Ο Δημήτρης Καραγεώργος είναι μαθηματικός με διδακτορικές σπουδές στην Αγγλία. Υπηρέτησε τριάντα πέντε χρόνια στη Δευτεροβάθμια Εκπαίδευση ως καθηγητής, γυμνασιάρχης, λυκειάρχης, σχολικός σύμβουλος, μόνιμος πάρεδρος και σύμβουλος του Παιδαγωγικού Ινστιτούτου.

Σήμερα διδάσκει ως επίκουρος καθηγητής στο Πανεπιστήμιο Αθηνών (τομέας Παιδαγωγικής του τμήματος Φ.Π.Ψ.) τα μαθήματα «Διδακτική των θετικών επιστημών» και «Μεθοδολογία έρευνας στις επιστήμες της αγωγής».

Στη μακρόχρονη εκπαιδευτική του θητεία δίδαξε το μάθημα των μαθηματικών σε όλες τις τάξεις του Γυμνασίου και Λυκείου, ασχολήθηκε με τη σύνταξη αναλυτικών προγραμμάτων, τη συγγραφή διδακτικών εγχειριδίων, που διδάσκονται στη Δευτεροβάθμια Εκπαίδευση, την παραγωγή υποστηρικτικού διδακτικού υλικού για τα μαθηματικά και την επιμόρφωση των εκπαιδευτικών.

Έχει γράψει εννιά βιβλία μαθηματικών και έχει δημοσιεύσει σε ελληνικά και ξένα περιοδικά περισσότερες από πενήντα εργασίες που αναφέρονται στα προηγούμενα αντικείμενα και στον εκπαιδευτικό σχεδιασμό.

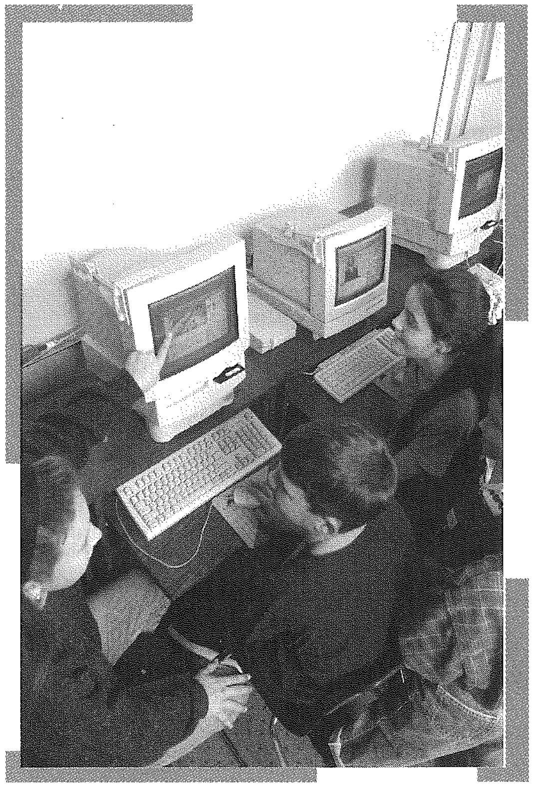
Ένα από τα θέματα που τον απασχόλησαν την τελευταία εικοσαετία ήταν και η εκπαιδευτική έρευνα. Ένα μέρος της προσπάθειάς του αυτής καταγράφεται σε αυτό το βιβλίο.

Δημήτρης Λ. Καραγεώργος

ΕΠΙΚΟΥΡΟΣ ΚΑΘΗΓΗΤΗΣ ΠΑΝΕΠΙΣΤΗΜΙΟΥ ΑΘΗΝΩΝ
Επιτ. ΣΥΜΒΟΥΛΟΣ ΤΟΥ ΠΑΙΔΑΓΩΓΙΚΟΥ ΙΝΣΤΙΤΟΥΤΟΥ

ΜΕΘΟΔΟΛΟΓΙΑ ΕΡΕΥΝΑΣ
στις επιστήμες της αγωγής

ΜΙΑ ΔΙΔΑΚΤΙΚΗ ΠΡΟΣΕΓΓΙΣΗ



Σαββάλας
ΕΚΔΟΣΕΙΣ

α) Για την κατηγοριακή κλίμακα μέτρησης τα κυριότερα στατιστικά κριτήρια που χρησιμοποιούνται είναι: η επικρατούσα τιμή, οι συχνότητες, τα ποσοστά, το ιστόγραμμα, το διωνυμικό κριτήριο z, το κριτήριο x^2 κ.λπ.
β) Για τη διατακτική κλίμακα μέτρησης χρησιμοποιούνται τα κριτήρια: Kendall's W, Spearman rho (ρ), Kolmogorov - Smirnov, Wilcoxon κ.λπ.
γ) Για την ισοδιαστημική - αναλογική κλίμακα μέτρησης χρησιμοποιούνται τα κριτήρια: μέση τιμή, διάμεσος, διασπορά, τυπική απόκλιση, Pearson r, κριτήριο t, ANOVA κ.λπ.

Τονίζουμε με έμφαση τα εξής: Τα δεδομένα που θα προκύψουν σε μια έρευνα και το είδος της επεξεργασίας που θα γίνει σε αυτά, πρέπει να απασχολεί τον ερευνητή σε όλα τα στάδια της ερευνητικής διαδικασίας και όχι στο τέλος της συγκέντρωσης των δεδομένων. Αυτό γίνεται γιατί ο τρόπος που θα διατυπώσει τους σκοπούς ή τις υποθέσεις της έρευνας, το είδος των δεδομένων που θα συγκεντρώσει, το μέσο συλλογής των δεδομένων που θα χρησιμοποιήσει κ.λπ. σχετίζονται άμεσα με την ανάλυση και την επεξεργασία των δεδομένων.

Υπενθυμίζουμε ότι τα δεδομένα μιας έρευνας μπορεί να είναι μόνο ποιοτικά ή μόνο ποσοτικά ή και τα δύο. Αν δεν έχουμε καθόλου αριθμητικά δεδομένα, είναι δυνατό να μην χρειαστεί η βοήθεια της Στατιστικής. Τα ποσοτικά δεδομένα απαιτούν κάποιας μορφής στατιστική επεξεργασία.

7.4 Ανάλυση δεδομένων με μεθόδους της Περιγραφικής Στατιστικής

Μετά τη συγκέντρωση των δεδομένων, την καταγραφή τους στα ειδικά έντυπα καταγραφής δεδομένων ερευνών και την εισαγωγή τους στον υπολογιστή, το πρώτο βήμα που πρέπει να κάνει ο ερευνητής στην ανάλυση δεδομένων είναι να περιγράψει τα δεδομένα συνοπτικά, χρησιμοποιώντας μεθόδους της Περιγραφικής Στατιστικής, οι οποίες θα του επιτρέψουν να εξάγει κάποια συμπεράσματα για το δείγμα των ατόμων που μελετά.

Υπενθυμίζουμε εδώ τον εξής ορισμό: «Στατιστική είναι η εφαρμοσμένη επιστήμη που ασχολείται με τις επιστημονικές μεθόδους σχεδιασμού μιας μελέτης, συλλογής, επεξεργασίας, παρουσίασης και ανάλυσης αριθμητικών δεδομένων, με σκοπό την εξαγωγή συμπερασμάτων για τη λήψη ορθών αποφάσεων.»

Με απλά λόγια η Στατιστική παρέχει διαδικασίες που επιτρέπουν επεξεργασίες:

α) για περιγραφή και παρουσίαση κάποιων δεδομένων (Περιγραφική Στατιστική) και

β) για συνθετότερους ελέγχους που οδηγούν σε εξαγωγή συμπερασμάτων για ολόκληρο τον πληθυσμό από τη μελέτη δειγμάτων του πληθυσμού αυτού (Επαγωγική Στατιστική).

Με τη χρήση της Περιγραφικής Στατιστικής ο ερευνητής έχει τις εξής δυνατότητες:

1. Να παρουσιάζει ποσοτικά δεδομένα με πίνακες συχνοτήτων, σχετικών συχνοτήτων, αθροιστικών συχνοτήτων και σχετικών αθροιστικών συχνοτήτων.
2. Να παρουσιάζει τα δεδομένα με γραφικές παραστάσεις και άλλα διαγράμματα.
3. Να υπολογίζει τα μέτρα θέσης και μέτρα διασποράς μιας κατανομής.
4. Να ελέγχει τη συσχέτιση μεταβλητών.
5. Να μεταφέρει δεδομένα από μια κλίμακα μέτρησης σε άλλη.

Τις δυο τελευταίες διαδικασίες τις αναπτύξαμε με συντομία στις παραγράφους 2.6 και 5.4, όπου μας ήταν απαραίτητες. Οι τρεις πρώτες έχουν διδαχθεί στο Γυμνάσιο και στο Λύκειο και πιστεύουμε ότι είναι γνωστές σε όλους τους ερευνητές. Για την πληρότητα του κειμένου θα τις επαναλάβουμε κι εδώ με συντομία.

- *Παρουσίαση πληροφοριών*

Μετά από τη συγκέντρωση και την ταξινόμηση των στοιχείων ακολουθεί η παρουσίασή τους μ' έναν τέτοιο τρόπο, ώστε να είναι εύκολο στον καθένα να τα κατανοήσει. Η παρουσίαση των πληροφοριών είναι σημαντική διαδικασία και γίνεται κυρίως με **πίνακες** και **διαγράμματα**.

Παλαιότερα η παρουσίαση γινόταν με αναλυτικές εκθέσεις, στις οποίες περιγράφονταν όλες οι λεπτομέρειες. Ο τρόπος αυτός έχει εγκαταλειφθεί, γιατί δεν παρουσιάζει τις πληροφορίες με συντομία ώστε να διευκολύνεται η χρήση τους.

- *Πίνακες πληροφοριών*

Τους πίνακες που συντάσσει η Στατιστική Υπηρεσία και χρησιμοποιούνται ως πηγές πληροφοριών τους λέμε **γενικούς**, ενώ τους πίνακες που συντάσσουμε για να παρουσιάσουμε μερικές μόνο πληροφορίες για έναν πληθυσμό τους λέμε **συνοπτικούς**. Όταν οι πίνακες μας δίνουν πληροφορίες για έναν πληθυσμό που τα στοιχεία του εξετάζονται ως προς ένα και μόνο χαρακτηριστικό, λέγονται **απλής εισόδου** (πίνακας 1), ενώ όταν τα στοιχεία του πληθυσμού εξετάζονται ως προς δύο χαρακτηριστικά, λέγονται **διπλής εισόδου** (πίνακας 2).

Πίνακας 1

Αριθμός παιδιών	Οικογένειες
0	2
1	1
2	8
3	1
Σύνολο	12

Εξέταση των 12 οικογενειών μιας πολυκατοικίας ως προς τον αριθμό των παιδιών τους.

Πίνακας 2

Φύλο	Χρώμα ματιών				Σύνολο
	Μαύρα	Καστανά	Πράσινα	Μπλε	
Αγόρια	12	15	9	4	40
Κορίτσια	18	20	15	7	60
Σύνολο	**30**	**35**	**24**	**11**	**100**

Κατανομή 100 μαθητών ως προς το φύλο και το χρώμα των ματιών τους.

Χρησιμοποιώντας τους παραπάνω πίνακες μπορούμε να δώσουμε πληροφορίες πολύ πιο συνοπτικά και περιεκτικά απ' ό,τι αν χρησιμοποιούσαμε συνεχή γραπτό λόγο.

Όταν συντάσσουμε έναν πίνακα, πρέπει να τηρούμε ορισμένους κανόνες, ώστε να μπορεί ο οποιοσδήποτε να καταλάβει τι περιγράφει ο πίνακας. Μερικοί από αυτούς τους κανόνες είναι:

1. Ο τίτλος του πίνακα γράφεται πάνω ή κάτω από αυτόν και φανερώνει περιληπτικά το αντικείμενο στο οποίο αναφέρεται.
2. Στο κάτω μέρος συνήθως γράφουμε την πηγή απ' όπου πήραμε τα στοιχεία.
3. Σημειώνουμε στην κορυφή κάθε στήλης και στην αρχή κάθε γραμμής το είδος των πληροφοριών.

- *Πίνακες συχνοτήτων*

Οι ετήσιοι βαθμοί στα Μαθηματικά 30 μαθητών της Β΄ Λυκείου είναι: 18, 17, 14, 8, 10, 11, 9, 11, 13, 14, 16, 17, 19, 20, 18, 17, 15, 16, 16, 17, 14, 13, 15, 16, 15, 17, 16, 12, 18, 15. Ο τρόπος της παρουσίασης είναι όπως είναι γραμμένοι οι μαθητές, με αλφαβητική σειρά, στον κατάλογο του καθηγητή τους.

Εδώ δεν μας ενδιαφέρει ποιος μαθητής πήρε τον α ή τον β βαθμό, αλλά θέλουμε να μελετήσουμε την επίδοση των μαθητών στα Μαθηματικά και να βγάλουμε διάφορα συμπεράσματα. Η μελέτη μας γίνεται πιο εύκολη αν παρουσιάσουμε τα στοιχεία (τους βαθμούς) από το μικρότερο προς το μεγαλύτερο. Έτσι έχουμε: 8, 9, 10, 11, 11, 12, 13, 13, 14, 14, 14, 15, 15, 15, 15, 16, 16, 16, 16, 16, 17, 17, 17, 17, 17, 18, 18, 18, 19, 20.

Μια τέτοια παρουσίαση στατιστικών στοιχείων λέγεται *διάταξη των στοιχείων*. Όταν όμως το πλήθος των στοιχείων είναι μεγάλο, η διάταξή τους είναι επίπονη και γι' αυτό παραλείπεται. Τότε εφαρμόζουμε τη μέθοδο της διαλογής, την οποία εκθέτουμε παρακάτω.

Εύλογο είναι να ρωτήσουμε: Πόσοι μαθητές πήραν 8, πόσοι πήραν 10, πόσοι πήραν 15, πόσοι πήραν 17; Όλες αυτές οι ερωτήσεις μπορούν να απαντηθούν εύκολα όταν τα προηγούμενα στοιχεία τα παρουσιάσουμε με τον επόμενο πίνακα.

Πίνακας 3

Τιμές μεταβλητής (βαθμός) x_i	Διαλογή	Αριθμός παρατηρήσεων (συχνότητα) ν_i
8	|	1
9	|	1
10	|	1
11	||	2
12	|	1
13	||	2
14	|||	3
15	||||	4
16	~~||||~~	5
17	~~||||~~	5
18	|||	3
19	|	1
20	|	1
Σύνολο		**30**

Έτσι βλέπουμε αμέσως ότι 15 πήραν 4 μαθητές, 17 πήραν 5 μαθητές κτλ. Ο αριθμός 4 λέγεται **συχνότητα** της τιμής 15, ο αριθμός 5 συχνότητα της τιμής 17 κτλ.

Μπορούμε τώρα να διατυπώσουμε τον ορισμό:

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

Είναι φανερό ότι το άθροισμα όλων των συχνοτήτων είναι ίσο με το μέγεθος ν του δείγματος, δηλαδή

$$\nu_1 + \nu_2 + \ldots + \nu_\kappa = \sum_{i=1}^{\kappa} \nu_i = \nu$$

Ο πίνακας 3 λέγεται πίνακας **συχνοτήτων** ή **κατανομή συχνοτήτων**. Είναι πίνακας απλής εισόδου, γιατί τα στατιστικά στοιχεία (παρατηρήσεις) αναφέρονται σε μία μόνο ιδιότητα των στοιχείων του πληθυσμού. Η μεταβλητή X (βαθμός) είναι ασυνεχής και το πλήθος των διαφορετικών τιμών της είναι μικρό.

Πολλές φορές ακούμε και την ερώτηση: Τι ποσοστό % των μαθητών πήρε 8; Για να απαντήσουμε στην ερώτηση αυτή, πρέπει να διαιρέσουμε τη συχνότητα της τιμής 8, που είναι 1, με το πλήθος των παρατηρήσεων, δηλαδή $\frac{1}{30} \simeq 0{,}03333$. Ο αριθμός αυτός λέγεται **σχετική συχνότητα** της τιμής 8. Έτσι διατυπώνουμε τον ακόλουθο ορισμό:

***Σχετική συχνότητα** (f_i) μιας τιμής x_i της μεταβλητής X λέγεται το πηλίκο της συχνότητάς της ν_i προς το πλήθος ν των παρατηρήσεων, δηλαδή*

$$f_i = \frac{\nu_i}{\nu} = \frac{\nu_i}{\nu_1 + \nu_2 + \ldots + \nu_\kappa} = \frac{\nu_i}{\sum\limits_{\rho=1}^{\kappa} \nu_\rho}, \quad i = 1, 2, \ldots, \kappa$$

Το άθροισμα των σχετικών συχνοτήτων είναι ίσο με τη μονάδα. Πράγματι:

$$\sum_{i=1}^{\kappa} f_i = f_1 + f_2 + \ldots + f_\kappa = \frac{\nu_1}{\nu} + \frac{\nu_2}{\nu} + \ldots + \frac{\nu_\kappa}{\nu} = \frac{\nu_1 + \nu_2 + \ldots + \nu_\kappa}{\nu} = \frac{\nu}{\nu} = 1$$

Η σχετική συχνότητα εκφράζεται συνήθως σε ποσοστό επί τοις εκατό. Έτσι η σχετική συχνότητα της τιμής 8 είναι $\frac{1}{30} = 0{,}03333\ldots$ ή 3,33%, ενώ της τιμής 18 είναι $\frac{3}{30} = 0{,}1$ ή 10%. Την κατανομή σχετικών συχνοτήτων μπορούμε να την παρουσιάσουμε συμπληρώνοντας τον πίνακα 3 όπως φαίνεται στον πίνακα 4 που ακολουθεί.

Πίνακας 4

Βαθμός (x_i)	Συχνότητα (ν_i)	Σχετική συχνότητα (f_i%)
8	1	3,33
9	1	3,33
10	1	3,33
11	2	6,67
12	1	3,33
13	2	6,67
14	3	10,00
15	4	13,34
16	5	16,67
17	5	16,67
18	3	10,00
19	1	3,33
20	1	3,33
Σύνολο	**30**	**100,00**

Εύλογο είναι να μας ρωτήσει κάποιος: Πόσοι μαθητές πήραν βαθμό κάτω από τη βάση, δηλαδή κάτω από 10; Πόσοι μαθητές πήραν το πολύ 14; Πόσοι μαθητές αρίστευσαν;

Για να απαντήσουμε στην πρώτη ερώτηση, πρέπει να προσθέσουμε τις συχνότητες των τιμών 8 και 9, δηλαδή $1 + 1 = 2$. Ίδια θα είναι η απάντηση κι αν αλλάξουμε την ερώτηση ως εξής: Πόσοι μαθητές πήραν το πολύ 9;

Για να απαντήσουμε στη δεύτερη ερώτηση, πρέπει να προσθέσουμε τις συχνότητες των τιμών 8, 9, 10, 11, 12, 13 και 14, δηλαδή $1 + 1 + 1 + 2 + 1 + 2 + 3 = 11$.

Ο αριθμός 11, που φανερώνει πόσοι μαθητές πήραν το πολύ 14, λέγεται **αθροιστική συχνότητα** της τιμής 14. Συνεπώς:

***Αθροιστική συχνότητα** μιας τιμής x_i λέγεται το άθροισμα των συχνοτήτων ν_i των τιμών που είναι μικρότερες από την τιμή αυτή ή ίσες, δηλαδή*

$$N_i = \nu_1 + \nu_2 + \nu_3 + \ldots + \nu_i$$

Είναι ακόμη πολύ πιθανό να ακούσουμε τις ακόλουθες ερωτήσεις: Τι ποσοστό μαθητών πήρε βαθμό κάτω από τη βάση; Τι ποσοστό μαθητών πήρε το πολύ 14; Είναι φανερό ότι η απάντηση στο πρώτο ερώτημα είναι $3{,}33 + 3{,}33 = 6{,}66\%$, ενώ στο δεύτερο είναι

$$3{,}33 + 3{,}33 + 3{,}33 + 6{,}67 + 3{,}33 + 6{,}67 + 10{,}00 = 36{,}66\%$$

Ο αριθμός 6,66% λέγεται **σχετική αθροιστική συχνότητα** της τιμής $x_2 = 9$. Συνεπώς:

Σχετική αθροιστική συχνότητα *μιας τιμής x_i λέγεται το άθροισμα των σχετικών συχνοτήτων $f_i\%$ των τιμών που είναι μικρότερες από αυτή ή ίσες, δηλαδή*

$$F_i = f_1 + f_2 + \ldots + f_i$$

Είναι φανερό ότι ισχύουν οι σχέσεις:

$$\nu_1 = N_1,\ \nu_2 = N_2 - N_1,\ \ldots,\ \nu_\kappa = N_\kappa - N_{\kappa-1}$$

$$f_1 = F_1,\ f_2 = F_2 - F_1,\ \ldots,\ f_\kappa = F_\kappa - F_{\kappa-1}$$

Ο πίνακας 5 που ακολουθεί σχηματίστηκε με συμπλήρωση του πίνακα 4 με τις στήλες της αθροιστικής και της σχετικής αθροιστικής συχνότητας.

Πίνακας 5

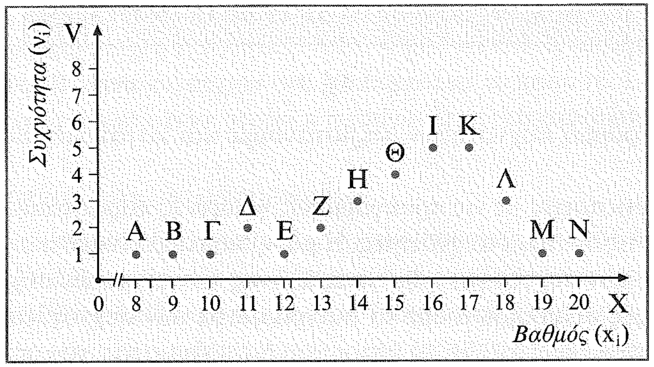
Βαθμός (x_i)	**Συχνότητα (ν_i)**	**Αθροιστική συχνότητα (N_i)**	**Σχετική συχνότητα ($f_i\%$)**	**Σχετική αθροιστική συχνότητα ($F_i\%$)**
8	1	1	3,33	3,33
9	1	2	3,33	6,66
10	1	3	3,33	9,99
11	2	5	6,67	16,66
12	1	6	3,33	19,99
13	2	8	6,67	26,66
14	3	11	10,00	36,66
15	4	15	13,34	50,00
16	5	20	16,67	66,67
17	5	25	16,67	83,34
18	3	28	10,00	93,34
19	1	29	3,33	96,67
20	1	30	3,33	100,00
Σύνολο			**100,00**	

- *Γραφική παράσταση κατανομής συχνοτήτων*

Θα δούμε τώρα πώς μπορούμε να παρουσιάσουμε γραφικά την κατανομή συχνοτήτων της βαθμολογίας των 30 μαθητών, την οποία παρουσιάσαμε προηγουμένως με πίνακες.

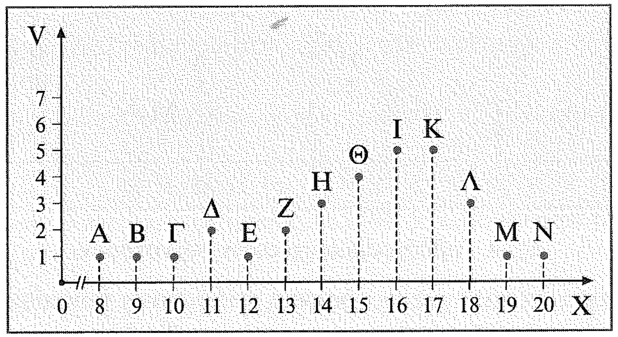
Είναι φανερό ότι εδώ έχουμε μια συνάρτηση με μια ανεξάρτητη μεταβλητή X, που παίρνει τιμές $x_1, x_2, \ldots, x_\kappa$, και με μια εξαρτημένη μεταβλητή V, που παίρνει τιμές $\nu_1, \nu_2, \ldots, \nu_\kappa$ αντίστοιχα. Μπορούμε λοιπόν να θεωρήσουμε ένα ορθογώνιο σύστημα αξόνων και στον οριζόντιο άξονα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

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



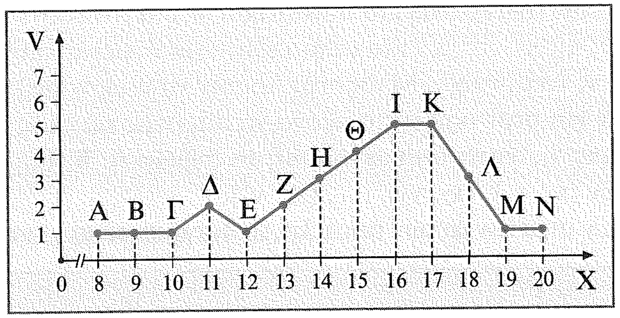
Σχ. 1. Διάγραμμα συχνοτήτων που αντιστοιχεί στον πίνακα 3.

Συνήθως ως διάγραμμα συχνοτήτων παίρνουμε όχι τα σημεία Α, Β, Γ, ..., Ν, αλλά τα ευθύγραμμα τμήματα που παριστάνουν τις τεταγμένες αυτών των σημείων. Τότε το διάγραμμα λέγεται **ραβδόγραμμα συχνοτήτων** (σχ. 2).



Σχ. 2. Ραβδόγραμμα συχνοτήτων.

Πολλές φορές ενώνουμε τα σημεία Α, Β, Γ, . . . , Ν και παίρνουμε μια τεθλασμένη γραμμή (σχ. 3). Κάτι τέτοιο όμως πρέπει να αποφεύγεται, γιατί τα σημεία αυτά είναι μεμονωμένα και επομένως κατασκευάζοντας την πολυγωνική γραμμή δίνουμε την εσφαλμένη εντύπωση ότι η γραμμή αυτή παριστάνει την κατανομή.



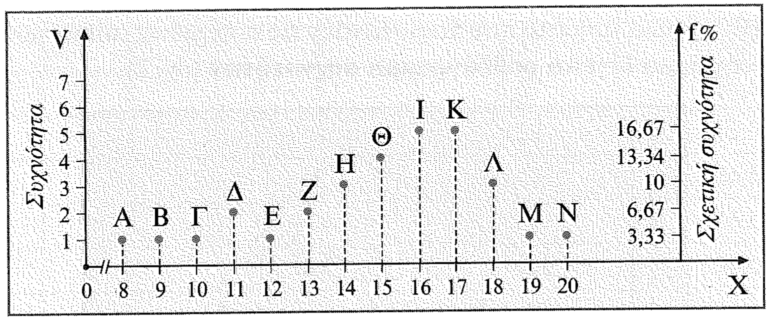
Σχ. 3

Με ραβδόγραμμα συχνοτήτων παριστάνουμε και κατανομές ποιοτικών μεταβλητών.

Ακριβώς αντίστοιχες με τις προηγούμενες γραφικές παραστάσεις μπορούμε να έχουμε και για τις σχετικές συχνότητες. Η μόνη διαφορά είναι ότι ο y - άξονας, που είναι ο άξονας των συχνοτήτων, θα γίνει άξονας των σχετικών συχνοτήτων.

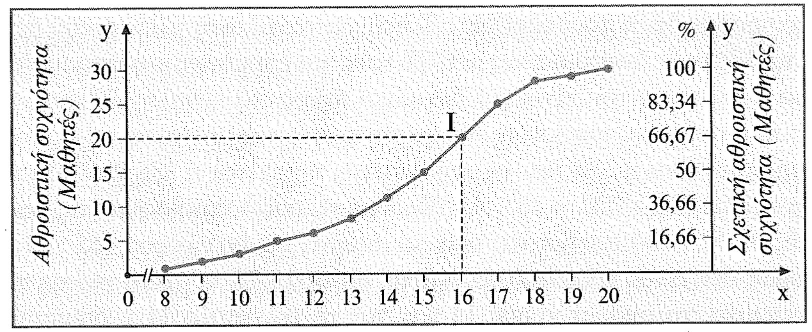
Πολλές φορές συμπληρώνουμε τα διαγράμματα των συχνοτήτων μ' έναν δεύτερο άξονα τεταγμένων στο δεξιό μέρος του διαγράμματος, τον οποίο παίρνουμε ως άξονα των σχετικών συχνοτήτων (σχ. 4).

Παρόμοια κατασκευάζουμε το ραβδόγραμμα σχετικών συχνοτήτων.



Σχ. 4. Ραβδόγραμμα σχετικών συχνοτήτων.

Μπορούμε τώρα να κάνουμε τη γραφική παράσταση των αθροιστικών συχνοτήτων και των σχετικών αθροιστικών συχνοτήτων. Τα δύο αυτά διαγράμματα τα δίνουμε στο σχήμα 5, παίρνοντας ως y - άξονα αριστερά την αθροιστική συχνότητα και δεξιά τη σχετική αθροιστική συχνότητα.



Σχ. 5. Διάγραμμα αθροιστικών συχνοτήτων και σχετικών αθροιστικών συχνοτήτων.

Στο σχήμα 5 βλέπουμε ότι στο σημείο Ι απεικονίζεται το ζεύγος (16, 20) για το διάγραμμα της αθροιστικής συχνότητας και το ζεύγος (16, 66,67) για το διάγραμμα της σχετικής αθροιστικής συχνότητας. Αυτό σημαίνει ότι 20 μαθητές ή το 66,67% των μαθητών της Β' Λυκείου πήραν στα Μαθηματικά το πολύ 16. Για την πολυγωνική γραμμή που σχηματίστηκε ισχύει η προηγούμενη παρατήρηση.

- *Ομαδοποίηση των παρατηρήσεων*

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

Έτσι, αν έχουμε 200 μαθητές και τους μελετάμε ως προς το ύψος τους, τότε δεν είναι εύκολο να φτιάξουμε πίνακα γράφοντας όλες τις τιμές της μεταβλητής «ύψος», γιατί αυτό μπορεί να πάρει κάθε τιμή, π.χ. από 140 cm μέχρι 190 cm. Γι' αυτό χωρίζουμε το διάστημα [140, 190) σε υποδιαστήματα (κλάσεις), π.χ. [140, 150), [150, 160) κτλ., και υπολογίζουμε τον αριθμό των μαθητών που ανήκουν σε κάθε κλάση.

Η όλη εργασία φαίνεται καλύτερα στο ακόλουθο παράδειγμα:

Παράδειγμα

Η κατανάλωση ηλεκτρικού ρεύματος σε kWh 50 οικογενειών το δεύτερο δίμηνο του 2001 ήταν η εξής:

308, **220**, 232, 290, 350, 375, 340, 240, 245, 290, 320, 346, 370, 235, 260, 272, 296, **385**, 349, 273, 238, 242, 256, 267, 278, 282, 297, 306, 343, 317, 325, 334, 341, 357, 378, 363, 382, 250, 270, 227, 280, 283, 294, 298, 300, 302, 241, 266, 274, 275

Παρατηρούμε ότι η ποσοτική μεταβλητή «κατανάλωση ηλεκτρικού ρεύματος» παίρνει πολλές τιμές διαφορετικές μεταξύ τους και επομένως μια παρουσίασή τους σε πίνακα αντίστοιχο του πίνακα 3 δεν είναι πρακτικά εύκολη. Γι' αυτό κάνουμε ομαδοποίηση των παρατηρήσεων. Ψάχνοντας στις παραπάνω τιμές βρίσκουμε τη μικρότερη, που είναι η 220, και τη μεγαλύτερη, που είναι η 385, δηλαδή **το εύρος των τιμών** είναι 385 – 220 = 165. Αν θέλουμε να ομαδοποιήσουμε τις παρατηρήσεις αυτές σε 12 υποδιαστήματα, τότε το καθένα θα έχει εύρος 165 : 12 = 13,75. Επειδή όμως δεν είναι πρακτικό να έχουμε εύρος 13,75, γι' αυτό για τα πρώτα 11 υποδιαστήματα παίρνουμε εύρος 14 και για το τελευταίο 11 ή διευρύνουμε το τελευταίο υποδιάστημα κατά 3 μονάδες, δηλαδή το ανώτερο όριο το παίρνουμε 388, οπότε όλα τα υποδιαστήματα έχουν εύρος 14.

Το μέσο μιας κλάσης λέγεται **κεντρική τιμή** της κλάσης.

Έτσι κατασκευάζουμε τον επόμενο πίνακα.

Πίνακας 6

Η κατανάλωση ηλεκτρικού ρεύματος σε kWh 50 οικογενειών.

kWh Κλάσεις	Διαλογή	Κεντρικές τιμές κλάσεων (x_i)	Συχνότητα (ν_i)	Αθροιστική συχνότητα (N_i)	Σχετική συχνότητα (f_i%)	Σχετική αθροιστική συχνότητα (F_i%)
[220, 234)	|||	227	3	3	6	6
[234, 248)	~~||||~~ |	241	6	9	12	18
[248, 262)	|||	255	3	12	6	24
[262, 276)	~~||||~~ ||	269	7	19	14	38
[276, 290)	||||	283	4	23	8	46
[290, 304)	~~||||~~ |||	297	8	31	16	62
[304, 318)	|||	311	3	34	6	68
[318, 332)	||	325	2	36	4	72
[332, 346)	||||	339	4	40	8	80
[346, 360)	||||	353	4	44	8	88
[360, 374)	||	367	2	46	4	92
[374, 388)	||||	381	4	50	8	100
Σύνολο			**50**		**100**	

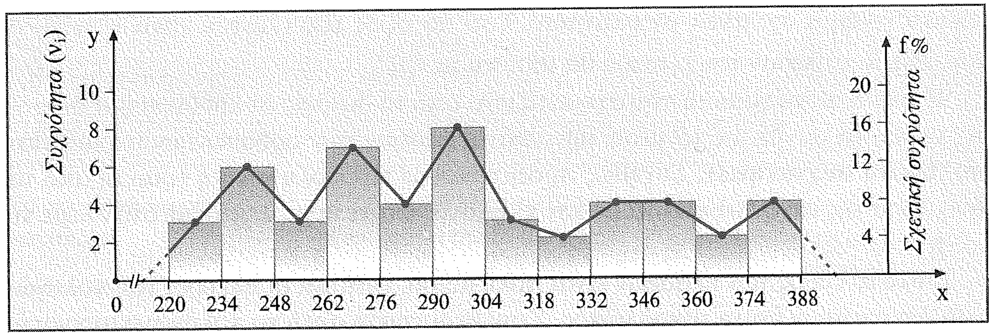
Η ομαδοποίηση αυτή μας βοηθά να κάνουμε διάφορες παρατηρήσεις και να βγάλουμε εύκολα διάφορα συμπεράσματα.

Έτσι, εύκολα βλέπουμε τώρα πόσες οικογένειες κατανάλωσαν αυτό το δίμηνο λιγότερα από 290 kWh (23 οικογένειες) και τι ποσοστό οικογενειών (%) κατανάλωσαν λιγότερα από 360 kWh (88%).

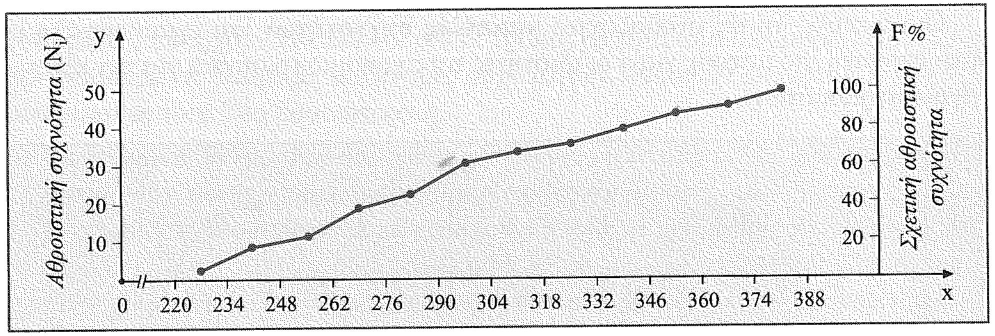
Τονίζουμε εδώ ότι οι συχνότητες αναφέρονται στην κλάση και όχι στις μεμονωμένες τιμές της μεταβλητής. Για παράδειγμα, το ότι η κλάση [262, 276) έχει συχνότητα 7 σημαίνει ότι 7 τιμές της μεταβλητής βρίσκονται σε αυτή την κλάση. Παρακάτω, για να διευκολύνουμε τους υπολογισμούς, θα λέμε ότι η κεντρική τιμή της κλάσης, δηλαδή το 269, έχει τη συχνότητα 7 της κλάσης.

Αυτό βέβαια είναι ένα μειονέκτημα της ομαδοποίησης, αλλά εξυπηρετεί τη μελέτη.

Με βάση τον πίνακα 6 κατασκευάζουμε τις γραφικές παραστάσεις συχνοτήτων και σχετικών συχνοτήτων, αθροιστικών συχνοτήτων και σχετικών αθροιστικών συχνοτήτων των ομαδοποιημένων παρατηρήσεων (σχ. 6 και 7).



Σχ. 6. Διάγραμμα συχνοτήτων και σχετικών συχνοτήτων ομαδοποιημένων παρατηρήσεων.



Σχ. 7. Εδώ δίνουμε το διάγραμμα των αθροιστικών και των σχετικών αθροιστικών συχνοτήτων του προηγούμενου παραδείγματος.

Το σχήμα 6, που αποτελείται από συνεχόμενα ορθογώνια, λέγεται **ιστόγραμμα**. Κάθε ορθογώνιο έχει βάση ίση με το πλάτος της κλάσης της μεταβλητής και εμβαδόν ίσο με την αντίστοιχη συχνότητα της κλάσης όταν το πλάτος της κλάσης ληφθεί ως μονάδα μέτρησης.

Γενικά, για να κατασκευάσουμε τους ιστούς σ' ένα ιστόγραμμα συχνοτήτων, προσέχουμε κάθε ορθογώνιο να έχει βάση ίση με το πλάτος της κλάσης και ύψος τέτοιο, ώστε **το εμβαδόν του ορθογωνίου να ισούται με τη συχνότητα της κλάσης αυτής**, δηλαδή $E = v_i = c_i \cdot \upsilon_i$, οπότε $\upsilon_i = \frac{v_i}{c_i}$, όπου y_i το ύψος του i - ιστού, που έχει συχνότητα n_i και εύρος κλάσης c_i. Τότε το άθροισμα των εμβαδών όλων των ιστών (ή του αντίστοιχου πολυγώνου συχνοτήτων) θα ισούται με το μέγεθος ν του δείγματος (ή του πληθυσμού).

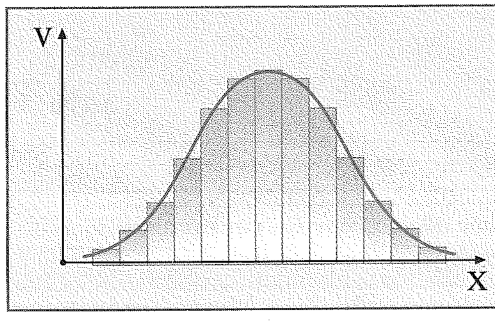
Αν, όπως είπαμε, οι κλάσεις είναι ίσου πλάτους και το πλάτος της κλάσης ληφθεί ως μονάδα μέτρησης, τότε τα ύψη των ιστών θα είναι ίσα με τη συχνότητά τους, δηλαδή με το εμβαδόν του ιστού.

Τα ίδια ακριβώς ισχύουν και όταν αντί του ιστογράμματος συχνοτήτων έχουμε το ιστόγραμμα σχετικών συχνοτήτων. Τότε το ύψος του i - ιστού είναι $\upsilon_i^* = \frac{f_i\%}{c_i}$, δηλαδή το εμβαδόν του i - ιστού θα ισούται με $f_i\%$.

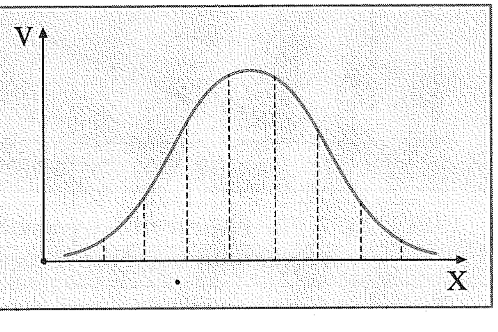
Η τεθλασμένη γραμμή που σχηματίζεται από τα διαδοχικά ευθύγραμμα τμήματα τα οποία συνδέουν τα μέσα των επάνω βάσεων των ορθογωνίων ονομάζεται **πολύγωνο συχνοτήτων**. Συνήθως προεκτείνουμε την πολυγωνική γραμμή από τα δύο άκρα της μέχρι να κλείσει με τον x - άξονα, οπότε κυριολεκτούμε όταν λέμε τη λέξη πολύγωνο.

Όταν το εύρος των κλάσεων είναι πολύ μικρό, τότε και οι πλευρές του πολυγώνου συχνοτήτων έχουν μικρό μήκος. Στην περίπτωση αυτή το πολύγωνο συχνοτήτων μοιάζει με καμπύλη γραμμή (σχ. 6α). Είναι φανερό ότι τα πολύγωνα συχνοτήτων των συνεχών μεταβλητών είναι καμπύλες (σχ. 6β).

Οι καμπύλες αυτές ονομάζονται **καμπύλες συχνοτήτων** και έχουν μεγάλη εφαρμογή στη Στατιστική, γιατί οι ιδιότητές τους χρησιμοποιούνται για την εξαγωγή συμπερασμάτων.

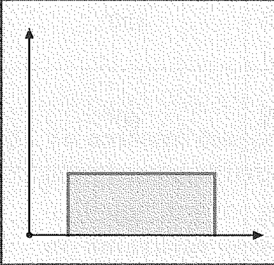


Σχ. 6α. Πολύγωνο συχνοτήτων όταν έχουμε μεγάλο αριθμό κλάσεων με μικρό εύρος.

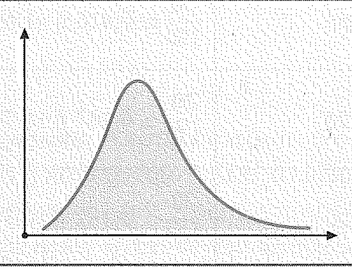


Σχ. 6β. Κατανομή συχνοτήτων συνεχούς μεταβλητής.

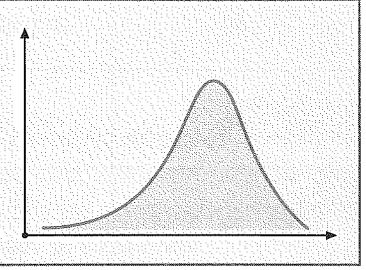
Το σχήμα μιας καμπύλης συχνοτήτων μας δείχνει πώς είναι κατανεμημένες οι παρατηρήσεις που έχουμε συγκεντρώσει για ένα φαινόμενο. Στις καμπύλες αυτές, ανάλογα με το σχήμα τους, δίνονται και συγκεκριμένα ονόματα για να μπορούν όλοι όσοι ασχολούνται με τη Στατιστική να έχουν κοινή γλώσσα επικοινωνίας. Για παράδειγμα, οι καμπύλες των σχημάτων 6α και 6β έχουν «κωνοειδές» σχήμα. Η κατανομή συχνοτήτων που έχει τέτοιο σχήμα λέγεται **κανονική κατανομή** (normal distribution) και παίζει τον σπουδαιότερο, από όλες τις άλλες κατανομές, ρόλο στη Στατιστική. Άλλες κατανομές που συναντάμε είναι αυτές που φαίνονται στα επόμενα σχήματα.



Σχ. 6γ. Ομοιόμορφη κατανομή.



Σχ. 6δ. Λοξή προς τα δεξιά (skew symmetric to the right) θετική ασυμμετρία.



Σχ. 6ε. Λοξή προς τ' αριστερά (skew symmetric to the left) αρνητική ασυμμετρία.

• *Άλλες γραφικές παραστάσεις*

Εκτός από τις προηγούμενες γραφικές παραστάσεις (ραβδογράμματα και ιστογράμματα), υπάρχουν και πολλοί άλλοι τρόποι με τους οποίους μπορούμε να παρουσιάσουμε τις διάφορες κατανομές συχνοτήτων, σχετικών συχνοτήτων κτλ. Θα αναφέρουμε τους δύο βασικότερους.

α) ***Τα κυκλικά διαγράμματα.*** Αυτά χρησιμοποιούνται κυρίως για τη γραφική απεικόνιση των σχετικών συχνοτήτων ποιοτικών μεταβλητών. Τις σχετικές συχνότητες τις παριστάνουμε με κυκλικούς τομείς ενός ολόκληρου κύκλου ή ημικυκλίου. Τότε ολόκληρος ο κύκλος (ή το ημικύκλιο) παριστάνει τη σχετική συχνότητα 100%. Το σχήμα που προκύπτει με τον τρόπο αυτό λέγεται **κυκλικό (ή ημικυκλικό) διάγραμμα**.

Παράδειγμα

Ο πίνακας 7 που ακολουθεί παρουσιάζει, στις δύο πρώτες στήλες του, την κατανομή των 120 οικογενειών ενός χωριού σε πέντε κατηγορίες ανάλογα με τα εισοδήματά τους. Η 1η κατηγορία είναι οι πλουσιότερες οικογένειες και η 5η οι φτωχότερες.

Πίνακας 7

Κατηγορία (x_i)	Συχνότητα (v_i)	Σχετική συχνότητα %	Γωνία κυκλικού τομέα σε μοίρες
1η	12	$\frac{12}{120} = 0{,}1$ ή 10%	36°
2η	18	$\frac{18}{120} = 0{,}15$ ή 15%	54°
3η	50	$\frac{50}{120} \simeq 0{,}416$ ή 41,6%	150°
4η	20	$\frac{20}{120} \simeq 0{,}167$ ή 16,7%	60°
5η	20	$\frac{20}{120} \simeq 0{,}167$ ή 16,7%	60°
Σύνολο	**120**	**100**	**360°**

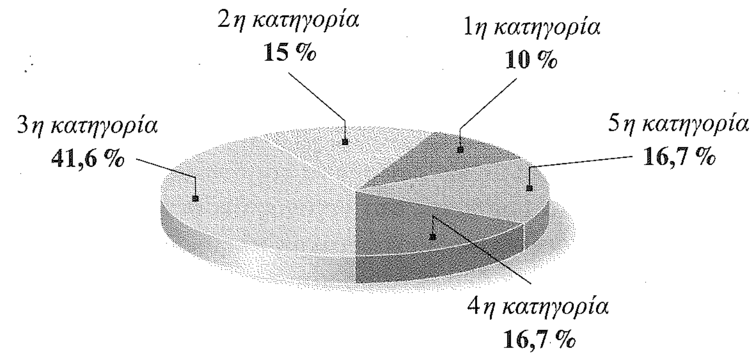
Για να απεικονίσουμε τις συχνότητες v_i αυτού του πίνακα με κυκλικό διάγραμμα, εργαζόμαστε ως εξής:

Μετατρέπουμε τον αριθμό των οικογενειών κάθε κατηγορίας σε ποσοστό επί του συνόλου, π.χ. η 1η είναι $\frac{12}{120} = 0{,}1$ ή 10% επί του συνόλου. Κατασκευάζουμε έναν κυκλικό δίσκο, ο οποίος θεωρείται σαν κυκλικός τομέας γωνίας 360°, και υποθέτουμε ότι το ολικό εμβαδόν του παριστάνει τον συνολικό αριθμό των 120 οικογενειών. Τότε το ποσοστό των οικογενειών κάθε κατηγορίας θα παριστάνεται με ανάλογο κυκλικό τομέα. Έτσι χωρίζουμε τις 360° σε μέρη ανάλογα των ποσοστών αυτών και βρίσκουμε τις γωνίες των κυκλικών τομέων. Επομένως οι γωνίες αυτές θα είναι:

$$\frac{12}{120} \cdot 360° = 36°, \quad \frac{18}{120} \cdot 360° = 54°, \quad \frac{50}{120} \cdot 360° = 150°, \quad \frac{20}{120} \cdot 360° = 60°$$

Έτσι συμπληρώνουμε τις επόμενες στήλες του πίνακα 7 και κατασκευάζουμε και το σχήμα 8.

Με ανάλογο τρόπο κατασκευάζονται και τα ημικυκλικά διαγράμματα.



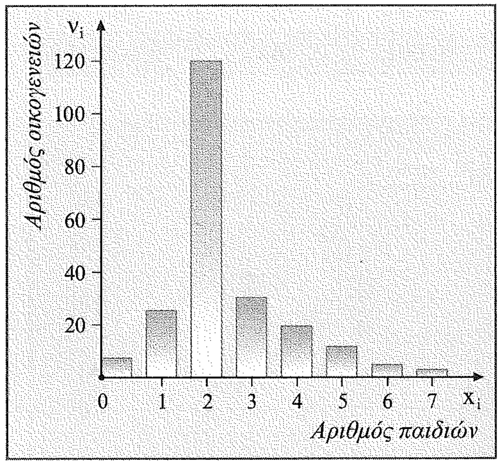
Σχ. 8

β) ***Τα ακιδωτά διαγράμματα.*** Αποτελούνται από ορθογώνια που είναι τοποθετημένα στον οριζόντιο ή κάθετο άξονα και το μήκος τους είναι ανάλογο με τους αριθμούς που αντιπροσωπεύουν. Χρησιμοποιούνται **για τη γραφική παράσταση ποιοτικών μεταβλητών ή ασυνεχών ποσοτικών μεταβλητών.**

Δίνουμε μια κατανομή στον πίνακα 8 και το ακιδωτό της διάγραμμα στο σχήμα 9.

Πίνακας 8

Αριθμός παιδιών (x_i)	Αριθμός οικογενειών (v_i)
0	8
1	25
2	120
3	30
4	20
5	12
6	5
7	4
Άθροισμα	**224**



Σχ. 9

• *Παράμετροι θέσης και παράμετροι διασποράς μιας κατανομής*

Στις προηγούμενες παραγράφους είδαμε ότι είναι δύσκολο να μελετήσουμε στατιστικά στοιχεία χωρίς να τα ταξινομήσουμε. Έτσι η πρώτη μας φροντίδα μετά τη συλλογή των στοιχείων είναι η ταξινόμησή τους και η παρουσίασή τους με πίνακες συχνοτήτων και με εποπτικές εικόνες. Οι πίνακες αυτοί περιορίζουν αρκετά τις δυσκολίες μελέτης, αλλά εξακολουθούν να παρουσιάζουν μερικές δυσκολίες.

Η μεγαλύτερη δυσκολία βρίσκεται στο γεγονός ότι δεν μπορούμε να συγκρίνουμε εύκολα ομοειδείς έρευνες σε διαφορετικούς πληθυσμούς.

Για παράδειγμα, αν έχουμε δύο πίνακες που παρουσιάζουν την επίδοση των μαθητών δύο τμημάτων της Β′ Λυκείου στα Μαθηματικά (πίνακες 8α και 8β), τότε δεν μπορούμε μόνο με τους πίνακες αυτούς να αποφανθούμε ποιο τμήμα έχει την καλύτερη επίδοση.

Δεν μπορούμε επίσης να συγκρίνουμε εύκολα ομοειδείς έρευνες σε διαφορετικούς πληθυσμούς ούτε με τα αντίστοιχα ιστογράμματά τους (σχ. 10 και 10α), γιατί, όπως βλέπουμε, ενώ η βαθμολογία και των δύο τμημάτων είναι συγκεντρωμένη γύρω στο 16, το τμήμα B_2 παρουσιάζει διαφορετική διασπορά από το B_1 (π.χ. στο B_2 δεν υπάρχουν καθόλου οι βαθμοί 9 και 10, ενώ έχει διπλάσιους αριστούχους, δηλαδή μαθητές με βαθμό 19 και 20).

Πίνακας 8α

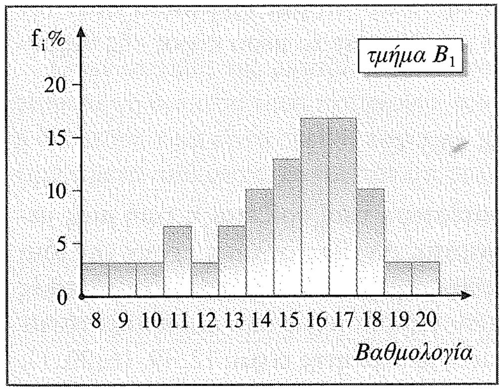
Βαθμολογία μαθητών του τμήματος B_1 στα Μαθηματικά. (Είναι ο πίνακας 4.)

Βαθμός (x_i)	**Συχνότητα (v_i)**	**Σχετική συχνότητα (f_i%)**
8	1	3,33
9	1	3,33
10	1	3,33
11	2	6,67
12	1	3,33
13	2	6,67*
14	3	10,00
15	4	13,34
16	5	16,67
17	5	16,67
18	3	10,00
19	1	3,33
20	1	3,33
Σύνολο	**30**	**100,00**

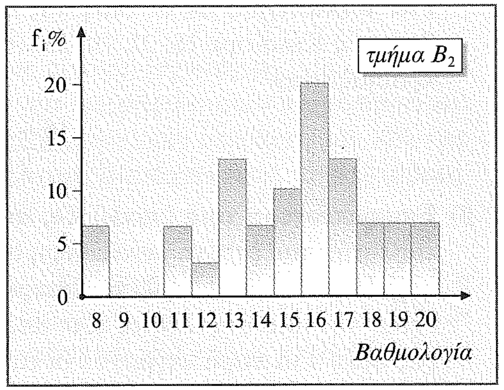
Πίνακας 8β

Βαθμολογία μαθητών του τμήματος B_2 στα Μαθηματικά.

Βαθμός (x_i)	**Συχνότητα (v_i)**	**Σχετική συχνότητα (f_i%)**
8	2	6,67
9	0	0,00
10	0	0,00
11	2	6,67
12	1	3,33
13	4	13,34
14	2	6,67*
15	3	10,00
16	6	20,00
17	4	13,34
18	2	6,66*
19	2	6,66
20	2	6,66
Σύνολο	**30**	**100,00**



Σχ. 10



Σχ. 10α

---

* Στη συχνότητα 2 αντιστοιχίζουμε σχετική συχνότητα 6,67 ή 6,66. Αυτό γίνεται για να έχουμε άθροισμα σχετικών συχνοτήτων 100.

Γι' αυτό λοιπόν χρειάζονται ορισμένα στοιχεία που θα μας βοηθήσουν να απαντούμε σε παρόμοια προβλήματα. Τα στοιχεία αυτά είναι αριθμοί που ονομάζονται **παράμετροι** ή ***γενικά χαρακτηριστικά*** ή **μέτρα** της κατανομής. Έχουμε τις **παραμέτρους θέσης** ή **μέτρα θέσης** (location measures), με τις οποίες προσδιορίζεται η θέση της κατανομής πάνω στον άξονα, και τις **παραμέτρους διασποράς** ή **μέτρα μεταβλητότητας** (measures of variability), με τις οποίες προσδιορίζεται η διασπορά της κατανομής.

Όταν λέμε ότι με την παράμετρο θέσης προσδιορίζεται η θέση της κατανομής πάνω στον άξονα, εννοούμε ότι βρίσκουμε έναν αριθμό πάνω στον άξονα, γύρω από τον οποίο τείνουν να συγκεντρωθούν οι τιμές της μεταβλητής. Ενώ με την παράμετρο διασποράς εννοούμε έναν αριθμό που δείχνει πώς «εξαπλώνονται» οι τιμές της μεταβλητής γύρω από την κεντρική τιμή.

Συνήθως τα μέτρα θέσης και διασποράς μιας κατανομής, όταν χρησιμοποιούνται για την περιγραφή ενός πληθυσμού, ονομάζονται **παράμετροι** (parameters), ενώ όταν χρησιμοποιούνται για την περιγραφή ενός δείγματος, ονομάζονται **στατιστικά** (statistics).

- *Παράμετροι θέσης*

Κυριότερες παράμετροι θέσης είναι οι παρακάτω:

α) **Μέσο εύρος** (mean range). Το μέσο εύρος ορίζεται ως εξής:

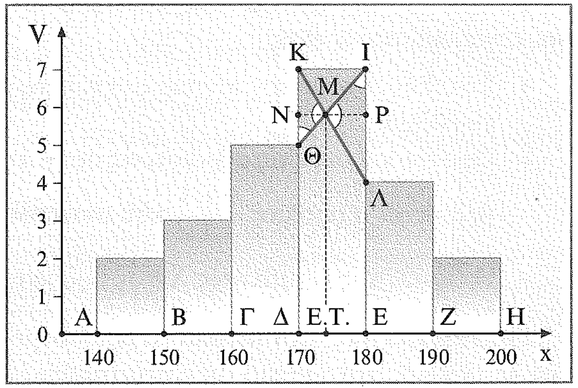
$$\text{Μέσο εύρος} = \frac{\text{Μικρότερη τιμή} + \text{Μεγαλύτερη τιμή}}{2}$$

Για παράδειγμα, αν οι τιμές της θερμοκρασίας για μια εβδομάδα ήταν (σε βαθμούς Κελσίου): 18, 22, 20, 23, 16, 18 και 21, τότε το μέσο εύρος της θερμοκρασίας ήταν

$$\frac{16+23}{2} = 19{,}5\,^{\circ}\mathrm{C}$$

β) **Επικρατέστερη τιμή** (mode). Ως επικρατέστερη τιμή ορίζεται η τιμή μιας μεταβλητής που έχει τη μεγαλύτερη συχνότητα. Αν όλες οι τιμές έχουν την ίδια συχνότητα, τότε δεν ορίζεται επικρατέστερη τιμή. Αν δύο ή περισσότερες τιμές έχουν την ίδια συχνότητα, τότε υπάρχουν περισσότερες από μία επικρατέστερες τιμές. Για παράδειγμα, αν οι τιμές μιας μεταβλητής είναι: 12, 19, 16, 20, 17, 22, 13 και 19, τότε η επικρατέστερη τιμή είναι η 19, ενώ αν οι τιμές είναι: 4, 7, 7, 5, 3, 8, 3 και 9, τότε υπάρχουν δύο επικρατέστερες τιμές, η 3 και η 7.

Όταν έχουμε ομαδοποιημένα δεδομένα σε κλάσεις ίσου πλάτους, στις οποίες θεωρούμε ότι οι παρατηρήσεις είναι ομοιόμορφα κατανεμημένες, τότε βρίσκουμε πρώτα την κλάση i που έχει τη μεγαλύτερη συχνότητα **(επικρατέστερη κλάση)** και στη συνέχεια υπολογίζουμε την επικρατέστερη τιμή (Ε.Τ.) όπως φαίνεται στο διάγραμμα κατανομής συχνοτήτων (σχ. 10β).



Σχ. 10β

γ) **Διάμεσος** (median). Αν οι τιμές μιας μεταβλητής έχουν διαταχθεί από τη μικρότερη προς τη μεγαλύτερη, η διάμεσος τιμή ορίζεται ως η μεσαία τιμή, αν το πλήθος των τιμών είναι περιττό και ως το ημιάθροισμα των δύο μεσαίων τιμών, αν το πλήθος είναι άρτιο. Για παράδειγμα:

1. Αν οι τιμές μιας μεταβλητής είναι: 18, 20, 23, 30, 32, 36 και 40, η διάμεσος είναι το 30. Οι τιμές εδώ δόθηκαν κατά τάξη μεγέθους από τη μικρότερη προς τη μεγαλύτερη. Αν αυτό δεν συμβαίνει, τότε διατάσσουμε πρώτα τις τιμές και μετά βρίσκουμε τη διάμεσο.
2. Αν οι τιμές μιας μεταβλητής είναι: 12, 6, 8, 14, 17, 19 και 8, τότε τις διατάσσουμε και έχουμε: 6, 8, 8, 9, 12, 14, 17 και 19. Βλέπουμε εδώ ότι έχουμε δύο μεσαίους όρους, το 9 και το 12. Η διάμεσος τώρα είναι το ημιάθροισμα $\frac{9+12}{2} = \frac{21}{2} = 10{,}5$.

Όταν οι τιμές της μεταβλητής δίνονται σε πίνακα συχνοτήτων, για να υπολογίσουμε τη διάμεσο τιμή, παίρνουμε τον πίνακα των αθροιστικών συχνοτήτων και η διάμεσος τιμή είναι εκείνη που έχει 50% των τιμών πριν από αυτή και 50% των τιμών μετά από αυτή. Εννοείται ότι οι τιμές έχουν διαταχθεί κατά τάξη μεγέθους.

Πίνακας 9

Αριθμός παιδιών (x_i)	**Οικογένειες (v_i)**	**Αθροιστική συχνότητα**
0	2	2
1	1	3
2	8	11
3	1	12
Σύνολο	**12**	

Στο παράδειγμα του πίνακα 9 η διάμεσος τιμή είναι 2. Αυτό βέβαια φαίνεται καλύτερα αν έχουμε όλες τις τιμές γραμμένες σε διάταξη, δηλαδή 0, 0, 1, 2, 2, 2, 2, 2, 2, 2, 2 και 3, οπότε η διάμεσος είναι $\frac{2+2}{2} = 2$.

Γενικά, για να υπολογίσουμε τη διάμεσο σε **ν ταξινομημένα διακριτά δεδομένα**, εργαζόμαστε ως εξής:

- Συμπληρώνουμε τον πίνακα συχνοτήτων με τη στήλη των αθροιστικών συχνοτήτων.
- Προσδιορίζουμε τις αθροιστικές συχνότητες N_{i-1} και N_i μεταξύ των οποίων βρίσκεται ο αριθμός $\frac{\nu+1}{2}$ που αντιστοιχεί στη θέση της διαμέσου.
- Η τιμή x_i της μεταβλητής X που βρίσκεται στη $\frac{\nu+1}{2}$ θέση είναι η διάμεσος.

Για παράδειγμα, στον πίνακα 9α βρίσκουμε τη διάμεσο των τιμών της μεταβλητής X: «βαθμός στο μάθημα της Στατιστικής» λέγοντας ότι είναι η τιμή που αφήνει το 50% των παρατηρήσεων κάτω από αυτή και το άλλο 50% πάνω από αυτή. Θα είναι δηλαδή η τιμή της μεταβλητής που αντιστοιχεί στην $\frac{\nu+1}{2}$ θέση, δηλαδή στην «50,5η» θέση. Με αυτό εννοούμε ότι η διάμεσος θα βρίσκεται ανάμεσα στην 50η και στην 51η παρατήρηση, επειδή το πλήθος των τιμών είναι άρτιο και δεν υπάρχει μεσαία παρατήρηση. Από την αθροιστική συχνότητα βλέπουμε ότι υπάρχουν 35 σπουδαστές με βαθμό Στατιστικής μέχρι και 5 και 55 σπουδαστές με βαθμό μέχρι και 6, οπότε ο βαθμός του 50ου όσο και του 51ου σπουδαστή θα είναι 6. Άρα η διάμεσος είναι η τιμή της μεταβλητής X που έχει αθροιστική συχνότητα $N_6 = 55$ και είναι $\delta = \frac{6+6}{2} = 6$.

Πίνακας 9α

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

Όταν **οι τιμές της μεταβλητής είναι ομαδοποιημένες**, μπορούμε να θεωρήσουμε ως θέση της διαμέσου τη $\frac{\nu}{2}$ θέση, όπου ν το πλήθος των τιμών της μεταβλητής. Για παράδειγμα, ας υπολογίσουμε τη διάμεσο των τιμών της μεταβλητής X: «ετήσια τηλεφωνικά έξοδα», που δίνονται στον πίνακα 9β για τις 1200 οικογένειες μιας κωμόπολης.

Πίνακας 9β

Τηλεφωνικά έξοδα (x_i σε ευρώ)	Αριθμός οικογενειών (ν_i)	Αθροιστική συχνότητα (N_i)
[0, 20)	100	100
[20, 40)	200	300
[40, 60)	250	550
[60, 80)	250	800
[80, 100)	200	1000
[100, 120)	100	1100
[120, 140)	60	1160
[140, 160)	20	1180
[160, 180)	10	1190
[180, 200)	10	1200
Σύνολο	**1200**	

Η θέση της διαμέσου είναι τέτοια, ώστε οι $50\% \cdot 1200 = 600$ παρατηρήσεις να είναι μικρότερες από αυτή. Άρα η διάμεσος θα βρίσκεται ανάμεσα στην 600η και την 601η παρατήρηση, αφού το πλήθος των τιμών είναι άρτιο.

Από τη στήλη των αθροιστικών συχνοτήτων βλέπουμε ότι ο αριθμός 600 βρίσκεται ανάμεσα στις αθροιστικές συχνότητες $N_3 = 550$ και $N_4 = 800$. Άρα η διάμεσος περιέχεται στην κλάση $[\alpha_{i-1}, \alpha_i) = [60, 80)$, στην οποία αντιστοιχεί η αθροιστική συχνότητα $N_4 = 800$. Υποθέτοντας ότι οι παρατηρήσεις σε κάθε κλάση κατανέμονται ομοιόμορφα, είναι εύκολο να υπολογίσουμε τη διάμεσο με την «απλή μέθοδο των τριών». Μέχρι την τιμή $\alpha_{i-1} = 60$ έχουμε 550 παρατηρήσεις (τιμές). Για να φθάσουμε τις 600, που αντιστοιχεί η διάμεσος τιμή, θέλουμε ακόμη 50, δηλαδή $\frac{\nu}{2} - N_{i-1} = 600 - 550 = 50$. Γι' αυτές τις 50 θα βρούμε πόσο εύρος καταλαμβάνουν από την επόμενη κλάση. Έχουμε:

Σε εύρος κλάσης c = 20 κατανέμονται ομοιόμορφα 250 παρατηρήσεις (τιμές)
Σε εύρος κλάσης x = ; » » οι 50 »

$$\frac{20}{x} = \frac{250}{50} \Leftrightarrow x = 50 \cdot \frac{20}{250} = 50 \cdot \frac{2}{25} = 4$$

Άρα η διάμεσος τιμή είναι $\delta = 60 + 4 = 64$, δηλαδή

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

όπου: α_{i-1} το κατώτερο όριο της κλάσης που περιέχει τη διάμεσο,
ν_i η συχνότητα της κλάσης που περιέχει τη διάμεσο,
c το πλάτος της κλάσης (οι κλάσεις είναι ίσου πλάτους),
N_{i-1} η αθροιστική συχνότητα της προηγούμενης κλάσης,
ν το πλήθος των παρατηρήσεων (τιμών).

δ) **Μέση τιμή** ή **αριθμητικός μέσος**. Η μέση τιμή ορίζεται ως εξής:

$$\text{M.T.} = \frac{\text{Άθροισμα των τιμών}}{\text{Πλήθος των τιμών}}$$

και συμβολικά:

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_\nu}{\nu}$$

Με $\overline{x}$ συμβολίζουμε τη μέση τιμή των τιμών $x_1, x_2, \ldots, x_\nu$ μιας μεταβλητής και με ν το πλήθος των τιμών της.

Για παράδειγμα, αν οι βαθμοί ενός μαθητή της Γ΄ Γυμνασίου στα Μαθηματικά είναι: το 1ο τρίμηνο 14, το 2ο τρίμηνο 15 και το 3ο τρίμηνο 19, τότε η μέση τιμή ή όπως συνήθως λέμε ο μέσος όρος είναι

$$\overline{x} = \frac{14 + 15 + 19}{3} = 16$$

Όταν έχουμε τις τιμές $x_1, x_2, \ldots, x_\kappa$ και τις αντίστοιχες συχνότητές τους $\nu_1, \nu_2, \ldots, \nu_\kappa$, τότε η μέση τιμή βρίσκεται από τον τύπο:

$$\overline{x} = \frac{\nu_1 x_1 + \nu_2 x_2 + \ldots + \nu_\kappa x_\kappa}{\nu_1 + \nu_2 + \ldots + \nu_\kappa} = \frac{\nu_1 x_1 + \nu_2 x_2 + \ldots + \nu_\kappa x_\kappa}{\nu}$$

Για παράδειγμα, η μέση τιμή της βαθμολογίας των 30 μαθητών (πίνακας 3) είναι

$$\overline{x} = \frac{1 \cdot 8 + 1 \cdot 9 + 1 \cdot 10 + 2 \cdot 11 + 1 \cdot 12 + 2 \cdot 13 + 3 \cdot 14 + 4 \cdot 15}{30} +$$

$$+ \frac{5 \cdot 16 + 5 \cdot 17 + 3 \cdot 18 + 1 \cdot 19 + 1 \cdot 20}{30}, \quad \text{δηλαδή} \quad \overline{x} = 14{,}9$$

Για ευκολία συμπληρώνουμε τον πίνακα συχνοτήτων με μία ακόμη στήλη, τη $\nu_i x_i$, οπότε οι υπολογισμοί απλουστεύονται.

Ο προηγούμενος τύπος γράφεται και ως εξής:

$$\overline{x} = \frac{\nu_1 x_1 + \nu_2 x_2 + \ldots + \nu_\kappa x_\kappa}{\nu} = x_1 \cdot \frac{\nu_1}{\nu} + x_2 \cdot \frac{\nu_2}{\nu} + \ldots + x_\kappa \cdot \frac{\nu_\kappa}{\nu}$$

δηλαδή

$$\overline{x} = x_1 f_1 + x_2 f_2 + \ldots + x_\kappa f_\kappa = \sum_{i=1}^{\kappa} x_i f_i$$

Όταν έχουμε πίνακα συχνοτήτων με ομαδοποιημένες παρατηρήσεις, για να υπολογίσουμε τη μέση τιμή, παίρνουμε ως τιμές της μεταβλητής X τις κεντρικές τιμές των κλάσεων. Έτσι από τον πίνακα 6 μπορούμε να υπολογίσουμε τη μέση τιμή της κατανάλωσης του ηλεκτρικού ρεύματος σε kWh ως εξής:

$$\overline{x} = \frac{3 \cdot 227 + 6 \cdot 241 + 3 \cdot 255 + 7 \cdot 269 + 4 \cdot 283 + 8 \cdot 297 + 3 \cdot 311}{50} +$$

$$+ \frac{2 \cdot 325 + 4 \cdot 339 + 4 \cdot 353 + 2 \cdot 367 + 4 \cdot 381}{50}, \quad \text{δηλαδή} \quad \overline{x} = \frac{14.892}{50} = 297{,}84$$

ε) **Τεταρτημόρια** (quartiles). Τεταρτημόρια λέμε τις τιμές της μεταβλητής που χωρίζουν το σύνολο των τιμών της σε τέσσερις (4) ισοπληθείς ομάδες, όταν οι τιμές της μεταβλητής τοποθετηθούν σε αύξουσα διάταξη.

Έτσι, το **πρώτο τεταρτημόριο** Q_1 θα είναι η τιμή της μεταβλητής που θα έχει κάτω από αυτή το 25% των παρατηρήσεων (τιμών).

Το **δεύτερο τεταρτημόριο** Q_2 θα είναι η τιμή της μεταβλητής που θα έχει κάτω από αυτή το 50% των παρατηρήσεων. Αυτή όμως η τιμή είναι η διάμεσος, δηλαδή $Q_2 = \delta$.

Τέλος, το **τρίτο τεταρτημόριο** Q_3 θα είναι η τιμή της μεταβλητής που θα έχει κάτω από αυτή το 75% των παρατηρήσεων (τιμών).

Σύμφωνα με ό,τι είπαμε προηγουμένως για τη διάμεσο δ, αν ν είναι το πλήθος των διατεταγμένων παρατηρήσεων, τότε η θέση που κατέχει το πρώτο τεταρτημόριο Q_1 είναι η $\frac{\nu+1}{4}$, ενώ η θέση που κατέχει το τρίτο τεταρτημόριο Q_3 είναι η $\frac{3(\nu+1)}{4}$.

Ο υπολογισμός των Q_1 και Q_3 σε ομαδοποιημένα δεδομένα γίνεται ακριβώς όπως και της διαμέσου δ ή γραφικά με τη βοήθεια του διαγράμματος των αθροιστικών συχνοτήτων. Τα τεταρτημόρια Q_1 και Q_3 σε ομαδοποιημένα δεδομένα δεχόμαστε ότι βρίσκονται στις θέσεις $\frac{\nu}{4}$ και $\frac{3\nu}{4}$ αντίστοιχα. Έτσι μπορούμε να χρησιμοποιήσουμε για την εύρεσή τους τύπους αντίστοιχους με αυτόν που χρησιμοποιήσαμε για τη διάμεσο δ, δηλαδή:

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

και

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

Για παράδειγμα, για τα δεδομένα του πίνακα 9β μπορούμε να βρούμε τα Q_1 και Q_3, δηλαδή τις τιμές της μεταβλητής που έχουν κάτω από αυτές το 25% και το 75% των παρατηρήσεων αντιστοίχως.

Έχουμε $\frac{\nu}{4} = \frac{1200}{4} = 300$ και $\frac{3\nu}{4} = \frac{3 \cdot 1200}{4} = 900$. Ο αριθμός $\frac{\nu}{4} = 300$ αντιστοιχεί ακριβώς στην αθροιστική συχνότητα $N_2 = 300$, οπότε το Q_1 θα είναι ακριβώς 40, δηλαδή $Q_1 = 40$. Αυτό σημαίνει ότι το 25% των οικογενειών της κωμόπολης έχουν τηλεφωνικά έξοδα το πολύ 40 ευρώ ετησίως.

Για το Q_3 χρειαζόμαστε υπολογισμούς όπως και για τη διάμεσο. Το Q_3 θα είναι η τιμή της 900ης παρατήρησης, η οποία περιέχεται στην κλάση $[\alpha_{i-1}, \alpha_i) = [80, 100)$. Έτσι έχουμε:

$$\frac{3\nu}{4} - N_{i-1} = 900 - 800 = 100, \quad c = 20 \quad \text{και} \quad \nu_i = 200, \quad \text{οπότε}$$

$$Q_3 = 80 + 100 \cdot \frac{20}{200} = 80 + 10 = 90$$

Παρατηρήσεις

1. Πολλές φορές οι τιμές $x_1, x_2, \ldots, x_\nu$ μιας μεταβλητής X έχουν διαφορετική βαρύτητα. Στις περιπτώσεις αυτές αντί του αριθμητικού μέσου χρησιμοποιούμε τον **σταθμισμένο αριθμητικό μέσο ή σταθμικό μέσο** (weighted mean). Δηλαδή, αν στις τιμές $x_1, x_2, \ldots, x_\nu$ της μεταβλητής X αντιστοιχούν συντελεστές στάθμισης (βαρύτητας) $w_1, w_2, \ldots, w_\nu$, τότε ο σταθμικός μέσος θα είναι

$$\bar{x}' = \frac{w_1x_1 + w_2x_2 + \ldots + w_\nu x_\nu}{w_1 + w_2 + \ldots + w_\nu} = \frac{\sum_{i=1}^{\nu} w_i x_i}{\sum_{i=1}^{\nu} w_i}$$

Για παράδειγμα, προκειμένου να εισαχθεί ένας απόφοιτος Ενιαίου Λυκείου στην τριτοβάθμια εκπαίδευση, συνυπολογίζονται ο βαθμός x_1 του απολυτηρίου με συντελεστή (βαρύτητα) $w_1 = 7$, ο βαθμός x_2 στο τεστ δεξιοτήτων με συντελεστή $w_2 = 1{,}5$, ο βαθμός x_3 στο 1ο βασικό μάθημα με συντελεστή $w_3 = 1$ και ο βαθμός x_4 στο 2ο βασικό μάθημα με συντελεστή $w_4 = 0{,}5$. Ας υποθέσουμε ότι ένας μαθητής έχει βαθμό απολυτηρίου $x_1 = 15$, βαθμό στο τεστ δεξιοτήτων $x_2 = 16$, βαθμό στο 1ο βασικό μάθημα $x_3 = 18$ και βαθμό στο 2ο βασικό μάθημα $x_4 = 16$. Τότε ο σταθμικός μέσος της επίδοσής του θα είναι

$$\bar{x}' = \frac{7 \cdot 15 + 1{,}5 \cdot 16 + 1 \cdot 18 + 0{,}5 \cdot 16}{7 + 1{,}5 + 1 + 0{,}5} = \frac{105 + 24 + 18 + 8}{10} = \frac{155}{10} = 15{,}5$$

2. Σε πολλά συγγράμματα Στατιστικής υπάρχει ο ακόλουθος διαχωρισμός στον συμβολισμό της μέσης τιμής. Όταν εξετάζουμε ένα πεπερασμένο δείγμα μεγέθους ν με τιμές $x_1, x_2, \ldots, x_\nu$ για τη μεταβλητή X, τότε η μέση τιμή (που λέγεται και **δειγματική**) συμβολίζεται, όπως προαναφέραμε, με $\bar{x}$. Όταν όμως εξετάζουμε ολόκληρο τον πληθυσμό μεγέθους N, τότε η μέση τιμή (που λέγεται και **πληθυσμιακή**) συμβολίζεται με μ, δηλαδή

$$\mu = \frac{x_1 + x_2 + \ldots + x_N}{N} \text{ ή } \mu = \frac{x_1\nu_1 + x_2\nu_2 + \ldots + x_\kappa\nu_\kappa}{\nu_1 + \nu_2 + \ldots + \nu_\kappa} \text{ για ομαδοποιημένες κατανομές.}$$

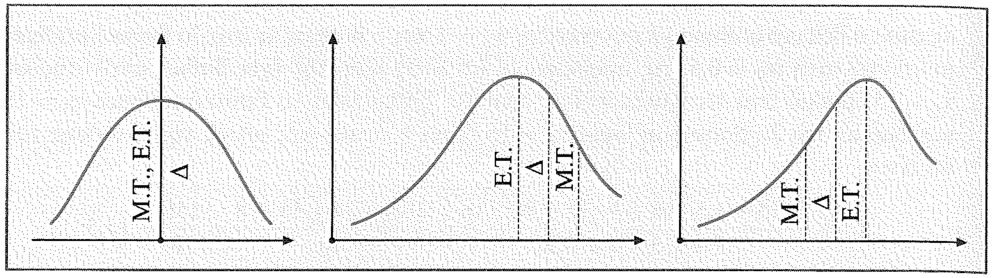
Στη συνέχεια θα χρησιμοποιούμε το σύμβολο $\bar{x}$, εκτός αν η φύση του θέματος που μελετάμε επιβάλλει να χρησιμοποιήσουμε το μ, οπότε θα δηλώνεται με σαφήνεια.

Αν θέλουμε να συγκρίνουμε τις τέσσερις παραμέτρους θέσης που αναφέραμε προηγουμένως, μπορούμε να πούμε τα εξής:

Το μέσο εύρος είναι εύκολο να υπολογιστεί, αλλά για τον υπολογισμό του χρησιμοποιούνται μόνο οι δύο ακραίες τιμές της κατανομής και καμιά άλλη πληροφορία της. Έτσι δεν μπορούμε να εμπιστευθούμε το μέσο εύρος για αντικειμενική εκπροσώπηση της κατανομής. Η επικρατέστερη τιμή είναι ικανοποιητικό μέτρο θέσης αν η κατανομή είναι συμμετρική. Αν δεν συμβαίνει αυτό, τότε η επικρατέστερη τιμή δεν μπορεί να είναι αντιπροσωπευτικός αριθμός της κατανομής. Η διάμεσος δεν επηρεάζεται καθόλου από τις ακραίες τιμές και για τον υπολογισμό της λαμβάνονται υπόψη όλες οι τιμές. Η μέση τιμή επηρεάζεται από τις ακραίες τιμές και λαμβάνει υπόψη της όλες τις τιμές.

Έτσι θα μπορούσαμε να πούμε ότι η διάμεσος είναι το καλύτερο μέτρο θέσης μιας κατανομής, αλλά η δυσκολία υπολογισμού της σε μερικές περιπτώσεις (ομαδοποιημένες παρατηρήσεις) μας αναγκάζει να χρησιμοποιούμε περισσότερο τη μέση τιμή, η οποία έχει και μεγάλη εφαρμογή στη στατιστική ανάλυση. Βέβαια κανένας δεν μας δεσμεύει να χρησιμοποιήσουμε όποιο μέτρο θέσης νομίζουμε ότι είναι καλύτερο για μια συγκεκριμένη κατανομή.

Στα επόμενα σχήματα φαίνεται η σχέση μεταξύ των μέτρων θέσης και των διάφορων τύπων των διαγραμμάτων συχνότητας.



Σχ. 10γ. M.T. = Μέση τιμή, E.T. = Επικρατέστερη τιμή, Δ = Διάμεσος

Παρατηρούμε ότι στο πρώτο σχήμα, όπου το διάγραμμα συχνοτήτων έχει άξονα συμμετρίας, οι τρεις παράμετροι συμπίπτουν και επομένως δίνουν την ίδια πληροφορία. Στα δύο άλλα σχήματα, όπου οι τρεις παράμετροι δεν συμπίπτουν, οι πληροφορίες είναι διαφορετικές. Η πιο αξιόπιστη από αυτές φαίνεται ότι είναι η διάμεσος.

Αν θέλουμε να έχουμε μια σαφέστερη εικόνα για τα πλεονεκτήματα και τα μειονεκτήματα των τριών κυριότερων μέτρων θέσης, τότε μπορούμε να φτιάξουμε τον επόμενο συγκριτικό πίνακα:

Σύγκριση μέσης τιμής, διαμέσου και επικρατούσας τιμής

Πλεονεκτήματα	**Μειονεκτήματα**
Μέση τιμή	
• Για τον υπολογισμό της χρησιμοποιούνται όλες οι τιμές. • Είναι μοναδική για κάθε σύνολο δεδομένων. • Είναι εύκολα κατανοητή. • Ο υπολογισμός της είναι σχετικά εύκολος. • Έχει μεγάλη εφαρμογή για περαιτέρω στατιστική ανάλυση.	• Επηρεάζεται πολύ από ακραίες τιμές. • Μπορεί να μην αντιστοιχεί σε κάποια τιμή της μεταβλητής. Όταν η X είναι διακριτή με ακέραιες τιμές, τότε η μέση τιμή μπορεί να μην είναι ακέραιος. • Δεν υπολογίζεται για ποιοτικά δεδομένα. • Είναι δύσκολος ο υπολογισμός της σε ομαδοποιημένα δεδομένα με ανοιχτές τις ακραίες κλάσεις.
Διάμεσος	
• Είναι εύκολα κατανοητή. • Δεν επηρεάζεται από ακραίες τιμές. • Υπολογίζεται και στην περίπτωση που οι ακραίες κλάσεις είναι ανοιχτές. • Ο υπολογισμός της είναι απλός. • Είναι μοναδική σε κάθε σύνολο δεδομένων.	• Δεν χρησιμοποιούνται όλες οι τιμές για τον υπολογισμό της. • Είναι δύσκολη η εφαρμογή της για περαιτέρω στατιστική ανάλυση. • Δεν υπολογίζεται για ποιοτικά δεδομένα. • Για τον υπολογισμό της μπορεί να χρειαστεί παρεμβολή.
Επικρατούσα τιμή	
• Υπολογίζεται εύκολα όταν δεν έχουμε ομαδοποιημένα δεδομένα. • Είναι εύκολα κατανοητή. • Υπολογίζεται και από ελλιπή δεδομένα. • Δεν επηρεάζεται από ακραίες τιμές. • Εφαρμόζεται και σε ποιοτικά δεδομένα.	• Δεν χρησιμοποιούνται όλες οι τιμές. • Δεν χρησιμοποιείται εύκολα για περαιτέρω στατιστική ανάλυση. • Δεν ορίζεται πάντα μονοσήμαντα. Μπορούμε να έχουμε πολλές κορυφές ή και καμία.

• *Παράμετροι διασποράς*

Στην προηγούμενη παράγραφο ασχοληθήκαμε με τις παραμέτρους θέσης, που είναι αριθμοί οι οποίοι μας δίνουν πληροφορίες για το πού είναι το «κέντρο» των τιμών μιας μεταβλητής ή ενός δείγματος.

Ας πάρουμε τη βαθμολογία τριών μαθητών στα 10 μαθήματά τους.

A: 14, 14, 15, 15, 16, 16, 17, 17, 18, 18 $\bar{x} = 16$
B: 10, 10, 15, 15, 16, 16, 19, 19, 20, 20 $\bar{x} = 16$
Γ: 8, 10, 16, 16, 16, 16, 19, 19, 20, 20 $\bar{x} = 16$

Ο καθένας από τους τρεις έχει μέσο όρο βαθμολογίας 16. Είναι όμως προφανές ότι η διασπορά των βαθμών διαφέρει σημαντικά. Στον μαθητή A όλοι οι βαθμοί συγκεντρώνονται γύρω στο 16 με απόσταση δύο μονάδων από αυτόν. Στον B υπάρχουν τιμές που απέχουν από το 16 μέχρι 6 μονάδες, ενώ στον Γ μέχρι και 8 μονάδες. Είναι φανερό λοιπόν ότι μόνη της η μέση τιμή $\bar{x}$ δεν μπορεί να μας δώσει σαφή εικόνα μιας κατανομής. Επομένως, εκτός από τον $\bar{x}$, θα ήταν επιθυμητό να έχουμε έναν ακόμη αριθμό που να φανερώνει τον τρόπο διασποράς των τιμών της μεταβλητής. Αυτοί οι δύο αριθμοί (και οι δύο μαζί) θα μας έδιναν καλύτερες πληροφορίες για την κατανομή των τιμών (πληροφοριών).

Οι αριθμοί που μας πληροφορούν για τον τρόπο διασποράς των τιμών λέγονται **παράμετροι** ή **μέτρα διασποράς**. Αυτοί είναι το εύρος, η μέση απόλυτη απόκλιση, η διακύμανση και η τυπική απόκλιση.

α) **Το εύρος**. Είναι το απλούστερο μέτρο διασποράς. Είναι η διαφορά της μικρότερης τιμής από τη μεγαλύτερη και υπολογίζεται εύκολα όταν οι τιμές έχουν διαταχθεί από τη μικρότερη προς τη μεγαλύτερη. Για παράδειγμα, το εύρος της βαθμολογίας του μαθητή Α είναι $18 - 14 = 4$, του Β είναι $20 - 10 = 10$ και του Γ είναι $20 - 8 = 12$. Η σημαντικότερη αντίρρηση για να χρησιμοποιούμε το εύρος ως παράμετρο διασποράς είναι ότι λαμβάνουμε υπόψη μόνο δύο τιμές για τον υπολογισμό του και δεν χρησιμοποιούμε καθόλου τις άλλες τιμές.

β) **Η διακύμανση και η τυπική απόκλιση**. Ο βαθμός συγκέντρωσης των τιμών $x_1, x_2, \dots, x_\nu$ μιας μεταβλητής Χ γύρω από τη μέση τιμή $\bar{x}$ προσδιορίζεται καλύτερα με τη βοήθεια των θετικών αριθμών $(x_i - \bar{x})^2$, οι οποίοι είναι καταλληλότεροι από τις απόλυτες τιμές για αλγεβρικό λογισμό. Έτσι ορίζουμε ως **διακύμανση της μεταβλητής Χ τη μέση τιμή των τετραγώνων των αποκλίσεων των τιμών από τον αριθμητικό της μέσο και τη συμβολίζουμε με S^2**, δηλαδή

$$S^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_\nu - \bar{x})^2}{\nu} \quad (1)$$

Εδώ υπάρχει ένα μικρό πρόβλημα. Οι μονάδες στις οποίες εκφράζεται η διακύμανση είναι τα τετράγωνα των μονάδων στις οποίες εκφράζονται οι τιμές της μεταβλητής. Γι' αυτό, αντί της διακύμανσης, χρησιμοποιούμε ως μέτρο διασποράς την τετραγωνική ρίζα της διακύμανσης, που τη λέμε **τυπική απόκλιση**, δηλαδή

$$S = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_\nu - \bar{x})^2}{\nu}} \quad (2)$$

Για τον υπολογισμό της διακύμανσης από τον πίνακα συχνοτήτων χρησιμοποιούμε τον τύπο

$$S^2 = \frac{\nu_1(x_1 - \bar{x})^2 + \nu_2(x_2 - \bar{x})^2 + \dots + \nu_\kappa(x_\kappa - \bar{x})^2}{\nu_1 + \nu_2 + \dots + \nu_\kappa} \quad (3)$$

οπότε η τετραγωνική του ρίζα μας δίνει την τυπική απόκλιση S.

Στις ομαδοποιημένες παρατηρήσεις χρησιμοποιούμε και πάλι τον τύπο 3 για το S^2, αλλά όπου x_i παίρνουμε τις κεντρικές τιμές των κλάσεων.

Οι υπολογισμοί στις περιπτώσεις αυτές γίνονται ευκολότερα αν χρησιμοποιήσουμε πίνακα συχνοτήτων, αλλά συμπληρωμένο με δύο ακόμη στήλες, τη $(x_i - \bar{x})^2$ και τη $\nu_i(x_i - \bar{x})^2$.

Παράδειγμα

Η διακύμανση της βαθμολογίας του μαθητή Α του προηγούμενου παραδείγματος είναι

$$S^2 = \frac{(14-16)^2 + (14-16)^2 + \dots + (18-16)^2}{10} = \frac{20}{10} = 2$$

και η τυπική απόκλιση είναι $S = \sqrt{2} \simeq 1{,}41$.

Οι ίδιες παράμετροι για τον μαθητή Γ είναι

$$S^2 = \frac{(8-16)^2 + (10-16)^2 + \dots + (20-16)^2}{10} = \frac{150}{10} = 15 \quad \text{και} \quad S = \sqrt{15} \simeq 3{,}87$$

Βλέπουμε ότι η βαθμολογία του μαθητή με τη μικρότερη διασπορά ως προς τη μέση τιμή έχει και τη μικρότερη τυπική απόκλιση.

Αν θέλουμε να συγκρίνουμε τις παραμέτρους διασποράς που αναφέραμε προηγουμένως, μπορούμε να πούμε τα εξής:

Το **εύρος** είναι πολύ απλό στον υπολογισμό, χρησιμοποιείται στον έλεγχο ποιότητας και μπορεί να χρησιμοποιηθεί για την εκτίμηση της τυπικής απόκλισης, αλλά δεν θεωρείται αξιόπιστο μέτρο διασποράς, ακριβώς γιατί βασίζεται μόνο σε δύο ακραίες τιμές. Επίσης δεν χρησιμοποιείται για περαιτέρω στατιστική ανάλυση.

Η **διακύμανση** και η **τυπική απόκλιση** είναι τα σπουδαιότερα μέτρα διασποράς, γιατί για τον υπολογισμό τους λαμβάνονται υπόψη όλες οι παρατηρήσεις (τιμές) και έχουν μεγάλη εφαρμογή στη στατιστική συμπερασματολογία. Το μειονέκτημα της διακύμανσης, ότι δηλαδή δεν εκφράζεται στις ίδιες μονάδες με τις οποίες εκφράζονται οι τιμές της μεταβλητής, εξαλείφεται με την τυπική απόκλιση. Ούτε οι περισσότερες αλγεβρικές πράξεις που απαιτούνται για τον υπολογισμό τους αποτελούν μειονέκτημα στη σημερινή εποχή που χρησιμοποιούμε κάθε είδους υπολογιστές.

Αν θέλουμε να έχουμε μια σαφέστερη εικόνα για τα πλεονεκτήματα και τα μειονεκτήματα των κυριότερων μέτρων διασποράς, τότε μπορούμε να φτιάξουμε τον επόμενο συγκριτικό πίνακα:

Σύγκριση μέτρων διασποράς

Πλεονεκτήματα	Μειονεκτήματα
Εύρος	
• Είναι πολύ απλό στον υπολογισμό. • Χρησιμοποιείται αρκετά στον έλεγχο ποιότητας. • Μπορεί να χρησιμοποιηθεί για την εκτίμηση της τυπικής απόκλισης.	• Δεν θεωρείται αξιόπιστο μέτρο διασποράς, επειδή βασίζεται μόνο στις δύο ακραίες παρατηρήσεις. • Δεν χρησιμοποιείται για περαιτέρω στατιστική ανάλυση.
Διακύμανση και τυπική απόκλιση	
• Λαμβάνονται υπόψη για τον υπολογισμό τους όλες οι παρατηρήσεις. • Έχουν μεγάλη εφαρμογή στη στατιστική συμπερασματολογία. • Σε κανονικές κατανομές το 68,3%, το 95,4% και το 99,7% των παρατηρήσεων βρίσκονται στα διαστήματα $\bar{x} \pm S$, $\bar{x} \pm 2S$ και $\bar{x} \pm 3S$ αντίστοιχα.	• Το κυριότερο μειονέκτημα της διακύμανσης είναι ότι δεν εκφράζεται στις ίδιες μονάδες με τις τιμές. Το μειονέκτημα αυτό παύει να υπάρχει με τη χρησιμοποίηση της τυπικής απόκλισης. • Απαιτούνται περισσότερες αλγεβρικές πράξεις για τον υπολογισμό τους απ' ό,τι για τον υπολογισμό των άλλων μέτρων.

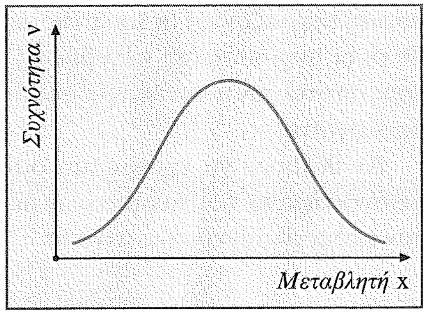
Η τυπική απόκλιση είναι το σπουδαιότερο μέτρο (παράμετρος) διασποράς, γιατί χρησιμοποιεί όλες τις τιμές της μεταβλητής και εκφράζεται με την ίδια μονάδα που εκφράζονται και οι τιμές.

Η μέση τιμή $\bar{x}$ και η τυπική απόκλιση S χρησιμοποιούν όλες τις τιμές. Η μέση τιμή δίνει το κέντρο της κατανομής και η τυπική απόκλιση δίνει ένα μέτρο της διασποράς της κατανομής.

Ας δούμε τώρα τη χρησιμότητα της μέσης τιμής και της τυπικής απόκλισης στην πιο γνωστή μας κατανομή συχνοτήτων, την **κανονική κατανομή**.

Μια κατανομή συχνοτήτων λέμε ότι είναι κανονική όταν το διάγραμμα συχνοτήτων της έχει τη μορφή του σχήματος 11.

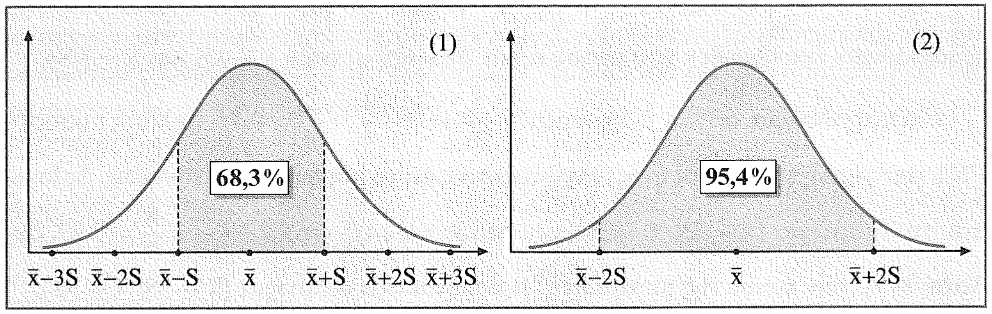
Αν τώρα η κατανομή ενός δείγματος είναι κανονική με μέση τιμή $\bar{x}$ και τυπική απόκλιση S, τότε θα ισχύουν πάντοτε τα εξής:



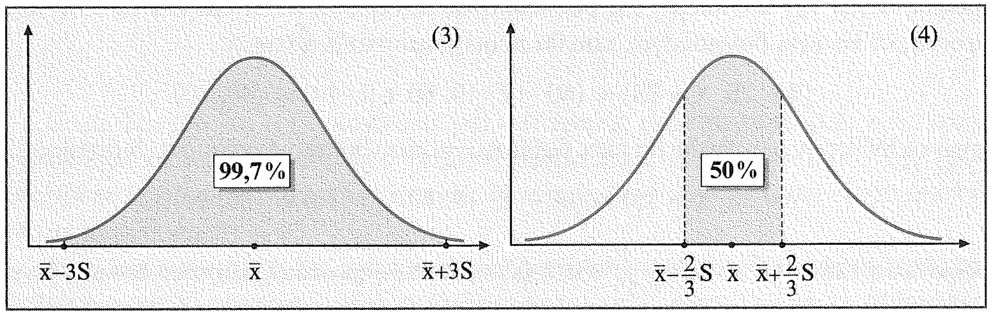
Σχ. 11

- **Το 68,3% των μεσαίων τιμών βρίσκεται μεταξύ $\bar{x} - S$ και $\bar{x} + S$.**
- **Το 95,4% των μεσαίων τιμών βρίσκεται μεταξύ $\bar{x} - 2S$ και $\bar{x} + 2S$.**
- **Το 99,7% των μεσαίων τιμών βρίσκεται μεταξύ $\bar{x} - 3S$ και $\bar{x} + 3S$.**
- **Το 50% των μεσαίων τιμών βρίσκεται μεταξύ $\bar{x} - \frac{2}{3}S$ και $\bar{x} + \frac{2}{3}S$.**

Τα συμπεράσματα αυτά αποδίδονται εποπτικά με τα επόμενα σχήματα.



Σχ. 12



Σχ. 13

Παράδειγμα

Θεωρούμε όλους τους μηνιαίους μισθούς των εργαζομένων σ' ένα εργοστάσιο και υπολογίζουμε ότι η μέση τιμή των μισθών είναι $\bar{x} = 1000$ € η τυπική απόκλιση $S = 200$ € και ότι η κατανομή των μισθών είναι κανονική.

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

i) Το 68,3% των εργαζομένων παίρνει μισθούς 1000 ± 200, δηλαδή μεταξύ 800 € και 1200 €.

ii) Το 95,4% των εργαζομένων παίρνει μισθούς $1000 \pm 2 \cdot 200$, δηλαδή μεταξύ 600 € και 1400 €.

iii) Το 99,7% των εργαζομένων παίρνει μισθούς $1000 \pm 3 \cdot 200$, δηλαδή μεταξύ 400 € και 1600 €.

Τονίζουμε ότι όλα τα παραπάνω ποσοστά ισχύουν όταν η κατανομή είναι **κανονική**. Όταν η κατανομή δεν είναι κανονική, τότε ο Chebyshev απέδειξε την ακόλουθη πρόταση, που είναι γνωστή ως **θεώρημα του Chebyshev**: **«Για οποιοδήποτε σύνολο δεδομένων (πληθυσμού ή δείγματος) και για οποιαδήποτε σταθερά λ (με $\lambda > 1$), ποσοστό τουλάχιστον $1 - \frac{1}{\lambda^2}$ των δεδομένων βρίσκεται εντός λ τυπικών αποκλίσεων εκατέρωθεν του μέσου όρου, δηλαδή μεταξύ $\bar{x} - \lambda S$ και $\bar{x} + \lambda S$»**.

Για παράδειγμα, αν $\lambda = 2$, έχουμε $1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4} = 0{,}75$, δηλαδή ποσοστό 75% των τιμών της μεταβλητής, ανεξάρτητα από τη μορφή της κατανομής, βρίσκεται στο διάστημα $(\bar{x} - 2S,\ \bar{x} + 2S)$, ενώ για $\lambda = 3$ ποσοστό $1 - \frac{1}{3^2} = \frac{8}{9} = 0{,}89$, δηλαδή 89%, των τιμών της μεταβλητής βρίσκεται στο διάστημα $(\bar{x} - 3S,\ \bar{x} + 3S)$.

Επομένως, σύμφωνα με το θεώρημα του Chebyshev, αν η επίδοση των φοιτητών στη Στατιστική έχει παραμέτρους $\bar{x} = 60$ και $S = 10$, τότε χωρίς καμιά άλλη πληροφορία για την κατανομή της βαθμολογίας μπορούμε να πούμε ότι το 75% των φοιτητών θα έχει βαθμολογία που θα περιέχεται στο διάστημα

$$(\bar{x} - 2S,\ \bar{x} + 2S) = (60 - 2 \cdot 10,\ 60 + 2 \cdot 10) = (40,\ 80)$$

ενώ το 89% των φοιτητών θα έχει βαθμολογία που θα περιέχεται στο διάστημα

$$(\bar{x} - 3S,\ \bar{x} + 3S) = (60 - 3 \cdot 10,\ 60 + 3 \cdot 10) = (30,\ 90)$$

Σημείωση: Για $\lambda = 1$ είναι $1 - \frac{1}{1^2} = 0$, δηλαδή το θεώρημα του Chebyshev δεν μας δίνει πληροφορίες για το ποσοστό των τιμών της μεταβλητής που βρίσκεται στο διάστημα $(\bar{x} - S,\ \bar{x} + S)$.

• *Μέτρα ασυμμετρίας και μέτρα κυρτότητας*

Τα μέτρα θέσης και τα μέτρα διασποράς που μελετήσαμε προηγουμένως χρησιμοποιούνται ευρύτατα στις πρακτικές εφαρμογές για να δούμε γύρω από ποιον αριθμό (θέση) συγκεντρώνονται οι τιμές της μεταβλητής και με ποιον βαθμό συγκέντρωσης - μεταβλητότητας (διασποράς) γίνεται αυτό.

Πολλές φορές όμως είναι ανάγκη, εκτός από τη θέση και τη διασπορά, να γνωρίζουμε και τη **μορφολογία** μιας κατανομής, δηλαδή την **ασυμμετρία** και την **κυρτότητά** της. Η ασυμμετρία αναφέρεται στην απόκλιση του διαγράμματος της κατανομής συχνοτήτων από το διάγραμμα της κανονικής κατανομής, ενώ η κυρτότητα αναφέρεται στο πόσο πεπλατυσμένο είναι το διάγραμμα της κατανομής συχνοτήτων.

Οι αριθμοί που καθορίζουν τον βαθμό ασυμμετρίας μιας κατανομής λέγονται **μέτρα ασυμμετρίας**, ενώ οι αριθμοί που καθορίζουν τον βαθμό κυρτότητας μιας κατανομής λέγονται **μέτρα κυρτότητας**.

Εδώ θα δούμε από ένα τέτοιο μέτρο για κάθε περίπτωση. Τα μέτρα αυτά οφείλονται στον Karl Pearson.

α) **Μέτρο (δείκτης ή συντελεστής) ασυμμετρίας του Pearson**. Αν μ και σ είναι η μέση τιμή και η τυπική απόκλιση των τιμών μιας μεταβλητής (**πληθυσμιακές και όχι δειγματικές παράμετροι**), τότε ορίζουμε ως παράμετρο (μέτρο) ασυμμετρίας της κατανομής συχνοτήτων τον αριθμό

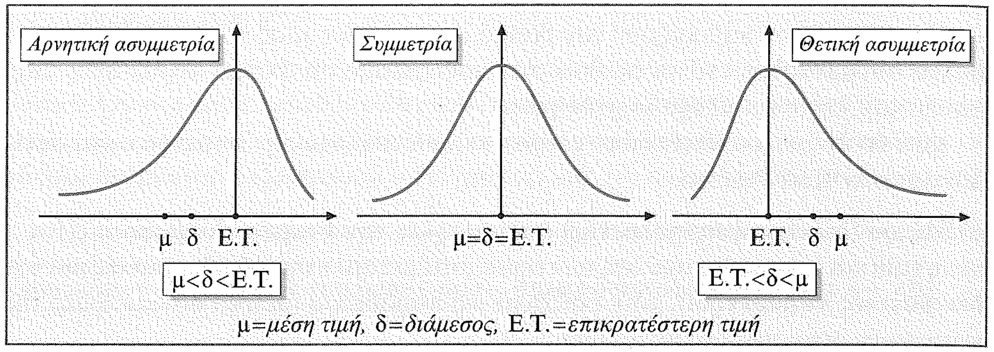
$$\beta_1 = \frac{\mu_3}{\sigma^3}, \quad \text{όπου} \quad \mu_3 = \frac{\sum_{i=1}^{N}(x_i - \mu)^3}{N} \quad \text{ή}$$

$$\mu_3 = \frac{\sum_{i=1}^{\kappa} v_i(x_i - \mu)^3}{\sum_{i=1}^{\kappa} v_i} \quad \text{και } v_i \text{ η συχνότητα της τιμής } x_i$$

Είναι φανερό ότι αν η καμπύλη της κατανομής συχνοτήτων είναι κανονική, δηλαδή αν οι ισαπέχουσες από τη μέση τιμή μ τιμές x_i παρουσιάζουν την ίδια συχνότητα, τότε οι θετικές και αρνητικές διαφορές $(x_i - \mu)^3$ έχουν άθροισμα μηδέν, οπότε $\mu_3 = 0$ και $\beta_1 = 0$.

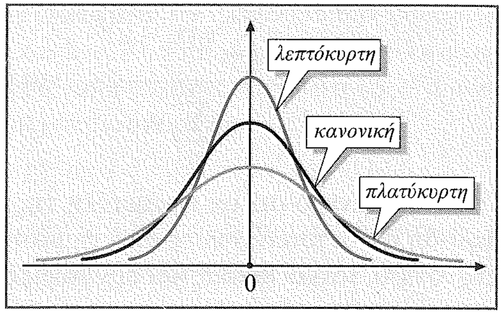
Αν όμως η καμπύλη της κατανομής συχνοτήτων παρουσιάζει «**ουρά**» προς τα δεξιά ή προς τ' αριστερά, τότε οι διαφορές $(x_i - \mu)^3$ είναι θετικές ή αρνητικές αντίστοιχα, οπότε $\mu_3 > 0$ ή $\mu_3 < 0$. Σε αυτή την περίπτωση θα έχουμε $\beta_1 > 0$ (θετική ασυμμετρία) ή $\beta_1 < 0$ (αρνητική ασυμμετρία) αντίστοιχα.

Το επόμενο σχήμα αποδίδει γραφικά τα συμπεράσματα αυτά.



Σχ. 14

β) **Μέτρο κυρτότητας του Pearson.** Για να χαρακτηρίσουμε την κυρτότητα του διαγράμματος (της καμπύλης) μιας κατανομής συχνοτήτων, το συγκρίνουμε και πάλι με το διάγραμμα της κανονικής κατανομής. Έτσι, μια καμπύλη κατανομής συχνοτήτων η οποία παρουσιάζει «αιχμηρότητα» **μεγαλύτερη**, **ίση** ή **μικρότερη** από αυτή της κανονικής κατανομής χαρακτηρίζεται αντίστοιχα ως **λεπτόκυρτη**, **κανονική** ή **πλατύκυρτη**, όπως φαίνεται στο διπλανό σχήμα.



Σχ. 15

Ορίζουμε ως παράμετρο (μέτρο) κυρτότητας της κατανομής συχνοτήτων (δείκτη κυρτότητας του Pearson) τον αριθμό:

$$\beta_2 = \frac{\mu_4}{\sigma^4}, \quad \text{όπου} \quad \mu_4 = \frac{\sum_{i=1}^{N}(x_i - \mu)^4}{N} \quad \text{ή} \quad \mu_4 = \frac{\sum_{i=1}^{\kappa} v_i(x_i - \mu)^4}{\sum_{i=1}^{\kappa} v_i}$$

Αποδεικνύεται ότι ο συντελεστής (δείκτης) β_2 για κανονικές κατανομές είναι πάντοτε τρία (δηλαδή $\beta_2 = 3$), για λεπτόκυρτες κατανομές είναι $\beta_2 > 3$ και για πλατύκυρτες είναι $\beta_2 < 3$.

Είναι φανερό ότι όσο μεγαλύτερη είναι η διαφορά $\beta_2 - 3$, τόσο «αιχμηρότερη» είναι η καμπύλη της κατανομής συχνοτήτων.

Σημείωση: Τα παραπάνω συμπεράσματα ισχύουν για τις περισσότερες κατανομές που μελετάμε, αλλά υπάρχουν και κατανομές για τις οποίες δεν ισχύουν. Εμείς δεν θα ασχοληθούμε με τέτοιες κατανομές.

Σε προηγούμενη παράγραφο είδαμε ότι αν έχουμε μια κανονική κατανομή ή κατανομή Gauss, τότε το 68,3% των παρατηρήσεων θα βρίσκεται στο διάστημα $(\bar{x} - S, \bar{x} + S)$. Αυτό μπορεί να χρησιμοποιηθεί ως μερικός δείκτης του βαθμού κύρτωσης της κατανομής. Δηλαδή, αν σε μια κατανομή το 68,3% δεν περιέχεται σε αυτό το διάστημα, τότε η κατανομή δεν είναι κανονική. Αν το ποσοστό που περιέχεται στο διάστημα αυτό είναι μικρότερο του 68,3%, τότε η κατανομή είναι πλατύκυρτη, ενώ αν είναι μεγαλύτερο του 68,3%, η κατανομή είναι λεπτόκυρτη.

Πρέπει να παρατηρήσουμε ότι η κύρτωση αναφέρεται σε συμμετρικές κατανομές.

Σχόλιο: Τα σημαντικότερα από τα μέτρα θέσης και διασποράς (μέση τιμή και τυπική απόκλιση) καθώς και οι δείκτες ασυμμετρίας και κυρτότητας μιας κατανομής συχνοτήτων ανήκουν σε μια ευρύτερη οικογένεια παραμέτρων που ονομάζονται **ροπές**.

Οι ροπές ορίζονται είτε ως προς την αρχή των δεδομένων (δηλαδή το μηδέν) είτε ως προς το μέσο (κεντρική ροπή). Επίσης οι ροπές διαφοροποιούνται αν τα δεδομένα είναι ταξινομημένα σε κατανομές συχνοτήτων ή αταξινόμητα.

Ορίζουμε ως ροπή t-τάξης ως προς την αρχή μηδέν τον αριθμό $V_t = \frac{\sum_{i=1}^{N} x_i^t}{N}$, όταν τα δεδομένα είναι αταξινόμητα, και τον αριθμό $V_t = \frac{\sum_{i=1}^{\kappa} v_i x_i^t}{\sum_{i=1}^{\kappa} v_i}$, όταν τα δεδομένα είναι ταξινομημένα σε κατανομές συχνοτήτων.

Ορίζουμε ως ροπή t-τάξης ως προς τη μέση τιμή (t-κεντρική ροπή) τον αριθμό

$$\mu_t = \frac{\sum_{i=1}^{N}(x_i - \mu)^t}{N}$$

όταν τα δεδομένα είναι αταξινόμητα, και τον αριθμό $\mu_t = \frac{\sum_{i=1}^{\kappa} v_i(x_i - \mu)^t}{\sum_{i=1}^{\kappa} v_i}$, όταν τα δεδομένα είναι ταξινομημένα σε κατανομές συχνοτήτων.

Είναι φανερό ότι:

- Η μηδενική κεντρική ροπή $(t = 0)$ είναι $\mu_0 = 1$.
- Η πρώτη κεντρική ροπή $(t = 1)$ είναι $\mu_1 = 0$.
- Η δεύτερη κεντρική ροπή $(t = 2)$ είναι $\mu_2 = \sigma^2$, δηλαδή ίση με τη διακύμανση.
- Η τρίτη και η τέταρτη κεντρική ροπή είναι οι αριθμοί μ_3 και μ_4, οι οποίοι παρουσιάζονται στους συντελεστές ασυμμετρίας και κυρτότητας αντιστοίχως.

7.5 Ανάλυση δεδομένων με μεθόδους της Επαγωγικής Στατιστικής

Η Επαγωγική Στατιστική είναι ο δεύτερος μεγάλος κλάδος της Στατιστικής (ο πρώτος είναι η Περιγραφική Στατιστική) που μας βοηθά να κάνουμε ανάλυση των δεδομένων μιας έρευνας. Η Επαγωγική Στατιστική αναφέρεται σε ορισμένους τύπους διαδικασιών που επιτρέπουν στους ερευνητές να εξάγουν συμπεράσματα για όλο τον πληθυσμό, βασισμένα σε ευρήματα από ένα αντιπροσωπευτικό δείγμα του πληθυσμού. Στο 3ο κεφάλαιο μιλήσαμε αναλυτικά για το πώς επιτυγχάνουμε αντιπροσωπευτικό δείγμα ενός πληθυσμού, για τη δειγματική κατανομή του μέσου $\bar{x}$ και για το τυπικό σφάλμα. Επισημάνθηκε στο κεφάλαιο αυτό ότι όταν ένα δείγμα είναι αντιπροσωπευτικό του πληθυσμού, τότε όλα τα χαρακτηριστικά του πληθυσμού υποτίθεται ότι θα είναι και στο δείγμα και μάλιστα στον ίδιο βαθμό. Φυσικά, ούτε και η τυχαία δειγματοληψία μάς εξασφαλίζει απόλυτα την αντιπροσωπευτικότητα ενός δείγματος, αλλά με αυτή τη διαδικασία επιτυγχάνουμε το αντιπροσωπευτικότερο δυνατό. Έτσι είμαστε βέβαιοι ότι τα ευρήματα από το δείγμα προσεγγίζουν καλύτερα τα χαρακτηριστικά του πληθυσμού.

Μια δεύτερη σημαντική προσφορά της Επαγωγής Στατιστικής είναι ο έλεγχος υποθέσεων και ο έλεγχος στατιστικής σημαντικότητας, τα οποία αναπτύξαμε σε έκταση στην § 2.7.

Η εξαγωγή συμπερασμάτων για τον πληθυσμό από τα ευρήματα τυχαίων δειγμάτων, ο έλεγχος υποθέσεων και ο έλεγχος της στατιστικής σημαντικότητας είναι ό,τι αναφέρουμε ως Επαγωγική Στατιστική.

Μια τρίτη μεγάλη προσφορά της Επαγωγικής Στατιστικής είναι η χρησιμοποίηση παραμετρικών και μη παραμετρικών τεχνικών για την ανάλυση δεδομένων. Το t - τεστ, που είναι παραμετρικό τεστ, χρησιμοποιείται για να δούμε αν οι διαφορές μεταξύ των μέσων τιμών δύο δειγμάτων είναι σημαντικές.

Υπάρχουν δύο τύποι του t - τεστ. Είναι το t - τεστ για ανεξάρτητους μέσους και το t - τεστ για συσχετιζόμενους μέσους. Στην πρώτη κατηγορία το t - τεστ χρησιμοποιείται για να συγκριθούν οι μέσοι όροι των αποτελεσμάτων δύο ανεξάρτητων ομάδων. Για παράδειγμα, αν δύο τμήματα της Α' τάξης του Γυμνασίου διδαχθούν για ένα τρίμηνο την ίδια ύλη Μαθηματικών αλλά με διαφορετικές μεθόδους διδασκαλίας και στο τέλος εξεταστούν στο ίδιο τεστ, τότε τα αποτελέσματα των δύο ομάδων μπορούν να συγκριθούν με το t - τεστ.

Το t - τεστ για συσχετιζόμενους μέσους χρησιμοποιείται για να συγκριθούν οι μέσοι όροι των αποτελεσμάτων της ίδιας ομάδας πριν και μετά τη «θεραπεία» (την εφαρμογή μιας νέας διδακτικής μεθόδου). Χρησιμοποιείται επίσης και όταν τα ίδια υποκείμενα λαμβάνουν δύο διαφορετικές «θεραπείες» (π.χ. διδακτικές μεθόδους).

Στην § 6.4 μιλήσαμε για την ανάλυση της διασποράς (ANOVA) και είπαμε ότι χρησιμοποιείται προκειμένου να διαπιστώσουμε αν υπάρχουν σημαντικές διαφορές μεταξύ των μέσων τιμών περισσοτέρων των δύο ομάδων.

Τέλος στην § 6.3 αναπτύξαμε τη χρήση του x^2 - τεστ (The Chi - Square test) για την ανάλυση δεδομένων τα οποία παρουσιάζονται σε κατηγορίες. Το x^2 - τεστ βασίζεται στη σύγκριση των πραγματικών και των αναμενόμενων συχνοτήτων που επιτυγχάνονται.

Υπάρχουν και πολλοί ακόμη τομείς της εκπαιδευτικής έρευνας στους οποίους προσφέρει τις υπηρεσίες της η Επαγωγική Στατιστική, αλλά πιστεύουμε ότι όσα αναφέραμε και αναπτύσσονται στις προηγούμενες παραγράφους είναι αρκετά για μια πρώτη μύηση στην ανάλυση δεδομένων με μεθόδους της Επαγωγικής Στατιστικής.

- ***Ερωτήσεις κατανόησης***

1. Τι εννοούμε με τη φράση: «Ανάλυση δεδομένων έρευνας»;
2. Πώς γίνεται ο έλεγχος των δεδομένων;
3. Πού βασίζεται η επιλογή κατάλληλων στατιστικών κριτηρίων;
4. Πώς βοηθά η Περιγραφική Στατιστική στην ανάλυση δεδομένων;
5. Γιατί είναι απαραίτητα τα μέτρα θέσης και διασποράς μιας κατανομής;
6. Τι ορίζουμε ως μέση τιμή $\bar{x}$ των τιμών $x_1, x_2, \ldots, x_\nu$ της μεταβλητής X;
7. Η μέση τιμή $\bar{x}$ σε αταξινόμητα δεδομένα υπολογίζεται από τον τύπο $\bar{x} = \ldots$, ενώ σε ταξινομημένα δεδομένα από τον τύπο $\bar{x} = \ldots$
8. Τι είναι η διάμεσος τιμή μιας κατανομής;
9. Τι είναι η τυπική απόκλιση μιας κατανομής;
10. Η μέση τιμή των αριθμών 3, 6, 7, α και 14 είναι 8. Η τυπική απόκλιση είναι:
    ☐ 6,5 ☐ 1,5 ☐ 3,74 ☐ 2
11. Να αντιστοιχίσετε κατάλληλα τα παρακάτω:

α) 5, 7, 8, 10, 13, 14 •	• 1.	$\bar{x} = 9$
β) 1, 2, 8, 9, 9, 25 •	• 2.	$\delta = 9$
γ) 1, 2, 9, 12, 12, 18 •	• 3.	Ε.Τ. = 9

12. Να γράψετε το γράμμα Θ ή το γράμμα Δ μπροστά από κάθε όρο, ανάλογα με το αν εκφράζει μέτρο θέσης ή μέτρο διασποράς.
    ☐ διάμεσος ☐ επικρατούσα τιμή ☐ διακύμανση

☐ τυπική απόκλιση ☐ μέση τιμή ☐ εύρος

13. Ως μέτρο ασυμμετρίας μιας κατανομής συχνοτήτων ορίζουμε τον αριθμό $\beta_1 = \ldots$, ενώ ως μέτρο κυρτότητας του Pearson τον αριθμό $\beta_2 = \ldots$

14. Τα μέτρα διασποράς είναι σημαντικά γιατί ....

15. Πώς βοηθά η Επαγωγική Στατιστική στην ανάλυση δεδομένων;

ΚΕΦΑΛΑΙΟ 8

Αποτελέσματα, συμπεράσματα και ανακοίνωση μιας έρευνας

8.1 Παρουσίαση των αποτελεσμάτων μιας έρευνας

Μετά την ανάλυση των δεδομένων με τη βοήθεια της Στατιστικής (Περιγραφικής και Επαγωγικής) και με τη χρήση σήμερα του ηλεκτρονικού υπολογιστή ο ερευνητής παρουσιάζει τα αποτελέσματα της έρευνας με κατάλληλο τρόπο (με πίνακες, με γραφικές ή άλλες παραστάσεις), έτσι ώστε να είναι κατανοητά ακόμη και από μη ειδικούς.

Η παρουσίαση των αποτελεσμάτων με πίνακες ή γραφικές παραστάσεις είναι σημαντική γιατί παρέχεται έτσι η δυνατότητα να βλέπουμε συγκεντρωμένα τα ευρήματα της έρευνας. Ειδικότερα, παρουσιάζονται οι διαφορές, οι ομοιότητες, οι σχέσεις και οι τάσεις των στοιχείων που παρουσιάζουν τα αποτελέσματα, πάνω στα οποία θα στηριχθεί ο ερευνητής για να διατυπώσει τα συμπεράσματά του ή να δώσει απαντήσεις στα ερευνητικά ερωτήματα ή να θεμελιώσει τη λύση του προβλήματος.

Οι πίνακες και οι διάφορες γραφικές παραστάσεις, όταν κατασκευάζονται προσεκτικά και δίνονται οι απαραίτητες επεξηγήσεις κατασκευής τους, διαβάζονται ευκολότερα και μεταδίδουν τις πληροφορίες συντομότερα και πιο αποτελεσματικά απ' ό,τι δίνονται με τα διάφορα κείμενα και τις εκθέσεις που περιγράφουν το περιεχόμενό τους. Όταν ο ερευνητής κατασκευάζει έναν πίνακα ή ένα διάγραμμα, πρέπει να έχει κατά νου τις υποθέσεις της έρευνας και ότι σκοπός του είναι με αυτές τις κατασκευές να κάνει τον έλεγχό τους. Η απλότητα στην κατασκευή ενός πίνακα είναι αναγκαία. Όταν ο αναγνώστης δυσκολεύεται να διαβάσει έναν πίνακα ή να κατανοήσει ένα διάγραμμα, τότε η παρουσίαση των αποτελεσμάτων με αυτόν τον τρόπο είναι αποτυχημένη. Ένας πίνακας καλά κατασκευασμένος δίνει όλες τις πληροφο-