ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΜΕΣΟΛΟΓΓΙΟΥ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ & ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ

Π Τ Υ Χ Ι Α Κ Η Ε Ρ Γ Α Σ Ι Α

ΣΤΑΤΙΣΤΙΚΑ ΜΕΤΡΑ ΚΑΙ ΟΙ ΕΦΑΡΜΟΓΕΣ ΤΟΥΣ ΣΤΙΣ ΕΠΙΧΕΙΡΗΣΕΙΣ ΚΑΙ ΤΗΝ ΟΙΚΟΝΟΜΙΑ

ΣΩΤΗΡΙΑ ΜΕΝΔΕΛΟΠΟΥΛΟΥ

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

ΣΤΑΤΙΣΤΙΚΑ ΜΕΤΡΑ ΚΑΙ ΟΙ ΕΦΑΡΜΟΓΕΣ ΤΟΥΣ ΣΤΙΣ ΕΠΙΧΕΙΡΗΣΕΙΣ ΚΑΙ ΤΗΝ ΟΙΚΟΝΟΜΙΑ

ΣΩΤΗΡΙΑ ΜΕΝΔΕΛΟΠΟΥΛΟΥ (Α.Μ. 15031)

sotimand@logistiki.teimes.gr

ΕΙΣΗΓΗΤΗΣ:

ΜΕΓΑΡΙΤΗΣ ΑΘΑΝΑΣΙΟΣ

Μ Ε Σ Ο Λ Ο Γ Γ Ι   2 0 1 3

<u>ΕΠΙΣΗΜΑΝΣΗ</u>

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 των σπουδαστών που την επιμελήθηκαν και δεν απηχούν κατ' ανάγκη τη γνώμη του εισηγητή εκπαιδευτικού, του Εκπαιδευτικού Προσωπικού του Τμήματος Λογιστικής ή του Α.Τ.Ε.Ι. Μεσολογγίου.

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΠΙΣΗΜΑΝΣΗ 4

ΠΕΡΙΕΧΟΜΕΝΑ 5-6

ΕΥΧΑΡΙΣΤΙΕΣ 7

ΠΡΟΛΟΓΟΣ..8

ΠΕΡΙΛΗΨΗ…………………………………………………………………..………9

ΚΕΦΑΛΑΙΟ 1ο 10

1. ΕΙΣΑΓΩΓΗ 10-11

1.1. ΕΙΣΑΓΩΓΙΚΟΙ ΟΡΙΣΜΟΙ 10-11

1.1.1. Στατιστική 10-11

1.1.2. Επιχείρηση.. 11

1.1.3. Οικονομία... 12

1.2. Ο ΡΟΛΟΣ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ ΣΤΗΝ ΕΠΙΧΕΙΡΗΣΗ............................ …12-14

1.3. Στατιστικά μέτρα..14-15

ΚΕΦΑΛΑΙΟ 2ο 16-18

2. ΣΤΑΤΙΣΤΙΚΗ.. 16-18

2.1. ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ ΚΑΙ ΕΞΕΛΙΞΗ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ.................... 16-18

2.2. Ο ΡΟΛΟΣ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ ΣΤΗΝ ΟΙΚΟΝΟΜΙΑ.................................. 18-19

2.3. ΤΟ ΚΡΑΤΟΣ ΚΑΙ Η ΣΤΑΤΙΣΤΙΚΗ... 19-20

2.4. ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ……………………………..20-23

2.4.1. Πληθυσμός-μεταβλητές-δείγμα……………………………..……………20-21

2.4.2. Συχνότητες-πίνακες συχνοτήτων …………………………………………….22

2.4.3. Ομαδοποιημένα δεδομένα……………………………………………….23

2.4.4. Στατιστικά διαγράμματα……………………………………………………..23

2.5. ΕΙΔΗ ΔΙΑΓΡΑΜΜΑΤΩΝ 24

2.5.1. Ραβδογράμματα...23-24

2.5.2. Κυκλικά διαγράμματα 24

2.5.3. Χρονολογικά διαγράμματα…………………………………………… .24

2.5.4. Ημιλογαριθμικά και λογαριθμικά διαγράμματα... 25

2.5.5. Τα χαρτογράμματα 25-26

2.5.6. Ιστογράμματα και καμπύλες συχνοτήτων 26

2.6.ΣΤΑΤΙΣΤΙΚΕΣ ΕΚΘΕΣΕΙΣ Η ΑΝΑΦΟΡΕΣ 26-27
2.7.ΔΕΙΓΜΑΤΟΛΗΨΙΑ.. 27-29
ΚΕΦΑΛΑΙΟ 3ο ΣΤΑΤΙΣΤΙΚΑ ΜΕΤΡΑ ΚΕΝΤΡΙΚΗΣ ΤΑΣΗΣ ΚΑΙ ΘΕΣΗΣ………………30
3.1. ΜΕΣΟΙ ΚΕΝΤΡΙΚΗΣ ΤΑΣΗΣ……………………………………………………….30
3.1.1. Ο μέσος αριθμητικός ή μέση τιμή 30-32
3.1.2. Ο μέσος γεωμετρικός .... …..33-35
3.1.3. Ο μέσος αρμονικός... 35-36
3.2. ΠΑΡΑΜΕΤΡΟΙ ΘΕΣΕΩΣ 36
3.2.1. Διάμεσος...37-38
3.2.2. Τεταρτημόρια-δεκατημόρια-εκατοστημόρια 38
Υπολογισμοί τεταρτημόριων..38-39
Υπολογισμοί δεκατημόριων-εκατοστημορίων...39
Επικρατούσα τιμή ή κορυφή 40-41
3.3. ΣΥΓΚΡΙΣΗ ΜΕΣΗΣ ΤΙΜΗΣ, ΔΙΑΜΕΣΟΥ, ΕΠΙΚΡΑΤΟΥΣΑΣ ΤΙΜΗΣ...............41-42
ΚΕΦΑΛΑΙΟ 4ο ΣΤΑΤΙΣΤΙΚΑ ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ……………………………..43-44
4.1.ΤΟ ΕΥΡΟΣ.. 44-45
4.1.1. ΤΟ ΗΜΙΕΝΔΟΤΕΤΑΡΤΗΜΟΡΙΑΚΟ ΕΥΡΟΣ 45
4.1.2. ΜΈΣΗ ΑΠΟΛΥΤΗ ΑΠΟΚΛΙΣΗ...45
4.2. ΔΙΑΚΥΜΑΝΣΗ ……………………………..…………………………46-47
4.3. ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ……………………………………………………47-49
4.4. ΘΕΩΡΗΜΑ CHEBYSHEN…………………………………………………49-50
4.5.ΣΥΝΤΕΛΕΣΤΗΣ ΜΕΤΑΒΛΗΤΙΚΟΤΗΤΑΣ……………………………..….50-51
4.6.ΜΕΣΗ ΔΙΑΦΟΡΑ ΤΟΥ GINI………………………………………………….51
4.7 ΘΗΚΟΓΡΑΜΜΑ……………………………………………………………....51-52
ΚΕΦΑΛΑΙΟ 5ο ΕΦΑΡΜΟΓΕΣ ΤΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΜΕΤΡΩΝ……………………….53-80
ΣΥΜΠΕΡΑΣΜΑΤΑ...81
ΒΙΒΛΙΟΓΡΑΦΙΑ.. 82-83

ΕΥΧΑΡΙΣΤΙΕΣ

Ως φοιτήτρια του ΤΕΙ Μεσολογγίου του Τμήματος Λογιστικής Δυτικής Ελλάδος, θα ήθελα να ευχαριστήσω κάποιους ανθρώπους για την ασύγκριτη, ανεκτίμητη αλλά και ουσιώδης βοήθεια που μου προσέφεραν ώστε να τελειοποιήσω αυτόν τον τόσο μακρύ δρόμο της απονομής του πτυχίου μου. Ευχαριστώ λοιπόν τον αξιοσέβαστο καθηγητή μου κύριο Αθανάσιο Μεγαρίτη για την εξαιρετική συνεργασία που είχαμε αλλά και για την πολύτιμη βοήθεια που μου πρόσφερε για την δημιουργία και ολοκλήρωση της πτυχιακής μου εργασίας. Ήτανε κοντά μου σε κάθε δύσκολο κομμάτι της εργασίας μου και αξίζει να αναφέρω ότι η επικοινωνία μου μαζί του ήτανε συνεπής και άμεση κάθε φορά που τον χρειαζόμουν. Επίσης θα ήθελα να ευχαριστήσω τους γονείς μου και την γιαγιά μου για την οικονομική και συναισθηματική κάλυψη και ασφάλεια που μου έδιναν με όλη τους την αγάπη και αφοσίωση. Τους εκτιμώ πολύ για τον κόπο, την κούραση και όσα στερήθηκαν για να με μορφώσουν και να μου δώσουν τα κατάλληλα και απαραίτητα εφόδια για να αποκατασταθώ πλήρως επαγγελματικά στην μετέπειτα ζωή μου, παίρνοντας ένα καλό πτυχίο σε συνάρτηση με ένα καλύτερο αύριο. Εγώ από την πλευρά μου τους ανταποδίδω με ένα καλό βαθμό πτυχίου, περίπου στο εφτά μισή, νομίζω ότι είναι το καλύτερο που μπορούσα να κάνω μαζί με το πλήρες ενδιαφέρον μου στη σχολή διότι αποφοιτώ ακριβώς στα χρόνια της κανονικής λήψης πτυχίου από το ΤΕΙ Μεσολογγίου.

ΠΡΟΛΟΓΟΣ

Η παρούσα πτυχιακή εργασία πραγματοποιήθηκε στο Τεχνολογικό Εκπαιδευτικό Ίδρυμα Μεσολογγίου, στο τμήμα λογιστικής. Σκοπός της εργασίας είναι να παρουσιάσει τα βασικά στατιστικά μέτρα και τις εφαρμογές τους στις επιχειρήσεις και την οικονομία ως μία ευρεία αλλά και ως μία ειδική έννοια. Θα αναλυθούν όλες οι βασικές έννοιες της στατιστικής, τα μέτρα αναλυτικά μελετημένα, τι γνώσεις είναι ανάγκη να έχουν οι ερευνητές, πως να δρουν σε κάθε περίπτωση, που εφαρμόζεται η στατιστική, τι ρίσκα έχει, πως εξελίσσεται, πόσο αληθείς είναι τα συμπεράσματα της, τι πληροφορίες έχει και από πού τις παίρνει και τέλος πόσο κοστίζει μία έρευνα.

ΠΕΡΙΛΗΨΗ

Στη συγκεκριμένη πτυχιακή εργασία με θέμα: στατιστικά μέτρα και οι εφαρμογές τους στις επιχειρήσεις και την οικονομία, έχω αναπτύξει το θέμα σε πέντε κεφάλαια. Τα δύο πρώτα αποτελούν το θεωρητικό μέρος της εργασίας ενώ τα εναπομείναντα τρία περιλαμβάνουν το πρακτικό μέρος της. Το πρώτο κεφάλαιο περιέχει τις εισαγωγικές έννοιες, τον ορισμό της στατιστικής, της επιχείρησης και της οικονομίας και τέλος τα στατιστικά μέτρα. Είναι κυρίως εισαγωγικό κεφάλαιο. Το δεύτερο κεφάλαιο αναφέρει τον ορισμό της στατιστικής σε πιο αναλυτική μορφή, την ιστορία, το ξεκίνημα και την εξέλιξη της από παλαιότερη εποχή. Επίσης στο κεφάλαιο αυτό δίνονται τα σημαντικότερα ονόματα των επιστημόνων που συντέλεσαν στη διαδοχική και εξελικτική πορεία της στατιστικής. Επιπροσθέτως, επικεντρώνεται στο ποιος είναι ο ρόλος της στατιστικής στην οικονομία, στη σχέση της με το κράτος, στις βασικές της έννοιες και στο τέλος του κεφαλαίου γράφονται τα είδη διαγραμμάτων, οι στατιστικές εκθέσεις ή αναφορές και η δειγματοληψία. Στο τρίτο κεφάλαιο αναλύονται μεθοδικά τα μέτρα κεντρικής τάσης και θέσης αφού προηγουμένως έχουν αναφερθεί οι έννοιες τους. Το κεφάλαιο αυτό αρχίζει με τους μέσους κεντρικής τάσης και τελειώνει με τις παραμέτρους θέσεως. Στο επόμενο κεφάλαιο, δηλαδή στο τέταρτο δίνονται με παραδείγματα τα στατιστικά μέτρα διασποράς. Επίσης μέσα υπάρχει το θεώρημα του Chebyshev και η μέση διαφορά του Gini. Στο τελευταίο κεφάλαιο δηλαδή το πέμπτο, φτάνουμε πια στο άμεσα πρακτικό κομμάτι της εργασίας. Στην ουσία αναλύονται διεξοδικά με εφαρμογές όλα τα προηγούμενα κεφάλαια. Αυτό το κεφάλαιο περιορίζεται μόνο σε εφαρμογές και ερμηνεύει τους τύπους που χρησιμοποιήθηκαν στις εφαρμογές αυτές. Αφορά περισσότερο το δεύτερο σκέλος του θέματος της πτυχιακής εργασίας και ελάχιστα το πρώτο σκέλος το οποίο κατά κύριο λόγο είναι θεωρητικό.

ΚΕΦΑΛΑΙΟ 1^{o}

ΕΙΣΑΓΩΓΗ

1.1. ΕΙΣΑΓΩΓΙΚΟΙ ΟΡΙΣΜΟΙ

1.1.1. Στατιστική

Η Στατιστική[1] είναι η συστηματική απαρίθμηση και παρουσίαση αριθμητικών δεδομένων ή στοιχείων τα οποία προέρχονται από πολλές παρατηρήσεις ή μετρήσεις και αναφέρονται σε ιδιότητες ή χαρακτηριστικά συγκεκριμένων αντικειμένων και γεγονότων που μετράμε και μελετάμε. Ανάλογα με το αντικείμενο και το γεγονός στο οποίο αναφέρονται αριθμητικά δεδομένα ή στατιστικά στοιχεία η στατιστική παίρνει μια ειδική ονομασία[2].

Είναι επιστήμη και έχει ευρεία έννοια. Είναι η επιστήμη που ασχολείται με συγκέντρωση, επεξεργασία, παρουσίαση, μελέτη και ανάλυση των παρατηρήσεων και μετρήσεων που αναφέρονται σε χαρακτηριστικές ιδιότητες ενός συγκεκριμένου αντικειμένου ή γεγονότος όποια και αν είναι η φύση του. Περιλαμβάνει τις μεθόδους συλλογής, οργάνωσης, επεξεργασίας και παρουσίασης των συγκεντρωθέντων στοιχείων όσο και τις μεθόδους ανάλυσης και ερμηνείας αριθμητικών δεδομένων για την εξαγωγή λογικών και τεκμηριωμένο επιχειρηματολογία συμπερασμάτων που θα βοηθούσαν στην ορθή και σωστή λήψη αποφάσεων και διαδικασιών. Στατιστική είναι η επιστήμη που ασχολείται με τις επιστημονικές μεθόδους συλλογής, επεξεργασίας, παρουσίασης, ανάλυσης και ερμηνείας αριθμητικών δεδομένων, ώστε να καταλήξουμε στη διατύπωση συμπερασμάτων τα οποία είναι χρήσιμα στη επιλογή μιας απόφασης στάδια της είναι διαφόρων μονάδων μιας μεγάλης ομάδας ή άψυχων αντικειμένων για να ληφθέν μια καλή και ορθή απόφαση. Αυτά είναι:

⇒ Συγκέντρωση των αναγκαίων στοιχείων.

---

[1] Αποστολόπουλος Θ., «Περιγραφική Στατιστική Επιχειρήσεων», σελ. 19-21.

[2] Υπάρχει η επιχειρηματική στατιστική, η οικονομική, η γεωργική, η δημογραφική (η καταγραφή του ανθρώπινου πληθυσμού μετά από κάποια χρόνια), ιατρική, εργατικού δυναμικού κλπ, εννοούμε αριθμητικά δεδομένα που αναφέρονται σε αυτούς τους κλάδους και τομείς.

⇒ Μεθοδική επεξεργασία και παρουσίαση των στατιστικών στοιχείων με αριθμητικούς πίνακες και γραφικές παραστάσεις.

⇒ Ανάλυση των αριθμητικών δεδομένων για εξαγωγή τεκμηριωμένων συμπερασμάτων.

Η στατιστική γίνεται για την μελέτη και πρόβλεψη σεισμών, ιατρικές και βιολογικές έρευνες, ανακαλύψεις κλπ.

1.1.2. Επιχείρηση

Επιχείρηση είναι μία οικονομική μονάδα η οποία αποκομίζει χρηματικούς πόρους και παραγωγικούς συντελεστές από εξεζητημένες χρηματοδοτικές πηγές πχ προμηθευτές, τράπεζες, πιστωτικά ιδρύματα με σκοπό να συντονίσει και να συνδυάσει τους οικονομικούς, παραγωγικούς συντελεστές για να πετύχει το μεγαλύτερο και μέγιστο δυνατό κέρδος. Είναι δύσκολη η εύρεση των παραγωγικών συντελεστών γιατί είναι σε στενότητα. Παραγωγικοί συντελεστές είναι τα χρήματα, η γη, η εργασία, τα μηχανήματα, τα κτίρια, τα μεταφορικά μέσα, η τεχνολογία και τεχνογνωσία είναι μερικά από αυτά. Οι κατηγορίες των επιχειρήσεων είναι:

⇒ Οι ιδιωτικές,

⇒ Οι δημόσιες και

⇒ Οι μικτές.

Οι επιχειρήσεις έχουν νομική προσωπικότητα αυτές είναι:

➢ Οι ΑΕ, ΕΠΕ (ανώνυμες εταιρίες, εταιρίες περιορισμένης ευθύνης) κεφαλαιουχικές

➢ ΟΕ, ΕΕ (ομόρρυθμες και ετερόρρυθμες εταιρίες) προσωπικές

➢ αφανής

➢ συμπλοιοκτησία

➢ Ναυτιλιακές εταιρείες

➢ Συνεταιρισμοί.

Οι επιχειρήσεις ως προς το μέγεθος τους:

✓ Μικρές

✓ Μεσαίες

✓ Μεγάλες

✓ Κολοσσοί.

1.1.3. Οικονομία

Η οικονομία είναι μία κοινωνική και οικονομική επιστήμη η οποία διαχειρίζεται τα οικονομικά αλλά και τις ανθρώπινες και καταναλωτικές ανάγκες. Η οικονομία μπορεί να είναι είτε στην καθημερινότητα μας ως προς τα νοικοκυριά τα χρήματα που έχουμε είτε σε κράτος, χώρα, πόλη, περιοχή δηλαδή η οικονομική της κατάσταση και ανάπτυξη ή απαξίωση ακόμα η οικονομία μπορεί να αφορά και επιχειρήσεις στο εσωτερικό τους, η λειτουργία της, ο σκοπός της και η ύπαρξη της αλληλοεπηρεάζονται και αλληλεπιδρούν με την οικονομία. Η επιστήμη αυτή μελετάει και ασχολείται με τα ζωή ενός αγαθού από την ύπαρξη του, τη γέννηση του μέχρι και την απαξίωση και απαθλίωση του. Οι ανθρώπινες ανάγκες είναι βασικό συστατικό της οικονομικής ευημερίας. Η οικονομία συνδυάζει και σχετίζει τη διαδικασία της διανομής, της επιβίωσης, της λειτουργίας και ανάπτυξης της αλλά και του καταναλωτισμού. Η καταναλωτική συμπεριφορά είναι το σημείο αναφοράς της, χωρίς αυτή δεν υπάρχει η οικονομική επιστήμη οικονομία χωρίζεται σε δύο κλάδους:

- Μικροοικονομία και
- Μακροοικονομία.

Η μικροοικονομία είναι η επιστήμη που ασχολείται με τη συνολική μελέτη και ανάλυση των επιχειρήσεων, νοικοκυριών και ατόμων για τη σωστή λήψη αποφάσεων, λειτουργίας και διαδικασίας της οικονομίας τόσο στις επιχειρήσεις, τα νοικοκυριά και το κράτος. Η μακροοικονομία ασχολείται με τη μία οικονομική μονάδα, την επιχείρηση, το νοικοκυριό και το άτομο. Προσεγγίζει σαν ατομική μονάδα τη μελέτη και ανάλυση της οικονομίας για τη διεξαγωγή συμπερασμάτων αλλά και λύσεων για την ήδη υπάρχουσα κατάσταση της οικονομίας αλλά και την μελλοντική της άνοδο και ανάπτυξη σε περίπτωση καθοδικής πορείας της οικονομίας μιας χώρας.

1.2. Ο ΡΟΛΟΣ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ ΣΤΗΝ ΕΠΙΧΕΙΡΗΣΗ

Εφαρμόζεται στον κλάδο των εμπορικών και βιομηχανικών επιχειρήσεων. Η έρευνα αγοράς γίνεται με δειγματοληψία για την καλύτερη δυνατή μελέτη του τι θέλει, τι ζητάει ο καταναλωτής τη δεδομένη περίοδο σύμφωνα με τις ανάγκες του. Μέσα στις μεγάλες επιχειρήσεις υπάρχουνε τμήματα στατιστικά που έχουν αντικείμενο της ασχολίας τους τις προβλέψεις των εισροών και εκροών, των επιχειρηματικών, οικονομικών και παραγωγικών συντελεστών και γενικά την όλη οικονομική κατάσταση την επιχείρησης. Κάθε επιχείρηση έχει κάποιο τμήμα που ασχολείται με τα στατιστικά στοιχεία των εργαζομένων της επιχείρησης πχ ηλικία, φύλο, αποδοτικότητα, ώρες εργασίας, παραγωγικότητα, αμοιβές, επίπεδο μόρφωσης, ασθένειες, ομάδες, συνεργασιμότητα κ.α. Παρακολουθούνται και οι κίνδυνοι που έχει η επιχείρηση ίσως ζημιές, καθόλου παραγωγικότητα, αποδοτικότητα, ρευστότητα, διάθεση κεφαλαίου ή αλλιώς υπόλοιπο, τι χρειάζεται από χρήματα, κεφάλαιο, περιουσιακά στοιχεία ενεργητικού, α' ύλες, πόσος είναι ο ανταγωνισμός, ποιότητα παραγόμενων αγαθών, τιμές, κόστος παραγωγής, δείκτες κεφαλαίου, προβλέψεις κ.α. Προσδιορίζει την προσοχή ή την άμεση αλλαγή αν κάτι δεν πάει καλά.

Τα επιχειρηματικά σχέδια ξεκινάνε με στατιστικά δεδομένα πληροφορίες και στοιχεία και επηρεάζουν την τελική απόφαση. Αν θέλει μία επιχείρηση να πάρει δάνειο η τράπεζα ελέγχει τα στατιστικά της στοιχεία για να το δώσει έτσι θα ξέρει εάν θα είναι σε θέσει να εξοφλήσει και να αποπληρώσει το δάνειο που έχει πάρει από αυτή. Αν η επιχείρηση δεν έχει στατιστικά στοιχεία τότε η τράπεζα δεν θα εγκρίνει το δάνειο και δεν θα της το παρέχει.

Ο στατιστικός ποιοτικός έλεγχος έχει σαν στόχο να εμποδίσει από την παραγωγή αποκλειόμενα και ελαττωματικά προϊόντα στη διαδικασία της παραγωγής και όχι στο τελικό προϊόν. Στην παραγωγή μπορεί να διορθώσει ή να ρυθμίσει γρήγορα την παραγωγική διαδικασία οπότε την παραγωγή κακών προϊόντων, αυτός είναι και ο πιο σημαντικός και ο πιο βαρύτατος στόχος του να εντοπίσει τους λόγους απόκλισης από τις προδιαγραφές, να τους απομονώσει και να τους μειώσει η να τους εξαφανίσει. Ας μην ξεχάσουμε να αναφέρουμε ότι η στατιστική πρωτοεμφανίστηκε στις ΗΠΑ Ηνωμένες Πολιτείες Αμερικής στις αρχές του 20^{ού} αιώνα ως συνέπεια του μεγάλου όγκου των εργαζομένων αλλά και των πολλών εργασιών και ασχολιών που ήθελε μια μεγάλη επιχείρηση. Αυτοί ήτανε οι προαναφερόμενοι ρόλοι της στατιστικής των επιχειρήσεων.

Η στατιστική είναι ένα "σπρώξιμο" για τις επιχειρήσεις και την οικονομία για να ελιχθούν και να επιβιώσουν. Το πιο περίφημο όμως είναι ότι η στατιστική εφαρμόζεται και σε άλλους κλάδους όπως η ιατρική, στη γεωργία, στην βιομηχανία, στη βιοτεχνία, στο εμπόριο, στην οικογενειακή κατάσταση, στο βάρος, στο φύλο, στην ηλικία, στο ανάστημα, στο χρώμα, στο ρατσισμό, στις καιρικές συνθήκες και φαινόμενα, στις προβλέψεις σεισμών, σε έμψυχων και άψυχων αντικειμένων, σε βαθμούς μαθητών, σε εισαγωγές φοιτητών σε πανεπιστήμια, σε μόρια πανελληνίων, σε αρρώστιες ή ασθένειες, σε νομισματική ισοτιμία, σε γενικό επίπεδο τιμών, σε εθνικό εισόδημα, σε οικονομικές διακυμάνσεις, σε απασχόληση, σε παραγωγικότητα, σε κατάρτιση δεικτών οικονομικής δραστηριότητας, σε εθνικές δαπάνες, σε εθνικούς πόρους, της μετανάστευσης κ.α. Καλό θα ήταν να αναφέρουμε περιληπτικά και κάποια από τα στάδια της:

⇒ Συλλογή: με κάποιες μεθόδους και τεχνικές.

⇒ Επεξεργασία και παρουσίαση: συγκεντρώνουμε τα στοιχεία και τα δεδομένα που έχουμε βρει και τα βάζουμε σε τάξη και σειρά. Η παρουσίαση γίνεται με πίνακες, χρονοδιαγράμματα, πίτες, ραβδογράμματα, με άλλα διαγράμματα κ.α.

⇒ Ανάλυση: Είναι ο υπολογισμός των απαραίτητων στοιχείων και γεγονότων που έχουμε μαζέψει και βοηθούν στη μελέτη της συμπεριφοράς των δημογραφικών, κοινωνικών, οικονομικών, επιχειρησιακών κα συνθηκών που θέλουμε να αναλύσουμε και να μελετήσουμε.

⇒ Ερμηνεία: Σκοπός αυτών είναι η εφαρμογή και πρακτική μιας έρευνας και γι' αυτό χρησιμοποιούμε στατιστικά μέτρα και εργαλεία για να πάρουνε αποφάσεις, για κοινωνικές, οικονομικές, επιχειρησιακές πολιτικές οι οποίες είναι ορθές λειτουργικά και πρακτικά αλλά και για προγραμματισμό[3].

1.3. ΣΤΑΤΙΣΤΙΚΑ ΜΕΤΡΑ

Το πρώτο και βασικό στάδιο της στατιστικής ανάλυσης ενός πληθυσμού ή δείγματος από τη γνώμη μιας ιδιότητας, ύστερα από τη συλλογή των στατιστικών στοιχείων, είναι η εμφάνιση και η ταξινόμηση των πολυάριθμων παρατηρήσεων σε μορφή κατανομής συχνοτήτων ή πινάκων.

Στόχος της εικόνας των στατιστικών δεδομένων σε μορφή περιληπτικών ή συνοπτικών πινάκων συχνοτήτων είναι ο περιορισμός του όγκου των μετρήσεων ή παρατηρήσεων που έχουν συγκεντρωθεί και μαζευτεί, και η εύκολη μελέτη, ανάλυση και περιγραφή του περιεχομένου και της δομής του πληθυσμού ή του δείγματος που μας

[3] Τα στατιστικά δεδομένα που βγαίνουν από τη μελέτη φαινομένων ονομάζονται στατιστικά δεδομένα ή στατιστικά στοιχεία ή παρατηρήσεις.

ενδιαφέρει και ερευνούμε. Αλλά και με τη μορφή αυτή των κατανομών συχνοτήτων (πινάκων), τα αρχικά επεξεργασθέντα και συγκεντρωθέντα στατιστικά στοιχεία συνεχίζουν να έχουν μία σύνθετη έκφραση γιατί οι πίνακες συχνοτήτων δεν μπορούν να μείνουν και να συγκρατηθούν στη μνήμη του αναγνώστη και να διευκολύνουν τις συγκρίσεις των διάφορων κατανομών συχνοτήτων.

Γι’ αυτό το σκοπό συμπυκνώνονται και συγκεφαλαιώνονται. Αυτό γίνεται με την αντικατάσταση των πινάκων συχνοτήτων ή του συνόλου των δεδομένων σε κάποιους ορισμένους αντιπροσωπευτικούς αριθμούς οι οποίοι ονομάζονται στατιστικοί παράμετροι, που προσδιορίζουν την:

- Κεντρική τάση,
- Τη θέση,
- Τη διασπορά των παρατηρήσεων και
- Τη μορφολογία του πληθυσμού.

ΚΕΦΑΛΑΙΟ 2^{ο}

ΣΤΑΤΙΣΤΙΚΗ

2.1. ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ ΚΑΙ ΕΞΕΛΙΞΗ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ

Η λέξη στατιστική προέρχεται από τη λατινική λέξη status που σημαίνει κράτος. Ο περίφημος καθηγητής της στατιστικής στην Α.Σ.Ο.Ε.Ε Κωνσταντίνος Αθανασιάδης πιθανολογεί την προέλευση της λατινικής λέξεως status από την ελληνική λέξη στατίζειν, η οποία σημαίνει δημιουργεί και διατυπώνειν. Ο Achenwall με τη λέξη statistika εννοούσε τη συλλογή αριθμητικών δεδομένων τα οποία θεωρούνται κρατικά απόρρητα για τις κρατικές ανάγκες (πληθυσμός, έκταση, στρατιωτική δύναμη, παραγωγή κα). Αργότερα άρχισε να εισάγεται και σε όλες τις γλώσσες η λέξη στατιστική. Δηλώνει αρχικά συλλογή στοιχείων και κρατικές ανάγκες (έκταση, παραγωγή, πληθυσμό κλπ.). Έχει εξακριβωθεί ότι η πρώτη απογραφή πληθυσμού έγινε στην Κίνα από τον αυτοκράτορα Υάο το έτος 2238 π.Χ ενώ στους Ρωμαίους η πρώτη απογραφή πληθυσμού έγινε επί Ρωμύλου (753-715 π.Χ) και η τελευταία από τον αυτοκράτορα Βεσπιανό το 73 μ.Χ. Στην Αγγλία η πρώτη καθολική απογραφή του πλούτου και του πληθυσμού γενικά έγινε το 1085 από το Γουλιέλμο τον κατακτητή.

Το 1583 γράφεται από τον Nr. Sansovino το πρώτο βιβλίο στατιστικού περιεχομένου και λίγο αργότερα εισάγεται από τον Konring (1606-1681) η στατιστική στην ανώτερη παιδεία. Την ίδια εποχή εμφανίζεται το ενδιαφέρον για τις ασφάλειες ζωής και ο περίφημος άγγλος αστρονόμος E.Halley (1693), χρησιμοποιώντας τα ληξιαρχικά βιβλία γεννήσεως και θανάτων της πόλεως Breslaou, παρουσιάζει τον πρώτο πίνακα θνησιμότητας. Το ρεύμα αυτών των δημογραφικών μελετών επεκτείνεται και στη Γερμανία, όπου ο πράκτορας Siissmilvh (1707-1767) συγκεντρώνει ληξιαρχικά βιβλία των εφημερίδων της Πρωσίας και καταλήγει το 1741, στο συμπέρασμα ότι το ποσοστό γέννησης των αγορίων είναι 51% και των κοριτσιών είναι 49%, ενώ τα δύο φύλα έχουν ίσα ποσοστά κατά την εποχή του γάμου. Για το συγγραφέα το φαινόμενο αυτό δεν είναι τυχαίο γεγονός, αλλά νόμος θείας προέλευσης που αποσκοπεί στη διαιώνιση του είδους. Μέχρι την εποχή αυτή, η στατιστική έχει περιγραφικό χαρακτήρα και ασχολείται κυρίως με θέματα δημογραφίας.

Η στατιστική θα ξεφύγει από τον περιγραφικό χαρακτήρα με ανάπτυξη ενός νέου κλάδου, του λογισμού των πιθανοτήτων, ο οποίος προήλθε από την μελέτη τυχερών παιχνιδιών (χαρακτηριστική μάλιστα είναι η αλληλογραφία ανάμεσα στους Γάλλους μαθηματικούς Pascal και Nermat, με αφορμή τα ερωτήματα που έθεσε ο Pascal ο ιππότης De Mere για τα παιχνίδια κύβου). Από τους θεμελιωτές του λογισμού των πιθανοτήτων αναφέρουμε τον Bernoulli, ο οποίος στο βιβλίο του «Η τέχνη των προβλέψεων» διατυπώνει τον περίφημο νόμο των μεγάλων αριθμών, και το Γάλλο μαθηματικό Laplace, στον οποίο οφείλεται η εφαρμογή στο νόμο του λογισμού των πιθανοτήτων στην σπουδή των φυσικών φαινομένων με πολυσύνθετες αιτίες. Στη νέα αυτή περίοδο της στατιστικής ο Βέλγος αστρονόμος Quetelet επεκτείνει την εφαρμογή της στατιστικής στη σπουδή της φυσικών, διανοητικών και ηθικών ιδιοτήτων του ανθρώπου και παίρνει την πρωτοβουλία για τη σύγκλιση του διεθνούς συνεδρίου στατιστικής που έγινε στις Βρυξέλλες το 1853, ενώ αργότερα ο N. Galton εφαρμόζει τη στατιστική στη Βιολογία και, ειδικότερα στα προβλήματα της κληρονομικότητας. Η προσπάθεια του Galton συνεχίστηκε από τον Άγγλο μαθηματικό Karl Pearson, στους οποίους οφείλεται κατά πολύ η σημερινή ανάπτυξη και θέση της στατιστικής. Ως αρχή της στατιστικής (συλλογή αριθμητικών δεδομένων) θεωρούνται οι απογραφές του πληθυσμού στη αρχαιότητα. Οι Αιγύπτιοι (3500 π.Χ), οι Κινέζοι (2300 π.Χ), οι Ασσύριοι, οι Βαβυλώνιοι, οι Χαλδαίοι άφησαν αποδείξεις ότι συγκέντρωναν στατιστικά δεδομένα. Η Αγία Γραφή (κατά Λουκά Ευαγγέλιο, Κεφ. Β') αναφέρει ότι ο Ιωσήφ και η Μαρία πήγαν στη Βηθλεέμ, γιατί επρόκειτο να γίνει η απογραφή του πληθυσμού από τους Ρωμαίους την εποχή που γεννήθηκε ο Χριστός. Στατιστικά στοιχεία για:

⇒ τον πληθυσμό,
⇒ τη φορολογία,
⇒ το εισόδημα,
⇒ τις στρατιωτικές δυνάμεις,

αναφέρονται και γράφονται και στα έργα του Ηρόδοτου, του Θουκυδίδη, του Ξενοφώντα και του Αριστοτέλη.

Η στατιστική συστηματοποιείται και γίνεται αυτοτελής επιστήμη στα μέσα του 17^{ου} αιώνα από τον Conring (1600-1681), ο οποίος καθιέρωσε και τη στατιστική στην ανώτατη παιδεία. Την εποχή αυτή η στατιστική έχει να κάνει με δημογραφικά στοιχεία και γεγονότα μετρημένα σε διάφορα μεγέθη. Πρωτοπόροι και θεμελιωτές της στατιστικής είναι οι Άγγλοι William Petty, John Graunt και ο Γερμανός πάστορας και καθηγητής Sussmilch, οι οποίοι έγραψαν σοβαρές και από τις πιο καλές δημογραφικές μελέτες. Αξίζει να αποτυπωθεί ότι από τα τέλη του 17^{ου} αιώνα αφού η στατιστική έχει περάσει από πολλά εξελικτικά στάδια φτάνει και προάγεται στην πλήρη μαθηματικοποίηση από τις μαθηματικές ιδιοφυίες που είναι:

- Το Γάλλο Laplace (1749-1827),
- Το Γερμανό Gauss (1777-1855),
- Των Γάλλων Poisson,
- Quetelet,
- Cournot
- Laspeyeres.

Από την εποχή του Cournot και του Nrancis Galton (1886) αρχίζει η συνεχής μαθηματικοποίηση της στατιστικής. Συνεχιστές του Galton υπήρξαν οι Άγγλοι Karl Pearson και R.A Nisher, οι οποίοι θεωρούνται οι θεμελιωτές της νεότερης στατιστικής.

2.2. Ο ΡΟΛΟΣ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ ΣΤΗΝ ΟΙΚΟΝΟΜΙΑ

Ο ρόλος που παίζει η στατιστική στην οικονομία είναι αρκετά ενδιαφέρον και σημαντικός. Από την αρχαιότητα επηρεαζόταν η οικονομία από τη στατιστική. Η στατιστική καθορίζει στην οικονομία τα εξής:

- Τον πληθωρισμό,
- Τη γενική αυξομείωση τιμών
- Το χρηματιστήριο
- Τον προϋπολογισμό
- Την σύγκριση μιας οικονομίας με μία άλλη
- Πόσο είναι η διαφορά την μιας οικονομίας από την άλλη

Για το χρηματιστήριο η στατιστική κάνει τα εξής:

- προβλέπει πόσο θα αυξηθούν οι μετοχές, τα κεφάλαια και τα ομόλογα,
- προβλέπει πόσο θα μειωθούν τα κεφάλαια, οι μετοχές και τα ομόλογα,
- προβλέπει ποιά θα είναι η ισοτιμία των νομισμάτων κάθε χώρας,
- προβλέπει αν πρέπει να δανείσει ο πιστωτής σε αγοραστές,
- προβλέπει αν οι μετοχές, τα ομόλογα και τα κεφάλαια θα μείνουν στην αξία που είναι,
- υπολογίζει πόση διαφορά είχε μία μετοχή πριν και μετά την αύξηση ή την μείωση που έχει,
- προβλέπει αν πρέπει να επενδύσουμε και γιατί.

Επίσης συμβάλλει η στατιστική σημαντικά σε θέματα αγορών και επενδύσεων, καθώς αξιολογεί κάποιες σοβαρές αξίες. Με αυτόν τον τρόπο αποφασίζουν οι αγοραστές εάν θα αγοράσουν πχ κτίρια, οικόπεδα, σπίτια κ.α. Άρα προβλέπεται μία πιθανή αγορά αλλά και η εξέλιξη αυτής μετά από κάποιο χρονικό διάστημα, για να ξέρει ο αγοραστής αν η αγορά του ήτανε επιτυχής.

2.3. ΤΟ ΚΡΑΤΟΣ ΚΑΙ Η ΣΤΑΤΙΣΤΙΚΗ

Ένα κράτος, μία χώρα είναι αναγκαίο να ξέρει:

- τον πληθυσμό,
- το φύλο,
- την ηλικία,
- το επάγγελμα,
- τον τουρισμό,
- τους μετανάστες,
- την εκπαίδευση,
- την οικοδομική δραστηριότητα,
- τις επιχειρήσεις,
- τη βιοτεχνία,
- τη βιομηχανία,
- το εμπόριο
- τις εδαφικές εκτάσεις,
- τα κτίρια,

- την οικονομία της χώρας (πληθυσμό, δημόσιο χρέος, ανεργία, ΑΕΠ, δείκτες τιμών, ελλείμματα, του κρατικού προϋπολογισμού και επιπτώσεις στο δημόσιο χρέος, οικονομικές και κοινωνικές, συνέπεια της ανεργίας, πιστωτική πολιτική, δημόσιες δαπάνες κα),
- όσο και τα διοικητικά και κοινωνικά φαινόμενα της χώρας (διοίκηση, ασφάλεια, πρόνοια, υγεία, εργασία, εγκληματικότητα, δικαιοσύνη).

Για όλους αυτούς τους λόγους το κράτος έχει δικό του μεμονωμένο και ξεχωριστό τμήμα στατιστικής το οποίο είναι μία ολόκληρη κολοσσιαία υπηρεσία, η οποία έχει πλούσιο στατιστικό σύστημα με κεντρικό αρχείο στατιστικών στοιχείων, από το οποίο λαμβάνει και παίρνει τις απαραίτητες και χρήσιμες πληροφορίες κάθε υψηλό στέλεχος του κράτους, κάθε ερευνητής και κάθε ενδιαφερόμενος. Με τη γνησιότητα των στατιστικών στοιχείων δεν πρέπει να εμπλέκονται πολιτικά και άλλα συμφέροντα ή σκοπιμότητες γι' αυτό στη μέση μπαίνει η Ευρωπαϊκή Ένωση η οποία επιβάλλει οικονομικές και κοινωνικές επιβαρύνσεις και κυρώσεις σε αυτούς που θα καταχραστούν τα στατιστικά στοιχεία.

2.4. ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ

2.4.1. Πληθυσμός – Μεταβλητές – Δείγμα

Ένα σύνολο του οποίου τα στοιχεία μελετάμε ως προς κάποιο χαρακτηριστικό τους, λέγεται πληθυσμός. Ο πληθυσμός είναι αυτό που θέλουμε να μελετήσουμε και να αναλύσουμε. Είναι ένα σύνολο που μας ενδιαφέρει για να μελετήσουμε. Χρησιμοποιείται και ο όρος ολότητα. Τέτοια σύνολα μπορεί να είναι εκλογική ομάδα, φοιτητές, ετήσια αλιεύματα μιας περιοχής ή χώρας, η ημερήσια παραγωγή οχημάτων, αυτοκινήτων μιας αυτοκινητοβιομηχανίας. Ο Στατιστικός πληθυσμός μπορεί επίσης να είναι ένα σύνολο επαναλαμβανόμενων μετρήσεων που έχουν προηγηθεί με παρατήρηση ή πείραμα, ή τιμών ή σχέσεων και ιδιοτήτων που αποκαλύπτονται στο στάδιο της στατιστικής επεξεργασίας τους. Κάθε στατιστικός πληθυσμός αποτελείται από στατιστικές μονάδες. Μερικές φορές είναι πολυπληθής με αποτέλεσμα να επιμερίζεται. Οπότε έχουμε το στατιστικό υποπληθυσμό ή στατιστικό υποσύνολο.

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

Τις μεταβλητές τις διακρίνουμε σε:

1) Ποιοτικές
2) Ποσοτικές

Οι ποιοτικές μεταβλητές είναι αυτές που σχετίζονται με κάποιο ποιοτικό χαρακτηριστικό όπως η ηλικία, το φύλο, η περιοχή καταγωγής, το επίπεδο μόρφωσης, το επίπεδο σχολείου των παιδιών, το επίπεδο σε ψυχολογίας, το επίπεδο σε χρήμα που έχει ο κάθε άνθρωπος, πλούσιος, πολύ πλούσιος, πάμπλουτος, φτωχό πολύ φτωχός, πάρα πολύ φτωχός.

Άλλες είναι οι ποσοτικές, των οποίων οι τιμές έχουν αριθμητικές ιδιότητες και εκφράζονται με μία μονάδα μέτρησης παράδειγμα το εισόδημα, το βάρος, ο αριθμός των παιδιών μιας οικογένειας, κάποιου σχολείου, το ύψος, το ελάττωμα σε κάποιο προϊόν, αγαθό.

Τι ποσοτικές μεταβλητές τις διακρίνουμε σε:

1) Διακριτές
2) Συνεχείς

Το σύνολο των δυνατών τιμών των διακριτών μεταβλητών είναι υποσύνολο των φυσικών αριθμών όπως ο αριθμός τροχαίων, αεροπορικών, ναυτιλιακών ατυχημάτων, ο αριθμός επιβατών σε αεροπορική πτήση, σε λεωφορεία, σε τραίνα και μετρό. Είναι αυτές που μπορούν να πάρουν πεπερασμένο και αριθμήσιμο πλήθος τιμών.

Το σύνολο των δυνατών τιμών των συνεχών μεταβλητών είναι ένα συνεχές υποσύνολο των πραγματικών αριθμών όπως το βάρος, η ηλικία, η πίεση κ.α.

Δείγμα είναι ένα υποσύνολο ενός πληθυσμού ή παρατηρηθέντων αποτελεσμάτων μιας διαδικασίας για μια χρονική περίοδο. Τα στοιχεία ενός δείγματος ονομάζονται δειγματικές ή δειγματοληπτικές μονάδες.

2.4.2. Συχνότητες – Πίνακες συχνοτήτων

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

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

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

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

Για τις σχετικές συχνότητες ισχύουν:

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

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

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

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

2.4.3. Ομαδοποιημένα δεδομένα

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

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

2.4.4. Στατιστικά διαγράμματα

Η στατιστική χρησιμοποιεί τις γραφικές παραστάσεις για να δώσει σε αφηρημένους αριθμούς μια συγκεκριμένη μορφή για να μας δώσει να καταλάβουμε άμεσα και καλύτερα μέσω γεωμετρικού σχήματος το φαινόμενο που θέλουμε να μελετήσουμε. Πρέπει όμως να είναι ιδιαίτερα καλά σχεδιασμένη έτσι ώστε να μένει στην μνήμη του ερευνητή - αναγνώστη αλλά και για να είναι πιο ευκολονόητη και ευκολοδιάβαστη από έναν αριθμητικό πίνακα. Επίσης πρέπει να μην διαθέτει τόσες πολλές λεπτομέρειες για να μην χάνει την ουσία και το περιεχόμενο της, πολλές λεπτομέρειες και αναλύσεις έχουν μόνο οι πίνακες. Το στατιστικό διάγραμμα περιλαμβάνει και έχει εκτός από το διάγραμμα και:

1. Τον τίτλο στο πάνω μέρος του σχήματος,
2. Την κλίμακα των τιμών των μεγεθών που απεικονίζονται στον κατακόρυφο άξονα των Υ και τον οριζόντιο άξονα των Χ,
3. Την ένδειξη των πηγών,
4. Υπόμνημα το οποίο συνήθως γράφεται κάτω δεξιά από το διάγραμμα και εξηγεί τις διάφορες γραμμές που περιέχει η γραφική παράσταση.

2.5. ΕΙΔΗ ΔΙΑΓΡΑΜΜΑΤΩΝ

2.5.1. Ραβδογράμματα

Τα διαγράμματα αυτά χρησιμοποιούνται για τη γραφική παράσταση ποιοτικών μεταβλητών, ποσοτικών ασυνεχών μεταβλητών και για την απεικόνιση της διαχρονικής εξέλιξης κάποιου φαινομένου. Η δημιουργία των ραβδογραμμάτων πραγματοποιείται με τη χάραξη, στον κάθετο ή στον οριζόντιο άξονα, ορθογώνιων παραλληλόγραμμων, το μήκος των οποίων είναι ανάλογο με τις συχνότητες που αντιπροσωπεύουν.

2.5.2. Κυκλικά διαγράμματα

Τα κυκλικά διαγράμματα χρησιμοποιούνται για τη γραφική απεικόνιση των σχετικών συχνοτήτων, τόσο των ποιοτικών όσο και των ποσοτικών μεταβλητών όταν οι διάφορες τιμές της μεταβλητής είναι λίγες. Για να κατασκευάσουμε ένα κυκλικό διάγραμμα χωρίζουμε τον κυκλικό δίσκο σε κυκλικούς τομείς των οποίων οι επίκεντρες γωνίες βαίνουν σε τόξα α_i ανάλογα με τις συχνότητες ν_i των τιμών της μεταβλητής.

Επειδή τα ποσά α_i και ν_i είναι ανάλογα, ισχύει:

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

2.5.3. Χρονολογικά διαγράμματα

Τα διαγράμματα αυτά είναι για να δείχνουν στοιχεία σε χρονολογική σειρά. Χρονολογική σειρά είναι μία σειρά παρατηρήσεων η γεγονότων που παίρνονται σε ίσα χρονικά διαστήματα πχ ο πυρετός, η καθημερινή τιμή των μετοχών, η θερμοκρασία της ημέρας και της νύχτας, άλλες ασθένειες, οι γεννήσεις, τα κέρδη μιας επιχείρησης ή και οι ζημιές ακόμη και η πορεία της, οι βαθμοί των μαθητών κ.α. Για τη γραφική παράσταση μιας χρονολογικής σειράς τοποθετούμε στον οριζόντιο άξονα των ορθογώνιων συντεταγμένων τις χρονικές στιγμές και στον κάθετο άξονα τις τιμές της μεταβλητής που θέλουμε να μελετήσουμε. Τα χρονοδιαγράμματα φτιάχνονται ή με τη μορφή τεθλασμένης γραμμής ή σε μορφή ορθογωνίων παραλληλογράμμων. Είναι πολύ χρήσιμα και σημαντικά για τη χρονική σύγκριση δύο ή περισσότερων μεταβλητών. Για παράδειγμα τα κέρδη κάποιας επιχείρησης ανάλογα με το χρόνο αυτό.

2.5.4. Ημιλογαριθμικά και λογαριθμικά διαγράμματα

Τα ημιλογαριθμικά και λογαριθμικά διαγράμματα χρησιμοποιούνται:

- Όταν οι τιμές της μεταβλητής που θέλουμε να δείξουμε γραφικά για κάτι είναι πολύ μικροί ή πολύ μεγάλοι.
- Όταν μας απασχολεί η ποσοστιαία μεταβολή των τιμών της μεταβλητής και όχι για την απόλυτη μεταβολή πχ αν έχουμε στοιχεία για μία σειρά χρόνων, του εθνικού εισοδήματος, πιο πολύ μας ενδιαφέρει πόσο αυξήθηκε ή μειώθηκε, ελαττώθηκε το εισόδημα κατά ποσοστό μεταξύ δύο ή παραπάνω χρονικών περιόδων, παρά να ξέρουμε το απόλυτο μέγεθος της αύξησης ή της μείωσης. Έτσι σε αυτές τις περιπτώσεις αντί για φυσικούς αριθμούς να βάζουμε τους λογάριθμους για να έχουμε το λόγο μεταβολής.

Εάν στον κάθετο μόνο άξονα τα αριθμητικά στοιχεία έχουν αντικατασταθεί με τους λογάριθμους τους, τότε τα διαγράμματα αυτά λέγονται ημιαλγοριθμικά διαγράμματα και χρησιμοποιούνται κατά βάση για τη γραφική απεικόνιση χρονολογικών σειρών. Αν όμως όπως στο κάθετο αλλά και στον οριζόντιο άξονα, η κλίμακα των απόλυτων τιμών έχει αντικατασταθεί με λογαριθμική κλίμακα, τότε το διάγραμμα επονομάζεται λογαριθμικό διάγραμμα και χρησιμοποιείται κυρίως για τη γραφική παράσταση της κατανομής των εισοδημάτων, γαιών κ.λπ. Στο εμπόριο υπάρχει ειδικό χαρτί για τη δημιουργία και κατασκευή λογαριθμικών και ημιαλγοριθμικών παραστάσεων, διαγραμμάτων.

2.5.5. Τα χαρτογράμματα

Τα χαρτογράμματα είναι γραφικές παραστάσεις στατιστικών στοιχείων σε γεωγραφικούς ή τοπογραφικούς χάρτες. Για να παραστήσουμε τις τιμές της μεταβλητής του φαινομένου το οποίο θέλουμε να αναλύσουμε στις διάφορες περιοχές, χρωματίζουμε στο χάρτη διάφορες περιοχές, ανάλογα με την τιμή της μεταβλητής, με διάφορα χρώματα. Κάτω από το χαρτόγραμμα γράφουμε το υπόμνημα που αναφέρει τι δείχνει κάθε χρωματισμός. Για παράδειγμα το μπλε δείχνει τα νησιά του Αιγαίου, το πράσινο δείχνει τα νησιά του Ιονίου, το μπεζ δείχνει τους νομούς, το πορτοκαλί δείχνει τις πόλεις, το καφέ δείχνει τα χωριά και σε όλα αυτά βλέπουμε τους ρυθμούς αυξομείωσης τους κάθε χρόνο με τις απογραφές που γίνονται σε αυτά.

2.5.6. Ιστογράμματα και καμπύλες συχνοτήτων

Η γραφική παράσταση ενός πίνακα συχνοτήτων με ομαδοποιημένα δεδομένα ονομάζεται ιστόγραμμα συχνοτήτων. Εάν στο ιστόγραμμα 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.

Το ιστόγραμμα μας πληροφορεί για τη μορφή μιας κατανομής συχνοτήτων. Εάν υποθέσουμε ότι ο αριθμός των κλάσεων για μία συνεχή μεταβλητή είναι αρκετά μεγάλος και ότι το πλάτος των κλάσεων είναι αρκετά μικρό, τότε το πολύγωνο συχνοτήτων τείνει να πάρει τη μορφή μιας ομαλής καμπύλης, η οποία ονομάζεται καμπύλη συχνοτήτων. Οι καμπύλες συχνοτήτων έχουν τεράστια σημασία για τη στατιστική, γιατί απ’ αυτές βγαίνουν χρήσιμα και μεγάλα συμπεράσματα πχ μια καμπύλη συχνοτήτων μας πληροφορεί, για τη μορφή της εξεταζόμενης κατανομής συχνοτήτων.

2.6. ΣΤΑΤΙΣΤΙΚΕΣ ΕΚΘΕΣΕΙΣ Η ΑΝΑΦΟΡΕΣ

Ένας άλλος τρόπος για την παρουσίαση στατιστικών στοιχείων και δεδομένων ανάλογα με το σκοπό που έχουμε και επιδιώκουμε, είναι οι εκθέσεις ή αναφορές στο κείμενο των οποίων γράφονται τα βασικότερα και κυριότερα σημεία των αποτελεσμάτων, σχολιάζεται η σημασία των αποτελεσμάτων αυτών, γίνονται οι παρατηρήσεις από αυτόν που συνέταξε την έκθεση και αναφέρεται πολλές φορές σύντομη σημείωση της στατιστικής τεχνικής που ακολούθησε η έρευνα. Ο τρόπος αυτός παρουσίασης σε μορφή αναφοράς έχει κάποια μειονεκτήματα, γιατί αυτός που διαβάζει είναι αναγκασμένος να διαβάσει ολόκληρη την έκθεση πολύ συγκεντρωμένα για να συγκρίνει τα διάφορα αριθμητικά στοιχεία και να του μείνουν, να τα συγκρατήσει στη μνήμη του αυτά που τον ενδιαφέρουν. Συχνά μέσα στο κείμενο των εκθέσεων δημοσιεύονται και μερικοί περιληπτικοί πίνακες.

2.7. ΔΕΙΓΜΑΤΟΛΗΨΙΑ

Εάν ο πληθυσμός που θέλουμε να μελετήσουμε έχει μεγάλο πλήθος στατιστικών μονάδων και αν η μελέτη των ιδιοτήτων καταστρέφει τις μονάδες πληθυσμού που εξετάζουμε τότε η γενική απογραφή είναι πρακτικά αδύνατη. Τότε χρησιμοποιούμε τη δειγματοληπτική μέθοδο που συνίσταται στην προσπάθεια να ανακαλύψουμε τις ιδιότητες ενός πληθυσμού εξετάζοντας από αυτόν αποκλειστικά ένα τυχαίο δείγμα, το οποίο επιλέγουμε κατά τέτοιο τρόπο ώστε οι πληροφορίες, οι εκτιμήσεις και τα συμπεράσματα που θα πάρουμε και λάβουμε από αυτό να ισχύουν για το σύνολο του πληθυσμού στο οποίο ανήκει το αντιπροσωπευτικό δείγμα.

Πλεονεκτήματα:

- Μεγαλύτερη ταχύτητα πληροφοριών,
- Μεγαλύτερη ακρίβεια και αποτελεσματικότητα,
- Μεγαλύτερη ευχέρεια εφαρμογής,
- Χαμηλό κόστος,
- Αδυναμία εφαρμογής της γενικής απογραφής.

Μειονεκτήματα:

- Εάν οι μονάδες του πληθυσμού που ερευνούμε εμφανίζονται πολύ σπάνια, πρέπει να μελετήσουμε ένα σημαντικά μεγάλο δείγμα, αν θέλουμε και επιθυμούμε να έχουμε αξιόπιστη εκτίμηση των χαρακτηριστικών ιδιοτήτων του πληθυσμού,
- Ο σχεδιασμός και η εκτέλεση της δειγματοληψίας χρειάζονται ιδιαίτερη και μεγάλη προσοχή και θα πρέπει να ακολουθήσει αυστηρά η θεωρητική διαδικασία που επιβάλλεται για την επιλογή του δείγματος και της στατιστικής ανάλυσης των αποτελεσμάτων της δειγματοληψίας,
- Τα δειγματοληπτικά σφάλματα.

- Η κακή σχεδίαση και εκτέλεση της δειγματοληψίας, η μη αντιπροσωπευτικότητα του δείγματος, η μη κατάλληλη διενέργεια του δείγματος της δειγματοληψίας και τα ανεπαρκή δεδομένα και πληροφορίες οδηγούν σε αποτυχία της μερικής έρευνας.

Οι μέθοδοι συλλογής στις δειγματοληπτικές έρευνες είναι:

- Η παρατήρηση,
- Οι προσωπικές συνεντεύξεις
- Ταχυδρομική αποστολή του ερωτηματολογίου.

Η εκλογή της μεθόδου επηρεάζεται:

- Από το επιθυμητό βαθμό ακρίβειας των αποτελεσμάτων,
- Από τα χρονικά και χρηματικά περιθώρια της έρευνας,
- Από τη μεταβλητικότητα και αλλαγή των μονάδων του πληθυσμού που μελετάμε και ερευνάμε,
- Από την ύπαρξη καταλόγων των μονάδων πληθυσμού,
- Από τη δυνατότητα και ευχέρεια που έχει ο πληθυσμός να μπορεί να διαιρεθεί σε υποπληθυσμούς σε μεγάλη ομοιογένεια.

Για να αποφασιστεί μια δειγματοληπτική έρευνα πρέπει να ληφθούν διάφορες, και ορισμένες αποφάσεις τεχνικής φύσης και έρευνες από πριν. Για να γίνει αυτό ένας στατιστικός να συνεργαστεί με μεγάλη εμπειρία στη διεξαγωγή των ερευνών με διάφορους επιστήμονες ειδικούς και όχι τυχαίους που είναι σχετικοί με το εξεταζόμενο πρόβλημα. Είναι η λεγόμενη ομάδα συνεργασίας, οι οποίοι είναι υπεύθυνοι και για το σχεδιασμό, την ανάλυση και την εκτέλεση των αποτελεσμάτων της έρευνας. Επίσης θα είναι σωστό να δείξουν ποια είναι η φύση του προς την εξέταση του προβλήματος και την κατανόηση και διατύπωση των αντικειμενικών σκοπών της. Για να γίνει αυτό πραγματικότητα είναι σοφό να μελετήσει όμοιες έρευνες που έχουν γίνει στο παρελθόν στο εσωτερικό της χώρας της Ελλάδας και στο εξωτερικό. Όμως πρέπει να κάνουν και προϋπολογισμό σχετικό με τις έρευνες που κάνουν. Για τη δειγματοληπτική έρευνα βασικά μέρη και θέση έχουν τα:

- Η κατάρτιση του πλαισίου,
- Το μέγεθος του δείγματος,
- Η μέθοδος δειγματοληψίας,
- Το ερωτηματολόγιο,
- Ο προϋπολογισμός της έρευνας,

- Η δοκιμαστική έρευνα,
- Η προετοιμασία των απογραφέων,
- Χρονοδιάγραμμα.

Στη δειγματοληψία υπάρχουν σφάλματα. Το σφάλμα ορίζεται γενικά από τη διαφορά μεταξύ μια εκτιμήσεως που βγαίνει από ένα δείγμα και της αντίστοιχης παραμέτρου που εξάγεται και βγαίνει από μια απογραφή. Τα σφάλματα χωρίζονται σε δειγματοληπτικά και μη δειγματοληπτικά. Όταν μία στατιστική έρευνα γίνεται με απογραφή ή με δειγματοληψία έρχονται τα λεγόμενα μη δειγματοληπτικά σφάλματα, τα οποία οφείλονται:

- Στη διαφορά ορισμών,
- Στις αναληθείς απαντήσεις των απογραφόμενων,
- Στην ατελή σύνταξη, τον σχεδιασμό του ερωτηματολογίου,
- Στις λαθεμένες μετρήσεις,
- Στις λαθεμένες καταχωρήσεις στις γραμμές και στις στήλες του ερωτηματολογίου από τους ερευνητές.

Δειγματοληπτικό σφάλμα είναι η διαφορά μεταξύ μιας στατιστικής εκτιμήσεως που εξάγεται και βγαίνει από ένα δείγμα και της αντίστοιχης στατιστικής παραμέτρου που προκύπτει με απογραφή.

Το άθροισμα των δειγματοληπτικών και μη δειγματοληπτικών σφαλμάτων αποτελεί και είναι το συνολικό σφάλμα μιας στατιστικής έρευνας, το οποίο πρέπει και είναι ανάγκη να ελαχιστοποιηθεί σε κάθε τύπου έρευνας. Τα σφάλματα ελαχιστοποιούνται και ελαττώνονται με:

- Με τη σωστή σύνταξη του ερωτηματολογίου,
- Με την καλή εκπαίδευση των ερευνητών,
- Με τον καλό σχεδιασμό της έρευνας και
- Τον καλό έλεγχο κατά τη διάρκεια της συλλογής των στατιστικών πληροφοριών.

ΚΕΦΑΛΑΙΟ 3^{ο}

ΣΤΑΤΙΣΤΙΚΑ ΜΕΤΡΑ ΚΕΝΤΡΙΚΗΣ ΤΑΣΗΣ ΚΑΙ ΘΕΣΗΣ

Ο σκοπός των μέτρων κεντρικής τάσης είναι να συνοψίσουν έναν πληθυσμό από τη άποψη ορισμένων χαρακτήρων μιας σειράς παρατηρήσεων ή μετρήσεων με ορισμένους αντιπροσωπευτικούς αριθμούς, που ονομάζονται στατιστικές παράμετροι και μετράνε την κεντρική τάση των δεδομένων. Τα μέτρα κεντρικής τάσης προσδιορίζουν ένα κεντρικό σημείο γύρω από το οποίο τείνουν να συγκεντρώνονται τα δεδομένα. Επίσης, λέγονται και μέτρα θέσης διότι δίνουν περιληπτικά τη θέση των δεδομένων πάνω στην ευθεία των πραγματικών αριθμών.

3.1. ΜΕΣΟΙ ΚΕΝΤΡΙΚΗΣ ΤΑΣΗΣ

Οι μέσοι κεντρικής τάσης είναι οι:

- Μέσος αριθμητικός,
- Μέσος γεωμετρικός,
- Μέσος αρμονικός.

3.1.1. Ο μέσος αριθμητικός ή μέση τιμή

Εάν έχουμε μία σειρά από παρατηρήσεις, δεδομένα, μετρήσεις που σχετίζονται σε μια χαρακτηριστική ιδιότητα ενός πληθυσμού, τότε ο μέσος αριθμητικός είναι το πηλίκο του αθροίσματος των τιμών των παρατηρήσεων ή μετρήσεων δια του πλήθους των μονάδων του πληθυσμού ή του δείγματος. Το μέσο αριθμητικό το διακρίνουμε σε:

- Αστάθμητο και
- Σταθμικό.

Τον αστάθμητο μέσο αριθμητικό τον έχουμε όταν όλες οι τιμές των δεδομένων έχουν την ίδια βαρύτητα, δηλαδή κάθε τιμή της εξεταζόμενης μεταβλητής έχει συντελεστή βαρύτητας τη μονάδα = 1. Το ίδιο συμβαίνει και για το σταθμικό μέσο με τη μόνη διαφορά ότι κάθε τιμή έχει διαφορετικό συντελεστή βαρύτητας.

Έστω $x_1, x_2, \ldots, x_k$ οι τιμές μιας μεταβλητής X, που αφορά τα στοιχεία ενός δείγματος μεγέθους ν, $k \leq \nu$. Ο αριθμητικός μέσος είναι:

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

Ο παραπάνω τύπος ισχύει και στην περίπτωση που τα δεδομένα είναι ομαδοποιημένα σε ν τάξεις ενός πίνακα συχνοτήτων και η τάξη i έχει κεντρική τιμή x_i και συχνότητα f_i.

Έστω τώρα ότι σε κάθε τιμή x_i έχουμε επισυνάψει ένα συντελεστή (βάρος) w_i ο οποίος είναι ο βαθμός σπουδαιότητας της x_i, τότε ο σταθμικός μέσος των τιμών $x_1, x_2, \ldots, x_k$ δίνεται από τον τύπο:

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

- **Γενικός μέσος δύο ή περισσότερων κατανομών**

Αν έχουμε τους μέσους δύο ή περισσότερων κατανομών συχνοτήτων, δηλαδή αν $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$ είναι οι μέσοι των k κατανομών συχνοτήτων και $N_1, N_2, \ldots, N_k$ οι αντίστοιχες συχνότητες αυτών, τότε ο γενικός μέσος όρος υπολογίζεται βάσει του τύπου:

$$\bar{x} = \frac{\bar{x}_1 \bar{N}_1 + \bar{x}_2 \bar{N}_2 + \cdots + \bar{x}_k \bar{N}_k}{\bar{N}_1 + \bar{N}_2 + \cdots + \bar{N}_k}.$$

- **Ιδιότητες μέσου αριθμητικού**

- Αν προσθέσουμε σε όλες τις τιμές μιας μεταβλητής x_i μια σταθερή ποσότητα α, τότε και ο μέσος αριθμητικός τους αυξάνεται κατά τη σταθερή αυτή ποσότητα.
- Αν αφαιρέσουμε από όλες τις τιμές μιας μεταβλητής x_i με μια σταθερή ποσότητα α, τότε και ο μέσος αριθμητικός τους ελαττώνεται κατά τη σταθερή αυτή ποσότητα.
- Αν πολλαπλασιάσουμε όλες τις τιμές μιας μεταβλητής x_i με μια σταθερή ποσότητα α, τότε και ο μέσος αριθμητικός τους πολλαπλασιάζεται με τη σταθερή αυτή ποσότητα.
- Το αλγεβρικό άθροισμα των αποκλίσεων (διαφορών) του μέσου αριθμητικού από κάθε τιμή της μεταβλητής x_i είναι 0 μηδέν. Δηλαδή ισχύει η σχέση:

$$\sum_{i=1}^{k} \nu_i(x_i - \bar{x}) = 0.$$

- Η μέση τιμή είναι ποσότητα που:
  1. Υπολογίζεται εύκολα.
  2. Είναι μονοσήμαντα ορισμένη.
  3. Λαμβάνει υπ’ όψιν στον υπολογισμό της όλες τις παρατηρήσεις.
  4. Επηρεάζεται από ακραίες τιμές (πολύ μικρές ή μεγάλες τιμές).
  5. Δεν συμπίπτει απαραίτητα με μια από τις τιμές της μεταβλητής.
  6. Βρίσκεται πάντοτε ανάμεσα στη μικρότερη και τη μεγαλύτερη τιμή της μεταβλητής.

- Στην περίπτωση ομαδοποιημένων παρατηρήσεων, η μέση τιμή που υπολογίζουμε με τη βοήθεια των μέσων κάθε κλάσης είναι μία εκτίμηση της πραγματικής μέσης τιμής. Το σφάλμα που κάνουμε όμως κατά την εκτίμηση αυτή δεν είναι ιδιαίτερα μεγάλο και εξαρτάται από το πόσο συμμετρική, γύρω από τη μέση τιμή, είναι η κατανομή συχνοτήτων.
- Ο μέσος αριθμητικός δεν μπορεί να υπολογιστεί στις περιπτώσεις ανοικτών κατανομών συχνοτήτων.

3.1.2. Ο μέσος γεωμετρικός

Έστω $x_1, x_2, \ldots, x_k \geq 0$ οι τιμές μιας μεταβλητής X, που αφορά τα στοιχεία ενός δείγματος μεγέθους ν, $k \leq \nu$. Ο γεωμετρικός μέσος δίνεται από τον τύπο:

$$\bar{x}_g = \sqrt[\nu]{{x_1}^{\nu_1} \cdot {x_2}^{\nu_2} \cdot \ldots \cdot {x_k}^{\nu_k}}.$$

Για τον υπολογισμό του μέσου γεωμετρικού χρησιμοποιούμε λογαρίθμους. Λογαριθμούμε και τα δύο μέλη της παραπάνω σχέσης και έχουμε:

$$\log \bar{x}_g = \log \sqrt[\nu]{{x_1}^{\nu_1} \cdot {x_2}^{\nu_2} \cdot \ldots \cdot {x_k}^{\nu_k}} = \log({x_1}^{\nu_1} \cdot {x_2}^{\nu_2} \cdot \ldots \cdot {x_k}^{\nu_k})^{\frac{1}{\nu}}$$

$$= \frac{1}{\nu} \cdot \log({x_1}^{\nu_1} \cdot {x_2}^{\nu_2} \cdot \ldots \cdot {x_k}^{\nu_k}) = \frac{1}{\nu} \cdot \sum_{i=1}^{k} \log({x_i}^{\nu_i}) = \frac{1}{\nu} \cdot \sum_{i=1}^{k} \nu_i \log(x_i)\,.$$

Από τη σχέση αυτή συνεπάγεται ότι: Ο λογάριθμος του μέσου γεωμετρικού ισούται με το μέσο αριθμητικό των λογαρίθμων των τιμών της μεταβλητής X.

- **Ιδιότητες μέσου γεωμετρικού**

- Ο γεωμετρικός μέσος:
  1. Υπολογίζεται μόνο για μεταβλητές με θετικές τιμές.
  2. Επηρεάζεται, όπως και η μέση τιμή (αν και θεωρητικά λιγότερο) από την ύπαρξη ακραίων τιμών.
  3. Πολλές φορές, αν το πλήθος των τιμών της μεταβλητής είναι μεγάλο, (λόγω του ότι ο υπολογισμός της ν-οστής ρίζας είναι δύσκολος), υπολογίζεται με τη βοήθεια των λογαρίθμων.
- Είναι μέσος που χρησιμοποιείται σπάνια, και κυρίως για τον υπολογισμό μέσων ποσοστών μεταβολής και την κατάρτιση αριθμοδεικτών.

- **Κυριότερες εφαρμογές του μέσου γεωμετρικού**

Ο μέσος γεωμετρικός χρησιμοποιείται κυρίως για τον υπολογισμό της μέσης ποσοστιαίας μεταβολής χρονολογικών σειρών αλλά και για την κατάρτιση αριθμοδεικτών. Είναι ένα από τα βασικότερα εργαλεία της στατιστικής επιστήμης με τη οποία ασχολούμαστε αλλά και μελετάμε εκτενέστερα. Αν για παράδειγμα έχουμε μια χρονολογική σειρά n όρων,

$$k_1, k_2, \dots, k_n$$

οι οποίοι σχηματίζουν γεωμετρική πρόοδο, είναι δυνατόν να υπολογίσουμε το λόγο της προόδου. Ο λόγος αυτός είναι το μέσο ποσοστό αυξήσεως (ή μειώσεως) της σειράς.

Το μέσο ποσοστό μεταβολής της χρονολογικής σειράς μπορεί να υπολογιστεί με δύο τρόπους:

α) Με τον τύπο του ανατοκισμού

$$k_n = k_1(1 + \mathrm{i})^{n-1}$$

όπου:

k_1= Πρώτος όρος της χρονολογικής σειράς

k_n= τελευταίος όρος της χρονολογικής σειράς

n = πλήθος όρων της χρονολογικής σειράς

i = μέσο ποσοστό μεταβολής χρονολογικής σειράς (δηλαδή ο μέσος γεωμετρικός)

Το μέσο ποσοστό (= i) θα βρεθεί αν λύσουμε την παραπάνω σχέση ως προς i.

Δηλαδή:

$$1 + \mathrm{i} = \sqrt[n-1]{\frac{k_n}{k_1}}.$$

β) Μπορούμε να μετατρέψουμε τους όρους της χρονολογικής σειράς σε αλυσιδωτούς αριθμοδείκτες. Δηλαδή, διαιρούμε κάθε όρο της σειράς με τον προηγούμενο του και έχουμε:

$$\frac{k_2}{k_1}, \frac{k_3}{k_2}, \dots, \frac{k_n}{k_{n-1}}.$$

Ο μέσος γεωμετρικός των παραπάνω αριθμοδεικτών θα είναι:

$$G = 1 + \mathrm{i} = \sqrt[n-1]{\frac{k_2}{k_1} \cdot \frac{k_3}{k_2} \cdot \ldots \cdot \frac{k_n}{k_{n-1}}} = \sqrt[n-1]{\frac{k_n}{k_1}}.$$

3.1.3. Ο μέσος αρμονικός

Ο μέσος αρμονικός είναι ο αντίστροφος του μέσου αριθμητικού των αντιστρόφων τιμών της μεταβλητής.

Έστω $x_1, x_2, \ldots, x_k \neq 0$ οι τιμές μιας μεταβλητής X, που αφορά τα στοιχεία ενός δείγματος μεγέθους ν, $k \leq \nu$. Ο γεωμετρικός μέσος δίνεται από τον τύπο:

$$\bar{x}_h = \frac{\nu}{\frac{\nu_1}{x_1} + \frac{\nu_2}{x_2} + \cdots + \frac{\nu_k}{x_k}}.$$

Παρατηρούμε ότι:

$$\frac{1}{\bar{x}_h} = \frac{1}{\nu} \cdot \sum_{i=1}^{k} \nu_i \frac{1}{x_i}.$$

- **Ιδιότητες μέσου αρμονικού**

- Ο αρμονικός μέσος:
  1. Υπολογίζεται σχετικά εύκολα.
  2. Λαμβάνει υπ’ όψιν όλες τις τιμές της μεταβλητής.
  3. Επηρεάζεται από ακραίες τιμές.
  4. Δεν υπολογίζεται για μεταβλητές που παίρνουν την τιμή 0.

- Είναι μέσος που χρησιμοποιείται σπάνια, και κυρίως σε προβλήματα υπολογισμού μέσων ποσοστών χρόνου και σε προβλήματα αποστάσεων.

- **Σύγκριση αριθμητικού, γεωμετρικού και αρμονικού μέσου**

Ισχύει γενικά ότι:

$$\bar{x}_h \leq \bar{x}_g \leq \bar{x}$$

3.2. ΠΑΡΑΜΕΤΡΟΙ ΘΕΣΕΩΣ

Όπως προαναφέραμε το σπουδαιότερο στατιστικό μέτρο μιας ομάδας δεδομένων είναι ο μέσος αριθμητικός. Δημιουργείται όμως ένα ερώτημα αν ο μέσος αριθμητικός είναι πάντοτε το ιδανικότερο και αντιπροσωπευτικότερο στατιστικό μέτρο. Η απάντηση σε αυτό το ερώτημα είναι η εξής: αν έχουμε μια κατανομή συχνοτήτων της οποίας η καμπύλη συχνοτήτων είναι συμμετρική ή προσεγγίζει τη συμμετρική μορφή, τότε μπορούμε να πούμε ότι ο μέσος αριθμητικός είναι το πιο αντιπροσωπευτικό στατιστικό μέτρο. Αυτό εξαρτάται και από άλλους παράγοντες τους οποίους θα συναντήσουμε πιο κάτω. Αν όμως η καμπύλη συχνοτήτων έχει ασυμμετρική μορφή, τότε ο μέσος αριθμητικός δεν μπορεί να θεωρηθεί αντιπροσωπευτικό στατιστικό μέτρο, γιατί υψηλό για παράδειγμα μέσο εισόδημα αντιπροσωπεύει μικρό αριθμό εισοδηματιών και δεν μας δίνει πληροφορίες για το εισόδημα της μεγάλης μάζας των εισοδηματιών. Όταν λοιπόν, μία κατανομή είναι ασυμμετρική, τότε ο μέσος αριθμητικός δεν είναι ευσταθές και αξιόπιστο στατιστικό μέτρο για τη λήψη ορθών και σωστών αποφάσεων. Υπάρχουν άλλα στατιστικά μέτρα (στατιστικές παράμετροι), τα οποία δίνουν αξιόπιστες πληροφορίες για τον εξεταζόμενο στατιστικό πληθυσμό. Τα στατιστικά μέτρα αυτά είναι:

- Η διάμεσος,
- Τα τεταρτημόρια,
- Τα δεκατημόρια,
- Τα εκατοστημόρια και
- Η επικρατούσα τιμή ή κορυφή.

3.2.1. Διάμεσος

Η διάμεσος είναι το σημαντικότερο και σπουδαιότερο στατιστικό μέτρο θέσεως και συμβολίζεται με το M. Όταν οι τιμές μιας μεταβλητής X ταξινομηθούν κατά τη φυσική τους τάξη (από τη μικρότερη προς τη μεγαλύτερη), τότε η διάμεσος είναι η τιμή εκείνη της μεταβλητής, η οποία κατέχει την κεντρική θέση. Με άλλα λόγια:

Διάμεσος είναι η τιμή της μεταβλητής, η οποία χωρίζει το σύνολο των τιμών της μεταβλητής σε δύο ισοπληθείς ομάδες. Συνεπώς, κάτω της διαμέσου βρίσκονται τα 50% των τιμών της μεταβλητής και άνω της διαμέσου τα υπόλοιπα 50% αυτών.

Για να υπολογίσουμε τη διάμεσο ταξινομούμε τα δεδομένα κατά αύξουσα σειρά 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 μεσαίων παρατηρήσεων.

Στην περίπτωση ομαδοποιημένων δεδομένων σε ν κλάσεις, ο υπολογισμός της διαμέσου γίνεται ως εξής:

1) Υπολογίζουμε τις αθροιστικές συχνότητες: $N_1, N_2, \dots, N_\nu$.
2) Εντοπίζουμε μεταξύ ποιων αθροιστικών συχνοτήτων N_{i-1} και N_i βρίσκεται ο αριθμός $\nu/2$. Εάν ο αριθμός $\nu/2$ συμπίπτει με μία από τις αθροιστικές συχνότητες, τότε ως διάμεσος ορίζεται το άνω άκρο της αντίστοιχης κλάσης.
3) Χρησιμοποιούμε τον τύπο:

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

όπου:

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

ν_i : η συχνότητα της τάξης i της διαμέσου,

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

δ : το πλάτος της τάξης i της διαμέσου.

- **Ιδιότητες διαμέσου**

- Η τιμή της διαμέσου δεν επηρεάζεται από τις ακραίες τιμές της μεταβλητής όπως ο μέσος αριθμητικός. Αν, λοιπόν σε μια ομάδα αριθμητικών δεδομένων υπάρχουν ακραίες τιμές, τότε η διάμεσος είναι αντιπροσωπευτικότερο στατιστικό μέτρο και πρέπει να προτιμάται.
- Η διάμεσος μπορεί να υπολογιστεί και όταν ακόμα μια κατανομή συχνοτήτων είναι ανοιχτή, αρκεί η τάξη που εντοπίζεται η διάμεσος να είναι κλειστή.
- Η διάμεσος μπορεί να υπολογιστεί και γραφικά.
- Στην περίπτωση ποιοτικών δεδομένων, η διάμεσος δεν έχει νόημα.

3.2.2 Τεταρτημόρια – δεκατημόρια – εκατοστημόρια

Είπαμε πιο πάνω ότι, εκτός από τη διάμεσο υπάρχουν και άλλα στατιστικά μέτρα θέσεως. Παράμετροι θέσεως, συναφείς με τη διάμεσο, είναι τα τεταρτημόρια, τα δεκατημόρια και τα εκατοστημόρια. Ενώ η διάμεσος υποδιαιρεί το συνολικό αριθμό των τιμών μιας μεταβλητής σε δύο ισοπληθείς ομάδες, τα τεταρτημόρια υποδιαιρούν το σύνολο των τιμών σε τέσσερις ισοπληθείς ομάδες, τα δεκατημόρια σε δέκα ισοπληθείς ομάδες και τα εκατοστημόρια σε εκατό ισοπληθείς ομάδες.

I. Υπολογισμός τεταρτημορίων

Το πρώτο τεταρτημόριο συμβολίζεται με Q_1 και είναι η τιμή εκείνης της μεταβλητής κάτω της οποίας βρίσκονται τα 25% του συνόλου των τιμών της μεταβλητής. Το δεύτερο τεταρτημόρια είναι η διάμεσος. Το τρίτο τεταρτημόριο συμβολίζεται με Q_3 και είναι η τιμή εκείνη της μεταβλητής κάτω της οποίας εντοπίζονται τα 75% του συνόλου των τιμών της μεταβλητής και άνω του Q_3 τα υπόλοιπα 25% των τιμών. Για τον υπολογισμό των τεταρτημορίων χρησιμοποιούμε τους εξής τύπους:

α) Για αταξινόμητα δεδομένα, εάν ν είναι το πλήθος των διατεταγμένων παρατηρήσεων, τότε το πρώτο τεταρτημόριο εντοπίζεται στη θέση

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

ενώ το τρίτο τεταρτημόριο εντοπίζεται στη θέση

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

β) Για ταξινομημένα δεδομένα σε κατανομή συχνοτήτων χρησιμοποιούνται οι εξής τύποι:

$$Q_1 = \alpha_{i-1} + \frac{\delta}{\nu_i}\left(\frac{\nu}{4} - N_{i-1}\right),$$

$$Q_3 = \alpha_{i-1} + \frac{\delta}{\nu_i}\left(\frac{3\nu}{4} - N_{i-1}\right).$$

II. Υπολογισμός δεκατημορίων

α) Για αταξινόμητα δεδομένα, εάν ν είναι το πλήθος των διατεταγμένων παρατηρήσεων, τότε το k - δεκατημόριο εντοπίζεται στη θέση

$$\frac{(\nu+1)k}{10}, \qquad k = 1,2,\dots,9.$$

β) Για ταξινομημένα δεδομένα σε κατανομή συχνοτήτων χρησιμοποιούνται οι εξής τύποι:

$$D_k = \alpha_{i-1} + \frac{\delta}{\nu_i}\left(\frac{k\nu}{10} - N_{i-1}\right).$$

III. Υπολογισμός εκατοστημορίων

α) Για αταξινόμητα δεδομένα, εάν ν είναι το πλήθος των διατεταγμένων παρατηρήσεων, τότε το k - εκατοστημόριο εντοπίζεται στη θέση

$$\frac{(\nu+1)k}{100}, \qquad k = 1,2,\dots,99.$$

β) Για ταξινομημένα δεδομένα σε κατανομή συχνοτήτων χρησιμοποιούνται οι εξής τύποι:

$$P_k = \alpha_{i-1} + \frac{\delta}{\nu_i}\left(\frac{k\nu}{100} - N_{i-1}\right).$$

3.2.3 Επικρατούσα τιμή ή κορυφή

Επικρατούσα τιμή ή κορυφή είναι η τιμή εκείνη της μεταβλητής που αντιστοιχεί στη μεγαλύτερη συχνότητα της κατανομής, γι 'αυτό ονομάζεται και σημείο μεγαλύτερης συχνότητας και συμβολίζεται με το M_0. Η επικρατούσα τιμή έχει σημασία στην περίπτωση που έχουμε ένα μεγάλο αριθμό δεδομένων και είναι δυνατός ο υπολογισμός της σε περίπτωση που έχουμε μονοκόρυφη κατανομή συχνοτήτων, ενώ μια κατανομή με δύο σημεία μεγίστης συχνότητας λέγεται δικόρυφη ή διπλοκόρυφη κατανομή. Η επικρατούσα τιμή δεν έχει έννοια στις δικόρυφες κατανομές, εκτός αν τα δεδομένα τα ταξινομήσουμε εκ νέου σε μεγαλύτερο ή μικρότερο αριθμό τάξεων και εμφανισθεί μονοκόρυφη κατανομή συχνοτήτων. Το χαρακτηριστικό γνώρισμα της επικρατούσας τιμής είναι ότι δεν επηρεάζεται από τις ακραίες τιμές (πολύ μεγάλες ή πολύ μικρές) μιας εξεταζόμενης μεταβλητής και μπορεί να υπολογιστεί και στις περιπτώσεις ανοικτών κατανομών, όταν δηλαδή δεν υπάρχει το κατώτατο όριο της πρώτης τάξης και το ανώτατο όριο της τελευταίας τάξης μιας κατανομής συχνοτήτων. Επίσης εφαρμόζεται πολύ αν η κατανομή συχνοτήτων είναι ασυμμετρική.

Η επικρατούσα τιμή υπολογίζεται από τον τύπο:

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

όπου:

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

δ : το πλάτος της τάξεως με τη μεγαλύτερη συχνότητα.

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

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

- **Ιδιότητες επικρατούσας τιμής**

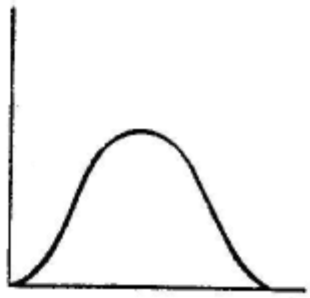
- Η επικρατούσα τιμή δεν επηρεάζεται από τις ακραίες τιμές της μεταβλητής.
- Η επικρατούσα τιμή μπορεί να υπολογιστεί και όταν ακόμα μια κατανομή συχνοτήτων είναι ανοιχτή, αρκεί η επικρατούσα κλάση να μην είναι η ανοικτή κλάση.

- Η επικρατούσα τιμή πολλές φορές δεν είναι μοναδική (έχουμε δηλαδή περισσότερες από μια επικρατούσες τιμές) και άλλες φορές δεν υπάρχει.
- Στην περίπτωση ποιοτικών δεδομένων, η επικρατούσα τιμή έχει νόημα.
-

3.3. ΣΥΓΚΡΙΣΗ ΜΕΣΗΣ ΤΙΜΗΣ, ΔΙΑΜΕΣΟΥ, ΕΠΙΚΡΑΤΟΥΣΑΣ ΤΙΜΗΣ

1) Αν η κατανομή των δεδομένων είναι **συμμετρική**, δηλαδή το ιστόγραμμα (ή το πολύγωνο συχνοτήτων) έχει μορφή ανάλογη με αυτήν του Διαγράμματος 4.1, τότε

$$M = M_0 = \overline{x}.$$



Διάγραμμα 4.1: Συμμετρία

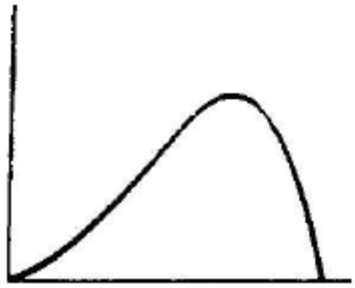
2) Εάν η κατανομή δεν είναι συμμετρική, αλλά παρουσιάζει τη μορφή του Διαγράμματος 4.2, είναι όπως λέμε **θετικά ασύμμετρη** (μικρές τιμές παρουσιάζονται πιο συχνά από ότι μεγαλύτερες και η ουρά της κατανομής βρίσκεται στο δεξί μέρος), τότε

$$M_0 < M < \overline{x}.$$

Διάγραμμα 4.2: Θετική ασυμμετρία

3) Εάν η κατανομή δεν είναι συμμετρική, αλλά παρουσιάζει τη μορφή του Διαγράμματος 4.3, είναι δηλαδή **αρνητικά ασύμμετρη** (μεγάλες τιμές παρουσιάζονται πιο συχνά από ότι οι μικρότερες και η ουρά της κατανομής βρίσκεται στο αριστερό μέρος), τότε

$$\bar{x} < M < M_0.$$



Διάγραμμα 4.3: Αρνητική ασσυμετρία

ΚΕΦΑΛΑΙΟ 4^{o}

ΣΤΑΤΙΣΤΙΚΑ ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ

Ο μέσος αριθμητικός, ο μέσος γεωμετρικός, η διάμεσος, η επικρατούσα τιμή και τα υπόλοιπα στατιστικά μέτρα τα οποία έχουμε ήδη εξετάσει είναι ιδιαίτερα χρήσιμα για τη συνοπτική περιγραφή ενός συνόλου δεδομένων, με την προϋπόθεση ότι τα μέτρα κεντρικής τάσης και θέσης να είναι αντιπροσωπευτικά της εξεταζόμενης κατανομής. Η γνώση όμως ενός μέσου όρου δεν αρκεί για να περιγράψει πλήρως την κατανομή των παρατηρήσεων, τη λήψη σωστών αποφάσεων και τη σύγκριση διαφόρων κατανομών συχνοτήτων. Η χρησιμότητα και η αξιοπιστία των παραμέτρων τάσης και θέσης ως δείκτες αντιπροσώπευσης και συνοπτικής παρουσίασης ενός συνόλου δεδομένων εξαρτάται από τη μεταβλητότητα (διασπορά) των τιμών των δεδομένων γύρω από τους μέσους και βασικά ως προς το μέσο αριθμητικό. Ο βαθμός κατά τον οποίο οι διάφορες τιμές ενός πληθυσμού ή ενός δείγματος τείνουν να είναι συγκεντρωμένες ή διασπαρμένες γύρω από τον μέσο αριθμητικό ονομάζεται διασπορά (μεταβλητότητα).

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

Παράδειγμα 1:

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

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

Πίνακας 5.1

Οι μέσες τιμές των εσόδων είναι:

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

και

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

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

Τα συνηθέστερα μέτρα που συνήθως χρησιμοποιούμε στη στατιστική για τη μέτρηση της μεταβλητότητας είναι:

- Το εύρος,
- Το ενδοτεταρτημοριακό εύρος,
- Η μέση απόλυτη απόκλιση,
- Η διακύμανση,
- Η τυπική απόκλιση,
- Ο συντελεστής μεταβλητότητας,
- Η μέση διαφορά του Gini.

4.1.ΤΟ ΕΥΡΟΣ

Το εύρος (*R*) είναι το πιο απλό μέτρο διασποράς και μας δίνει τη διαφορά ανάμεσα στη μεγαλύτερη και τη μικρότερη τιμή μιας σειράς παρατηρήσεων.

Το πλεονέκτημα του Εύρους είναι η ευκολία υπολογισμού του, το δε μειονέκτημα είναι ότι για τον υπολογισμό του λαμβάνονται υπόψη μόνο δυο τιμές του δείγματος, οι ακραίες. Στην περίπτωση όπου οι δυο ακραίες τιμές βρίσκονται σε μεγάλη απόσταση από τις υπόλοιπες, τότε η εικόνα της διασποράς είναι ψευδής. Παρά το σημαντικό μειονέκτημά του, το εύρος του δείγματος χρησιμοποιείται συχνά, κυρίως σε περιπτώσεις που οι ακραίες τιμές δεν απέχουν σημαντικά από τις υπόλοιπες τιμές του δείγματος.

4.1.1ΤΟ ΕΝΔΟΤΕΤΑΡΤΗΜΟΡΙΑΚΟ ΕΥΡΟΣ

Το ενδοτεταρτημοριακό εύρος H είναι η διαφορά του πρώτου τεταρτημορίου Q_1 από τρίτο Q_3:

$$H = Q_3 - Q_1$$

Το ενδοτεταρτημοριακό εύρος, μας δίνει το εύρος του 50% των κεντρικών παρατηρήσεων του δείγματος, απαλλαγμένο από το 25% των μικρότερων και το 25% των μεγαλύτερων παρατηρήσεων. Όσο μικρότερη είναι η τιμή του H, τόσο μεγαλύτερη είναι η συγκέντρωση των τιμών της μεταβλητής. Το πλεονέκτημα του ενδοτεταρτημοριακού εύρους είναι το ότι δεν επηρεάζεται από τις ακραίες τιμές του δείγματος, αφού αυτές δεν λαμβάνονται υπ’ όψιν στον υπολογισμό του, πράγμα που συμβαίνει με το εύρος R.

4.1.2 ΜΕΣΗ ΑΠΟΛΥΤΗ ΑΠΟΚΛΙΣΗ

Η μέση απόλυτη απόκλιση ορίζεται ως ο μέσος αριθμητικός όλων των διαφορών των τιμών μιας μεταβλητής από το μέσο αριθμητικό της μεταβλητής αυτής.

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

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

4.2. ΔΙΑΚΥΜΑΝΣΗ

Για να αποφύγουμε τη διαδικασία εκείνη κατά την οποία δεν λαμβάνουμε υπόψη μας τα σημεία των αρνητικών διαφορών μεταξύ του μέσου αριθμητικού και των τιμών της μεταβλητής, όπως συμβαίνει στη μέση απόλυτη απόκλιση, μπορούμε αντί των απόλυτων τιμών να λάβουμε υπόψη μας τα τετράγωνα των αποκλίσεων των τιμών της μεταβλητής από το μέσο αριθμητικό, ώστε οι ποσότητες $(x_i - \bar{x})^2$ να είναι πάντα θετικές.

Ο μέσος αριθμητικός των τετραγώνων των αποκλίσεων των τιμών της μεταβλητής X από το μέσο αριθμητικό είναι επίσης ένα μέτρο διασποράς και ονομάζεται διακύμανση και είναι το πιο γνωστό αριθμητικό μέτρο για τον υπολογισμό του ρίσκου στα χρηματοοικονομικά.

Έστω $x_1, x_2, \dots, x_k$ οι τιμές μιας μεταβλητής X, που αφορά τα στοιχεία ενός δείγματος μεγέθους ν, $k \leq \nu$ και $\bar{x}$ η μέση τιμή. Η διακύμανση δίνεται από τον τύπο[4]:

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

Ο τύπος της διακύμανσης, μπορεί με κατάλληλες πράξεις να πάρει και την ισοδύναμη μορφή:

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

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

[4] Στο σημείο αυτό θα πρέπει να αναφέρουμε ότι στη βιβλιογραφία συναντάται για τη διακύμανση και ο τύπος:

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

αλλά δεν προτιμάται, αφού δίνει μεγαλύτερο σφάλμα εκτίμησης της διακύμανσης του πληθυσμού.

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

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

Συνεπώς,

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

- **Ιδιότητες διακύμανσης**

- Η διακύμανση μιας σταθερής ποσότητας α είναι ίση με το μηδέν.
- Εάν προσθέσουμε (ή αφαιρέσουμε) σε όλες τις τιμές μιας μεταβλητής X μια σταθερή ποσότητα α, τότε η διακύμανση παραμένει αμετάβλητη.
- Εάν πολλαπλασιάσουμε (ή διαιρέσουμε) όλες τις τιμές μιας μεταβλητής X με μια σταθερή ποσότητα β, τότε η διακύμανση πολλαπλασιάζεται (διαιρείται) με β^2.
- Εάν μια μεταβλητή Y είναι γραμμικός μετασχηματισμός της μεταβλητής X, δηλαδή εάν $Y = \alpha + \beta X$, τότε $s_Y^2 = \beta^2 s_X^2$.

4.3. ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ

Η διακύμανση εκφράζεται σε μονάδες, οι οποίες είναι τα τετράγωνα των αρχικών μονάδων της εξεταζόμενης μεταβλητής παραδείγματος χάρη αν το εισόδημα μιας κατηγορίας εργαζομένων εκφράζεται σε ευρώ, η διακύμανση εκφράζεται σε μονάδα το ευρώ, στο τετράγωνο, ομοίως αν η μεταβλητή εκφράζεται σε εκατοστά, η διακύμανση εκφράζεται σε εκατοστά στο τετράγωνο και λοιπά. Για να έχουμε όμως ένα δείκτη ο οποίος μετράει τη μεταβλητικότητα μιας μεταβλητής και να εκφράζεται στις ίδιες μονάδες που εκφράζεται η μεταβλητή μας, παίρνουμε την τετραγωνική ρίζα της διακύμανσης. Το μέτρο αυτό ονομάζεται τυπική απόκλιση και είναι από τα βασικά μέτρα που χρησιμοποιούμε συνήθως στην πράξη για τη μελέτη της μεταβλητότητας.

Ώστε, τυπική απόκλιση s ορίζεται η τετραγωνική ρίζα του μέσου αριθμητικού των τετραγώνων των αποκλίσεων των τιμών της μεταβλητής X, από το μέσο αριθμητικό τους. Δηλαδή, η τυπική απόκλιση, δίνεται από τον τύπο:

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

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

- **Ιδιότητες τυπικής απόκλισης**

- Η τυπική απόκλιση μιας σταθερής ποσότητας α είναι ίση με το μηδέν.
- Εάν προσθέσουμε (ή αφαιρέσουμε) σε όλες τις τιμές μιας μεταβλητής X μια σταθερή ποσότητα α, τότε η τυπκή απόκλιση παραμένει αμετάβλητη.
- Εάν πολλαπλασιάσουμε (ή διαιρέσουμε) όλες τις τιμές μιας μεταβλητής X με μια σταθερή ποσότητα $\beta > 0$, τότε η τυπική απόκλιση πολλαπλασιάζεται (διαιρείται) με β.

- **Τυποποιημένες αποκλίσεις**

Έστω ότι $x_1, x_2, \dots, x_k$ είναι οι τιμές μιας μεταβλητής X με μέση τιμή $\bar{x}$ και τυπική απόκλιση s. Εάν από κάθε τιμή x_i αφαιρέσουμε το $\bar{x}$, τότε δημιουργούνται οι αποκλίσεις

$$x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_k - \bar{x}.$$

Εάν τώρα κάθε μία από τις παραπάνω αποκλίσεις διαιρεθεί με την τυπική απόκλιση, τότε προκύπτουν οι αποκλίσεις:

$$z_1 = \frac{x_1 - \bar{x}}{s}, z_2 = \frac{x_2 - \bar{x}}{s}, \dots, z_k = \frac{x_k - \bar{x}}{s}$$

οι οποίες ονομάζονται τυποποιημένες αποκλίσεις και η μεταβλητή Z ονομάζεται τυποποιημένη μεταβλητή. Οι τυποποιημένες μεταβλητές έχουν τεράστια χρησιμότητα στην επαγωγική στατιστική. Στην περιγραφική στατιστική χρησιμεύουν όταν θέλουμε να συγκρίνουμε δύο τιμές από διαφορετικές σειρές δεδομένων και όταν οι σειρές:

i) διαφέρουν ως προς το $\bar{x}$ ή την s, ή και τα δύο ή
ii) εκφράζονται σε διαφορετικές μονάδες μετρήσεως.

4.4. ΘΕΩΡΗΜΑ CHEBYSHEV

Για οποιοδήποτε σύνολο δεδομένων και για οποιαδήποτε σταθερά $\lambda > 1$, το ποσοστό των δεδομένων που βρίσκεται στο διάστημα $(\bar{x} - \lambda s, \bar{x} + \lambda s)$ είναι τουλάχιστον

$$1 - \frac{1}{\lambda^2}.$$

λ	Διάστημα	Ποσοστό δεδομένων στο διάστημα αυτό
1	$(\bar{x} - s, \bar{x} + s)$	$\geq 0\%$
2	$(\bar{x} - 2s, \bar{x} + 2s)$	$\geq 75\%$
2.5	$(\bar{x} - 2.5s, \bar{x} + 2.5s)$	$\geq 84\%$
3	$(\bar{x} - 3s, \bar{x} + 3s)$	$\geq 89\%$

Πίνακας 5.2

Παράδειγμα 2:

Θεωρούμε όλους τους μηνιαίους μισθούς των 2000 εργαζομένων σ' ένα εργοστάσιο και υπολογίζουμε ότι η μέση τιμή των μισθών είναι $\bar{x} = 1200$ ευρώ και η τυπική απόκλιση $s = 160$ ευρώ.

Σύμφωνα με τα προηγούμενα, μπορούμε να αποφανθούμε με βεβαιότητα ότι:

1) Το 75% των εργαζομένων παίρνει μισθό 1200 ± 320. Δηλαδή, μεταξύ 880 ευρώ και 1520 ευρώ αμείβονται οι 1500 εργαζόμενοι του εργοστασίου.
2) Το 84% των εργαζομένων παίρνει μισθό 1200 ± 400. Δηλαδή, μεταξύ 800 ευρώ και 1600 ευρώ αμείβονται οι 1680 εργαζόμενοι του εργοστασίου.
3) Το 89% των εργαζομένων παίρνει μισθό 1200 ± 480. Δηλαδή μεταξύ 720 ευρώ και 1680 ευρώ αμείβονται οι 1780 εργαζόμενοι του εργοστασίου.

Όταν τα δεδομένα έχουν μια συμμετρική κατανομή με μια μόνο κορυφή, σχήμα το οποίο ταιριάζει στην κανονική κατανομή ισχύει ο παρακάτω εμπειρικός κανόνας:

- Το διάστημα $(\bar{x} - s, \bar{x} + s)$ περιέχει περίπου το 68% των δεδομένων.
- Το διάστημα $(\bar{x} - 2s, \bar{x} + 2s)$ περιέχει περίπου το 95% των δεδομένων.
- Το διάστημα $(\bar{x} - 3s, \bar{x} + 3s)$ περιέχει σχεδόν όλα τα δεδομένα.

4.5. ΣΥΝΤΕΛΕΣΤΗΣ ΜΕΤΑΒΛΗΤΟΤΗΤΑΣ

Η τυπική απόκλιση, η οποία θεωρείται ως το κυρίως χρησιμοποιούμενο μέτρο για τη μέτρηση της διασποράς, εκφράζεται στις ίδιες μονάδες με τις οποίες εκφράζεται η τυχαία μεταβλητή και μας δίνει την απόλυτη διασπορά των τιμών της τυχαίας μεταβλητής από τον μέσο αριθμητικό. Η χρησιμοποίηση όμως της τυπικής απόκλισης και των άλλων μέτρων διασποράς είναι σε αρκετές περιπτώσεις αδύνατη και σε άλλες περιπτώσεις πολύ περιορισμένη. Αυτό συμβαίνει όταν θέλουμε να συγκρίνουμε δύο κατανομές οι οποίες εκφράζονται σε διαφορετικές μονάδες (μέτρα, κιλά, ευρώ, εκατοστά και λοιπά), ή όταν οι μέσοι αριθμητικοί δύο διαφορετικών μεταβλητών, έστω και αν εκφράζονται στις ίδιες μονάδες, διαφέρουν πάρα πολύ μεταξύ τους. Τότε τα μέτρα της απόλυτης διασποράς δεν μας εξυπηρετούν και χρησιμοποιούμε τη σχετική διασπορά. Το βασικό μέτρο της σχετικής διασποράς είναι ο συντελεστής μεταβλητικότητας. Ο συντελεστής αυτός είναι ανεξάρτητος από τις μονάδες μέτρησης που χρησιμοποιούμε και, επομένως, επιτρέπει τη σύγκριση τόσο των ομοειδών όσο και των ετεροειδών κατανομών. Ο συντελεστής μεταβλητικότητας ορίζεται ως ο λόγος της τυπικής απόκλισης προς τον αριθμητικό μέσο των παρατηρήσεων, δηλαδή:

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

Επομένως, ο συντελεστής μεταβλητικότητας που είναι το πηλίκο τη τυπικής απόκλισης μιας κατανομής προς τον αριθμητικό μέσο αυτής εκφράζει την τυπική απόκλιση ως ποσοστό επί τοις εκατό του μέσου αριθμητικού $\bar{x}$ και είναι ένας καθαρός αριθμός, χωρίς μονάδες μέτρησης.

Συμπερασματικά μπορούμε να πούμε ότι όσο πιο μικρός είναι ο συντελεστής μεταβλητότητας, τόσο μεγαλύτερη ομοιογένεια υπάρχει στις τιμές της μεταβλητής. Γενικά δεχόμαστε ότι ένα δείγμα τιμών μιας μεταβλητής θα είναι ομοιογενές, εάν ο συντελεστής μεταβλητότητας δεν είναι μεγαλύτερος του 10%. Θα πρέπει όμως να τονίσουμε ότι ο συντελεστής μεταβλητότητας δεν ενδείκνυται να χρησιμοποιείται όταν η μέση τιμή είναι κοντά στο μηδέν.

4.6. ΜΕΣΗ ΔΙΑΦΟΡΑ ΤΟΥ GINI

Η μέση διαφορά του Gini είναι ο μέσος αριθμητικός των απολύτων τιμών των διαφορών τις οποίες μπορούμε να σχηματίσουμε συνδυάζοντας ανά δύο, με όλους τους δυνατούς τρόπους, όλες τις τιμές μιας μεταβλητής. Κυρίως στα οικονομικά φαινόμενα μας ενδιαφέρει να γνωρίζουμε τη απόκλιση που παρουσιάζει η τιμή μιας μεταβλητής σε σχέση με τις τιμές των άλλων τιμών μιας μεταβλητής και όχι τις αποκλίσεις κάθε τιμής από το μέσο αριθμητικό. Στην περίπτωση αυτή, το στατιστικό μέτρο που χρησιμοποιούμε για να μετρήσουμε τη διακύμανση ονομάζεται μέση διαφορά. Το μέτρο αυτό της διασποράς το πρότεινε ο καθηγητής του τμήματος Στατιστικών και Αναλογιστικών Επιστημών του Πανεπιστημίου της Ρώμης C. Gini.

Έστω ότι έχουμε τις παρατηρήσεις $x_1, x_2, \dots, x_\nu$. Για τον υπολογισμό της μέσης διαφοράς του Gini χρησιμοποιούμε τον τύπο:

$$GMD = \frac{1}{\nu(\nu-1)} \sum_{i=1}^{\nu} \sum_{j=1}^{\nu} |x_i - x_j|.$$

4.7. ΘΗΚΟΓΡΑΜΜΑ

Γνωρίζουμε ότι για να μπορεί κανείς να ερμηνεύσει ορθά τα μέτρα θέση είναι απαραίτητο να διαθέτει πείρα και εμπειρία για να μπορέσει, να συνοψίζει, να συνδυάζει και να συμπυκνώνει όλες τις πληροφορίες που αυτά δίνουν για την κατανομή. Η διερευνητική ανάλυση δεδομένων, με μία έξυπνη συνάμα και εύκολη τεχνική βοηθάει να παρουσιαστούν τα κυριότερα μέτρα θέσης με τέτοιο τρόπο που να διευκολύνεται πολύ η εξαγωγή συμπερασμάτων για την κατανομή. Αυτό λέγεται θηκόγραμμα. Είναι γνωστό και ως διάγραμμα πέντε των αριθμών. Το θηκόγραμμα αποτελεί γραφικό τρόπο παρουσίασης πέντε περιληπτικών μέτρων μιας κατανομής ομαδοποιημένων δεδομένων, με συνδυασμό των οποίων είναι δυνατή η άντληση περισσότερων πληροφοριών από αυτήν που περιέχεται στα πέντε αυτά μέτρα. Πιο συγκεκριμένα είναι ένα ορθογώνιο με δύο κεραίες του οποίου η κατασκευή του είναι η εξής: η κάτω βάση του ορθογωνίου βρίσκεται στο Q_1 και η πάνω στο Q_3. Η διάμεσος Μ αναπαριστάνεται με ένα οριζόντιο ευθύγραμμο τμήμα μέσα στο ορθογώνιο. Το μήκος των βάσεων του ορθογωνίου λαμβάνεται αυθαίρετα. Η πάνω και η κάτω κεραία που έχουν τη μορφή Τ και ανεστραμμένου Τ αντίστοιχα, εκτείνονται μέχρι τις οριακές τιμές που μπορεί να είναι:

α) η μέγιστη και η ελάχιστη παρατήρηση,

β) η μεγαλύτερη παρατήρηση που είναι μικρότερη ή ίση από το ανώτερο εσωτερικό φράγμα $Q_3 + 1{,}5(Q_3 - Q_1)$ και η μικρότερη παρατήρηση που είναι μεγαλύτερη ή ίση από το κατώτερο εσωτερικό φράγμα $Q_1 - 1{,}5(Q_3 - Q_1)$,

γ) η μεγαλύτερη παρατήρηση που είναι μικρότερη ή ίση από το ανώτερο εξωτερικό φράγμα $Q_3 + 3(Q_3 - Q_1)$ και η μικρότερη παρατήρηση που είναι μεγαλύτερη ή ίση από το κατώτερο εξωτερικό φράγμα $Q_1 - 3(Q_3 - Q_1)$.

ΚΕΦΑΛΑΙΟ 5ο

ΕΦΑΡΜΟΓΕΣ ΤΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΜΕΤΡΩΝ

Σε αυτό το κεφάλαιο θα ασχοληθούμε και θα λύσουμε ασκήσεις, θα εφαρμόσουμε τους τύπους και τις μεθόδους που αναλύσαμε στα προηγούμενα κεφάλαιο και θα κατανοήσουμε πρακτικά τις έννοιες, τους τύπους και τους ορισμούς που ήδη ξέρουμε θεωρητικά. Είναι ουσιαστικά η μετάβαση από τη θεωρία στην πράξη για μια πιο ολοκληρωμένη και σφαιρική γνώση των στατιστικών μέτρων θέσης και διασποράς.

ΕΦΑΡΜΟΓΗ 5.1

Στον παρακάτω πίνακα δίνεται η ομαδοποιημένη κατανομή των καταθέσεων σε χιλιάδες ευρώ ενός συνόλου νοικοκυριών της Αττικής.

Κλάσεις	Συχνότητες ν_i
[100-150)	10
[150-200)	8
[200-250)	19
[250-300)	12
[300-350)	11

Πίνακας 6.1

1) Να κατασκευαστεί ο πίνακας συχνοτήτων και να συμπληρωθεί με τις στήλες των κεντρικών τιμών x_i, των αθροιστικών συχνοτήτων N_i, και των γινομένων $x_i \nu_i$.
2) Να βρεθεί:

   α) από πόσα νοικοκυριά προέρχονται οι καταθέσεις,

   β) η μέση τιμή των καταθέσεων.

ΛΥΣΗ

1) Ο πίνακας διαμορφώνεται ως εξής:

Κλάσεις	x_i	ν_i	N_i	$x_i \nu_i$
[100-150)	125	10	10	1250
[150-200)	175	8	18	1400
[200-250)	225	19	37	4275
[250-300)	275	12	49	3300
[300-350)	325	11	60	3575
Σύνολο		60		13800

Πίνακας 6.2

2) Από τον πίνακα προκύπτει ότι:

   α) Ο αριθμός των νοικοκυριών είναι ίσος με 60,

   β) Έχουμε:

$$\bar{x} = \frac{1}{60}\sum_{i=1}^{5} x_i \nu_i = \frac{1}{60} \cdot 13800 = 230.$$

Άρα, η μέση τιμή των καταθέσεων είναι 230000 ευρώ.

ΕΦΑΡΜΟΓΗ 5.2

Τα 16 μεγάλα τμήματα ενός μεγάλου εργοστασίου του Βόλου έχουν τους εξής εργαζόμενους:

41 37 38 40 39 41 41 37 39 39 38 38 40 39 37 39

1) Να κατασκευαστεί πίνακας σχετικών και αθροιστικών συχνοτήτων.
2) Να βρεθεί η μέση τιμή, διάμεσος, επικρατούσα τιμή, εύρος, διασπορά, τυπική απόκλιση για τα δεδομένα.
3) Να κατασκευαστεί ραβδόγραμμα και κυκλικό διάγραμμα σχετικών συχνοτήτων.

ΛΥΣΗ

1) Από τα δεδομένα του προβλήματος κατασκευάζουμε τον παρακάτω πίνακα συχνοτήτων:

Τιμές x_i	Συχνότητες ν_i	Σχετικές Συχνότητες f_i	Αθροιστικές Συχνότητες N_i	$x_i \nu_i$	x_i^2	$x_i^2 \nu_i$
37	3	0,1875	3	111	1369	4107
38	3	0,1875	6 (= 3+3)	114	1444	4332
39	5	0,3125	11 (= 6+5)	195	1521	7605
40	2	0,125	13	80	1600	3200
41	3	0,1875	16	123	1681	5043
Σύνολο	16	1		$\sum_{i=1}^{5} x_i \nu_i = 623$		$\sum_{i=1}^{5} x_i^2 \nu_i = 24287$

Πίνακας 6.3

2) Η μέση τιμή είναι:

$$\bar{x} = \frac{1}{16}\sum_{i=1}^{5} x_i \nu_i = \frac{1}{60} \cdot 623 = 38{,}93.$$

Η διάμεσος είναι:

$$M = 39$$

(άρτιο πλήθος παρατηρήσεων, 16, άρα η διάμεσος είναι η $8^{\eta} + 9^{\eta}$ παρατήρηση δια 2).

Η επικρατούσα τιμή είναι:

$$M_0 = 39.$$

Το εύρος είναι:

$$R = 41 - 37 = 4.$$

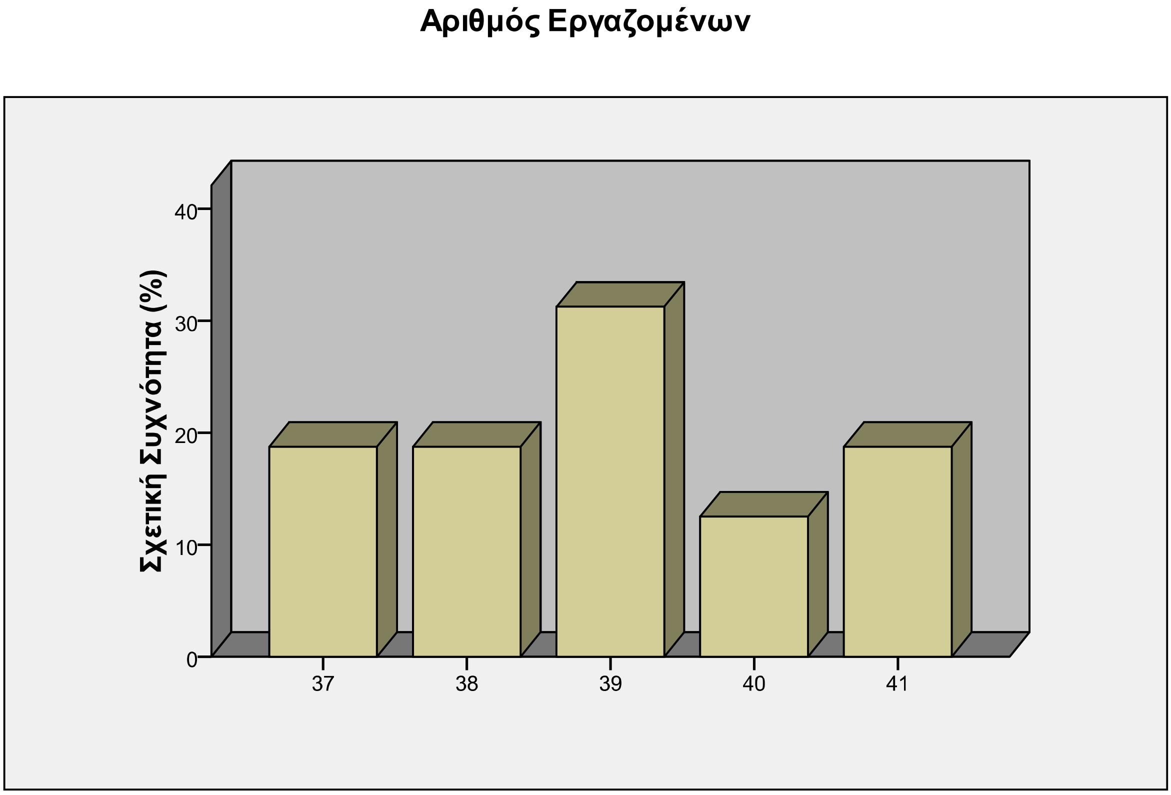
Η διασπορά είναι:

$$s^2 = \frac{1}{15}\left[\sum_{i=1}^{5} \nu_i x_i^2 - 16\bar{x}^2\right] = \frac{1}{15}\left(24287 - 16 \cdot 38{,}93^2\right) = 2{,}5521.$$

Η τυπική απόκλιση είναι:

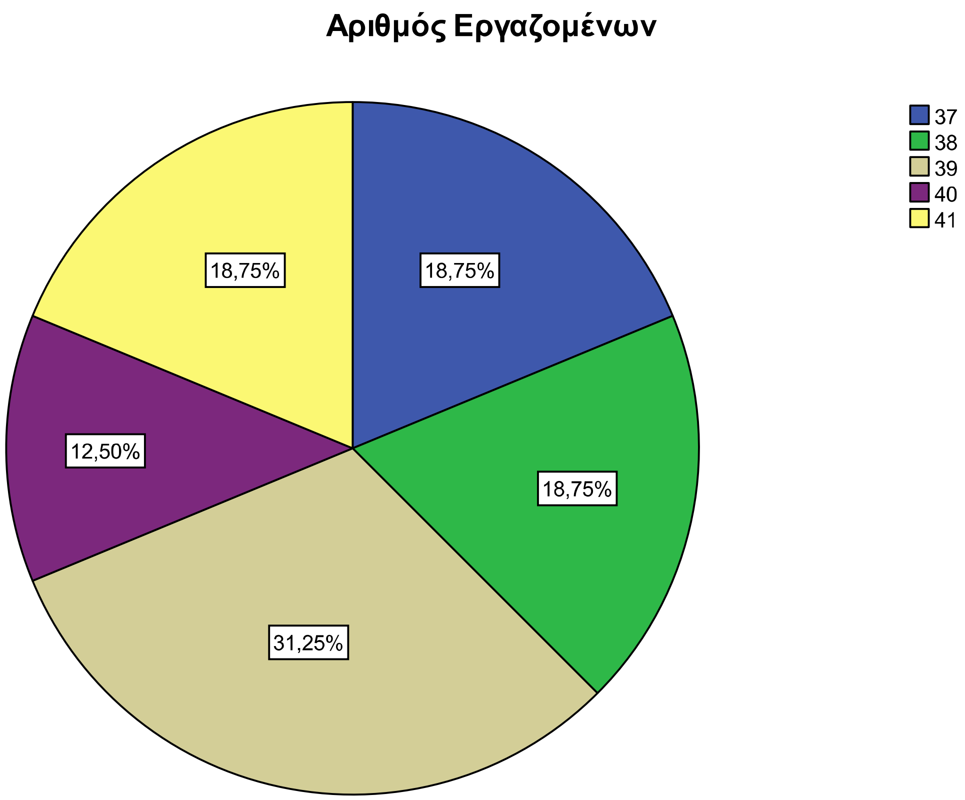
$$s = \sqrt{s^2} = \sqrt{2{,}5521} = 1{,}5975.$$

3) Το ραβδόγραμμα σχετικών συχνοτήτων δίνεται παρακάτω:



Διάγραμμα 5.1: ραβδόγραμμα

4) Το κυκλικό διάγραμμα σχετικών συχνοτήτων δίνεται παρακάτω:



Διάγραμμα 5.2: κυκλικό διάγραμμα

ΕΦΑΡΜΟΓΗ 5.3

Σε μια εταιρεία κινητής τηλεφωνίας, έγινε καταμέτρηση του χρόνου διάρκειας των τηλεφωνικών κλήσεων στο κέντρο εξυπηρέτησης πελατών, για μια εβδομάδα. Καταγράφηκαν συνολικά 1200 τηλεφωνικές κλήσεις που ομαδοποιήθηκαν σε κλάσεις των 2 λεπτών. Οι τιμές του δείγματος παρουσιάζονται στον ακόλουθο πίνακα συχνοτήτων:

Χρόνος διάρκειας κλήσεων x_i	**Αριθμός κλήσεων** ν_i
[0, 2)	100
[2, 4)	200
[4, 6)	250
[6, 8)	250
[8, 10)	200
[10, 12)	100
[12, 14)	60
[14, 16)	20
[16, 18)	10
[18, 20)	10
Σύνολο	**1200**

Πίνακας 6.4

1) Να υπολογισθούν τα τεταρτημόρια.
2) Να υπολογισθεί το ενδοτεταρτημοριακό εύρος.
3) Να βρεθεί η μέση τιμή, η διακύμανση και τυπική απόκλιση.

ΛΥΣΗ

1) Για να υπολογίσουμε τα τεταρτημόρια πρέπει πρώτα να βρούμε τις αθροιστικές συχνότητες. Ο πίνακας αθροιστικών συχνοτήτων δίνεται παρακάτω:

Χρόνος διάρκειας κλήσεων x_i	**Αριθμός κλήσεων** ν_i	**Αθροιστική Συχνότητα** N_i
[0, 2)	100	100
[2, 4)	200	300
[4, 6)	250	550
[6, 8)	250	800
[8, 10)	200	1000
[10, 12)	100	1100
[12, 14)	60	1160
[14, 16)	20	1180
[16, 18)	10	1190
[18, 20)	10	1200
Σύνολο	**1200**	

Πίνακας 6.5

Υπολογισμός πρώτου τεταρτημορίου.

Έχουμε:

$$\frac{N}{4} = \frac{1200}{4} = 300 = N_2.$$

Συνεπώς,

$$Q_1 = 4.$$

Υπολογισμός δεύτερου τεταρτημορίου (διαμέσου).

Ο αριθμός

$$\frac{\nu}{2} = \frac{1200}{2} = 600$$

βρίσκεται μεταξύ των αθροιστικών συχνοτήτων $N_3 = 550$ και $N_4 = 800$. Έχουμε:

$$\alpha_3 = 6, \nu_4 = 250, \delta = 2.$$

Επομένως η τιμή της διαμέσου είναι:

$$M = Q_2 = \alpha_3 + \frac{\delta}{\nu_4}\left(\frac{\nu}{2} - N_3\right) = 6 + \frac{2}{250}\left(\frac{1200}{2} - 550\right) = 6{,}4.$$

Υπολογισμός τρίτου τεταρτημορίου.

Ο αριθμός

$$\frac{3\nu}{4} = \frac{3600}{4} = 900$$

βρίσκεται μεταξύ των αθροιστικών συχνοτήτων $N_4 = 800$ και $N_5 = 1000$. Έχουμε:

$$\alpha_4 = 8, \nu_5 = 200, \delta = 2.$$

Επομένως η τιμή του τρίτου τεταρτημορίου είναι:

$$Q_3 = \alpha_4 + \frac{\delta}{\nu_5}\left(\frac{3\nu}{4} - N_4\right) = 8 + \frac{2}{200}\left(\frac{3600}{4} - 800\right) = 9.$$

2) Το ενδοτεταρτημοριακό εύρος του δείγματος των τηλεφωνικών κλήσεων στο κέντρο εξυπηρέτησης πελατών, της εταιρείας κινητής τηλεφωνίας είναι ίσο με:

$$H = Q_3 - Q_1 = 9 - 4 = 5 \text{ λεπτά της ώρας.}$$

3) Δημιουργούμε στον πίνακα των δεδομένων τις στήλες που χρειαζόμαστε για τον υπολογισμό των στατιστικών μέτρων.

Χρόνος διάρκειας κλήσεων x_i	**Αριθμός κλήσεων** ν_i	$x_i \nu_i$	$x_i^2 \nu_i$
[0, 2)	100	100	100
[2, 4)	200	600	1800
[4, 6)	250	1.250	6250
[6, 8)	250	1.750	12250
[8, 10)	200	1.800	16200
[10, 12)	100	1.100	12100
[12, 14)	60	780	10140
[14, 16)	20	300	4500
[16, 18)	10	170	2890
[18, 20)	10	190	3610
Σύνολο	**1200**	**8040**	**69840**

Πίνακας 6.6

Υπολογισμός μέσης τιμής.

Η μέση τιμή των τηλεφωνικών κλίσεων του τηλεφωνικού κέντρου είναι:

$$\bar{x} = \frac{1}{1200} \sum_{i=1}^{10} x_i \nu_i = \frac{8040}{1200} = 6{,}7.$$

Δηλαδή η μέση διάρκεια κάθε κλήσης ήταν $\bar{x} = 6{,}7$ λεπτά.

Υπολογισμός διακύμανσης.

Για τη διακύμανση των τηλεφωνικών κλήσεων του δείγματος έχουμε:

$$s^2 = \frac{1}{1199}\left[\sum_{i=1}^{10} \nu_i x_i^2 - 1200\bar{x}^2\right] = \frac{1}{1199}\left(69849 - 1200 \cdot 6{,}7^2\right) = 13{,}3286.$$

Δηλαδή η διακύμανση των τηλεφωνικών κλήσεων του τηλεφωνικού κέντρου ήταν

$$s^2 = 13{,}3286.$$

Υπολογισμός τυπικής απόκλισης.

Η τυπική απόκλιση των τηλεφωνικών κλήσεων είναι:

$$s = \sqrt{s^2} = \sqrt{13{,}3286} \cong 3{,}65.$$

Συνεπώς, για τις τηλεφωνικές κλήσης που δέχεται το κέντρο εξυπηρέτησης πελατών της εταιρείας κινητής τηλεφωνίας, διαπιστώνουμε ότι έχουν μέση χρονική διάρκεια $\bar{x} = 6{,}7$ λεπτά και τυπική απόκλιση $s = 3{,}65$ λεπτά.

ΕΦΑΡΜΟΓΗ 5.4

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

Πίνακας 6.7

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

ΛΥΣΗ

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

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

Πίνακας 6.8

Η νέα μέση τιμή των πωλήσεων της εταιρείας θα είναι:

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

Η διακύμανση των πωλήσεων θα είναι:

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

Συνεπώς, η νέα τυπική απόκλιση είναι:

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

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

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

και

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

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

(1) Έχουμε σημαντική αύξηση της μέσης τιμής των πωλήσεων από 90000000 ευρώ πέρυσι 130000000 ευρώ φέτος.

(2) Η διασπορά των πωλήσεων φέτος παρουσιάζει καλύτερη εικόνα, δηλαδή έχουμε καλύτερη ομοιογένεια ως προς τα έσοδα, αφού:

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

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

Πίνακας 6.9

ΕΦΑΡΜΟΓΗ 5.5

Δίνεται η κατανομή 1000 υπαλλήλων σύμφωνα με το ημερομίσθιο τους.

Μισθοί	Εργάτες
[30,35)	9
[35,40)	108
[40,45)	488
[45,50)	230
[50,55)	112
[55,60)	30
[60,65)	16
[65,70)	7

Πίνακας 6.10

Να υπολογισθούν τα βασικά μέτρα θέσης και διασποράς και να ερμηνευθούν τα αποτελέσματα.

ΛΥΣΗ

Δημιουργούμε στον πίνακα των δεδομένων τις στήλες που χρειαζόμαστε για τον υπολογισμό των στατιστικών μέτρων.

Τάξεις-μισθοί	ν_i	x_i	$x_i \nu_i$	N_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot \nu_i$
30-35	9	32,5	292,5	9	-12,585	158,38222	1425,4399
35-40	108	37,5	4050	117	-7,585	57,532225	6213,4803
40-45	488	42,5	20740	605	-2,585	6,682225	3260,9258
45-50	230	47,5	10925	835	2,415	5,832225	1341,4117
50-55	112	52,5	5880	947	7,415	54,982225	6158,0092
55-60	30	57,5	1725	977	12,415	154,13222	4623,9666
60-65	16	62,5	1000	994	17,415	303,28222	4852,5155
65-70	7	67,5	472,5	1000	22,415	502,43222	3517,0255
Σύνολο	**1000**	**-**	**45085**	**-**	**-**	**-**	**31392,772**

Πίνακας 6.11

Υπολογισμός μέσης τιμής.

Η μέση τιμή είναι:

$$\bar{x} = \frac{1}{1000}\sum_{i=1}^{8} x_i \nu_i = \frac{45085}{1000} = 45{,}085.$$

ΕΡΜΗΝΕΙΑ: Ο μέσος όρος των μισθών των 1000 υπαλλήλων είναι 45,085 ευρώ.

Υπολογισμός επικρατούσας τιμής.

Η επικρατούσα τιμή υπολογίζεται από τον τύπο:

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

όπου:

$\alpha_{i-1} = 40$: το κατώτερο όριο της τάξεως στην οποία αντιστοιχεί η μεγαλύτερη συχνότητα ($\nu_i = 488$).

$\delta = 5$: το πλάτος της τάξεως με τη μεγαλύτερη συχνότητα.

$\Delta_1 = \nu_i - \nu_{i-1} = 488 - 108 = 380$: διαφορά μεγαλύτερης συχνότητας μείον προηγούμενη.

$\Delta_2 = \nu_i - \nu_{i+1} = 488 - 230 = 258$: διαφορά μεγαλύτερης συχνότητα μείον επόμενη.

Επομένως:

$$M_0 = 40 + 5 \cdot \frac{380}{380 - 258} = 40 + \frac{1900}{638} = 42{,}978056.$$

ΕΡΜΗΝΕΙΑ: Ο μισθός που επικρατεί (εμφανίζεται πιο πολλές φορές) είναι

$$M_0 = 42{,}9780564.$$

Υπολογισμός διαμέσου.

Η διάμεσος υπολογίζεται από τον τύπο:

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

όπου:

$\alpha_{i-1} = 40$: το κατώτερο όριο της τάξης $i = 3$ της διαμέσου,

$\nu_i = 488$: η συχνότητα της τάξης i της διαμέσου,

$N_{i-1} = 117$: η αθροιστική συχνότητα της τάξης $i - 1$ η οποία προηγείται της τάξης i,

$\delta = 5$: το πλάτος της τάξης i της διαμέσου.

Επομένως:

$$M = 40 + \frac{5}{488} \cdot [500 - 117] = 40 + \frac{5}{488} \cdot 383 = 43{,}9241803.$$

ΕΡΜΗΝΕΙΑ: Εάν ταξινομήσουμε τους μισθούς από το μικρότερο στο μεγαλύτερο, τότε η διάμεσος είναι ο μισθός που κατέχει την κεντρική θέση και χωρίζει το σύνολο των υπαλλήλων σε δύο ίσου πλήθους ομάδες. Συνεπώς: το πρώτο 50% των υπαλλήλων παίρνουν μισθό μέχρι και 43,9241803 ευρώ ενώ το υπόλοιπο 50% παίρνουν μισθό άνω των 43,9241803 ευρώ μισθό.

Υπολογισμός τεταρτημορίων.

$$Q_1 = \alpha_{i-1} + \frac{\delta}{\nu_i}\left(\frac{\nu}{4} - N_{i-1}\right) = 40 + \frac{5}{488} \cdot [250 - 117] = 40 + \frac{665}{488} = 40 + 1{,}3627049$$
$$= 41{,}3627049.$$

$$Q_3 = \alpha_{i-1} + \frac{\delta}{\nu_i}\left(\frac{3\nu}{4} - N_{i-1}\right) = 45 + \frac{5}{230} \cdot [750 - 605] = 45 + \frac{725}{230} = 45 + 3{,}1521739$$
$$= 48{,}1521739.$$

ΕΡΜΗΝΕΙΑ: Εάν ταξινομηθούν οι μεταβλητές κατά αύξουσα σειρά από τη μικρότερη στη μεγαλύτερη, τότε τα τεταρτημόρια χωρίζουν το πλήθος τους σε τέσσερις ίσες ομάδες. Για παράδειγμα, το πρώτο 25% των μεταβλητών του δείγματος έχουν τιμή μέχρι και 41,3627049 ευρώ ενώ το υπόλοιπο 75% του δείγματος έχει τιμή πάνω από 41,3627049 ευρώ.

Υπολογισμός δεκατημορίων.

$$D_4 = \alpha_{i-1} + \frac{\delta}{\nu_i}\left(\frac{4\nu}{10} - N_{i-1}\right) = 40 + \frac{5}{488} \cdot [400 - 117] = 40 + \frac{1415}{488} = 40 + 2{,}8995901 = 42{,}8995901.$$

$$D_8 = \alpha_{i-1} + \frac{\delta}{\nu_i}\left(\frac{8\nu}{10} - N_{i-1}\right) = 45 + \frac{5}{230} \cdot [800 - 605] = 45 + \frac{975}{230} = 45 + 4{,}2391304 = 49{,}2391304.$$

ΕΡΜΗΝΕΙΑ: Εάν ταξινομηθούν οι τιμές των μεταβλητών από τη μικρότερη στη μεγαλύτερη, τότε τα δεκατημόρια χωρίζουν το πλήθος των μεταβλητών σε δέκα ίσες ομάδες. Για παράδειγμα, το πρώτο 40% του δείγματος έχει τιμή μέχρι και 42,8995901 ενώ το υπόλοιπο 60% έχει τιμή πάνω από 42,8995901.

Υπολογισμός εκατοστημορίων.

$$P_{43} = \alpha_{i-1} + \frac{\delta}{\nu_i}\left(\frac{43\nu}{100} - N_{i-1}\right) = 40 + \frac{5}{117} \cdot [430 - 177] = 40 + \frac{1265}{117} = 40 + 7{,}14689266 = 47{,}14689266.$$

$$P_{89} = \alpha_{i-1} + \frac{\delta}{\nu_i}\left(\frac{89\nu}{100} - N_{i-1}\right) = 50 + \frac{5}{112} \cdot [890 - 835] = 50 + \frac{275}{112} = 50 + 2{,}45535714 = 52{,}45535714.$$

ΕΡΜΗΝΕΙΑ: Εάν ταξινομηθούν οι τιμές των μεταβλητών από τη μικρότερη στη μεγαλύτερη, τότε τα εκατοστημόρια χωρίζουν το πλήθος των μεταβλητών σε εκατό ίσες ομάδες. Για παράδειγμα, το 43% των υπαλλήλων παίρνει μέχρι 47,14689266 ευρώ ενώ το υπόλοιπο 67% παίρνει πάνω από 47,14689266 ευρώ.

Υπολογισμός τυπικής απόκλισης.

Για να βρούμε την τυπική απόκλιση πρέπει πρώτα να βρούμε τη διακύμανση άρα έχουμε:

$$s^2 = \frac{1}{999}\sum_{i=1}^{8} v_i(x_i - \bar{x})^2 = \frac{31392{,}772}{999} = 31{,}4241962.$$

Συνεπώς, η τυπική απόκλιση είναι:

$$s = \sqrt{31{,}4241962} = 5{,}605728873.$$

ΕΡΜΗΝΕΙΑ: Η τιμή $s = 5{,}605728873$ σημαίνει ότι η αξία των ημερομισθίων κάθε υπαλλήλου διαφέρει (αποκλίνει) κατά μέσο όρο από τη μέση τιμή των ημερομισθίων κατά 5,605728873.

Υπολογισμός συντελεστή μεταβλητότητας.

$$CV = \frac{s}{\bar{x}} = \frac{5{,}605728873}{45{,}085} = 0{,}1243368942.$$

ΕΡΜΗΝΕΙΑ: Επειδή

$$CV = 0{,}1243368942 > 10\% = 0{,}1,$$

το δείγμα δεν είναι ομοιογενές.

ΕΦΑΡΜΟΓΗ 5.6

Η μέση τιμή του μηνιαίου μισθού των υπαλλήλων μιας επιχείρησης είναι 800 ευρώ.

i. Αν ο μισθός κάθε υπαλλήλου αυξηθεί κατά 100 ευρώ, ποια μεταβολή θα επέλθει στη μέση τιμή του;
ii. Αν αυξηθούν οι μισθοί όλων των εργαζομένων κατά 10% ποια μεταβολή θα επέλθει στη μέση τιμή του;

ΛΥΣΗ

i. Όπως γνωρίζουμε σύμφωνα με την πρώτη ιδιότητα της μέσης τιμής, αν σε όλες τις τιμές μιας μεταβλητής X προσθέσουμε μια σταθερή ποσότητα k στην περίπτωση μας 100 ευρώ, τότε ο $\bar{x}$ αυξάνεται κατά τη σταθερή αυτή ποσότητα k. Δηλαδή η μέση τιμή του μηνιαίου μισθού των υπαλλήλων αυτής της επιχείρησης θα αυξηθεί κατά 100 ευρώ και κατά συνέπεια θα είναι:

$$\bar{x} = 800 + 100 = 900 \text{ ευρώ.}$$

ii. Σύμφωνα με τις ιδιότητες της μέσης τιμής, η νέα μέση τιμή θα αυξηθεί κατά $10\% \cdot 800 = 80$ ευρώ.

ΕΦΑΡΜΟΓΗ 5.7

Ο υπεύθυνος πωλήσεων μιας εταιρείας ηλεκτρονικών υπολογιστών ενδιαφέρεται να μελετήσει τις μηνιαίες πωλήσεις σε κάποιο από τα καταστήματα που διαθέτει η εταιρεία. Τα παρακάτω δεδομένα αναφέρονται στις μηνιαίες πωλήσεις, σε τεμάχια υπολογιστών, τους τελευταίους 48 μήνες ενός καταστήματος της εταιρείας.

25	26	32	21	29	31	27	23	34	29
32	34	35	31	36	37	41	44	46	41
56	54	45	51	53	48	50	55	49	57
67	65	59	55	57	54	59	61	54	59
67	56	64	61	59	64	62	58		

α) Να υπολογιστούν ο αριθμητικός μέσος, η επικρατούσα τιμή, η διάμεσος, το εύρος, το ενδοτεταρτημοριακό εύρος και η τυπική απόκλιση των δεδομένων,

β) Να κατασκευαστεί το θηκόγραμμα των μηνιαίων πωλήσεων και με βάση αυτό να εξαχθούν συμπεράσματα για τη μορφή της κατανομής τους,

γ) Από τα στοιχεία που διαθέτει ο υπεύθυνος πωλήσεων παρατηρεί ότι οι μηνιαίες πωλήσεις ενός άλλου κεντρικού καταστήματος είναι οι διπλάσιες από τις αντίστοιχες πωλήσεις στο συγκεκριμένο κατάστημα. Να υπολογιστούν ο αριθμητικός μέσος, η τυπική απόκλιση, το εύρος και η διακύμανση στο κεντρικό κατάστημα.

δ) Σε ποιο από τα δύο καταστήματα παρατηρείται μεγαλύτερη μεταβλητότητα στις μηνιαίες πωλήσεις;

ΛΥΣΗ

Δημιουργούμε στον πίνακα των δεδομένων τις στήλες που χρειαζόμαστε για τον υπολογισμό των στατιστικών μέτρων.

i	x_i	ν_i	x_i^2	$\nu_i x_i^2$
1	21	1	441	441
2	23	1	529	529

3	25	1	625	625
4	26	1	676	676
5	27	1	729	729
6	29	2	841	1682
7	31	2	961	1922
8	32	2	1024	2048
9	34	2	1156	2312
10	35	1	1225	1225
11	36	1	1296	1296
12	37	1	1369	1369
13	41	2	1681	3362
14	44	1	1936	1936
15	45	1	2025	2025
16	46	1	2116	2116
17	48	1	2304	2304
18	49	1	2401	2401
19	50	1	2500	2500
20	51	1	2601	2601
21	53	1	2809	2809
22	54	3	2916	8748
23	55	2	3025	6050
24	56	2	3136	6272
25	57	2	3249	6498
26	58	1	3364	3364
27	59	4	3481	13924
28	61	2	3721	7442
29	62	1	3844	3844
30	64	2	4096	8192
31	65	1	4225	4225
32	67	2	4489	8978
Σύνολο		48		114445

Πίνακας 6.12

α) Η μέση τιμή είναι:

$$\bar{x} = \frac{1}{48}\sum_{i=1}^{32} x_i v_i = \frac{2253}{48} = 46{,}9375.$$

Επειδή η επικρατούσα τιμή είναι η τιμή με τη μεγαλύτερη συχνότητα εμφάνισης εδώ ισούται με την τιμή 59 η οποία εμφανίζεται τέσσερις φορές, δηλαδή $M_0 = 59$.

Για να υπολογίσουμε τη διάμεσο, βάζουμε τα δεδομένα σε αύξουσα σειρά.

21	41	56
23	41	57
25	44	57
26	45	58
27	46	59
29	48	59
29	49	59
31	50	59
31	51	61
32	53	61
32	54	62
34	54	64
34	54	64
35	55	65
36	55	67
37	56	67

Πίνακας 6.13

Επομένως, η διάμεσος είναι:

$$M = \frac{50 + 51}{2} = 50{,}5.$$

Το εύρος ισούται με τη διαφορά μεγαλύτερη παρατήρηση - μικρότερη παρατήρηση, δηλαδή

$$R = X_{max} - X_{min} = 67 - 21 = 46.$$

Για να υπολογίσουμε το ενδοτεταρτημοριακό εύρος, πρώτα θα βρούμε τα τεταρτημόρια Q_1 και Q_3.

Πρώτο τεταρτημόριο Q_1:

Για την εύρεση του Q_1 έχουμε

$$\frac{i(\nu+1)}{4} = \frac{1(48+1)}{4} = \frac{49}{4} = 12{,}25.$$

Άρα,

$$Q_1 = 34 + 0{,}25(34 - 34) = 34.$$

Τρίτο τεταρτημόριο Q_3:

Για την εύρεση του Q_3 έχουμε

$$\frac{i(\nu+1)}{4} = \frac{3(48+1)}{4} = \frac{3 \cdot 49}{4} = 36{,}75.$$

Άρα.

$$Q_3 = 58 + 0{,}75(59 - 58) = 58{,}75.$$

Το ενδοτεταρτημοριακό εύρος ισούται με

$$H = Q_3 - Q_1 = 58{,}75 - 34 = 24{,}75.$$

Πρώτα υπολογίζουμε την διακύμανση s^2 και στη συνέχεια την τυπική απόκλιση s.

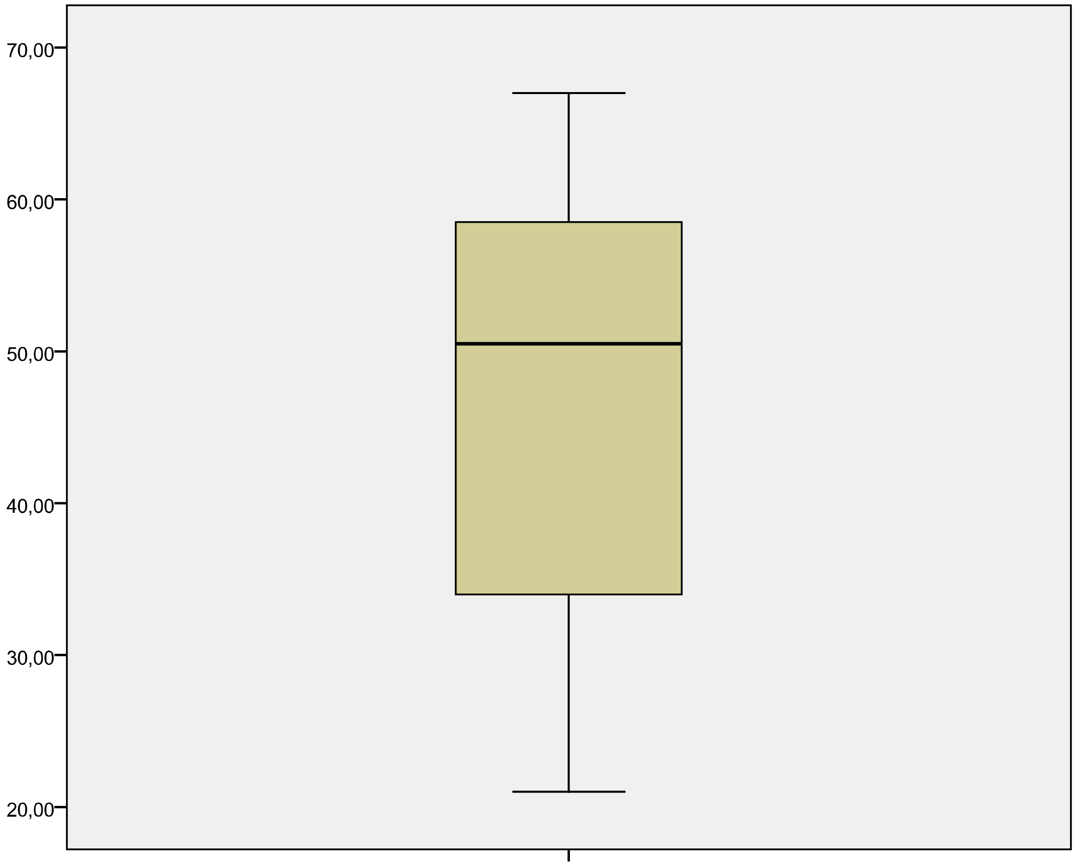
Έχουμε:

$$s^2 = \frac{1}{47}\left[\sum_{i=1}^{32} \nu_i x_i^2 - 48\bar{x}^2\right] = \frac{114445 - 48 \cdot 46,9375^2}{47} = 184,996.$$

Άρα:

$$s = \sqrt{s^2} = \sqrt{184,996} = 13,601.$$

β) Το θηκόγραμμα δίνεται παρακάτω:



Διάγραμμα 6.1: θηκόγραμμα

Από το θηκόγραμμα συμπεραίνουμε ότι η κατανομή των δεδομένων παρουσιάζει αρνητική ασυμμετρία.

γ) Οι μετασχηματισμένες τιμές είναι

$$Y_i = 2 \cdot X_i, , \mathrm{i} = 1,2, \dots ,48.$$

Αριθμητικός μέσος $\bar{y}$:

$$\bar{y} = 2 \cdot \bar{x} = 2 \cdot 46,9375 = 93,8750.$$

Διακύμανση s_Y^2:

$$s_Y^2 = 2^2 \cdot s_X^2 = 4 \cdot 184{,}996 = 739{,}984.$$

Τυπική απόκλιση s_Y:

$$s_Y = 2 \cdot s_X = 2 \cdot 13{,}601 = 27{,}202.$$

Εύρος R_Y:

Το εύρος ισούται με τη διαφορά μεγαλύτερη παρατήρηση - μικρότερη παρατήρηση, δηλαδή:

$$R_Y = Y_{max} - Y_{min} = 2X_{max} - 2X_{min} = 2 \cdot (X_{max} - X_{min}) = 2R_Y = 2 \cdot 46 = 92.$$

δ) Για να εξετάσουμε σε ποιο από τα δύο δείγματα παρατηρείται μεγαλύτερη μεταβλητότητα των μηνιαίων πωλήσεων θα εξετάσουμε ποιο δείγμα έχει μεγαλύτερο συντελεστή μεταβλητότητας. Έχουμε λοιπόν ότι:

$$CV_Y = \frac{s_Y}{\bar{y}} \cdot 100\% = \frac{2s_X}{\bar{x}} \cdot 100\% = \frac{s_X}{\bar{x}} \cdot 100\% = CV_X = \frac{13{,}601}{46{,}9375} = 28{,}9768\%.$$

Παρατηρούμε ότι και τα δύο δείγματα έχουν την ίδια μεταβλητότητα στις μηνιαίες πωλήσεις όπως αποδείξαμε και στον προηγούμενο τύπο.

ΕΦΑΡΜΟΓΗ 5.8

Οι ετήσιες πωλήσεις (Χ) των 15 βιοτεχνιών (αλυσίδα καταστημάτων) που έχει μεγάλη βιοτεχνική επιχείρηση δίνονται σε χιλιάδες ευρώ από τον παρακάτω πίνακα:

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

Πίνακας 6.14

Ζητείται να υπολογιστούν:

α) η μέση τιμή των πωλήσεων των 15 καταστημάτων,

β) το εύρος μεταβολής,

γ) η διακύμανση και η τυπική απόκλιση,

δ) ο συντελεστής μεταβλητότητας.

Μια άλλη επιχείρηση παρουσίασε μέση τιμή πωλήσεων 70 χιλιάδες ευρώ και τυπική απόκλιση 17 με άλλη επιχείρηση-αλυσίδας καταστημάτων-του ίδιου βιοτεχνικού κλάδου. Να συγκρίνεται και να σχολιάσετε τη δραστηριότητα των δύο επιχειρήσεων.

ΛΥΣΗ

α) Κατασκευάζουμε τον πίνακα:

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

Πίνακας 6.15

Έχουμε:

$$\bar{x} = \frac{1}{15}\sum_{i=1}^{7} x_i \nu_i = \frac{1160}{15} = 77{,}333.$$

Συνεπώς, η μέση τιμή των πωλήσεων των 15 καταστημάτων είναι 77333 ευρώ.

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

$$R = 110 - 50 = 60.$$

γ) Συνθέτουμε τον παρακάτω πίνακα αριθμητικών υπολογισμών:

x_i	ν_i	x_i^2	$\nu_i x_i^2$
50	1	2500	2500
60	2	3600	7200
70	4	4900	19600
80	4	6400	25600
90	2	8100	16200
100	1	10000	10000
110	1	12100	12100
Σύνολο	15		93200

Πίνακας 6.16

Έχουμε:

$$s^2 = \frac{1}{14}\left[\sum_{i=1}^{7} \nu_i x_i^2 - 15\bar{x}^2\right] = \frac{93200 - 15 \cdot 77{,}333^2}{14} = 233.$$

Άρα:

$$s = \sqrt{s^2} = \sqrt{233} = 15{,}264.$$

δ) Έχουμε:

$$CV = \frac{s}{\bar{x}} = \frac{15{,}264}{77{,}333} = 19{,}738.$$

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

$$CV' = \frac{17}{70} = 24{,}286.$$

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 ($19{,}738 < 24{,}286$).

ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην παραπάνω πτυχιακή εργασία με θέμα στατιστικά μέτρα και οι εφαρμογές τους στις επιχειρήσεις και την οικονομία αναλύσαμε, ερμηνεύσαμε και επεξεργαστήκαμε γνώσεις που αφορούν το αντικείμενο της στατιστικής ως επιστήμη. Ασχοληθήκαμε με την έννοια της στατιστικής, τα μέτρα θέσης και διασποράς της, το πώς λειτουργεί η στατιστική, που εφαρμόζεται και γιατί εκεί, πόσο ευρύ φάσμα λειτουργικότητας καταλαμβάνει η επιστήμη αυτή σε διαφόρους επιστημονικούς κλάδους και πως μπορεί να συσχετιστεί μαζί τους. Από όλα τα παραπάνω καταλαβαίνουμε ότι η επιστήμη αυτή είναι άπειρη και ότι αυτά που καταγράφηκαν στην εργασία αποτελούν απλά ένα μέρος της. Πιο πάνω αναπτύχθηκαν πέντε κεφάλαια όσον το δυνατόν εκτενέστερα στα οποία αποτυπώθηκαν τόσο θεωρίες όσο και εφαρμογές οι οποίες αφορούν τη δυναμική υπόσταση της επιστήμης αυτή. Η στατιστική είναι μια επιστήμη που απασχολεί όλο τον κόσμο διότι από αυτή απορρέουν πολλά συμπεράσματα μέσα από τύπους και πάρα πολλές αναλύσεις. Μέσω αυτής λύνονται πολλά προβλήματα και δίνονται απαντήσεις σε πολλά ερωτήματα τα οποία μπορεί να αφορούν: τις επιχειρήσεις, τα νοικοκυριά, την οικονομία, την πολιτική, τις τράπεζες, τους φορείς, την κοινωνία, αλλά και το κράτος. Βοηθάει στην ανάπτυξη των προαναφερθέντων διότι η στατιστική πριν συμβεί ένα γεγονός, πριν βγει ένα αποτέλεσμα είναι ικανή να γνωρίζει ποιο είναι αυτό. Για παράδειγμα στις εκλογές η στατιστική υπηρεσία μέσω των στατιστικών μεθόδων και εργαλείων που διαθέτει μπορεί να προβλέψει τα αποτελέσματα των εκλογών . Τέλος είδαμε πολλούς τύπους στα μέτρα θέσης και διασποράς, τους ερμηνεύσαμε και τους συγκρίναμε ως προς τη σημαντικότητα και τη λειτουργικότητα που διαθέτει ο κάθε ένας τύπος από αυτούς.

ΒΙΒΛΙΟΓΡΑΦΙΑ

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

[2] Αποστολόπουλος Κ., Αποστολόπουλος Θ., *Στατιστική επιχειρήσεων: Περιγραφική και επαγωγική στατιστική*, Σύγχρονη Εκδοτική, 2004.

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

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

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

[6] Πανάρετος Ι., Ξεκαλάκης Ε., *Εισαγωγή στη Στατιστική Σκέψη*, *Τόμος Ι*, Αθήνα, 1993.

ΗΛΕΚΤΡΟΝΙΚΕΣ ΠΗΓΕΣ

[1] http://digitalschool.minedu.gov.gr/modules/ebook/show.php/DSGL-C100/493/3203,13012/

Μαθηματικά και στοιχεία στατιστικής, ηλεκτρονικό βιβλίο της Γ' λυκείου.

[2] http://el.wikipedia.org/wiki/%CE%95%CF%80%CE%B9%CF%87%CE%B5%CE%AF%CF%81%CE%B7%CF%83%CE%B7

Επιχείρηση, τελευταία τροποποίηση 05:09, 22 Ιουνίου 2013, Βικιπαίδεια.

[3] http://el.wikipedia.org/wiki/%CE%A3%CF%84%CE%B1%CF%84%CE%B9%CF%83%CF%84%CE%B9%CE%BA%CF%8C%CF%82_%CF%80%CE%BB%CE%B7%CE%B8%CF%85%CF%83%CE%BC%CF%8C%CF%82

Στατιστικός πληθυσμός, τελευταία τροποποίηση 22:22, 31 Μαρτίου 2013, Βικιπαίδεια.

[4]
https://el.wikipedia.org/wiki/%CE%9F%CE%B9%CE%BA%CE%BF%CE%BD%CE%BF%CE%BC%CE%AF%CE%B1

Οικονομία, τελευταία τροποποίηση 14:07, 26 Απριλίου 2013, Βικιπαίδεια.

[5]
http://el.wikipedia.org/wiki/%CE%A3%CF%84%CE%B1%CF%84%CE%B9%CF%83%CF%84%CE%B9%CE%BA%CE%AE

Στατιστική, τελευταία τροποποίηση 20:04, 7 Μαρτίου 2013, Βικιπαίδεια.

[6] http://www.e-conometrics.gr/2013/02/18/vasikes-ennoies/

Βασικές έννοιες, 18 Φεβρουαρίου 2013, εταιρεία e-conometrics.

[7] http://www.math.upatras.gr/~adk/lectures/ida/lab2/slides2.pdf

Ανάλυση δεδομένων, Ελευθέριος Αγγελής.

[8] http://dl.dropboxusercontent.com/u/76699276/statistiki/book.swf

Τα μαθηματικά λίγο αλλιώς, τελευταία τροποποίηση 1 Μαΐου 2013, Τάκης Τσακαλάκος.

[9] http://www.aua.gr/gpapadopoulos/files/perigrafiki091.pdf

Εργαστήριο μαθηματικών & στατιστικής, Γεώργιος Παπαδόπουλος.