συμμετέχει στο πρόγραμμα, καθώς και τα κιλά (y) που έχασε στο αντίστοιχο διάστημα εμφανίζονται στον παρακάτω πίνακα:

Αριθμός εβδομάδων (x)	3	2	1	4	5
Ελάττωση βάρους σε κιλά (y)	3	2,5	2	4,5	5,5

Ζητείται:

i) Να βρείτε την εξίσωση της ευθείας των ελαχίστων τετραγώνων.
ii) Πόσες εβδομάδες πρέπει κάποιος να συμμετέχει στο πρόγραμμα για να χάσει 10 κιλά;

Ομάδα Β΄

1. Ένας ερευνητής μελετά τον αριθμό των ηλιακών κηλίδων(x) και τον αριθμό ατυχημάτων(y) σε μια χώρα για πέντε συνεχή χρόνια. Σύμφωνα με τη θεωρία του ο αριθμός των ηλιακών κηλίδων συνδέεται με κάποιο τρόπο με τον αριθμό των ατυχημάτων. Οι παρατηρήσεις του φαίνονται στον πίνακα που ακολουθεί:

Αριθμός ηλιακών κηλίδων (x)	3	1	2	5	4
Αριθμός ατυχημάτων σε χιλιάδες (y)	9	5	7	14	10

Ζητείται:

i) Να κατασκευαστεί το διάγραμμα διασποράς των δύο μεταβλητών.
ii) Μπορεί στο διάγραμμα αυτό να προσαρμοστεί ευθεία γραμμή; Αν ναι, να προσδιοριστεί η εξίσωσή της.
iii) Να βρεθεί ο συντελεστής συσχέτισης. Τι παρατηρείτε;
iv) Αν δεν παρατηρηθούν ηλιακές κηλίδες, πόσα ατυχήματα προβλέπετε ότι θα συμβούν;

2. Μία εταιρεία έκανε μελέτη της ποσότητας καλαμποκιού που παράγεται από μια συγκεκριμένη καλλιεργήσιμη έκταση σε σχέση με την ποσότητα λιπά-

σματος που χρησιμοποιήθηκε για την παραγωγή. Τα αποτελέσματα φαίνονται στον πίνακα που ακολουθεί:

Ποσότητα λιπάσματος (x) σε κιλά	0,3	0,6	0,9	1,2	1,5	1,8	2,1	2,4
Ποσότητα καλαμποκιού (y) σε κιλά	10	15	30	35	25	30	50	45

Ζητείται:

i) Να παρασταθεί γραφικά η εξαρτημένη μεταβλητή σαν συνάρτηση της ανεξάρτητης μεταβλητής.
ii) Να βρεθεί η ευθεία ελαχίστων τετραγώνων για τα συγκεκριμένα δεδομένα.
iii) Να βρεθεί η ποσότητα καλαμποκιού, αν δεν χρησιμοποιηθεί λίπασμα.
iv) Αν είχαν χρησιμοποιηθεί 3 κιλά λίπασμα, ποια θα είναι η καλύτερη εκτίμηση της ποσότητας καλαμποκιού που θα είχε παραχθεί;

3. Μία εταιρεία θέλει να προβλέψει την απόδοση των πωλητών που εκπαιδεύει. Στην αρχή της διμηνιαίας εκπαίδευσης 10 υποψήφιοι πωλητές απάντησαν σε έλεγχο δεξιοτήτων. Οι επιδόσεις τους παρουσιάζονται στον παρακάτω πίνακα. Η εταιρεία κράτησε αρχεία των πωλήσεων κάθε υπαλλήλου (μεταβλητή Υ) σε ορισμένο χρονικό διάστημα.

Επίδοση στον έλεγχο (x)	18	26	28	34	36	42	48	52	54	60
Αριθμός πωλήσεων (y)	54	64	54	62	68	70	76	66	76	74

Ζητείται:

i) Να παρασταθούν γραφικά τα δεδομένα του πίνακα και να βρεθεί η ευθεία ελαχίστων τετραγώνων.
ii) Ποιος ο αριθμός των πωλήσεων που περιμένετε από 3 πωλητές, των οποίων τα αποτελέσματα στον έλεγχο δεξιοτήτων ήταν 40, 50 και 70 αντίστοιχα;

ΦΥΛΛΟ ΑΥΤΟΕΛΕΓΧΟΥ

1. Στη γραμμική παλινδρόμηση η σχέση μεταξύ της ανεξάρτητης μεταβλητής και της εξαρτημένης είναι αυτή της:

I. Ευθείας γραμμής. ❒
II. Καμπύλης γραμμής. ❒

2. Ο συντελεστής γραμμικής παλινδρόμησης $\hat{\beta}$ μας δείχνει:

I. Την ποσοστιαία μεταβολή που επέρχεται στην εξαρτημένη μεταβλητή από μεταβολή της ανεξάρτητης κατά 1%. ❒
II. Την ποσοστιαία μεταβολή που επέρχεται στην εξαρτημένη μεταβλητή από μεταβολή της ανεξάρτητης κατά 10%. ❒
III. Τη μεταβολή που επέρχεται στην εξαρτημένη μεταβλητή από μοναδιαία μεταβολή της ανεξάρτητης μεταβλητής. ❒

3. Ο συντελεστής γραμμικής παλινδρόμησης $\hat{\beta}$ έχει πρόσημο:

I. Πάντα θετικό. ❒
II. Πάντα αρνητικό. ❒
III. Είτε θετικό, είτε αρνητικό. ❒

4. Στην εξίσωση παλινδρόμησης συμβολίζουμε με $\hat{y}$

I. Τις πραγματικές τιμές της εξαρτημένης μεταβλητής. ❒
II. Τις «θεωρητικές» τιμές της εξαρτημένης μεταβλητής, που μας δίνει η εξίσωση παλινδρόμησης. ❒
III. Και τις πραγματικές και τις θεωρητικές τιμές της εξαρτημένης μεταβλητής. ❒
IV. Τις τιμές της ανεξάρτητης μεταβλητής. ❒

5. Η παράμετρος της γραμμικής παλινδρόμησης δίνεται από τον τύπο:
I. $\hat{\alpha} = y - \hat{\beta} x$ ❒
II. $\hat{\alpha} = \overline{y} - \hat{\beta} \overline{x}$ ❒
III. $\hat{\alpha} = \hat{y} - \hat{\beta} x$ ❒
IV. $\hat{\alpha} = y - \hat{\beta} \overline{x}$ ❒

6. Αν στην εξίσωση γραμμικής παλινδρόμησης είναι το x = 0, τότε θα έχουμε:

I. $\hat{y} = \hat{\beta}$ ❒
II. $\hat{y} = \hat{\alpha}$ ❒
III. $\hat{y} = -\beta$ ❒

7. Η παλινδρόμηση και η συσχέτιση χρησιμοποιούνται για τη μελέτη:

I. Μόνο μονομεταβλητών πληθυσμών. ❒
II. Μόνο διμεταβλητών πληθυσμών. ❒
III. Μόνο πολυμεταβλητών πληθυσμών. ❒
IV. Είτε διμεταβλητών είτε πολυμεταβλητών πληθυσμών. ❒

8. Ο συντελεστής γραμμικής συσχέτισης παίρνει τιμές:

I. Από 0 έως 1. ❒
II. Από -1 έως 0. ❒
III. Από -1 έως 1. ❒
IV. Από $-\infty$ έως $+\infty$ ❒

9. Ο συντελεστής γραμμικής συσχέτισης r και ο συντελεστής παλινδρόμησης $\hat{\beta}$ έχουν:

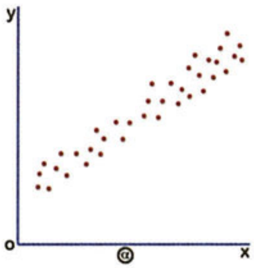
I. Πάντοτε το ίδιο πρόσημο. ❒
II. Πάντοτε διαφορετικό πρόσημο. ❒
III. Άλλοτε το ίδιο και άλλοτε διαφορετικό πρόσημο. ❒

10. Για ένα σύνολο δεδομένων οι εκτιμήτριες ελαχίστων τετραγώνων της μεταβλητής Y πάνω στη X είναι $\hat{\alpha} = -2$ και $\hat{\beta} = 1$.

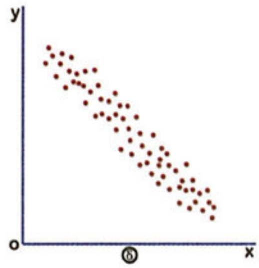
Ποια τιμή παίρνει η Y, όταν η τιμή της X γίνει ίση με 0;

Α: 1 ❒ Β:0 ❒ Γ: -2 ❒ Δ: 2 ❒

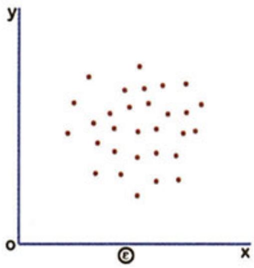
11. Η σχέση που υπάρχει μεταξύ των μεταβλητών είναι:



I. Θετική. ❒
II. Αρνητική. ❒
III. Δεν υπάρχει συσχέτιση. ❒



I. Θετική. ❒
II. Αρνητική. ❒
III. Δεν υπάρχει συσχέτιση. ❒



I. Θετική. ❒
II. Αρνητική. ❒
III. Δεν υπάρχει συσχέτιση. ❒

12. Ποιος από τους παρακάτω συντελεστές συσχέτισης αντιστοιχεί στην ισχυρότερη συσχέτιση;

Α. 0,5 ❒ Β. $\frac{\sqrt{4}}{2}$ ❒ Γ. $\frac{1}{3}$ ❒ Δ. $\frac{5}{1000}$ ❒ Ε. $\frac{-99}{1000}$ ❒

13. Αν για ένα σύνολο δεδομένων οι εκτιμήτριες ελαχίστων τετραγώνων είναι $\hat{\alpha} = \frac{2}{3}$ και $\hat{\beta} = -3$, τότε πόσο μεταβάλλεται το y, αν το x μεταβληθεί κατά 2 μονάδες;

Α. αυξάνεται κατά 2 ❒
Β. αυξάνεται κατά 3 ❒
Γ. μειώνεται κατά 3 ❒
Δ. μειώνεται κατά 2/3 ❒
Ε. μειώνεται κατά 6 ❒

14. Αν η ευθεία ελαχίστων τετραγώνων της μεταβλητής Υ πάνω στη Χ έχει εξίσωση y = 2x - 10, τότε ο συντελεστής συσχέτισης μεταξύ των μεταβλητών Χ και Υ είναι:

Α = 2 ❒ Β = -10 ❒ Γ = -1 ❒ Δ = θετικός ❒ Ε = αρνητικός ❒

ΚΕΦΑΛΑΙΟ 7ο

ΧΡΟΝΟΛΟΓΙΚΕΣ ΣΕΙΡΕΣ

❖ Μελέτες στις χρονολογικές σειρές έγιναν για πρώτη φορά στις ΗΠΑ αμέσως μετά τον Α΄ Παγκόσμιο Πόλεμο. Ο άνθρωπος που θεωρείται γενικά σαν πρωτεργάτης της μελέτης των χρονολογικών σειρών είναι ο Αμερικανός οικονομολόγος – στατιστικός Warren M. Pursons, επικεφαλής της επιτροπής οικονομικών ερευνών στο Πανεπιστήμιο του Harvard, ο οποίος δημοσίευσε το 1919 σχετικό άρθρο στο οικονομικό περιοδικό «The review of economic statistics».

❖ *«Με τη στατιστική μπόρεσα, όπως ο κάθε οικονομολόγος, να θεμελιώσω τις θεωρίες μου· και τούτο, διότι η στατιστική παρέχει στον οικονομολόγο τα γεγονότα, με τα οποία στηρίζει τις θεωρίες του ή αποδεικνύει αυτές»*

Alfred Marshall (1842-1924)

ΚΕΦΑΛΑΙΟ 7ο

ΧΡΟΝΟΛΟΓΙΚΕΣ ΣΕΙΡΕΣ

7.1 Γενικά
Α. Έννοια και χρησιμότητα

Σήμερα, περισσότερο από άλλοτε, έχει καταστεί αναγκαία προϋπόθεση για την επιτυχή λειτουργία και επιβίωση μιας επιχείρησης, ενός οργανισμού ή ενός κράτους η σύνταξη οικονομικών προγραμμάτων δράσης τους. Δηλαδή, μια επιχείρηση, με βάση τα δεδομένα των προηγούμενων ετών, προϋπολογίζει τα έξοδα και τα έσοδα που θα έχει, όπως και άλλα οικονομικά μεγέθη, για το επόμενο έτος, την επόμενη πενταετία ή και δεκαετία. Το ίδιο κάνει και μία τράπεζα, ένας ασφαλιστικός οργανισμός ή ένα κράτος.

Όλο και περισσότεροι οικονομολόγοι σήμερα, δέχονται ότι το παιχνίδι του ανταγωνισμού έχει πλέον μεταφερθεί σε μεγάλο βαθμό από το παρόν στο μεσοπρόθεσμο και στο μακροπρόθεσμο πεδίο ανταγωνισμού. Οι επιχειρήσεις, οι οργανισμοί, αλλά και τα σύγχρονα κράτη που θέλουν να επιβιώσουν και να έχουν προοπτική οικονομικής ανάπτυξης, οφείλουν να μελετούν το παρελθόν και να καταγράφουν το παρόν, ώστε να μπορούν να προβλέπουν το μέλλον.

Φυσικά τα παραπάνω δεν αποτελούν κάτι νέο. Τα στοιχειωδώς οργανωμένα κράτη, εδώ και πολλά χρόνια, προβαίνουν σε προϋπολογισμούς πενταετούς ή και δεκαετούς διάρκειας. Ένα κράτος, αν δεν προβλέψει τον αριθμό των μαθητών για την επόμενη πενταετία ή και δεκαετία, θα αντιμετωπίσει στη συνέχεια σωρεία προβλημάτων, όπως προβλήματα υλικοτεχνικής υποδομής, προσωπικού κτλ. Παρόμοια προβλήματα θα συμβούν, αν ένα κράτος δεν προβλέψει τον αριθμό των οχημάτων που κυκλοφορούν, ιδιαίτερα στις μεγαλουπόλεις, όπου υπάρχει οξύ κυκλοφοριακό πρόβλημα ή αν δεν φροντίσει να βρει κονδύλια για επενδύσεις σε οδικά δίκτυα κτλ. Όλα αυτά επηρεάζουν δυσμενώς και σε μεγάλο βαθμό την ομαλή οικονομική λειτουργία μιας χώρας και αποτελούν τροχοπέδη για την οικονομική της ανάπτυξη.

Ένα από τα σημαντικά εργαλεία που χρησιμοποιούν σήμερα οι οικονομολόγοι, για να κάνουν ικανοποιητικές προβλέψεις στα οικονομικά μεγέθη και έτσι να δημιουργούν έναν επιτυχημένο οικονομικό προγραμματισμό, είναι η μελέτη και ανάλυση των χρονολογικών σειρών (time series analysis).

Στο ερώτημα **τι είναι χρονολογικές σειρές (time series), πολύ απλά θα απαντούσαμε** ότι είναι αυτό που αφήνει να εννοηθεί η ίδια η ονομασία τους, δηλαδή, **μια σειρά από τιμές που παίρνει μια μεταβλητή που εξετάζουμε στη διάρκεια του χρόνου και συνήθως σε ίσα διαδοχικά χρονικά διαστήματα.** Για παράδειγμα, τα ετήσια έσοδα μιας επιχείρησης τα τελευταία 10 χρόνια, ο αριθμός των ετήσιων τροχαίων ατυχημάτων την τελευταία δεκαπενταετία κτλ.

Συμπερασματικά λοιπόν μπορούμε να πούμε ότι **ο βασικός σκοπός της μελέτης και της περιγραφής της διαχρονικής εξέλιξης** (δηλαδή, εξέλιξης στη διάρκεια του χρόνου) των τιμών **μιας μεταβλητής είναι: η απόκτηση επαρκών γνώσεων από την επιστημονική μελέτη της συμπεριφοράς της εξεταζόμενης μεταβλητής στο παρελθόν, ώστε να κάνουμε μια ικανοποιητικού βαθμού πρόβλεψη για τη συμπεριφορά της στο μέλλον.**

Β. Χρονολογικά διαγράμματα

Στο κεφάλαιο αυτό θα ασχοληθούμε με τις βασικές στατιστικές μεθόδους που χρησιμοποιούνται συνήθως στην πράξη για τη μελέτη των χρονολογικών σειρών και θα δούμε πώς μπορούμε να κάνουμε προβλέψεις για το μέλλον εξετάζοντας και αναλύοντας κυρίως οικονομικού χαρακτήρα χρονολογικά δεδομένα. Οι χρονολογικές σειρές που αναφέρονται στη διαχρονική εξέλιξη ενός οικονομικού φαινομένου ονομάζονται οικονομικές χρονολογικές σειρές (economic time series).

Τη μεταβλητή που μελετάμε, την οποία θεωρούμε συνάρτηση του χρόνου, θα τη συμβολίζουμε με Y και τη μεταβλητή χρόνος με t. Όπως γίνεται φανερό, οι παρατηρήσεις μας θα αποτελούνται από ζεύγη τιμών της μορφής (t_1, y_1), (t_2, y_2) … (t_i, y_i), … , (t_ν, y_ν).

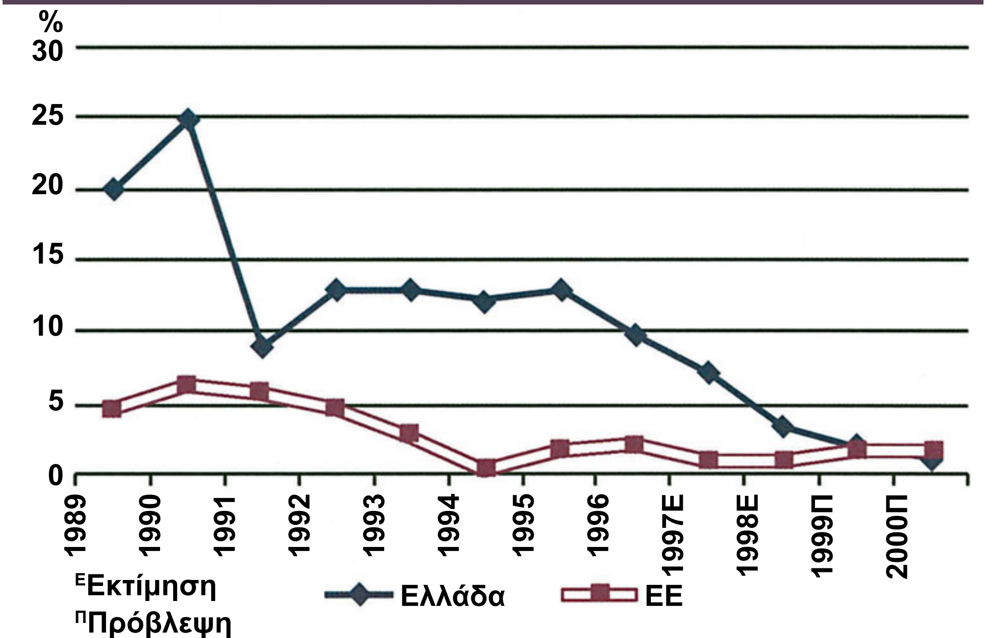
Η γραφική απεικόνιση των χρονολογικών σειρών γίνεται με τη βοήθεια των **χρονολογικών διαγραμμάτων ή χρονοδιαγραμμάτων** (time series diagrams).

Τα χρονοδιαγράμματα είναι πολύ απλά στην κατασκευή τους. Τον οριζόντιο άξονα τον χρησιμοποιούμε πάντοτε για τη μέτρηση της μεταβλητής χρόνος t και τον κάθετο για τις τιμές y_i της εκάστοτε εξεταζόμενης μεταβλητής Y. Φυσικά οι παρατηρήσεις μας αντιστοιχίζονται με τα προκύπτοντα ζεύγη τιμών (t_i, y_i).

Διαγράμματα τέτοιας μορφής και μάλιστα οικονομικά χρονοδιαγράμματα είναι τα 7.1, 7.2, 7.3, 7.4 και 7.6.

Ονομαστικό Κόστος Εργασίας ανά Μονάδα Προϊόντος 1989-2000 (Ετήσια αύξηση): Ελλάδα – Ευρωπαϊκή Ένωση

Διάγραμμα Α3.1: Ονομαστικό Κόστος Εργασίας ανά Μονάδα Προϊόντος 1989-2000 (Ετήσια αύξηση): Ελλάδα - Ευρωπαϊκή Ένωση



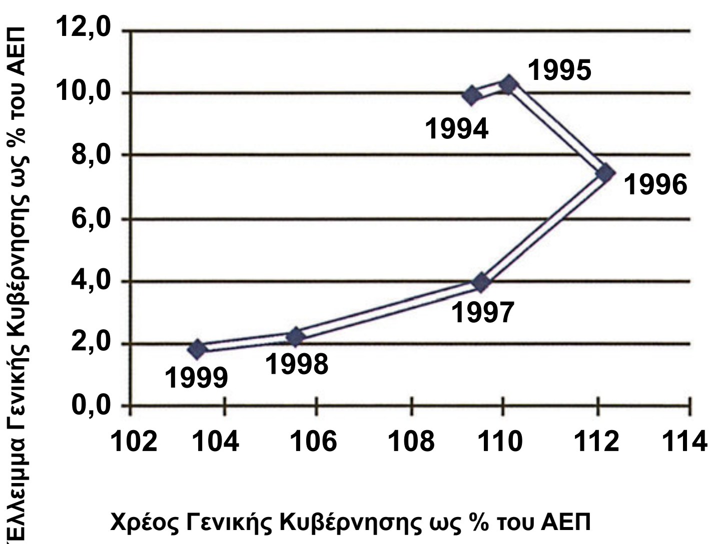
Πηγή: Eurostat και DGII

(Σχ. 7.1) Οικονομικό χρονοδιάγραμμα

Τα χρονολογικά διαγράμματα χρησιμοποιούνται καθημερινά, γιατί αποτελούν ένα πολύ καλό διαφημιστικό μέσο των επιχειρήσεων για την επιρροή των καταναλωτών. Οι επιχειρήσεις θέλουν με τρόπο απλό, σαφή και σύντομο, να δείξουν, για παράδειγμα, τη διαχρονική αύξηση των πωλήσεών τους (υπονοώντας την αυξανόμενη από έτος σε έτος προτίμηση των καταναλωτών στα προϊόντα τους) ή να δείξουν τη διαχρονική αυξητική εξέλιξη της μετοχής τους στο Χρηματιστήριο Αξιών Αθηνών (Χ.Α.Α.). Ακόμη και το ίδιο το Χ.Α.Α. με τη βοήθεια χρονοδιαγραμμάτων δείχνει την εξέλιξη του γενικού δείκτη του ή των επιμέρους κλαδικών χρηματιστηριακών δεικτών του (τραπεζών, ασφαλειών, κατασκευαστικών εταιρειών, κ.ά.). Σήμερα με τη βοήθεια των ηλεκτρονικών υπολογιστών (Η/Υ) κατασκευάζουμε χρονοδιαγράμματα πιο εντυπωσιακά από τα κλασσικά, τα οποία προσελκύουν περισσότερο την προσοχή του καταναλωτή. Τέτοια είναι τα χρονοδιαγράμματα 7.8, 7.9.

Εξέλιξη Ελλείμματος και Χρέους Γενικής Κυβέρνησης ως % του ΑΕΠ

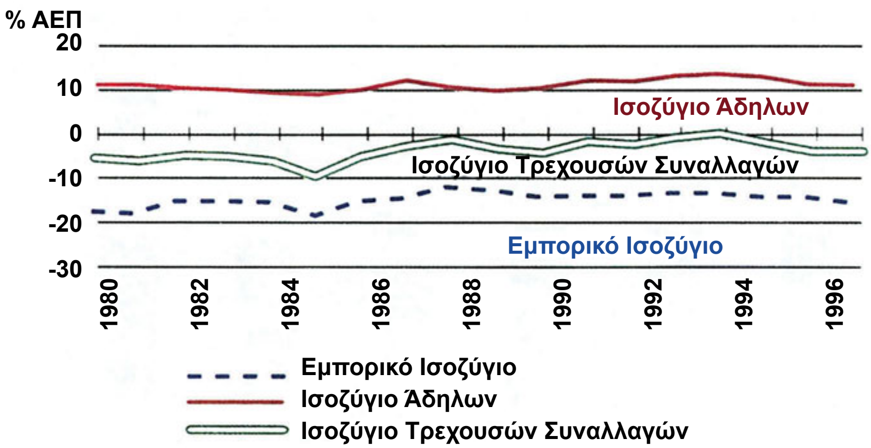
Διάγραμμα Α7.3: Εξέλιξη Ελλείμματος και Χρέους Γενικής Κυβέρνησης ως % του ΑΕΠ



Πηγή: Εξαμηνιαία έκθεση ΥΠΕΘΟ, Δεκ. 1998

(Σχ. 7.2) Οικονομικό χρονοδιάγραμμα

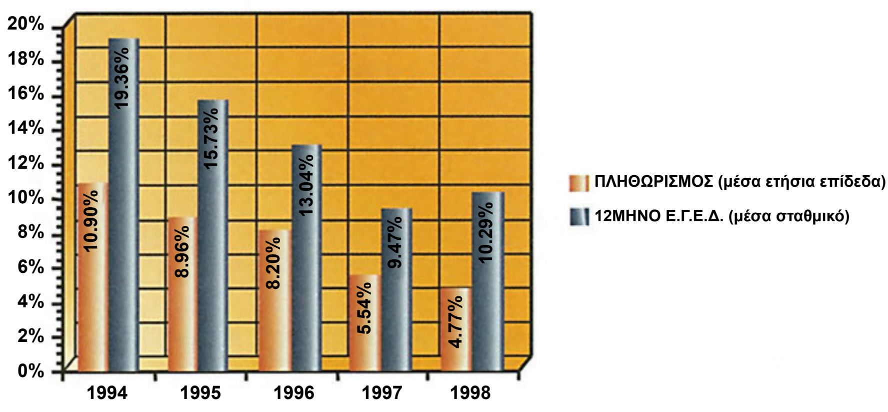
Ισοζύγιο Πληρωμών 1980-1997%



(Σχ. 7.3) Οικονομικό χρονοδιάγραμμα ***Πηγή:*** *Τράπεζα της Ελλάδος*

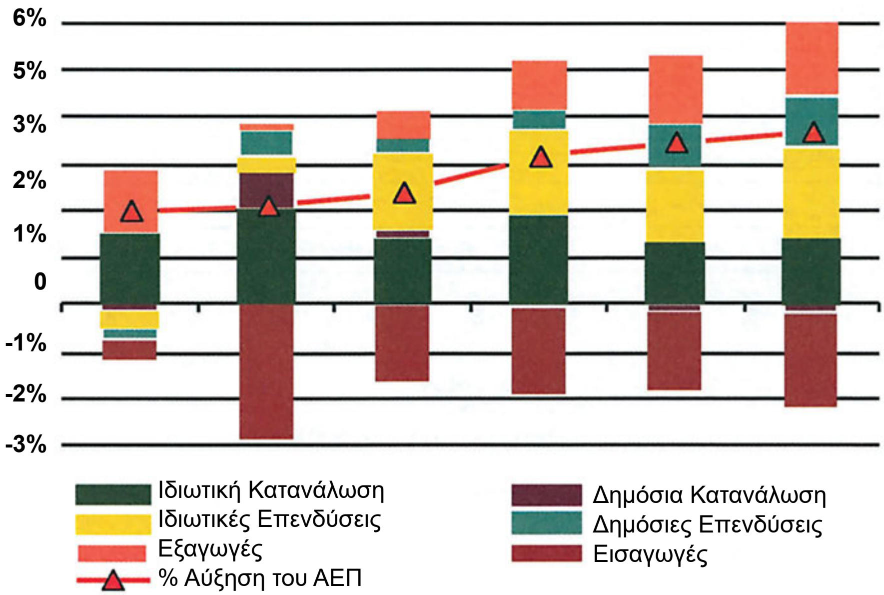
Συγκριτική Πορεία Πληθωρισμού – Επιτοκίου 12μηνου Ε.Γ.Ε.Δ. (1994-1998)

(Ε.Γ.Ε.Δ. = Έντοκα Γραμμάτια Ελληνικού Δημοσίου)



(Σχ. 7.4) Οικονομικό χρονοδιάγραμμα ***Πηγή:*** *Ε.Τ.Ε*

Συμβολή στην αύξηση του ΑΕΠ



Πηγή: ΥΠΕΘΟ, υπολογισμοί Ε.Τ.Ε.

(Σχ. 7.5) Χρονοδιάγραμμα

7.2 Συνιστώσες χρονολογικών σειρών

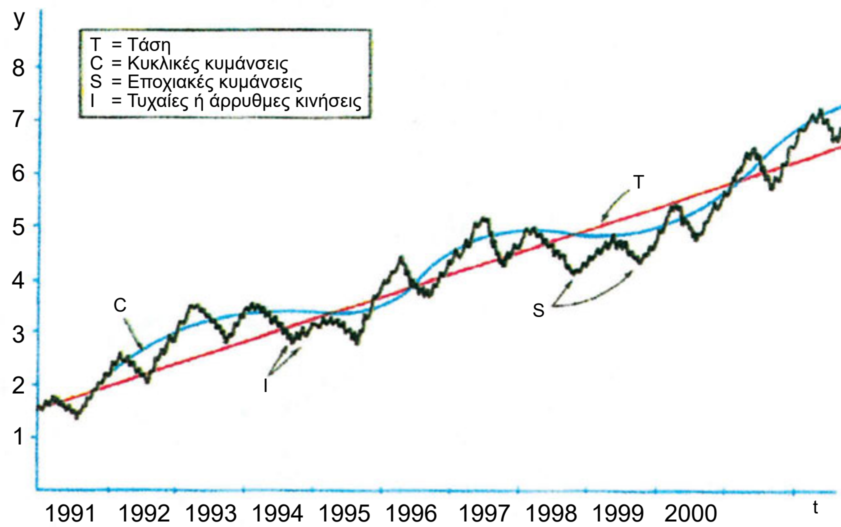
Η διεξοδική μελέτη των χρονολογικών σειρών, δηλαδή η συστηματική παρακολούθηση της διαχρονικής εξέλιξης ενός φαινομένου, μας αποκάλυψαν ότι οι τιμές μιας μεταβλητής σχηματίζονται κυρίως από την επίδραση τεσσάρων συνιστωσών (time – series components). Αυτές είναι:

Α. Η μακροχρόνια τάση ή απλά τάση (secural trend), που συμβολίζεται με **Τ**.

Β. Οι κυκλικές κυμάνσεις (cyclical fluctuactions), που συμβολίζονται με **C.**

Γ. Οι εποχικές κυμάνσεις (seasonal fluctuactions), που συμβολίζονται με **S.**

Δ. Οι τυχαίες ή άρρυθμες κινήσεις (random or irregular movements), που συμβολίζονται με **I**.



(Σχ. 7.6) Οικονομικό χρονοδιάγραμμα

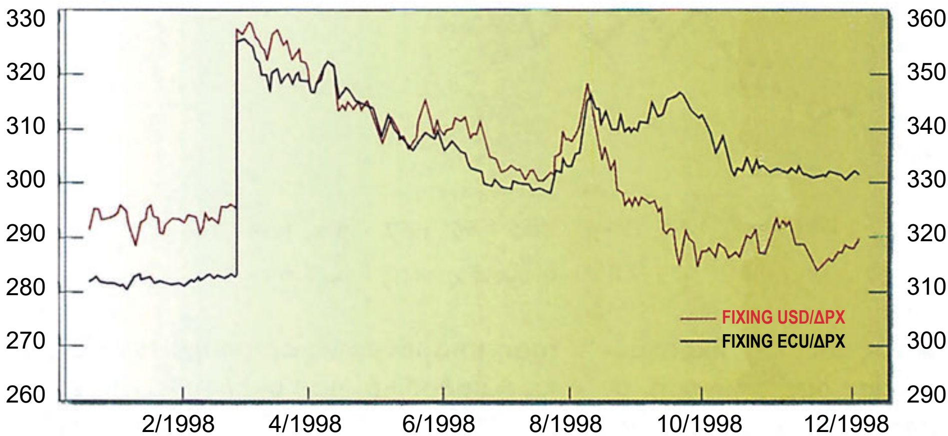
Α. Ως τάση ή μακροχρόνια τάση μπορούμε να ορίσουμε την μεγάλης διάρκειας ομαλή κίνηση, ανοδική ή καθοδική, που ακολουθεί μια χρονολογική σειρά. Πιο απλά, θα μπορούσαμε να πούμε ότι η τάση αποτελεί μια μακροχρόνια δυναμική κίνηση που οδηγεί τα χρονολογικά δεδομένα προς μία κατεύθυνση ανοδική ή καθοδική. Αν μία χρονολογική σειρά κινείται παράλληλα προς τον οριζόντιο άξονα του χρόνου t, τότε είναι φανερό, ότι αυτή θα είναι μηδενικής τάσης.

Οι αιτίες που επηρεάζουν αυξητικά την τάση μιας χρονολογικής σειράς είναι πολλές, όπως η τεχνολογική πρόοδος που οδηγεί σε αύξηση της παραγωγής προϊόντων, η αύξηση του πληθυσμού, η αύξηση του εισοδήματος του πληθυσμού κ.ά. Υπάρχουν όμως και αιτίες που επηρεάζουν μειωτικά μία τάση, όπως η εμφάνιση ενός νέου προϊόντος που αντικαθιστά ένα παλαιό, η μείωση του πληθυσμού μιας χώρας κ.ά.

Β. Ως κυκλικές κινήσεις, κυμάνσεις, ορίζουμε τις περιοδικές κινήσεις, των οποίων η χρονική διάρκεια είναι μεγαλύτερη από ένα έτος και εμφανίζουν για κάποιο χρονικό διάστημα κίνηση ανοδική (περίοδος αυξήσεων) και ακολούθως καθοδική (περίοδος μειώσεων). Οι κυκλικές διακυμάνσεις είναι συνηθισμένο φαινόμενο στο χώρο των οικονομικών επιστημών και συνδέονται συνήθως με τη βιομηχανική παραγωγή, την αγροτική παραγωγή, την απασχόληση και γενικά με την όλη οικονομική δραστηριότητα. Σε όλους μας είναι γνωστές οι περίοδοι υφέσεων και ανόδου της οικονομίας.

Η αξιολόγηση μιας χρονολογικής σειράς στην πράξη είναι ιδιαίτερα δύσκολη. Απαιτούνται παρατηρήσεις πολλών ετών, ενώ το μήκος της περιόδου των κυκλικών κυμάνσεων δεν είναι σταθερό και μερικές φορές δύσκολα ξεχωρίζει κανείς το καθοδικό από το ανοδικό τμήμα του κύκλου.

ΠΟΡΕΙΑ ΙΣΟΤΙΜΙΩΝ FIXING ECU/ΔΡΧ & USD/ΔΡΧ 1998

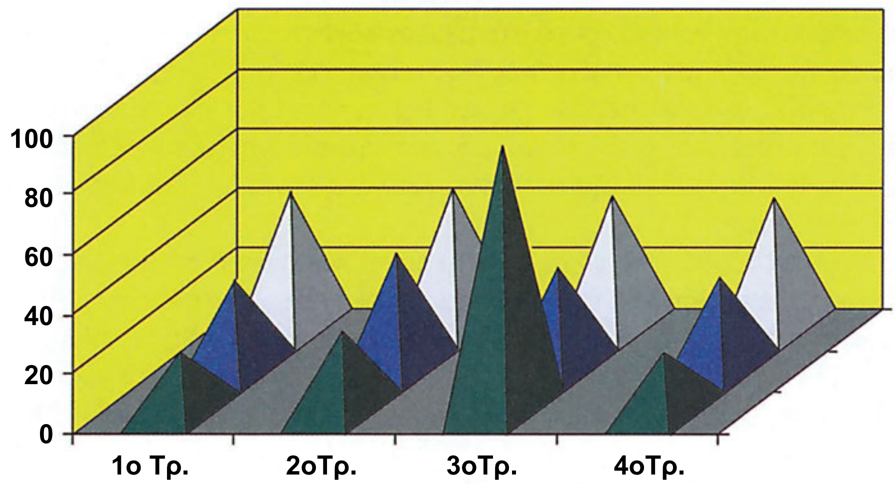


(Σχ. 7.7) Χρονοδιάγραμμα ***Πηγή: ΥΠΕΘΟ*** *1999*

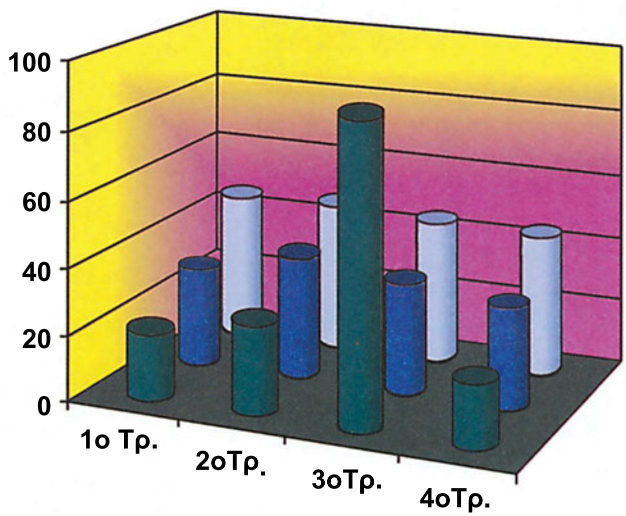
Γ. Ως εποχικές κυμάνσεις ορίζουμε τις περιοδικές κινήσεις που εξαντλούν μέσα σ' ένα έτος όλες τις ανοδικές ή καθοδικές τάσεις τους. Οι εποχικές κυμάνσεις είναι από τα πιο συνηθισμένα φαινόμενα στην οικονομία. Όλοι γνωρίζουμε ότι σε δύο συγκεκριμένες περιόδους το χρόνο έχουμε αύξηση των πωλήσεων λόγω εκπτώσεων και ότι τα γεωργικά προϊόντα προσφέρονται σε υπερεπάρκεια σε ορισμένες χρονικές περιόδους του έτους.

Για να βγάλουμε κάποια συμπεράσματα από τις εποχικές κυμάνσεις απαιτείται συνήθως η χρησιμοποίηση μηνιαίων ή τριμηνιαίων δεδομένων.

Δ. Τέλος έχουμε τις τυχαίες ή άρρυθμες κινήσεις, οι οποίες δεν παρουσιάζουν καμία κανονικότητα ή κάποιο ρυθμό. Αυτές τις κινήσεις μπορούμε να τις διακρίνουμε σε δύο κατηγορίες: i) Στις συμπτωματικές κινήσεις που οφείλονται σε σοβαρά, αλλά συμπτωματικά ή απρόβλεπτα γεγονότα και ii) τις καθαρά τυχαίες κινήσεις, οι οποίες οφείλονται σε άγνωστους παράγοντες, δηλαδή όπως συνηθίζουμε να λέμε στην τύχη. Στο διάγραμμα 7.8 απεικονίζονται και τα τέσσερα είδη των συνιστωσών που προαναφέρθηκαν.



(Σχ. 7.8) Χρονολογικό διάγραμμα με κώνους



(Σχ. 7.9) Χρονολογικό διάγραμμα με ράβδους

7.3 Η Γραμμική τάση

Ο προσδιορισμός της τάσης μιας χρονολογικής σειράς γίνεται αφενός, για να μπορούμε να την προβάλλουμε στο μέλλον, κάνοντας έτσι προβλέψεις, και αφετέρου, αφαιρώντας την ήδη προσδιορισμένη τάση από μια χρονολογική σειρά εντοπίζουμε ευκολότερα τις άλλες κυμάνσεις. Επιπλέον, έχοντας προσδιορίσει την τάση μιας χρονολογικής σειράς επισημαίνονται ευκολότερα οι παράγοντες που την επηρεάζουν, γίνονται εύκολα συγκρίσεις με τάσεις άλλων χρονοσειρών κτλ.

Ο συνηθέστερος τρόπος μελέτης μιας χρονολογικής σειράς είναι αυτός που γίνεται με τη βοήθεια της μεθόδου των ελαχίστων τετραγώνων που γνωρίσαμε στο προηγούμενο κεφάλαιο. Αν λοιπόν η διαχρονική εξέλιξη της μεταβλητής Υ που εξετάζουμε μπορεί σε ικανοποιητικό βαθμό να εκφρασθεί από μία σχέση της μορφής:

$$\mathbf{y = \alpha + \beta t}$$

τότε ο προσδιορισμός της ζητούμενης τάσης θα γίνει με τη μέθοδο των ελαχίστων τετραγώνων. Από το γνωστό σύστημα των κανονικών εξισώσεων που συναντήσαμε στην παλινδρόμηση:

$$\sum y = \nu\hat{\alpha} + \hat{\beta}\sum t$$

$$\sum yt = \hat{\alpha}\sum t + \hat{\beta}\sum t^2$$

αν πάρουμε σαν αρχή στον οριζόντιο άξονα του χρόνου t, (t = 0) το κεντρικό έτος της χρονολογικής σειράς, τότε έχουμε $\sum t = 0$ και το παραπάνω σύστημα κανονικών εξισώσεων απλοποιείται και γίνεται:

$$\sum y = \nu\hat{\alpha}$$

$$\sum yt = \hat{\beta}\sum t^2$$

από το οποίο λαμβάνουμε :

$$\hat{\alpha} = \frac{\sum y}{\nu} \quad \text{και} \quad \hat{\beta} = \frac{\sum yt}{\sum t^2}$$

Με αυτό τον τρόπο παρατηρούμε ότι ο υπολογισμός των παραμέτρων $\hat{\alpha}$ και $\hat{\beta}$ είναι εμφανώς ευκολότερος. Ως προς το συντελεστή διεύθυνσης $\hat{\beta}$ πρέπει να αναφέρουμε ότι:

Αν $\hat{\beta} > 0$, τότε η χρονολογική σειρά παρουσιάζει ανοδική τάση.

Αν $\hat{\beta} < 0$, τότε η χρονολογική σειρά παρουσιάζει καθοδική τάση.

Αν $\hat{\beta} = 0$, τότε η χρονολογική σειρά κινείται παράλληλα προς τον άξονα t, δηλαδή είναι σταθερή και έχει μηδενική τάση.

Για να εφαρμόσουμε τους προηγούμενους τύπους και να έχουμε $\sum t = 0$, αν έχουμε περιττό αριθμό ετών, τότε επιλέγουμε ως αρχή του άξονα της μεταβλητής του χρόνου (t = 0) το μεσαίο (κεντρικό) έτος της χρονολογικής σειράς. Για παράδειγμα, αν έχουμε την παρακάτω χρονοσειρά των 7 ετών, θα επιλέξουμε ως αρχή του άξονα t, δηλαδή t = 0, το μεσαίο έτος 1998:

-1____-2_____-1_____0_____1____2______3
1995 1996 1997 1998 1999 2000 2001
$\Downarrow$
μεσαίο έτος

Στην περίπτωση που έχουμε άρτιο αριθμό ετών λαμβάνουμε ως διάμεσο έτος της χρονολογικής σειράς το ημιάθροισμα των δύο μεσαίων ετών. Για παράδειγμα, αν έχουμε την παρακάτω χρονοσειρά των έξι ετών, λαμβάνουμε ως διάμεσο έτος το ημιάθροισμα των ετών 1997 και 1998.

-2,5 -1,5 -0,5 0,5 1,5 2,5
1995 1996 1997 $\uparrow$ 1998 1999 2000
διάμεσο έτος

Στην επόμενη παράγραφο αναλύονται δύο παραδείγματα από το χώρο της οικονομίας και των επιχειρήσεων για να κατανοηθούν καλύτερα τα όσα αναφέραμε.

7.4 Η χρησιμότητα των χρονολογικών σειρών στην οικονομική δραστηριότητα

Η ανάλυση των χρονολογικών σειρών αποτελεί ένα από τα πλέον συχνά χρησιμοποιούμενα εργαλεία των οικονομικών αναλυτών και των επιχειρησιακών ερευνητών. Για παράδειγμα, μία επιχείρηση οφείλει να γνωρίζει το ύψος των πωλήσεών της για τα επόμενα δύο ή πέντε και δέκα χρόνια, ώστε να μπορεί να προγραμματίσει πολλά άλλα πράγματα, όπως τις επενδύσεις κεφαλαίου, τις διαφημιστικές της δαπάνες, τη μείωση ή την αύξηση του προσωπικού της στο σύνολο και κατά τμήματα κ.α. Για παρόμοιους λόγους μια ασφαλιστική εταιρεία θέλει να γνωρίζει την πορεία των τροχαίων ατυχημάτων για τα επόμενα πέντε ή δέκα χρόνια ή τον αριθμό ανασφάλιστων αυτοκινήτων για την ίδια περίοδο. Ένας χρηματοοικονομικός αναλυτής που ενδιαφέρεται για την πορεία του γενικού δείκτη του χρηματιστηρίου και των επιμέρους χρηματιστηριακών δεικτών αναγκαστικά θα πρέπει να αναλύσει τη διαχρονική πορεία τους, ώστε να λάβει όσο γίνεται πιο ασφαλείς αποφάσεις.

Ας δούμε δύο εφαρμογές από το χώρο της διοίκησης των επιχειρήσεων.

7.5 Εφαρμογές στις επιχειρήσεις

- ***Εφαρμογή 1η - Περιττός αριθμός ετών***

Περίπτωση μελέτης (case study) οικονομικών δεδομένων για τη σύνταξη πενταετούς προγράμματος οικονομικής δράσης.

Η εταιρεία πετρελαιοειδών G and B – Interoil προβληματίζεται ως προς το ύψος των επενδύσεων που πρέπει να κάνει σε κεφαλαιουχικό εξοπλισμό, σε μετεκπαίδευση προσωπικού, σε κλείσιμο νέων μακροπρόθεσμων συμβολαίων συνεργασίας με διυλιστήρια κτλ. για την επόμενη πενταετία.

Επειδή η πραγματοποίηση όλων των παραπάνω στόχων είναι πολύ δαπανηρή, ο οικονομικός αναλυτής που συντάσσει το πρόγραμμα οικονομικής δράσης θέλει να γνωρίζει το ύψος των απαραίτητων κεφαλαίων που απαιτούνται κατά έτος στον πενταετή σχεδιασμό. Έτσι αποφασίζει ότι είναι απαραίτητο, έστω και με ικανοποιητική προσέγγιση, να γνωρίζει:

α) Ποιο θα είναι το ετήσιο ύψος των τόνων πετρελαιοειδών που θα πουληθούν την ερχόμενη πενταετία.

β) Ποια είναι η μέση ετήσια αύξηση των πωλήσεων από έτος σε έτος.

γ) Ποια θα είναι η αύξηση των πωλήσεων το 2005 σε σχέση με το 2000.

Για όλους τους παραπάνω λόγους αποφάσισε να μελετήσει τις ετήσιες πωλήσεις της εταιρείας στη διάρκεια της τελευταίας επταετίας:

Πίνακας 7.1

Έτη	1994	1995	1996	1997	1998	1999	2000
Πωλήσεις (y)	10	15	20	25	30	30	40

(σ.σ: οι πωλήσεις δίνονται σε εκατοντάδες τόνους)

Προσδιορίζουμε αρχικά την ευθύγραμμη τάση ακολουθώντας τα παρακάτω βήματα:

Βήμα Ι

Κατασκευάζουμε την στήλη (3) του πίνακα 7.2, δηλαδή την t, αντιστοιχίζοντας το έτος 1997 (Κεντρικό Έτος) στην υποδιαίρεση μηδέν του άξονα t, επειδή έχουμε περιττό αριθμό ετών.

Βήμα ΙΙ

Κατασκευάζουμε τη στήλη (4), πολλαπλασιάζοντας τους όρους της στήλης (2) επί τους αντίστοιχους όρους της στήλης (3). Κατόπιν αθροίζοντας τους όρους της στήλης (4) βρίσκουμε $\sum yt = 130$, επίσης $\sum y = 170$.

Βήμα ΙΙΙ

Υψώνοντας στο τετράγωνο κάθε όρο της στήλης (3) δημιουργούμε τους αντίστοιχους όρους της στήλης (5). Αθροίζοντας κατόπιν τους όρους της στήλης αυτής βρίσκουμε $\sum t^2 = 28$.

Πίνακας 7.2

Στήλες δεδομένων		**Στήλες υπολογισμών**		
(1)	(2)	(3)	(4)	(5)
Έτη	Πωλήσεις y	t	yt	t^2
1994	10	-3	-30	9
1995	15	-2	-30	4
1996	20	-1	-20	1
Κ.Ε. 1997	25	0	0	0
1998	30	1	30	1
1999	30	2	60	4
2000	40	3	120	9
Σύνολο	170		130	28

Κ.Ε. = Κεντρικό Έτος

Βήμα IV

Υπολογισμός των παραμέτρων $\hat{\alpha}$ και $\hat{\beta}$.

Αντικαθιστώντας τα αποτελέσματα που βρήκαμε παραπάνω στους τύπους υπολογισμού των $\hat{\alpha}$ και $\hat{\beta}$ έχουμε:

$$\hat{\alpha} = \frac{\sum y}{\nu} = \frac{170}{7} = 24,28$$

και

$$\hat{\beta} = \frac{\sum yt}{\sum t^2} = \frac{130}{28} = 4,643$$

Κατά συνέπεια η ζητούμενη εξίσωση ευθύγραμμης τάσης είναι:

$\hat{y} = 24,28 + 4,643\ t$.

Άρα θα έχουμε:

1. Οι ετήσιες πωλήσεις σε εκατοντάδες τόνους για τα επόμενα έτη, λαμβάνοντας υπόψη ότι τα έτη μετρώνται από το έτος 1997 (t = 0), θα είναι:

Για το 2001, $\hat{y} = 24,28 + 4,64\ .\ 4 = 42,84$

Για το 2002, $\hat{y} = 24,28 + 4,64\ .\ 5 = 47,48$

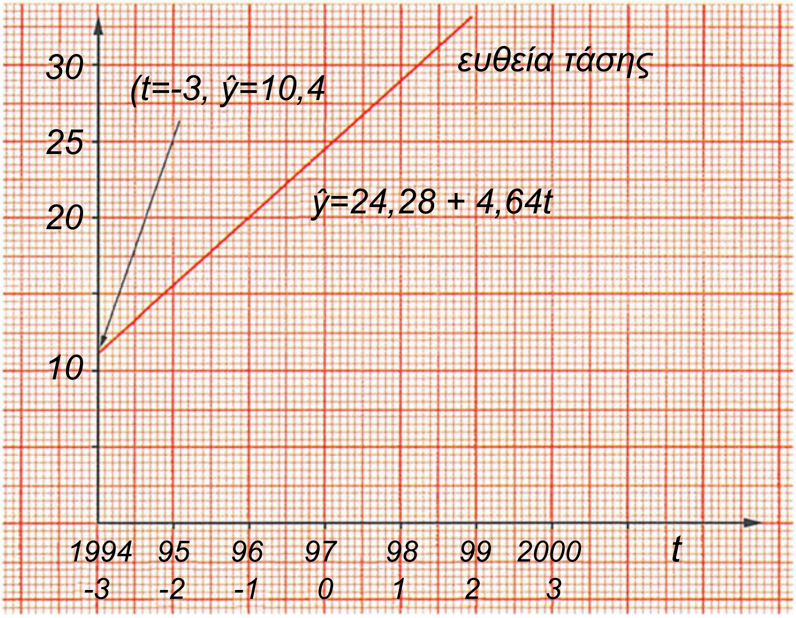
:

για το 2005, $\hat{y} = 24,28 + 4,64\ .\ 8 = 61,40$

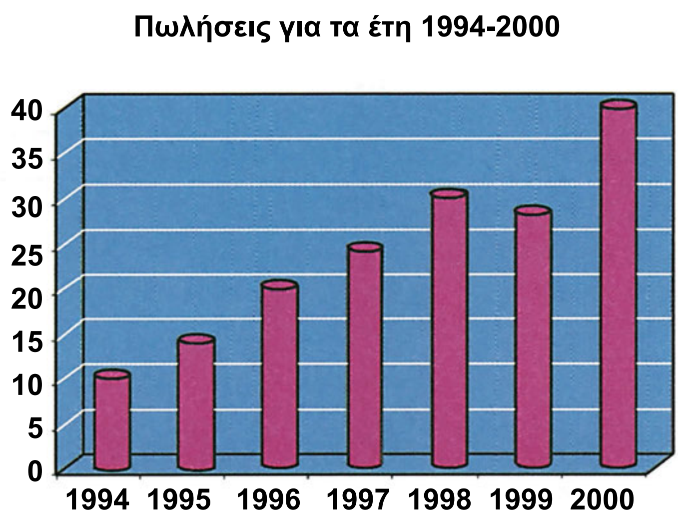
2. Επειδή ο συντελεστής διεύθυνσης ή γωνιακός συντελεστής $\hat{\beta}$ μας λέει πόσο μεταβάλλεται η εξαρτημένη μεταβλητή Y σε μία μοναδιαία μεταβολή (κατά ένα έτος) της t, είναι προφανές ότι κατά την εξεταζόμενη περίοδο η μέση ετήσια αύξηση των πωλήσεων ήταν 464 τόνοι, διότι βρήκαμε $\hat{\beta} = 4,64$.

3. Παρατηρούμε ότι η επιχείρηση θα πρέπει να ετοιμάζεται, ώστε στο τέλος της ερχόμενης πενταετίας, το 2005, να ικανοποιήσει μία ζήτηση (πωλήσεις) ύψους 6.140 τόνων (61,4 εκατ/δων τόνων). Δηλαδή, περίπου μία αύξηση 53,5% σε σχέση με τις πωλήσεις του έτους 2.000, που είναι 4.000 τόνοι.

Πρέπει να τονιστεί ότι, επειδή αντιστοιχίσαμε το t = 0 στο έτος 1997, η ευθεία τάσης και ο άξονας τέμνονται στο σημείο (t = -3, y = 10,4), διότι y = 24,3 - 3 . 4,64 = 10,4. (Βλ. Σχ. 7.11).


30
25
20
10
(t=-3, ŷ=10,4
ευθεία τάσης
ŷ=24,28 + 4,64t
1994 95 96 97 98 99 2000
-3 -2 -1 0 1 2 3
t

(Σχ. 7.10)


Πωλήσεις για τα έτη 1994-2000
40
35
30
25
20
15
10
5
0
1994 1995 1996 1997 1998 1999 2000

(Σχ. 7.11)

• *Εφαρμογή 2η - Άρτιος αριθμός ετών*

Με βάση τα δεδομένα της πρώτης εφαρμογής αλλά με τη διαφορά ότι θα λάβουμε υπόψη μας ένα έτος ακόμη, το 1993, έχουμε άρτιο αριθμό ετών, δηλαδή 8 έτη.

Η διαδικασία υπολογισμού είναι περίπου η ίδια με την προηγούμενη, μόνο που εδώ επειδή έχουμε άρτιο αριθμό ετών, θα πάρουμε διάμεσο έτος (1996,5) το ημιάθροισμα των ετών 1996 και 1997, που είναι τα δύο μεσαία έτη. Έχουμε λοιπόν τα εξής βήματα:

Βήμα Ι

Σχηματίζουμε τη στήλη (3) του πίνακα 7.3 λαμβάνοντας υπόψη ως διάμεσο έτος το 1996,5.

Βήμα ΙΙ

Σχηματίζουμε τη στήλη (4) του παρακάτω πίνακα κατά τα γνωστά και βρίσκουμε ότι: Σyt = 177,5 , επίσης Σy = 195.

Βήμα ΙΙΙ

Σχηματίζουμε, όπως και στην προηγούμενη εφαρμογή, τη στήλη t^2 και βρίσκουμε ότι $\sum t^2 = 42$.

Πίνακας 7.3

Στήλες δεδομένων		Στήλες υπολογισμών		
(1) Έτη	(2) Πωλήσεις y	(3) t	(4) yt	(5) t^2
1993	5	-3,5	-17,5	12,25
1994	15	-2,5	-37,5	6,25
1995	50	-1,5	-30	2,25
1996	25	-0,5	-12,5	0,25
1997	30	0,5	15	0,25
1998	30	1,5	45	2,25
1999	30	2,5	75	6,25
2000	40	3,5	140	12,25
Σύνολο	195		177,5	42

Βήμα IV

Αντικαθιστούμε τα αποτελέσματα που βρήκαμε στους τύπους υπολογισμού των $\hat{\alpha}$ και $\hat{\beta}$ και έχουμε:

$$\hat{\alpha} = \frac{\sum y}{v} = \frac{195}{8} = 24,37$$

και

$$\hat{\beta} = \frac{\sum yt}{\sum t^2} = \frac{177,5}{42} = 4,23$$

Κατά συνέπεια η ζητούμενη εξίσωση ευθύγραμμης τάσης είναι:

$$\hat{y} = 24,37 + 4,23t$$

Αν κάνουμε μία πρόβλεψη για το έτος 2005 θα έχουμε:

$$\hat{y} = 24,37 + 4,23 \cdot 8,5 \cong 60,33$$

Δηλαδή, οι πωλήσεις θα είναι 60,33 εκατοντάδες τόνοι πετρελαιοειδών. Όπως παρατηρούμε και αυτή η εκτίμηση είναι πολύ κοντά στην προηγούμενη εκτίμηση (εφαρμογή Α) που βρήκαμε 61,4 εκατοντάδες τόνους για το έτος 2005.

ΑΝΑΚΕΦΑΛΑΙΩΣΗ

1. **Χρονολογική σειρά** είναι μια σειρά από τιμές που παίρνει μια μεταβλητή που εξετάζουμε στη διάρκεια του χρόνου και συνήθως σε ίσα χρονικά διαστήματα.

2. Βασικός **σκοπός της μελέτης της διαχρονικής εξέλιξης μιας μεταβλητής** είναι η απόκτηση επαρκούς γνώσης από την επιστημονική μελέτη της συμπεριφοράς της εξεταζόμενης μεταβλητής στο παρελθόν, ώστε να διαμορφώσουμε μια **ικανοποιητικού βαθμού πρόβλεψη** για τη συμπεριφορά της στο μέλλον.

3. Με τα **χρονογράμματα ή χρονοδιαγράμματα** απεικονίζουμε τη **διαχρονική εξέλιξη μιας μεταβλητής**. Αν η μεταβλητή είναι οικονομικού χαρακτήρα, αναφερόμαστε σε οικονομικά χρονοδιαγράμματα.

4. **Οι συνιστώσες** μιας χρονοσειράς είναι:
α) η τάση
β) οι κυκλικές κυμάνσεις
γ) οι εποχικές κυμάνσεις
δ) οι τυχαίες ή άρρυθμες κινήσεις.

5. Η ύπαρξη κάποιας ή κάποιων συνιστωσών των χρονολογικών σειρών είναι φαινόμενο πολύ συνηθισμένο στο χώρο της οικονομίας.

6. Η μέθοδος των **ελαχίστων τετραγώνων** αποτελεί στην πράξη την πλέον συνηθισμένη μέθοδο προσδιορισμού της τάσης.

7. Η απλή μορφή του συστήματος των κανονικών εξισώσεων στην περίπτωση της ευθύγραμμης τάσης είναι:

$$\sum y = \nu\hat{\alpha}$$

$$\sum y \cdot t = \hat{\beta} \sum t^2$$

8. Από το απλοποιημένο σύστημα κανονικών εξισώσεων έχουμε:

$\hat{\alpha} = \frac{\sum y}{\nu}$ και $\hat{\beta} = \frac{\sum yt}{\sum t^2}$

9. **Ο συντελεστής διεύθυνσης ή γωνιακός συντελεστής $\hat{\beta}$** ή παράμετρος $\hat{\beta}$ της ευθύγραμμης τάσης, **μας δίνει την ετήσια μεταβολή της τάσης.**

10. Για την παράμετρο $\hat{\beta}$ της $\hat{y} = \hat{\alpha} + \hat{\beta}t$ ισχύει:

Αν $\hat{\beta} > 0$, τότε η χρονολογική σειρά έχει **τάση ανοδική.**
Αν $\hat{\beta} < 0$, τότε η χρονολογική σειρά έχει **τάση καθοδική.**
Αν $\hat{\beta} = 0$, τότε έχει **μηδενική τάση** στη χρονολογική σειρά.

ΕΡΩΤΗΣΕΙΣ ΚΑΤΑΝΟΗΣΗΣ

1. Ως χρονολογική σειρά ορίζουμε .... από τιμές .... που εξετάζουμε...................... χρόνου και συνήθως ....

2. Ο βασικός σκοπός της .... της διαχρονικής .... μιας .... είναι να αποκτήσουμε μια ικανοποιητική .... για τη .... στο μέλλον.

3. Οι συνιστώσες μιας χρονολογικής σειράς είναι:

α. .... β. ....

γ. .... δ. ....

4. Η ύπαρξη χρονολογικών συνιστωσών είναι φαινόμενο συνηθισμένο στο χώρο της οικονομίας.

Σ ☐ Λ ☐

5. Να αναφέρετε συνοπτικά τη χρησιμότητα των χρονολογικών σειρών για τις επιχειρήσεις και το χώρο της οικονομίας γενικά.

6. Η μακροχρόνια γραμμική τάση προσδιορίζεται από την εξίσωση:

....

7. Περιγράψτε τις κυκλικές διακυμάνσεις.

8. Περιγράψτε τις εποχικές διακυμάνσεις.

9. Περιγράψτε τις τυχαίες ή άρρυθμες κινήσεις.

10. Συνήθως προσδιορίζουμε στην πράξη την ευθύγραμμη τάση με την

....

....

11. Το απλοποιημένο σύστημα των κανονικών εξισώσεων που χρησιμοποιούμε για τον προσδιορισμό της ευθύγραμμης τάσης είναι:

α. $\sum xy = \nu\hat{\alpha}$
$\sum yt = \hat{\beta}\sum xt^2$

β. $\sum ty = \nu\hat{\alpha}$
$\sum xy = \hat{\beta}\sum t^2$

γ. $\sum y = \nu\hat{\alpha}$
$\sum yt = \hat{\beta}\sum t^2$

δ. $\sum y = t\hat{\alpha}$
$\sum yt = \hat{\beta}\sum t$

12. Ο συντελεστής $\hat{\beta}$ της $\hat{y} = \hat{\alpha} + \hat{\beta}t$ μας δείχνει πόσο μεταβάλλεται

....σε μία..

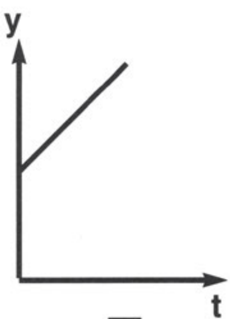
13. Οι τύποι υπολογισμού που προκύπτουν από το απλοποιημένο σύστημα κανονικών εξισώσεων:

Για την παράμετρο $\hat{\alpha}$ είναι: ……………………………………………..

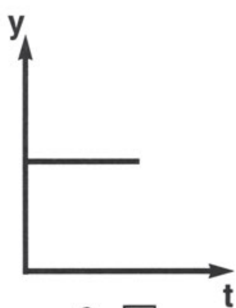
Για την παράμετρο $\hat{\beta}$ είναι: ……………………………………………..

14. Η τάση μιας χρονολογικής σειράς, όταν $\hat{\beta} > 0$, είναι:
α. αυξητική β. φθίνουσα γ. μηδενική δ. είτε αύξουσα είτε φθίνουσα.

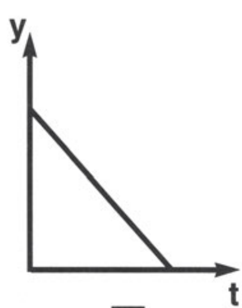
15. Ποια από τις χρονολογικές σειρές έχει τάση φθίνουσα:



α ☐



β ☐



γ ☐

ΑΣΚΗΣΕΙΣ

Ομάδα Α

1. Ο παρακάτω πίνακας δίνει την ετήσια παραγωγή σε τόνους ενός προϊόντος από μία επιχείρηση κατά τη διάρκεια των ετών 1996-2000.

Έτη	Προϊόν
1996	5
1997	10
1998	10
1999	15
2000	20

Ζητείται με τη βοήθεια της μεθόδου των ελαχίστων τετραγώνων να προσδιοριστεί η ευθύγραμμη τάση.

2. Τα κέρδη προ φόρων μιας επιχείρησης σε δεκάδες χιλιάδες ευρώ κατά τα έτη 1994-2000 δίνονται από τον παρακάτω πίνακα:

Έτη	Κέρδη
1995	20
1996	30
1997	30
1998	40
1999	50
2000	60

Ζητείται να γίνει το σχετικό χρονοδιάγραμμα.

3. Με βάση τα δεδομένα της άσκησης 2 να προσδιοριστεί (με τη μέθοδο των ελαχίστων τετραγώνων) η ευθύγραμμη τάση $\hat{y} = \hat{\alpha} + \hat{\beta}t$.

4. Ο αριθμός των ετήσιων τροχαίων ατυχημάτων μιας διοικητικής περιφέρειας κατά τα έτη 1994 έως και 2000 δίνεται από τον πίνακα που ακολουθεί:

Έτη	**Αριθμός ατυχημάτων**
1994	350
1995	400
1996	400
1997	450
1998	500
1999	550
2000	600

Να προσαρμοστεί στα παραπάνω δεδομένα η ευθύγραμμη τάση $\hat{y} = \hat{\alpha} + \hat{\beta}t$.
Επίσης να γίνει μια πρόβλεψη για τον αριθμό ατυχημάτων το έτος 2002.

5. Τα γενικά έξοδα σε εκατομ. ευρώ ετησίως της εταιρείας Αστήρ-3Α κατά τα έτη 1992 μέχρι και 1999 φαίνονται στον παρακάτω πίνακα:

Έτη	**Έξοδα**
1992	0,5
1993	0,5
1994	1
1995	1
1996	1,5
1997	2
1998	2,5
1999	3

Να προσδιοριστεί:

α) η ευθύγραμμη τάση $\hat{y} = \hat{\alpha} + \hat{\beta}t$.

β) να γίνει πρόβλεψη εξόδων για τα έτη 2002 και 2004.

Ομάδα Β

1. Οι διαφημιστικές δαπάνες της εταιρείας Ελλάς-Κόσμος κατά τα έτη 1993 μέχρι 1999 δίνονται από τον παρακάτω πίνακα (σε δεκάδες χιλιάδες ευρώ).

Έτη	Δαπάνες
1993	8
1994	10
1995	10
1996	12
1997	14
1998	16
1999	20

Ζητείται:

α) να προσδιοριστεί η ευθύγραμμη τάση $\hat{y} = \hat{\alpha} + \hat{\beta}t$.

β) με τη βοήθεια της εξίσωσης $\hat{y} = \hat{\alpha} + \hat{\beta}t$ να εκτιμήσετε την τάση για τα έτη 1993-1999.

γ) να βρείτε τις διαφορές μεταξύ των πραγματικών τιμών y και των τιμών $\hat{y}$ που μας έδωσε η $\hat{y} = \hat{\alpha} + \hat{\beta}t$ για τα έτη 1993 μέχρι και 1999.

δ) να σχολιαστεί το αποτέλεσμα του ερωτήματος (γ), δηλαδή οι διαφορές $(y - \hat{y})$.

2. Ο παρακάτω πίνακας παρουσιάζει τις αγορές που έγιναν με πιστωτικές κάρτες στην Ελλάδα κατά τα έτη 1991 μέχρι και 1997. Τα ποσά είναι σε ευρώ και έχουν στρογγυλοποιηθεί:

Έτη	Ποσά
1991	23.478
1992	29.347
1993	61.629
1994	129.127
1995	85.106
1996	179.017
1997	132.061

Πηγή: *Δελτίο Ελληνικών Τραπεζών*

Ζητείται:

α) να προσδιοριστεί η ευθεία τάσης $\hat{y} = \hat{\alpha} + \hat{\beta}t$.

β) με βάση την $\hat{y} = \hat{\alpha} + \hat{\beta}t$ να εκτιμηθούν οι τιμές της $\hat{y}$ για τα έτη 1998 και 2000.

γ) ποιος ήταν ο ρυθμός της μέσης ετήσιας μεταβολής των αγορών με πιστωτικές κάρτες κατά την εξεταζόμενη περίοδο;

δ) να γίνει το σχετικό διάγραμμα και να χαραχθεί πάνω σ' αυτό η ευθεία τάσης.

3. Ο παρακάτω πίνακας παρουσιάζει το σύνολο ασφαλίστρων, ως ποσοστό επί του Α.Ε.Π, κατά τα έτη 1990 μέχρι και το 1997, που καταβλήθηκαν στα διάφορα ταμεία κοινωνικής ασφάλισης:

Έτη	Ασφάλιστρα (%)
1989	1,10
1990	1,25
1991	1,30
1992	1,46
1993	1,59
1994	1,61
1995	1,65
1996	1,72

Πηγή: *8η Συν/ση για την Υγεία και τις πολ. υγείας - 1999*

Ζητείται:

α) να γίνει ο προσδιορισμός της ευθύγραμμης τάσης $\hat{y} = \hat{\alpha} + \hat{\beta}t$

β) να βρεθεί η μέση ετήσια μεταβολή των ασφαλίστρων κατά την περίοδο 1989-1996

γ) να γίνει μια πρόβλεψη για το 2000

δ) να εκτιμηθούν οι τιμές της Y για τα έτη 1989 μέχρι 1996 με βάση την εξίσωση $\hat{y} = \hat{\alpha} + \hat{\beta}t$

ε) να σχολιασθεί το αποτέλεσμα του ερωτήματος (δ), δηλαδή οι διαφορές $(y - \hat{y})$.

4. Η συνολική ιδιωτική δαπάνη υγείας κατά τη χρονική περίοδο 1990 μέχρι και 1997 δίνεται από τον παρακάτω πίνακα:

Έτη	Ιδιωτική δαπάνη υγείας
1990	1,1
1991	1,3
1992	1,6
1993	2,0
1994	2,9
1995	3
1996	3,3
1997	3,5

Πηγή: *8η Συν/ση για την υγεία και τις πολ. υγείας – 1999.*

Τα ποσά εκφράζουν εκατομ. ευρώ.

Ζητείται:

α) να γίνει ο προσδιορισμός της ευθείας τάσης $\hat{y} = \hat{\alpha} + \hat{\beta}t$

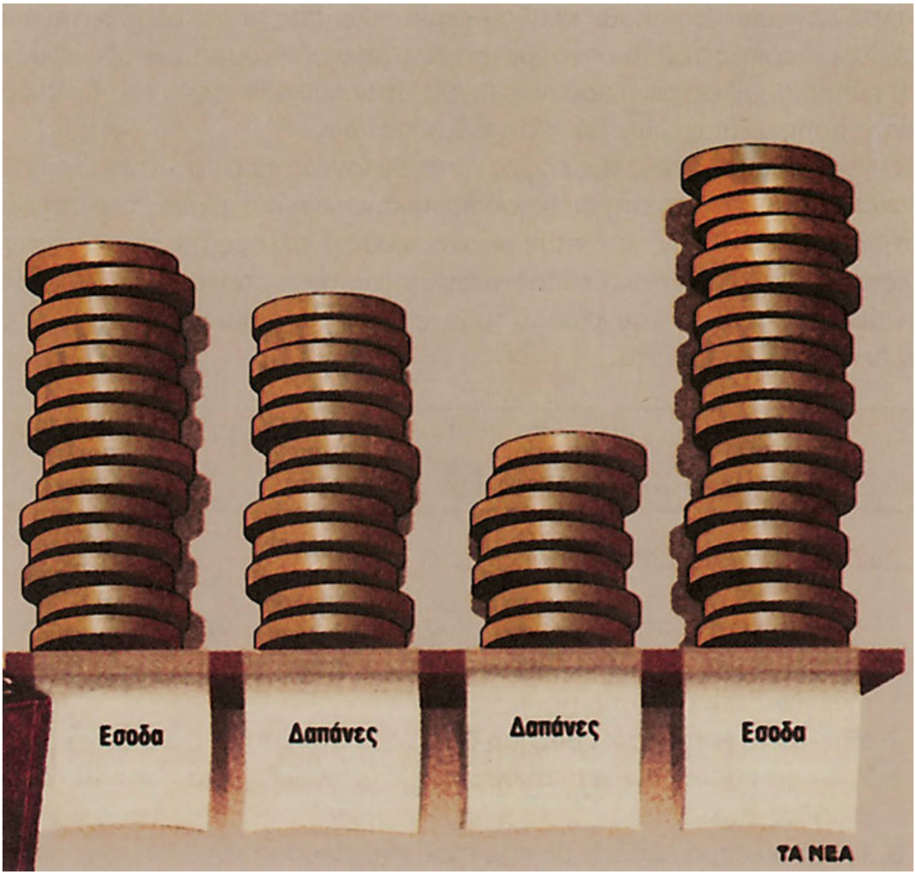
β) να γίνει πρόβλεψη για τα έτη 1999 και 2002

γ) να γίνει με βάση την $\hat{y} = \hat{\alpha} + \hat{\beta}t$ εκτίμηση των τιμών της y για τα έτη 1990 μέχρι και 1997

δ) να βρεθούν οι διαφορές των πραγματικών τιμών y της χρονολογικής σειράς από τις αντίστοιχες $\hat{y}$ που μας δίνει η $\hat{y} = \hat{\alpha} + \hat{\beta}t$

ε) να γίνει το σχετικό διάγραμμα και να χαραχθεί πάνω σ' αυτό η καμπύλη τάσης.

ΓΕΝΙΚΗ ΕΦΑΡΜΟΓΗ



Περίπτωση μελέτης (case study) επιχείρησης δύο εκμεταλλεύσεων.

Η Ανώνυμη Εταιρεία «Ευρωπαϊκή Ηλεκτρεμπορική – 3Α» διαθέτει δύο κλάδους εκμετάλλευσης, τον κλάδο (X) εμπορίας ηλεκτρικών συσκευών και τον κλάδο (Y) συντήρησης και επισκευής ηλεκτρικών συσκευών (service). Ο κύκλος εργασιών του κάθε κλάδου εκμετάλλευσης παρακολουθείται χωριστά. Στη διάρκεια του προηγούμενου οικονομικού έτους τα έσοδα του κλάδου εμπορίας ηλεκτρικών συσκευών (X) ήταν 500.000 ευρώ, του δε κλάδου συντήρησης – επισκευών (Y) ήταν 100.000 ευρώ.

Ο γενικός διευθυντής της επιχείρησης θέλοντας να διερευνήσει το βαθμό αποτελεσματικότητας στη διοίκηση και οικονομική διαχείριση της εταιρείας ζήτησε, μεταξύ άλλων, από τους οικονομικούς αναλυτές της επιχείρησης να μελετήσουν τα αντίστοιχα δεδομένα πωλήσεων του προηγούμενου έτους των 12 πρώτων εταιρειών του κλάδου τους και να απαντήσουν, αν μπορούν, στα παρακάτω 13 ερωτήματα.

x	35	36	38	40	40	42	44	45	48	50	55	60
y	7	7	8	8	8	8,5	8,5	9	10	10	10	12

(Δεδομένα σε δεκάδες χιλιάδες ευρώ)

1. Αν υπάρχει και σε ποιο βαθμό συσχέτιση μεταξύ των εσόδων από πωλήσεις ηλεκτρικών συσκευών (x) και των εσόδων της υπηρεσίας επισκευών – συντήρησης (y).
2. Μπορεί να εκτιμηθεί τι αύξηση θα είχαμε στα έσοδα της εκμετάλλευσης Y, αν τα έσοδα της εκμετάλλευσης X ήταν αυξημένα κατά 10 δεκάδες χιλιάδες ευρώ;
3. Τι αποτέλεσμα εσόδων θα είχε η εκμετάλλευση Y, αν κάποια χρονιά η X είχε μηδενικά έσοδα;
4. Ποια θα ήταν η αναμενόμενη μεταβολή στα έσοδα της εκμετάλλευσης Y, αν η εκμετάλλευση X είχε μεταβολή εσόδων κατά 1 δεκάδα χιλιάδα ευρώ;
5. Να βρεθεί σ’ αυτή την ομάδα των 12 επιχειρήσεων η τυπική απόκλιση εσόδων για την εκμετάλλευση X.
6. Να βρεθεί η τυπική απόκλιση εσόδων και για την εκμετάλλευση Y.
7. Να βρεθεί ο CV των εσόδων για την εκμετάλλευση X.
8. Να βρεθεί ο CV των εσόδων για την εκμετάλλευση Y.
9. Ποια από τις δύο εκμεταλλεύσεις παρουσιάζει μεγάλη διασπορά εσόδων;

10. Να βρεθεί το διάμεσο ύψος εσόδων της εκμετάλλευσης X (εμπορίας ηλεκτρικών συσκευών).
11. Να βρεθεί το διάμεσο ύψος εσόδων της εκμετάλλευσης Y (επισκευών – συντήρησης).
12. Να γίνει συνοπτική ερμηνεία για όλα τα αποτελέσματα των προηγούμενων περιπτώσεων από την 1 μέχρι και την 11.
13. Να γίνουν τα σχετικά διαγράμματα για την καλύτερη παρουσίαση του ζητήματος.

Πίνακας Π 1

Στήλες δεδομένων		**Στήλες Υπολογισμών**		
(1)	**(2)**	**(3)**	**(4)**	**(5)**
x	**y**	**xy**	**x^2**	**y^2**
35	7	245	1225	49
36	7	252	1296	49
38	8	304	1444	64
40	8	320	1600	64
40	8	320	1600	64
42	8,5	357	1764	72,25
44	8,5	374	1936	72,25
45	9	405	2025	81
48	10	480	2304	100
50	10	500	2500	100
55	10	550	3025	100
60	12	720	3600	144
Σύνολο 533	**106**	**4827**	**24319**	**959,5**

Λύση:

1. Είναι προφανές ότι το ερώτημα 1 απαντάται, αν υπολογίσουμε το συντελεστή γραμμικής συσχέτισης r. Ακολουθώντας τα γνωστά βήματα υπολογισμού έχουμε:

Βήμα Ι: $\Sigma x = 533$
$\Sigma y = 106$

Βήμα ΙΙ: $\Sigma xy = 4827$
$\Sigma x^2 = 24319$
$\Sigma y^2 = 959,5$

Βήμα ΙΙΙ: Υπολογισμός του r

$$r = \frac{\nu\sum xy - \left(\sum x\right)\cdot\left(\sum y\right)}{\sqrt{\nu\sum x^2 - \left(\sum x\right)^2}\cdot\sqrt{\nu\sum y^2 - \left(\sum y\right)^2}}$$

$$r = \frac{12\cdot 4827 - 533\cdot 106}{\sqrt{12\cdot 24319 - 533^2}\cdot\sqrt{12\cdot 959,5 - 106^2}}$$

r = 0,97 ή
r = 97%

Είναι λοιπόν φανερό ότι μεταξύ των δύο εκμεταλλεύσεων υπάρχει θετική συσχέτιση και μάλιστα εντονότατη, αφού το r είναι πολύ κοντά στη μονάδα.

2. Είναι φανερό ότι για την απάντηση αυτή απαιτείται ο προσδιορισμός της ευθείας παλινδρόμησης. Εργαζόμενοι κατά τα γνωστά έχουμε:

Βήμα Ι: Υπολογισμός των δύο μέσων αριθμητικών

$$\overline{x} = \frac{\sum x}{\nu} = \frac{533}{12} = 44,41$$

$$\overline{y} = \frac{\sum y}{\nu} = \frac{106}{12} = 8,83$$

Βήμα ΙΙ: $\Sigma xy = 4827$

$\Sigma x^2 = 24319$

Βήμα ΙΙΙ:

$$\hat{\beta} = \frac{\nu\sum xy - \left(\sum x\right)\cdot\left(\sum y\right)}{\nu\sum x^2 - \left(\sum x\right)^2}$$

$$\hat{\beta} = \frac{12\cdot 4827 - 533\cdot 106}{12\cdot 24319 - 533^2}$$

Βήμα IV: $\hat{\beta} = 0{,}18$

$\hat{\alpha} = 8{,}83 - 0{,}18\cdot 44{,}41$

$\hat{\alpha} = 8{,}83 - 7{,}99 = 0{,}84$

Επομένως έχουμε: $\hat{y} = 0{,}84 + 0{,}18x$

Αν λοιπόν η εκμετάλλευση Χ της επιχείρησης είχε έσοδα αυξημένα κατά 10 δεκάδες χιλιάδες ευρώ, δηλαδή αντί για 50 δεκάδες χιλιάδες ευρώ είχε 50 + 10 = 60 δεκάδες χιλιάδες ευρώ, τότε τα έσοδα της Υ θα ήταν: $\hat{y} = 0{,}836 + 0{,}18\cdot 60 = 11{,}64$ δεκάδες χιλιάδες ευρώ ή 116.400 ευρώ.

3. Αν η εκμετάλλευση Χ είχε μηδενικά έσοδα, δηλαδή $x = 0$, τότε η Υ εκμετάλλευση θα είχε έσοδα ίσα με $\hat{y} = 0{,}836 + 0{,}18\cdot 0 = 0{,}836$ δεκάδες χιλιάδες ευρώ, δηλαδή όσο είναι η σταθερά $\hat{\alpha}$.

4. Προφανώς επειδή τα δεδομένα είναι σε δεκάδες χιλιάδες ευρώ, ο γωνιακός συντελεστής $\hat{\beta}$ εκφράζει τη μεταβολή που επέρχεται στην τιμή της μεταβλητής Υ από μία μοναδιαία μεταβολή (10.000 ευρώ) της τιμής της μεταβλητής x. Τότε, η μεταβολή στα έσοδα της Υ, θα ήταν ίση με το $\hat{\beta}$ (1.800 ευρώ).

5. Η τυπική απόκλιση των τιμών της Χ δίνεται από το γνωστό τύπο:

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{\nu}}$$

Σχηματίζουμε λοιπόν τον παρακάτω πίνακα Π2, κατά τα γνωστά, στη συνέχεια αντικαθιστούμε το άθροισμα των στοιχείων της στήλης (2) στον παραπάνω τύπο και βρίσκουμε:

Πίνακας Π 2

(1)	**(2)**	**(3)**	**(4)**
$(\mathbf{x}-\bar{\mathbf{x}})$	$(\mathbf{x}-\bar{\mathbf{x}})^2$	$(\mathbf{y}-\bar{\mathbf{y}})$	$(\mathbf{y}-\bar{\mathbf{y}})^2$
35 - 44,41	88,54	7 – 8,8	3,24
36 – 44,41	70,72	7 – 8,8	3,24
38 – 44,41	41,09	8 – 8,8	0,64
40 – 44,41	19,45	8 – 8,8	0,64
40 – 44,41	19,45	8 – 8,8	0,64
42 – 44,41	5,81	8,5 – 8,8	0,09
44 – 44,41	0,17	8,5 – 8,8	0,09
45 – 44,41	0,35	9 – 8,8	0,04
48 – 44,41	12,88	10 – 8,8	1,44
50 – 44,41	31,25	10 – 8,8	1,44
55 – 44,41	112,15	10 – 8,8	1,44
60 – 44,41	243,05	12 – 8,8	10,24
Σύνολο	**644,91**		**23,18**

$$s_x = \sqrt{\frac{644{,}91}{12}} = \sqrt{53{,}74}$$

$$s_x = 7{,}33$$

Από τον ίδιο πίνακα βρίσκουμε το άθροισμα της στήλης (4) και το αντικαθιστούμε στον τύπο, οπότε έχουμε:

$$s_y = \sqrt{\frac{\sum(y-\bar{y})^2}{v}}$$

$$s_y = \sqrt{\frac{23,18}{12}} = \sqrt{1,93}$$

$s_y = 1,39$

Παρατηρούμε ότι τα έσοδα της εκμετάλλευσης Y παρουσιάζουν πολύ μικρότερη διασπορά από αυτή που παρουσιάζουν τα έσοδα της εκμετάλλευσης X, αφού $s_x = 7,33 > s_y = 1,38$. Όμως, όπως έχουμε αναφέρει, για να είμαστε σίγουροι θα πρέπει να βρούμε και τους αντίστοιχους συντελεστές μεταβλητότητας CV.

7. Με βάση το γνωστό τύπο $CV_x = \frac{s_x}{\bar{x}}$ θα έχουμε για την περίπτωση της X ότι:

$$CV_x = \frac{s_x}{\bar{x}} = \frac{7,33}{44,4} = 0,165$$

ή

$CV_x = 16,5\%$

8. Ομοίως και για την περίπτωση της Y, έχουμε:

$$CV_y = \frac{s_y}{\bar{y}} = \frac{1,38}{8,83} = 0,156$$

ή

$CV_y = 15,6\%$

9. Παρατηρούμε ότι οι συντελεστές μεταβλητότητας CV_y και CV_x είναι πολύ κοντά ο ένας στον άλλο. Θα λέγαμε λοιπόν ότι η διασπορά των εσόδων της Y, είναι σχεδόν ίδια με αυτή της X. Προφανώς είναι λογικό και επόμενο να συμβαίνει αυτό, επειδή υπάρχει μια υψηλότατη συσχέτιση, όπως είδαμε, $r = 0,97$ με αποτέλεσμα οι τιμές της Y να επηρεάζονται εντονότατα από αυτές της X. Ακόμη παρατηρούμε ότι η διασπορά είναι μικρή και για τις δύο εκμεταλλεύσεις, αφού είναι μικρές οι τιμές του συντελεστή μεταβλητότητας (CV).

10. Το διάμεσο ύψος εσόδων (άρτιος αριθμός επιχειρήσεων) θα βρεθεί, αφού κατατάξουμε σε αύξουσα σειρά τις παρατηρήσεις – έσοδα, και θα είναι μεταξύ του έκτου και έβδομου όρου. Δηλαδή η θέση της διαμέσου εντοπίζεται στο $\frac{\nu+1}{2}=\frac{12+1}{2}=6{,}5$

Ως προς την περίπτωση λοιπόν της Χ η διάμεσος είναι:

$$\delta_x = \frac{42+44}{2}=\frac{86}{2}$$

δ_x = 43 δεκάδες χιλιάδες ευρώ ή
δ_x = 430.000 ευρώ

Δηλαδή το 50% (6 επιχειρήσεις) έχουν εκμεταλλεύσεις Χ με έσοδα από 43 δεκάδες χιλιάδες ευρώ και άνω.

11 και 12.
Ομοίως για την εκμετάλλευση Υ έχουμε:

$$\delta_y = \frac{8{,}5+8{,}5}{2}=\frac{17}{2}$$

δ_y = 8,5 δεκάδες χιλιάδες ευρώ ή
δ_y = 85.000 ευρώ.

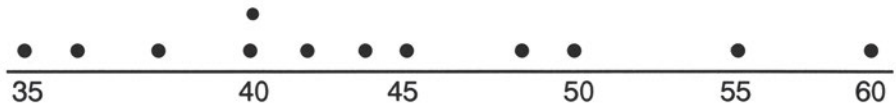
Δηλαδή οι εκμεταλλεύσεις Υ, 6 επιχειρήσεων έχουν έσοδα από 85.000 ευρώ και άνω και άλλες 6 έχουν από 85.000 ευρώ και κάτω.

Ακόμη παρατηρώντας προσεκτικά βλέπουμε ότι και για τις δύο εκμεταλλεύσεις ο μέσος και η διάμεσος είναι πάρα πολύ κοντά. Για την εκμετάλλευση Χ έχω μέσο $\bar{x}$ = 44,41 και διάμεσο 43 και για την εκμετάλλευση Υ έχω $\bar{y}$ = 8,83 και διάμεσο 8,5.

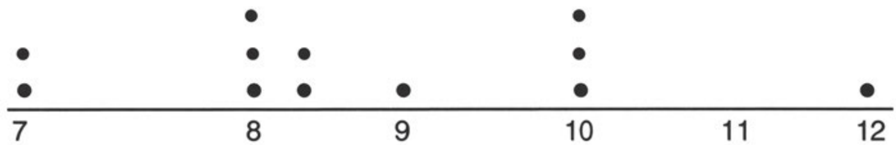
Επίσης παρατηρούμε ότι και οι δύο εκμεταλλεύσεις της επιχείρησης κατέχουν, ως προς τα έσοδά τους, μια καλή θέση, αφού ανήκουν και οι δύο στο 50% των εκμεταλλεύσεων με τα υψηλά έσοδα. Δηλαδή τα έσοδα της Χ ήταν 500.000 ευρώ, ενώ το διάμεσο ύψος εσόδων της Χ είναι 430.000 ευρώ. Για την Υ τα έσοδα ήταν 100.000 ευρώ, ενώ το διάμεσο ύψος εσόδων της Υ είναι 85.000 ευρώ.

ΔΙΑΓΡΑΜΜΑΤΑ

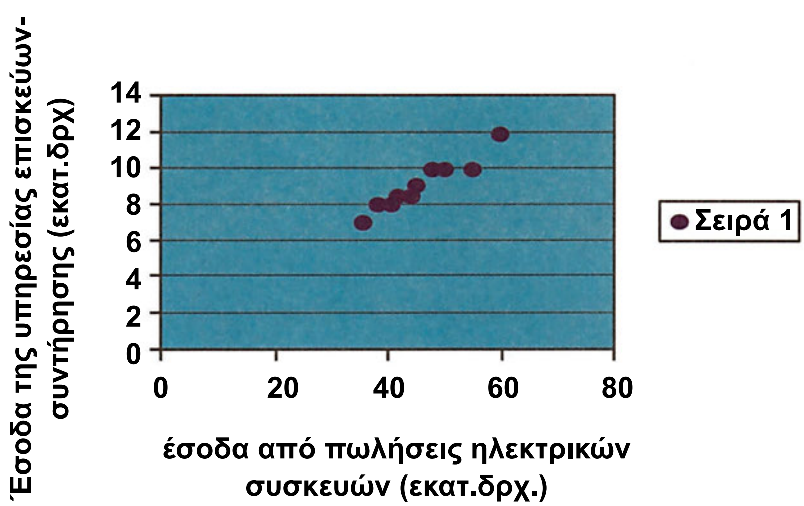
1) Σημειόγραμμα της μεταβλητής X (έσοδα από πωλήσεις ηλεκτρικών συσκευών)



2) Σημειόγραμμα της μεταβλητής Y (έσοδα από επισκευές ηλεκτρικών συσκευών)



3) Διάγραμμα διασποράς των X και Y



ΠΑΡΑΡΤΗΜΑ

Α. Συμβολισμοί – Στατιστικό Τυπολόγιο
Β. Λεξιλόγιο
Γ. Βιβλιογραφία

Α. ΣΥΜΒΟΛΑ – ΣΤΑΤΙΣΤΙΚΟ ΤΥΠΟΛΟΓΙΟ

$Α_1$ ΣΥΜΒΟΛΑ

$\hat{\alpha}$	το σημείο τομής του κάθετου άξονα y από την ευθεία παλινδρόμησης
α_i	το ανώτερο όριο i κλάσης μιας κατανομής
α_{i-1}	το κατώτερο όριο της κλάσης μιας κατανομής
$\hat{\beta}$	ο συντελεστής διεύθυνσης της ευθείας παλινδρόμησης ή γωνιακός συντελεστής (συντελεστής παλινδρόμησης)
CV	ο συντελεστής μεταβλητότητας
c_i	το εύρος (πλάτος) της i κλάσης μιας κατανομής
Ε.Τ.	επικρατούσα τιμή
ε	ο όρος τυχαίο σφάλμα του μοντέλου παλινδρόμησης, $\varepsilon = \hat{y} - y$
f	σχετική συχνότητα (βλ. πίνακες κατανομής συχνοτήτων)
F	η αθροιστική συχνότητα (βλ. διάμεσο, τεταρτημόρια)
Q_1, Q_2, Q_3	το πρώτο, το δεύτερο και το τρίτο τεταρτημόριο
δ	η διάμεσος τιμή του x
M και m	η μεγαλύτερη και η μικρότερη αντίστοιχα τιμή της μεταβλητής X
Μ.Α.Α.	η μέση απόλυτη απόκλιση
Μ.Ε.	μέσο εύρος
μ	ο μέσος αριθμητικός του εξεταζόμενου πληθυσμού
μ_x	η μέση τιμή της μεταβλητής X – περίπτωση γραμμικής παλινδρόμησης πληθυσμού
μ_y	η μέση τιμή της μεταβλητής Y – περίπτωση γραμμικής παλινδρόμησης πληθυσμού
v_i	η συχνότητα της i κλάσης
N_i	η αθροιστική συχνότητα της i κλάσης
v	ο αριθμός των παρατηρήσεων του εξεταζόμενου δείγματος
r	ο συντελεστής γραμμικής συσχέτισης
R	το εύρος
σ	η τυπική απόκλιση πληθυσμού

σ^2	η διακύμανση πληθυσμού
s	η τυπική απόκλιση δείγματος
s^2	η διακύμανση δείγματος
Σ	το άθροισμα
$\Sigma\nu_i$	το πλήθος των παρατηρήσεων του εξεταζόμενου πληθυσμού
X	α) μεταβλητή, β) ανεξάρτητη μεταβλητή της εξίσωσης της ευθείας παλινδρόμησης
$\bar{x}$	ο μέσος αριθμητικός δείγματος
Y	α) μεταβλητή, β) εξαρτημένη μεταβλητή της ευθείας της εξίσωσης παλινδρόμησης
$\hat{y}$	α) η εκτιμώμενη ή προβλεπόμενη τιμή που δίνεται από την εξίσωση παλινδρόμησης, β) η προβλεπόμενη τιμή που δίνεται από την εξίσωση της ευθείας τάσης.

A_2 ΣΤΑΤΙΣΤΙΚΟ ΤΥΠΟΛΟΓΙΟ

Ι. Συχνότητες

$$f_i = \frac{v_i}{v}$$

$$f_i\% = \frac{v_i}{v} \cdot 100$$

ΙΙ. Μέτρα Θέσης
Α. Μέσος Αριθμητικός

$$v = v_1 + v_2 + v_\kappa = \sum_{i=1}^{\kappa} v_i$$

α) Τύπος υπολογισμού μέσου αριθμητικού ($\bar{x}$).

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_v}{v}$$

β) Τύπος υπολογισμού του μέσου αριθμητικού ($\bar{x}$) από πίνακα συχνοτήτων:

$$\bar{x} = \frac{v_1 x_1 + v_2 x_2 + \ldots + v_\kappa x_\kappa}{v_1 + v_2 + \ldots + v_\kappa}$$

$$\bar{x} = \frac{\sum vx}{v}$$

Β. Διάμεσος

Τύπος υπολογισμού της διαμέσου (ομαδοποιημένες παρατηρήσεις σε κλάσεις):

$$\delta = \alpha_{i-1} + \frac{\frac{\nu}{2} - N_{i-1}}{\nu_i} \cdot c$$

Γ. Τεταρτημόρια

(Ομαδοποιημένες παρατηρήσεις κατά κλάσεις).
Τύπος υπολογισμού πρώτου (Q_1) και τρίτου (Q_3) τεταρτημορίου:

$$Q_1 = \alpha_{i-1} + \frac{\frac{\nu}{4} - N_{i-1}}{\nu_i} \cdot c$$

$$Q_3 = \alpha_{i-1} + \frac{\frac{3\nu}{4} - N_{i-1}}{\nu_i} \cdot c$$

III. Μέτρα διασποράς

Α. Διακύμανση

α) Τύπος υπολογισμού διακύμανσης s^2.

$$s^2 = \frac{\sum (x - \bar{x})^2}{\nu}$$

$$s^2 = \frac{\sum x^2}{\nu} - \left(\frac{\sum x}{\nu}\right)^2$$

$$s^2 = \frac{\sum x^2}{\nu} - \bar{x}^2$$

β) Τύπος υπολογισμού διακύμανσης s^2 από πίνακα συχνοτήτων:

$$s^2 = \frac{\sum v(x-\bar{x})^2}{v}$$

$$s^2 = \frac{\sum vx^2}{v} - \left(\frac{\sum vx}{v}\right)^2$$

$$s^2 = \frac{\sum vx^2}{v} - \bar{x}^2$$

Β. Συντελεστής μεταβλητότητας.

Τύπος υπολογισμού μεταβλητότητας:

i) Πληθυσμού: $CV = \frac{\sigma}{\mu} \cdot 100$

ii) Δείγματος: $CV = \frac{s}{\bar{x}} \cdot 100$

IV. Παλινδρόμηση – Συσχέτιση

Α. Παλινδρόμηση

α) Σύστημα κανονικών εξισώσεων:

$$\sum y = v\hat{\alpha} + \hat{\beta}\sum x$$

$$\sum xy = \hat{\alpha}\sum x + \hat{\beta}\sum x^2$$

β) Τύποι υπολογισμών των παραμέτρων $\hat{\alpha}$ και $\hat{\beta}$:

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

$$\hat{\beta} = \frac{v\sum xy - (\sum x)(\sum y)}{v\sum x^2 - (\sum x)^2}$$

ή

$$\hat{\beta} = \frac{\sum xy - \nu\bar{x}\cdot\bar{y}}{\sum x^2 - \nu\bar{x}^2}$$

Β. Συσχέτιση

Τύπος υπολογισμού του συντελεστή συσχέτισης r:

$$r = \frac{\nu\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\nu\sum x^2 - \left(\sum x\right)^2}\cdot\sqrt{\nu\sum y^2 - \left(\sum y\right)^2}}$$

V. Χρονολογικές σειρές

Α. Απλοποιημένο σύστημα κανονικών εξισώσεων:

$$\sum y = \nu\hat{\alpha}$$

$$\sum yt = \hat{\beta}\sum t^2$$

Β. Τύποι υπολογισμών των $\hat{\alpha}$ και $\hat{\beta}$:

$$\hat{\alpha} = \frac{\sum y}{\nu}$$

$$\hat{\beta} = \frac{\sum yt}{\sum t^2}$$

$\hat{\beta} > 0$, η χρονολογική σειρά έχει ανοδική τάση.
$\hat{\beta} < 0$, η χρονολογική σειρά έχει καθοδική τάση.
$\hat{\beta} = 0$, η χρονολογική σειρά έχει μηδενική τάση.

Β. ΛΕΞΙΛΟΓΙΟ

Β. ΛΕΞΙΛΟΓΙΟ

ΕΛΛΗΝΙΚΗ	ΑΓΓΛΙΚΗ
Α	
άθροισμα	= sum
αθροιστικός	= cumulative
αθροιστικά διαγράμματα	= cumulative diagrams
αθροιστική κατανομή	= cumulative frequency distribution
ακατέργαστα δεδομένα	= raw data
ακιδωτά διαγράμματα (ραβδογράμματα)	= bar charts or bar diagrams
ακρίβεια	= precision
ανάλυση	= analysis
ανάλυση διακυμάνσεως	= variance analysis
αναποτελεσματικός	= inefficient
ανεξαρτησία	= independence
ανεξάρτητα γεγονότα	= independent events
ανεξάρτητες μεταβλητές	= independent variables
άνισος	= unequal
ανισότητα	= inequality
ανοικτό διάστημα	= open internal
ανοχή	= tolerance
αντιστροφή	= inversion
αντίστροφος	= inverse
αντιπροσωπευτικός	= representative
αξία	= value
αξιοπιστία	= reliability
αποδιδόμενη στη Χ	= ascribable to X
αποδοτικότητα	= efficiency
απόκλιση	= deviation

απόλυτος	= absolute
αποτέλεσμα	= effect
άπειρο	= infinity
άπειρος	= infinite
άπειρος πληθυσμός	= infinite population
αριθμητικός	= arithmetic
αριθμητικός μέσος όρος	= arithmetic mean
αριθμοδείκτης	= index number
αριθμοδείκτης του Laspeyres	= index numbers Laspeyres
αριθμοδείκτης του Paasche	= index numbers Paasche
αριθμοδείκτης ποσότητας	= quantity index number
αρνητικός	= negative
αρνητικός αριθμός	= negative number
άρτιος αριθμός	= even number
αρχές	= principles
αρχή	= origin
αρχικός - πρωτότυπος	= original
αρχικά δεδομένα	= original data
αστάθμητος αριθμοδείκτης τιμών	= unweighted price index
ασυμπτωματικός	= asymptotic
ασυνεχής κατανομή	= discontinuous distribution
ασυσχέτιστες μεταβλητές	= uncorrelated variables
αταξινόμητα δεδομένα	= ungrouped data

Β

βασικό σύνολο	= universal set

Γ

Γενίκευση	= generalization
Γενικός	= general
γενικός δείκτης	= aggregative index
γενικός τιμάριθμος χονδρικής πώλησης	= general wholesale price index
γραφική μέθοδος	= graphic method

γραφική παράσταση	= graph
γραμμή	= line
γραμμική εξίσωση παλινδρόμησης	= linear regression equation
γραμμική συνάρτηση	= linear function
γραμμική συσχέτιση	= linear correlation
γραμμική σχέση	= linear relationship
γραμμικός	= linear
γραμμικότητα	= linearity

Δ

δεδομένα	= data
δείγμα	= sample
δειγματοληψία	= sampling
δείκτες αξίας	= value indices
δείκτης ισοτιμίας	= parity index
δείκτης μέσης αξίας	= unit value index
δείκτης όγκου	= volume index, quantum index
δείκτης τιμών	= price index
δείκτης τιμών καταναλωτή	= price consumer index
δείκτης τιμών λιανικής πώλησης	= retail price index
δεκατημόριο	= decile
διάγραμμα	= chart, diagram
διακριτή μεταβλητή	= discrete variable
διακριτός, διακεκριμένος	= discrete
διακύμανση	= fluctuation, variance
διάμεσος	= medium
διασπορά	= dispersion
διάσταση	= dimension
διάστημα	= internal
διάστημα κλάσης	= class internal
διάταξη	= permutation

Ε

Εθνική Στατιστική Υπηρεσία	= National Statistical Service

ειδοδιαγράμματα	= pictograms or pictographs
εκατοστιαία σημεία	= percentiles or percentile values
εκατοστημόριο	= percentile
εκθετική καμπύλη	= exponential curve
εκθετική παλινδρόμηση	= exponential regression
εκθετικός	= exponential
εκτιμηθείσα εξίσωση παλινδρόμησης	= estimated regression equation
εκτίμηση	= estimation
εκτίμηση με τη μέθοδο ελάχιστων τετραγώνων	= least square estimate
ελάχιστος	= minimum
έλεγχος ποιότητας	= quality control
εξάρτηση	= dependence
εξαρτημένη μεταβλητή	= dependence variable
εξίσωση παλινδρόμησης	= regression equation
επικρατούσα τιμή της μεταβλητής ή τύπος	= mode
επίπεδο	= plane
εποχική μεταβολή	= seasonal variation
επιχείρηση	= operation
επιχειρηματική πρόγνωση	= business forecasting
επιχειρηματικός κύκλος	= business cycle
ερμηνεία	= interpretation
ερμηνευτική μεταβλητή	= explanatory variable
ετήσιος μέσος	= annual mean
ευθεία	= line
ευθεία ελάχιστων τετραγώνων	= least squares line
ευθεία παλινδρόμησης	= regression line

Ζ

ζεύγος	= pair

Θ

Θεμελιώδης, βασικός	= fundamental

Θεωρητικός	= theoretical

Ι

Ιδιότητα	= property
Ιστόγραμμα	= histogram
ιστορικές ή χρονολογικές σειρές	= historical series or time series

Κ

καμπύλη	= curve
καμπύλη κατανομής	= distribution curve
καμπύλη παλινδρόμησης	= regression curve
καμπύλη συχνοτήτων	= frequency curve
καμπυλόγραμμος παλινδρόμηση	= curvilinear regression
κανονικές εξισώσεις	= normal equations
κανονική καμπύλη	= normal curve
κανονικός	= normal
κατανομή	= rule
κανόνας	= distribution
κατανομή συχνότητας	= frequency distribution
κατάταξη δεδομένων	= classification of data
κατηγορία, κλάση	= category
κεντρική τάση	= central tendency
κλάση (τάξη)	= class
κλειστό διάστημα	= closed interval
κλίμακα	= scale
κλίση	= slope
κομμάτι, τμήμα	= block
κυκλικό διάγραμμα	= pie diagram
κωδικοποιημένη μέθοδος	= coding method

Λ

λόγος	= ration

Μ

μέγεθος δείγματος	= sample size
μέγεθος πληθυσμού	= population size
μεγιστοποίηση	= maximizing
μέγιστος	= maximum
μέθοδος	= method
μέθοδος ελάχιστων τετραγώνων	= least squares method
μέσα	= means
μέση απόκλιση	= average deviation, mean deviation
μέση παραγωγή	= average production
μέση τιμή	= mean
μέση τιμή δείγματος	= sample mean
μέσος όρος	= average
μεταβλητότητα	= variation
μεταβλητή	= variable
μετασχηματισμός	= transformation
μέτρα διασποράς	= measures of dispersion
μέτρα κεντρικής τάσης	= measures of central tendency
μετρήσιμος	= measurable
μετρήσιμο σύνολο	= measurable set
μέτρηση	= measurement
μέτρο	= measure
μη γραμμική σχέση	= non linear relationship
μη ευθύγραμμη συσχέτιση	= non linear correlation
μηνιαίος δείκτης	= monthly index
μονάδα	= unit
μορφή, σχήμα	= shape

Ο

οικονομική στατιστική	= economic statistic
ολικός, σύνολο	= total
ορθογώνιος	= orthogonal
όρια διαστήματος τάξης	= class limits
όρια τάξης	= class boundaries
όριο, σύνορο	= boundary

όριο τάξης, κατώτερο, ανώτερο	= boundary lower, upper
όρος, συνθήκη	= condition
όρος	= term

Π

παλινδρόμηση	= regression
παραβολή	= parabola
παραβολική παλινδρόμηση	= parabolic regression
παράγοντας	= factor
παραγωγή	= production
παράμετρος	= parameter
παράμετρος διασποράς	= dispersion parameters
παρατήρηση	= observation
παρατηρούμενη συχνότητα	= observed frequency
πεπερασμένος	= finite
πεπερασμένος πληθυσμός	= finite population
περιγραφικό	= descriptive
περιγραφική στατιστική	= descriptive statistics
περίοδος βάσης και έτος βάσης	= base period or base year
περιττός	= odd
περιττός αριθμός	= odd number
πηλίκο	= quotient
πιθανότητα	= probability
πίνακας	= table
πίνακας συχνότητας	= frequency table
πίνακας ταξινόμησης	= classification table
πλάτος ή διάστημα τάξης	= width or class interval
πληθυσμός	= population
ποιότητα	= quality
ποιοτικά δεδομένα	= attribute data
πολύ σημαντικό αποτέλεσμα	= highly significant result
πολύγωνο	= polygon
πολύγωνο αθροιστικής συχνότητας	= cumulative frequency polygon
πολύγωνο συχνότητας	= frequency polygon or polygon graph
πολύπλοκος, σύνθετος	= complex
ποσοστιαία μεταβολή	= percentage change

ποσοστό	= percentage
ποσότητα	= quantity
πρόβλεψη	= prediction
προσαρμογή	= fitting
προσαρμογή καμπύλης	= curve fitting
προσαρμοσθείσα γραμμή	= fitted line
προσαρμοστική ικανότητα	= goodness of fit
προσέγγιση	= approximation
προσωρινά στοιχεία	= provisional data

Σ

σημείο	= point
σταθερά	= constant
σταθμικοί αριθμοδείκτες τιμών	= weighted price indices
σταθμικός δείκτης σχετικών τιμών	= weighted relative price index
σταθμικός δείκτης τιμών	= weighted aggregative price index
στατιστικά δεδομένα	= statistical data
στατιστική ανάλυση	= statistical analysis
στατιστική απόφαση	= statistical decision
στατιστική υπόθεση	= statistical hypothesis
στατιστική	= statistics
στατιστική επιχειρήσεων	= business statistics
στατιστικό μέτρο θέσεως	= measure of location
στήλη	= column
στοιχεία στατιστικής	= elements of statistics
στοιχειώδες γεγονός	= elementary event
στοιχειώδης	= elementary
συλλογή	= collection
συλλογή δεδομένων	= gathering of data
σύμβολο αθροίσεως	= summation signs
συμβολισμός	= notation
σύμβολο	= symbol
συμπέρασμα	= inference
συνάρτηση	= function
συνεχής	= continuous
συνεχής μεταβλητή	= continuous variable
συνιστώσα	= component

συντελεστές στάθμισης	= weights
συντελεστής	= coefficient
συντελεστής γραμμικής συσχέτισης	= linear correlation coefficient
συντελεστής μεταβλητότητας	= coefficient of variability
συντελεστής παλινδρόμησης	= coefficient of regression
συντελεστής συσχέτισης	= coefficient of correlation
συντεταγμένες	= coordinates
συσχέτιση	= correlation
συχνότητα	= frequency
συχνότητα τάξης	= class frequency
σφάλμα	= error
σφάλμα δειγματοληψίας	= error of sampling
σχέση	= relation, relationship
σχέση ποσότητας και τιμής	= quantity price relation
σχετικός	= relative
σχετική συχνότητα	= relative frequency
σχετική τιμή	= price relative

Τ

ταξινόμηση	= classification
ταξινομημένα σε ομάδες δεδομένα	= grouped data
τέλεια (πλήρης) γραμμική συσχέτιση	= perfect linear correlation
τετμημένη	= abscissa
τεχνική δειγματοληψίας	= sampling technique
τιμή, αξία	= value
τομή	= intercept
τομή (με τον άξονα y)	= intercept (only y – axis)
τρίγωνο	= triangle
τυπικό αποτέλεσμα	= standard score
τυπική μονάδα	= standard unit
τύπος	= formula, type
τυχαίος	= random
τυχαίος αριθμός	= random number
τυχαιότητα	= randomization

Υ

υπόδειγμα	= model
υποσύνολο	= subset
υπολογισμός	= calculation

Φ

φύλλο	= sheet

Χ

χαρτοδιαγράμματα	= map charts
χρονοδιάγραμμα	= time chart of time series chart
χρονολογικές σειρές	= time series
χρόνος	= series
χώρος	= space

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Αδαμόπουλος – Δαμιανού – Σβέρκος, *Μαθηματικά Γενικής Παιδείας, Ο.Ε.Δ.Β.,* Αθήνα 1999.

2. Αθανασόπουλος Α. Δημήτριος, *Περιγραφική Στατιστική,* 3 τόμοι, Εκδόσεις Α. Σταμούλης, Πειραιάς 1989.

3. Αθανασόπουλος Α. Δημήτριος – Μπένος Κ. Βασίλειος, *Εφαρμογές Στατιστικής,* Εκδόσεις Α. Σταμούλης, Πειραιάς 1990.

4. Αλεξανδρόπουλος Αν. – Κατωπόδης Επ. – Παλιάτσος Αθ. – Πρεζεράκος Ν., *Στατιστική,* Σύγχρονη Εκδοτική, Αθήνα 1994.

5. Γεωργιακώδης Φώτιος – Γιαλαμάς Βασίλειος – Δίκαρος Δημήτρης – Κόκλα Άννα-Μαρία, *Εισαγωγική στη Στατιστική,* Γ΄ Λυκείου, Αθήνα 1999.

6. Καλαματιανού Α., *Κοινωνική Στατιστική,* Εκδόσεις «Το Οικονομικό».

7. Καραγεώργος Δημήτρης – Κατσουλάκης Α. – Καπετάνου Ε., *Μαθηματικά Ενιαίου Πολυκλαδικού Λυκείου* Ο.Ε.Δ.Β.

8. Κιόχος Α. Πέτρος, *Στατιστική,* Εκδόσεις «Ευγενιδίου Ιδρύματος», Αθήνα 1987.

9. Παπακωνσταντίνου Ευάγγελος – Καΐτσας Γεώργιος, *Στατιστική,* Β΄ τάξη Τ.Ε.Λ., Εκδόσεις «Ευγενιδίου Ιδρύματος», Αθήνα 1995.

10. Παπακωνσταντίνου Ευάγγελος – Καΐτσας Γεώργιος, *Εφαρμογές Στατιστικής,* «Βιβλιοεκδοτική Αναστασάκη», Αθήνα 1997.

11. Παρασκευόπουλος Ι., *Στοιχεία Περιγραφικής και Επαγωγικής Στατιστικής,* Αθήνα 1984.

ΞΕΝΗ

1. Calot G., *Cours de statistique descriptive.*

2. *Economist, Numbers Guide*. The Economist Books Ltd and Hamish Hamilton Ltd. Great Britain 1993.

3. Eppen, G. D., Gould, G. J. and Schmidt, C. P. *, Introductory Management Science*, 3rd ed. Englewood Cliffs, N. J.: Prentice hall, Inc., 1991.

4. Freund, E.J., Williams, J.F. and Perles, M.B., *Elementary Business Statistics*, sixth edition, N. Jersey: Prentice Hall Intern. Inc., 1993.

5. Gawshaw J. – Chambers J., *A concise course in A Level Statistics*, 1994.

6. Jefrey, R. C., *The Logic of Decision*, 3rd ed. Chicago: University of Chicago, Press, 1990.

7. Lecaillon J., *Cours de statistique*. Licence 1re A.

8. Mann, P.S., *Introductory Statistics*. John Wiley and Sons, Inc., USA, 1992.

9. Madsen R. – Moeschberger M., *Statistical concepts with applications to business and economics*, 1986.

10. Morice E. et Chartier F., *Méthode statistique. Deuxiéme partie: Analyse statistique*.

11. *Problem Solver Statistics*, REA’s , 1991.

Κωδικός βιβλίου: 0-24-0199
ISBN 978-960-06-2976-7

ΙΤΥΕ
"ΔΙΟΦΑΝΤΟΣ"
ινστιτουτο
τεχνολογιας
υπολογιστων & εκδοσεων

(01) 000000 0 24 0199 7