ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΜΕΣΟΛΟΓΓΙΟΥ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ & ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Ο ρόλος των στατιστικών μέτρων διασποράς για την λήψη ορθών επιχειρηματικών αποφάσεων

ΛΑΜΠΡΑΚΗ ΕΜΜ. ΜΑΡΙΑ

ΜΙΧΑΛΟΠΟΥΛΟΥ ΜΙΧ. ΧΡΙΣΤΙΝΑ

ΕΙΣΗΓΗΤΗΣ
ΒΙΚΤΩΡΙΑ ΜΠΟΤΑ

ΜΕΣΟΛΟΓΓΙ 2010

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΜΕΣΟΛΟΓΓΙΟΥ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ & ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Ο ρόλος των στατιστικών μέτρων διασποράς για την λήψη ορθών επιχειρηματικών αποφάσεων

ΛΑΜΠΡΑΚΗ ΕΜΜ. ΜΑΡΙΑ (Α.Μ 12007)

ΜΙΧΑΛΟΠΟΥΛΟΥ ΜΙΧ. ΧΡΙΣΤΙΝΑ (13504)

ΕΙΣΗΓΗΤΗΣ
ΒΙΚΤΩΡΙΑ ΜΠΟΤΑ

ΜΕΣΟΛΟΓΓΙ 2010

ΕΙΣΑΓΩΓΗ

Η παρούσα εργασία έχει ως αντικείμενο το ρολό των στατιστικών μέτρων διασποράς στην σωστή λήψη επιχειρηματικών αποφάσεων.

Η ανάλυση του θέματος χωρίζεται σε 3 κεφάλαια.

Το πρώτο κεφάλαιο αναφέρει την ιστορική εξέλιξη της στατιστικής επιστήμης και την χρησιμότητα της μέσα στην επιχείρηση .

Το δεύτερο κεφάλαιο αναφέρονται Μέθοδοι συλλογής στατιστικών στοιχείων ο τρόπος Παρουσίαση Στατιστικών στοιχείων, οι γραφικές παραστάσεις, και τέλος η Ομαδοποίηση των Παρατηρήσεων.

Στο τρίτο κεφάλαιο αναφέρουμε τον τρόπο επεξεργασίας των στατιστικών δεδομένων και τα μέτρα κεντρικής τάσης.

Τέλος στο τέταρτο κεφάλαιο αναλύουμε τα μέτρα διασποράς και αναφέρουμε παραδείγματα.

ΠΕΡΙΕΧΟΜΕΝΑ

	ΣΕΛΙΔΑ
1. Η ιστορική εξέλιξη της Στατιστικής επιστήμης	1
1.2 Ορισμός και βασικές έννοιες της Στατιστικής	2
1.3 Στατιστική και επιχείρηση	8
2. Μέθοδοι συλλογής στατιστικών στοιχείων	10
2.1 Παρουσίαση Στατιστικών στοιχείων	12
2.1.2 Στατιστικοί Πίνακες	13
2.1.3 Γραφική Παράσταση Κατανομής Συχνοτήτων	14
2.2 Ομαδοποίηση των Παρατηρήσεων	20
2.2.1 Κλάσεις Ίσου Πλάτους	24
3.Επεξεργασια στατιστικών δεδομένων	30
3.1 Μέτρα Κεντρικής Τάσης	30
3.1.1 Αριθμητικός Μέσος – Arithmetic Mean	31
3.1.1.1 Αστάθμητος Αριθμητικός Μέσος	31
3.1.1.2 Ιδιότητες Αριθμητικού Μέσου	37
3.1.1.3 Γενικά Χαρακτηριστικά Αριθμητικού Μέσου	38
3.2 Διάμεσος - Median	39
3.3 Τεταρτημόρια (quartiles)	45
3.4 Δεκατημόρια - Εκατοσημόρια	47
3.5 Χαρακτηριστικά Διαμέσου, Τεταρτημορίου κλπ.	51
3.6 Σημείο Μέγιστης Συχνότητας – Τύπο- επικρατούσα τιμή – Mode	52
3.6.1 Γενικά χαρακτηριστικά Τύπου	55
4. Μέτρα Διασποράς – Dispersion	57
4.1 Εύρος - Range	58
4.1.2 Ενδοτεταρτημοριακό Εύρος – Interquartile Range	59

4.2. Μέση Απόλυτη Απόκλιση – Mean Deviation 61

4.3 Διακύμανση και Τυπική Απόκλιση – Variance and Standard Deviation 64

4.3.2 Διόρθωση Διακύμανσης κατά Sheppard 72

4.4 Σχετική Διασπορά – Relative Dispersion 74

4.5 Μέση διαφορά του Gini 77

Επαναληπτικές ασκήσεις 78

1. Η ιστορική εξέλιξη της Στατιστικής επιστήμης

Οι πρώτες προσπάθειες συλλογής στατιστικών στοιχείων ανάγονται στην αρχαιότητα. Απογραφές του πληθυσμού έκαναν όλοι σχεδόν οι λαοί του αρχαίου κόσμου, όπως οι Κινέζοι οι Αιγύπτιοι οι Έλληνες και άλλοι. Ο όρος Στατιστική προέρχεται από την Λατινική λέξη status, που σημαίνει κράτος και η οποία προέρχεται από το ελληνικό ρήμα στατίζω που σημαίνει διαπιστώνω, προσδιορίζω.

Στατιστικά στοιχεία για τον πληθυσμό, τις στρατιωτικές δυνάμεις, το εισόδημα, τη φορολογία αναφέρονται στα έργα του Ηροδότου, του Θουκυδίδη, του Ξενοφώντα και του Αριστοτέλη.

Οι Ρωμαίοι συγκέντρωναν στατιστικά στοιχεία για όλους σχεδόν τους τομείς της ανθρώπινης δραστηριότητας, όπως π.χ. δημογραφικά στοιχεία, δηλαδή στοιχεία που αναφέρονταν στις γεννήσεις, τους θανάτους και σε άλλα χαρακτηριστικά του πληθυσμού, αλλά και στοιχεία που αναφέρονταν σε φόρους, πολέμους, γεωργική δραστηριότητα, αθλητικά γεγονότα κλπ.

Η Αγία Γραφή αναφέρει ότι ο Ιωσήφ και η Μαρία πήγαν στη Βηθλεέμ, για τη γενική απογραφή του πληθυσμού που έγινε από τους Ρωμαίους κατά την περίοδο που γεννήθηκε ο Χριστός.

Η στατιστική ως αυτοτελής επιστήμη παρουσιάζεται από το $17^{ο}$ αιώνα. Τότε άρχισε να διαμορφώνεται ένας νέος κλάδος που προήλθε από τη μελέτη τυχερών παιχνιδιών, γνωστός σήμερα ως θεωρία των πιθανοτήτων. Η θεωρία των πιθανοτήτων θεμελιώθηκε κυρίως από τους Bernoulli,Gauss, Laplace. Έτσι κατά το τέλος του $19^{ου}$ αιώνα η Στατιστική είχε αποκτήσει το κατάλληλο θεωρητικό υπόβαθρο για πάρα πέρα ανάπτυξή της.

1.2 Ορισμός και βασικές έννοιες της Στατιστικής

Τώρα αν θα θέλαμε να δώσουμε τον **ορισμό** της **Στατιστικής**, θα λέγαμε ότι είναι η επιστήμη που αναπτύσσει και προσφέρει σε κάθε ενδιαφερόμενο μεθόδους:

- **Συλλογής**
- **Ταξινόμησης**
- **Επεξεργασίας**
- **Παρουσίασης**
- **Ανάλυσης** πληροφοριών, πάνω σε οποιαδήποτε

δραστηριότητα του ανθρώπου ή φαινομένου της φύσης με σκοπό της εξαγωγή συμπερασμάτων που χρησιμεύουν στην λήψη **αποφάσεων**.

Σχήμα 1

Πρόβλημα

Συλλογή των απαιτούμενων πληροφοριών

Επεξεργασία και οργάνωση των δεδομένων

Συστηματική παρουσίαση των στατιστικών στοιχείων

Ανάλυση των δεδομένων και εξαγωγή συμπερασμάτων

Άρα Στατιστική είναι ένα σύνολο αρχών και μεθοδολογιών για:

- το σχεδιασμό της διαδικασίας συλλογής δεδομένων
- τη συνοπτική και αποτελεσματική παρουσίασή τους
- την ανάλυση και εξαγωγή αντίστοιχων συμπερασμάτων.

Ο κλάδος της Στατιστικής που ασχολείται με τον πρώτο στόχο λέγεται **σχεδιασμός πειραμάτων** (experimental design) ενώ, με τον δεύτερο ασχολείται η **περιγραφική στατιστική** (descriptive statistics), που αποτελεί και το αντικείμενο μελέτης μας στη συνέχεια. Τέλος, η **επαγωγική στατιστική** ή **στατιστική συμπερασματολογία** (inferential statistics) περιλαμβάνει τις μεθόδους με τις οποίες γίνεται η προσέγγιση των χαρακτηριστικών ενός μεγάλου συνόλου δεδομένων, με τη μελέτη των χαρακτηριστικών ενός μικρού υποσυνόλου των δεδομένων.

Έτσι αν, για παράδειγμα, ο Διευθυντής ενός σχολείου εξετάζοντας ένα δείγμα 100 απουσιών των μαθητών από το σύνολο των απουσιών ενός τριμήνου αναφέρει στο σύλλογο των καθηγητών ότι 20 από τις 100 απουσίες είναι αδικαιολόγητες, τότε απλώς **περιγράφει** αυτό που παρατήρησε. Αν όμως αναφέρει ότι το 20% των απουσιών είναι αδικαιολόγητες, τότε **συμπεραίνει** ότι το ποσοστό των απουσιών όλων των μαθητών του σχολείου είναι (περίπου) το ίδιο με αυτό του δείγματος. Προβαίνει δηλαδή σε μια **επαγωγή** από το δείγμα στον πληθυσμό.

Η Στατιστική σήμερα χρησιμοποιείται ευρύτατα σε όλους σχεδόν τους τομείς της ανθρώπινης δραστηριότητας. Βασικές έννοιες της Στατιστικής έχουν εισχωρήσει και ενσωματωθεί σε όλες τις επιστήμες. Από τις Ανθρωπιστικές, Νομικές και Κοινωνικές Επιστήμες (Αρχαιολογία, Λαογραφία, Κοινωνιολογία, Δημογραφία, …), τις Φυσικές Επιστήμες (Φυσική, Χημεία, Αστρονομία, …), τις Επιστήμες Υγείας (Ιατρική, Φαρμακευτική, Βιολογία,…), τις Τεχνολογικές Επιστήμες (Μηχανολογία, Τοπογραφία, Ναυπηγική, …) μέχρι τις Επιστήμες Οικονομίας και Διοίκησης (Οικονομικά, Χρηματιστηριακά, Διαφήμηση, Marketing, …), βλέπουμε να υπεισέρχεται η Στατιστική είτε με την αρχική περιγραφική μορφή της είτε με τις προηγμένες αναλυτικές τεχνικές της.

Η ανάλυση στατιστικών ερευνών είναι το κυριότερο εργαλείο έρευνας σε ένα μεγάλο φάσμα εφαρμογών των παραπάνω επιστημών. Οι έρευνες των ανθρώπινων πληθυσμών (συχνά αναφερόμενες και ως **δημοσκοπήσεις**) αποτελούν σπουδαίες πηγές βασικής γνώσης των κοινωνικών επιστημών. Οικονομολόγοι, ψυχολόγοι, κοινωνιολόγοι και πολιτικοί επιστήμονες μελετούν ποικίλα θέματα όπως πρότυπα εσόδων-εξόδων των οικογενειών και των επιχειρήσεων, την επίδραση της επαγγελματικής απασχόλησης των γυναικών στην οικογενειακή ζωή, τις συγκοινωνιακές και ταξιδιωτικές συνήθειες των κατοίκων μιας πόλης, τις προτιμήσεις των ψηφοφόρων για τους υποψηφίους και τις θέσεις τους. Πολλά προβλήματα που αντιμετωπίζουν σήμερα οι επιχειρήσεις αφορούν τη διατήρηση, αντικατάσταση ή το κρίσιμο σημείο αντοχής συσκευών ή προσωπικού.

Ο διευθυντής μιας βιομηχανίας πρέπει να είναι σε θέση να κατανοεί στατιστικές έρευνες που αφορούν την ποιότητα του προϊόντος και την αποδοτικότητα της παραγωγικής διαδικασίας. Πρέπει επίσης να αντιλαμβάνεται την αποτελεσματικότητα της διαφήμισης και τις προτιμήσεις του καταναλωτή σε μια έρευνα αγοράς. Συμβουλευόμενος και τον στατιστικό μπορεί να πάρει σωστές αποφάσεις αναφορικά με την επέκταση ή μη της επιχείρησης. Σήμερα κάθε γιατρός πρέπει να έχει βασικές γνώσεις Στατιστικής που θα τον βοηθήσουν τόσο στην έρευνα όσο και στην καθημερινή άσκηση του κάθε μορφής και είδους ιατρικού ή βιοϊατρικού, γενικότερα, επαγγέλματος. Η Εθνική Στατιστική Υπηρεσία κάθε χώρας διενεργεί σε τακτά χρονικά διαστήματα δειγματοληπτικές έρευνες, για να πάρει πληροφορίες για τον πληθωρισμό, την απασχόληση και την ανεργία στη χώρα. Ανάλογα με τα αποτελέσματα διαμορφώνεται και η κυβερνητική πολιτική στα θέματα αυτά.

Πέρα από όλα αυτά, διαπιστώνουμε ολοένα και περισσότερο να γίνεται χρήση μεθόδων της Στατιστικής για την υποστήριξη διάφορων θέσεων. Ακόμα και σε τηλεοπτικές αντιπαραθέσεις (κυρίως σε προεκλογικές περιόδους) βλέπουμε τους συνομιλητές να κάνουν χρήση αριθμών, στατιστικών στοιχείων, γραφημάτων και διαγραμμάτων, για να δώσουν εγκυρότητα στις απόψεις τους και να πείσουν για τα λεγόμενά τους.

Παραπάνω έχουν αναφερθεί ελάχιστα από τα πεδία εφαρμογών της Στατιστικής. Προφανώς μια λεπτομερής περιγραφή όλων των εφαρμογών δεν είναι δυνατή. Η μελέτη όμως και η γνώση της Στατιστικής βοηθά όχι μόνο στη σωστή χρήση των γνωστών μεθόδων αλλά και στην ανάπτυξη νέων τεχνικών για την αποτελεσματικότερη εξαγωγή χρήσιμων συμπερασμάτων.

Για να μπορέσουμε να κατανοήσουμε όσα θα συναντήσουμε στα επόμενα κεφαλαία είναι απαραίτητο να γνωρίσουμε ορισμένες βασικές έννοιες της Στατιστικής.

1. **Δεδομένα:** είναι τα στοιχεία τα οποία συλλέγονται, αναλύονται και συνοψίζονται για παρουσίαση και ερμηνεία. Δεδομένα αποκτώνται με τη συλλογή των μετρήσεων κάθε μεταβλητής για κάθε στοιχείο και μπορεί να είναι αριθμοί, λέξεις ή και σύμβολα. Το σύνολο των μετρήσεων που συγκεντρώθηκαν για κάθε στοιχείο ονομάζεται **παρατήρηση**. Κατά τη συλλογή των δεδομένων υπάρχει πιθανότητα να πάρουμε λανθασμένα στοιχεία. Είναι προτιμότερο να χρησιμοποιούμε, για σωστότερα αποτελέσματα, μόνο αυτά για τα οποία είμαστε σίγουροι.
2. **Άτομο- Στοιχείο :** είναι η μονάδα βάσεως, την οποία παρατηρεί ο στατιστικολόγος. Πχ ένα ημερομίσθιο σε μια επιχείρηση, ένας εργαζόμενος σε μια επιχείρηση ή ακόμα και ένα αντικείμενο ή μια γνώμη.
3. **Πληθυσμός** : είναι το σύνολο των εξεταζόμενων ατόμων σε μία ερευνά. Έτσι πληθυσμός μπορεί να είναι το σύνολο των ξενοδοχείων της χώρας ή το σύνολο των ποδοσφαιρικών ομάδων της Ελλάδας ή ακόμη το σύνολο των ημερομισθίων μιας επιχείρησης. Ο στατιστικός πληθυσμός μπορεί να είναι **άπειρος,** δηλαδή να περιλαμβάνει άπειρο πλήθος ατόμων ή **πεπερασμένος,** δηλαδή να περιλαμβάνει ορισμένο και αριθμήσιμο πλήθος ατόμων.
4. **Δείγμα** : είναι κάθε γνήσιο υποσύνολο του πληθυσμού. Πχ το σύνολο ξενοδοχείων του νόμου Καβάλας ή το σύνολο των ομάδων ποδοσφαίρου της Α Εθνικής είναι ένα δείγμα των πληθυσμών που αναφέραμε παραπάνω.

5. **Πείραμα :** είναι η σχεδιασμένη ενέργεια η οποία έχει ως αποτέλεσμα τη δημιουργία σετ δεδομένων.
6. **Μεταβλητή :** είναι τα χαρακτηριστικά ως προς τα οποία εξετάζουμε έναν πληθυσμό και τις συμβολίζουμε συνήθως με τα κεφαλαία γράμματα. Οι δυνατές τιμές που μπορεί να πάρει μια μεταβλητή λέγονται **τιμές της μεταβλητής**., Τις μεταβλητές τις διακρίνουμε:

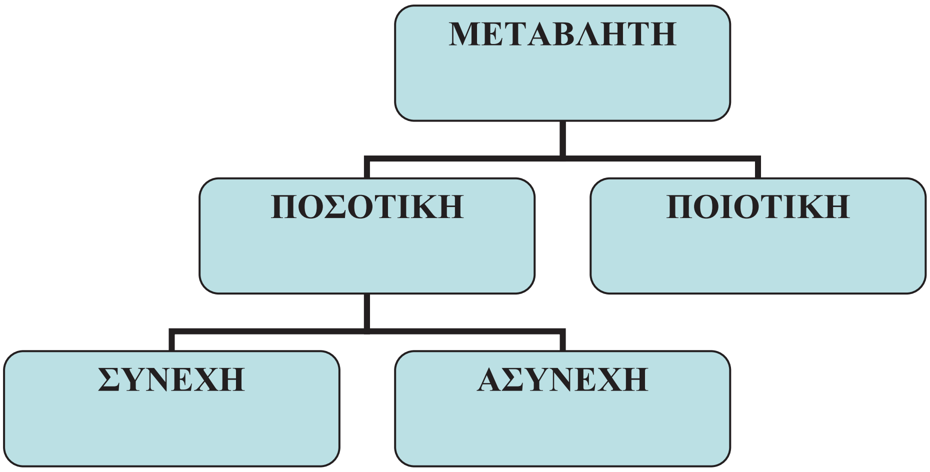
1. Σε **ποιοτικές** ή **κατηγορικές** μεταβλητές, των οποίων οι τιμές τους δεν είναι αριθμοί. Τέτοιες είναι, για παράδειγμα, η ομάδα αίματος , το φύλο όπως επίσης και η οικονομική κατάσταση και η υγεία των ανθρώπων (που μπορεί να χαρακτηριστεί ως κακή, μέτρια, καλή ή πολύ καλή).

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

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

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

Σχήμα 2



1.3 Στατιστική και επιχείρηση

Η χρήση της στατιστικής είναι ευρύτατα διαδεδομένη και εφαρμόσιμη, ουσιαστικά σε όλες τις λειτουργίες μιας επιχείρησης, όπως το Μάρκετινγκ και οι Πωλήσεις, η Παραγωγή, ο Έλεγχος Αποθεμάτων, η Χρηματοοικονομική Διοίκηση και η Λογιστική

Η στατιστική δεν είναι σε καμία περίπτωση μια μαθηματική μεθοδολογία, που μπορεί να «ντύσει» τα οποιαδήποτε δεδομένα με «επιστημονικοφάνεια», αποδεικνύοντας σχεδόν οτιδήποτε θα ήθελε κάποιος να αποδείξει. Απεναντίας, όπως θα δειχθεί στη συνέχεια, η Στατιστική Επιστήμη έχει πολύ αυστηρούς κανόνες που διασφαλίζουν ότι μια πληροφόρηση δεν στηρίζεται σε ελλιπή δεδομένα. Αυτό, βεβαίως, δεν αποκλείει μερικές φορές την παραβίαση των κανόνων χρήσης των στατιστικών τεχνικών (από άγνοια ή και εσκεμμένα) και την παρουσίαση αβάσιμων συμπερασμάτων.

Η Στατιστική (Περιγραφική και Επαγωγική) που αναφέρεται σε δεδομένα πάνω στην :

- Απασχόληση
- Παραγωγή
- Τιμές
- Πωλήσεις κλπ

είναι πολύ χρήσιμη στις βασικές διοικητικές εργασίες όπως :

- Προοπτικές
- Εκτίμηση συμπεριφοράς
- Εντοπισμό αδυναμιών,

αποτελεί τον ειδικό κλάδο της Στατιστικής των Επιχειρήσεων που θα μας απασχολήσει στην συνέχεια .

Εφαρμογές της Περιγραφικής Στατιστικής στις Επιχειρήσεις και στην Οικονομία :

✓ Η αποτύπωση ποιοτικών και ποσοτικών χαρακτηριστικών της επιχείρησης. Μια επιχείρηση θέλει να έχει κάθε στιγμή διαθέσιμα στοιχεία σχετικά με τον όγκο των παραγόμενων προϊόντων, τις πωλήσεις, τις προμήθειες, τις ηλικίες και το φύλο των εργαζόμενων, τις αμοιβές κτλ (απλά Περιγραφικά Στατιστικά μέτρα).

✓ Η προσπάθεια προσδιορισμού των παραγόντων οι οποίοι επιδρούν στις πωλήσεις ή στην τιμή ενός προϊόντος ή στην απόδοση των εργαζόμενων. Θέλουμε να ξέρουμε, για παράδειγμα, αν απόδοση του εργαζόμενου εξαρτάται από το μισθό του, την ηλικία του τη μόρφωσή του κτλ (Απλή ή πολλαπλή παλινδρόμηση- συσχέτιση).

- ✓ Η μεταβολή στην παραγωγικότητα, την απόδοση, τα κέρδη, την τιμή, την ποσότητα. Θέλουμε να γνωρίζουμε, για παράδειγμα, την ποσοστιαία μεταβολή στην παραγωγικότητα ή τα κέρδη της επιχείρησης μας (αριθμοδείκτες) .

Η Στατιστική είναι ένα εργαλείο που, αν χρησιμοποιηθεί σωστά, μπορεί να δώσει μια ακριβέστερη εικόνα της πραγματικότητας και μια πιο σαφή κατάσταση.

2. Μέθοδοι συλλογής στατιστικών στοιχείων

Το πρώτο στάδιο για τη στατιστική μελέτη ενός προβλήματος, είναι η συλλογή των αριθμητικών στοιχείων, δηλαδή η συγκέντρωση των παρατηρήσεων που αναφέρονται στις μεταβλητές, ως προς τις οποίες πρόκριτε να εξετάσουμε το πληθυσμό.

Η συλλογή του στατιστικού υλικού αποτελεί θεμέλιο πάνω στο οποίο στηρίζεται κάθε στατιστική έρευνα. Αν τα στατιστικά δεδομένα είναι ακριβή τότε θα προκύψουν σωστά άρα και χρήσιμα συμπεράσματα.

Η συλλογή των στατιστικών δεδομένων γίνεται με δύο κυρίως μεθόδους, που είναι η **απογραφή** και η **δειγματοληψία.**

Όταν οι παρατηρήσεις προκύπτουν από όλες τις μονάδες του πληθυσμού η έρευνα χαρακτηρίζετε ως απογραφή.

Όταν οι παρατηρήσεις προκύπτουν από ένα υποσύνολο του αρχικού πληθυσμού με σκοπό τη γενίκευση των συμπερασμάτων της περιορισμένης αυτής έρευνας για ολόκληρο τον πληθυσμό , τότε η έρευνα ονομάζεται **δειγματοληψία** και το εξεταζόμενο υποσύνολο **δείγμα.**

Σε περιπτώσεις στατιστικών ερευνών, η απογραφή μεγάλων κυρίως πληθυσμών στοιχίζει σημαντικά σε χρόνο και σε χρήμα, γεγονός που αποθαρρύνει τη μέθοδο της απογραφής πχ η ακροαματικότητα ενός σταθμού, η δημοτικότητα ενός πολιτικού, το ποσοστό των ψηφοφόρων ενός κόμματος κτλ προκύπτουν από δειγματοληπτικές έρευνες.

Η επιλογή του αντιπροσωπευτικού δείγματος είναι “εκ των ων ουκ άνευ”. Αποτελεί πολύ σοβαρή και δύσκολη διαδικασία. Ο κακός σχεδιασμός και η εκτέλεση της στατιστικής έρευνας, η μη αντιπροσωπευτικότητα του δείγματος, ο μη σωστός καθορισμός του μεγέθους του δείγματος αποτελούν μερικά βασικά μειονεκτήματα στη διαδικασία επιλογής ενός δείγματος. Από την άλλη πλευρά, στις απογραφές απαιτείται συνήθως μεγάλος αριθμός απογραφέων. Παρουσιάζεται έτσι η ανάγκη πρόσληψης και εκπαίδευσης μεγάλου αριθμού υπαλλήλων. Λόγω του μεγάλου χρόνου και κυρίως των σημαντικών εξόδων που απαιτούνται, πολλές φορές χρησιμοποιούνται ανεπαρκώς εκπαιδευμένοι απογραφείς με κίνδυνο να σημειώνονται λάθη οφειλόμενα σ’αυτούς.

Αξίζει να σημειωθεί ότι μία “προσεκτική” επιλογή μικρότερου δείγματος είναι δυνατόν να δώσει καλύτερα αποτελέσματα από ένα μεγαλύτερο δείγμα που δεν έχει εκλεγεί κατάλληλα. Ενδεικτικό είναι το παράδειγμα των προεδρικών εκλογών των ΗΠΑ το 1936. Το περιοδικό Literary Digest χρησιμοποιώντας δείγμα 2.400.000 ατόμων πρόβλεψε νίκη του Landon με ποσοστό 57%. Αντίθετα, το δημοσκοπικό γραφείο του G. Gallup χρησιμοποιώντας δείγμα 50.000 ατόμων πρόβλεψε το σωστό αποτέλεσμα που ήταν νίκη του Roosvelt με ποσοστό 62%! Η παταγώδης αποτυχία της δημοσκόπησης του περιοδικού οφειλόταν στο γεγονός ότι το δείγμα που επελέγη δεν ήταν αντιπροσωπευτικό του πληθυσμού.

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

2.1 Παρουσίαση Στατιστικών στοιχείων

Σε μερικές περιπτώσεις, ο αντικειμενικός σκοπός της Στατιστικής είναι η κατά περιγραφικό τρόπο διερεύνηση του στατιστικού υλικού, η παρουσίαση των στατιστικών μέτρων και η εξαγωγή συμπερασμάτων, τα οποία αναφέρονται μόνο σε ένα τμήμα του στατιστικού πληθυσμού χωρίς να γενικεύονται τα συμπεράσματα για το σύνολο του πληθυσμού. Στις περιπτώσεις αυτές λέμε ότι η διερεύνηση των στατιστικών δεδομένων γίνεται με μεθόδους και τεχνικές της Περιγραφικής Στατιστικής.

Οι κυριότερες μέθοδοι της Περιγραφικής Στατιστικής είναι :

i. **Η ταξινόμηση** των στατιστικών δεδομένων σε πίνακες Κατανομών Συχνοτήτων και η απεικόνιση τους με Στατιστικά Διαγράμματα.

ii. **Ο υπολογισμός** των βασικών στατιστικών μέτρων των διαφόρων κατανομών συχνοτήτων (πχ μέσων όρων, μέτρα διασποράς, ασυμμετρίας κτλ)

iii. **Η ανάλυση των χρονολογικών σειρών**, για τη διαπίστωση, περιγραφή και μελέτη της διαχρονικής εξέλιξης διαφόρων φαινομένων.

2.1.2 Στατιστικοί Πίνακες

Μετά τη συλλογή των στατιστικών δεδομένων είναι αναγκαία η κατασκευή συνοπτικών **πινάκων** ή **γραφικών παραστάσεων**, ώστε να είναι εύκολη η κατανόησή τους και η εξαγωγή σωστών συμπερασμάτων. Η παρουσίαση των στατιστικών δεδομένων σε πίνακες γίνεται με την κατάλληλη τοποθέτηση των πληροφοριών σε γραμμές και στήλες, με τρόπο που να διευκολύνεται η σύγκριση των στοιχείων και η καλύτερη ενημέρωση του αναγνώστη σχετικά με τη δομή του πληθυσμού που ερευνάμε.

Οι πίνακες διακρίνονται στους:

α) **γενικούς πίνακες**, οι οποίοι περιέχουν όλες τις πληροφορίες που προκύπτουν από μία στατιστική έρευνα (συνήθως με αρκετά λεπτομερειακά στοιχεία) και αποτελούν πηγές στατιστικών πληροφοριών στη διάθεση των επιστημόνων-ερευνητών για παραπέρα ανάλυση και εξαγωγή συμπερασμάτων,

β) **ειδικούς πίνακες**, οι οποίοι είναι συνοπτικοί και σαφείς. Τα στοιχεία τους συνήθως έχουν ληφθεί από τους γενικούς πίνακες.

Κάθε πίνακας που έχει κατασκευαστεί σωστά πρέπει να περιέχει:

α) τον **τίτλο**, που γράφεται στο επάνω μέρος του πίνακα και δηλώνει με σαφήνεια και συνοπτικά το περιεχόμενο του πίνακα,

β) τις **επικεφαλίδες** των γραμμών και στηλών, που δείχνουν συνοπτικά τη φύση και τις μονάδες μέτρησης των δεδομένων,

γ) το **κύριο σώμα** (κορμό), που περιέχει διαχωρισμένα μέσα στις γραμμές και στις στήλες τα στατιστικά δεδομένα, δ) την **πηγή**, που γράφεται στο κάτω μέρος του πίνακα και δείχνει την προέλευση των στατιστικών στοιχείων, έτσι ώστε ο αναγνώστης να ανατρέχει σ'αυτήν, όταν επιθυμεί, για επαλήθευση στοιχείων ή για λήψη περισσότερων πληροφοριών.

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

Πίνακας 1
Εργατικά ατυχήματα κατά ομάδες ηλικιών
Έτη 1990-94

Ηλικία	1990	1991	1992	1993	1994
Κάτω των	18	9	10	16	5
15	731	564	437	735	442
15-19	3323	2785	2755	2981	2696
20-24	4277	3921	4246	3881	3717
25-29	3952	3700	3388	3348	3282
30-34	3589	3146	3233	3230	3000
35-39	3237	2803	2911	2880	2903
40-44	2839	2593	2784	2608	2403
45-49	2727	2564	2286	2095	1877
50-54	2304	2230	2185	1699	1664
55-59	728	720	688	420	523
60-64	121	150	140	66	96
65-69					
Σύνολο	27846	25185	25063	23959	22608

Πηγή: ΙΚΑ, Ελληνικό Ινστιτούτο Υγιεινής και Ασφάλειας της Εργασίας

2.1.3 Γραφική Παράσταση Κατανομής Συχνοτήτων

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

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

α) Ραβδόγραμμα

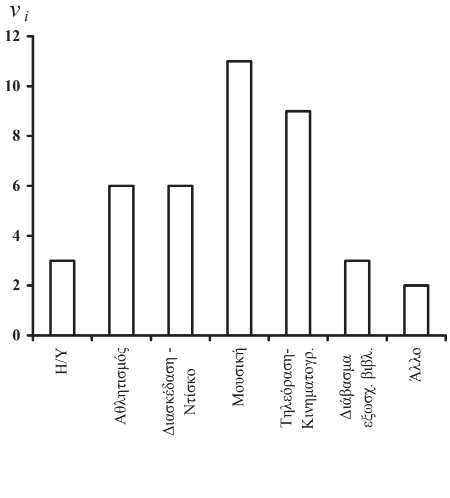
Το **ραβδόγραμμα** (barchart) χρησιμοποιείται για τη γραφική παράσταση των τιμών μιας ποιοτικής μεταβλητής. Το ραβδόγραμμα αποτελείται από ορθογώνιες στήλες που οι βάσεις τους βρίσκονται πάνω στον οριζόντιο ή τον κατακόρυφο άξονα. Σε κάθε τιμή της μεταβλητής X αντιστοιχεί μια ορθογώνια στήλη της οποίας το ύψος είναι ίσο με την αντίστοιχη συχνότητα ή σχετική συχνότητα. Έτσι έχουμε αντίστοιχα το **ραβδόγραμμα συχνοτήτων** και το **ραβδόγραμμα σχετικών συχνοτήτων**. Τόσο η απόσταση μεταξύ των στηλών όσο και το μήκος των βάσεών τους καθορίζονται αυθαίρετα. Στον πίνακα 7 έχουμε την κατανομή συχνοτήτων της μεταβλητής X: "απασχόληση στον ελεύθερο χρόνο" και στα σχήματα 1(α), (β) τα αντίστοιχα ραβδογράμματα συχνοτήτων και σχετικών συχνοτήτων.

Μερικές φορές σε ένα ραβδόγραμμα συχνοτήτων ο ρόλος των δύο αξόνων είναι δυνατόν να αντιστραφεί, όπως φαίνεται στο σχήμα 1(β), που παριστάνεται το ραβδόγραμμα σχετικών συχνοτήτων της ίδιας μεταβλητής. Αν θέλουμε να συγκρίνουμε τον τρόπο που περνούν τον ελεύθερο χρόνο τους τα αγόρια και τα κορίτσια, τότε κατασκευάζουμε το ραβδόγραμμα σχετικών συχνοτήτων του σχήματος 1(γ), όπως προκύπτει από τον πίνακα 4.

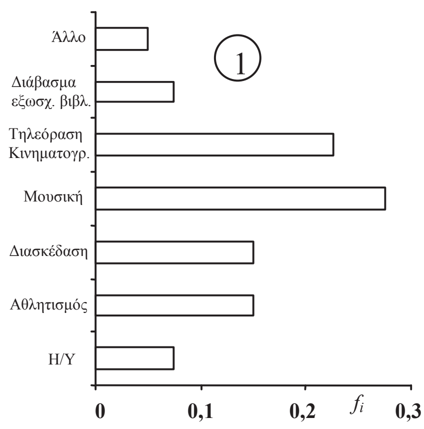
Πίνακας 7

Κατανομή συχνοτήτων για την απασχόληση στον ελεύθερο χρόνο τους των εργαζομένων μιας επιχείρησης του πίνακα 4.

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

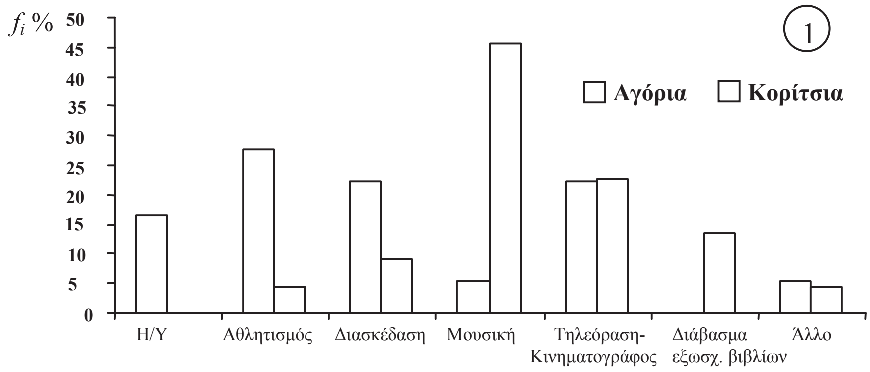


(α)



(β)

Ραβδόγραμμα συχνοτήτων (α) και σχετικών συχνοτήτων (β) για την απασχόληση των εργαζομένων μιας επιχείρησης του πίνακα 7.



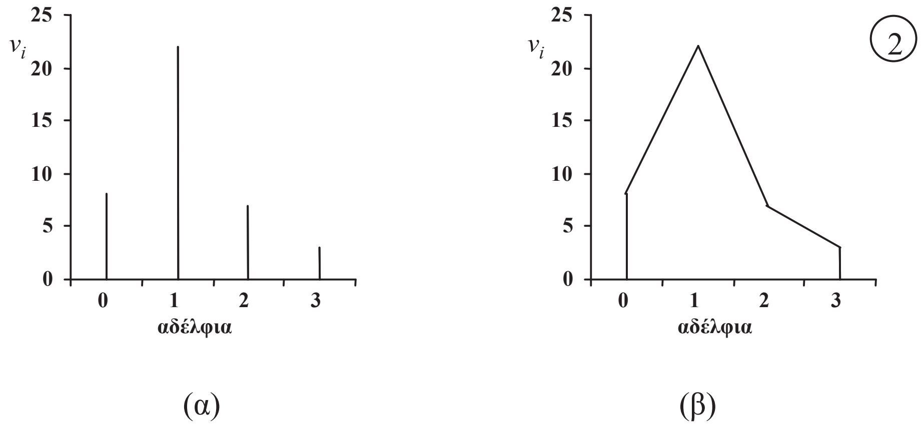
(γ)

Ραβδόγραμμα σχετικών συχνοτήτων για την απασχόληση των εργαζομένων του πίνακα 4 ανάλογα με το φύλο.

β) Διάγραμμα Συχνοτήτων

Στην περίπτωση που έχουμε μια ποσοτική μεταβλητή αντί του ραβδογράμματος χρησιμοποιείται το **διάγραμμα συχνοτήτων** (line diagram). Αυτό μοιάζει με το ραβδόγραμμα με μόνη διαφορά ότι αντί να χρησιμοποιούμε συμπαγή ορθογώνια υψώνουμε σε κάθε x_i (υποθέτοντας ότι $x_1 < x_2 < ... < x_\kappa$) μία κάθετη γραμμή με μήκος ίσο προς την αντίστοιχη συχνότητα, όπως φαίνεται στο σχήμα 2(α). Μπορούμε επίσης αντί των συχνοτήτων v_i στον κάθετο άξονα να βάλουμε τις σχετικές συχνότητες f_i, οπότε έχουμε το **διάγραμμα σχετικών συχνοτήτων**.

Ενώνοντας τα σημεία (x_i, v_i) ή (x_i, f_i) έχουμε το λεγόμενο **πολύγωνο συχνοτήτων** ή **πολύγωνο σχετικών συχνοτήτων**, αντίστοιχα, που μας δίνουν μια γενική ιδέα για τη μεταβολή της συχνότητας ή της σχετικής συχνότητας όσο μεγαλώνει η τιμή της μεταβλητής που εξετάζουμε, βλέπε σχήμα 2(β).



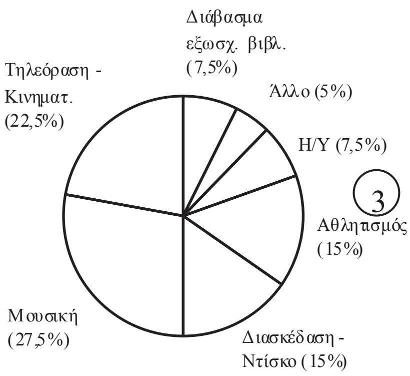
Διάγραμμα συχνοτήτων (α) και πολύγωνο συχνοτήτων (β) για τη μεταβλητή "αριθμός αδελφών" του πίνακα 4.

γ) Κυκλικό Διάγραμμα

Το **κυκλικό διάγραμμα** (piechart) χρησιμοποιείται για τη γραφική παράσταση τόσο των ποιοτικών όσο και των ποσοτικών δεδομένων, όταν οι διαφορετικές τιμές της μεταβλητής είναι σχετικά λίγες. Το κυκλικό διάγραμμα είναι ένας κυκλικός δίσκος χωρισμένος σε κυκλικούς τομείς, τα εμβαδά ή ,ισοδύναμα, τα τόξα των οποίων είναι ανάλογα προς τις αντίστοιχες συχνότητες v_i ή τις σχετικές συχνότητες f_i των τιμών x_i της μεταβλητής. Αν συμβολίσουμε με α_i το αντίστοιχο τόξο ενός κυκλικού τμήματος στο κυκλικό διάγραμμα συχνοτήτων, τότε

$$\alpha_i = v_i \frac{360^o}{v} = 360^o f_i \quad \text{για} \quad i = 1,2,...,\kappa \,.$$

Στο σχήμα 3 παριστάνεται το αντίστοιχο κυκλικό διάγραμμα σχετικών συχνοτήτων της "απασχόλησης των εργαζομένων μιας επιχείρησης" για τα δεδομένα του πίνακα 4.



Κυκλικό διάγραμμα σχετικών συχνοτήτων της απασχόλησης των εργαζομενών για τα δεδομένα του πίνακα 4.

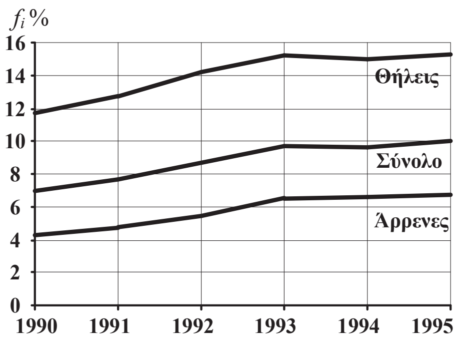
δ) Σημειόγραμμα

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

χρόνος (σε λεπτά)

ε) Χρονόγραμμα.

Το χρονόγραμμα ή χρονολογικό διάγραμμα χρησιμοποιείται για τη γραφική απεικόνιση της διαχρονικής εξέλιξης ενός οικονομικού, δημογραφικού ή άλλου μεγέθους. Ο οριζόντιος άξονας χρησιμοποιείται συνήθως ως άξονας μέτρησης του χρόνου και ο κάθετος ως άξονας μέτρησης της εξεταζόμενης μεταβλητής. Στο σχήμα 5 έχουμε το χρονόγραμμα του ποσοστού ανεργίας στη χώρα μας από το 1990 έως το 1995. (Πηγή ΕΣΥΕ).



5

Παρατηρούμε ότι στο γυναικείο πληθυσμό υπάρχει συστηματικά μεγαλύτερο ποσοστό ανεργίας, γύρω στις 8 εκατοστιαίες μονάδες. Στο διάστημα 1993-95 το ποσοστό ανεργίας έχει σταθεροποιηθεί γύρω στο 6,5% για τους άνδρες και γύρω στο 15% για τις γυναίκες.

2.2 Ομαδοποίηση των Παρατηρήσεων

Οι πίνακες συχνοτήτων και κατ’αναλογίαν τα αντίστοιχα διαγράμματα είναι δύσκολο να κατασκευαστούν, όταν το πλήθος των τιμών μιας μεταβλητής είναι αρκετά μεγάλο. Αυτό μπορεί να συμβεί είτε στην περίπτωση μιας διακριτής μεταβλητής είτε, πολύ περισσότερο, στην περίπτωση μιας συνεχούς μεταβλητής, όπου αυτή μπορεί να πάρει οποιαδήποτε τιμή στο διάστημα ορισμού της. Σ’αυτές τις περιπτώσεις είναι απαραίτητο να ταξινομηθούν (ομαδοποιηθούν) τα δεδομένα σε μικρό πλήθος ομάδων, που ονομάζονται και **κλάσεις** (class intervals), έτσι ώστε κάθε τιμή να ανήκει μόνο σε μία κλάση. Τα άκρα των κλάσεων καλούνται **όρια των κλάσεων** (class boundaries). Συνήθως υιοθετούμε την περίπτωση που μια κλάση περιέχει το κάτω άκρο της (κλειστή αριστερά) αλλά όχι το άνω άκρο της (ανοικτή

δεξιά), δηλαδή που οι κλάσεις είναι της μορφής [,). Οι παρατηρήσεις κάθε κλάσης θεωρούνται όμοιες, οπότε μπορούν να "αντιπροσωπευθούν" από τις **κεντρικές τιμές**, τα κέντρα δηλαδή κάθε κλάσης.

- Το πρώτο βήμα στην ομαδοποίηση των δεδομένων είναι η εκλογή του αριθμού κ των ομάδων ή κλάσεων. Ο αριθμός αυτός συνήθως ορίζεται αυθαίρετα από τον ερευνητή σύμφωνα με την πείρα του. Γενικά όμως μπορεί να χρησιμοποιηθεί ως οδηγός ο παρακάτω πίνακας:

Μέγεθος δείγματος ν	Αριθμός κλάσεων κ	Μέγεθος δείγματος ν	Αριθμός κλάσεων κ
< 20	5	$200-400$	9
$20-50$	6	$400-700$	10
$50-100$	7	$700-1000$	11
$100-200$	8	≥ 1000	12

- Το δεύτερο βήμα είναι ο προσδιορισμός του πλάτους των κλάσεων. **Πλάτος μιας κλάσης** ονομάζεται η διαφορά του κατωτέρου από το ανώτερο όριο της κλάσης. Στην πλειονότητα των πρακτικών εφαρμογών οι κλάσεις έχουν το ίδιο πλάτος. Φυσικά υπάρχουν και περιπτώσεις όπου επιβάλλεται οι κλάσεις να έχουν άνισο πλάτος, όπως, για παράδειγμα, στις κατανομές εισοδήματος, ημερών απεργίας κτλ. Για να

• κατασκευάσουμε ισοπλατείς κλάσεις, χρησιμοποιούμε το **εύρος** (range) R του δείγματος, δηλαδή τη διαφορά της μικρότερης παρατήρησης από τη μεγαλύτερη παρατήρηση του συνολικού δείγματος. Τότε υπολογίζουμε το πλάτος c των κλάσεων διαιρώντας το εύρος R διά του αριθμού των κλάσεων κ, στρογγυλεύοντας, αν χρειαστεί για λόγους διευκόλυνσης, πάντα προς τα πάνω.

• Το επόμενο βήμα είναι η κατασκευή των κλάσεων. Ξεκινώντας από την μικρότερη παρατήρηση, ή για πρακτικούς λόγους λίγο πιο κάτω από την μικρότερη παρατήρηση, και προσθέτοντας κάθε φορά το πλάτος c δημιουργούμε τις κ κλάσεις. Αυτονόητο είναι ότι η μεγαλύτερη τιμή του δείγματος θα (πρέπει να) ανήκει οπωσδήποτε στην τελευταία κλάση.

• Τέλος, γίνεται η **διαλογή** των παρατηρήσεων. Το πλήθος των παρατηρήσεων ν_i που προκύπτουν από τη διαλογή για την κλάση i καλείται **συχνότητα της κλάσης** αυτής ή **συχνότητα της κεντρικής τιμής** x_i , $i = 1,2,...,\kappa$.

Έστω, για παράδειγμα, ότι από τα δεδομένα του πίνακα 4 εξετάζουμε το ύψος των μαθητών. Το ύψος των μαθητών, όπως έχει καταγραφεί με τη σειρά, δίνεται στον παρακάτω πίνακα 8.

Πίνακας 8

Το ύψος (σε cm) των μαθητών της Γ΄ Λυκείου, όπως έχει καταγραφεί στον πίνακα 4.

Σε αγκύλες έχουμε τη μικρότερη και τη μεγαλύτερη τιμή.

170	180	178	165	170	168	175	175	173	162
160	170	167	177	180	170	182	178	165	178
[151]	175	172	173	167	187	170	180	178	[191
176	169	167	166	179	178	180	164	170	173

Παρατηρούμε ότι το εύρος του δείγματος είναι $R = 191 - 156 = 35$. Επειδή έχουμε $\nu = 40$ παρατηρήσεις, χρησιμοποιούμε $\kappa = 6$ κλάσεις. Το πλάτος των κλάσεων είναι $c = R/\kappa = 35/6 = 5{,}83 \approx 6$. Αν θεωρήσουμε ως αρχή της πρώτης κλάσης το 156, θα έχουμε τον επόμενο πίνακα 9.

Πρέπει να προσεχτεί ότι:

- Καμία παρατήρηση δεν μπορεί να μείνει έξω από κάποια κλάση.
- Οι κεντρικές τιμές διαφέρουν μεταξύ τους όσο και το πλάτος των κλάσεων, που εδώ είναι ίσο με 6.
- Μία παρατήρηση που συμπίπτει με το άνω άκρο μιας κλάσης θα τοποθετηθεί κατά τη διαλογή στην αμέσως επόμενη κλάση. Για παράδειγμα, ο μαθητής με ύψος 180 θα τοποθετηθεί στην πέμπτη κλάση [180, 186).

Πίνακας 9

Κατανομές συχνοτήτων (απόλυτων, σχετικών, αθροιστικών) για τα δεδομένα του πίνακα 8.

Κλάσεις [-)	Κεντρικές τιμές x_i	Διαλογή	Συχν. ν_i	Σχετική Συχνότητα $f_i\%$	Αθρ.συχν. N_i	Αθρ. Σχετ. Συχν. $F_i\%$
156-162	159	||	2	5,0	2	5,0
162-168	165	𝍸 |||	8	20,0	10	25,0
168-174	171	𝍸 𝍸	12	30,0	22	55,0
174-180	177	||	11	27,5	33	82,5
180-186	183	𝍸 𝍸	5	12,5	38	95,0
186-192	189	|	2	5,0	40	100,0
		𝍸				
		||				
	Σύνολο	—	40	100	—	—

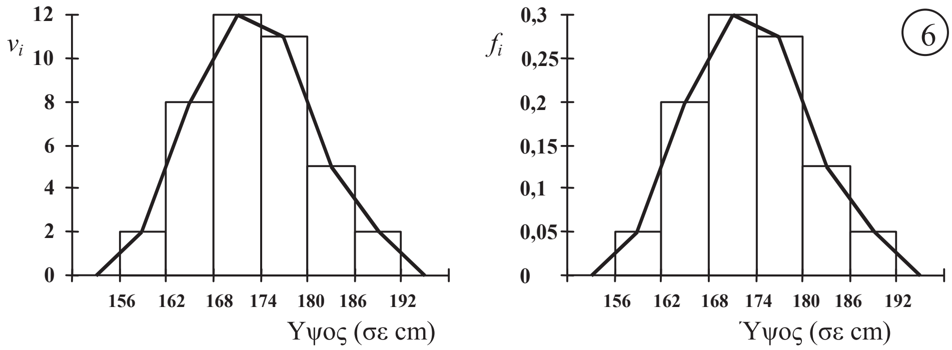
Ιστόγραμμα Συχνοτήτων

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

2.2.1 Κλάσεις Ίσου Πλάτους

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

Αν στα ιστογράμματα συχνοτήτων θεωρήσουμε δύο ακόμη υποθετικές κλάσεις, στην αρχή και στο τέλος, με συχνότητα μηδέν και στη συνέχεια ενώσουμε τα μέσα των άνω βάσεων των ορθογωνίων, σχηματίζεται το λεγόμενο **πολύγωνο συχνοτήτων** (frequency polygon). 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, (βλέπε σχήμα 6).

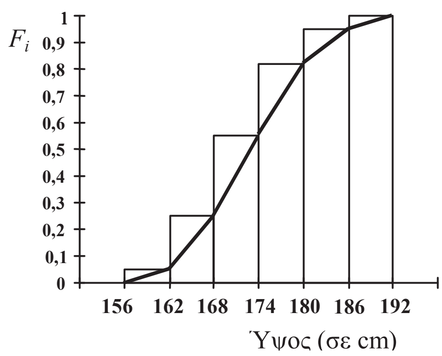


(α) (β)

Ιστόγραμμα και πολύγωνο (α) συχνοτήτων και (β) σχετικών συχνοτήτων για τα δεδομένα του πίνακα 9.

Με τον ίδιο τρόπο κατασκευάζονται και τα **ιστογράμματα αθροιστικών συχνοτήτων** και **αθροιστικών σχετι-κών συχνοτήτων**. Αν ενώσουμε σε ένα ιστόγραμμα αθροιστικών συχνοτήτων τα **δεξιά άκρα** (όχι μέσα) των άνω βάσεων των ορθογωνίων με ευθύγραμμα τμήματα βρίσκουμε το **πολύγωνο αθροιστικών συχνοτήτων** (ogive) της κατανομής. Στο σχήμα 7 παριστάνεται το ιστόγραμμα και το πολύγωνο αθροιστικών σχετικών συχνοτήτων για το ύψος των μαθητών του πίνακα 9 .

7



β) Κλάσεις Άνισου Πλάτους

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

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

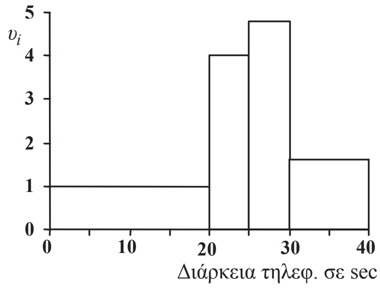
Διάρκεια τηλεφ. σε sec	Συχνότητα ν_i
0-20	20
20-25	20
25-30	24
30-40	16
Σύνολο	$\nu = 80$

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

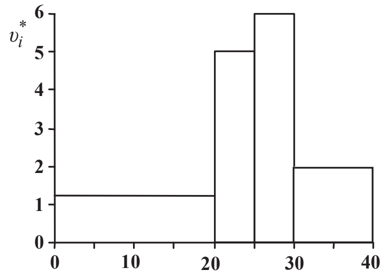
Διάρκεια τηλεφ. σε sec.	Πλάτος κλάσης c_i	Συχνότητα ν_i	Ύψος $\upsilon_i = \frac{\nu_i}{c_i}$	Ύψος $\upsilon_i^* = \frac{f_i\%}{c_i}$
0-20	20	20	1,0	1,25
20-25	5	20	4,0	5,00
25-30	5	24	4,8	6,00
30-40	10	16	1,6	2,00

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

8



(α)



(β)

Ιστόγραμμα συχνοτήτων (α) και σχετικών συχνοτήτων (β)

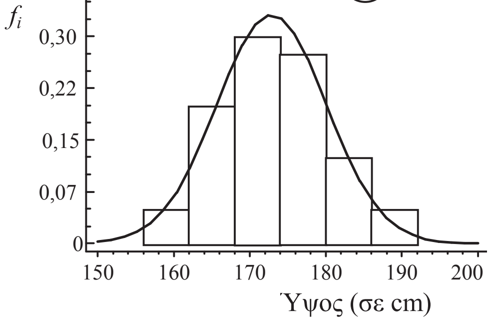
της διάρκειας τηλεφωνημάτων.

Με ανάλογο τρόπο κατασκευάζεται και το ιστόγραμμα σχετικών συχνοτήτων, (σχήμα 8(β)) αρκεί να χρησιμοποιήσουμε ως ύψος των ορθογωνίων το λόγο των σχετικών συχνοτήτων προς το πλάτος των κλάσεων, δηλαδή $\upsilon_i^* = \dfrac{f_i\%}{c_i}$.

Καμπύλες Συχνοτήτων

Εάν υποθέσουμε ότι ο αριθμός των κλάσεων για μια συνεχή μεταβλητή είναι αρκετά μεγάλος (τείνει στο άπειρο) και ότι το πλάτος των κλάσεων είναι αρκετά μικρό (τείνει στο μηδέν), τότε η πολυγωνική γραμμή συχνοτήτων τείνει

9



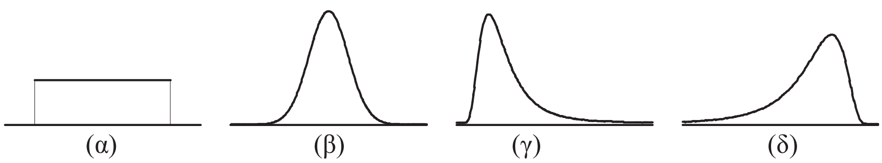
να πάρει τη μορφή μιας ομαλής καμπύλης, η οποία ονομάζεται **καμπύλη συχνοτήτων** (frequency curve), όπως δείχνει το σχήμα 9. Οι καμπύλες συχνοτήτων έχουν μεγάλη εφαρμογή στη Στατιστική, όπου οι ιδιότητες τους μπορούν να χρησιμοποιηθούν για την εξαγωγή χρήσιμων συμπερασμάτων.

Καμπύλη συχνοτήτων για το ύψος

των μαθητών του πίνακα 4

Η μορφή μιας κατανομής συχνοτήτων εξαρτάται από το πώς είναι κατανεμημένες οι παρατηρήσεις σε όλη την έκταση του εύρους τους. Μερικές χαρακτηριστικές καμπύλες συχνοτήτων που συναντάμε συχνά στις εφαρμογές δίνονται στο σχήμα 10. Η κατανομή (β), με "κωδωνοειδή" μορφή λέγεται **κανονική κατανομή** (normal distribution) και παίζει σπουδαίο ρόλο στη Στατιστική. Όταν οι παρατηρήσεις "κατανέμονται" ομοιόμορφα σε ένα διάστημα [α, β], όπως στην κατανομή (α), η κατανομή λέγεται **ομοιόμορφη**. Όταν οι παρατηρήσεις δεν είναι συμμετρικά κατανεμημένες, η κατανομή λέγεται ασύμμετρη με θετική ασυμμετρία όπως στην κατανομή (γ) ή αρνητική ασυμμετρία όπως στην κατανομή (δ).

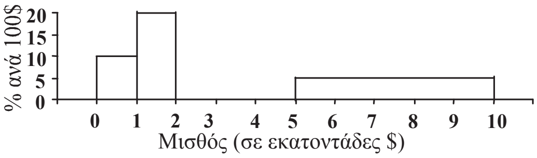
10



Μερικές χαρακτηριστικές κατανομές συχνοτήτων

Παράδειγμα 1.

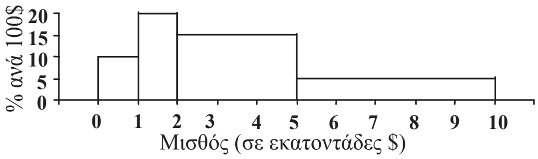
Στο διπλανό ιστόγραμμα σχετικών συχνοτήτων σβήστηκε κατά λάθος



το ορθογώνιο της κλάσης [2-5). Εάν είναι γνωστό ότι δεν υπάρχει μισθός άνω των 1000€, να κατασκευάσετε το ορθογώνιο αυτό.

ΛΥΣΗ

Επειδή έχουμε ένα ιστόγραμμα σχετικών συχνοτήτων ($f_i\%$), το άθροισμα των εμβαδών όλων των ορθογωνίων θα πρέπει να ισούται με 100. Το εμβαδόν του πρώτου ορθογωνίου είναι $E_1 = (1-0)\cdot 10 = 10$, του δεύτερου ορθογωνίου $E_2 = (2-1)\cdot 20 = 20$,



και του τέταρτου $E_4 = (10-5)\cdot 5 = 25$. Άρα, το εμβαδόν του τρίτου ορθογωνίου θα είναι $E_3 = 100-(10+20+25) = 45$. Επειδή το πλάτος του ορθογωνίου είναι $5-2=3$, το ύψος του θα είναι $45/3 = 15$, όπως φαίνεται στο διπλανό σχήμα.

3.ΕΠΕΞΕΡΓΑΣΙΑ ΣΤΑΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Τα στατιστικά στοιχεία τα οποία έχουμε συλλέξει, τα ταξινομήσαμε και τα παρουσιάσαμε. Τώρα πρέπει και να τα επεξεργαστούμε για να πάρουμε αποτελέσματα τα οποία αφού ερμηνεύσουμε θα τα χρησιμοποιήσουμε για τη λήψη αποφάσεων.

Η επεξεργασία αυτή μπορεί να γίνει με διάφορα Στατιστικά Μέτρα τα οποία διακρίνονται σε:

1. Μέτρα Κεντρικής Τάσης – Central Tendency
2. Μέτρα Διασποράς – Dispersion
3. Μέτρα Ασυμμετρίας – Skewness και
4. Μέτρα Κύρτωσης – Kurtosis

3.1 Μέτρα Κεντρικής Τάσης

Measures of Central Tendency

Τα Μέτρα Κεντρικής Τάσης είναι:

- Ο Αριθμητικός Μέσος – Arithmetic Mean
- Η Διάμεσος – Median
- Τα τεταρτημόρια – Quartiles
- Τα δεκατημόρια
- Τα εκατοστημόρια – Percentiles

Μπορεί οι μέσοι και ειδικά ο αριθμητικός να είναι σπουδαία στατιστικά μέτρα, όμως για να έχουμε μία πλήρη εικόνα της συμπεριφοράς του πληθυσμού ή του δείγματος και να πάρουμε σωστές αποφάσεις χρειάζεται να υπολογίσουμε και άλλα στατιστικά μέτρα. Όταν ειδικά η καμπύλη συχνοτήτων της κατανομής είναι πολύ ασύμμετρη, τότε επιβάλλεται η επιπλέον μελέτη και άλλων μέτρων κεντρικής τάσης, από τα προαναφερθέντα, σπουδαιότερα των οποίων είναι η Διάμεσος και ο Τύπος.

3.1.1 Αριθμητικός Μέσος – Arithmetic Mean

Ο πλέον γνωστός και εύχρηστος από τους Μέσους είναι ο Αριθμητικός Μέσος, τον οποίο όλοι γνωρίζουμε και σίγουρα κάποια φορά, τον έχουμε χρησιμοποιήσει είτε για να δούμε πόσα χρήματα ξοδεύουμε κατά μέσο όρο κάθε μέρα είτε γιατί θέλαμε να δούμε τη μέση βαθμολογία του τριμήνου στις τάξεις του Λυκείου.

Σ' αυτή την παράγραφο θα αναφερθούμε αναλυτικά στον Αριθμητικό Μέσο, προσπαθώντας να ξεχωρίσουμε τις διάφορες περιπτώσεις όπως αυτές παρουσιάζονται με βάση την μορφή και το είδος της μεταβλητής. Θα δούμε επίσης τις ιδιότητες, τα πλεονεκτήματα και τα μειονεκτήματα.

3.1.1.1 Αστάθμητος Αριθμητικός Μέσος

Όταν όλες οι τιμές της μεταβλητής έχουν την ίδια βαρύτητα στη διαμόρφωση του τελικού αποτελέσματος χρησιμοποιούμε τον Αστάθμητο (Unweigthed) Αριθμητικό Μέσο.

i) Αστάθμητος Αριθμητικός Μέσος αταξινόμητων παρατηρήσεων

Όταν οι παρατηρήσεις είναι λίγες, είναι πολύ εύκολο να τις επεξεργαστούμε χωρίς να χρειαστεί πίνακας κατανομής συχνοτήτων. Σ' αυτή την περίπτωση ο υπολογισμός του Αριθμητικού Μέσου των τιμών $\chi_1, \chi_2, \chi_3, \ldots.. \chi_\nu$, είναι μία απλή πρόθεση όλων των τιμών και εν συνεχεία διαίρεση αυτού του αθροίσματος διά του πλήθους των τιμών και εν συνεχεία διαίρεση αυτού του αθροίσματος διά του πλήθους των τιμών, όπως φαίνεται παρακάτω.

Μέσος πληθυσμού: $\mu = \dfrac{X_1 + X_2 \ldots + X_n}{n} = \dfrac{\sum_{i=1}^{n} X_i}{n}$ (1)

όπου **μ** : Αριθμητικός Μέσος

$\sum_{i=1}^{n} X_i$: Σύμβολο αθροίσματος από i=1 έως n

n: πλήθος στοιχείων πληθυσμού

και X_i : Οι τιμές της μεταβλητής

Μέσος Δείγματος : $\overline{X} = \dfrac{X_1 + X_2 \ldots + X_\nu}{\sum_{i=1}^{\nu} \nu_i} = \dfrac{\sum_{i=1}^{\nu} X_i}{\sum_{i=1}^{\nu} \nu_i}$ (2)

όπου $\overline{X}$: Αριθμητικός Μέσος

$\sum_{i=1}^{\nu} X_i$: Σύμβολο αθροίσματος από i=1 έως ν

ν_i : Αριθμός των στοιχείων του δείγματος

και X_i : Οι τιμές της μεταβλητής

Παράδειγμα: Τα προσωπικά μας έξοδα ενός υπαλλήλου, στη διάρκεια μίας εβδομάδας είναι 6, 7, 5, 8, 9 ,11 και 5 Euro. Προκειμένου να υπολογίσουμε τον ημερήσιο μέσο όρο αυτών των εξόδων χρησιμοποιούμε τον τύπο 2 και έχουμε:

$$\overline{X} = \frac{6+7+5+8+9+11+5}{7} = 7{,}29 \text{ Euro}$$

ii) Αστάθμητος Αριθμητικός μέσος ταξινομημένων παρατηρήσεων

Όταν οι παρατηρήσεις είναι πολλές, είναι απαραίτητο, όπως είπαμε σε προηγούμενα κεφάλαια, να τις ταξινομήσουμε και να δημιουργήσουμε ένα πίνακα κατανομής συχνοτήτων. Από τον πίνακα αυτό πλέον θα υπολογίσουμε τον αριθμητικό μέσο χρησιμοποιώντας τον παρακάτω τύπο:

$$\bar{X} = \frac{X_1F_1 + X_2F_2 + ... + X_\nu F_\nu}{F_1 + F_2 + ... + F_\nu} = \frac{\sum_{i=1}^{\nu} X_i F_i}{\sum_{i=1}^{\nu} F_i} \qquad (4)$$

Όπου $\bar{X}$: Αριθμητικός Μέσος

$\sum_{i=1}^{\nu}$: Σύμβολο αθροίσματος από i=1 έως ν

και F_i : Συχνότητα (αριθμός παρατηρήσεων) κάθε τιμής.

Παράδειγμα: Αν θελήσουμε να βρούμε τον μέσο αριθμό υπαλλήλων στο σύνολο των 30 επιχειρήσεων του παρακάτω πίνακα , θα πρέπει να εφαρμόσουμε τον τύπο (4)ως εξής:

Αριθμός υπαλλήλων X_i	Αριθμός επιχειρήσεων F_i	X_iF_i
0	4	0
1	7	7
2	9	18
3	5	15
4	2	8
5	3	15
ΣΥΝΟΛΟ	30	63

$$\bar{X} = \frac{63}{12} = 2,1 \cong 2 \text{ περίπου}$$

υπάλληλοι ανά επιχείρηση

iii) Αστάθμητος Αριθμητικός Μέσος Ομαδοποιημένων Παρατηρήσεων.

Όταν οι τιμές της μεταβλητής είναι πολλές, όπως γνωρίζουμε, κάνουμε ομαδοποίηση σε ομάδες οι οποίες μπορεί να έχουν ίσο πλάτος ή και άνισο. Στην περίπτωση αυτή και πριν υπολογίσουμε τον αριθμητικό μέσο πρέπει να δημιουργήσουμε μία στήλη στην οποία θα αναγράψουμε την κεντρική τιμή του κάθε διαστήματος η οποία πλέον θα αντικαθιστά ολόκληρο το διάστημα στους υπολογισμούς για τον προσδιορισμό του μέσου. Η κεντρική τιμή του κάθε διαστήματος προκύπτει από τον τύπο:

$\frac{μέγαλο.άκρο.διαστ. + μικρό.άκρο.διαστ.}{2}$. η τιμή αυτή είναι η κεντρική τιμή του διαστήματος και όχι η μέση τιμή αυτού. Έτσι και το αποτέλεσμα που θα πάρουμε θα είναι λίγο διαφορετικό από αυτό που θα παίρναμε αν τις ίδιες παρατηρήσεις τις επεξεργαζόμασταν χωρίς ομαδοποίηση. Η διαφορά όμως είναι πολύ μικρή και θεωρούμε ότι τα οφέλη από την ομαδοποίηση είναι πολύ μεγαλύτερα στην ολοκληρωμένη επεξεργασία.

Παράδειγμα 1^o: Στον παρακάτω πίνακα θα χρησιμοποιήσουμε τον τύπο (4) αφού συμπληρώσουμε τις απαραίτητες στήλες.

Πινακας (Α)

Εβδομαδιαίοι Μισθοί σε ευρώ ($χ_i$)	Αριθμός εργατών-Συχνότητες (F_i)	Κέντρο ομάδας X_i	F_iX_i
140-230	4	185	740
230-320	9	275	2475
320-410	16	365	5840
410-500	35	455	15925
500-590	16	545	8720
590-680	13	635	8255
680-770	5	725	3625
770-880	2	815	1630
ΣΥΝΟΛΟ	100		

Από τα στοιχεία του προηγούμενου πίνακα θα έχουμε: $\overline{X} = \frac{47210}{100} = 472{,}1$.

Αν αντί της ομαδοποίησης προχωρήσουμε στον υπολογισμό του αριθμητικού μέσου από τις αταξινόμητες παρατηρήσεις της , τότε χρησιμοποιώντας τον τύπο (1) θα έχουμε: $\overline{X} = \frac{47254}{100} = 472,54$, που είναι και το πιο σωστό αποτέλεσμα.

Πολλές φορές ο τύπος (4) είναι δύσκολο να χρησιμοποιηθεί λόγω των μεγάλων αριθμών οι οποίοι προκύπτουν από τα γινόμενα F_i X_i και για τον λόγο αυτό μπορούμε να χρησιμοποιούμε ένα άλλο τύπο ο οποίος δίνει μικρότερα ενδιάμεσα αριθμητικά αποτελέσματα, τον ίδιο όμως αριθμητικό μέσο. Ο τύπος αυτός έχει τη μορφή: $\overline{X} = X_0 + \frac{\delta . \sum_{i=1}^{\nu} X_i F_i}{\sum_{i=1}^{\nu} F_i}$ (5)

Όπου : X_0 : Το κέντρο του διαστήματος στο οποίο βάλαμε 0

δ: το πλάτος του διαστήματος στο οποίο βάλαμε 0 και

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

Όπως καταλαβαίνουμε πρέπει να δημιουργήσουμε μία στήλη ξ, η διαμόρφωση της οποίας εξαρτάται από τη θέση στην οποία θα βάλουμε το 0. Η θέση αυτή επιλέγεται αυθαίρετα, αν και συνήθως προτιμούμε την ομάδα με τη μεγαλύτερη συχνότητα. Είναι ευνόητο ότι ο καθένας μπορεί αν επιλέξει διαφορετική θέση και να προκύψουν κατ' αρχάς διαφορετικά ενδιάμεσα αποτελέσματα, όμως το τελικό αποτέλεσμα του αριθμητικού μέσου θα πρέπει να είναι το ίδιο ανεξαρτήτως επιλογής.

Παράδειγμα 2ο: Θα εφαρμόσουμε τον τύπο (5) στον παρακάτω πίνακα(Β).

Πινάκας (Β)

Εβδομαδιαίοι Μισθοί σε ευρώ (χ_i)	Αριθμός εργατών-Συχνότητες (F_i)	Κέντρο ομάδας X_i	ξ_i	$F_i\xi_i$
150-250	5	200	3	-15
250-350	13	300	2	-26
350-450	20	400	1	-20
450-550	35	500	0	0
550-650	18	600	1	18
650-750	7	700	2	14
750-850	2	800	3	6
ΣΥΝΟΛΟ	100			-23

$$\bar{X} = 500 + \frac{100(-23)}{100} = 477$$

Το αποτέλεσμα που βρήκαμε από τον Πίνακα (Β) είναι αρκετά διαφορετικό από αυτό που βρήκαμε από τον Πίνακα (Α), αν και πρόκειται για επεξεργασία των ίδιων στοιχείων. Αυτό οφείλεται στη διαφορετική ομαδοποίηση και φυσικά γίνεται αντιληπτό ότι καλύτερη ομαδοποίηση είχαμε στον πίνακα (Α). όπου και το αποτέλεσμα είναι 472,1 πολύ κοντά στο πραγματικό που είναι 472,54.

Η χρησιμότητα αυτού του τύπου θα γίνει περισσότερο κατανοητή όταν σε επόμενες παραγράφους θα πρέπει να υπολογίσουμε τα μέτρα διασποράς, όπου οι πράξει είναι πιο περίπλοκες.

Ιδιαίτερη προσοχή χρειάζεται η εφαρμογή του τύπου (5) όταν οι ομάδες δεν έχουν το ίδιο πλάτος και η στήλη ξ είναι πιο δύσκολο να διαμορφωθεί.

Παράδειγμα 3^o: Η εφαρμογή του τύπου (5) στον παρακάτω Πίνακα είναι αυτής της περίπτωσης και θέλει ιδιαίτερη προσοχή.

Πληθυσμός- X_i	Αριθμός Κοινοτήτων F_i	Κέντρο ομάδας X_i	ξ_ι	$F_i\xi_i$
0-2000	32	1000	-3.5	-112
2001-3000	13	2500	-2	-26
3001-4000	5	3500	-1	-5
4001-5000	3	4500	0	0
5001-7000	5	6000	1,5	7,5
7000-10000	1	8500	4	4
ΣΥΝΟΛΟ	59			-131,5

Παρατηρούμε δηλαδή ότι η στήλη του ξ_i, δεν αυξάνεται από τη θέση του 0 και προς τα κάτω κατά μία μονάδα (ελαττώνεται από τη θέση του 0 και προς τα επάνω) όπως στις περιπτώσεις που έχουμε ομάδες με ίσα πλάτη. Στην περίπτωση που είδαμε έχουμε μία διαμόρφωση της στήλης εξαρτώμενη αποκλειστικά από τα πλάτη των διαστημάτων. Έτσι ο αριθμητικός μέσος είναι: $\overline{X} = 4500 + \dfrac{1000 \cdot (-131.5)}{59} = 2271,20$ κάτοικοι ανά κοινότητα.

3.1.1.2 Ιδιότητες Αριθμητικού Μέσου

Οι σπουδαιότερες ιδιότητες του αριθμητικού μέσου, χωρίς απόδειξη, είναι αυτές που αναφέρονται παρακάτω.

- Το άθροισμα των διαφορών όλων των τιμών της μεταβλητής από τον αριθμητικό μέσο είναι μηδέν. Δηλαδή $\sum_{i=1}^{\nu}(\overline{X} - X_i) = 0$. Δεν ισχύει στις ομαδοποιημένες κατανομές.
- Ο αριθμητικός μέσος οποιασδήποτε κατανομής βρίσκεται μεταξύ της μικρότερης και της μεγαλύτερης τιμής αυτής. $\min X_i \le \overline{X} \le \max X_i$

- Αν σε όλες τις τιμές της μεταβλητής προσθέσουμε ή αφαιρέσουμε τον ίδιο σταθερό αριθμό και ο αριθμητικός μέσος αυξάνεται ή ελαττώνεται κατά τον ίδιο αριθμό κ. Δηλαδή $\overline{Y} = \overline{X} \pm k$.
- Αν όλες τις τιμές της μεταβλητής τις πολλαπλασιάσουμε ή τις διαιρέσουμε με τον ίδιο σταθερό αριθμό λ, τότε και ο αριθμητικός μέσος πολλαπλασιάζεται ή διαιρείται με τον ίδιο σταθερό αριθμό λ. Δηλαδή $\overline{Y} = \lambda \cdot \overline{X}$ ή $\overline{Y} = \overline{X} \cdot \lambda$.

3.1.1.3 Γενικά Χαρακτηριστικά Αριθμητικού Μέσου

- Είναι μία παράμετρος η οποία είναι καλά ορισμένη, υπολογίζεται εύκολα και γίνεται κατανοητή και από τον απλό αναγνώστη.
- Επηρεάζεται από όλες τις τιμές της μεταβλητής.
- Μπορεί να χρησιμοποιηθεί για περαιτέρω στατιστική ανάλυση. Για παράδειγμα ο αριθμητικός μέσος τυχαίου δείγματος μπορεί να χρησιμοποιηθεί ως εκτιμητής της μέσης τιμής του πληθυσμού.
- Αντιμετωπίζει περιπτώσεις ποσοτικής μεταβλητής κάθε μορφής.
- Επηρεάζεται πολύ από τις ακραίες τιμές, σε σημείο που μειώνεται αισθητά η αξιοπιστία του στο σύνολο των τιμών.
- Μπορεί να πάρει μία τιμή η οποία δεν ανήκει στο σύνολο τιμών της μεταβλητής και αυτό συμβαίνει κυρίως όταν η μεταβλητής είναι ασυνεχής.
- Δεν επαρκεί όταν η κατανομή παρουσιάζει μεγάλη ανομοιογένεια.
- Δεν μπορεί να εφαρμοστεί σε μία κατανομή ανοιχτή.

!!!. Συχνά σε ένα πίνακα κατανομής συχνοτήτων συμβαίνει να υπάρχουν λίγες, πολύ μεγάλες ή και πολύ μικρές τιμές. Ο αριθμητικός μέσος, ως γνωστόν, επηρεάζεται έντονα από αυτές με αποτέλεσμα να μην περιγράφει κατά αντικειμενικό τρόπο την πραγματική κατάσταση. Για να αποφύγουμε την επίδραση αυτών των τιμών τις αφαιρούμε και υπολογίζουμε τον αριθμητικό μέσο των υπολοίπων τιμών. Αυτόν τον αριθμητικό μέσο τον αποκαλούμε Τακτοποιημένο μέσο – Trimmed mean και συνοδεύεται από ένα συντελεστή α, ο οποίος δηλώνει το ποσοστό των ακραίων τιμών οι οποίες θα αφαιρεθούν. Έτσι ένας 6% τακτοποιημένος μέσος σημαίνει ότι προέκυψε αφού πρώτα αφαιρέθηκε το 6% των χαμηλών και το 6% των υψηλών τιμών της κατανομής. Συνεπώς, ο μέσος προέκυψε από το υπόλοιπο 88% των τιμών.

3.2 Διάμεσος - Median

Η Διάμεσος αποτελεί το σπουδαιότερο, μετά τον αριθμητικό μέσο, στατιστικό μέτρο κεντρικής τάσης. Χωρίζει την κατανομή σε δύο ισοπληθείς ομάδες και η τιμή της είναι η κεντρική τιμή της μεταβλητής. Έτσι, το 50% των τιμών της μεταβλητής είναι μικρότερες από αυτή και το υπόλοιπο 50% μεγαλύτερες. Αν θελήσουμε δηλαδή να μάθουμε μέχρι ποιο ποσό φτάνει ο μισθός του 50% των χαμηλόμισθων εργατών μίας βιομηχανίας πρέπει να υπολογίσουμε τη διάμεσο, αφού πρώτα κατατάξουμε τους μισθούς σε αύξουσα τάξη. Η τιμή της διαμέσου θα είναι το ποσό που θέλουμε.

i) Διάμεσος αταξινόμητων παρατηρήσεων

Στην περίπτωση των ολίγων και κατά συνέπεια αταξινόμητων παρατηρήσεων προέχει η κατ' αύξουσα τάξη τακτοποίηση αυτών. Στη συνέχεια επιλέγουμε τη θέση της διαμέσου από τη σχέση $\frac{\nu+1}{2}$ και τη τιμή η οποία βρίσκεται στη θέση αυτή είναι η διάμεσος. Αν το ν είναι περιττός αριθμός, τότε η θέση είναι ακέραιος αριθμός και δεν υπάρχει κανένα πρόβλημα. Αν όμως το ν είναι άρτιος αριθμός, τότε η θέση είναι δεκαδικός αριθμός και διάμεσος είναι το ημιάθροισμα των τιμών που βρίσκονται στην προηγούμενη και στην επόμενη θέση.

Παράδειγμα 1^{o}: Οι αριθμοί που δίνονται παρακάτω εκφράζουν την ηλικία των εργατών μίας μικρής επιχείρησης: 35,56,53,29,43,54,28,39,42,58,26,31,51.

Αν θελήσουμε να υπολογίσουμε τη διάμεσο των παρακάτω τιμών, πρέπει πρώτα να τους τοποθετήσουμε από τον μικρότερο προς τον μεγαλύτερο. Έτσι θα έχουμε τον παρακάτω πίνακα.

1^{η}	2^{η}	3^{η}	4^{η}	5^{η}	6^{η}	7^{η}	8^{η}	9^{η}	10^{η}	11^{η}	12^{η}	13^{η}
26	28	29	31	35	39	42	43	51	53	54	56	58

Από τη σχέση $\frac{\nu+1}{2}$ με αντικαταστάτη έχουμε $\frac{13+1}{2}=7$. Ο αριθμός 7 δείχνει τη θέση της διαμέσου. Στη θέση αυτή στον παραπάνω πίνακα αντιστοιχεί η τιμή 42.

Δηλαδή το 50% των εργατών έχει ηλικία μικρότερη των 42 ετών και το υπόλοιπο 50% ηλικία μεγαλύτερη 42 ετών.

Παράδειγμα 2^{o}: Αν στο προηγούμενο παράδειγμα προσθέσουμε την ηλικία ενός ακόμη εργάτη, έστω 59 ετών, θα έχουμε έναν άλλο πίνακα.

1^{η}	2^{η}	3^{η}	4^{η}	5^{η}	6^{η}	7^{η}	8^{η}	9^{η}	10^{η}	11^{η}	12^{η}	13^{η}	14^{η}
26	28	29	31	35	39	42	43	51	53	54	56	58	59

Όπου από τη σχέση $\frac{\nu+1}{2}$ με αντικατάσταση θα έχουμε $\frac{14+1}{2}=7,5$. Σ' αυτή την περίπτωση διάμεσος είναι το ημιάθροισμα των ηλικιών που βρίσκονται στην 7^{η} και την 8^{η} θέση, δηλαδή $\frac{42+43}{2}=42,5$.

ii) Διάμεσος Ταξινομημένων παρατηρήσεων

Στην περίπτωση που οι παρατηρήσεις καταγραφούν σε ένα πίνακα κατανομής συχνοτήτων, φαινόμενο συνηθισμένο όταν η μεταβλητή είναι ασυνεχής, δημιουργούμε τη στήλη των αθροιστικών συχνοτήτων. Στη συνέχεια προσδιορίζουμε τη θέση της διαμέσου από τη σχέση $\frac{\sum_{i=1}^{\nu} F_i}{2}$ ή $\frac{N}{2}$, βρίσκουμε τις δύο διαδοχικές αθροιστικές συχνότητες μεταξύ των οποίων βρίσκεται η τιμή που υπολογίσαμε και παίρνουμε για διάμεσο την τιμή της μεταβλητής που αντιστοιχεί στη μεγαλύτερη από τις δύο.

Παράδειγμα 1$^{\text{o}}$: Ο παρακάτω πίνακας δίνει την κατανομή των ορθογραφικών σφαλμάτων σε 100 γραπτά μαθητών, στους οποίου υπαγορεύτηκε το ίδιο κείμενο.

Αριθμός σφαλμάτων X_i	Αριθμός μαθητών F_i	Αθροιστική συχνότητα
0	12	12
1	27	$\left.\begin{matrix}39\\68\end{matrix}\right\} \frac{N}{2}=\frac{100}{2}=50$
2 ←	29	
3	19	87
4	8	95
5	4	99
6	1	100
ΣΥΝΟΛΟ	100	

Από τη σχέση $\frac{N}{2}=\frac{100}{2}=50$ προσδιορίζουμε τη θέση της διαμέσου. Στη συνέχεια βρίσκουμε τις δύο αθροιστικές συχνότητες που περιέχουν τον αριθμό 50. αυτές είναι οι αριθμοί 38 και 68. Η τιμή της μεταβλητής που αντιστοιχεί στο μεγαλύτερο, δηλαδή στον αριθμό 68, είναι 2 και είναι η διάμεσος. Καταλαβαίνουμε λοιπόν ότι το 50% των μαθητών έκανε το πολύ 2 σφάλματα και το υπόλοιπο 50% τουλάχιστον δύο σφάλματα.

Αν η θέση της διαμέσου συμπίπτει με μία αθροιστική συχνότητα, τότε Διάμεσος είναι η τιμή της μεταβλητής που αντιστοιχεί σ' αυτή την αθροιστική συχνότητα.

Παράδειγμα 2^o: Ο πίνακας που ακολουθεί, αναγράφει τον αριθμό των παιδιών 70 εργατών μίας επιχείρησης.

Αριθμός παιδιών X_i	Αριθμός εργατών F_i	Αθροιστική συχνότητα
0	7	7
1	10	17
2 ←	18	(70/2)=35
3	20	55
4	10	65
5	5	70
ΣΥΝΟΛΟ	70	

Εδώ παρατηρούμε ότι η θέση της διαμέσου είναι $\frac{70}{2}=35$ και συμπίπτει με μία αθροιστική συχνότητα στην οποία αντιστοιχεί η τιμή 2 που είναι η διάμεσος.

iii) Διάμεσος ομαδοποιημένων παρατηρήσεων.

Όταν οι παρατηρήσεις ομαδοποιηθούν, ο υπολογισμός της διαμέσου είναι περισσότερο πολύπλοκος και απαιτεί μεγαλύτερη προσοχή.

Για να αποφύγουμε τα λάθη, καλό είναι να ακολουθήσουμε την παρακάτω διαδικασία:

1. Συμπληρώνουμε τη στήλη των αθροιστικών συχνοτήτων
2. Προσδιορίζουμε τη θέση της διαμέσου από τη σχέση $\frac{N}{2}$
3. Βρίσκουμε τις δύο διαδοχικές αθροιστικές συχνότητες που περιέχουν τον αριθμό που βρήκαμε.
4. Τραβάμε μία γραμμή μεταξύ αυτών.
5. Επιλέγουμε την ομάδα κάτω από τη γραμμή.
6. Αντικαθιστούμε στον τύπο $M_{1/2} = X_i + \frac{\delta}{F_i}\left(\frac{N}{2} - \Phi_i\right)$ (6)

Όπου: $\mathbf{M_{1/2}}$: Διάμεσος

$\mathbf{X_i}$: Η μικρότερη τιμή της ομάδας κάτω από τη γραμμή

δ: Το πλάτος της ομάδας κάτω από τη γραμμή

$\mathbf{F_i}$: Η συχνότητα της ομάδα κάτω από τη γραμμή

N/2: Η θέση της διαμέσου

$\mathbf{\Phi_i}$: Η αθροιστική συχνότητα πάνω από τη γραμμή και

7. ελέγχουμε αν η τιμή που βρήκαμε ανήκει στην ομάδα κάτω από τη γραμμή.

Παράδειγμα 1^{ο}: Εφαρμογή στο παράδειγμα του πίνακα

Ομάδες μισθών Σε ευρώ X_i	Αριθμός εργατών-συχνότητα F_i	Αθροιστική Συχνότητα $Φ_i$
$[150-250)$	5	5
$[250-350)$	13	18
$[350-450)$	20	38 50
$\rightarrow [450-550) \leftarrow$	35	73
$[550-650)$	18	91
$[650-750)$	7	98
$[750-850)$	2	100
ΣΥΝΟΛΟ	100	

Από τη σχέση $\frac{N}{2}=\frac{100}{2}=50$ προσδιορίζουμε τη θέση της διαμέσου. Στη στήλη των αθροιστικών συχνοτήτων το 50 βρίσκεται μεταξύ του 38 και του 73. Άρα η ομάδα στην οποία θα βρίσκεται η Διάμεσος θα είναι η [450-550). Εφαρμόζοντας όσα αναφέραμε παραπάνω θα έχουμε Χ=450, δ=100, F=35, N/2=50, Φ=38 και με αντικατάσταση στον τύπο (6) θα πάρουμε:

$M_{1/2}=X_i+\frac{\delta}{F_i}\left(\frac{N}{2}-Φ_i\right)=450+\frac{100}{35}\left(\frac{100}{2}-38\right)=484,28$ που είναι η ζητούμενη διάμεσος αυτής της κατανομής και η οποία κατ' αρχάς τουλάχιστον είναι σωστή μια και βρίσκεται στο σωστό διάστημα. Το συμπέρασμα, από την τιμή που βρήκαμε, είναι ότι το 50% των εργατών έχουν μισθό μέχρι 484,28 euro, και το υπόλοιπο 50% μισθό μεγαλύτερο ή ίσο του ποσού αυτού.

Όταν η θέση της διαμέσου συμπίπτει με μία συγκεκριμένη τιμή της στήλης των αθροιστικών συχνοτήτων, τότε Διάμεσος είναι η μεγαλύτερη τιμή του διαστήματος στο οποίο αντιστοιχεί η θέση που βρήκαμε.

Όταν η θέση είναι μικρότερη και από την πρώτη τιμή της στήλης των αθροιστικών συχνοτήτων, τότε εφαρμόζουμε τον τύπο (1) βάζοντας στη θέση του Φ_i, το μηδέν.

3.3 Τεταρτημόρια (quartiles)

Εφαρμογή ανάλογη της Διαμέσου έχουν και τα τεταρτημόρια. Με αυτά χωρίζουμε την κατανομή σε τέσσερα ίσα τμήματα του 25% το καθένα. Έτσι το πρώτο τεταρτημόριο αναφέρεται στο πρώτο 25%, το δεύτερο στο 50% (διάμεσος) και το τρίτο στο 75% του συνόλου των παρατηρήσεων.

Σε περίπτωση αταξινόμητων ή απλά ταξινομημένων παρατηρήσεων εφαρμόζουμε ακριβώς την ίδια τεχνική με αυτή των αντίστοιχων περιπτώσεων της Διαμέσου, αλλάζοντας μόνο τη θέση.

Οι τύποι των τεταρτημόριων στις ομαδοποιημένες παρατηρήσεις είναι όμοιοι με αυτόν της διαμέσου και το μόνο που διαφέρει είναι η θέση. Έτσι θα έχουμε:

$$M_{1/4} = X_i + \frac{\delta}{F_i}\left(\frac{N}{4} - \Phi_i\right) \qquad \text{(7) Πρώτο τεταρτημόριο}$$

$$M_{3/4} = X_i + \frac{\delta}{F_i}\left(\frac{3N}{4} - \Phi_i\right) \qquad \text{(8) τρίτο τεταρτημόριο}$$

Η εφαρμογή αυτών δε διαφέρει σε τίποτα από την εφαρμογή του τύπου της διαμέσου, ακόμη και για τις ειδικές περιπτώσεις που αναφέραμε.

Παράδειγμα: Στον παρακάτω πίνακα θα έχουμε την ευκαιρία να εφαρμόσουμε στην πράξη αυτά που αναφέραμε.

Ομάδες μισθών σε ευρώ X_i	Αριθμός εργατών Συχνότητα F_i	Αθροιστική Συχνότητα- Φ_i
$[140-230)$	4	4
$[230-320)$	9	13
$\rightarrow[320-410)\leftarrow$	16	29
$\rightarrow[410-500)\leftarrow$	35	64
$\rightarrow[500-590)\leftarrow$	16	80
$[590-680)$	13	93
$[680-770)$	5	98
$[770-860)$	2	100
ΣΥΝΟΛΟ	100	

Έτσι οι θέσεις των τριών τεταρτημορίων είναι αντίστοιχα:

$$\frac{N}{4}=\frac{100}{4}=25,\frac{2N}{4}=\frac{N}{2}=50,\frac{3N}{4}=\frac{3\cdot 100}{4}=75$$

Και οι τιμές αυτών : $M_{1/4}=320+\frac{90}{16}(25-13)=387,5$

$$M_{1/2}=410+\frac{90}{35}(50-30)=461,42$$

$$M_{3/4}=500+\frac{90}{16}(75-64)=561,875$$

Αν θελήσουμε να ερμηνεύσουμε αυτά τα αποτελέσματα, θα δούμε ότι το 25% των εργατών έχει μισθό από 140 έως 387,5 Euro και επομένως το υπόλοιπο 75% μισθό μεγαλύτερο ή ίσιο αυτού του ποσού αλλά μέχρι τα 860 Euro. Το 50% μισθό από 140 έως 461,42 Euro και το υπόλοιπο 50% μεγαλύτερο ή ίσο αυτού, ενώ το 75% αυτών από 140 έως 561,875 Euro και το υπόλοιπο 25% μεγαλύτερο ή ίσο αυτού. Επίσης μισθό από 387,5 έως 561,875 Euro το 75%-25%=50% των εργατών.

3.4 Δεκατημόρια - Εκατοσημόρια

Για τους ίδιους λόγους που χρησιμοποιούμε τα τεταρτημόρια μπορούμε να χρησιμοποιήσουμε τα δεκατημόρια και τα εκατοστημόρια, τα οποία χωρίζουν την κατανομή σε τμήματα του 10% ή και του 1%. Έτσι παίρνουμε πληροφορίες για τη συμπεριφορά της μεταβλητής σε μικρότερα τμήματα. Σε σχέση με τον τρόπο εφαρμογής αυτών σε αταξινόμητες ή απλά ταξινομημένες παρατηρήσεις, το μόνο που αλλάζει, είναι η θέση του αντίστοιχου δεκατημορίου ή εκατοστημορίου.

Για ομαδοποιημένες παρατηρήσεις ο τύπος θα έχει τη μορφή:

$M_{d/10} = X_i + \frac{\delta}{F_i}\left(\frac{d \cdot N}{10} - \Phi_i\right)$ (9) με d=1,2,3,…,10 για τα **δεκατημόρια** και

$M_{P/100} = X_i + \frac{\delta}{F_i}\left(\frac{p \cdot N}{100} - \Phi_i\right)$ (10) με p=1,2,3,…,100 για **εκατοστημόρια**.

Παράδειγμα 1^{o}: Στον πίνακα που ακολουθεί παρουσιάζονται οι ηλικίες των 352 ενήλικων γυναικών ενός εργοστασίου .

Ομάδες ηλικιών X_i	Γυναίκες F_i	Αθροιστική συχνότητα Φ_ι
$\rightarrow 18-23 \leftarrow$	51	51
23-28	37	88
$\rightarrow 28-23 \leftarrow$	59	147
33-38	70	217
$\rightarrow 38-43 \leftarrow$	66	283
43-48	31	314
48-53	17	331
53-58	13	344
58-63	6	350
63-68	2	352
ΣΥΝΟΛΟ	352	

Αν θελήσουμε να υπολογίσουμε το πρώτο, το τρίτο και το έβδομο δεκατημόριο θα πρέπει πρώτα από όλα να προσδιορίσουμε τη θέση αυτών. Έτσι θα έχουμε:

$\frac{N}{10}=\frac{352}{10}=35,2$ $\frac{3N}{10}=\frac{3\cdot 352}{10}=105,6$ και $\frac{7N}{10}=\frac{7\cdot 352}{10}=246,4$, αντίστοιχα.

Τα δεκατημόρια θα είναι :

$$M_{1/10}=18+\frac{5}{51}\left(\frac{352}{10}-0\right)=21,45$$

$$M_{3/10}=28+\frac{5}{59}\left(\frac{3\cdot 352}{10}-88\right)=36,95$$

και $$M_{7/10}=38+\frac{5}{66}\left(\frac{7\cdot 352}{10}-217\right)=40,23$$

Τα οποία αν θέλουμε να τα ερμηνεύσουμε, όπως και τα τεταρτημόρια, θα πρέπει να πούμε ότι το 10% των γυναικών έχουν ηλικία μικρότερη ή ίση των 21,45 ετών, ενώ το υπόλοιπο 90% μεγαλύτερη ή ίση των 21,45, το 30% ηλικία μικρότερη ή ίση των 36,95 ετών ενώ το υπόλοιπο 70% μεγαλύτερη ή ίση των 36,95, το 70% αυτών ηλικία μικρότερη ή ίση των 40,23 ετών ενώ το υπόλοιπο 30% μεγαλύτερη ή ίση των 40,23. Επίσης αν θελήσουμε να δούμε πόσες γυναίκες έχουν ηλικία μεταξύ 36,95 και 40,23 ετών πρέπει να πούμε 70%-30%=40% και στη συνέχεια 40%x352=140,8 ή 246,4-105,6=140,8.

Μία άλλη ερώτηση θα μπορούσε να είναι η εξής: Πόσες γυναίκες έχουν ηλικία μεγαλύτερη ή ίση των 40,23 ετών; Ξέρουμε ότι ο αριθμός αυτός αντιστοιχεί στο έβδομο δεκατημόριο, δηλαδή σε 246,4 γυναίκες. Οι υπόλοιπες από το σύνολο των 352, είναι η απάντηση. Δηλαδή 352-246,4=105,6.

Εδώ θα πρέπει να τονίσουμε ότι η Διάμεσος, τα Τεταρτημόρια, τα Δεκατημόρια και τα Εκατοστημόρια είναι παράμετροι οι οποίοι εκφράζουν την κατανομή από την κατώτερη τιμή αυτής μέχρι την τιμή που υπολογίσαμε. Από την τιμή που υπολογίσαμε και μέχρι την ανώτερη τιμή της κατανομής είναι το υπόλοιπο ποσοστό. Έστω, λοιπόν, ότι μία κατανομή έχει κατώτερο άκρο τον αριθμό 30 και ανώτερο τον αριθμό 120. Αν ο αριθμός 50 είναι το πρώτο τεταρτημόριο, τότε το 25%

των παρατηρήσεων έχουν τιμή από 30 μέχρι 50 και το υπόλοιπο 75% από 50 μέχρι 120.

Πολλές φορές θέλουμε να προσδιορίσουμε το πλήθος των παρατηρήσεων μέχρι ή μετά από μία συγκεκριμένη τιμή της μεταβλητής, η οποία δεν αποτελεί κάποια από τις παραμέτρους που εξετάσαμε.

Άλλοτε πάλι θέλουμε να προσδιορίσουμε την τιμή της μεταβλητής που αντιστοιχεί σε ένα ορισμένο πλήθος παρατηρήσεων.

Και στις δύο αυτές περιπτώσεις θα χρησιμοποιήσουμε ένα γενικό τύπο που θα εφαρμόζεται ανάλογα με τη μορφή της ερώτησης. Ο τύπος αυτός είναι στην ουσία ο ίδιος που χρησιμοποιήσαμε μέχρι τώρα στις διάφορες παραμέτρους που εξετάσαμε, όπου όμως θα είναι άγνωστη η θέση (δηλαδή το ποσοστό που εκφράζει η δοσμένη τιμή της μεταβλητής) ή θα είναι άγνωστη η τιμή της παραμέτρου (ενώ θα δίνεται η θέση).

Η μορφή του τύπου είναι η εξής: $M = X_i + \frac{\delta}{F_i}\left(a - \Phi_i\right)$ (11)

Όπου: Μ: Η τιμή της μεταβλητής που βρίσκεται στη θέση α.

α: η θέση στην οποία αντιστοιχεί η τιμή Μ

X_i: η μικρότερη τιμή της ομάδας κάτω από τη γραμμή

δ: Το πλάτος της ομάδας κάτω από τη γραμμή

F_i: Η συχνότητα της ομάδας κάτω από τη γραμμή

Και Φ_i: Η αθροιστική συχνότητα πάνω από τη γραμμή.

Στον παρακάτω πίνακα, ο οποίος παρουσιάζει τις επενδύσεις 200 βιοτεχνιών της Θράκης (σε χιλιάδες Euro) κατά το έτος 2000, θα εφαρμόσουμε τις παρατηρήσεις που αναφέραμε.

Επενδύσεις (σε χιλ. ευρώ) X_i	Αριθμός βιοτεχνιών Συχνότητες F_i	Αθροιστικη Συχνοτητα Φ_i
$[44-74)$	10	10
$[74-104)$	26	36
$\rightarrow[104-134)\leftarrow$	40	76
$[134-164)$	70	146
$\rightarrow[164-194)\leftarrow$	36	182
$\rightarrow[194-224)\leftarrow$	14	196
$[224-254)$	4	200
ΣΥΝΟΛΟ	200	

Παράδειγμα 2^{o}: Έστω ότι θέλουμε να δούμε από τον πίνακα 4.1.6β

Α) πόσες βιοτεχνίες επένδυσαν το 1995 ποσά μικρότερα ή ίσα των 124.000 euro.

Β) πόσες βιοτεχνίες επένδυσαν το 1995 ποσά μεγαλύτερα ή ίσα των 170.000 euro και

Γ) Σε ποιο ποσό ανέρχεται η $186^{η}$ σε τάξη μεγέθους επένδυση.

Οι απαντήσεις κατά σειρά θα είναι:

Α) Προσδιορίζουμε την ομάδα στην οποία ανήκει η τιμή 124 και τραβάμε γραμμή πάνω από αυτή. Στη συγκεκριμένη περίπτωση, η ομάδα είναι η 104-134. Στη συνέχεια εφαρμόζουμε τον τύπο (5) έχοντας ως άγνωστο το α και θέτοντας στη θέση του M=124, X_i=104, δ=30, F_i=40 και Φ_i=36. Έτσι θα έχουμε :

$$124=140+\frac{30}{40}(a-36)\Leftrightarrow 124-140=\frac{30a}{40}-\frac{30\cdot 36}{40}\Leftrightarrow$$
$$\Leftrightarrow 20+\frac{30\cdot 36}{40}=\frac{30a}{40}\Leftrightarrow 20+27=0,75a\Leftrightarrow 47=0,75a\Leftrightarrow a=63$$

Δηλαδή 63 βιοτεχνίες ανήκουν στη ζητούμενη κατηγορία.

Β) Προσδιορίζουμε την ομάδα στην οποία ανήκει η τιμή 170 και τραβάμε γραμμή πάνω από αυτή. Στη συγκεκριμένη περίπτωση, η ομάδα είναι η 164-194.

Στη συνέχεια εφαρμόζουμε τον τύπο (11) έχοντας ως άγνωστο το α και θέτοντας στη θέση του Μ=170, Χ=164, δ=30, F=36 και Φ=146. Έτσι θα έχουμε :

$$170 = 164 + \frac{30}{36}\left(a - 146\right) \Leftrightarrow 170 - 164 = \frac{30a}{36} - \frac{30 \cdot 146}{36} \Leftrightarrow$$

$$\Leftrightarrow 6 + \frac{30 \cdot 146}{36} = \frac{30a}{36} \Leftrightarrow 6 + 122 = 0,83a \Leftrightarrow 128 = 0,83a \Leftrightarrow a = 154$$

Δηλαδή οι βιοτεχνίες της ερώτησης β είναι 200-154=46

Γ) προσδιορίζουμε από τη στήλη των αθροιστικών συχνοτήτων την $186^{η}$ θέση. Στη συγκεκριμένη περίπτωση, η δοσμένη θέση βρίσκεται μεταξύ $182^{ης}$ και $196^{ης}$ θέσης. Τραβάμε γραμμή μεταξύ αυτών των θέσεων και η ομάδα στην οποία θα ανήκει η ζητούμενη επένδυση είναι η 194-224. Στη συνέχεια εφαρμόζουμε τον τύπο (11) , έχοντας ως άγνωστο το Μ και θέτοντας στη θέση του Χ=194, δ=30, F=14, Φ=182 και α=186. Έτσι θα έχουμε:

$$M = 194 + \frac{30}{14}\left(186 - 182\right) \Leftrightarrow M = 202,6 \text{ χιλιάδες ευρώ.}$$

Δηλαδή η $186^{η}$ σε μέγεθος επένδυση είναι 202,6 χιλιάδες Euro. Μπορούμε επίσης να πούμε ότι οι υπόλοιπες 14 είναι ίσου ή μεγαλύτερου ύψους από αυτό το ποσό.

3.5 Χαρακτηριστικά Διαμέσου, Τεταρτημορίου κλπ.

- Δεν επηρεάζονται από τις ακραίες τιμές της μεταβλητής όπως ο αριθμητικός μέσος.
- Εφαρμόζονται σε κάθε περίπτωση, ακόμη και αν η κατανομή είναι ανοικτή.
- Επίσης μπορεί να εφαρμοστούν και σε ποιοτικές μεταβλητές σε ιεραρχική κλίμακα.

Όλες οι παράμετροι αυτής της παραγράφου, όταν πρόκειται για ομαδοποιημένες κατανομές, πρέπει να βρίσκονται στην ομάδα κάτω από τη γραμμή. Πιο συγκεκριμένα αν:

1. (θέση παραμέτρου-προηγούμενη αθροιστική συχνότητα) < (επόμενη αθροιστική συχνότητα – θέση παραμέτρου) τότε η παράμετρος βρίσκεται πιο κοντά στο μικρό άκρο του διαστήματος κάτω από τη γραμμή.
2. (θέση παραμέτρου – προηγούμενη αθροιστική συχνότητα) > (επόμενη αθροιστική συχνότητα – θέση παραμέτρου) τότε η παράμετρος βρίσκεται πιο κοντά στο μεγάλο άκρο του διαστήματος και
3. (θέση παραμέτρου – προηγούμενη αθροιστική συχνότητα) = (επόμενη αθροιστική συχνότητα – θέση παραμέτρου) τότε η ζητούμενη παράμετρος είναι η κεντρική τιμή του διαστήματος.

3.6 Σημείο Μέγιστης Συχνότητας – Τύπο- επικρατούσα τιμή – Mode

Το σημείο μέγιστης συχνότητας είναι μία παράμετρος η οποία εκφράζει την τιμή της μεταβλητής με τη μεγαλύτερη συχνότητα. Μας δείχνει, δηλαδή, την τιμή εκείνη η οποία συναντάται συχνότερα στην κατανομή που εξετάζουμε.

Είναι μία από τις λίγες παραμέτρους οι οποίες εφαρμόζονται και σε ποιοτικές και κατηγορικές μεταβλητές.

i) Τύπος ποιοτικής και κατηγορικής μεταβλητής

Σε μία ποιοτική ή κατηγορική μεταβλητή προσδιορίζουμε στη στήλη των συχνοτήτων την τιμή της μεταβλητής με τη μεγαλύτερη συχνότητα και αυτή η τιμή είναι το σημείο μέγιστης συχνότητας.

Αν η κατανομή είναι δίκορφη, δηλαδή η μέγιστη συχνότητα εμφανίζεται σε δύο τιμές, τότε τα σημεία μέγιστης συχνότητας είναι δύο.

ii) Τύπος ποσοτικής μεταβλητής

iii) Τύπος αταξινόμητων – ταξινομημένων παρατηρήσεων

Αν οι παρατηρήσεις είναι αταξινόμητες, είναι απαραίτητο πρώτα να ταξινομηθούν σε πίνακα κατανομής συχνοτήτων και στη συνέχει από αυτόν, όπως και προηγουμένως, να υπολογίσουμε το σημείο μέγιστης συχνότητας.

iv) Τύπος ομαδοποιημένων παρατηρήσεων

Όταν οι παρατηρήσεις είναι ομαδοποιημένες, σε ομάδες με ίδιο πλάτος, εντοπίζουμε την ομάδα με τη μεγαλύτερη συχνότητα από τη στήλη των συχνοτήτων και μέσα σ' αυτή θα βρούμε το σημείο μέγιστης συχνότητας εφαρμόζοντας τον τύπο:

$$M_0 = X_i + \frac{\delta \cdot \Delta_1}{\Delta_1 + \Delta_2} \quad (12)$$

Όπου X_ι: το μικρό άκρο του διαστήματος με τη μεγαλύτερη συχνότητα

δ: το πλάτος του διαστήματος με τη μεγαλύτερη συχνότητα

Δ_1: Η διαφορά της συχνότητας της προηγούμενης ομάδας από τη μέγιστη συχνότητα και

Δ_2: Η διαφορά της συχνότητας της επόμενης ομάδας από τη μέγιστη συχνότητα.

Παράδειγμα 1^{o}: Η ομαδοποιημένη κατανομή του επόμενου πίνακα παρουσιάζει τις ημερήσιες πωλήσεις (σε Euro) 100 Fast – Food.

0μαδες X_i	F_i
700-800	4
800-900	7
900-1000	8
1000-1100	10
1100-1200	12
$\rightarrow 1200-1300 \leftarrow$	17
1300-1400	13
1400-1500	10
1500-1600	9
1600-1700	7
1700-1800	2
1800-1900	1
ΣΥΝΟΛΟ	100

Η μεγαλύτερη συχνότητα είναι 17 και η ομάδα στην οποία αντιστοιχεί είναι η 1.200-1.300. Έτσι λοιπόν θα έχουμε: X_i=1.200, δ=100, Δ_1=17-12=5, και Δ_2=17-13=4.

Αντικαθιστώντας στον τύπο (12) έχουμε:

$$M_0 = 1200 + \frac{100 \cdot (17-12)}{(17-12)+(17-13)} = 1255,56€$$

! Ιδιαίτερη προσοχή χρειάζεται ο υπολογισμός του Τύπου, όταν οι ομάδες δεν έχουν ίσα πλάτη. Στην περίπτωση αυτή, πρέπει πρώτα να δημιουργήσουμε τη στήλη των τυποποιημένων συχνοτήτων F_i, στη συνέχεια να προσδιορίσουμε την ομάδα με τη μεγαλύτερη συχνότητα στη στήλη F_i και να εφαρμόσουμε τον τύπο (12) στη στήλη F_i των συχνοτήτων θέτοντας όπου Δ_1 και Δ_2 $|\Delta_1|$ και $|\Delta_2|$ αντίστοιχα.

Παράδειγμα 2: Ο παρακάτω πίνακας παρουσιάζει τους εβδομαδιαίους μισθούς των 560 εργατών μιας βιομηχανίας μετάλλου σε Euro.

Μισθοί X_i	F_i	δ_i	$\delta_i=\delta_i/\delta$	$F'_i=F_i/\delta'_i$
120-140	50	20	20/20=1	50/1=50
140-160	90	20	20/20=1	90/1=90
160-200	110	40	40/20=2	110/2=55
200-250	190	50	50/20=2,5	190/2,5=76
250-300	70	50	50/20=2,5	70/2,5=28
300-400	40	100	100/20=5	40/5=8
400-500	10	100	100/20=5	10/5=2
ΣΥΝΟΛΟ	560			

Επειδή οι ομάδες δεν έχουν το ίδιο πλάτος, θα εφαρμόσουμε όσα ανέφερα προηγουμένως και θα έχουμε:

$M_o = 140 + 20\frac{|90-50|}{|90-50|+|90-110|} = 140 + 20 \cdot 0,67 = 153,33$ Θα είναι ο συνηθέστερος μισθός.

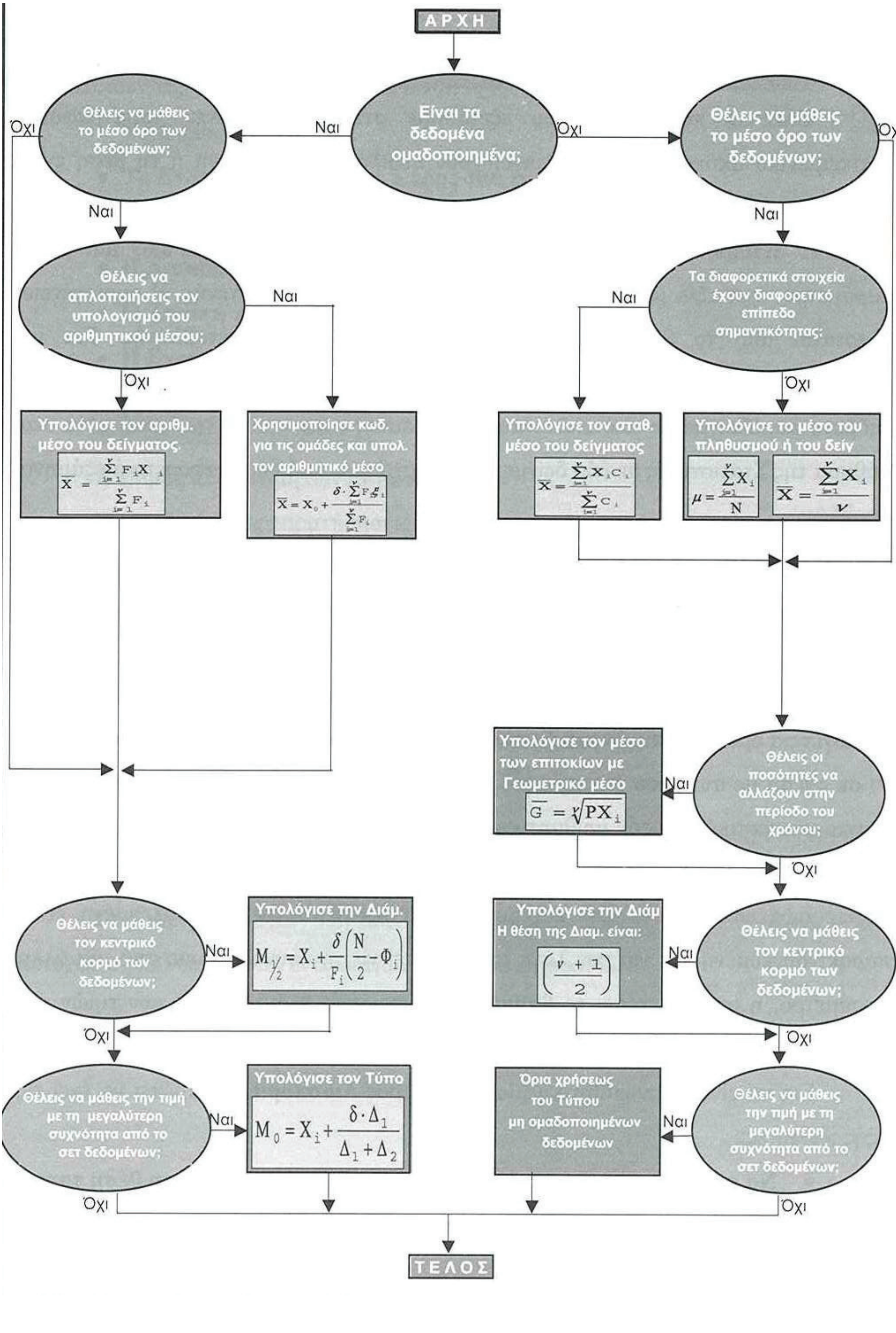
3.6.1 Γενικά χαρακτηριστικά Τύπου

Σαν γενικά χαρακτηριστικά του τύπου μπορούμε να αναφέρουμε τα εξής:

- Ο Τύπος μπορεί να χρησιμοποιηθεί σε ποσοτικές αλλά και σε ποιοτικές μεταβλητές.
- Δεν επηρεάζεται από τις ακραίες τιμές της κατανομής.
- Μπορεί να εφαρμοστεί ακόμη και σε ανοικτές κατανομές.
- Σε πολύ ασύμμετρες κατανομές είναι πιο αντιπροσωπευτικός από τον Αριθμητικό μέσο.
- Σε κατανομές με δύο ή και περισσότερα σημεία μέγιστης συχνότητας είναι δύσκολο να υπολογιστεί και να παρασταθεί.

Αν σε μία ομαδοποιημένη κατανομή (ίσες ή άνισες ομάδες) η μέγιστη συχνότητα εμφανίζεται δύο ή περισσότερες φορές, μπορούμε να δεχτούμε ότι έχουμε δύο ή περισσότερους Τύπους. Μπορούμε όμως να κάνουμε και σύμπτυξη των ομάδων, αν αυτό είναι εφικτό από το πλήθος αυτών και στη συνέχεια από τη νέα μορφή της κατανομής να υπολογίσουμε τον Τύπο. Επίσης μία λύση είναι η εκ νέου ομαδοποίηση, σε διαφορετικό αριθμό ομάδων, που όμως απαιτεί περισσότερο κόπο.

!!! Σε ομαδοποιημένες κατανομές η τιμή του Τύπου βρίσκεται στην ομάδα με τη μεγαλύτερη συχνότητα και είναι κοντά στο μικρό άκρο του διαστήματος, αν $\Delta_1 < \Delta_2$, ενώ είναι κοντά στο μεγάλο άκρο του διαστήματος αν $\Delta_1>\Delta_2$. Τέλος αν $\Delta_1=\Delta_2$ ο Τύπος είναι η κεντρική τιμή αυτού του διαστήματος.


ΑΡΧΗ
Είναι τα δεδομένα ομαδοποιημένα;
Ναι
Όχι
Θέλεις να μάθεις το μέσο όρο των δεδομένων;
Ναι
Όχι
Θέλεις να απλοποιήσεις τον υπολογισμό του αριθμητικού μέσου;
Ναι
Όχι
Υπολόγισε τον αριθμ. μέσο του δείγματος.
$\overline{X} = \frac{\sum_{i=1}^{\nu} F_i X_i}{\sum_{i=1}^{\nu} F_i}$
Χρησιμοποίησε κωδ. για τις ομάδες και υπολ. τον αριθμητικό μέσο
$\overline{X} = X_0 + \frac{\delta \cdot \sum_{i=1}^{\nu} F_i \xi_i}{\sum_{i=1}^{\nu} F_i}$
Θέλεις να μάθεις το μέσο όρο των δεδομένων;
Ναι
Όχι
Τα διαφορετικά στοιχεία έχουν διαφορετικό επίπεδο σημαντικότητας;
Ναι
Όχι
Υπολόγισε τον σταθ. μέσο του δείγματος
$\overline{X} = \frac{\sum_{i=1}^{\nu} X_i C_i}{\sum_{i=1}^{\nu} C_i}$
Υπολόγισε το μέσο του πληθυσμού ή του δείγ
$\mu = \frac{\sum_{i=1}^{\nu} X_i}{N}$
$\overline{X} = \frac{\sum_{i=1}^{\nu} X_i}{\nu}$
Θέλεις οι ποσότητες να αλλάζουν στην περίοδο του χρόνου;
Ναι
Όχι
Υπολόγισε τον μέσο των επιτοκίων με Γεωμετρικό μέσο
$\overline{G} = \sqrt[\nu]{PX_i}$
Θέλεις να μάθεις τον κεντρικό κορμό των δεδομένων;
Ναι
Όχι
Υπολόγισε την Διάμ.
$M_{1/2} = X_i + \frac{\delta}{F_i}\left(\frac{N}{2} - \Phi_i\right)$
Θέλεις να μάθεις τον κεντρικό κορμό των δεδομένων;
Ναι
Όχι
Υπολόγισε την Διάμ
Η θέση της Διαμ. είναι:
$\left(\frac{\nu + 1}{2}\right)$
Θέλεις να μάθεις την τιμή με τη μεγαλύτερη συχνότητα από το σετ δεδομένων;
Ναι
Όχι
Υπολόγισε τον Τύπο
$M_0 = X_i + \frac{\delta \cdot \Delta_1}{\Delta_1 + \Delta_2}$
Θέλεις να μάθεις την τιμή με τη μεγαλύτερη συχνότητα από το σετ δεδομένων;
Ναι
Όχι
Όρια χρήσεως του Τύπου μη ομαδοποιημένων δεδομένων
ΤΕΛΟΣ

4. Μέτρα Διασποράς – Dispersion

Όλες οι παράμετροι που εξετάσαμε στις προηγούμενες παραγράφους, αντικειμενικό σκοπό έχουν να περιγράψουν περιληπτικά την τάση ή τη θέση ενός πληθυσμού ή δείγματος.

Η αντιπροσώπευση όμως ενός ολόκληρου σετ δεδομένων από μία μονό παράμετρο, έχει αξία μόνο όταν η κατανομή είναι πολύ ομοιογενής. Η ομοιογένεια εξαρτάται από το βαθμό συγκέντρωσης των τιμών της μεταβλητής. Για να κατανοήσουμε καλύτερα την έννοια της ομοιογένειας θα παρακολουθήσουμε στο επόμενο παράδειγμα τα αποτελέσματα μιας έρευνας σχετικής με τη βαθμολογία στο μάθημα της Στατιστικής σε δύο δείγματα, προερχόμενα από δύο διαφορετικά εξάμηνα σπουδών.

Εαρινό Εξάμηνο: 3,4,5,7,8,9,7,6,7,8,3,4,5,6,7,8,9,4,7,5.

Χειμερινό Εξάμηνο: 9,8,3,4,5,5,1,0,9,8,7,6,5,9,7,9,10,10,2,5.

Η μέση βαθμολογία στα δύο δείγματα είναι $\overline{X}_E = 6,1$ και $\overline{X}_X = 6,1$. Παρατηρούμε δηλαδή ότι η μέση βαθμολογία και στα δύο εξάμηνα είναι ίδια. Η ομοιογένεια όμως των τιμών του Εαρινού εξαμήνου είναι σαφώς πιο μικρή (από 3 έως 9) σε σχέση με αυτή του Χειμερινού εξαμήνου (από 0 έως 10). Για το λόγο αυτό, η αντιπροσωπευτικότητα του αριθμητικού μέσου στο Εαρινό εξάμηνο είναι μεγαλύτερη σε σχέση με το Χειμερινό εξάμηνο. Όταν η κατανομή είναι ανομοιογενής, η αντιπροσώπευση από μία μόνο παράμετρο, όσο σωστά και αν είναι επιλεγμένη, είναι παρακινδυνευμένη και για το λόγο αυτό πρέπει να την συνοδεύουμε από μία άλλη παράμετρο, η οποία μετράει το βαθμό συγκέντρωσης ή διασποράς των τιμών της μεταβλητής.

Για να είναι ικανοποιητική μία παράμετρος Διασποράς θα πρέπει να έχει τις εξής ιδιότητες:

- Να επηρεάζεται από τις διαφορές των τιμών και όχι από τη θέση τους.
- Να μεταβάλλεται αντίστροφα με τη συγκέντρωση των τιμών γύρω από μία παράμετρο θέσης.

Τα κυριότερα μέτρα Διασποράς, τα οποία μεμονωμένα ή συνολικά ικανοποιούν τις παραπάνω ιδιότητες, είναι:

- Το Εύρος – Range
- Το Ενδοτεταρτημοριακό Εύρος – Interquartile Range
- Η Μέση Απόλυτη Απόκλιση – Mean Deviation
- Η Διακύμανση – Variance και
- Η Τυπική Απόκλιση – Standard Deviation
- Η Σχετική Διασπορά – Relative Dispersion

4.1 Εύρος - Range

i) Εύρος αταξινόμητων παρατηρήσεων

Σε αταξινόμητες παρατηρήσει, Εύρος είναι η διαφορά της μικρότερης από την μεγαλύτερη τιμή των παρατηρήσεων: R=Max(X)-min(X) (20).

Παράδειγμα: Τα χρήματα τα οποία ξοδεύουν 10 σπουδαστές κατά τη διάρκεια ενός σπουδαστικού έτους, σε Euro, είναι:

2.650, 3.500, 2.055, 2.800, 3.380, 3.670, 4.100, 1.760, 2.200, 1.910.

Το Εύρος των τιμών αυτών είναι R=4.100-1.760=2.340 Euro.

ii) Εύρος ομαδοποιημένων παρατηρήσεων

Σε μία ομαδοποιημένη κατανομή, Εύρος είναι η διαφορά της μικρότερης τιμής της πρώτης ομάδας από τη μεγαλύτερη τιμή της τελευταίας ομάδας.

Παράδειγμα: Στον Πίνακα 1Α παρουσιάζονται οι βαθμολογίες που πέτυχαν οι 680 υποψήφιοι ενός διαγωνισμού του Δημοσίου.

ΠΙΝΑΚΑΣ 1Α

Ομάδες X_i	Συχνότητα F_i
20-29	100
30-39	300
40-49	120
50-59	240
60-69	50
70-79	50
ΣΥΝΟΛΟ	860

Το Εύρος αυτής της ομαδοποιημένης κατανομής είναι R=79-20=59.

iii) Χαρακτηριστικά γνωρίσματα του Εύρους:

- Βασικό πλεονέκτημα του Εύρους είναι ότι υπολογίζεται εύκολα και είναι κατανοητό και από τον μη ειδικό.
- Είναι ιδανικό μέτρο μέτρησης της διασποράς σε μεγέθη στα οποία μας ενδιαφέρει η κατώτερη και ανώτερη τιμή, όπως οι τιμές των μετοχών, οι θερμοκρασίες μίας ημέρας, κλπ.
- Στο στατιστικό έλεγχο ποιότητας ενός προϊόντος, συνηθίζεται η λήψη πολλών μικρών δειγμάτων από τα οποία είναι εύκολο να υπολογιστεί το Εύρος. Από το εύρος αυτών των δειγμάτων παίρνουμε μία εικόνα για τη μεταβλητότητα της ποιότητας του υπό έλεγχο προϊόντος.
- Συνίσταται η εφαρμογή του σε ολιγοπληθή δείγματα.
- Θεωρείται πολύ γενικό μέτρο μέτρησης της διασποράς.

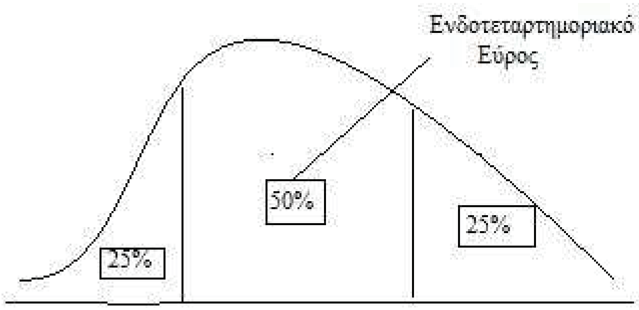
Βασικό μειονέκτημα του Εύρους είναι ότι επηρεάζεται η διαμόρφωσή του μόνο από δύο τιμές, τη μεγαλύτερη και τη μικρότερη. Αν μία από αυτές είναι πολύ μεγάλη ή πολύ μικρή τότε επηρεάζεται σημαντικά το μέγεθός του.

4.1.2 Ενδοτεταρτημοριακό Εύρος – Interquartile Range

Το Ενδοτεταρτημοριακό Εύρος είναι η διαφορά ανάμεσα στο τρίτο και το πρώτο τεταρτημόριο. $Q = M_{3/4} - M_{1/4}$ (13)

Η διαφορά αυτή προσδιορίζει το μήκος του διαστήματος στο οποίο κατανέμονται οι μισές και μάλιστα κεντρικές τιμές της μεταβλητής. Είναι ευνόητο ότι όσο μικρότερο είναι αυτό το μήκος τόσο μεγαλύτερη είναι η συγκέντρωση των τιμών της κατανομής.

Στο παρακάτω σχήμα φαίνεται παραστατικά το Q από τη διαφορά των τεταρτημορίων.



$M_{1/4}$ $M_{3/4}$

Παράδειγμα 1^{o}: Στον Πίνακα 2Α, ο οποίος παρουσιάζει τον αριθμό των παιδιών 30 οικογενειών, θα υπολογίσουμε το Q, αφού πρώτα βρούμε το τρίτο και το πρώτο τεταρτημόριο.

Πίνακας 2Α

Αριθμός παιδιών X_i	Αριθμός οικογενειών Συχνότητα F_i	Αθροιστική Συχνότητα Φ_i
	2	27
5	3	30
ΣΥΝΟΛΟ	30	

7,5= θέση 1^{ου} τεταρτημορίου

22,5= θέση 3^{ου} τεταρτημορίου

Η θέση του πρώτου και του τρίτου τεταρτημορίου είναι 30/4=7,5 και (3-30)/4=22,5 αντίστοιχα. Οι δε τιμές αυτών θα είναι: $M_{1/2}=1$ και $M_{3/4}=3$ σύμφωνα με όσα αναφέραμε στη σχετική παράγραφο. Το Q για αυτές τις τιμές θα είναι Q=3-1=2.

Χαρακτηριστικά μμ μ γνωρίσματα Ενδοτεταρτημοριακού Εύρους.

- Δεν επηρεάζεται από τις ακραίες τιμές της μεταβλητής.
- Μπορεί να υπολογιστεί και σε ανοιχτές κατανομές.
- Όσο μικρότερο είναι τόσο μεγαλύτερη είναι η συγκέντρωση των τιμών της μεταβλητής.
- Χρησιμοποιείται συνήθως όταν σαν παράμετρο θέσης χρησιμοποιούμε τη Διάμεσο.
- Δε μας δίνει καμία πληροφορία για την κατανομή του 50% των τιμών μέσα στο διάστημα, ενώ αγνοεί τελείως το 25% των μικρότερων τιμών του πρώτου τεταρτημορίου και το 25% των μεγαλύτερων τιμών του τρίτου τεταρτημορίου.

4.2. Μέση Απόλυτη Απόκλιση – Mean Deviation

Σαν μέση απόλυτη απόκλιση ορίζουμε τον μέσο των διαφορών κατ' απόλυτο τιμή, όλων των τιμών της μεταβλητής από τον αριθμητικό μέσο.

i) Μ.Α.Α. Αταξινόμητων παρατηρήσεων.

Όταν οι παρατηρήσεις είναι αταξινόμητες, η μέση απόκλιση δίνεται από τον τύπο: $M.A.A = \dfrac{\sum\left|X_i - \overline{X}\right|}{N}$ (14) όπου $\left|X_i - \overline{X}\right|$ είναι οι διαφορές των τιμών της μεταβλητής από τον αριθμητικό μέσο κατ’ απόλυτο τιμή, γιατί σύμφωνα με την ιδιότητα του αριθμητικού μέσου το $\sum\left(X_i - \overline{X}\right)$.

Παράδειγμα: Αν οι τιμές της μεταβλητής είναι 5,6,7,8 και 9, τότε ο

$\overline{X} = \dfrac{5+6+7+8+9}{5} = 7$ και $M.A.A = \dfrac{\left|5-7\right| + \left|6-7\right| + \left|7-7\right| + \left|8-7\right| + \left|9-7\right|}{5} = 1,2$

ii) Μ.Α.Α. Ταξινομημένων – Ομαδοποιημένων παρατηρήσεων

Όταν οι παρατηρήσεις είναι ταξινομημένες, ο τύπος που δίνει την Μ.Α.Α. είναι ο επόμενος $M.A.A = \frac{\sum F_i \left| X_i - \bar{X} \right|}{\sum F_i}$ (15)

Σε περίπτωση ομαδοποιημένων παρατηρήσεων, χρησιμοποιούμε τον ίδιο τύπο βάζοντας στη θέση του X_i την κεντρική τιμή του κάθε διαστήματος, όπως σε ανάλογες περιπτώσεις υπολογισμού άλλων παραμέτρων ομαδοποιημένων κατανομών.

Παράδειγμα 1^{ο}: Στον παρακάτω πίνακα παρουσιάζεται ο αριθμός των υπάλληλων 30 επιχειρήσεων. Αφού πρώτα υπολογίσουμε τον μέσο αριθμό παιδιών ανά οικογένεια, θα προσδιορίσουμε τη διασπορά γύρω από αυτό το μέσο με τη Μ.Α.Α.

Αριθμός υπαλ.	Αριθμος επιχ F_i	F_iX_i	$\left\lvert X_i - \bar{X} \right\rvert$	$F_i \left\lvert X_i - \bar{X} \right\rvert$
0	4	$0 \cdot 4 = 0$	$\lvert 0 - 2,07 \rvert = 2,07$	$4 \cdot 2,07 = 8,28$
1	7	$1 \cdot 7 = 7$	$\lvert 1 - 2,07 \rvert = 1,07$	$7 \cdot 1,07 = 7,49$
2	9	$2 \cdot 9 = 18$	$\lvert 2 - 2,07 \rvert = 0,07$	$9 \cdot 0,07 = 0,63$
3	5	$3 \cdot 5 = 15$	$\lvert 3 - 2,07 \rvert = 0,93$	$5 \cdot 0,93 = 4,65$
4	2	$4 \cdot 2 = 8$	$\lvert 4 - 2,07 \rvert = 1,93$	$2 \cdot 1,93 = 3,86$
5	3	$5 \cdot 3 = 15$	$\lvert 5 - 2,07 \rvert = 2,93$	$3 \cdot 2,93 = 8,79$
ΣΥΝΟΛΟ	30	62	9	33,70

Έχουμε $\bar{X} = 62/30 = 2,07 \cong 2$ υπάλληλοι ανά επιχείρηση και Μ.Α.Α.=33,7/30=1,21 $\cong$ 1 υπάλληλος.

Δηλαδή κατά μέσο όρο, ο αριθμός των παιδιών κάθε οικογένειας διαφέρει από τον μέσο όρο των υπαλλήλων κατά 1 υπάλληλο.

Παράδειγμα 2^{o}: Στον παρακάτω Πίνακα δίνονται οι εβδομαδιαίοι μισθοί 370 υπαλλήλων μίας επιχείρησης σε Euro και θέλουμε να δούμε όχι μόνο το μισθό που κατά μέσο όρο παίρνουν οι υπάλληλοι, αλλά και την κατά μέσο όρο απόκλιση των μισθών τους από το μέσο μισθό.

Ομάδες μισθών X_i	F_i	Κέντρο ομάδας X_i	F_iX_i	$\lvert X_i - \bar{X} \rvert$	$F_i \lvert X_i - \bar{X} \rvert$
120-140	30	130	3900	$\lvert 130-190 \rvert = 60$	$30 \cdot 60 = 1800$
140-160	50	150	7500	$\lvert 150-190 \rvert = 40$	$50 \cdot 40 = 2000$
160-180	60	170	10200	$\lvert 170-190 \rvert = 20$	$60 \cdot 20 = 1200$
180-200	90	190	17100	$\lvert 190-190 \rvert = 0$	$90 \cdot 0 = 0$
200-220	60	210	12600	$\lvert 210-190 \rvert = 20$	$60 \cdot 20 = 1200$
220-240	50	230	11500	$\lvert 230-190 \rvert = 40$	$50 \cdot 40 = 2000$
240-260	30	250	7500	$\lvert 250-190 \rvert = 60$	$30 \cdot 60 = 1800$
ΣΥΝΟΛΟ	370		70300	240	10000

Ο μέσος μηνιαίος μισθός είναι $\bar{X}$ =70.300/370=190 Euro. Η Μ.Α.Α σύμφωνα με τον τύπο (15) θα είναι Μ.Α.Α=10.000/370=27 Euro που σημαίνει ότι κατά μέσο όρο οι μισθοί των υπαλλήλων διαφέρουν από το μέσο μισθό 27 Euro.

iii) Χαρακτηριστικά γνωρίσματα της Μ.Α.Α.

- Διαμορφώνεται από όλες τις τιμές της μεταβλητής, σε αντίθεση με τις προηγούμενες που στηρίζονται μόνο σε δύο τιμές.
- Είναι απόλυτα κατανοητό μέτρο διασποράς.
- Δεν προσφέρεται για περαιτέρω στατιστικού υπολογισμούς και αυτό είναι το βασικό μειονέκτημά της.

4.3 Διακύμανση και Τυπική Απόκλιση – Variance and Standard Deviation

Η Διακύμανση και η Τυπική Απόκλιση είναι οι δύο σπουδαιότερες παράμετροι Διασποράς. Η χρησιμότητά τους στη διενέργεια στατιστικών ερευνών είναι πολύ μεγάλη.

Με τον όρο **Διακύμανση,** εννοούμε τον αριθμητικό μέσο των τετραγώνων των αποκλίσεων όλων των τιμών της μεταβλητής από τον αριθμητικό μέσο. Επειδή όμως εκφράζεται σε μονάδες οι οποίες είναι τα τετράγωνα των μονάδων μέτρησης της μεταβλητής, για να επανέλθουμε στην αρχική και σωστή μονάδα μέτρησης υπολογίζουμε την τετραγωνική ρίζα της διακύμανσης, εισάγοντας έτσι μία νέα παράμετρο την οποία ονομάζουμε Τυπική Απόκλιση.

Θα συμβολίζουμε τη διακύμανση του πληθυσμού με σ^2 ενώ του δείγματος με s^2. Αντίστοιχα, η τυπική απόκλιση του πληθυσμού συμβολίζεται με σ και του δείγματος με s. Μπορούμε επίσης να συμβολίσουμε τη διακύμανση με V, οπότε ισχύει: $V = \sigma^2$ ή $\sigma = \sqrt{V}$

i) **Διακύμανση – Τυπική Απόκλιση αταξινόμητων παρατηρήσεων**

Όταν οι παρατηρήσεις είναι λίγες, δεν υπάρχει λόγος να ταξινομηθούν ή να ομαδοποιηθούν.

Για τον υπολογισμό των δύο παραμέτρων διακρίνουμε τρεις περιπτώσεις:

α) Δεδομένα που αποτελούν τον πληθυσμό

Σε αυτή την περίπτωση χρησιμοποιούμε τους παρακάτω τύπους:

$$\sigma^2 = \frac{\sum_{\iota=1}^{N}\left(X_\iota - \mu\right)^2}{N} \quad (16) \text{ ή } \sigma^2 = \frac{1}{N}\sum_{\iota=1}^{N} X_\iota^2 - \mu^2 \quad (17)$$ **Διακύμανση**

και

$$\sigma = \sqrt{\frac{\sum_{\iota=1}^{N}\left(X_{\iota}-\mu\right)^{2}}{N}} \quad (18) \text{ ή } \quad \sigma = \sqrt{\frac{1}{N}\sum_{\iota=1}^{N}X_{\iota}^{2}-\mu^{2}} \quad (19)$$ **Τυπική απόκλιση**

Όπου: σ^2: Διακύμανση του πληθυσμού

σ: Τυπική Απόκλιση του πληθυσμού

μ: Μέση τιμή του πληθυσμού

μ^2: Τετράγωνο μέσης τιμής πληθυσμού

X_{ι}^{2}: Τετράγωνο των τιμών της μεταβλητής

N: Το πλήθος των ατόμων του πληθυσμού

$\sum_{\iota}^{N}\left(X_{\iota}-\mu\right)^{2}$: Άθροισμα τετραγώνων των διαφορών όλων των τιμών της μεταβλητής από το μέσο του πληθυσμού.

Παράδειγμα 1^o: Γνωρίζοντας ότι τις 7 μέρες μια εβδομάδας έχουμε ατομικά έξοδα 6,7,5,8,9,11 και 5 Euro μπορούμε εύκολα να υπολογίσουμε ότι κατά μέσο όρο ξοδεύουμε:

$$\mu = \frac{6+7+8+9+11+5}{7} = 7,29€$$

Για να υπολογίσουμε τώρα τη διακύμανση και την τυπική απόκλιση, αν υποθέσουμε ότι αυτές οι επτά μέρες αποτελούν τον πληθυσμό, θα χρησιμοποιήσουμε τους τύπους (16) ή (17) και (18) ή (19) , αφού πρώτα δημιουργήσουμε τα τετράγωνα των τιμών X_ι και των διαφορών X_ι-μ_ι όπως φαίνεται παρακάτω:

X_i	$(X_i-\mu)^2$	X_i^2
2000	$(6-7,29)^2=1,66$	36
2500	$(7-7,29)^2=0,08$	49
1750	$(5-7,29)^2=5,24$	25
2700	$(8-7,29)^2=0,50$	64
3000	$(9-7,29)^2=2,92$	81
3750	$(11-7,29)^2=13,76$	121
1800	$(5-7,29)^2=5,24$	25
ΣΥΝΟΛΟ	29,4	401

Έτσι θα έχουμε: $\sigma^2=\dfrac{29,4}{7}=4,2$ ή $\sigma^2=\dfrac{401}{7}-7,29^2=4,2$

και $\sigma=\sqrt{\dfrac{29,4}{7}}=2,05$ ή $\sigma=\sqrt{\dfrac{401}{7}-7,29^2}=2,05€$.

β) Δεδομένα που αποτελούν μεγάλο δείγμα ενός πληθυσμού

Για λόγους που αναφέραμε στην Εισαγωγή, στις στατιστικές έρευνες, συνήθως χρησιμοποιούμε δείγματα. Αν το μέγεθος του δείγματος είναι μεγάλο (τουλάχιστον 30 μονάδες), τότε οι τύποι που χρησιμοποιούμε διαφέρουν από τους προηγούμενους μόνον ως προς τους συμβολισμούς. Έτσι θα έχουμε:

$$s^2=\frac{\sum_{i=1}^{N}(X_i-\bar{X})^2}{N} \quad (20) \quad \text{ή} \quad s^2=\frac{1}{N}\sum_{i=1}^{N}X_i^2-\bar{X}^2 \quad (21)$$ **Διακύμανση**

$$s=\sqrt{\frac{\sum_{i=1}^{N}(X_i-\bar{X})^2}{N}} \quad (22) \quad \text{ή} \quad s=\sqrt{\frac{1}{N}\sum_{i=1}^{N}X_i^2-\bar{X}^2} \quad (23)$$ **Τυπική απόκλιση**

Όπου S^2: Διακύμανση του δείγματος

S: Τυπική Απόκλιση του δείγματος

$\bar{X}$: Μέση τιμή του δείγματος

$\bar{X}^2$: Τετράγωνο μέσης τιμής δείγματος

X_i^2 : Τετράγωνο των τιμών της μεταβλητής

Ν: Το πλήθος των ατόμων του δείγματος

$\sum_{i=1}^{N}(X_i - \bar{X})^2$: Άθροισμα τετραγώνων των διαφορών όλων των τιμών της μεταβλητής από τον μέσον του δείγματος.

ii) Διακύμανση και Τυπική Απόκλιση Ταξινομημένων – ομαδοποιημένων παρατηρήσεων

Όταν οι παρατηρήσεις είναι πολλές και η μορφή της μεταβλητής τέτοια που μας αναγκάζει να δημιουργήσουμε ένα πίνακα κατανομής συχνοτήτων χρησιμοποιούμε τους παρακάτω τύπους:

ΠΛΗΘΥΣΜΟΣ

$$\sigma^2 = \frac{\sum_{i=1}^{\nu} F_i (X_i - \mu)^2}{\sum_{i=1}^{\nu} F_i} \quad (30) \text{ ή } \quad \sigma^2 = \frac{\sum_{i=1}^{\nu} F_i X_i^2}{\sum_{i=1}^{\nu} F_i} - \mu^2 \quad (31) \quad \text{Διακύμανση}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{\nu} F_i (X_i - \mu)^2}{\sum_{i=1}^{\nu} F_i}} \quad (32) \text{ ή } \quad \sigma = \sqrt{\frac{\sum_{i=1}^{\nu} F_i X_i^2}{\sum_{i=1}^{\nu} F_i} - \mu^2} \quad (33) \quad \text{τυπική απόκλιση}$$

ΜΕΓΑΛΟ ΔΕΙΓΜΑ

ή $s^2 = \dfrac{\sum_{i=1}^{N} F_i (X_i - \bar{X})^2}{N}$ (34) ή $s^2 = \dfrac{\sum_{i=1}^{\nu} F_i X_i^2}{\sum_{i=1}^{\nu} F_i} - \bar{X}^2$ (35) Διακύμανση

$s = \sqrt{\dfrac{\sum_{i=1}^{N} F_i (X_i - \bar{X})^2}{N}}$ (36) ή $s = \sqrt{\dfrac{\sum_{i=1}^{\nu} F_i X_i^2}{\sum_{i=1}^{\nu} F_i} - \bar{X}^2}$ (37) Τυπική απόκλιση

Παράδειγμα 1ο: Στον παρακάτω πίνακα, δίνεται ο αριθμός των ημερών θερινής αδείας των 25 υπαλλήλων μίας επιχείρησης. Υποθέτοντας ότι αυτοί οι 25 υπάλληλοι είναι ο πληθυσμός, αφού υπολογίσουμε τον αριθμητικό μέσο, θα βρούμε τη διακύμανση και την τυπική απόκλιση.

X_i	F_i	F_iX_i	$F_iX_i^2$	$(X_i-\mu)^2$	$F_i(X_i-\mu)^2$
21	2	$21\cdot 2 = 42$	$42\cdot 21 = 882$	$(21-24,36)^2 = 11,3$	$2\cdot 11,3 = 22,6$
22	3	$22\cdot 3 = 66$	$66\cdot 22 = 1452$	$(22-24,36)^2$=5,6	$3\cdot 5,6 = 16,8$
23	2	$23\cdot 2 = 46$	$46\cdot 23 = 1058$	$(23\text{-}24,36)^2$=1,8	$2\cdot 1,8 = 3,6$
24	5	$24\cdot 5 = 120$	$120\cdot 24 = 2880$	$(24\text{-}24,36)^2$=0,13	$5\cdot 0,13 = 0,65$
25	6	$25\cdot 6 = 150$	$150\cdot 25 = 3750$	$(25\text{-}24,36)^2$=0,4	$6\cdot 0,4 = 2,4$
26	4	$26\cdot 4 = 104$	$104\cdot 26 = 2704$	$(26\text{-}24,36)^2$=2,7	$4\cdot 2,7 = 10,8$
27	3	$27\cdot 3 = 81$	$81\cdot 27 = 2187$	$(27\text{-}24,36)^2$=6,9	$3\cdot 6,9 = 20,7$
ΣΥΝΟΛΟ	25	609	14913	28,83	77,5

Έχουμε $\mu = \dfrac{609}{25} = 24,36$ ημέρες θερινής αδείας και από τους τύπους (30) ή (31) θα έχουμε διακύμανση: $\sigma^2 = \dfrac{77,55}{25} = 3,102$ ή $\sigma^2 = \dfrac{14913}{25} - 24,36^2 = 3,102$. Ομοίως από τους τύπους (32) ή (33) θα βρούμε τυπική απόκλιση:

$\sigma = \sqrt{\dfrac{77,55}{25}} = 1,76$ ή $\sigma = \sqrt{\dfrac{14913}{25} - 24,36^2} = 1,76$ ημέρες θερινής άδειας.

Με τους ίδιους τύπους θα δουλέψουμε και όταν οι παρατηρήσεις είναι ομαδοποιημένες, βάζοντας στη θέση του X_i την κεντρική τιμής της ομάδας, όπως κάναμε και για τον υπολογισμό του αριθμητικού μέσου. Να θυμηθούμε βέβαια ότι αυτή η πράξη δημιουργεί ένα λάθος στον υπολογισμό των παραμέτρων και βέβαια διαφορετικό αποτέλεσμα από αυτό που θα είχαμε αν δεν ομαδοποιούμε τα δεδομένα.

Πολλές φορές όταν οι παρατηρήσεις είναι ομαδοποιημένες, όπως και στην περίπτωση του αριθμητικού μέσου, χρησιμοποιούμε τον έμμεσο τρόπο υπολογισμού της διακύμανσης και της τυπικής απόκλισης. Ο έμμεσος τρόπος συνιστάται κυρίως όταν οι τιμές της μεταβλητής αλλά και οι συχνότητες είναι πολύ μεγάλες. Έτσι έχουμε μικρότερα αριθμητικά αποτελέσματα και συνεπώς πιο εύκολες πράξεις.

Οι τύποι που δίνουν τη διακύμανση είναι οι επόμενοι:

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

για μεγάλο δείγμα (αν πρόκειται για πληθυσμό η μόνο αλλαγή στον τύπο είναι ο συμβολισμός σ αντί S).

Όπου δ: Το πλάτος του διαστήματος το οποίο επιλέξαμε σαν βάση

$\xi_\iota = \dfrac{X_i - X_0}{\delta}$ με X_i= κεντρική τιμή της κάθε ομάδας

X_0: Κεντρική τιμή της ομάδας την οποία επιλέξαμε σαν βάση και

$\sum_{i=1}^{\nu} F_i$: Το πλήθος των τιμών του πληθυσμού ή του δείγματος και

$$s = \delta \cdot \sqrt{ \frac{\sum_{i=1}^{\nu} F_i \xi_i^2}{\sum F_i} - \left(\frac{\sum_{i=1}^{\nu} F_i \xi_i}{\sum_{i=1}^{\nu} F_i} \right)^2 }$$

Παράδειγμα 2ο: Στον πίνακα 2.2.ε παρουσιάζονται οι εβδομαδιαίοι μισθοί των 100 εργατών μίας επιχείρησης. Θα υπολογίσουμε τη διακύμανση και την τυπική απόκλιση των μισθών με τον έμμεσο τρόπο για να δούμε τον τρόπο εφαρμογής του.

X_i	F_i	Κέντρο X_i	ξ_i	$F_i\xi_i$	$F_i\xi_i^2$
[150-250)	5	200	(200-500)/100=-3	5(-30)=-15	(-15)(-3)=45
[250-350)	13	300	(300-500)/100=-2	13(-2)=-26	(-2)(-26)=52
[350-450)	20	400	(400-500)/100=-1	20(-1)=-20	(-1)(-20)=20
[450-550)	35	500	(500-500)/100=0	35-0=0	0-0=0
[550-650)	18	600	(600-500)/100=1	18*1=18	1*18=18
[650-750)	7	700	(700-500)/100=2	7*2=14	2*14=28
[750-850)	2	800	(800-500)/100=3	2*3=6	3*6=18
ΣΥΝΟΛΟ	100			-23	181

Με αντικατάσταση στους τύπους τον παραπάνω τύπων θα έχουμε:

$$s^2 = 100^2\left[\frac{181}{100} - \left(\frac{23}{100}\right)^2\right] = 17.511 \text{ και } s = 100\sqrt{\frac{181}{100} - \left(\frac{-23}{100}\right)^2} = 132{,}56€.$$

Ιδιαίτερη προσοχή χρειάζεται η εφαρμογή αυτών των τύπων όταν οι ομάδες δεν έχουν ίσα πλάτη. Σαν δ χρησιμοποιούμε το πλάτος της ομάδας στην οποία βάζουμε το μηδέν. Στο επόμενο παράδειγμα θα ασχοληθούμε με τέτοια περίπτωση.

Παράδειγμα 3ο: Στον πίνακα έχουμε τον πληθυσμό των νέων Δήμων που προέκυψαν από συνενώσεις μικρών κοινοτήτων. Οι ομάδες των πληθυσμών δεν είναι ίσες.

Πληθυσμός X_i	F_i	Κέντρο X_i	ξ_i	$F_i\xi_i$	$F_i\xi_i^2$
0-2000	32	1000	(1000-4500)/1000=3,5	32(-3,5)=-112	(-3,5)(112)=392
2001-3000	13	2500	(2500-4500)/1000=-2	13(-2)=-26	(-2)(-26)=52
3001-4000	5	3500	(3500-4500)/1000=-1	5(-1)=-5	(-1)(-5)=5
4001-5000	3	4500	(4500-4500)/1000=0	3*0=0	0*0=0
5001-7000	5	6000	(6000-4500)/1000=1,5	5*1,5=7,5	1,5*7,5=11,25
7001-10000	1	8500	8500-4500)/1000=4	1*4=4	4*4=16
ΣΥΝΟΛΟ	59			-131,5	476,25

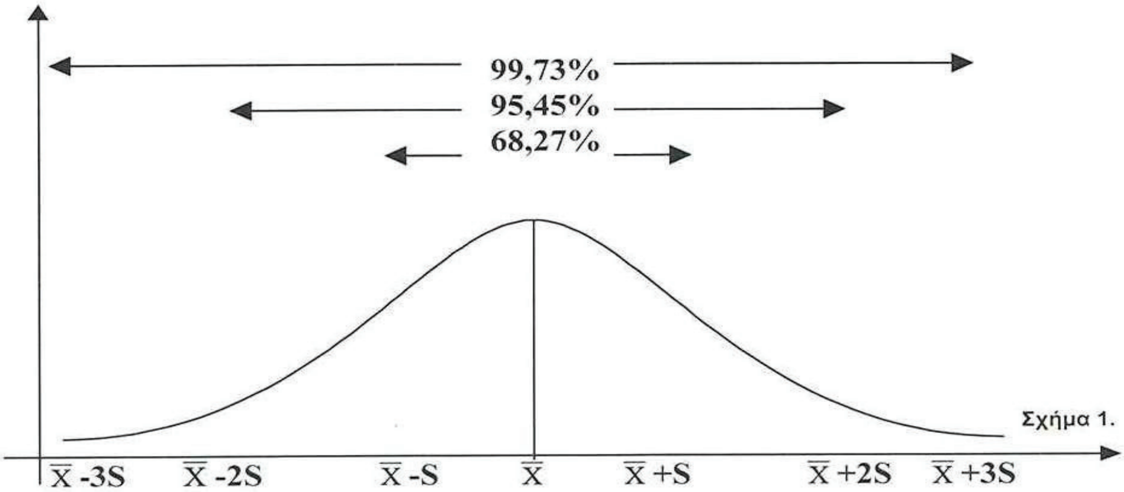
Εφαρμόζοντας τους τύπους θα έχουμε

$$s^2 = 1000^2 \left[\frac{476,25}{59} - \left(\frac{-131,5}{59} \right)^2 \right] = 31020400 \text{ και}$$

$$s = 1000 \sqrt{\frac{476,25}{59} - \left(\frac{-131,5}{59} \right)^2} = 1761,8 \text{ κάτοικοι ανά δήμο}$$

Ιδιότητες διακύμανσης και τυπικής απόκλισης

- Αν όλες οι τιμές της μεταβλητής είναι ίσες, τότε η διακύμανση και η τυπική απόκλιση είναι 0.
- Αν σε όλες τις τιμές της μεταβλητής προσθέσουμε ή αφαιρέσουμε μία σταθερή ποσότητα α, τότε η διακύμανση και η τυπική απόκλιση παραμένουν σταθερές.
- Αν όλες τις τιμές της μεταβλητής τις πολλαπλασιάσουμε ή τις διαιρέσουμε με μία σταθερή ποσότητα α, τότε η διακύμανση πολλαπλασιάζεται ή διαιρείται με α και η τυπική απόκλιση με α.
- Ειδικά σε μία κανονική κατανομή $\bar{X} = M_{1/2} = M_0$ το 68,27% των παρατηρήσεων βρίσκεται στο διάστημα $\bar{X} - S, \bar{X} + S$, το 95,45% των παρατηρήσεων στο διάστημα $\left(\bar{X} - 2S, \bar{X} + 2S\right)$ και το 99,73% των παρατηρήσεων στο διάστημα $\bar{X} - 3S, \bar{X} + 3S$.



- Μεταξύ των παραμέτρων διασποράς υπάρχουν οι παρακάτω εμπειρικές σχέσεις.

  Μέση Απόλυτη Απόκλιση = $\frac{4}{5}$ Τυπικής Απόκλισης

  Ενδοτεταρτημοριακό Εύρος = $\frac{4}{3}$ Τυπικής Απόκλισης

Η Τυπική Απόκλιση όσο μικρότερη είναι, τόσο περισσότερο συγκεντρωμένα είναι τα δεδομένα γύρω από τον αριθμητικό μέσο.

Δεν μπορούμε να συγκρίνουμε δύο κατανομές που εξετάζουν την ίδια μεταβλητή, έχοντας σαν μέτρο σύγκρισης την τυπική απόκλιση, παρά μόνο, αν έχουν τον ίδιο αριθμητικό μέσο.

4.3.2 Διόρθωση Διακύμανσης κατά Sheppard

Κατά τον υπολογισμό της διακύμανσης σε ομαδοποιημένη κατανομή, επειδή αντικαθιστούμε την ομάδα με την κεντρική της τιμή, κάνουμε κάποιο λάθος. Για να διορθώσουμε αυτό το λάθος χρησιμοποιούμε τον παρακάτω τύπο:

$\sigma^2 = \sigma^2 - \frac{\delta^2}{12}$ για τον πληθυσμό, με σ^2: Διορθωμένη διακύμανση και

$s^2 = s^2 - \frac{\delta^2}{12}$ για το δείγμα με s^2: Διορθωμένη διακύμανση δείγματος.

Οι τύποι αυτοί αποτελούν το αποτέλεσμα της έρευνας του W.F. Sheppard και για το λόγο αυτό φέρουν και το όνομά του.

Οι στατιστικολόγοι διαφωνούν για το πού και πότε πρέπει να χρησιμοποιούμε αυτή τη διόρθωση. Οπωσδήποτε όμως πριν κάνουμε χρήση αυτής θα πρέπει να εξετάζουμε πολύ καλά την κατάσταση.

Γενικά θα μπορούσαμε να πούμε ότι δίνει τα καλύτερα δυνατά αποτελέσματα όταν:

- Η κατανομή είναι συμμετρική και
- Οι ομάδες έχουν ίσα πλάτη.

Πρέπει να αποφεύγεται τελείως η χρήση της όταν:

- Οι ομάδες δεν έχουν ίσα πλάτη
- Η κατανομή δεν είναι συμμετρική και
- Το πλήθος των παρατηρήσεων είναι μεγαλύτερο του 500.

Παράδειγμα: Στον που είχαμε τις ομάδες μισθών των 100 εργατών μίας επιχείρησης. Με εφαρμογή των τύπων βρήκαμε ότι: S=17.571 και S=132,56.

Αν θέλουμε να υπολογίσουμε τη διορθωμένη διακύμανση και τη διορθωμένη τυπική απόκλιση, έχοντας υπ' όψη ότι το πλάτος των ομάδων είναι δ=100, με αντικατάσταση στον τύπο θα έχουμε:

$s_0^2 = s^2 - \frac{\delta^2}{12} = 17571 - \frac{100^2}{12} = 16737,67$ και συνεπώς $s_0 = \sqrt{16737,67} = 129,37$

4.4 Σχετική Διασπορά – Relative Dispersion

Οι παράμετροι διασποράς που είδαμε μέχρι τώρα και κυρίως η τυπική απόκλιση, εκφράζονται σε μονάδες μέτρησης αντίστοιχες της μεταβλητής την οποία εξετάζουν. Έτσι η σύγκριση της διασποράς δύο ή περισσότερων κατανομών που εξετάζουν διαφορετικά χαρακτηριστικά δεν είναι δυνατόν να γίνει με βάση αυτές.

Είναι λοιπόν ανάγκη να εισάγουμε μια νέα παράμετρο η οποία θα εκφράζει τη διασπορά αόριστα, χωρίς να αναφέρεται σε κάποια συγκεκριμένη μονάδα μέτρησης. Η παράμετρος αυτή είναι η σχετική διασπορά, η οποία εκφράζεται σαν ποσοστό επί τοις %.

Κύριος εκφραστής της σχετικής διασποράς είναι ο Συντελεστής Μεταβλητότητας (Coefficient of Variation) ο οποίος δίνεται από τους παρακάτω τύπους:

$C.V = \frac{\sigma}{\mu}100$ Συντελεστής Μεταβλητότητας πληθυσμού και

$C.V = \frac{s}{\overline{X}}100$ Συντελεστής Μεταβλητότητας δείγματος.

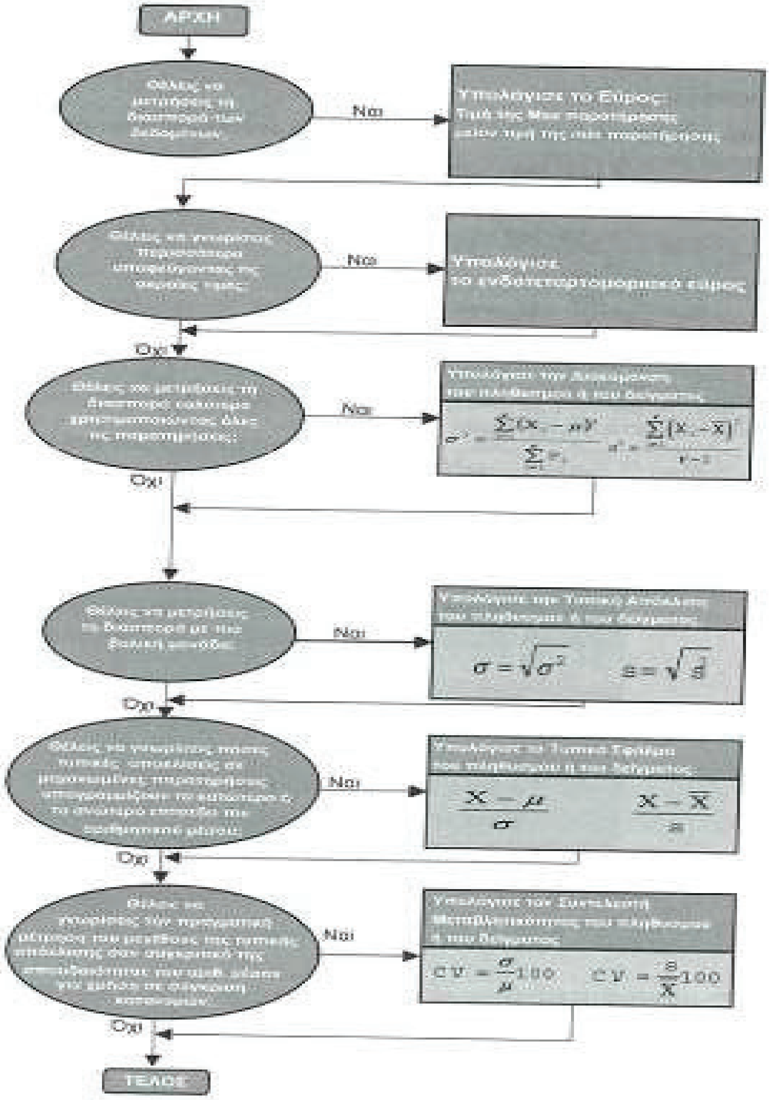
Παράδειγμα: Ένα εργοστάσιο παράγει δύο τύπους τηλεοπτικών αγωγών Α και Β. Οι αγωγοί έχουν μέση διάρκεια ζωής 1.495 και 1.875 ώρες αντίστοιχα και τυπική απόκλιση 280 και 310 ώρες αντίστοιχα. Ποιος τύπος αγωγού έχει καλύτερη σχετική διασπορά;

Αν θεωρήσουμε ότι τα νούμερα που δίνονται ανήκουν σε δείγματα, θα έχουμε:

$\overline{X}_A = 1495, \overline{X}_B = 1875$ και $s_A = 280, s_B = 310$.

Για να απαντήσουμε στο ερώτημα θα υπολογίσουμε με τον τύπο το συντελεστή μεταβλητότητας για τους δύο τύπους αγωγών και όπου προκύψει μικρότερο αποτέλεσμα εκεί θα έχουμε καλύτερη σχετική διασπορά.

Έχουμε: $C.V(A)=\frac{280}{1495}\cdot 100=18,73\%, C.V(B)=\frac{310}{1875}\cdot 100=16,53\%$ και με βάση αυτά τα αποτελέσματα μπορούμε να πούμε ότι ο Β τύπος αγωγού είναι καλύτερος.


ΑΡΧΗ
Ναι
Υπολόγισε το Εύρος:
Ναι
Υπολόγισε
το ενδοτεταρτομοριακό εύρος
Όχι
Ναι
Όχι
Ναι
$\sigma = \sqrt{\sigma^2}$
$s = \sqrt{s^2}$
Όχι
Ναι
$\frac{X - \mu}{\sigma}$
$\frac{X - \bar{X}}{s}$
Όχι
Ναι
$CV = \frac{\sigma}{\mu}100$
$CV = \frac{s}{\bar{X}}100$
Όχι
ΤΕΛΟΣ

4.5 Μέση διαφορά του Gini

Ως μέση διαφορά του Gini είναι και αυτή ένα μέτρο διασποράς που χρησιμοποιείται στις περιπτώσεις που ενδιαφερόμαστε για τις αποκλίσεις κάθε τιμής της μεταβλητής από τις υπόλοιπες, συμπεριλαμβανομένων και των ιδίων, οι οποίες προφανώς είναι ίσες με το μηδέν.

Άρα σε ομαδοποιημένες παρατηρήσεις –περίπτωση συνεχούς μεταβλητής- και με τάξεις ίσου πλάτους τότε το d δίνεται από την σχέση :

$$d = \frac{2\delta}{v^2}\sum_{i=1}^{v}(v_i - F_i)F_i$$

Όπου

δ : το πλάτος των τάξεων

ν: το πλήθος τον τιμών

F_i : οι αθροιστική συχνότητα

Συντελεστής Gini είναι ένα μέτρο της στατιστικής διασποράς το οποίο κυρίως χρησιμοποιημένος ως μέτρο της ανισότητας της εισοδηματικής διανομής ή ανισότητα της διανομής πλούτου. Ένας χαμηλός συντελεστής Gini δείχνει περισσότερη ίση διανομή εισοδήματος ή πλούτου, ενώ ένας υψηλός συντελεστής Gini δείχνει περισσότερη άνιση διανομή. 0 αντιστοιχούν στην τέλεια ισότητα (η καθεμία που έχει ακριβώς το ίδιο εισόδημα) και 1 αντιστοιχεί στην τέλεια ανισότητα (όπου ένα άτομο έχει όλο το εισόδημα, ενώ έχει μηδέν εισόδημα). Ο συντελεστής Gini απαιτεί ότι κανένας δεν έχει ένα αρνητικό καθαρό εισόδημα ή έναν πλούτο. Παγκοσμίως, οι συντελεστές Gini κυμαίνονται από περίπου 0.249 στην Ιαπωνία ως 0.707 στη Ναμίμπια.

Επαναληπτική άσκηση

Η κατανομή 164 βιοτεχνιών ανάλογα με τον αριθμό των απασχολούμενων εργατών έχει ως εξής:

Αριθμός εργατών	Αριθμός Βιοτεχν.	X_i	F_iX_i	$\lvert X_i - \overline{X} \rvert$	Φ_i	$F_i \lvert X_i - \overline{X} \rvert$	F_i-Φ_i
1-5	18	3	54	9,78	18	29,34	146
6-10	25	8	200	4,78	43	119,50	121
11-15	67	13	871	0	110	0	54
16-20	54	18	972	5,22	164	281,88	0
Σύνολο	164		2097	19,78		430,74	321

Ζητούνται : η μέση τιμή, διάμεσος, επικρατούσα τιμή, 1^{o} τεταρτημόριο, 3^{o} δεκατημόριο, ενδοταρτιμοριακιο εύρος, εύρος, μέση απόλυτη τιμή, τυπική απόκλιση, διακύμανση και μέση διαφορά του Gini.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i V_i}{n} = \frac{2097}{164} = 13$$

$M_{1/2} = X_i + \frac{\delta}{F_i}\left(\frac{N}{2} - \Phi_i\right) = 11 + \frac{4}{67}(82-43) = 13,32$ άρα οι μέσοι εργάτες είναι 13,32

$Q_{1/4} = X_i + \frac{\delta}{F_i}\left(\frac{N}{4} - \Phi_i\right) = 8 + \frac{4}{25}(41-18) = 11,68$ άρα το 25% έχει από 1 έως 11,68 εργάτες.

$Q_{3/10} = X_i + \frac{\delta}{F_i}\left(\frac{3N}{10} - \Phi_i\right) = 11 + \frac{4}{67}(49,2-43) = 11,37$ άρα το 30% έχει από 1 έως 11,37 εργάτες.

$M_o = X_i + \frac{\delta \cdot \Delta_1}{\Delta_1 + \Delta_2} = 11 + \frac{4 \cdot 42}{13+42} = 14,05$ άρα η μεγίστη τιμή είναι 14,5

$$Q = Q_{3/4} - Q_{1/4}$$

$Q_{3/4} = 16 + \frac{4}{54}(123-110) = 16,91$ άρα θα έχουμε

Q=16,91-11,68=5,23

Εύρος=F_{max}-F_{min}=20-1=19

M.A.A=430/164=2,62 που σημαίνει ότι κατά μέσο όρο οι εργάτες των βιοτεχνιών διαφέρουν από τον μέσο όρο εργατών κατά 2,62 εργάτες.

$$s^2 = \frac{\sum_{i=1}^{N} F_i (X_i - \overline{X})^2}{N} = \frac{(19,78)^2}{164} = 2,38$$

$\sqrt{s} = \sqrt{2,38}$ =1,54 εργάτες ανά βιοτεχνία

$$C.V = \frac{s}{\overline{X}} = \frac{1,54}{13} = 0,11$$

$$d = \frac{2\delta}{v^2} \sum_{i=1}^{v} (v_i - F_i) F_i = \frac{2 \cdot 4}{164^2} 321 \cdot 164 = 15,40$$

Επαναληπτική άσκηση 2

Ο παρακάτω πίνακας δίνει τις ημερήσιες αποδοχές των υπαλλήλων μιας επιχείρησης σε ευρώ.

Ημερήσιες Αποδοχές	F_i	X_i	F_iX_i	ξ_i	$F_i\xi_i$	Φ_i	$X_i - \overline{X}$	$F_i(X_i - \overline{X})$	F_i-Φ_i
20-24	18	22	396	-2,5	-45	18	-6.59	-118.62	162
25-29	120	27	3240	0	0	138	-1.59	-190.80	42
30-34	18	32	576	2,5	45	156	3.41	61.38	24
35-39	12	37	444	5	60	168	8.41	100.92	12
40-44	6	42	252	7,5	45	174	13.41	80.46	6
45-49	6	47	282	10	60	180	18.41	110.46	0
ΣΥΝΟΛΟ	180		5190		165		35.46	43.80	246

Ζητούνται α) μέση τιμή β) διάμεσος γ) επικρατούσα τιμή δ) Q_3, D_2, C_{20} ε) τυπική απόκλιση ζ) διακύμανση και μέση διαφορά Gini.

α) $\overline{X} = \dfrac{\sum_{i=1}^{n} X_i F_i}{n} = \dfrac{5190}{180} = 28.83$ μέσος ορός ημερομισθίου

β) $M = X_i + \dfrac{\delta}{F_i}\left[\dfrac{N}{2} - \Phi_i\right] = 27 + \dfrac{4}{180}\left[\dfrac{180}{2} - 18\right] = 28.59$

$\left(\dfrac{N}{2} = \dfrac{180}{2} = 90\right.$ άρα η διάμεσος εντοπίζεται στο Φ_i που ανήκει το 90 $\left.\right)$.

που σημαίνει ότι το 50% των εργαζομένων έχει ημερομίσθιο μέχρι 28,59€ και το άλλο 50% πάνω από 25,59€.

γ) $M_o = X_i + \delta \dfrac{\Delta_1}{\Delta_1 + \Delta_2} = 25 + 4\dfrac{102}{204} = 25 + 8 = 33$

Δ_1=120-18=102

Δ_2=120-18=102

δ) $Q_3 = X_i + \dfrac{\delta}{v_i}\left(\dfrac{KN}{4} - \Phi_i\right) = 21 + \dfrac{4}{120}(135 - 18) = 21 + 3.86 = 24.86$€ που σημαίνει ότι το 25% έχει ημερομίσθιο έως 24,86€ και το υπόλοιπο 65% έχει πάνω από αυτό.

$\dfrac{KN}{4} = 135$ άρα το X_i=21

$D_2 = X_i + \dfrac{\delta}{v_i}\left(\dfrac{KN}{10} - \Phi_i\right) = 21 + \dfrac{4}{120}(36 - 18) = 21 + 0.59 = 21.59$€

$\dfrac{KN}{10} = 36$

$C_{10} = X_i + \dfrac{\delta}{v_i}\left(\dfrac{KN}{100} - \Phi_i\right) = 20 + 0 = 20$€

$\dfrac{KN}{100} = 18$

ε) $s^2 = \dfrac{\sum_{i=1}^{N} F_i (X_i - \bar{X})^2}{N}$ =43.80/180=0.24

ζ) $\sqrt{s} = \sqrt{0.24} = 0.49$

$$d = \frac{2\delta}{v^2} \sum_{i=1}^{v} (v_i - F_i) F_i = \frac{2 \cdot 4}{180^2} 246 \cdot 180 = 10.93$$

ΒΙΒΛΙΟΓΡΑΦΙΑ

- **Στατιστική - Μέθοδοι ανάλυσης για επιχειρηματικές αποφάσεις,** Εκδότης: Rosili ,Συγγραφέας: Χαλικιάς, Ιωάννης Γ., Τόπος Έκδοσης: Αθήνα 2003
- **Στατιστική - Θεωρία και εφαρμογές**, Εκδότης: Ζήτη ,Συγγραφέας: Μπόρα - Σέντα, . Κολυβά - Μαχαίρα Φωτεινή ,Τόπος Έκδοσης: Θεσσαλονίκη 1998
- **Στατιστική επιχειρήσεων - Περιγραφική και επαγωγική στατιστική**, Εκδότης: Σύγχρονη Εκδοτική, Συγγραφέας: Αποστολόπουλος, Κωνσταντίνος Θ. , Τόπος Έκδοσης: Αθήνα 2004
- **Στατιστική** Εκδότης: Interbooks, Συγγραφέας: Κιόχος, Πέτρος . Τόπος Έκδοσης: Αθήνα 1993
- **Περιγραφική Στατιστική** Συγγραφείς: Ευστάθιος Δημητριάδης Εκδότης: Κριτική ΑΘΗΝΑ 2002
- **Περιγραφική στατιστική** Συγγραφείς: Πέτρος Κιόχος Εκδότης: Interbooks ΑΘΗΝΑ 1993