ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΜΕΣΟΛΟΓΓΙΟΥ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ & ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ

Π Τ Υ Χ Ι Α Κ Η   Ε Ρ Γ Α Σ Ι Α

Η ΣΤΑΤΙΣΤΙΚΗ ΩΣ ΕΡΓΑΛΕΙΟ ΑΝΑΠΤΥΞΗΣ ΤΩΝ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΜΠΟΥΡΔΟΥ ΚΩΝ/ΝΟΣ ΒΑΣΙΛΙΚΗ - ΠΑΓΩΝΑ

ΕΙΣΗΓΗΤΗΣ:

ΜΕΓΑΡΙΤΗΣ ΑΘΑΝΑΣΙΟΣ

Μ Ε Σ Ο Λ Ο Γ Γ Ι   2 0 1 3

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΜΕΣΟΛΟΓΓΙΟΥ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ & ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ

Π Τ Υ Χ Ι Α Κ Η Ε Ρ Γ Α Σ Ι Α

Η ΣΤΑΤΙΣΤΙΚΗ ΩΣ ΕΡΓΑΛΕΙΟ ΑΝΑΠΤΥΞΗΣ ΤΩΝ
ΕΠΙΧΕΙΡΗΣΕΩΝ

ΜΠΟΥΡΔΟΥ ΚΩΝ. ΒΑΣΙΛΙΚΗ - ΠΑΓΩΝΑ (Α.Μ. 14677)

vasibour@logistiki.teimes.gr

ΕΙΣΗΓΗΤΗΣ:

ΜΕΓΑΡΙΤΗΣ ΑΘΑΝΑΣΙΟΣ

Μ Ε Σ Ο Λ Ο Γ Γ Ι 2 0 1 3

Τ.Ε.Ι.
ΜΕΣΟΛΟΓΓΙΟΥ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ & ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ

ΕΙΣΗΓΗΤΙΚΗ ΕΚΘΕΣΗ ΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

υποβάλλεται ενσωματωμένη σε κάθε αντίτυπο της Πτυχιακής Εργασίας

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

Η ΣΤΑΤΙΣΤΙΚΗ ΩΣ ΕΡΓΑΛΕΙΟ ΑΝΑΠΤΥΞΗΣ ΤΩΝ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΦΟΙΤΗΤΕΣ

ΕΠΩΝΥΜΟ	ΟΝΟΜΑ	ΑΡ. ΜΗΤΡ.
ΜΠΟΥΡΔΟΥ	ΒΑΣΙΛΙΚΗ-ΠΑΓΩΝΑ	14677

ΕΚΘΕΣΗ ΕΓΚΡΙΣΗΣ ΕΙΣΗΓΗΤΗ ΕΚΠΑΙΔΕΥΤΙΚΟΥ *καταχώρηση από τον εισηγητή*

Η παρούσα πτυχιακή εργασία πραγματοποιήθηκε υπό την καθοδήγησή μου. Το θέμα της εργασίας είναι η Στατιστική και ο ρόλος της στην ανάπτυξη των επιχειρήσεων. Ιδιαίτερη έμφαση δίνεται στις εφαρμογές της Στατιστικής στο χώρο των επιχειρήσεων. Κατά τη γνώμη μου, το θέμα αντιμετωπίστηκε επαρκώς, η δομή είναι ικανοποιητική και συνεπώς εισηγούμαι την έγκρισή της.

ΗΜΕΡΟΜΗΝΙΑ 25/5/2013

ΥΠΟΓΡΑΦΗ

Απαραίτητη η υπογραφή του εισηγητή εκπαιδευτικού ……………………..……………….

ΠΡΟΛΟΓΟΣ – ΕΙΣΑΓΩΓΗ

Η πτυχιακή εργασία αυτή καλύπτει την ύλη του μαθήματος «Στατιστική Επιχειρήσεων». Για τη συγγραφή αυτής της εργασίας λήφθηκε υπόψη η ελληνική εμπειρία για παρόμοια εγχειρίδια και γενικότερες επιστημονικές απόψεις για την κατανόηση της Στατιστικής στη Γενική και Επαγγελματική Εκπαίδευση.

Η Στατιστική είναι ο κλάδος των εφαρμοσμένων μαθηματικών, ο οποίος βασίζεται σε ένα σύνολο αρχών και μεθοδολογιών για:

Το σχεδιασμό της διαδικασίας συλλογής δεδομένων, τη συνοπτική και αποτελεσματική παρουσίαση των δεδομένων, την ανάλυση και εξαγωγή συμπερασμάτων από τα δεδομένα.

Η στατιστική είναι μια ατελής επαγωγή. Από τις ιδιότητες του μέρους εξάγει συμπεράσματα για το όλον.

Σκοπός της εργασίας είναι να αποκτήσει ο αναγνώστης τα απαραίτητα αναλυτικά εργαλεία για το σχεδιασμό και την υλοποίηση ερευνών αγοράς καθώς και η κατανόηση των βασικών τεχνικών της Στατιστικής.

Η ανάπτυξη της ύλης έγινε με τέτοιο τρόπο, ώστε:

α) Να ανταποκρίνεται στις νοητικές δυνατότητες, στις ανάγκες και στα ενδιαφέροντα των φοιτητών, για τους οποίους προορίζεται, β) να βοηθήσει στην ανάπτυξη της κριτικής ικανότητας και δημιουργικότητας των φοιτητών και γ) να δείξει τη χρησιμότητα της Στατιστικής σε διάφορες άλλες επιστήμες, καθώς και στην αντιμετώπιση πραγματικών προβλημάτων.

Το περιεχόμενο της πτυχιακής εργασίας χωρίζεται σε έξι κεφάλαια. Στο πρώτο κεφάλαιο αναπτύσσονται με συντομία ο σκοπός, η ιστορία και το αντικείμενο της στατιστικής, καθώς και ορισμοί βασικών όρων που θα χρησιμοποιηθούν στα επόμενα κεφάλαια.

Στο δεύτερο κεφάλαιο αναπτύσσονται οι μέθοδοι συλλογής και παρουσίασης των στατιστικών δεδομένων από το χώρο των επιχειρήσεων καθώς και η χρησιμοποίηση τους στις επιχειρήσεις.

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

Στο τέταρτο κεφάλαιο παρουσιάζονται τα βασικά μέτρα διασποράς μιας κατανομής, ο τρόπος υπολογισμού τους και η χρησιμότητα τους στην ερμηνεία της συμπεριφοράς του φαινομένου που μελετάμε.

Στο πέμπτο 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, καθώς και το συντελεστή προσδιορισμού.

Τέλος, στο έκτο κεφάλαιο αναφέρεται ο ρόλος της στατιστικής στις επιχειρήσεις καθώς και εφαρμογές βασισμένες στην ύλη των προηγούμενων κεφαλαίων που αποδεικνύουν την χρησιμότητα της στατιστικής, τόσο στην καθημερινή ζωή μας, όσο και στην ανάπτυξη των επιχειρήσεων.

ΕΠΙΣΗΜΑΝΣΗ

Οι διαπιστώσεις, τα αποτελέσματα, τα συμπεράσματα και οι πιθανές προτάσεις της παρούσας Πτυχιακής Εργασίας, εκτός των αναφορών που σημαίνονται ως λήμματα, αποτελούν προσωπικές θεωρητικές ή εμπειρικές διαπιστώσεις της φοιτήτριας που την επιμελήθηκε και δεν απηχούν κατ' ανάγκη τη γνώμη του εισηγητή εκπαιδευτικού, ή του Εκπαιδευτικού Προσωπικού του Τμήματος Λογιστικής ή του Α.Τ.Ε.Ι. Μεσολογγίου.

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 4

Τα βασικά μέτρα διασποράς για την επιχείρηση

ΚΕΦΑΛΑΙΟ 5

Απλή Γραμμική Παλινδρόμηση

ΚΕΦΑΛΑΙΟ 6

Στατιστική και Επιχειρήσεις

ΠΕΡΙΛΗΨΗ

Ως έννοια η Στατιστική είναι ένα εργαλείο για να δημιουργήσουμε μια νέα αντίληψη από ένα σύνολο αριθμών, είναι ένας τρόπος με τον οποίο αντλούμε πληροφορίες από δεδομένα. Γίνεται αναφορά στις βασικές στατιστικές έννοιες και ορισμούς, καθώς και σε μεθόδους συλλογής αριθμητικών πληροφοριών από το χώρο των επιχειρήσεων. Ακόμη πραγματοποιείται ταξινόμηση και παρουσίαση στατιστικών στοιχείων. Επίσης, επισημαίνονται τα μέτρα θέσης και διασποράς για την επιχείρηση. Ταυτόχρονα, τονίζεται η μεθοδολογία για τον υπολογισμό και την ερμηνεία των θεμελιωδών τεχνικών συσχέτισης και παλινδρόμησης οικονομικών μεταβλητών. Δίνεται έμφαση σε παραδείγματα και εφαρμογές από το χώρο των επιχειρήσεων, ώστε να γίνει φανερό ότι η Στατιστική είναι απαραίτητο εργαλείο στη σύγχρονη διοίκηση των επιχειρήσεων.

ΚΕΦΑΛΑΙΟ 1

ΣΤΑΤΙΣΤΙΚΗ

1.1 ΟΡΙΣΜΟΣ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

Η στατιστική είναι ο κλάδος της επιστήμης ο οποίος ασχολείται με τις μεθόδους που είναι κατάλληλες για τη συλλογή, την οργάνωση, την παρουσίαση και την ανάλυση δεδομένων, όπως επίσης, για τη χρησιμοποίηση των πληροφοριών που περιέχουν τα δεδομένα, με το σκοπό να συνταχθούν συμπεράσματα και να ληφθούν αποφάσεις, ειδικά όταν υπάρχει το στοιχείο της αβεβαιότητας. Βασίζεται στη χρήση της στατιστικής θεωρίας, ενός κλάδου των εφαρμοσμένων μαθηματικών. Η στατιστική ορίζεται επίσης σαν «η επιστήμη της λήψης αποφάσεων όταν υπάρχει η αβεβαιότητα».

Οι μέθοδοι της Στατιστικής χρησιμοποιούνται σε πολλά πεδία, όπως στην Τεχνική, στο Εμπόριο, στη Βιολογία, στην Ιατρική, στη Γεωργία, στη Ψυχολογία και Εκπαίδευση. Πρέπει να τονισθεί ότι οι μέθοδοι της Στατιστικής βοηθούν μόνο στη λήψη αποφάσεων. Πολύ σπάνια οι μέθοδοι αυτές πρέπει να χρησιμοποιούνται σαν μοναδική βάση για μια απόφαση.

Συμπερασματικά, μπορούμε να δώσουμε ως ορισμό της ΄΄ΣΤΑΤΙΣΤΙΚΗΣ΄΄[1] το συνεθέστερο και πλέον γνωστό ορισμό του R. Fisher (1890-1962)[2], πατέρα της σύγχρονης Στατιστικής:

Στατιστική είναι ένα σύνολο αρχών και μεθοδολογιών για:

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

[1] << Η στατιστική είναι η επιστήμη που μελετά τα γεγονότα κατά μάζας, δηλαδή κατά πολυάριθμους ομοειδείς μονάδας >>, W. Lexis (1837-1914), Γερμανός στατιστικός.

[2] 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.

1.2 ΙΣΤΟΡΙΑ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ

Η λέξη «Στατιστική» παράγεται από την λατινική λέξη «STATUS» που σημαίνει «κράτος».

Το φαινόμενο της ανάπτυξης των Μαθηματικών της θεωρίας Πιθανοτήτων, συνέβαλε κατά πολύ στη ανάπτυξη της σύγχρονης Στατιστικής. Επίσης, η ανάγκη των κυβερνήσεων να συλλέγουν πληροφορίες για τις βασικές δραστηριότητες των πολιτών στη Γεωργία και το Εμπόριο, όπως επίσης και για τον αριθμό των γεννήσεων και των θανάτων, ήταν ένας άλλος παράγοντας ανάπτυξης της Στατιστικής.

Η συλλογή δεδομένων γινόταν σε όλη τη διάρκεια της γραπτής ιστορίας. Κατά τη διάρκεια του Αιγυπτιακού, Ελληνικού και Ρωμαϊκού πολιτισμού, τα δεδομένα που αφορούσαν τους πολίτες μαζεύονταν για το σκοπό της επιβολής φορολογίας και στρατιωτικής θητείας.

Στους Βυζαντινούς χρόνους, συνήθως, τα μοναστήρια και οι διευθύνσεις εκκλησιαστικών περιφερειών κρατούσαν στοιχεία που αφορούσαν γεννήσεις, γάμους, θανάτους κτλ.

Στην Αμερική, διάφορα δεδομένα συλλέχθηκαν κατά τους χρόνους της αποικιοκρατίας και το 1970 το Ομοσπονδιακό Σύνταγμα καθιέρωσε τη διεξαγωγή απογραφής κάθε 10 χρόνια.

Φημισμένοι πιθανοθεωρητικοί και στατιστικοί επιστήμονες είναι: οι Άγγλοι R. Fisher, K. Pearson, M. Stone, ο Πολωνός J. Newman, οι Ρώσοι Kolmogorov, Chebyshev, Markov, οι Αμερικανοί Tukey, B. Efron, ο Γερμανός R. Von Mises και άλλοι.

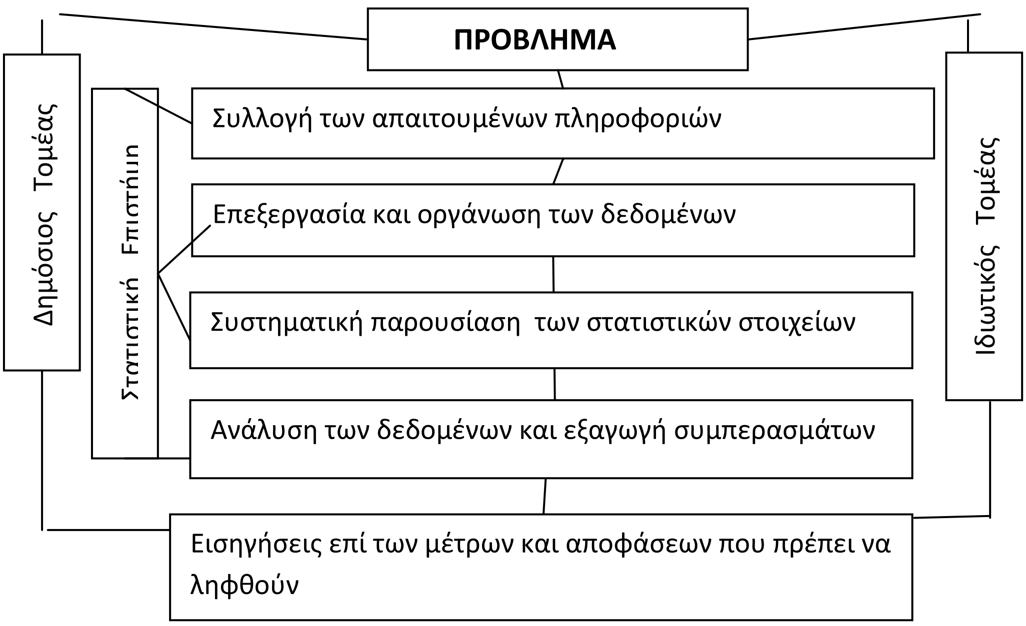
1.3 ΤΟ ΑΝΤΙΚΕΙΜΕΝΟ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ

Το αντικείμενο της στατιστικής επιστήμης συνίσταται σήμερα στην αποτελεσματική αξιοποίηση πληροφοριών μετά από κατάλληλη κατά περίπτωση συλλογή, επεξεργασία, οργάνωση, παρουσίαση και ανάλυση αριθμητικών δεδομένων, έτσι ώστε να διευκολύνεται είτε ο δημόσιος είτε ο ιδιωτικός τομέας στη σωστή λήψη μέτρων και αποφάσεων.

Η στατιστική επιστήμη με τη μορφή της Στατιστικής των Επιχειρήσεων είναι ένα χρήσιμο εργαλείο για τη μελέτη διαφόρων προβλημάτων της ιδιωτικής πρωτοβουλίας και ειδικότερα της επιχειρηματικής δραστηριότητας, όπως για την παρακολούθηση των αγορών, των πωλήσεων και των κερδών μιας επιχείρησης, για τον έλεγχο της ποιότητας των παραγόμενων προϊόντων, για την άσκηση σωστής τιμολογιακής, επενδυτικής, πιστωτικής πολιτικής κλπ.

Η παρεμβολή της στατιστικής στο κύκλωμα πρόβλημα-πληροφόρηση-λήψη απόφασης φαίνεται από το σχήμα που ακολουθεί:

Σχήμα 1



1.4 ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ – ΕΝΝΟΙΑ ΚΑΙ ΠΕΡΙΕΧΟΜΕΝΟ

Ανάλογα με τον τρόπο που χρησιμοποιούμε τη Στατιστική διακρίνουμε σήμερα δύο μεγάλους τομείς της: την Περιγραφική Στατιστική και την Επαγωγική Στατιστική.

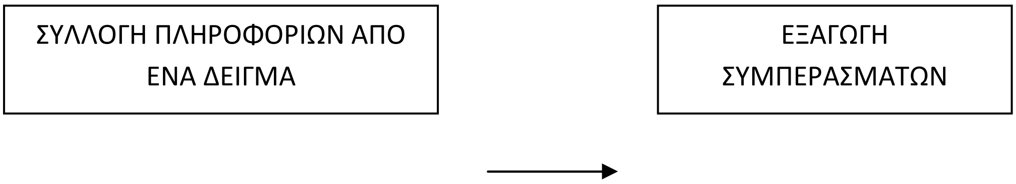
Ο πρώτος τομέας, η Περιγραφική Στατιστική περιλαμβάνει τις μεθόδους για τη συλλογή, ταξινόμηση, περιγραφή και παρουσίαση ενός συνόλου δεδομένων. Συνήθως τα δεδομένα που έχουμε στη διάθεσή μας σε μια στατιστική μελέτη είναι πολλά και σε ακατάστατη μορφή, με αποτέλεσμα να μην μπορούμε να διακρίνουμε την 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 χρησιμεύουν σε γενικεύσεις που αφορούν ένα ευρύτερο σύνολο δεδομένων (πληθυσμός).

1.5 ΕΠΑΓΩΓΙΚΗ ΣΤΑΤΙΣΤΙΚΗ – ΕΝΝΟΙΑ ΚΑΙ ΠΕΡΙΕΧΟΜΕΝΟ

Ο δεύτερος τομέας, η Επαγωγική Στατιστική ή Στατιστική Συμπερασματολογία, περιλαμβάνει μεθόδους που μας βοηθούν να καταλήξουμε σε συμπεράσματα από το μέρος (δείγμα) τα οποία μπορούμε να μελετήσουμε με τη βοήθεια των μεθόδων της Περιγραφικής Στατιστικής στο σύνολο (πληθυσμό) 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 για παράδειγμα μια φαρμακευτική εταιρεία από τη μελέτη της συμπεριφοράς ενός δείγματος (ομάδα Α και Β ασθενών) βγάζει το συμπέρασμα ποιο από τα δύο φάρμακα είναι το καλύτερο για να χρησιμοποιηθεί στο σύνολο των ασθενών. Γι ΄ αυτό πολλές φορές λέμε η Στατιστική δημιουργεί τις αποφάσεις.

Η διαδικασία αυτή μπορεί να αποδοθεί σχηματικά ως εξής:

Σχήμα 2



1.6 Η ΕΠΙΔΡΑΣΗ ΤΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ ΣΤΗΝ ΕΞΑΓΩΓΗ ΣΤΑΤΙΣΤΙΚΩΝ ΣΥΜΠΕΡΑΣΜΑΤΩΝ

Ο ηλεκτρονικός υπολογιστής έχει αυξήσει κατά πολύ τη χρήση των στατιστικών μεθόδων. Επειδή ο Η/Υ μπορεί να κάνει εκατομμύρια υπολογισμούς σε δευτερόλεπτα, κατέστησε εφικτή τη χρήση στατιστικών μεθόδων σε πεδία που αυτό ήταν προηγουμένως αδύνατο.

Μερικά από τα σπουδαιότερα στατιστικά πακέτα που χρησιμοποιούνται σήμερα σε μικρουπολογισμούς και μεγάλες μονάδες υπολογιστών, για την στατιστική ανάλυση δεδομένων, είναι τα εξής:

- STATGRAPHICS
- Statistical Package for the Social Sciences (SPSS)
- MICROSTAT
- MINITAB
- GB-STAT
- S-Plus
- SAS
- Numerical Algorithms Group for Statistics (NAG)
- SSP

1.7 ΠΛΗΘΥΣΜΟΣ – ΜΕΤΑΒΛΗΤΕΣ - ΔΕΙΓΜΑ

Ένα σύνολο του οποίου τα στοιχεία μελετάμε ως προς κάποιο χαρακτηριστικό τους, λέγεται **πληθυσμός.**

Το χαρακτηριστικό ως προς το οποίο μελετάμε τα στοιχεία ενός πληθυσμού, ονομάζεται **μεταβλητή.**

Τις μεταβλητές τις διακρίνουμε:

1) Σε **ποιοτικές** ή **κατηγορικές** μεταβλητές, των οποίων οι τιμές τους δεν είναι αριθμοί. Τέτοιες είναι, για παράδειγμα, η ομάδα αίματος, το φύλο, οι συνέπειες του καπνίσματος, όπως επίσης και η οικονομική κατάσταση και η υγεία των ανθρώπων.

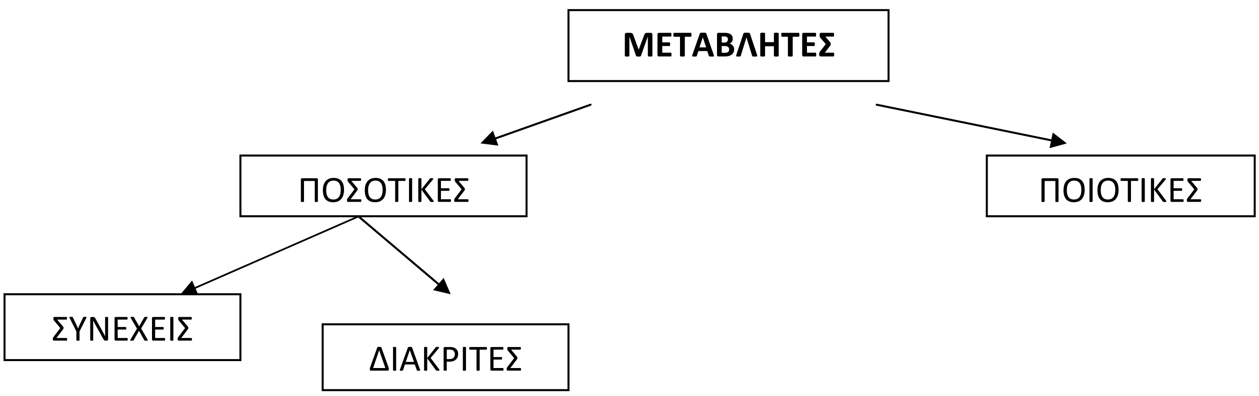
2) Σε **ποσοτικές** μεταβλητές, των οποίων οι τιμές είναι αριθμοί και διακρίνονται:

α) Σε **διακριτές** μεταβλητές, που παίρνουν μόνο ΄΄μεμονωμένες΄΄ τιμές. Τέτοιες μεταβλητές είναι, για παράδειγμα, ο αριθμός των υπαλλήλων μιας επιχείρησης (με τιμές 1,2,…), το αποτέλεσμα της ρίψης ενός ζαριού (με τιμές 1,2,…,6) κτλ.

β) Σε **συνεχείς** μεταβλητές, που μπορούν να πάρουν οποιαδήποτε τιμή ενός διαστήματος πραγματικών αριθμών (α, β). Τέτοιες μεταβλητές είναι το ύψος και το βάρος ενός ανθρώπου, η διάρκεια μιας τηλεφωνικής συνδιάλεξης κτλ.

Για την καλύτερη κατανόηση των προηγούμενων ορισμών παραθέτουμε το παρακάτω σχήμα.

Σχήμα 3



Ένας τρόπος για να πάρουμε τις απαραίτητες πληροφορίες που χρειαζόμαστε για κάποιο πληθυσμό είναι να εξετάσουμε όλα τα άτομα (στοιχεία) του πληθυσμού ως προς το χαρακτηριστικό που μας ενδιαφέρει. Η μέθοδος αυτή συλλογής των δεδομένων καλείται **απογραφή**[1].

Σε πολλές όμως περιπτώσεις η εξέταση όλων των μονάδων του πληθυσμού είναι δύσκολη ή ακόμα και αδύνατη. Επίσης, ο κόπος, ο χρόνος και τα έξοδα που χρειάζονται για τη διεξαγωγή μιας απογραφής είναι πολλές φορές αρκετά μεγάλα. Εξάλλου ένας κατασκευαστής εκρηκτικών μηχανισμών ή ηλεκτρικών λυχνιών είναι αδύνατο να δοκιμάσει όλους τους παραγόμενους μηχανισμούς, για να ελέγχει την αποτελεσματικότητα τους, ή όλες τις παραγόμενες λυχνίες για να ελέγχει το χρόνο ζωής τους.

Όπου λοιπόν η απογραφή είναι δύσκολη, αδύνατη ή οικονομικά και χρονικά ασύμφορη, ο ερευνητής μαζεύει πληροφορίες από κάποια μικρή ομάδα ή υποσύνολο του πληθυσμού, το οποίο καλείται **δείγμα**. Κάνει τις παρατηρήσεις του στο δείγμα αυτό και μετά γενικεύει τα συμπεράσματα του για ολόκληρο τον πληθυσμό. Τα συμπεράσματα όμως που θα προκύψουν από τη μελέτη του δείγματος θα είναι αξιόπιστα, θα ισχύουν δηλαδή με ικανοποιητική ακρίβεια για ολόκληρο τον πληθυσμό, αν η επιλογή του δείγματος γίνει με σωστό τρόπο, ώστε το δείγμα να είναι, όπως λέμε, **αντιπροσωπευτικό** του πληθυσμού.

Στην πράξη, ένα δείγμα θεωρείται αντιπροσωπευτικό ενός πληθυσμού, εάν έχει επιλεγεί με τέτοιο τρόπο, ώστε κάθε μονάδα του πληθυσμού να έχει την ίδια δυνατότητα να επιλεγεί.

Η επιλογή του αντιπροσωπευτικού δείγματος είναι ΄΄εκ των ων ουκ άνευ΄΄. Αποτελεί πολύ σοβαρή και δύσκολη διαδικασία. Ο κακός σχεδιασμός και η εκτέλεση της στατιστικής έρευνας, η μη αντιπροσωπευτικότητα του δείγματος, ο μη σωστός καθορισμός του μεγέθους του δείγματος αποτελούν μερικά βασικά μειονεκτήματα στη διαδικασία επιλογής ενός δείγματος.

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

Από την άλλη πλευρά, στις απογραφές απαιτείται συνήθως μεγάλος αριθμός απογραφέων.

Παρουσιάζεται έτσι η ανάγκη πρόσληψης και εκπαίδευσης μεγάλου αριθμού υπαλλήλων. Λόγω του μεγάλου χρόνου και κυρίως των σημαντικών εξόδων που απαιτούνται, πολλές φορές χρησιμοποιούνται ανεπαρκώς εκπαιδευμένοι απογραφείς με κίνδυνο να σημειώνονται λάθη οφειλόμενα σ' αυτούς.

Οι αρχές και οι μέθοδοι για τη συλλογή και ανάλυση δεδομένων από πεπερασμένους πληθυσμούς είναι το αντικείμενο της **Δειγματοληψίας**[2], που αποτελεί τη βάση της Στατιστικής. Γενικά, μπορούμε να πούμε ότι η οργάνωση της συλλογής και επεξεργασίας των σχετικών δεδομένων και πληροφοριών γίνεται κατά τρόπο που για δεδομένη ακρίβεια να επιτυγχάνεται το χαμηλότερο δυνατό κόστος ή, αντιστρόφως, να εξασφαλίζεται η μέγιστη δυνατή ακρίβεια την οποίαν επιτρέπουν τα μέσα που διαθέτουμε.

[2]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 το πληθυσμό.

1.8 ΠΙΝΑΚΕΣ ΣΥΧΝΟΤΗΤΩΝ

Ας υποθέσουμε ότι $x_1, x_2, ..., x_k$ είναι οι τιμές μιας μεταβλητής *Χ*, που αφορά τα άτομα ενός δείγματος μεγέθους v , $k \le v$. Στην τιμή x_i αντιστοιχίζεται η (απόλυτη) **συχνότητα** (frequency) v_i, δηλαδή ο φυσικός αριθμός που δείχνει πόσες φορές εμφανίζεται η τιμή x_i της εξεταζόμενης μεταβλητής *Χ* στο σύνολο των παρατηρήσεων. Είναι φανερό ότι το άθροισμα όλων των συχνοτήτων είναι ίσο με το μέγεθος *ν* του δείγματος , δηλαδή:

$$v_1 + v_2 + \cdots + v_k = v$$

Αν διαιρέσουμε τη συχνότητα v_i με το μέγεθος *ν* του δείγματος, προκύπτει η **σχετική συχνότητα** (relative frequency) f_i της τιμής x_i, δηλαδή

$$f_i = \frac{v_i}{v}, i = 1,2,....,k$$

Για τη σχετική συχνότητα ισχύουν οι ιδιότητες:

(i) $0 \le f_i \le 1$ για $i = 1,2,....,k$ αφού $0 \le v_i \le v$

(ii) $f_1 + f_2 + \cdots + f_k = 1$, αφού

$$f_1 + f_2 + \cdots + f_k = \frac{v_1}{v} + \frac{v_2}{v} + \cdots + \frac{v_k}{v} = \frac{v_1 + v_2 + \cdots + v_k}{v} = \frac{v}{v} = 1.$$

Συνήθως, τις σχετικές συχνότητες f_i τις εκφράζουμε επί τοις εκατό, οπότε συμβολίζονται με $f_i\% = 100 f_i$.

Οι ποσότητες x_i, v_i, f_i για ένα δείγμα συγκεντρώνονται σε ένα συνοπτικό πίνακα, που ονομάζεται πίνακας κατανομής συχνοτήτων ή απλά **πίνακας συχνοτήτων**.

Για μια μεταβλητή, το σύνολο των ζευγών (x_i, v_i) λέμε ότι αποτελεί την **κατανομή συχνοτήτων** και το σύνολο των ζευγών (x_i, f_i), ή των ζευγών $(x_i, f_i\%)$, την **κατανομή των σχετικών συχνοτήτων**.

Στην περίπτωση των **ποσοτικών μεταβλητών** εκτός από τις συχνότητες ν_i και f_i χρησιμοποιούνται συνήθως και οι λεγόμενες **αθροιστικές συχνότητες** N_i και οι **αθροιστικές σχετικές συχνότητες** F_i, οι οποίες εκφράζουν το πλήθος και το ποσοστό αντίστοιχα των παρατηρήσεων που είναι μικρότερες ή ίσες της τιμής x_i. Συχνά οι F_i πολλαπλασιάζονται επί 100 εκφραζόμενες έτσι επί τοις εκατό , δηλαδή $F_i\% = 100F_i$. Αν οι τιμές $x_1, x_2, ..., x_k$ μιας ποσοτικής μεταβλητής X είναι σε αύξουσα διάταξη, τότε η αθροιστική συχνότητα της τιμής x_i είναι $N_i = \nu_1 + \nu_2 + \cdots + \nu_i$. Όμοια, η αθροιστική σχετική συχνότητα είναι $F_i = f_1 + f_2 + \cdots + f_i$, για $i = 1,2,....,k$.

Είναι φανερό ότι ισχύουν οι σχέσεις:

$$\nu_1 = N_1, \nu_2 = N_2 - N_1, ..., \nu_k = N_k - N_{k-1}$$

$$f_1 = F_1, f_2 = F_2 - F_1, ..., f_k = F_k - F_{k-1}.$$

ΚΕΦΑΛΑΙΟ 2

ΣΥΛΛΟΓΗ ΚΑΙ ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΤΟ ΧΩΡΟ ΤΩΝ ΕΠΙΧΕΙΡΗΣΕΩΝ.

2.1 ΓΕΝΙΚΑ

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.

Ο κυριότερος φορέας συλλογής και δημοσίευσης στατιστικών στοιχείων στη χώρα μας είναι η Εθνική Στατιστική Υπηρεσία της Ελλάδας (Ε.Σ.Υ.Ε.). Η υπηρεσία αυτή έχει την έδρα της στην οδό Λυκούργου 14-16, κοντά στην πλατεία Ομονοίας, στην Αθήνα. Υπάγεται δε στο Υπουργείο Εθνικής Οικονομίας.

Τα σπουδαιότερα δημοσιεύματα της Ε.Σ.Υ.Ε. είναι:

α) Το μηνιαίο στατιστικό δελτίο.

β) Η (ετήσια) στατιστική επετηρίδα.

γ) Η (ετήσια) συνοπτική στατιστική επετηρίδα.

δ) Το εξωτερικό εμπόριο.

ε) Οι έρευνες των οικογενειακών προϋπολογισμών, κ.λπ.

Τα στατιστικά στοιχεία που παρέχονται από την Ε.Σ.Υ.Ε. καλύπτουν όλους τους τομείς της ανθρώπινης δραστηριότητας, όπως π.χ. την υγεία, την παιδεία, την κοινωνική ασφάλιση, τη δικαιοσύνη, τα δημόσια οικονομικά, το εμπόριο, την τουριστική κίνηση, την απασχόληση, τη γεωργία, τις τιμές, την οικονομική δραστηριότητα κ.α. Ακόμη παρέχονται πλήρη στοιχεία σχετικά με τις μεταβολές του πληθυσμού (δημογραφικά στοιχεία), καθώς και στοιχεία που αφορούν το κλίμα, το έδαφος, τη μεταποίηση, την ενέργεια κ.λπ. Για όλα τα προηγούμενα μεγέθη υπάρχουν αναλυτικά στοιχεία, τα οποία περιλαμβάνονται σε αυτοτελή δημοσιεύματα. Συστηματική σύνοψη στοιχείων περιλαμβάνεται στη ''Στατιστική επετηρίδα της Ελλάδος'' και στη ''Συνοπτική στατιστική επετηρίδα'', ενώ στο ''Μηνιαίο στατιστικό δελτίο'' δημοσιεύονται τα πλέον πρόσφατα, κάθε φορά, στατιστικά δεδομένα.

Αντίστοιχη υπηρεσία, με έδρα τις Βρυξέλλες και με το όνομα EUROSTAT καλύπτει τη συνολική δραστηριότητα των χωρών μελών της Ευρωπαϊκής Ένωσης .

Άλλοι φορείς συλλογής στατιστικών στοιχείων, με περιορισμένες όμως δυνατότητας, είναι οι διάφορες εταιρείες δημοσκόπησης οι οποίες εργάζονται κατόπιν παραγγελίας και για λογαριασμό των κομμάτων, των εφημερίδων, των καναλιών της τηλεόρασης κ.λπ. Επίσης σε συλλογή στοιχείων προβαίνουν τα τριτοβάθμια εκπαιδευτικά ιδρύματα, οι επιχειρήσεις, η Τοπική Αυτοδιοίκηση, αλλά και μεμονωμένοι ιδιώτες για χάρη της επιστημονικής έρευνας.

2.2 ΣΥΛΛΟΓΗ ΣΤΑΤΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Α) Η **απογραφή**, 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. Στη χώρα μας απογραφή του πληθυσμού γίνεται κάθε 10 χρόνια από την Ε.Σ.Υ.Ε, ενώ απογραφή των βιομηχανικών και εμπορικών επιχειρήσεων κάθε 5 χρόνια. Απογραφές γεωργίας, κτηνοτροφίας κ.τ.λ. δεν γίνονται σε καθορισμένο χρόνο.

Β) Η **δειγματοληψία** 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 ένα 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 κ.τ.λ.).

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 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.

2.3 ΤΑΞΙΝΟΜΗΣΗ ΚΑΙ ΠΑΡΟΥΣΙΑΣΗ ΣΤΑΤΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

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

- Με συνοπτικές εκθέσεις ή αναφορές
- Με στατιστικούς πίνακες
- Με διαγράμματα

2.3.1 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. Αυτές 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.

2.3.2 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 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.

2.3.3 Στατιστικά διαγράμματα

Οι στατιστικοί πίνακες κατανομής συχνοτήτων των μεταβλητών είναι ένας χρήσιμος τρόπος παρουσίασης των δεδομένων και διευκολύνει την επεξεργασίας 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 ή γραφικές απεικονίσεις, αποτελούν ένα 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

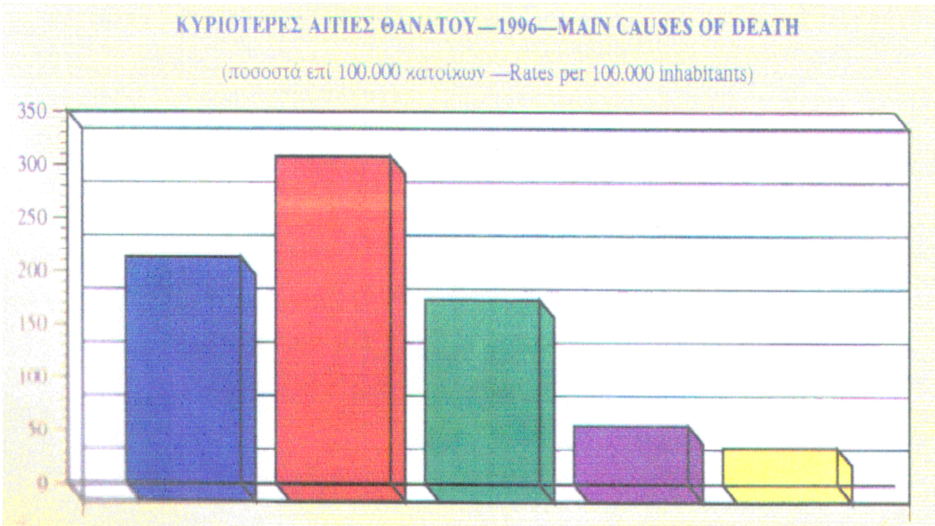
Δ) Την πηγή των δεδομένων.

Παρακάτω παραθέτουμε μερικά από τα κυριότερα είδη διαγραμμάτων :

- **<u>Ραβδόγραμμα</u>**

Το ραβδόγραμμα χρησιμοποιείται για τη γραφική παράσταση των τιμών μιας ποιοτικής μεταβλητής και αποτελείται από κατακόρυφα ή οριζόντια ορθογώνια πλάτους (με κενά μεταξύ τους) που λέγονται ράβδοι. Σε κάθε τιμή της μεταβλητής Χ αντιστοιχεί μια 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.

Πίνακας 1



Πηγή Ε.Σ.Υ.Ε, (2010).

- Νεοπλάσματα
- Καρδιακά Νοσήματα
- Νόσοι των αγγείων
- Νόσοι του αναπνευστικού συστήματος
- Ατυχήματα

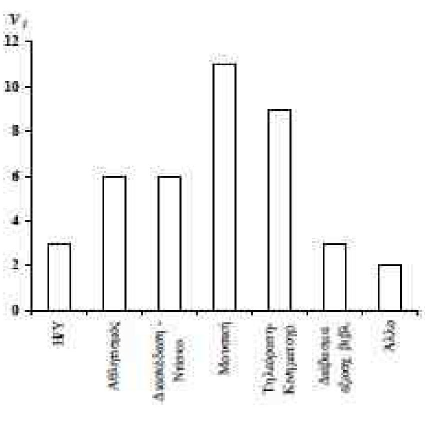
Πίνακας 2

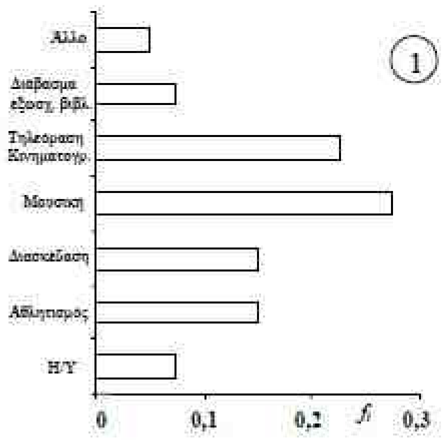
Κατανομή συχνοτήτων για την απασχόληση των φοιτητών στον ελεύθερο χρόνο τους.

I	**Απασχόληση** *x_i*	**Συχνότητα** *v_i*	**Σχετική συχνότητα** *f_i*	***Σχετική συχνότητα*** *f_i%*
1	Υπολογιστές	3	0,075	7,5
2	Αθλητισμός	6	0,150	15,0
3	Διασκέδαση	6	0,150	15,0
4	Μουσική	11	0,275	27,5
5	Τηλεόραση-Κινηματογράφος	9	0,225	22,5
6	Διάβασμα	3	0,075	7,5
7	Άλλο	2	0,050	5,0
Σύνολο		40	1,000	100,0

Μερικές φορές σε ένα ραβδόγραμμα συχνοτήτων ο ρόλος των δύο αξόνων είναι δυνατόν να αντιστραφεί, όπως φαίνεται στο Σχήμα 4 (β), που παριστάνεται το ραβδόγραμμα σχετικών συχνοτήτων της ίδιας μεταβλητής.

Σχήμα 4





(α) (β)

Ραβδόγραμμα συχνοτήτων (α) και σχετικών συχνοτήτων (β) για την απασχόληση των φοιτητών του πίνακα 2.

- **Κυκλικό διάγραμμα**

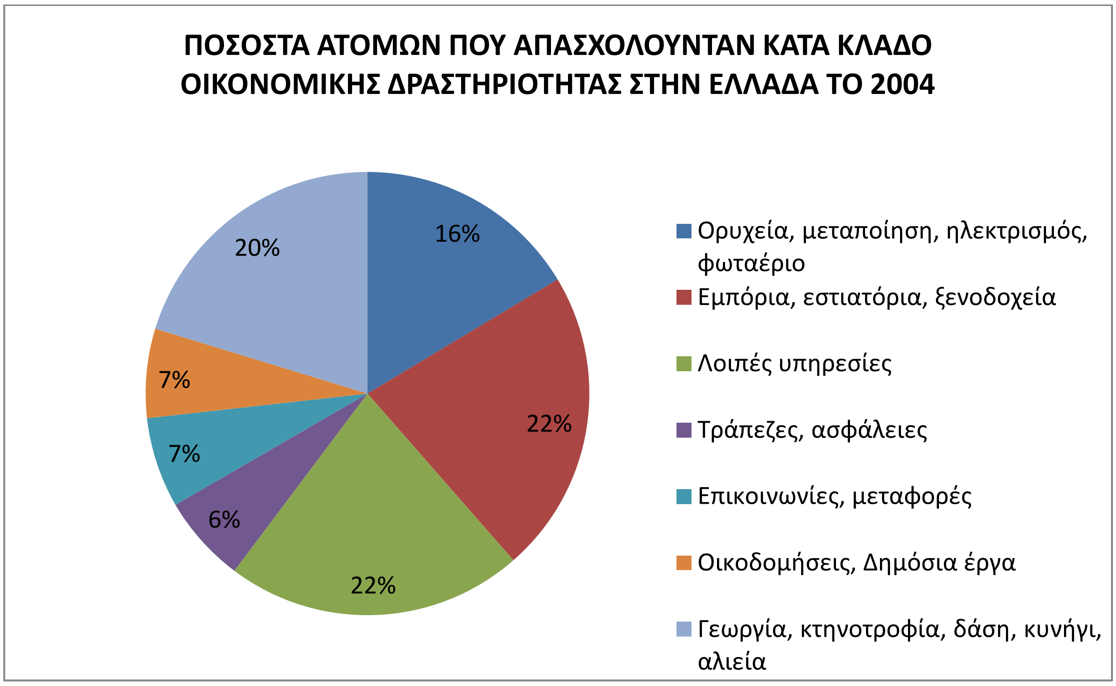
Τα κυκλικά διαγράμματα χρησιμοποιούνται κυρίως για τη γραφική απεικόνιση των σχετικών συχνοτήτων, τόσο των ποιοτικών όσο και των ποσοτικών μεταβλητών. Τις σχετικές συχνότητες τις παριστάνουμε με κυκλικούς τομείς ενός ολόκληρου κύκλου. Έτσι ολόκληρος ο κύκλος παριστάνει τη σχετική συχνότητα 100%. Το σχήμα 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^o (πλήρης κύκλος) αντιςτοιχεί στο σύνολο των παρατηρήσεων. Το τόξο a_i αντιστοιχεί στη συχνότητα v_i. Τα ποσά είναι ανάλογα οπότε:

$$\frac{a_i}{360^\circ} = \frac{v_i}{v}$$

ή ισοδύναμα $a_i = \frac{v_i}{v} = f_i.360^\circ, i = 1,2,\ldots,k.$

Πίνακας 3



Πηγή Ε.Σ.Υ.Ε, (2004).

- **Διάγραμμα συχνοτήτων**

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.

Μπορούμε επίσης αντί των συχνοτήτων v_i στον κάθετο άξονα να βάλουμε τις σχετικές συχνότητες f_i, οπότε έχουμε το διάγραμμα σχετικών συχνοτήτων. Ενώνοντας τα σημεία (x_i, v_i) ή (x_i, f_i) έχουμε το λεγόμενο **πολύγωνο συχνοτήτων** ή **πολύγωνο σχετικών συχνοτήτων**, αντίστοιχα που μας δίνουν μια γενική ιδέα για τη μεταβολή της συχνότητας ή της σχετικής συχνότητας όσο μεγαλώνει η τιμή της μεταβλητής που εξετάζουμε.

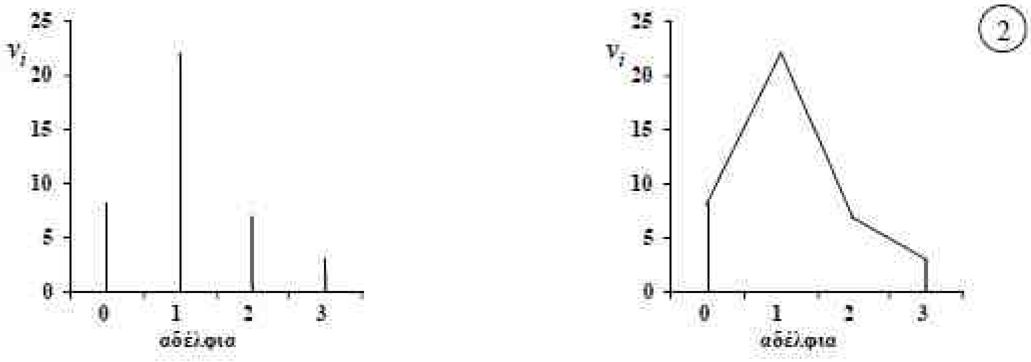
Πίνακας 4

Κατανομή συχνοτήτων της μεταβλητής X .

Αριθμός παιδιών ανά οικογένεια x_i	**Συχνότητα** v_i	**Σχετική συχνότητα** f_i	**Σχετική συχνότητα** $f_i\%$
0	8	0,200	20,0
1	22	0,550	55,0
2	7	0,175	17,5
3	3	0,075	7,5
ΣΥΝΟΛΟ	40	1,000	100,0

Είναι φανερό ότι εδώ έχουμε μια συνάρτηση με ανεξάρτητη μεταβλητή x, που παίρνει τις τιμές $x_1, x_2, ..., x_k$ και μια εξαρτημένη ν, που παίρνει τιμές $\nu_1, \nu_2, ..., \nu_k$, αντίστοιχα. Μπορούμε λοιπόν να πάρουμε ένα ορθογώνιο σύστημα αξόνων και στον οριζόντιο άξονα να πάρουμε τις τιμές x_i ενώ στο κατακόρυφο τις τιμές, οπότε έχουμε αμέσως τα σημεία $\mathrm{A}_1, \mathrm{A}_2, ..., \mathrm{A}_k$ που είναι εικόνες των ζευγών $(x_1, \nu_1), (x_2, \nu_2), \ldots, (x_k, \nu_k)$. Το σύνολο των σημείων αυτών αποτελεί το διάγραμμα συχνοτήτων.

Σχήμα 5



(α) (β)

Διάγραμμα συχνοτήτων (α) και πολύγωνο συχνοτήτων (β) για τη μεταβλητή ´´ αριθμός παιδιών ανά οικογένεια ´´.

- <u>**Σημειόγραμμα**</u>

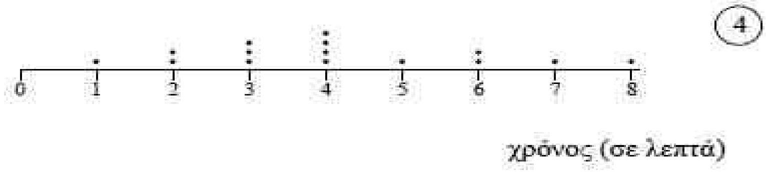
Όταν έχουμε λίγες παρατηρήσεις, η κατανομή τους μπορεί να περιγραφεί με το σημειόγραμμα, στο οποίο οι τιμές παριστάνονται γραφικά σαν σημεία υπεράνω ενός οριζόντιου άξονα.

Στο Σχήμα 6 έχουμε το σημειόγραμμα των χρόνων (σε λεπτά)

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

που χρειάστηκαν 15 φοιτητές, για να λύσουν ένα πρόβλημα.

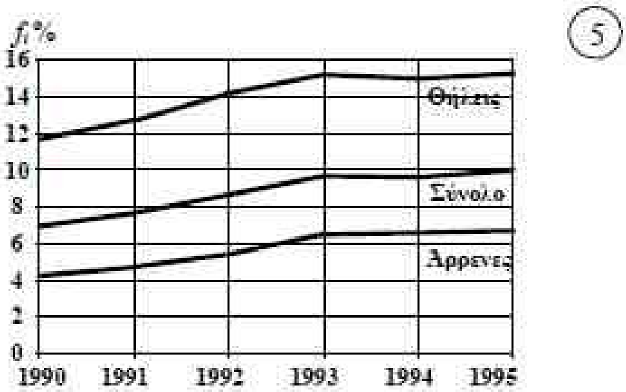
Σχήμα 6



- **<u>Χρονόγραμμα</u>**

Το χρονόγραμμα ή χρονολογικό διάγραμμα χρησιμοποιείται για τη γραφική απεικόνιση της διαχρονικής εξέλιξης ενός οικονομικού, δημογραφικού ή άλλου μεγέθους. Ο οριζόντιος άξονας χρησιμοποιείται συνήθως ως άξονας μέτρησης του χρόνου και ο κάθετος ως άξονας μέτρησης της εξεταζόμενης μεταβλητής . Στο Σχήμα 7 έχουμε το χρονόγραμμα του ποσοστού ανεργίας στη χώρα μας από το 1990 έως το 1995. (Πηγή Ε.Σ.Υ.Ε). Παρατηρούμε ότι στο γυναικείο πληθυσμό υπάρχει συστηματικά μεγαλύτερο ποσοστό ανεργίας, γύρω στις 8 εκατοστιαίες μονάδες. Στο διάστημα 1993-95 το ποσοστό ανεργίας έχει σταθεροποιηθεί γύρω στο 6,5% για τους άνδρες και γύρω στο 15% για τις γυναίκες.

Σχήμα 7



- **<u>Ιστόγραμμα Συχνοτήτων</u>**

Η αντίστοιχη γραφική παράσταση ενός πίνακα συχνοτήτων με ομαδοποιημένα δεδομένα γίνεται με το λεγόμενο ιστόγραμμα συχνοτήτων. Στον οριζόντιο άξονα ενός συστήματος ορθογωνίων αξόνων σημειώνουμε, με κατάλληλη κλίμακα, τα όρια των κλάσεων. Στη συνέχεια, κατασκευάζουμε διαδοχικά ορθογώνια (ιστούς), από καθένα από τα οποία έχει βάση ίση με το πλάτος της κλάσης και το ύψος τέτοιο, ώστε το εμβαδόν του ορθογωνίου να ισούται με τη συχνότητα της κλάσης αυτής.

Στον παρακάτω πίνακα συχνοτήτων δίνεται το ύψος (σε cm) 40 μαθητών της Γ΄ Λυκείου.

Πίνακας 5

ΚΑΤΑΝΟΜΕΣ ΣΥΧΝΟΤΗΤΩΝ

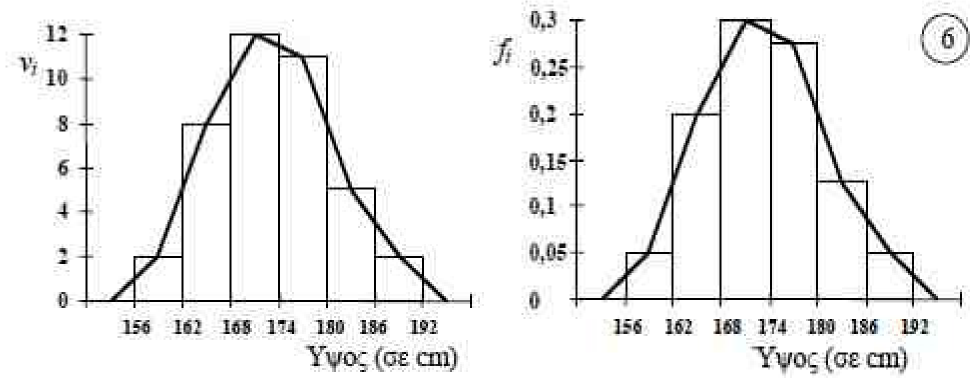
Κλάσεις [,)	**Κεντρικές Τιμές** x_i	v_i	$f_i\%$	N_i	$F_i\%$
156-162	159	2	5,0	2	5,0
162-168	165	8	20,0	10	25,0
168-174	171	12	30,0	22	55,0
174-180	177	11	27,5	33	82,5
180-186	183	5	12,5	38	95,0
186-192	189	2	5,0	40	100,0
	Σύνολο	40	100	-	-

α) Κλάσεις Ίσου Πλάτους

Θεωρώντας το πλάτος c ως μονάδα μέτρησης του χαρακτηριστικού στον οριζόντιο άξονα, το ύψος κάθε ορθογωνίου είναι ίσο προς τη συχνότητα της αντίστοιχης κλάσης, έτσι ώστε να ισχύει πάλι ότι το εμβαδόν των ορθογωνίων είναι ίσο με τις αντίστοιχες συχνότητες. Επομένως, στον κατακόρυφο άξονα σε ένα ιστόγραμμα συχνοτήτων βάζουμε τις συχνότητες. Με ανάλογο τρόπο κατασκευάζεται και το **ιστόγραμμα σχετικών συχνοτήτων**, οπότε στον κάθετο άξονα βάζουμε τις σχετικές συχνότητες.

Αν στα ιστογράμματα συχνοτήτων θεωρήσουμε δύο ακόμη υποθετικές κλάσεις, στην αρχή και στο τέλος, με συχνότητα μηδέν και στη συνέχεια ενώσουμε τα μέσα των άνω βάσεων των ορθογωνίων με ευθύγραμμα τμήματα, σχηματίζεται το λεγόμενο **πολύγωνο συχνοτήτων**. Το εμβαδόν του χωρίου που ορίζεται από το πολύγωνο συχνοτήτων και τον οριζόντιο άξονα είναι ίσο με το άθροισμα των συχνοτήτων, δηλαδή με το μέγεθος του δείγματος ν. Όμοια κατασκευάζεται από το ιστόγραμμα σχετικών συχνοτήτων και το **πολύγωνο σχετικών συχνοτήτων** με εμβαδόν ίσο με 1.

Σχήμα 8

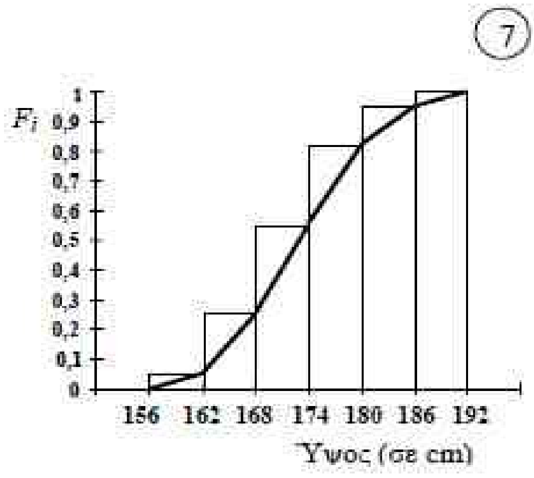


(α) (β)

Ιστόγραμμα και πολύγωνο (α) συχνοτήτων και (β) σχετικών συχνοτήτων για τα δεδομένα του Πίνακα 5.

Με τον ίδιο τρόπο κατασκευάζονται και τα **ιστογράμματα αθροιστικών συχνοτήτων** και **αθροιστικών σχετικών συχνοτήτων**. Αν ενώσουμε σε ένα ιστόγραμμα αθροιστικών συχνοτήτων τα **δεξιά άκρα** (όχι μέσα) των άνω βάσεων των ορθογωνίων με ευθύγραμμα τμήματα βρίσκουμε το **πολύγωνο αθροιστικών συχνοτήτων** της κατανομής. Στο Σχήμα 9 παριστάνεται το ιστόγραμμα και το πολύγωνο αθροιστικών σχετικών συχνοτήτων για το ύψος των μαθητών του Πίνακα 5.

Σχήμα 9



β) Κλάσεις Άνισου Πλάτους

Όπως προαναφέραμε, συνήθως επιλέγουμε κλάσεις ίσου πλάτους. Υπάρχουν όμως κα περιπτώσεις που είναι απαραίτητο να έχουμε κλάσεις διαφορετικού πλάτους όπως, για παράδειγμα, στη κατανάλωση νερού και ηλεκτρικού ρεύματος ή ακόμα και περιπτώσεις όπου οι συχνότητες σε κάποιες κλάσεις να είναι πολύ μικρές οπότε γίνεται συγχώνευση κλάσεων.

Έστω για παράδειγμα, η διάρκεια (σε sec) ν=80 τηλεφωνημάτων που έγιναν τυχαία από ένα κινητό τηλέφωνο, η οποία δίνεται στον παρακάτω πίνακα συχνοτήτων.

Πίνακας 6

Διάρκεια τηλ. σε (sec)	**Συχνότητα v_i**
0-20	20
20-25	20
25-30	24
30-40	16
Σύνολο	ν=80

Το αντίστοιχο ιστόγραμμα συχνοτήτων κατασκευάζεται πάλι, έτσι ώστε το εμβαδόν κάθε ορθογωνίου να ισούται με τη συχνότητα της αντίστοιχης κλάσης. Άρα, αν c_i είναι το πλάτος της κλάσης i με συχνότητα v_i , το ύψος του ορθογωνίου θα είναι $u_i = \frac{v_i}{c_i}, i = 1,2,...,k$. Επομένως, για την κατασκευή του ιστογράμματος συχνοτήτων χρειαζόμαστε τα πλάτη των κλάσεων και τα ύψη των ορθογωνίων.

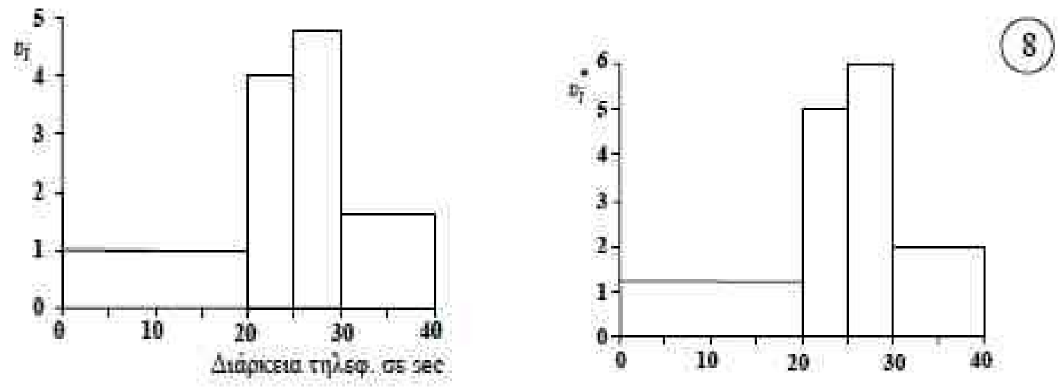
Αυτά δίνονται στον Πίνακα 7.

Πίνακας 7

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

Τότε το ιστόγραμμα συχνοτήτων δίνεται στο Σχήμα 10(α). Παρατηρούμε ότι το άθροισμα των εμβαδών όλων των ορθογωνίων είναι ίσο με το συνολικό μέγεθος δείγματος ν, όπως δηλαδή συμβαίνει και στο ιστόγραμμα με κλάσεις ίσου πλάτους.

Σχήμα 10



(α) (β)

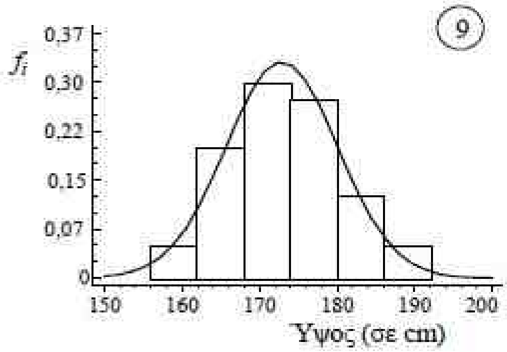
Ιστόγραμμα συχνοτήτων (α) και σχετικών συχνοτήτων (β) της διάρκειας τηλεφωνημάτων.

Με ανάλογο τρόπο κατασκευάζεται και το ιστόγραμμα σχετικών συχνοτήτων, (Σχήμα 10(β)) αρκεί να χρησιμοποιήσουμε ως ύψος των ορθογωνίων το λόγο των σχετικών συχνοτήτων προς το πλάτος των κλάσεων, δηλαδή $u_i = \frac{f_i \%}{c_i}$.

Καμπύλες Συχνοτήτων

Εάν υποθέσουμε ότι ο αριθμός των κλάσεων για μία συνεχή μεταβλητή είναι αρκετά μεγάλος (τείνει στο άπειρο) και ότι το πλάτος των κλάσεων είναι αρκετά μικρό (τείνει στο μηδέν), τότε η πολυγωνική γραμμή συχνοτήτων τείνει να πάρει τη μορφή μιας ομαλής καμπύλης, η οποία ονομάζεται **καμπύλη συχνοτήτων**, όπως δείχνει το Σχήμα 11. Οι καμπύλες συχνοτήτων έχουν μεγάλη εφαρμογή στη Στατιστική, όπου οι ιδιότητες τους μπορούν να χρησιμοποιηθούν για την εξαγωγή χρήσιμων συμπερασμάτων.

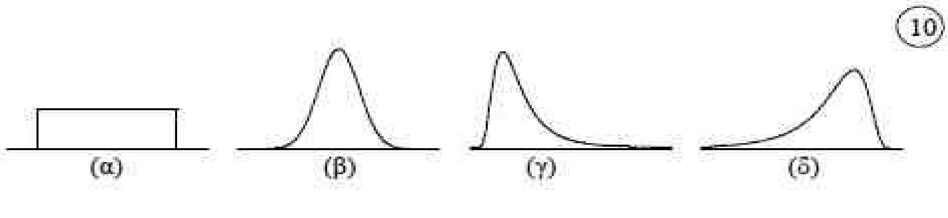
Σχήμα 11



Καμπύλη συχνοτήτων για το ύψος των μαθητών

Μερικές χαρακτηριστικές καμπύλες συχνοτήτων που συναντάμε συχνά στις εφαρμογές δίνονται στο Σχήμα 12. Η καμπύλη (β), με ΄΄κωδωνοειδή΄΄ μορφή λέγεται **κανονική κατανομή** και παίζει σπουδαίο ρόλο στη Στατιστική. Όταν οι παρατηρήσεις κατανέμονται ομοιόμορφα σε ένα διάστημα [α, β], όπως στη κατανομή (α), η κατανομή λέγεται **ομοιόμορφη**. Όταν οι παρατηρήσεις δεν είναι συμμετρικά κατανεμημένες, η κατανομή λέγεται ασύμμετρη με θετική ασυμμετρία όπως στη κατανομή (γ) ή αρνητική ασυμμετρία όπως στη κατανομή (δ).

Σχήμα 12



Μερικές χαρακτηριστικές κατανομές συχνοτήτων.

2.4 Η ΧΡΗΣΙΜΟΠΟΙΗΣΗ ΑΠΟ ΤΙΣ ΕΠΙΧΕΙΡΗΣΕΙΣ ΣΤΑΤΙΣΤΙΚΩΝ ΠΙΝΑΚΩΝ ΚΑΙ ΔΙΑΓΡΑΜΜΑΤΩΝ

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.

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 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.

ΚΕΦΑΛΑΙΟ 3

ΤΑ ΒΑΣΙΚΑ ΜΕΤΡΑ ΘΕΣΗΣ ΓΙΑ ΤΗΝ ΕΠΙΧΕΙΡΗΣΗ

3.1 ΕΝΝΟΙΑ ΚΑΙ ΣΗΜΑΣΙΑ ΤΩΝ ΜΕΤΡΩΝ ΘΕΣΗΣ

Ο υπολογισμός ενός μέτρου θέσης μίας μεταβλητής αποβλέπει στον καθορισμό μίας τιμής η οποία μπορεί να θεωρηθεί ως ένα σημείο συγκέντρωσης, προς το οποίο συνήθως τείνουν οι επί μέρους τιμές των μονάδων του πληθυσμού.

Τα πιο σημαντικά μέτρα θέσης είναι ο αριθμητικός μέσος, η διάμεσος και τα τεταρτημόρια. Άλλα μέτρα θέσης που χρησιμοποιούνται όμως λιγότερο στις εφαρμογές είναι ο γεωμετρικός μέσος, ο αρμονικός μέσος και η επικρατούσα τιμή. Για τον υπολογισμό των διαφόρων μέτρων θα διακρίνουμε τις εξής δύο περιπτώσεις.

α) Την περίπτωση των απλών δεδομένων:

$x_1, x_2, \ldots\ldots, x_\nu$

όπου οι ν το πλήθος μετρήσεις αναφέρονται μεμονωμένα στις ν μονάδες του πληθυσμού και

β) την περίπτωση των ομαδοποιημένων ή ταξινομημένων δεδομένων, όπου τα δεδομένα παρουσιάζονται σε πίνακα συχνοτήτων με το συμβολισμό που ακολουθεί:

Πίνακας 8

ΤΑΞΕΙΣ ΤΙΜΩΝ	ΚΕΝΤΡΙΚΕΣ ΤΙΜΕ	ΣΥΧΝΟΤΗΤΕΣ ΤΑΞΙΚΕΣ	ΣΥΧΝΟΤΗΤΕΣ ΑΘΡΟΙΣΤΙΚΕΣ
$[a_0, a_1)$	x_1	f_1	F_1
$[a_1, a_2)$	x_2	f_2	F_2
…	…	…	…
…	…	…	…
$[a_{i-1}, a_i)$	x_i	f_i	F_i
…	…	…	…
$[a_{k-1}, a_k)$	x_k	f_k	F_k

3.2 ΜΕΣΗ ΤΙΜΗ Η ΑΡΙΘΜΗΤΙΚΟΣ ΜΕΣΟΣ

Έστω X μία μεταβλητή και $x_1, x_2, ..., x_\nu$ οι ν το πλήθος παρατηρήσεις. Θα καλείται αριθμητικός μέσος[1] των παρατηρήσεων αυτών το πηλίκο:

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

Υπολογισμός μέσου ταξινομημένων δεδομένων.

Αν τα δεδομένα είναι ομαδοποιημένα σε ν τάξεις ενός πίνακα συχνοτήτων και η τάξη i έχει κεντρική τιμή x_i και συχνότητα f_i, τότε ο μέσος δίνεται από τον τύπο:

$$\overline{x} = \frac{f_1 x_1 + f_2 x_2 + ... + f_\nu x_\nu}{\nu} = \frac{1}{\nu} \cdot \sum_{i=1}^{\nu} f_i x_i$$

Υπολογισμός σταθμικού μέσου.

Δίνονται οι εξής τιμές της μεταβλητής $x_1, x_2, ..., x_\nu$ και επισυνάπτουμε σε κάθε τιμή x_i, έναν συντελεστή (βάρος) $w_i, i = 1,2,....,\nu$ ο οποίος είναι ο βαθμός σπουδαιότητας της μεταβλητής x_i. Τότε ο σταθμικός μέσος των μεταβλητών $x_1, x_2, ..., x_\nu$ δίνεται από τον τύπο:

$$\overline{x}_w = \frac{x_1 w_1 + x_2 w_2 + \cdots + x_\nu w_\nu}{w_1 + w_2 + \cdots + w_\nu}$$

[1] << 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.

3.2.1 Ιδιότητες της μέσης τιμής

1. Αν σε όλες τις τιμές μιας μεταβλητής X προσθέσουμε (ή αφαιρέσουμε) μια σταθερή ποσότητα, τότε ο $\bar{x}$ αυξάνεται (ή μειώνεται) κατά την ποσότητα αυτή.

2. Αν όλες οι τιμές της μεταβλητής είναι ίσες με μια σταθερά, τότε ο $\bar{x}$ ισούται με τη σταθερά αυτή.

3. Αν όλες οι τιμές της μεταβλητής πολλαπλασιασθούν επί μια σταθερά, τότε ο $\bar{x}$ πολλαπλασιάζεται με αυτή τη σταθερά.

4. Αν από όλες τις τιμές της μεταβλητής αφαιρέσουμε το $\bar{x}$, τότε το άθροισμα των διαφορών $x_i - \bar{x}$, ισούται με μηδέν. Δηλαδή:

$$\sum \left(x_i - \bar{x} \right) = 0$$

5. Αν ένας πληθυσμός χωρισθεί σε κ υποπληθυσμούς, που ο καθένας έχει $N_1, \cdots, N_k$ μονάδες και οι αντίστοιχοι μέσοι είναι $\bar{x}_1, \cdots, \bar{x}_k$, τότε ο μέσος όλου του πληθυσμού είναι:

$$\bar{x} = \frac{N_1 \bar{x}_1 + \cdots + N_k \bar{x}_k}{N_1 + \cdots + N_k}.$$

3.3 Η ΔΙΑΜΕΣΟΣ

Ένα άλλο μέτρο θέσης μίας κατανομής είναι η διάμεσος τιμή. Ως διάμεσος μιας ποσοτικής μεταβλητής X ορίζεται μια τιμή της μεταβλητής τέτοια, ώστε οι μισές παρατηρήσεις να είναι μικρότερες ή ίσες της τιμής αυτής και οι άλλες μισές μεγαλύτερες.

Στην περίπτωση απλών δεδομένων $x_1, x_2, ..., x_\nu$ ο υπολογισμός της διαμέσου γίνεται ως εξής:

Ταξινομούμε τα δεδομένα κατά αυξανόμενο μέγεθος και στην περίπτωση που το πλήθος των δεδομένων είναι περιττό, η μεσαία παρατήρηση είναι η διάμεσος. Στην περίπτωση που το πλήθος των δεδομένων είναι άρτιο, ως διάμεσος ορίζεται το ημιάθροισμα των δύο μέσων παρατηρήσεων.

Στην περίπτωση ομαδοποιημένων δεδομένων και εφόσον η μεταβλητή X είναι ασυνεχής, ο υπολογισμός της διαμέσου γίνεται ως εξής:

Κατ’ αρχήν καταρτίζουμε τις αθροιστικές συχνότητες της X και εντοπίζουμε μεταξύ ποιων αθροιστικών συχνοτήτων βρίσκεται ο αριθμός $\nu/2$ προκειμένου για απόλυτες συχνότητες ή 100/2=50 για σχετικές. Ως διάμεσος τιμή ορίζεται η τιμή της μεταβλητής που αντιστοιχεί στη μεγαλύτερη από τις δύο αθροιστικές συχνότητες μεταξύ των οποίων βρίσκεται ο αριθμός $\nu/2$ ή 50. Εάν ο αριθμός $\nu/2$ ή 50 συμπίπτει με μία από τις αθροιστικές συχνότητες τότε ως διάμεσος ορίζεται η αντίστοιχη τιμή της συχνότητας αυτής.

Στην περίπτωση που η μεταβλητή X είναι συνεχής, ο υπολογισμός της διαμέσου γίνεται με όμοια κατ’ αρχήν διαδικασία. Εντοπίζουμε δηλαδή μεταξύ ποιων αθροιστικών συχνοτήτων βρίσκεται ο αριθμός $\nu/2$ ή 50.

Ο υπολογισμός της διαμέσου τιμής γίνεται από τον τύπο:

$$M = L_i + \frac{(\nu/2 - F_{i-1})}{f_i} \cdot \delta$$

όπου

L_i : κατώτερο όριο της τάξης i της διαμέσου,

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

f_i : συχνότητα της τάξης i της διαμέσου,

δ : πλάτος της τάξης i της διαμέσου.

3.4 ΕΠΙΚΡΑΤΟΥΣΑ ΤΙΜΗ

Επικρατούσα τιμή των παρατηρήσεων $x_1, x_2, ..., x_v$ λέγεται η τιμή της μεταβλητής που έχει τη μεγαλύτερη συχνότητα εμφάνισης, δηλαδή είναι η τιμή που επαναλαμβάνεται τις περισσότερες φορές στον πληθυσμό ή στο δείγμα. Η επικρατούσα τιμή ενός πλήθος δεδομένων είναι δυνατόν να μην υπάρχει ή και όταν υπάρχει δεν είναι απαραίτητο να είναι και μοναδική.

Κάθε κατανομή που έχει μια επικρατούσα τιμή λέγεται μονοκόρυφη. Αν μια κατανομή έχει δύο επικρατούσες τιμές, λέγεται δικόρυφη. Στην περίπτωση αυτή και, εφόσον φυσικά η μεταβλητή είναι συνεχής, παρουσιάζει δύο τοπικά μέγιστα.

Στην περίπτωση κατά την οποία η μεταβλητή είναι συνεχής και η κατανομή σταθερού πλάτους, η επικρατούσα τιμή υπολογίζεται από τον τύπο :

$$T = a_{i-1} + \frac{(f_i - f_{i-1}) \cdot \delta}{2f_i - f_{i-1} - f_{i+1}}$$

όπου

a_{i-1}: το κατώτερο όριο της i τάξης (αυτής με τη μεγαλύτερη συχνότητα),

f_i : η συχνότητα της i τάξης ,

δ : το σταθερό πλάτος της i τάξης ,

f_{i-1}: Η συχνότητα της προηγούμενης i τάξης,

f_{i+1}: Η συχνότητα της επόμενης της i τάξης .

3.5 ΤΕΤΑΡΤΗΜΟΡΙΑ

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 σειρά. Δηλαδή, σε ένα σύνολο διατεταγμένων παρατηρήσεων σε αύξουσα σειρά 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 ως την τιμή της μεταβλητής, κάτω της οποίας βρίσκεται το πολύ το 50% του συνολικού αριθμού των παρατηρήσεων. Δηλαδή, το Q_2 είναι ίσο με τη διάμεσο.

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 $(v+1)/4$, η θέση που κατέχει το δεύτερο τεταρτημόριο Q_2 είναι $(v+1)/2$ και η θέση που κατέχει το τρίτο τεταρτημόριο Q_3 είναι $3(v+1)/4$.

Ανάλογα ορίζονται τα δεκατημόρια ή τα εκατοστημόρια. Θα καλείται, για παράδειγμα, τρίτο δεκατημόριο D_3 μιας μεταβλητής X, μια τιμή αυτής τέτοια ώστε το 30% των παρατηρήσεων να είναι μικρότερες ή ίσες της D_3.

Στην περίπτωση των ομαδοποιημένων δεδομένων, τα δεκατημόρια δίνονται από τον τύπο:

$$D_k = a_{i-1} + \frac{\delta_i}{f_i}\left(\frac{k\nu}{10} - F_{i-1}\right),\ k = 1, 2, \dots, 10$$

Τελείως ανάλογα ορίζονται τα εκατοστημόρια P_k από τον τύπο:

$$P_k = a_{i-1} + \frac{\delta}{f_i}\left(\frac{k\nu}{100} - F_{i-i}\right),\ k = 1, 2, \dots, 100$$

ΚΕΦΑΛΑΙΟ 4

ΤΑ ΒΑΣΙΚΑ ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ ΓΙΑ ΤΗΝ ΕΠΙΧΕΙΡΗΣΗ

4.1 ΓΕΝΙΚΑ

Όταν θέλουμε να έχουμε μια 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. Τα μέτρα διασποράς, ως στατιστικά μέτρα που μας πληροφορούν για το 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.

Τα σπουδαιότερα μέτρα διασποράς είναι η διακύμανση, τυπική απόκλιση, το εύρος, το ενδοτεταρτημοριακό εύρος και ο συντελεστής μεταβλητότητας.

4.2 ΔΙΑΚΥΜΑΝΣΗ

Έστω $x_1, x_2, ..., x_\nu$ οι ν το πλήθος των παρατηρήσεων μιας ποσοτικής μεταβλητής x και $\bar{x}$ ο αριθμητικός μέσος αυτών των παρατηρήσεων. Η διαφορά: $x_i - \bar{x}$ του μέσου από την τιμή x_i καλείται απόκλιση της i παρατήρησης, ενώ το τετράγωνο: $(x_i - \bar{x})^2$ καλείται τετραγωνική απόκλιση της i παρατήρησης.

Θα καλούμε διακύμανση s^2 των παρατηρήσεων (ή της μεταβλητής Χ) τον αριθμητικό μέσο των τετραγωνικών αποκλίσεων:

$$(x_i - \bar{x})^2,\ i = 1,2,....,\nu.$$

Δηλαδή:

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

Αν τα δεδομένα είναι ομαδοποιημένα σε k το πλήθος τάξεις με κεντρικές τιμές: $x_1, x_2, ..., x_k$, κάθε ένα από τα τετράγωνα: $\left(x_i - \overline{x}\right)^2$ εμφανίζεται f_i φορές και επομένως ως διακύμανση (σε ομαδοποιημένα δεδομένα) ορίζεται η ποσότητα:

$$s^2 = \frac{1}{\nu - 1}\sum_{i=1}^{k} f_i (x_i - \overline{x})^2$$

Γίνεται όμως φανερό ότι ο τύπος ορισμού της διακύμανσης οδηγεί συχνά σε πάρα πολλές αριθμητικές πράξεις, γι' αυτό αντί του τύπου ορισμού χρησιμοποιείται ο τύπος:

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

ή

$$s^2 = \frac{1}{\nu - 1}\left(\sum_{i=1}^{\nu} x_i^{\ 2} - \nu \overline{x}^2\right)$$

Σε ομαδοποιημένα δεδομένα ο τύπος παίρνει τη μορφή:

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

ή

$$s^2 = \frac{1}{\nu - 1}\left(\sum_{i=1}^{k} f_i x_i^{\ 2} - \nu \overline{x}^2\right)$$

4.2.1 Ιδιότητες της διακύμανσης

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

- Αν σε κάθε τιμή της μεταβλητής X πολλαπλασιασθεί (ή διαιρεθεί) με ένα σταθερό αριθμό (α), τότε η διακύμανση της μεταβλητής

$$Y = aX \left(Y = \frac{X}{a}, a \neq 0 \right)$$

πολλαπλασιάζεται (ή διαιρείται) με το τετράγωνο του αριθμού αυτού, δηλαδή:

$$s_Y^2 = a^2 s_X^2 \left(s_Y^2 = \frac{s_X^2}{a^2} \right).$$

- Αν οι τιμές της μεταβλητής X είναι ίσες μεταξύ τους, τότε $s^2 = 0$

4.3 ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ

Η τυπική απόκλιση έχει το σημαντικό πλεονέκτημα να μετράται στις ίδιες μονάδες που μετράται και η μεταβλητή X γεγονός που φυσικά δεν ισχύει για τη διακύμανση. Έτσι αν η μεταβλητή X μετράται π.χ. σε εκατοστά, η διακύμανση μετράται σε τετραγωνικά εκατοστά, ενώ η τυπική απόκλιση ως τετραγωνική ρίζα της διακύμανσης μετράται επίσης σε εκατοστά, δηλαδή με την ίδια μονάδα που μετράται και η X. Ως τυπική απόκλιση μιας μεταβλητής X ορίζεται η τετραγωνική ρίζα της διακύμανσης.

Η τυπική απόκλιση συμβολίζεται με s και ορίζεται ως η τετραγωνική ρίζα της διακύμανσης, δηλαδή:

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

4.4 ΕΥΡΟΣ ΔΕΔΟΜΕΝΩΝ

Εύρος δεδομένων (range) είναι ένα μέτρο διακύμανσης των δεδομένων το οποίο δίνεται από τον τύπο:

$$R = X_{MAX} - X_{MIN}$$

όπου

X_{MAX} : μέγιστη παρατήρηση

X_{MIN} : ελάχιστη παρατήρηση

Το πλεονέκτημα του Εύρους είναι η ευκολία υπολογισμού του, το δε μειονέκτημα είναι ότι δεν είναι συνάρτηση όλων των δεδομένων, όπως είναι τα άλλα μέτρα διασποράς, με συνέπεια να χάνονται πληροφορίες που βασίζονται στις τιμές των δεδομένων.

4.5 ΕΝΔΟΤΕΤΑΡΤΗΜΟΡΙΑΚΟ ΕΥΡΟΣ

Το **ενδοτεταρτημοριακό εύρος** (interquartile range) είναι η διαφορά του πρώτου τεταρτημορίου Q_1 από το τρίτο Q_3, δηλαδή:

$$Q = Q_3 - Q_1$$

Στο μεταξύ τους διάστημα περιλαμβάνεται το 50% των παρατηρήσεων. Επομένως όσο μικρότερο είναι αυτό το διάστημα, τόσο μεγαλύτερη θα είναι η συγκέντρωση των τιμών και άρα μικρότερη η διασπορά των τιμών της μεταβλητής.

4.6 ΣΥΝΤΕΛΕΣΤΗΣ ΜΕΤΑΒΛΗΤΟΤΗΤΑΣ

Ένα άλλο μέτρο διασποράς σχετικό με το μέσο που χρησιμοποιούμε και εκφράζεται επί τοις εκατό είναι ο συντελεστής μεταβλητότητας[1], το οποίο δίνεται από τον τύπο:

$$CV = \left(\frac{s}{\bar{x}} \right) \cdot 100\%$$

Από τον τύπο ορισμού του είναι φανερό ότι είναι καθαρός αριθμός δηλαδή απαλλαγμένος των μονάδων μέτρησης και συνεπώς η σύγκριση των τυπικών αποκλίσεων δύο πληθυσμών είναι πάντα δυνατή. Ο συντελεστής μεταβλητότητας είναι ένα σχετικό μέτρο διασποράς (σε σχέση προς το μέσο του πληθυσμού), όπως ακριβώς ένα σφάλμα που προκύπτει από κάποια μέτρηση είναι ´´μικρό´´ ή ´´μεγάλο´´ σε σχέση πάντα προς το πραγματικό μέτρο. Ένα δείγμα τιμών μιας μεταβλητής είναι ομοιογενές όταν ο *CV* είναι μικρότερος ή ίσος από 10%. Το μοναδικό μειονέκτημα του συντελεστή μεταβλητότητας είναι ότι παίρνει πολύ μεγάλες τιμές, όταν ο μέσος του πληθυσμού είναι κοντά στο μηδέν και φυσικά απειρίζεται στην περίπτωση μηδενικού μέσου. Γι' αυτό στις περιπτώσεις αυτές η χρήση του πρέπει να αποκλείεται.

[1] Ο συντελεστής μεταβλητότητας προτάθηκε από τον Karl Pearson.

ΚΕΦΑΛΑΙΟ 5

ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

5.1 ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

Συχνά συναντούμε προβλήματα στα οποία επιθυμούμε να περιγράψουμε τη σχέση μεταξύ δύο ή περισσότερων μεταβλητών. Παραδείγματα δύο μεταβλητών που έχουν μια σχέση μεταξύ τους είναι πολλά. Μερικά από αυτά είναι:

α) Το είδος ενός γεωργικού λιπάσματος το οποίο ένας γεωργός βάζει στο χωράφι του και το ύψος του σιταριού που αναπτύσσεται στο χωράφι αυτό.

β) Το ύψος της μετοχής μίας εταιρείας και η οικονομική βάση της εταιρείας αυτής.

γ) Οι διαφημιστικές δαπάνες μίας αλυσίδας καταστημάτων ειδών διατροφής και ο όγκος πωλήσεων των καταστημάτων σε ένα ορισμένο χρόνο.

Έτσι λοιπόν είναι ενδιαφέρον να εξεταστούν οι επιδράσεις που κάποιες μεταβλητές ασκούν σε κάποιες άλλες μεταβλητές. Η ύπαρξη μιας συναρτησιακής σχέσης (εξίσωσης) μεταξύ των μεταβλητών μπορεί να είναι εξαιρετικά πολύτιμη για την πρόβλεψη των τιμών μιας μεταβλητής από τις γνώσεις που διαθέτουμε για τις άλλες μεταβλητές, όταν ισχύουν κάποιες συγκεκριμένες συνθήκες.

Ο κλάδος της Στατιστικής που εξετάζει τη σχέση μεταξύ δύο ή περισσότερων μεταβλητών με απώτερο σκοπό τη πρόβλεψη μιας απ' αυτές μέσω των άλλων χαρακτηρίζεται με την ονομασία **ανάλυση παλινδρόμησης** (regression analysis)[1].

[1] Ιστορικά, ο όρος ´´regression´´ χρησιμοποιήθηκε για πρώτη φορά από τον Άγγλο ανθρωπολόγο Galton (1822-1911) το 1885. Με τη μελέτη του ύψους των παιδιών σε σχέση με το ύψος των γονέων τους διαπιστώθηκε ότι παιδιά ψηλών γονέων τείνουν, κατά μέσο όρο, να είναι κοντύτερα των γονέων τους, ενώ παιδιά κοντών γονέων τείνουν, κατά μέσο όρο, να γίνονται ψηλότερα των γονιών τους.

5.2 ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

Η απλούστερη περίπτωση παλινδρόμησης είναι η απλή γραμμική παλινδρόμηση, κατά την οποία υπάρχει μόνο μία ανεξάρτητη μεταβλητή X και η εξαρτημένη μεταβλητή Y, η οποία μπορεί να προσεγγιστεί ικανοποιητικά από μία γραμμική συνάρτηση του X. Η περίπτωση αυτή εμφανίζεται τόσο σε πειραματικές όσο και σε μη πειραματικές μελέτες.

Σε πειραματικές έρευνες ανεξάρτητη μεταβλητή X είναι εκείνη την οποία μπορούμε να ελέγξουμε, δηλαδή, να καθορίσουμε τις τιμές της (για παράδειγμα, το ύψος της διαφημιστικής δαπάνης ενός προϊόντος, ο αριθμός των λειτουργούντων ταμείων σε ένα υποκατάστημα τραπέζης, η ποσότητα λιπάσματος που χρησιμοποιείται σε μια καλλιέργεια, η θερμοκρασία επεξεργασίας ενός υλικού). Εξαρτημένη μεταβλητή Y είναι εκείνη στην οποία αντανακλάται το αποτέλεσμα των μεταβολών στις ανεξάρτητες μεταβλητές (για παράδειγμα, η ζήτηση ενός προϊόντος, ο χρόνος αναμονής των πελατών ενός υποκαταστήματος τραπέζης, η απόδοση μιας καλλιέργειας, η αντοχή ενός υλικού).

Σε μη πειραματικές έρευνες (δειγματοληψίες) η διάκριση μεταξύ ανεξάρτητων και εξαρτημένων μεταβλητών δεν είναι πάντοτε σαφής γιατί καμία μεταβλητή δεν είναι ελεγχόμενη αλλά όλες είναι τυχαίες (για παράδειγμα, το ύψος και το βάρος των φοιτητών, οι ώρες μελέτης των φοιτητών ενός πανεπιστημιακού τμήματος και η απόδοση τους σε ένα τεστ, οι εβδομάδες εμπειρίας ενός εργάτη σε μια επιχείρηση και ο αριθμός των ελαττωματικών προϊόντων που παράγει, ο αριθμός των πωλήσεων μουσικών cd σε μια περιοχή και ο αριθμός των νέων στην ίδια περιοχή).

5.3 ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ

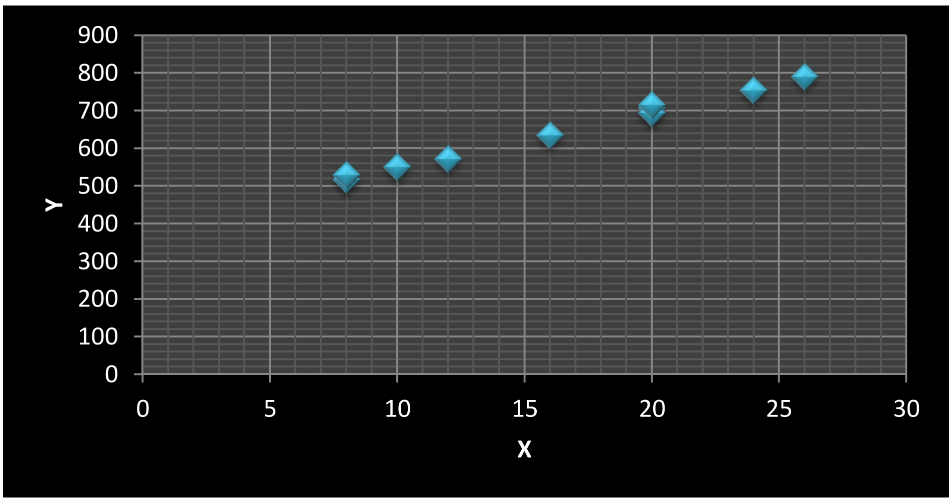
Ο παρακάτω πίνακας δίνει την ποσότητα του λιπάσματος (X) που χρησιμοποιείται σε ένα αγροτεμάχιο και την παραγωγή του (Y) σε 10 όμοια αγροτεμάχια (ίδιου εμβαδού, σε περιοχές που επικρατούν παρόμοιες συνθήκες κτλ), έτσι ώστε οι όποιες διαφοροποιήσεις παρατηρούνται στην παραγωγή των αγρών να οφείλονται κατά κύριο λόγο στις διαφορετικές ποσότητες λιπάσματος που χρησιμοποιήθηκαν.

Πίνακας 9

i	Ποσότητα λιπάσματος (σε εκατοντάδες κιλά) x_i	Παραγωγή (σε χιλιάδες κιλά) y_i
1	20	706
2	10	550
3	26	790
4	8	517
5	20	694
6	16	634
7	20	715
8	12	571
9	8	529
10	24	754

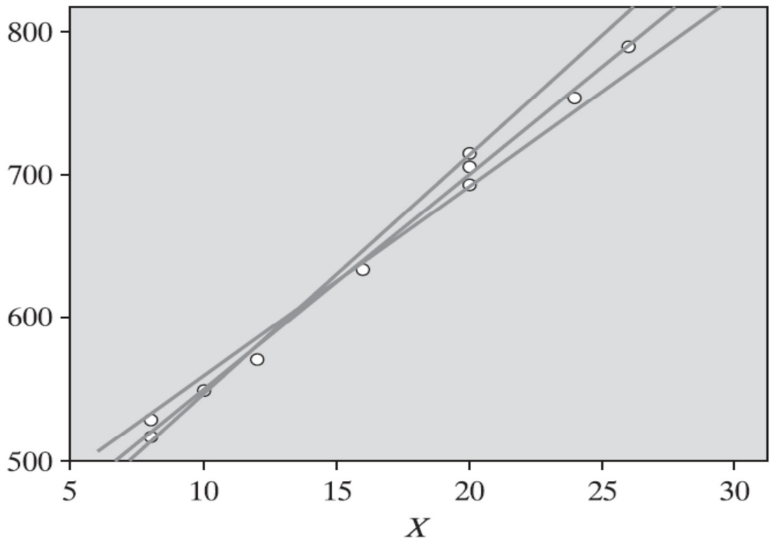
Στο παράδειγμα αυτό έχουμε την περίπτωση όπου για κάθε αγροτεμάχιο γίνονται δύο μετρήσεις. Δηλαδή το δείγμα αποτελείται από τα ζεύγη τιμών των συνεχών μεταβλητών x_i (ποσότητα λιπάσματος) και y_i (παραγωγή). Αν παραστήσουμε τα ζεύγη (x_i, y_i) των παρατηρήσεων σε ένα σύστημα ορθογώνιων αξόνων, παρατηρούμε ότι προκύπτει μία ΄΄διασπορά΄΄ των σημείων που αντιστοιχούν στα αγροτεμάχια που εξετάζουμε. Η παράσταση αυτή των σημείων καλείται διάγραμμα διασποράς.

Διάγραμμα 1



Η προσεκτική παρατήρηση ενός διαγράμματος διασποράς μπορεί να μας δώσει σημαντικές πληροφορίες για τη σχέση εξάρτησης που υπάρχει μεταξύ των μεταβλητών τις οποίες εξετάζουμε. Παρακάτω βλέπουμε πιθανές ευθείες προσαρμοσμένες με το μάτι για τα δεδομένα.

Διάγραμμα 2



.

5.3.1 ΕΥΘΕΙΑ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Από το διάγραμμα διασποράς του προηγούμενου παραδείγματος φαίνεται καθαρά ότι υπάρχει μία σχέση ανάμεσα στη ποσότητα λιπάσματος (x_i) και της παραγωγής (y_i) των 10 αγροτεμάχιων. Τα σημεία (x_i, y_i) είναι συγκεντρωμένα περίπου γύρω από μια ευθεία, δηλαδή η σχέση μεταξύ των *X* και *Y* είναι κατά προσέγγιση γραμμική. Θεωρούμε ως ανεξάρτητη μεταβλητή το (x_i) και ως εξαρτημένη μεταβλητή το(y_i), οπότε η ευθεία που θα προσαρμόζεται καλύτερα στα σημεία αυτά καλείται ευθεία παλινδρόμησης της *Y* πάνω στη *X*.

Όπως γνωρίζουμε, η εξίσωση μιας ευθείας δίνεται από την σχέση:

$$y = a + bx \quad (1)$$

όπου a και b είναι παράμετροι τις οποίες θέλουμε να υπολογίσουμε ή όπως λέμε, να ΄΄εκτιμήσουμε΄΄, έτσι ώστε η ευθεία που θα προκύψει να μας δίνει όσο το δυνατόν την καλύτερη περιγραφή της σχέσης (εξάρτησης) που υπάρχει μεταξύ των μεταβλητών *X* και *Y*. Η παράμετρος a μας δίνει το σημείο $(0, a)$ όπου η ευθεία τέμνει τον κάθετο άξονα και η παράμετρος b παριστάνει το συντελεστή διεύθυνσης της ευθείας.

Ο πιο εύκολος τρόπος χάραξης της ευθείας είναι αυτός που γίνεται ΄΄με το μάτι΄΄. Τέτοιες ευθείες έχουμε φέρει και στο παραπάνω σχήμα διασποράς.

- Επιλέγουμε δύο σημεία, έστω A_1(12,571) και A_2(20,715) πάνω στην ευθεία που φέραμε ΄΄με το μάτι΄΄.

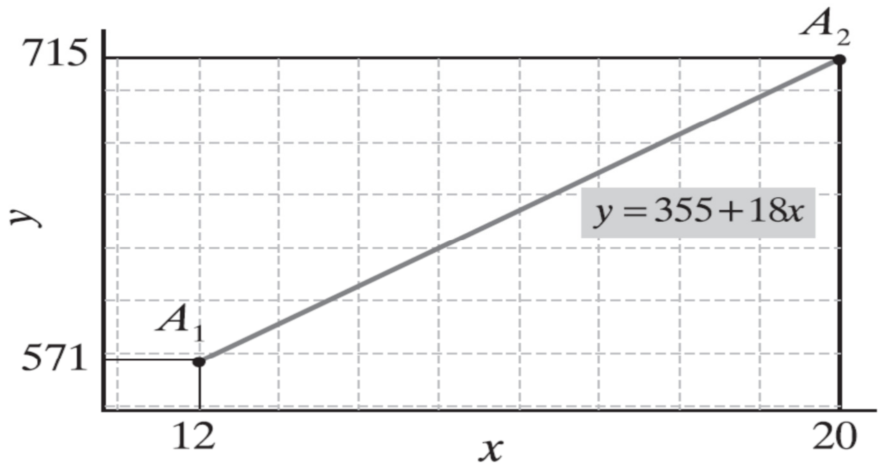
- Αντικαθιστούμε τις συντεταγμένες των σημείων αυτών στην (1), οπότε προκύπτει το σύστημα:

$$\begin{cases} 571 = a + 12b \\ 715 = \alpha + 20b \end{cases}$$

- Επιλύοντας το σύστημα αυτό βρίσκουμε $a = 355$ και $b = 18$, οπότε η εξίσωση της ευθείας (1) γίνεται:
  $y = 355 + 18x$.

Επομένως, η ευθεία που κατά τη γνώμη μας προσαρμόζεται καλύτερα στα σημεία του διαγράμματος διασποράς διέρχεται από το σημείο (0,355) και έχει συντελεστή διεύθυνσης 18.

Διάγραμμα 3



5.4 ΜΕΘΟΔΟΣ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ

Είδαμε ότι η πιο απλή διαδικασία προσαρμογής μιας ευθείας γραμμής σε ένα διάγραμμα διασποράς είναι με το μάτι. Αυτή όμως έχει πολλά μειονεκτήματα παρά την απλότητα της. Το κυριότερο είναι η έλλειψη αντικειμενικότητας, αφού διάφορα άτομα μπορούν να χαράξουν διαφορετικές μεταξύ τους ευθείες. Ακόμα και το ίδιο άτομο μπορεί να χαράζει διαφορετικές ευθείες κάθε φορά. Χρειαζόμαστε λοιπόν μια ακριβέστερη μέθοδο για την προσαρμογή μιας ευθείας γραμμής σε τέτοιου είδους δεδομένα.

Μια μέθοδος που χρησιμοποιείται για την εκτίμηση των παραμέτρων α και b, άρα και για την εύρεση της εξίσωσης της καλύτερης ευθείας που προσαρμόζεται στα δεδομένα, είναι η ΄΄ μέθοδος ελαχίστων τετραγώνων ΄΄.

Η πρώτη αναφορά με ολοκληρωμένη ανάπτυξη της μεθόδου των ελαχίστων τετραγώνων εμφανίζεται το 1805 σε μια εργασία του Γάλλου μαθηματικού Legendre, (1752-1833) και αμέσως μετά από το Γερμανό μαθηματικό Gauss, (1777-1855) στην αστρονομική του πραγματεία ΄΄ Theoria Motus ΄΄ για τον προσδιορισμό της τροχιάς του μικρού πλανήτη Δήμητρα. Μάλιστα εδώ ο Gauss αναφέρει ότι χρησιμοποίησε την αρχή των ελαχίστων τετραγώνων πριν από το 1794(σε ηλικία μόλις 17 ετών), έτσι ώστε να προηγείται του Legendre ως προς την ανακάλυψη αυτής της μεθόδου.

Ας θεωρήσουμε την ευθεία $y = 355 + 18x$ που κατά τη γνώμη μας προσαρμόζεται καλύτερα στα σημεία του διαγράμματος διασποράς του Πίνακα 1. Έτσι για παράδειγμα, για το 6 αγροτεμάχιο με ποσότητα λιπάσματος 16 εκατοντάδες κιλά έχουμε βρει από τον πίνακα παραγωγή 634 χιλιάδες κιλά, ενώ σύμφωνα με την ευθεία που φέραμε η παραγωγή του αναμένεται να είναι $355 + 18 \cdot 16 = 643$ χιλιάδες κιλά, έχουμε δηλαδή ένα σφάλμα $\varepsilon = 643 - 634 = 9$, δηλαδή 9 χιλιάδες κιλά λιγότερο από το αναμενόμενο. Ανάλογα σφάλματα εμφανίζονται και για τα υπόλοιπα αγροτεμάχια. Θα θέλαμε λοιπόν να βρούμε με κάποια μέθοδο εκείνη την ευθεία $y = a + bx$ έτσι ώστε τα σφάλματα που προκύπτουν να είναι όσο το δυνατόν μικρότερα.

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

δηλαδή το

$$\sum_{i=1}^{\nu} \varepsilon_i^2 = \sum_{i=1}^{\nu} \left(y_i - a - bx_i\right)^2 \quad (2)$$

να γίνεται ελάχιστο.

Οι τιμές των παραμέτρων a και b, που ελαχιστοποιούν την (2), καλούνται εκτιμήτριες ελαχίστων τετραγώνων, συμβολίζονται με $\hat{a}$ (a καπέλο) και $\hat{b}$ (b καπέλο), αντιστοίχως και αποδεικνύεται ότι δίνονται από τις σχέσεις:

$$\hat{b} = \frac{\nu \sum_{i=1}^{\nu} x_i y_i - \left(\sum_{i=1}^{\nu} x_i\right)\left(\sum_{i=1}^{\nu} y_i\right)}{\nu \sum_{i=1}^{\nu} {x_i}^2 - \left(\sum_{i=1}^{\nu} x_i\right)^2} \quad (3)$$

$$\hat{a} \equiv \overline{y} - \hat{b}\overline{x}$$

όπου $\overline{x} = \frac{1}{\nu}\sum_{i=1}^{\nu} x_i$ και $\overline{y} = \frac{1}{\nu}\sum_{i=1}^{\nu} y_i$.

Η ευθεία

$$\hat{y} = \hat{a} + \hat{b}x \quad (4)$$

καλείται ευθεία ελαχίστων τετραγώνων ή ευθεία παλινδρόμησης της *Y* (πάνω) στη *X*. Αντικαθιστώντας το $\hat{a} = \overline{y} - \hat{b}\overline{x}$ στη σχέση (4) βρίσκουμε την

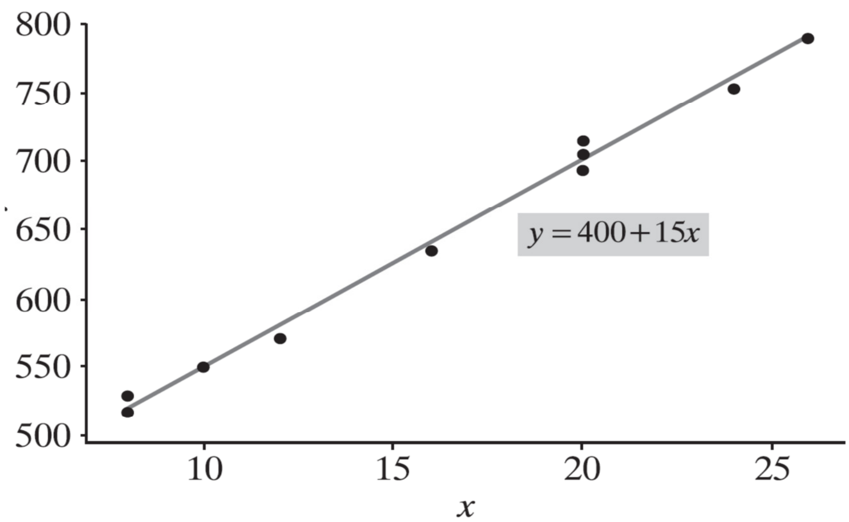
$$\hat{y} - \overline{y} = \hat{b}\left(x - \overline{x}\right),$$

η οποία φανερώνει ότι η ευθεία ελαχίστων τετραγώνων (3) διέρχεται από το σημείο με συντεταγμένες ($\overline{x}, \overline{y}$) και έχει συντελεστή διεύθυνσης το $\hat{b}$.

Αντικαθιστώντας τις τιμές x_i και y_i από τον Πίνακα 1 στις σχέσεις (3) βρίσκουμε $\hat{a} = 400$ και $\hat{b} = 15$. Οπότε η ευθεία ελαχίστων τετραγώνων που προσαρμόζεται καλύτερα στα δεδομένα είναι από τη σχέση (4), η

$$\hat{y} = 400 + 15x$$

Διάγραμμα 4



Παρατηρούμε ότι υπάρχει σημαντική διαφορά από την ευθεία $y = 355 + 18x$ που προσαρμόσαμε ΄΄με το μάτι΄΄ στο παραπάνω σχήμα

5.4.1 ΕΡΜΗΝΕΙΑ ΤΩΝ ΕΚΤΙΜΗΤΡΙΩΝ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ

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

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

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

Δηλαδή

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

Συνεπώς ο συντελεστής διεύθυνσης $\hat{b}$ της ευθείας $\hat{y} = \hat{a} + \hat{b}x$ παριστά τη μεταβολή της εξαρτημένης μεταβλητής Y όταν το X μεταβληθεί κατά μία μονάδα.

Έτσι, όταν το X αυξηθεί κατά μία μονάδα τότε το $\hat{y}$ αυξάνεται κατά $\hat{b}$ μονάδες όταν $\hat{b} > 0$ ή ελαττώνεται κατά $\hat{b}$ μονάδες όταν $\hat{b} < 0$.

5.5 ΣΥΝΤΕΛΕΣΤΗΣ ΓΡΑΜΜΙΚΗΣ ΣΥΣΧΕΤΙΣΗΣ

Ο συντελεστής γραμμικής συσχέτισης είναι ένα μέτρο που μας δίνει το μέγεθος της γραμμικής σχέσης ή το βαθμό συγκέντρωσης των σημείων του διαγράμματος διασποράς γύρω από την ευθεία παλινδρόμησης.

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

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

Ο συντελεστής συσχέτισης είναι καθαρός αριθμός, δηλαδή δεν εκφράζεται σε συγκεκριμένες μονάδες μέτρησης, επομένως είναι ανεξάρτητος των χρησιμοποιούμενων μονάδων μέτρησης των μεταβλητών X και Y. Επί πλέον ισχύει πάντοτε ότι:

$$-1 \le r \le 1$$

- $0 < r < 1$, τότε οι X και Y είναι θετικά γραμμικά συσχετισμένες.
- $-1 < r < 0$, τότε οι X και Y είναι αρνητικά γραμμικά συσχετισμένες.
- $r = 1$, τότε έχουμε τέλεια θετική γραμμική συσχέτιση και όλα τα σημεία βρίσκονται πάνω σε μια ευθεία με θετική κλίση, δηλαδή $\hat{y} = \hat{a} + \hat{b}x$, $\hat{b} > 0$.
- $r = -1$, τότε έχουμε τέλεια αρνητική γραμμική συσχέτιση και όλα τα σημεία βρίσκονται πάνω σε μια ευθεία με αρνητική κλίση, δηλαδή $\hat{y} = \hat{a} + \hat{b}x$, $\hat{b} < 0$.
- $r = 0$, τότε δεν υπάρχει γραμμική συσχέτιση μεταξύ των μεταβλητών. Οι μεταβλητές X και Y είναι γραμμικά ασυσχέτιστες.

Αποδεικνύεται ότι ο συντελεστής γραμμικής συσχέτισης r[1] δίνεται ισοδύναμα και από τον παρακάτω τύπο, η χρήση του οποίου διευκολύνει συχνά τους υπολογισμούς κυρίως στην περίπτωση που οι $\bar{x}, \bar{y}$ δεν είναι ακέραιοι:

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

[1] Ο συντελεστής γραμμικής συσχέτισης ονομάζεται και συντελεστής Pearson.

5.6 ΣΥΝΤΕΛΕΣΤΗΣ ΠΡΟΣΔΙΟΡΙΣΜΟΥ

Ο συντελεστής προσδιορισμού ορίζεται ως το τετράγωνο r^2 του συντελεστή γραμμικής συσχέτισης r. Αποδεικνύεται ότι ο συντελεστής προσδιορισμού δίνεται από τον τύπο:

$$r^2 = 1 - \frac{\sum_{i=1}^{v}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{v}\left(y_i - \overline{y}\right)^2}$$

και εκφράζει το ποσοστό της συνολικής μεταβλητότητας της εξαρτημένης μεταβλητής το οποίο εξηγείται από την ανεξάρτητη μεταβλητή.

Αν το r^2 είναι κοντά στο 1, τότε η γραμμική παλινδρόμηση ερμηνεύει ικανοποιητικά τη μεταβολή του *Y*.

Αν το r^2 είναι κοντά στο 0, τότε η εξίσωση της παλινδρόμησης δεν ερμηνεύει τη μεταβολή στο *Y*.

5.7 Η ΣΗΜΑΣΙΑ ΤΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ ΚΑΙ ΤΗΣ ΣΥΣΧΕΤΙΣΗΣ ΣΤΗ ΣΥΓΧΡΟΝΗ ΕΠΙΧΕΙΡΗΣΗ.

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 αποφάσεων. Χρησιμοποιούνται για τη ποσοτική διερεύνηση των σχέσεων, οι οποίες υπάρχουν μεταξύ των διαφόρων οικονομικών μεγεθών και βοηθούν σημαντικά στη χάραξη της οικονομικής πολιτικής που ασκείται από τις 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 $\hat{y} = \hat{a} + \hat{b}x$ 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 κλπ. Καταλαβαίνουμε λοιπόν ότι αποτελούν ένα πολύ σημαντικό όργανο εργασίας και έρευνας για το σημερινό οικονομολόγο, όπως και για άλλους ερευνητές και επιστήμονες.

ΚΕΦΑΛΑΙΟ 6
ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΕΠΙΧΕΙΡΗΣΕΙΣ

6.1 Ο ΡΟΛΟΣ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ ΣΤΙΣ ΕΠΙΧΕΙΡΗΣΕΙΣ

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, του κεφαλαιουχικού εξοπλισμού, τις αγορές και τις πωλήσεις, τα αποτελέσματα των διαφημιστικών δαπανών κ.τ.λ., 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 των επιχειρήσεων. Με τη βοήθεια της Στατιστικής, οι επιχειρήσεις μπορούν να κάνουν προβλέψεις για την εξέλιξη των πωλήσεων τους, να εξασφαλίσουν τον έλεγχο της ποιότητας των παραγόμενων προϊόντων, να ελέγχουν το κόστος παραγωγής κ.τ.λ.

Η ανάπτυξη ολοκληρωμένων επιστημονικών κλάδων, όπως είναι η Οικονομετρία, η Επιχειρησιακή Έρευνα, η Διαφήμιση κ.α., 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.

6.2 ΠΑΡΑΔΕΙΓΜΑΤΑ ΕΦΑΡΜΟΓΩΝ ΣΤΙΣ ΕΠΙΧΕΙΡΗΣΕΙΣ

<u>ΕΦΑΡΜΟΓΗ 1</u>

Οι μηνιαίες αποδοχές 100 υπαλλήλων μιας επιχείρησης κυμαίνονται μεταξύ 2000 και 3000 ευρώ. Ακόμη, γνωρίζουμε ότι:

10 υπάλληλοι αμείβονται με 2200 ευρώ και κάτω,

30 υπάλληλοι αμείβονται με 2400 ευρώ και κάτω,

40 υπάλληλοι έχουν μισθούς μεγαλύτερους από 2600 ευρώ και

10 υπάλληλοι έχουν μισθούς μεγαλύτερους από 2800 ευρώ.

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

ΛΥΣΗ

Έχοντας υπόψη μας τα δεδομένα της άσκησης σχηματίζουμε τον παρακάτω πίνακα:

Πίνακας 10

Μηνιαίες Αποδοχές	**Αριθμός Υπαλλήλων** v_i	**Κεντρικοί Όροι Κλάσεων** x_i	$v_i x_i$
2000-2200	10	2100	21000
2200-2400	20	2300	46000
2400-2600	30	2500	75000
2600-2800	30	2700	81000
2800-3000	10	2900	29000
ΣΥΝΟΛΟ	100		252000

α)Υπολογισμός μέσης τιμής:

Στον τύπο εύρεσης της μέσης τιμής θέτουμε:

$\sum_{i=1}^{5} v_i = 100$ και $\sum_{i=1}^{5} v_i x_i = 252000$ και βρίσκουμε:

$$\bar{x} = \frac{\sum_{i=1}^{5} v_i x_i}{\sum_{i=1}^{5} v_i} = \frac{252000}{100} = 2520 .$$

Δηλαδή, ο μέσος μηνιαίος μισθός των υπαλλήλων της επιχείρησης 2520 ευρώ.

β)Υπολογισμός της διαμέσου

Σχηματίζουμε τον παρακάτω πίνακα:

Πίνακας 11

Τάξεις μηνιαίων αποδοχών	**Αριθμός Υπαλλήλων** v_i	**Αθροιστικές συχνότητες** N_i	
2000-2200	10	10	
2200-2400	20	30	ν/4=25
2400-2600	30	60	ν/2=50
2600-2800	30	90	3ν/4=75
2800-3000	10	100	

Ο αριθμός ν/2=100/2=50, όπως βλέπουμε, βρίσκεται μεταξύ των αθροιστικών συχνοτήτων 30 και 60. Άρα, έχουμε:

$$M = L_i + \frac{(v/2 - N_{i-1})}{v_i}\cdot\delta = 2400 + \frac{(50-30)}{30}\cdot 200 = 2400 + \frac{20}{30}\cdot 200 = 2533{,}33$$

Δηλαδή, ο διάμεσος μηνιαίος μισθός είναι 2533,33 ευρώ περίπου.

γ)Υπολογισμός του πρώτου τεταρτημορίου Q_1

Ο αριθμός ν/4=25 όπως βλέπουμε, βρίσκεται μεταξύ των αθροιστικών συχνοτήτων 10 και 30, άρα έχουμε:

$$Q_1 = L_i + \frac{(v/4 + N_{i-1})}{v_i}\cdot\delta = 2200 + \frac{(25-10)}{20}200 = 2350$$

Δηλαδή, το 25% των υπαλλήλων έχει το πολύ μηνιαίες αποδοχές 2350 ευρώ και το υπόλοιπο 75% περισσότερες από 2350 ευρώ.

Υπολογισμός του τρίτου τεταρτημορίου Q_3

Εργαζόμενοι κατά τρόπο παρόμοιο όπως στην εύρεση του Q_1 έχουμε:

$$Q_1 = L_i + \frac{(3v/4 + N_{i-1})}{v_i} \cdot \delta = 2600 + \frac{(75-60)}{30} 200 = 2600 + 100 = 2700$$

Δηλαδή, το 75% των υπαλλήλων έχει το πολύ μηνιαίες αποδοχές 2700 ευρώ και το υπόλοιπο 25% περισσότερες από 2700 ευρώ.

ΕΦΑΡΜΟΓΗ 2

Οι ετήσιες πωλήσεις (*X*) των 15 καταστημάτων – παραρτημάτων (αλυσίδα καταστημάτων), που έχει μεγάλη εμπορική επιχείρηση, δίνονται σε εκατομμύρια ευρώ από τον παρακάτω πίνακα:

Πίνακας 12

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

Ζητείται να υπολογισθούν:

α) Το μέσο ετήσιο ύψος πωλήσεων των 15 καταστημάτων,
β) το εύρος μεταβολής ,
γ) η διακύμανση και η τυπική απόκλιση και
δ) ο συντελεστής μεταβλητότητας.
Μια άλλη επιχείρηση του ίδιου εμπορικού κλάδου παρουσίασε μέσο ετήσιο ποσό πωλήσεων 70 εκατομμύρια ευρώ και τυπική απόκλιση 17 εκατομμύρια. Να συγκρίνετε και να σχολιάσετε τη δραστηριότητα των δύο επιχειρήσεων.

ΛΥΣΗ

α) Υπολογισμός της μέσης τιμής

Κατασκευάζουμε τον πίνακα:

Πίνακας 13

x_i	v_i	$v_i x_i$
50	1	50
60	2	120
70	4	280
80	4	320
90	2	180
100	1	100
110	1	110
Σύνολο	15	1160

$$\overline{x} = \frac{\sum_{i=1}^{7} v_i x_i}{\sum_{i=1}^{7} v_i} = \frac{1160}{15} = 73{,}333$$

Συνεπώς, η μέση τιμή των πωλήσεων των 15 καταστημάτων είναι 77.330.000 ευρώ περίπου.

β) Το εύρος μεταβολής, ως γνωστό, είναι η διαφορά μεταξύ των δύο ακραίων τιμών της *X*, δηλαδή έχουμε:

Εύρος Μεταβολής: 110-50=60

Δηλαδή το εύρος μεταβολής είναι ίσο με 60.000.000 ευρώ.

γ) Υπολογισμός διακύμανσης και τυπικής απόκλισης.

Συνθέτουμε τον παρακάτω πίνακα αριθμητικών υπολογισμών:

Πίνακας 14

x_i	v_i	$x_i v_i$	$x_i^{\ 2}$	$v_i x_i^{\ 2}$
50	1	50	2.500	2.500
60	2	120	3.600	7.200
70	4	280	4.900	19.600
80	4	320	6.400	25.600
90	2	180	8.100	16.200
100	1	100	10.000	10.000
110	1	110	12.100	12.100
Σύνολο	15	1.160		93.200

Η διακύμανση είναι:

$$s^2 = \frac{1}{v-1}\left[\sum_{i=1}^{7} v_i x_i^{\ 2} - \frac{1}{v}\cdot\left(\sum_{i=1}^{7} v_i x_i\right)^2\right] = \frac{1}{14}\left(93.200 - \frac{1}{15}\cdot 1.160^2\right) = 249{,}524$$

Η τυπική απόκλιση θα είναι:

$$s = \sqrt{s^2} = \sqrt{249{,}524} = 15{,}796$$

Άρα, $s = 15{,}796$

δ) Υπολογισμός του συντελεστή μεταβλητότητας.

$$CV = \frac{s}{\bar{x}}100 \Rightarrow CV = \frac{15{,}796}{77{,}333}100 = 20{,}426$$
$$CV = 20{,}426\%$$

Για να προχωρήσουμε σε μια ασφαλή σύγκριση με άλλη επιχείρηση του ίδιου εμπορικού κλάδου είναι προφανές ότι απαιτείται ο υπολογισμός του CV της άλλης επιχείρησης, έτσι θα έχουμε:

$$CV = \frac{s}{\bar{x}}100 = \frac{17}{70}100 = 24{,}286$$
$$CV = 24{,}286\%$$

Σχετικά λοιπόν με τη διασπορά των πωλήσεων αυτών των δύο επιχειρήσεων μπορούμε να πούμε ότι οι πωλήσεις της επιχείρησης που εξετάσαμε παρουσιάζουν μικρότερη σχετικά διασπορά και καλύτερη ομοιογένεια σε σχέση με τις πωλήσεις της άλλης επιχείρησης, καθόσον έχουμε μικρότερο συντελεστή μεταβλητότητας (20,426< 24,286).

ΕΦΑΡΜΟΓΗ 3

Περίπτωση μελέτης αποτελέσματος μετά από αναδιοργάνωση επιχείρησης.

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

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

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

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

Ως προς τη περσινή χρήση γνωρίζουμε ότι η μέση τιμή των πωλήσεων ήταν $\bar{x}_1$ =90.000.000 ευρώ και η τυπική απόκλιση $s_1 = 10.000.000$ ευρώ. Ακόμη, οι χαμηλότερες πωλήσεις υποκαταστήματος ήταν 70.000.000 ευρώ και οι υψηλότερες 140.000.000 ευρώ. Ο γενικός διευθυντής συντάσσει αρχικά τον παρακάτω πίνακα και υπολογίζει τη νέα μέση τιμή και τη νέα τυπική απόκλιση:

Πίνακας 15

Κλάσεις Πωλήσεων	**Αριθμός Υποκ/των** v_i	x_i	x_i^2	$x_i v_i$	$v_i x_i^2$
100-110	1	105	11.025	105	11.025
110-120	3	115	13.225	345	39.675
120-130	6	125	15.625	750	93.750
130-140	6	135	18.225	814	109.350
140-150	3	145	21.025	435	63.075
150-160	1	155	24.025	155	24.025
Σύνολο	20			2.600	340.900

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

$$\bar{x}_2 = \frac{\sum_{i=1}^{6} v_i x_i}{\sum_{i=1}^{6} v_i} = \frac{2.600}{20} = 130 \text{ εκατομμύρια ευρώ.}$$

Η νέα τυπική απόκλιση s_2 θα είναι:

$$s_2 = \sqrt{\frac{1}{v-1}\left[\sum_{i=1}^{6} v_i x_i^2 - \frac{1}{v}\left(\sum_{i=1}^{6} v_i x_i\right)^2\right]} = \sqrt{\frac{1}{19}\left(340.900 - \frac{1}{20}\cdot 2.600^2\right)} = \sqrt{152,632} = 12,354$$

Άρα, $s_2 = 12{,}354$ εκατομμύρια 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 είναι μεγαλύτερη από της περσινής s_1=10, η διασπορά των εσόδων φέτος παρουσιάζει χειρότερη εικόνα. Για να είμαστε σίγουροι προχωράμε στον υπολογισμό και των αντίστοιχων CV_1 και CV_2 και έχουμε:

$$CV_1 = \frac{s_1}{x_1} = \frac{10}{90} = 0{,}111 \quad ή \quad CV_1 = 11{,}1\%$$

$$CV_2 = \frac{s_2}{\overline{x}} = \frac{12{,}354}{130} = 0{,}095 \quad ή \quad CV_1 = 9{,}5\%$$

Τιμές των στατιστικών μέτρων των 2 διαχειριστικών χρήσεων.

Πίνακας 16

Στατιστικά Μέτρα	**Περσινή χρήση (1)**	**Φετινή χρήση (2)**	**Μεταβολή των (1) και (2)**
Μέση Τιμή $\overline{x}$	$\overline{x}_1 = 90$	$\overline{x}_2 = 130$	Αύξηση
Τυπική Απόκλιση s	$s_1 = 10$	$s_2 = 12{,}354$	Αύξηση
Διακύμανση s^2	$s_1^2 = 100$	$s_2^2 = 152{,}621$	Αύξηση
Συντελ.Μεταβλητότητας CV	$CV_1 = 11{,}1\%$	$CV_2 = 9{,}5\%$	Μείωση
Εύρος Μεταβ. *Ε.Μ.*	*Ε.Μ.*$_1$=70	*Ε.Μ.*$_2$=51	Μείωση

<u>ΣΥΜΠΕΡΑΣΜΑΤΑ :</u>

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 90.000.000 ευρώ πέρυσι 130.000.000 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 $CV_2 = 9{,}5\% < CV_1 = 11{,}1\%$.

Ακόμη, έχουμε φέτος ένα εύρος μεταβολής ίσο με: 156-105=51 εκατομμύρια , το οποίο είναι αρκετά μικρότερο από το περσινό, που ήταν 140-70=70 εκατομμύρια ευρώ. Φυσικά η μελέτη δεν 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.

ΕΦΑΡΜΟΓΗ 4

Μελέτη προβλήματος πωλήσεων.

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 κ.τ.λ. Μια εμπορική επιχείρηση Α είχε την περασμένη χρονιά πωλήσεις εμπορευμάτων αξίας 500 εκατομμύρια ευρώ και επιστροφές εμπορευμάτων αξίας 50 εκατομμύρια ευρώ. Ο νέος γενικός διευθυντής θέλοντας να εξακριβώσει 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 παρακάτω πίνακα.

Πίνακας 17

ΑΞΙΑ ΠΩΛΗΣΕΩΝ x_i	20	30	40	40	50	50	60	70	80	90
ΑΞΙΑ ΕΠΙΣΤΡΟΦΩΝ y_i	1	3	3	4	5	6	6	8	9	10

(ΔΕΔΟΜΕΝΑ ΣΕ ΔΕΚΑΔΕΣ ΕΚΑΤΟΜΜΥΡΙΑ ΕΥΡΩ)

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 επιστρέφονται, αν οι πωλήσεις αυξάνονταν κατά μια μονάδα , δηλαδή κατά 10 εκατομμύρια ευρώ.

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

5.Ποια η μέση αξία των πωλήσεων και των επιστροφών κατά έτος αυτής της ομάδας επιχειρήσεων.

6.Ποια η διάμεση αξία των επιστροφών κατά έτος.

7.Όπου κρίνεται απαραίτητο να γίνει το σχετικό διάγραμμα.

ΛΥΣΗ

1. Για να υπολογίσουμε τον συντελεστή γραμμικής συσχέτισης των μεταβλητών «αξία πωλήσεων» (*X)* και «αξία επιστροφών» (*Y*) συμπληρώνουμε τον παρακάτω πίνακα δεδομένων:

Πίνακας 18

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

Αντικαθιστούμε στον τύπο υπολογισμού του συντελεστή συσχέτισης και έχουμε:

$$r = \frac{10\sum_{i=10}^{10} x_i y_i - \left(\sum_{i=1}^{10} x_i\right)\left(\sum_{i=10}^{10} y_i\right)}{\sqrt{10\sum_{i=1}^{10} x_i^{\,2} - \left(\sum_{i=1}^{10} x_i\right)^2}\sqrt{10\sum_{i=1}^{10} y_i^{\,2} - \left(\sum_{i=1}^{10} y_i\right)^2}} = \frac{10 \cdot 3480 - 530 \cdot 55}{\sqrt{10 \cdot 32500 - (530)^2}\sqrt{10 \cdot 377 - (55)^2}} =$$

$$= \frac{5650}{210 \cdot 27{,}295} = 0{,}986$$

Παρατηρούμε ότι υπάρχει ισχυρή συσχέτιση μεταξύ των μεταβλητών X και Y.

2. Θα υπολογίσουμε τις εκτιμήτριες ελαχίστων τετραγώνων $\hat{a}, \hat{b}$ και θα προσδιορίσουμε την εξίσωση της ευθείας παλινδρόμησης.

Από τα στοιχεία του πίνακα υπολογίζουμε το συντελεστή διεύθυνσης της ευθείας παλινδρόμησης.

$$\hat{b} = \frac{10\sum_{i=1}^{10} x_i y_i - \left(\sum_{i=1}^{10} x_i\right)\left(\sum_{i=1}^{10} y_i\right)}{10\sum_{i=1}^{10} x_i^{\,2} - \left(\sum_{i=1}^{10} x_i\right)^2} = \frac{5650}{44100} = 0{,}128$$

Επίσης υπολογίζουμε:

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

και

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

Οπότε από τη σχέση $\hat{a} = \overline{y} - \hat{b}\overline{x}$ προκύπτει ότι:

$$\hat{a} = 5{,}5 - 0{,}128 \cdot 53 \Rightarrow \hat{a} = -1{,}29$$

Άρα η ευθεία που προσαρμόζεται καλύτερα στα δεδομένα είναι:

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

Παρατηρούμε ότι, αν $\hat{y} = 0$, τότε $x \cong 10$ που σημαίνει ότι για πωλήσεις εμπορευμάτων αξίας περίπου μέχρι και 100 εκατομμύρια ευρώ δεν υπάρχουν επιστροφές εμπορευμάτων (υπενθυμίζουμε ότι τα δεδομένα είναι σε δεκάδες εκατομμυρίων).

3. Αν οι πωλήσεις αυξάνονταν κατά μια μονάδα τότε:

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

Συμπεραίνουμε ότι, αν οι πωλήσεις αυξάνονται κατά μια μονάδα δηλαδή κατά 10 εκατομμύρια ευρώ, τότε το ύψος των επιστρεφόμενων εμπορευμάτων θα αυξανόταν κατά 0,128 δεκάδες εκατομμύρια, δηλαδή 1,28 εκατομμύρια ευρώ όσο ακριβώς και η τιμή του συντελεστή $\hat{b}$, ($\hat{b} = 0,128$).

4. Αν οι ετήσιες πωλήσεις μιας επιχείρησης ήταν της τάξης των 35 εκατομμυρίων ευρώ, τότε το ύψος των επιστρεφόμενων εμπορευμάτων θα υπολογιζόταν από τον τύπο:

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

Άρα,

$$\hat{y} = -1,29 + 0,128 \cdot 35 = 3.19$$

Δηλαδή, η αξία των εμπορευμάτων που επιστράφηκαν θα ήταν 3.190.000 ευρώ.

5.Η μέση αξία των εμπορευμάτων που πουλήθηκαν κατά έτος από αυτή την ομάδα των επιχειρήσεων είναι 530.000.000 ($\bar{x}$ =53) και αντίστοιχα η μέση αξία των επιστροφών είναι 55.000.000 ($\bar{y}$ =5,5). Οι συντεταγμένες του σημείου (53,5.5) επαληθεύουν την ευθεία παλινδρόμησης.

6.Η διάμεση αξία των προϊόντων που επιστράφηκαν κατά έτος υπολογίζεται ως εξής:
Αξία επιστροφών : 1, 3, 3, 4, 5, 6, 6, 8, 9, 10.

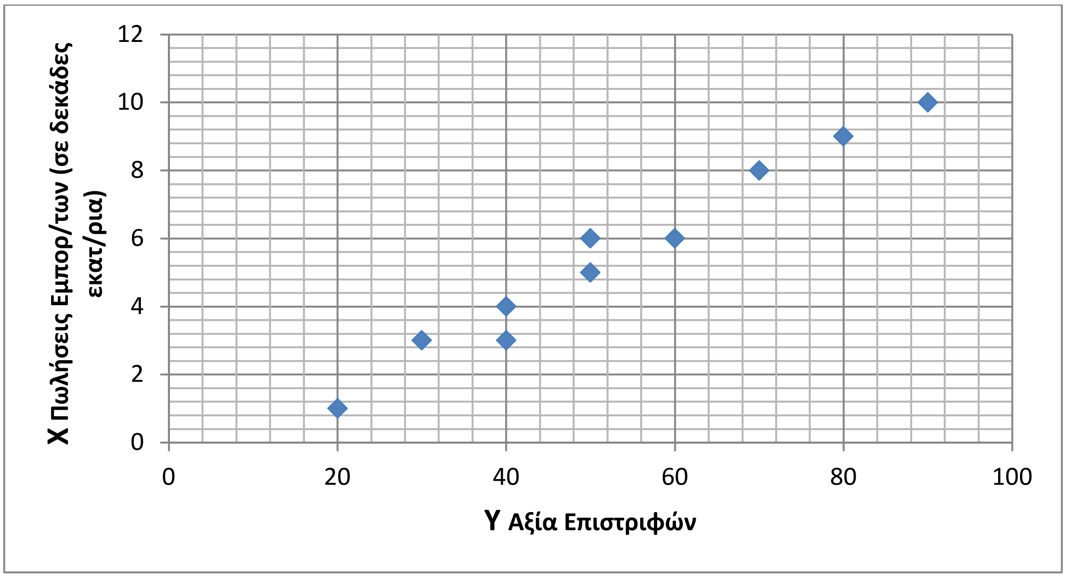
Ο αριθμός των επιχειρήσεων είναι 10 (άρτιος) άρα η διάμεσος θα βρίσκεται στην $\frac{10+1}{2}=5{,}5$ θέση, δηλαδή ανάμεσα στην $5^{η}$ και $6^{η}$ παρατήρηση και θα ισούται με $\frac{5+6}{2}=5{,}5$, δηλαδή αξία των επιστροφών θα είναι 55.000.000 ευρώ.

Παρατηρούμε ότι η μέση τιμή της αξίας των εμπορευμάτων των 10 επιχειρήσεων που επιστρέφονται κατά έτος είναι 55.000.000 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.

7.

ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ

Διάγραμμα 5



ΕΦΑΡΜΟΓΗ 5

Η αξία των επενδυμένων κεφαλαίων (X) και τα αντίστοιχα κέρδη (Y) ομάδας 10 ομοειδών βιομηχανικών επιχειρήσεων περιέχονται στον παρακάτω πίνακα (τα δεδομένα εκφράζουν εκατομμύρια ευρώ).

Πίνακας 19

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:

α) Να βρεθεί ο συντελεστής γραμμικής συσχέτισης *r* και να σχολιασθεί το αποτέλεσμα.

β) Να ερμηνευθεί η παράμετρος $\hat{b}$ της $\hat{y} = \hat{a} + \hat{b}x$.

ΛΥΣΗ

α)Σχηματίζουμε τον παρακάτω πίνακα:

Πίνακας 20

	Κεφάλαιο x_i	**Κέρδη** y_i	$x_i y_i$	x_i^2	y_i^2
1	80	15	1200	6400	225
2	90	20	1800	8100	400
3	100	25	2500	10000	625
4	100	25	3000	14400	625
5	150	30	4500	22500	900
6	180	30	5400	32400	900
7	200	40	8000	40000	1600
8	250	40	10000	62500	1600
9	300	60	18000	90000	3600
10	350	80	28000	122500	6400
Σύνολο	1820	365	82400	408800	16875

Υπολογισμός του συντελεστή παλινδρόμησης

Κατ ΄ αρχήν υπολογίζουμε τα αθροίσματα των στηλών των δεδομένων:

$$\sum_{i=1}^{10} x_i = 1820$$

και

$$\sum_{i=1}^{10} y_i = 365$$

Καθώς και τα αθροίσματα:

$$\sum_{i=1}^{10} x_i y_i = 82400$$

$$\sum_{i=1}^{10} x_i^{\,2} = 408800$$

και

$$\sum_{i=1}^{10} y_i^{\,2} = 16875$$

Τέλος αντικαθιστούμε τα αθροίσματα στον τύπο υπολογισμού *r*:

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

$$= \frac{10 \cdot 82400 - 1829 \cdot 365}{\sqrt{10 \cdot 408800 - (1820)^2}\sqrt{10 \cdot 16875 - (365)^2}}$$

$$= 0{,}962$$

Παρατηρούμε ότι σε αυτή την ομάδα των βιομηχανικών επενδύσεων υπάρχει μια θετική και πολύ έντονη συσχέτιση μεταξύ των επενδυμένων κεφαλαίων και των αντίστοιχων κερδών.

β)<u>Εύρεση της εξίσωσης παλινδρόμησης</u>

Υπολογίζουμε τα $\bar{x}, \bar{y}$:

$$\overline{x} = \frac{\sum_{i=1}^{10} x_i}{\nu} = \frac{1820}{10} = 182$$

$$\overline{y} = \frac{\sum_{i=1}^{10} y_i}{\nu} = \frac{365}{10} = 36{,}5$$

Αντικαθιστούμε στο τύπο υπολογισμού του $\hat{b}$:

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

$$= \frac{10 \cdot 82400 - 1820 \cdot 365}{10 \cdot 408800 - (1820)^2}$$

$$= \frac{824000 - 664300}{4088000 - 3312400}$$

$$= \frac{159700}{775600} = 0{,}206 \cong 0{,}2.$$

Επίσης, έχουμε:

$$\hat{a} = \overline{y} - \hat{b}\overline{x}$$

$$\hat{a} = 36{,}5 - 0{,}206 \cdot 182$$

$$\hat{a} \cong -1$$

Άρα, η ζητούμενη εξίσωση παλινδρόμησης θα είναι:

$$\hat{y} = -1 + 0{,}2x$$

Για παράδειγμα:

α) για $x = 35$ έχουμε $\hat{y} = -1 + 0{,}2 \cdot 35 = 6$

β) για $x = 70$ έχουμε $\hat{y} = -1 + 0{,}2 \cdot 70 = 13$

Σε μια αύξηση του επενδυμένου κεφαλαίου κατά μια μονάδα, δηλαδή ένα εκατομμύριο ευρώ τα κέρδη αναμένεται να αυξηθούν κατά 206.000 ευρώ, αφού το $\hat{b} = 0{,}206$.

ΕΦΑΡΜΟΓΗ 6

Ο παρακάτω πίνακας αναφέρεται σε 10 ομοειδείς εμπορικές επιχειρήσεις και δείχνει τις ετήσιες πωλήσεις τους (*Y*) και τις ετήσιες διαφημιστικές τους δαπάνες (*X*). (Τα δεδομένα εκφράζονται σε εκατομμύρια ευρώ):

Πίνακας 21

X	20	22	25	30	32	35	40	50	60	70
Y	250	270	260	350	330	300	370	480	450	500

Ζητείται:

α) Να βρεθεί ο συντελεστής γραμμικής συσχέτισης *r* και να σχολιασθεί το αποτέλεσμα.

β) Να προσδιορισθεί η εξίσωση γραμμικής παλινδρόμησης.

ΛΥΣΗ

α)

Πίνακας 22

	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	20	250	5000	400	62500
2	22	270	5940	484	72900
3	25	260	6500	625	67600
4	30	350	10500	900	122500
5	32	330	10560	1024	108900
6	35	300	10500	1225	90000
7	40	370	14800	1600	136900
8	50	480	24000	2500	230400
9	60	450	27000	3600	202500
10	70	500	35000	4900	250000
Σύνολο	384	3560	149800	17258	1344200

Αντικαθιστούμε στον τύπο υπολογισμού του συντελεστή συσχέτισης τα αθροίσματα:

$$\sum_{i=1}^{10} x_i = 384 , \sum_{i=1}^{10} y_i = 3.560 , \sum_{i=1}^{10} x_i y_i = 149800$$

$$\sum_{i=1}^{10} x_i^2 = 17.258 , \sum_{i=1}^{10} y_i^2 = 1.344.200$$

$$\begin{aligned} r &= \frac{10\sum_{i=1}^{10} x_i y_i - \left(\sum_{i=1}^{10} x_i\right)\left(\sum_{i=1}^{10} y_i\right)}{\sqrt{10\sum_{i=1}^{10} x_i^2 - \left(\sum_{i=1}^{10} x_i\right)^2}\sqrt{10\sum_{i=1}^{10} y_i^2 - \left(\sum_{i=1}^{10} y_i\right)^2}} \\ &= \frac{10\cdot 149.800 - 384 \cdot 3560}{\sqrt{10\cdot 17.258 - (384)^2}\sqrt{10\cdot 1.344.200 - (3.560)^2}} \\ &= \frac{1.498.000 - 1.367.040}{\sqrt{172.580 - 147.456}\sqrt{1.344.200 - 12.673.600}} \\ &= \frac{130.960}{158{,}506 \cdot 876{,}584} = 0{,}943. \end{aligned}$$

Από το παραπάνω αποτέλεσμα που βρήκαμε (r = 0,943) συμπεραίνουμε ότι υπάρχει θετική συσχέτιση μεταξύ των διαφημιστικών δαπανών (X) και των πωλήσεων (Y). Επίσης παρατηρούμε ότι αυτή η γραμμική συσχέτιση μεταξύ διαφημιστικών δαπανών και πωλήσεων είναι έντονη, γιατί το r εκτός από θετικό είναι πολύ κοντά στη μονάδα.

β)Προσδιορισμός της εξίσωσης γραμμικής παλινδρόμησης $\hat{y} = \hat{a} + \hat{b}x$.

Υπολογισμός της παραμέτρου $\hat{b}$

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

$$= \frac{10 \cdot 149.800 - 384 \cdot 3560}{10 \cdot 17.258 - (384)^2}$$

$$= \frac{1.498.000 - 1.367.040}{172.580 - 147.456} =$$

$$= \frac{130.960}{25.124} = 5,213.$$

Υπολογισμός της παραμέτρου $\hat{a}$

Αρχικά βρίσκουμε τους μέσους αριθμητικούς:

$$\overline{x} = \frac{\sum_{i=1}^{10} x_i}{v} = \frac{384}{10} = 38,4$$

$$\overline{y} = \frac{\sum_{i=1}^{10} y_i}{v} = \frac{3560}{10} = 356$$

Συνεπώς, έχουμε:

$$\hat{a} = \overline{y} - \hat{b}\overline{x}$$

ή

$$\hat{a} = 356 - 5,213 \cdot 38,4$$

ή

$$\hat{a} = 356 - 200,1792 = 155,82$$

Δηλαδή, η ζητούμενη εξίσωση παλινδρόμησης θα είναι:

$$\hat{y} = 155,82 + 5,213x\ .$$

Ερμηνεία των παραμέτρων $\hat{a}$ και $\hat{b}$ της εξίσωσης παλινδρόμησης.

Κατ ΄ αρχήν η παράμετρος $\hat{a} = 155,82$ μας δηλώνει ότι το ύψος των πωλήσεων μιας επιχείρησης, η οποία δεν προβαίνει σε καμία διαφημιστική δαπάνη, θα ήταν 155,82 εκατομμύρια ευρώ (155.820.000 ευρώ).

Η παράμετρος $\hat{b} = 5,213$ μας δηλώνει ότι σε αυτήν την ομάδα των επιχειρήσεων με αύξηση των διαφημιστικών δαπανών κατά μια μονάδα, δηλαδή κατά ένα εκατομμύριο ευρώ, αναμένεται να σημειωθεί αύξηση των πωλήσεων κατά 5,213 εκατομμύρια ευρώ (5.213.000 ευρώ).

ΕΦΑΡΜΟΓΗ 7

Μια 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 Y) σε ορισμένο χρονικό διάστημα.

Πίνακας 23

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:

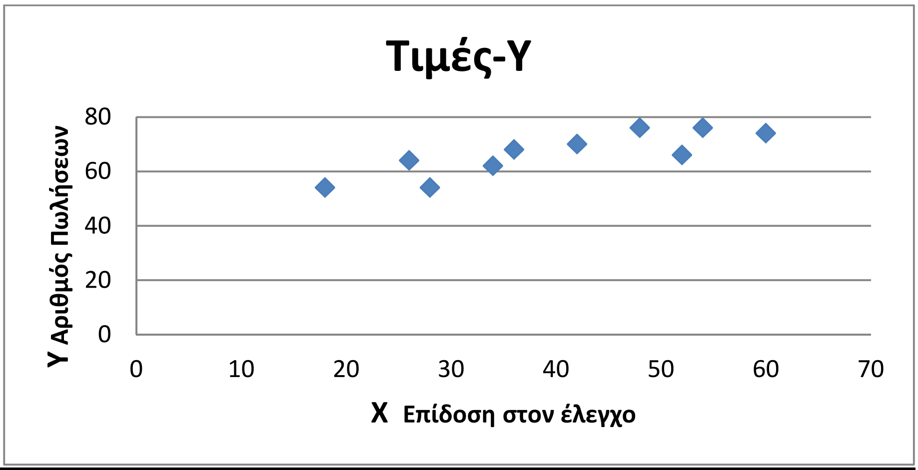
α) Να παρασταθούν γραφικά τα δεδομένα του πίνακα και να βρεθεί η ευθεία ελαχίστων τετραγώνων.

β) Ποιος ο αριθμός των πωλήσεων που περιμένετε από 3 πωλητές των οποίων τα αποτελέσματα στον έλεγχο δεξιοτήτων ήταν 40, 50 και 70 αντίστοιχα;

ΛΥΣΗ

α)

Διάγραμμα 6



Συμπληρώνουμε τον παρακάτω πίνακα:

Πίνακας 24

x_i	y_i	x_i^2	$x_i y_i$
18	54	324	972
26	64	676	1664
28	54	784	1512
34	62	1156	2108
36	68	1296	2448
42	70	1764	2940
48	76	2304	3648
52	66	2704	3432
54	76	2916	4104
60	74	3600	4440
$\sum_{i=1}^{10} x_i = 398$	$\sum_{i=1}^{10} y_i = 664$	$\sum_{i=1}^{10} x_i^2 = 17.524$	$\sum_{i=1}^{10} x_i y_i = 27.268$

Υπολογίζουμε τη μέση τιμή των μεταβλητών X και Y:

$$\overline{x}=\frac{\sum_{i=1}^{10}x_i}{\nu}=\frac{398}{10}=39{,}8$$

$$\overline{y}=\frac{\sum_{i=1}^{10}y_i}{\nu}=\frac{664}{10}=66{,}4$$

Υπολογίζουμε τις εκτιμήτριες $\hat{a}$ και $\hat{b}$ ως εξής:

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

$$=\frac{10\cdot 27.268-398\cdot 664}{10\cdot 17.524-(398)^2}=\frac{8.408}{16.836}\cong 0,5$$

$$\hat{a}=\overline{y}-\hat{b}\overline{x}=66,4-0,5\cdot 39,8=66,4-19,9=46,5$$

Άρα η ευθεία παλινδρόμησης είναι:

$$\hat{y}=46{,}5+0{,}5x\,.$$

β) Αν τα αποτελέσματα στο έλεγχο δεξιοτήτων ήταν 40, 50 και 70, τότε ο αριθμός των πωλήσεων που αναμένεται από τους τρείς πωλητές θα ήταν αντίστοιχα:

$x=40 \quad \hat{y}=46{,}5+0{,}5\cdot 40=66{,}5$

$x=50 \quad \hat{y}=46{,}5+0{,}5\cdot 50=71{,}5$

$x=70 \quad \hat{y}=46{,}5+0{,}5\cdot 70=81{,}5$

ΕΦΑΡΜΟΓΗ 8

Ο παρακάτω πίνακας δείχνει το σύνολο των εξαγωγών μιας επιχείρησης σε εκατομμύρια ευρώ σε δύο διαδοχικά χρόνια:

Πίνακας 25

ΈΤΟΣ	ΑΦΡΙΚΗ	ΑΜΕΡΙΚΗ	ΑΣΙΑ	ΕΥΡΩΠΗ
2011	80,4	120,2	150,6	230,8
2012	50,5	60,7	130,2	190,6

α) Κατασκευάστε δύο κυκλικά διαγράμματα που να απεικονίζουν τις εξαγωγές της επιχείρησης κατά τα έτη 2011 και 2012. Με τη βοήθεια των κυκλικών διαγραμμάτων να συγκρίνετε τις ετήσιες εξαγωγές της επιχείρησης.

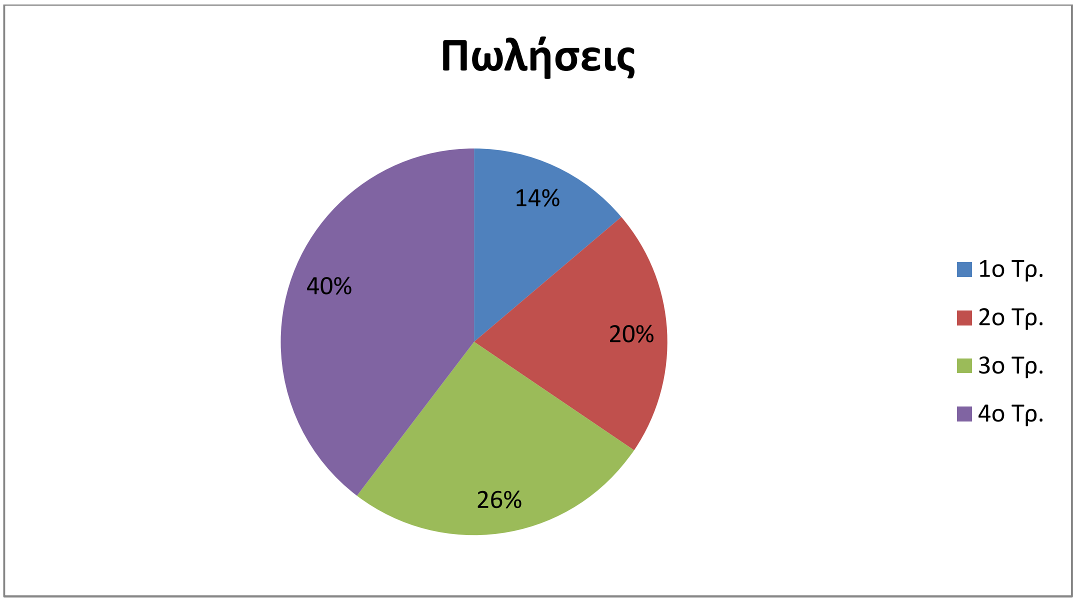
β) Πόσο τοις εκατό αυξήθηκαν ή μειώθηκαν οι εξαγωγές της επιχείρησης τα έτη 2011-2012 σε κάθε ήπειρο.

ΛΥΣΗ

α)Με τη βοήθεια των κυκλικών διαγραμμάτων παρατηρούμε ότι η ποσοστιαία σύνθεση των εξαγωγών παρέμεινε κατά τη διάρκεια των δύο ετών περίπου η ίδια.

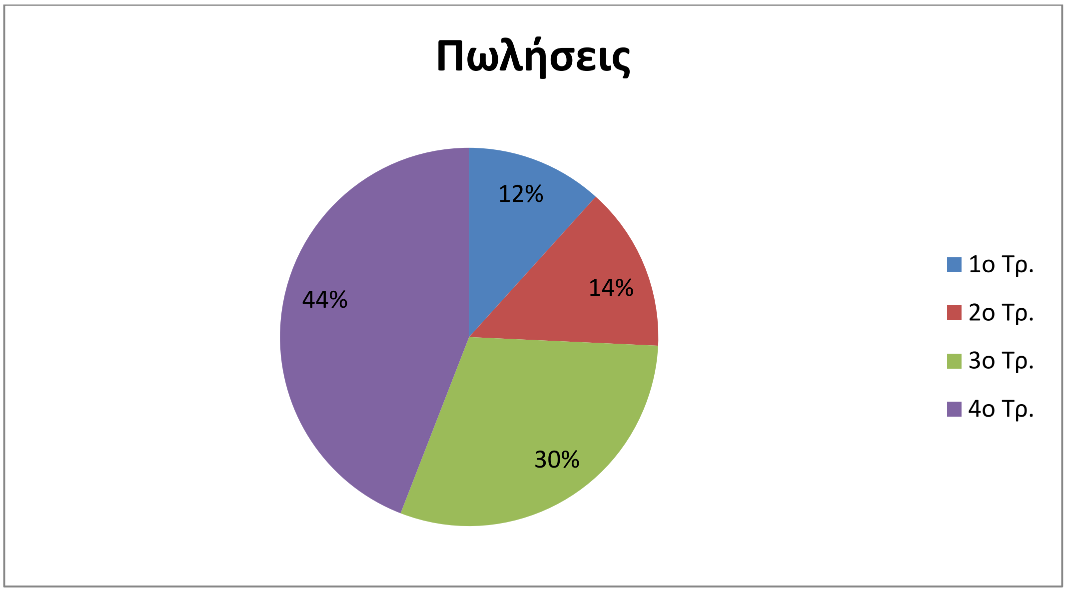
Πίνακας 26

Έτος 2011



Πίνακας 27

Έτος 2012



β)

Πίνακας 28

ΠΟΣΟΣΤΟ ΜΕΙΩΣΗΣ	**ΑΦΡΙΚΗ**	**ΑΜΕΡΙΚΗ**	**ΑΣΙΑ**	**ΕΥΡΩΠΗ**
2011-2012(%)	37,2	49,5	13,6	17,4

ΕΦΑΡΜΟΓΗ 9

Το ύψος των ετήσιων κερδών μιας ομάδας 20 επιχειρήσεων σε εκατομμύρια ευρώ είναι:

Πίνακας 29

ΚΕΡΔΗ	25	30	35	40	45	50
ΕΠΙΧΕΙΡΗΣΕΙΣ	3	3	6	5	2	1

Ζητείται να υπολογισθούν και να ερμηνευθούν:

α) Ο αριθμητικός μέσος

β) Η διάμεσος

γ) Το εύρος μεταβολής

δ) Η τυπική απόκλιση

ε) Ο συντελεστής μεταβλητότητας

<u>ΛΥΣΗ</u>

α) Υπολογισμός του αριθμητικού μέσου

Κατασκευάζουμε τον παρακάτω αριθμητικό πίνακα:

Πίνακας 30

x_i	v_i	$x_i v_i$
25	3	75
30	3	90
35	6	210
40	5	200
45	2	90
50	1	50
Σύνολο	20	715

Αντικαθιστώντας το άθροισμα

$$\sum_{i=1}^{6} x_i y_i = 715$$

στον τύπο υπολογισμού του μέσου αριθμητικού έχουμε:

$$\bar{x} = \frac{\sum_{i=1}^{6} v_i x_i}{v} = \frac{715}{20} = 35{,}75$$

Δηλαδή, ο μέσος όρος των ετήσιων κερδών αυτής της ομάδας επιχειρήσεων είναι 35.750.000 ευρώ.

β)Υπολογισμός της διαμέσου

Κατασκευάζουμε τον παρακάτω αριθμητικό πίνακα:

Πίνακας 31

x_i	v_i	N_i
25	3	3
30	3	6
35	6	12
40	5	17
45	2	19
50	1	20
	ν=20	

Όπως παρατηρούμε ο αριθμός ν/2=20/2=10 βρίσκεται μεταξύ των αθροιστικών συχνοτήτων 6 και 12 και συνεπώς η διάμεσος είναι ίση με 35, δηλαδή δ=35.000.000.

Το 50% λοιπόν των επιχειρήσεων έχει αποδοχές 350.000.000 ευρώ και κάτω και το υπόλοιπο 50% 350.000.000 ευρώ και πάνω.

γ)Το εύρος μεταβολής ως γνωστό είναι η διαφορά μεταξύ των δύο ακραίων τιμών που παίρνει η μεταβλητή X και στη προκειμένη περίπτωση είναι:

Εύρος μεταβολής:50-25=25

Άρα, τα ετήσια κέρδη της πιο κερδοφόρας επιχείρησης είναι διπλάσια από αυτά της λιγότερο κερδοφόρας επιχείρησης.

δ)Υπολογισμού της τυπικής απόκλισης

Κατασκευάζουμε τον παρακάτω πίνακα αριθμητικών υπολογισμών.

Πίνακας 32

x_i	v_i	$x_i v_i$	x_i^2	$v_i x_i^2$
25	3	75	625	1875
30	3	90	900	2700
35	6	210	1225	7350
40	5	200	1600	8000
45	2	90	2025	4050
50	1	50	2500	2500
	$v = 20$	$\sum_{i=1}^{6} v_i x_i = 715$		$\sum_{i=1}^{6} v_i x_i^2 = 26.475$

Η διακύμανση είναι:

$$s = \frac{1}{v-1}\left[\sum_{i=1}^{6} v_i x_i^{2} - \frac{1}{v}\cdot\left(\sum_{i=1}^{6} v_i x_i\right)^2\right] = \frac{1}{19}\left(26.475 - \frac{1}{20}\cdot 715^2\right) = 48{,}092$$

Άρα, η τυπική απόκλιση:

$$s = \sqrt{48{,}092} = 6{,}935$$

Δηλαδή, η τετραγωνική ρίζα της μέσης τιμής των τετραγώνων των αποκλίσεων των ετήσιων κερδών κάθε επιχείρησης από το μέσο ετήσιο κέρδος τους (35.750.00) είναι ίση με 6.935.000 ευρώ.

ε)Υπολογισμός συντελεστή μεταβλητότητας.

Όπως γνωρίζουμε ο συντελεστής μεταβλητής CV ορίζεται ως ο λόγος της s δια του $\bar{x}$ επί 100, δηλαδή έχουμε:

$$CV = \frac{s}{\bar{x}}100 = \frac{6{,}935}{35{,}75}100 = 19{,}4\%$$

Μπορούμε να πούμε ότι η σχετική διασπορά είναι ικανοποιητική, αφού το αποτέλεσμα που προέκυψε $CV = 19{,}4\%$, όπως αντιλαμβανόμαστε, δεν είναι ιδιαίτερα υψηλό.

ΕΦΑΡΜΟΓΗ 10

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

Ζητείται να υπολογισθούν και να ερμηνευθούν:

α)Η μέση τιμή

β)Η διακύμανση

γ)Η τυπική απόκλιση

ΛΥΣΗ

α)Υπολογισμός της μέσης τιμής

Κατασκευάζουμε τον παρακάτω πίνακα αριθμητικών υπολογισμών:

Πίνακας 33

ΚΛΑΣΕΙΣ	v_i	x_i	$x_i v_i$	x_i^2	$v_i x_i^2$
[0-20)	2	10	20	100	200
[20-40)	3	30	90	900	2700
[40-60)	8	50	400	2500	20000
[60-80)	10	70	700	4900	49000
[80-100)	4	90	360	8100	32400
[100-120)	2	110	220	12100	24200
[120-140)	1	130	130	16900	16900
ΣΥΝΟΛΟ	ν=30		$\sum_{i=1}^{7} x_i v_i = 1.920$		$\sum_{i=1}^{7} v_i x_i^2 = 145.400$

Η μέση τιμή είναι:

$$\overline{x} = \frac{\sum_{i=1}^{7} x_i v_i}{v} = \frac{1920}{30} = 64$$

Δηλαδή, αυτές οι 30 επιχειρήσεις στη χρονιά που εξετάζουμε δαπάνησαν κατά μέσο όρο 64.000.000 ευρώ για διαφημιστικά έξοδα.

β)Υπολογισμός της διακύμανσης

Η διακύμανση είναι:

$$s^2 = \frac{1}{v-1}\left[\sum_{i=1}^{7} v_i x_i^{\ 2} - \frac{1}{v}\cdot\left(\sum_{i=1}^{7} v_i x_i\right)^2\right] = \frac{1}{29}\left(145.400 - \frac{1}{30}\cdot 1920^2\right) = 776{,}552.$$

γ) Υπολογισμός της τυπικής απόκλισης:

Η τυπική απόκλιση είναι:

$$s = \sqrt{776{,}552} = 27{,}86668$$

Όπως γνωρίζουμε, όσο μικρότερες είναι οι τιμές της s^2 (και τις s φυσικά) τόσο πιο συγκεντρωμένες γύρω από τη μέση τιμή βρίσκονται οι παρατηρήσεις μας. Επίσης, η τετραγωνική ρίζα του μέσου αριθμητικού των τετραγώνων των αποκλίσεων των δεδομένων μας από τη μέση ετήσια δαπάνη διαφήμισης (64 εκατ ευρώ) ισούται με 27,9 εκατ. ευρώ περίπου.

ΕΦΑΡΜΟΓΗ 11

Περίπτωση μελέτης επιχείρησης δύο εκμεταλλεύσεων

Μια εταιρεία ηλεκτρονικών συσκευών διαθέτει δύο κλάδους εκμετάλλευσης, τον κλάδο (*X*) εμπορίας ηλεκτρονικών συσκευών και τον κλάδο (*Y)* συντήρησης και επισκευής ηλεκτρονικών συσκευών. Ο κύκλος εργασιών του κάθε κλάδου εκμετάλλευσης παρακολουθείται χωριστά. Στη διάρκεια του προηγούμενου οικονομικού έτους τα έσοδα του κλάδου εμπορίας ηλεκτρικών συσκευών (*X*) ήταν 50 εκατομμύρια ευρώ, του δε κλάδου συντήρησης – επισκευής (*Y*) ήταν 10 εκατομμύρια 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 στα παρακάτω ερωτήματα.

Πίνακας 34

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 εκατ. ευρώ)

1. 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 έχουμε στα έσοδα της εκμετάλλευσης *Y*, αν τα έσοδα της εκμετάλλευσης *X* ήταν αυξημένα κατά 10 εκατομμύρια 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 ένα εκατομμύριο ευρώ;

5. Να βρεθεί σε αυτή την ομάδα των 12 επιχειρήσεων η τυπική απόκλιση εσόδων για την εκμετάλλευση *X*.

6. Να βρεθεί η τυπική απόκλιση εσόδων 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 εμπορίας ηλεκτρικών συσκευών *X*.

11. Να βρεθεί το διάμεσο ύψος εσόδων της εκμετάλλευσης επισκευών-συντήρησης *Y*.

12. Να γίνει συνοπτική ερμηνεία για όλα τα αποτελέσματα των προηγούμενων περιπτώσεων.

13. Να γίνουν τα σχετικά διαγράμματα για την καλύτερη παρουσίαση του ζητήματος.

ΛΥΣΗ

1. Αρκεί να υπολογίσουμε το συντελεστή γραμμικής συσχέτισης *r*.

Κατασκευάζουμε τον παρακάτω πίνακα:

Πίνακας 35

x_i	y_i	$x_i y_i$	x_i^2	y_i^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
$\sum_{i=1}^{12} x_i = 533$	$\sum_{i=1}^{12} y_i = 106$	$\sum_{i=1}^{12} x_i y_{ii} = 4827$	$\sum_{i=1}^{12} x_i^{\,2} = 24319$	$\sum_{i=1}^{12} y_i^{\,2} = 959{,}5$

Έχουμε:

$$r = \frac{12\sum_{i=1}^{12} x_i y_i - \left(\sum_{i=1}^{12} x_i\right)\left(\sum_{i=1}^{12} y_i\right)}{\sqrt{12\sum_{i=1}^{12} x_i^{\,2} - \left(\sum_{i=1}^{12} x_i\right)^2}\ \sqrt{12\sum_{i=1}^{12} y^{\,2} - \left(\sum_{i=1}^{12} y_i\right)^2}}$$

$$= \frac{12 \cdot 4.827 - 533 \cdot 106}{\sqrt{12 \cdot 24.319 - 533^2}\sqrt{12 \cdot 959{,}5 - 106^2}} = 0{,}97$$

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:

$$\bar{x} = \frac{\sum_{i=1}^{12} x_i}{v} = \frac{533}{12} = 44{,}4167$$

$$\bar{y} = \frac{\sum_{i=1}^{12} y_i}{v} = \frac{106}{12} = 8{,}8333$$

$$\hat{b} = \frac{12\sum_{i=1}^{12} x_i y_i - \left(\sum_{i=1}^{12} x_i\right)\left(\sum_{i=1}^{12} y_i\right)}{12\sum_{i=1}^{12} x_i^{\,2} - \left(\sum_{i=1}^{12} x_i\right)^2}$$
$$= \frac{12\cdot 4.827 - 533\cdot 106}{12\cdot 24.319 - 533^2} = 0{,}184$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$
$$= 8{,}8333 - 0{,}184\cdot 44{,}4167$$
$$= 0{,}65$$

Επομένως, η ευθεία παλινδρόμησης είναι:

$$\hat{y} = 0{,}65 + 0{,}184\,x$$

Αν λοιπόν η εκμετάλλευση X της επιχείρησης είχε έσοδα αυξημένα κατά 10 εκατομμύρια, δηλαδή αντί για 50 εκατομμύρια είχε 50+10=60 εκατομμύρια, τότε τα έσοδα της Y θα ήταν:

$$\hat{y} = 0{,}65 + 0{,}184\cdot 60 = 11{,}69 \text{ εκατ. ευρώ.}$$

3. Αν η εκμετάλλευση X είχε μηδενικά έσοδα, δηλαδή X=0, τότε η Y εκμετάλλευση θα είχε έσοδα ίσα με

$$\hat{y} = 0{,}65 + 0{,}184 \cdot 0 = 0{,}65 \text{ εκατομμύρια ευρώ,}$$

δηλαδή, όσο είναι η σταθερά $\hat{a}$.

4. Προφανώς επειδή τα δεδομένα είναι σε εκατομμύρια ευρώ, ο γωνιακός συντελεστής $\hat{b}$ εκφράζει τη μεταβολή που επέρχεται στη *Y* από μία μοναδιαία μεταβολή (1 εκατομμύριο) της *X*. Τότε, η μεταβολή στα έσοδα της *Y*, θα ήταν ίση με το $\hat{b}$ (0,184 εκατομμύρια ευρώ ή 184.000 ευρώ).

5. Η τυπική απόκλιση των τιμών της Χ δίνεται από το γνωστό τύπο:

$$s_x = \sqrt{\frac{\sum_{i=1}^{12} (x - \bar{x})^2}{\nu - 1}}$$

Σχηματίζουμε λοιπόν τον παρακάτω πίνακα:

Πίνακας 36

$(x-\bar{x})$	$(x-\bar{x})^2$	$(y-\bar{y})$	$(y-\bar{y})^2$
-9,41	88,52	-1,8	3,24
-8,41	70,72	-1,8	3,24
-6,41	41,09	-0,8	0,64
-4,41	19,45	-0,8	0,64
-4,41	19,45	-0,8	0,64
-2,41	5,81	-0,3	0,09
-0,41	0,17	-0,3	0,09
0,59	0,35	0,2	0,04
3,59	12,88	1,2	1,44
5,59	31,25	1,2	1,44
10,59	112,15	1,2	1,44
15,59	243,05	3,2	10,24
	$\sum_{i=1}^{12} (x-\bar{x})^2 = 644{,}89$		$\sum_{i=1}^{12} (y-\bar{y})^2 = 23{,}18$

Άρα,

$$s_X = \sqrt{\frac{644,89}{11}} = 7,657$$

6. Από τον ίδιο πίνακα βρίσκουμε:

$$s_Y = \sqrt{\frac{23,18}{11}} = 1.452$$

Παρατηρούμε ότι τα έσοδα της εκμετάλλευσης *Y* παρουσιάζουν πολύ μικρότερη διασπορά από αυτήν που παρουσιάζουν τα έσοδα της εκμετάλλευσης *X*, αφού

$$s_X = 7,657 > s_Y = 1,452.$$

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:

$$CV_X = \frac{s_X}{\bar{x}} = \frac{7,657}{44,4} = 0,172 \text{ ή } 17,2\%.$$

8. Ομοίως και για την περίπτωση της Y, έχουμε:

$$CV_Y = \frac{s_Y}{\bar{y}} = \frac{1,452}{8,83} = 0,164 \text{ ή } 16,4\%.$$

9. Παρατηρούμε ότι οι συντελεστές μεταβλητότητας CV_X και CV_Y είναι πολύ κοντά ο ένας στον άλλον. 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 ικρές 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 του έβδομου όρου. Δηλαδή, η θέση της διαμέσου εντοπίζεται στο

$$\frac{\nu+1}{6}=\frac{12+1}{2}=6{,}5.$$

Ως προς την περίπτωση λοιπόν της X η διάμεσος είναι:

$$\frac{42+44}{2}=43 \text{ εκατ. ευρώ}$$

Δηλαδή το 50% έχουν εκμεταλλεύσεις X με έσοδα από 43 εκατ. ευρώ και άνω.

11. Ομοίως για την εκμετάλλευση Y έχουμε:

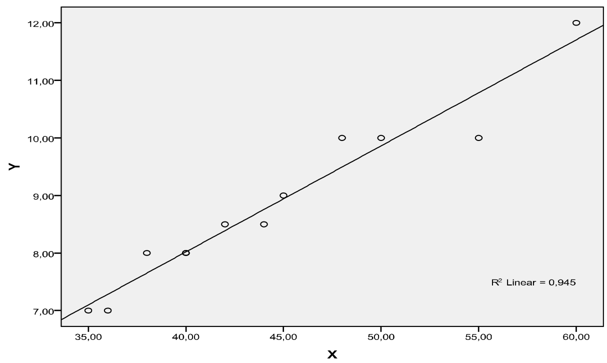
$$\frac{8{,}5+8{,}5}{2}=8{,}5 \text{ εκατ. ευρώ}$$

Δηλαδή οι εκμεταλλεύσεις Y 6 επιχειρήσεων έχουν έσοδα από 8,5 εκατ. και άνω και άλλες 6 έχουν από 8,5 εκατ. και κάτω.

12. Παρατηρώντας προσεκτικά βλέπουμε ότι και για τις δύο εκμεταλλεύσεις ο μέσος και η διάμεσος είναι πάρα πολύ κοντά. Για την εκμετάλλευση X έχω μέσο $\bar{x}=44{,}41$ και διάμεσο 43, και για την εκμετάλλευση Y έχω $\bar{y}=8{,}83$ και διάμεσο 8,5 εκατ. ευρώ. Επίσης, παρατηρούμε ότι και οι δύο εκμεταλλεύσεις της επιχείρησης κατέχουν, ως προς τα έσοδα τους μια καλή θέση, αφού ανήκουν και οι δύο στο 50% των εκμεταλλεύσεων με τα υψηλά έσοδα. Δηλαδή, τα έσοδα της X ήταν 50 εκατ. ευρώ, ενώ το διάμεσο ύψος εσόδων της X είναι 43 εκατ. ευρώ. Για την Y τα έσοδα ήταν 10 εκατ. ευρώ , ενώ το διάμεσο ύψος εσόδων της Y είναι 8,5 εκατ. ευρώ.

13. Διάγραμμα διασποράς των μεταβλητών X και Y.

Διάγραμμα 7



ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην σημερινή εποχή, κάθε φαινόμενο της φύσης και οποιαδήποτε ενέργεια του ανθρώπου αποτελεί πεδίο εφαρμογής της Στατιστικής. Ιδιαίτερα χρήσιμη είναι στις Κοινωνικές, Οικονομικές και Διοικητικές επιστήμες. Οι στατιστικές τεχνικές είναι απαραίτητες για την επιτυχή λειτουργία κάθε επιχείρησης γιατί παρέχουν στην Διοίκηση όλες τις απαραίτητες πληροφορίες που αφορούν την ίδια την επιχείρηση (εσωτερικό περιβάλλον) αλλά και τους ανταγωνιστές (εξωτερικό περιβάλλον). Παρέχει πολλές δυνατότητες, όσον αφορά στον προσδιορισμό της μεταβλητότητας στην αντιμετώπιση, στην πρόβλεψη, στο σχεδιασμό και στην λήψη αποφάσεων ενώ ταυτόχρονα μας εξασφαλίζει κέρδος χρόνου και χρήματος.

Από τις εφαρμογές που έχω παρουσιάσει προκύπτει ότι η Στατιστική είναι απαραίτητο εργαλείο στην σύγχρονη διοίκηση των επιχειρήσεων. Ως οργάνωση και διοίκηση (management) ορίζεται η διαδικασία του προγραμματισμού, της οργάνωσης, της διεύθυνσης και του ελέγχου που ασκούνται σε μια επιχείρηση, προκειμένου να επιτευχθούν αποτελεσματικά οι στόχοι τους.

Η επιστήμη της Διοίκησης Επιχειρήσεων, επειδή τα φαινόμενα που μελετά είναι πολυδιάστατα , έχει σχέση άμεση ή έμμεση με την οικονομική επιστήμη.

Η έννοια της οργάνωσης και διοίκησης συνδέεται άμεσα με την αποτελεσματικότητα της επιχείρησης. Στην πραγματικότητα αποτελεί τη βασική προϋπόθεση για την αποτελεσματική αξιοποίηση των πόρων που διαθέτει μια επιχείρηση και την επίτευξη των στόχων. Η συμβολή της οργάνωσης και διοίκησης στην αποτελεσματικότητα μιας επιχείρησης γίνεται πλήρως κατανοητή, αν σκεφτεί κανείς τα αποτελέσματα που θα υπήρχαν, για παράδειγμα, σε μια βιομηχανία, αν δεν υπήρχε καμία οργάνωση κανένας έλεγχος και κακή διοίκηση των εργαζομένων.

Η άμεση εφαρμογή της Στατιστικής στις επιχειρήσεις σε συνδυασμό με την Πληροφορική, οδήγησε στην ανάπτυξη συστημάτων για την παραγωγή και την οργάνωση λήψεων, περιλήψεων, σχέσεων και γενικεύσεων βασισμένων στα δεδομένα τα οποία έχουν τη δυνατότητα εισαγωγής, επεξεργασίας, ανάλυσης και παρουσίασης μεγάλου όγκου δεδομένων.

Πρέπει όμως να τονισθεί ότι η κακή χρήση της επιστήμης, είτε λόγω άγνοιας, είτε λόγω σκοπιμότητας, μπορεί να οδηγήσει σε λανθασμένα συμπεράσματα και συνεπώς σε λάθος επιλογές.

Αρχικά η χρήση του υπολογιστή και τα πρώτα προγράμματα εφαρμόστηκαν στη μισθοδοσία του προσωπικού και για την πληροφόρηση της διοίκησης, για παράδειγμα υπόλοιπο προμηθευτών, τραπεζών κ.τ.λ. Αργότερα επεκτάθηκε και σε άλλες βασικές δραστηριότητες της επιχείρησης και προέκυψαν έτσι οι Διαχειριστικές Εφαρμογές, που καλύπτουν ζωτικές λειτουργίες και υπολειτουργίες της, όπως Αγορές – Πωλήσεις, Λογιστήριο, Παραγωγή κ.τ.λ.

Οι εφαρμογές αυτές παρέχουν πολλά πλεονεκτήματα, όπως:

1) Μείωση του χρόνου διεκπεραίωσης των διαδικασιών με τη ταυτόχρονη καταχώρηση των στοιχείων στα αρχεία του υπολογιστή και στα εκδιδόμενα παραστατικά κάθε συναλλαγής.

2) Διασφάλιση της ακεραιότητας των στοιχείων από κάθε παρέμβαση.

3) Δυνατότητα χρήσης από την πλευρά των στελεχών των αποτελεσμάτων από την επεξεργασία των στοιχείων, για παράδειγμα διαγράμματα, πίνακες κ.τ.λ. και πολλά άλλα που συμβάλλουν αποφασιστικά στη λήψη σωστών αποφάσεων και στην εξυπηρέτηση του αντικειμενικού σκοπού της επιχείρησης.

Οι επιχειρήσεις χρησιμοποιούν ολοένα και περισσότερο στατιστικές μεθόδους για να μετατρέπουν δεδομένα σε πληροφορίες. Για ένα σπουδαστή που προετοιμάζεται να εισέλθει στο κόσμο των επιχειρήσεων δεν είναι αρκετό απλά να γνωρίζει ένα ετερόκλητο σύνολο από στατιστικές συναρτήσεις και τεχνικές, αλλά να έχει μια ολοκληρωμένη εικόνα των στατιστικών εννοιών και των εφαρμογών τους στον πραγματικό κόσμο. Γίνεται φανερό, ότι οι στατιστικές μέθοδοι αποτελούν ένα κρίσιμο εργαλείο για την σύγχρονη διοίκηση και οικονομία.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Α) Συγγράμματα

1. Αθανασοπούλου, Δ. , *Εισαγωγή στην περιγραφική στατιστική και θεωρία πιθανοτήτων*, Πειραιάς 1972.

2. Αποστολόπουλου, Θ. , *Στατιστική Επιχειρήσεων*, 1991.

3. Γναρδέλης, Χ. , *Εφαρμοσμένη Στατιστική*, Εκδόσεις Παπαζήση, Αθήνα 2003.

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

5. Κιόχος, Α. , Κιόχος, Π. , *Στατιστική για τις Επιχειρήσεις και την Οικονομία*, Εκδόσεις Κιόχου Ε. , 2010.

6. Πανάρετος, Ι. , Ξεκαλάκη, Ε. , *Εισαγωγή στη Στατιστική Σκέψη* (Τόμος 1 Περιγραφική Στατιστική), 1997.

7. Κοίλιας, Χ. , Παπαδήμας, Ο. , *Εφαρμοσμένη Στατιστική*, Εκδόσεις Νέων Τεχνολογιών, Αθήνα 1998.

8. Κάιτσα, Π. , Παπακωσταντίνου, Γ. , *Στατιστική*, Εκδόσεις Ιδρύματος Ευγενιδίου, Αθήνα 1995.

9. Φράγκος, Χ. , *Στατιστική Επιχειρήσεων*, εκδόσεις Σταμούλης Α. , Αθήνα 1998.

Β) Διαδίκτυο

1. Αθανασίου, Α. ,(2010) Ο Ρόλος της Στατιστικής στην επιχειρηματική ανάπτυξη.

Διαθέσιμο στο δικτυακό τόπο:

http://www.lib.teipat.gr/ptyxiakes/sdo/sdo_de/2006-2010/8921

(τελευταία πρόσβαση στις 30/06/2013)

2. Γκλεζάκος, Μ. (2009) Επιχειρήσεις – Οικονομία .

Διαθέσιμο στο δικτυακό τόπο:

http://www.digilib.lib.unipi.gr/gspace/handle/unipi/746

(τελευταία πρόσβαση στις 30/06/2013)

3. Αδαμόπουλος, Λ. , Δαμιανού, Χ. ,Σβέρκος, Α. , (1999) Μαθηματικά και Στοιχεία Στατιστικής.

Διαθέσιμο στο δικτυακό τόπο:

http://www.digitalschools.minedu.gov.gr/courses/DSGL-C100

(τελευταία πρόσβαση στις 30/06/2013)