ΚΕΦΑΛΑΙΟ 2
ΠΕΡΙΓΡΑΦΗ ΔΕΔΟΜΕΝΩΝ ΜΕ ΓΡΑΦΙΚΕΣ ΜΕΘΟΔΟΥΣ

2.1 Κατανομή Δεδομένων

Προκειμένου να χρησιμοποιήσει κανείς κατά τον καλύτερο δυνατό τρόπο τα διαθέσιμα δεδομένα χρειάζεται, σχεδόν πάντα, να συνοψίσει τα δεδομένα αυτά με κάποιο τρόπο. Υπάρχουν δύο βασικές μέθοδοι σύνοψης των δεδομένων. Μια μέθοδος είναι να παρουσιαστούν τα δεδομένα με κάποιο τρόπο από τη μικρότερη ως τη μεγαλύτερη τιμή τους με πίνακες ή γραφικές παραστάσεις. Ο τρόπος με τον οποίο ταξινομούνται τα δεδομένα ονομάζεται στη Στατιστική **κατανομή (distribution)** των δεδομένων αυτών.

Η κατανομή των δεδομένων είναι σημαντική γιατί αποκαλύπτει τρόπους με τους οποίους τα δεδομένα μεταβάλλονται. Ετσι μπορούμε να κατανοήσουμε καλύτερα την πηγή από την οποία προήλθαν τα υπό μελέτη δεδομένα.

Ο δεύτερος τρόπος συνοπτικής παρουσίασης των δεδομένων είναι η χρησιμοποίηση αριθμητικών ποσοτήτων που προέρχονται από αυτά, και που, όπως ήδη έχουμε πει, προκύπτουν ως τιμές συναρτήσεων που ονομάζονται στατιστικές συναρτήσεις, ή είναι τιμές παραμέτρων αν τα δεδομένα προέρχονται από ολόκληρο τον πληθυσμό. Για παράδειγμα, μπορούμε να χαρακτηρίσουμε την απόδοση μιας τάξης στις εξετάσεις ενός μαθήματος Στατιστικής δίνοντας μόνο τον μέσο βαθμό των φοιτητών.

Στο κεφάλαιο αυτό θα ασχοληθούμε με δυνατούς τρόπους παρουσίασης των δεδομένων με πίνακες και γραφικές παραστάσεις.

Στο επόμενο κεφάλαιο θα συζητήσουμε αριθμητικούς τρόπους παρουσίασης των δεδομένων.

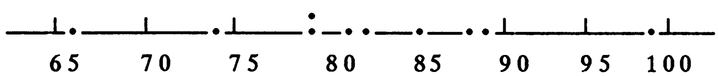
2.2 Διαγράμματα Σημείων (Dot Diagrams)

Οταν έχουμε ένα μικρό μόνο αριθμό παρατηρήσεων, η κατανομή τους μπορεί να περιγραφεί εύκολα με ένα **διάγραμμα σημείων (Dot Diagram)**. Ενα διάγραμμα σημείων είναι απλά η τοποθέτηση των διαθέσιμων τιμών πάνω σε ένα ευθύγραμμο τμήμα. Εάν υπάρχουν δύο ή περισσότερες τιμές οι οποίες συμπίπτουν, τοποθετούνται η μία πάνω στην άλλη.

Για παράδειγμα, έστω ότι στις εξετάσεις ενός μαθήματος Στατιστικής δέκα φοιτητές πήραν τους εξής βαθμούς:

66, 74, 78, 78, 80, 81, 85, 88, 89, 97

Για να κατασκευάσουμε ένα διάγραμμα σημείων χρειαζόμαστε ένα ευθύγραμμο τμήμα που να απεικονίζει τιμές από το 66 ως το 97. Στη συνέχεια σημειώνουμε τους δέκα διαθέσιμους βαθμούς στο ευθύγραμμο αυτό τμήμα χρησιμοποιώντας τελείες όπως φαίνεται στο σχήμα 2.2.1.



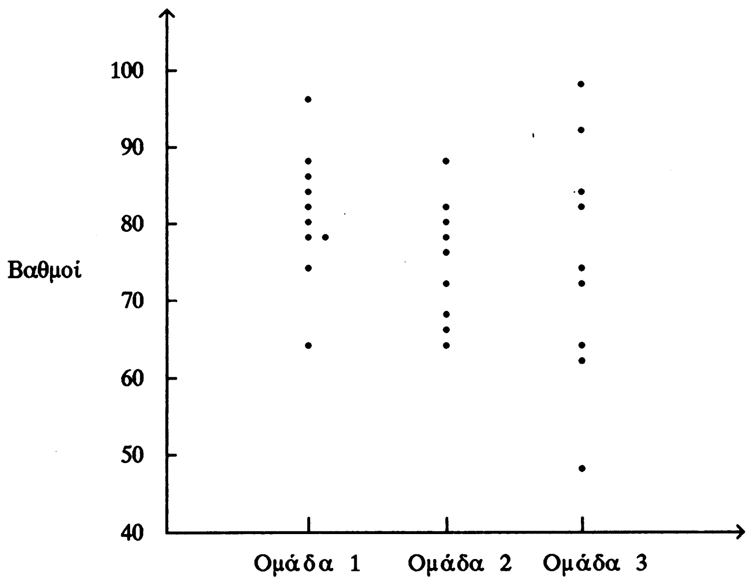
Σχήμα 2.2.1
Διάγραμμα σημείων των βαθμών 10 φοιτητών

Στο σχεδιασμό πειραμάτων πολλές φορές έχουμε έναν αριθμό από μικρά δείγματα που προέκυψαν κάτω από τις ίδιες συνθήκες. Στην περίπτωση αυτή, τα διαγράμματα σημείων αποτελούν έναν αποτελεσματικό τρόπο για να παρουσιασθούν πειραματικά αποτελέσματα, όπως θα δούμε και αργότερα.

Για παράδειγμα, έστω ότι ο ίδιος καθηγητής Στατιστικής έδωσε το ίδιο διαγώνισμα σε τρεις διαφορετικές ομάδες φοιτητών, κάθε μια από τις οποίες διδάχθηκε το μάθημα με διαφορετική μέθοδο. Τα αποτελέσματα που δόθηκαν ήδη αποτελούν τους βαθμούς της πρώτης ομάδας, ενώ οι βαθμοί για τις ομάδες 2 και 3 δίνονται στη συνέχεια.

Ομάδα 2	:	65	66	68	71	77	78	80	82	88
Ομάδα 3	:	48	61	62	71	75	82	84	91	99

Το σχήμα 2.2.2 που ακολουθεί μας δείχνει τη μεταβλητότητα των διαγραμμάτων σημείων για τις τρεις ομάδες φοιτητών στην ίδια γραφική παράσταση.



Σχήμα 2.2.2
Διαγράμματα σημείων για τις τρεις ομάδες φοιτητών

Συγκρίνοντας τα διαγράμματα σημείων και των τριών ομάδων φοιτητών όπως αυτά εμφανίζονται στο σχήμα 2.2.2 μπορούμε να βγάλουμε τα εξής προκαταρκτικά συμπεράσματα. Ο μέσος βαθμός των ομάδων 2 και 3 είναι, περίπου, ίδιος ενώ ο μέσος βαθμός της ομάδας

1 είναι υψηλότερος. Επίσης η μεταβλητότητα των βαθμών των φοιτητών της ομάδας 3 είναι μεγαλύτερη από όλες τις ομάδες ενώ αυτή της ομάδας 2 είναι η μικρότερη.

Εάν τα δεδομένα αποτελούνται από πολλές τιμές, τα διαγράμματα σημείων γίνονται πολύ πυκνά και όχι αποτελεσματικα. Στην περίπτωση αυτή, εάν επιθυμούμε να παρουσιάσουμε την κατανομή των τιμών, ο καλύτερος τρόπος είναι να τις τοποθετήσουμε σε ομάδες όμοιες τιμές δεδομένων. Αντί να κάνουμε το διάγραμμα των συγκεκριμένων τιμών, αναφέρουμε τις τιμές των δεδομένων όπως αυτές βρίσκονται σε κάθε ομάδα.

Ο κυριότερος τρόπος αυτής της μορφής παρουσίασης είναι τα διαγράμματα μίσχου - φύλλου (stem-and-leaf plots) που θα δούμε στη συνέχεια.

2.3 Διαγράμματα Μίσχου - Φύλλου (Stem-and-Leaf Plots)

Τα διαγράμματα μίσχου - φύλλου ή φυλλογράμματα (stem-and-leaf plots ή stem-and-leaf diagrams) είναι ένας πολύ απλός αλλά εξαιρετικά περιγραφικός τρόπος οργάνωσης και παρουσίασης δεδομένων με τρόπο που να περιγράφεται η κατανομή τους. Η μέθοδος αυτή περιγραφής δεδομένων οφείλεται στον Στατατιστικό John Tukey. Ο τρόπος αυτός παρουσίασης μπορεί να θεωρηθεί ως μία εναλλακτική μέθοδος για μια προκαταρκτική ανάλυση των στοιχείων της περιγραφής δεδομένων με ιστογράμματα, που θα δούμε στη συνέχεια. Πιο συγκεκριμένα αποτελεί ένα χρήσιμο πρώτο βήμα στην κατασκευή μιας κατανομής συχνότητας και ενός ιστογράμματος.

Ας θεωρήσουμε τους παρακάτω 20 βαθμούς που αναφέρονται στους βαθμούς των φοιτητών που παρακολούθησαν ένα συγκεκριμένο μάθημα Στατιστικής:

93 83 86 83 56 63 64 73 78 81 62 88 54 72 74 87 78 61 63 89

Παρατηρούμε ότι αν μας ενδιαφέρει να εξετάσουμε πως απέδωσαν οι φοιτητές στο διαγώνισμα αυτό, ακόμα και για μια μικρή τάξη 20 φοιτητών δεν είναι εύκολο να το κάνουμε αμέσως αν δεν χρησιμοποιήσουμε κάποια συγκεκριμένη μέθοδο παρουσίασης της κατανομής τους.

Ο τρόπος παρουσίασης της κατανομής των δεδομένων με ένα διάγραμμα μίσχου - φύλλου είναι ο εξής: χωρίζουμε κάθε αριθμό από τα δεδομένα σε δύο μέρη: το **μίσχο (stem)** και το **φύλλο (leaf)**. Στη συνέχεια κάνουμε μια κατακόρυφη γραμμή και αριστερά της τοποθετούμε το μίσχο (που στο συγκεκριμένο μας παράδειγμα αποτελείται από το ψηφίο των δεκάδων) και στα δεξιά της το φύλλο (που στο παράδειγμά μας αποτελείται από τις μονάδες των βαθμών εκείνων που έχουν ως δεκάδα τον αντίστοιχο μίσχο). Αυτό μας οδηγεί το σχήμα 2.3.1.

Μίσχος (stem)	**Φύλλο** (leaf)
5	6 4
6	3 4 2 1 3
7	3 8 2 4 8
8	3 6 3 1 8 7 9
9	3

Σχήμα 2.3.1

Διάγραμμα Μίσχου - Φύλλου των βαθμών 20 φοιτητών

Παρατηρούμε ότι η πρώτη τιμή μίσχου είναι το 5 (το ψηφίο των δεκάδων) και οι τιμές των φύλλων είναι 6 και 4 (τα ψηφία των μονάδων) που αντιστοιχούν στους βαθμούς 56 και 54. Η επόμενη τιμή μίσχου είναι 6 με τιμές φύλλων 3, 4, 2, 1 και 3 που αντιστοιχούν στους βαθμούς 63, 64, 62, 61 και 63. Η συνέχεια είναι προφανής.

Εν γένει, όλοι οι αριθμοί με την ίδια τιμή μίσχου τοποθετούνται στην ίδια γραμμή του διαγράμματος μίσχου-φύλλου με τιμές φύλλων τοποθετημένες στα δεξιά της κατακόρυφης γραμμής.

Για το συγκεκριμένο παράδειγμα που τα στοιχεία αποτελούνταν από διψήφιους αριθμούς, ο τρόπος χωρισμού των αριθμών σε μίσχους και φύλλα ήταν προφανής. Για στοιχεία που εκφράζονται με μεγαλύτερους αριθμούς είναι απαραίτητο να διαιρεθούν οι αριθμοί αυτοί με διαφορετικούς τρόπους. Για παράδειγμα, αν έχουμε τετραψήφιες παρατηρήσεις π.χ. τον αριθμό 1268. Κάποιος αναλυτής μπορεί να θεωρήσει ένα μίσχο με τον αριθμό 12 και ένα φύλλο με τον αριθμό 68. Εναλλακτικά, ο μίσχος μπορεί να αποτελείται από το ψηφία των χιλιάδων, οπότε θα είναι στην περίπτωσή μας ο αριθμός 1 και το φύλλο θα είναι ο αριθμός 268. Ο καθορισμός του μίσχου και του φύλλου είναι αυθαίρετος. Η επιλογή γίνεται έτσι ώστε τα δεδομένα παρουσιάζονται με τον καλύτερο τρόπο.

Το κατά πόσο τοποθετούμε τις τιμές των φύλλων σε κάθε γραμμή με τρόπο ώστε να αρχίζουν από τη μικρότερη και να καταλήγουν στη μεγαλύτερη ή να τις διατηρούμε με τη σειρά που καταγράφηκαν (όπως κάναμε στο παράδειγμά μας) είναι κάτι που εξαρτάται από την προτίμηση του ερευνητή. Είναι ίσως όμως περισσότερο πληροφοριακή η παρουσίαση όταν οι παρατηρήσεις τοποθετούνται με τη σειρά μεγέθους. Το μειονέκτημα σε κάτι τέτοιο είναι ότι, ίσως, γίνει λάθος κατά την τοποθέτηση των αριθμών.

Για μερικά σύνολα δεδομένων είναι ενδεχόμενο να οδηγηθούμε στο συμπέρασμα ότι υπάρχουν πάρα πολλά φύλλα σε σχέση με τον αριθμό των μίσχων. Για παράδειγμα, αν και οι 20 φοιτητές στο διαγώνισμα της Στατιστικής είχαν πάρει βαθμούς από 70 έως 90, η χρησιμοποίηση των ψηφίων των δεκάδων για τον καθορισμό των μίσχων θα μας έδινε μόνο δύο γραμμές. Για να πετύχουμε μια καλύτερη παρουσίαση της κατανομής των βαθμών, θα μπορούσαμε να χωρίσουμε τους βαθμούς με κοινό μίσχο σε δύο κατηγορίες: "υψηλή" και "χαμηλή". Ετσι, οι μίσχοι για τους βαθμούς στη δεκάδα του 80 θα μπορούσαν να είναι 8^+ για βαθμούς από 85 έως 89 και 8^- για βαθμούς από το 80 έως το 84. Το ίδιο για τους βαθμούς στη δεκάδα του 70 θα είχαμε 7^+ και 7^-. Με τον τρόπο αυτές θα είχαμε τέσσερις γραμμές βαθμών αντί για δύο γραμμές.

Σημείωση: Τα διαγράμματα μίσχου-φύλλου δεν είναι μόνο χρήσιμα γιατί δίνουν μια καλή εικόνα των δεδομένων που αποκαλύπτει πολλά από τα χαρακτηριστικά τους, αλλά και γιατί παρουσιάζουν τα δεδομένα αυτά με περισσότερο συνοπτικό τρόπο.

Για παράδειγμα, ο πίνακας που ακολουθεί αναφέρεται στα δρομολόγια τραίνων και στις αναχωρήσεις της Ιαπωνικής γραμμής Keihin Express Line με αφετηρία το σταθμό Yokohama. Ο πίνακας αυτός των δρομολογίων εμφανίζεται στο χρονοδιάγραμμα των δρομολογίων Sagami Tetsuto Company (έκδοση 1985, σελ.76). Οι διάφοροι κώδικες αναφέρονται στο είδος του τραίνου (υπερταχεία, επιβατηγό, κ.λ.π.) και σε διάφορες τοπικές στάσεις.

Πίνακας 2.3.1

Διάγραμμα μίσχου-φύλλου για τα δρομολόγια τραίνων

時	平日下り
5	06 18 31 40 46 58
6	04 12 18 21 30 38 41 49 55 59
7	03 08 14 17 23 26 30 35 38 40 45 47 49 54 56 58
8	03 06 09 18 20 22 28 30 32 38 40 42 50 52 54
9	00 02 04 10 12 14 20 22 24 29 31 33 41 43 50 53 57
10	01 03 07 11 12 17 20 22 26 29 34 37 40 45 49 54 57
11	00 05 08 12 17 19 25 28 32 37 39 45 48 52 57 59
12	05 08 12 17 19 25 28 32 37 39 45 48 52 57 59
13	05 08 12 17 19 25 28 32 37 39 45 48 52 57 59
14	05 08 12 17 19 25 28 32 37 39 45 48 52 57 59
15	05 08 12 17 19 25 28 32 37 39 45 48 52 57 59
16	05 08 09 16 18 21 27 29 32 38 40 42 48 50 52 59
17	01 04 10 12 14 19 22 24 26 30 32 34 36 40 43 45 47 53 55 57
18	01 03 05 07 13 15 17 21 23 25 28 33 35 37 41 43 45 48 53 55 57
19	01 04 06 08 13 15 17 20 23 25 27 32 34 36 40 43 45 47 51 53 55
20	00 02 04 10 12 14 19 21 23 30 32 34 39 41 46 50 52 58
21	01 06 09 11 18 21 26 29 31 38 41 46 50 51 58
22	01 09 11 17 21 29 32 39 44 51 53 59
23	04 10 14 21 30 36 47 54
24	03 15 21 23

○……特急
●……快特
▲……急行
黒字……各駅停車
無印……浦賀行
ハ……金沢八景行

上……上大岡行
ブ……金沢文庫行
み……三崎口行
ミ……三浦海岸行
ク……久里浜行
ズ……新逗子行

Στον πίνακα αυτό εμφανίζονται οι χρόνοι αναχωρήσεων για όλα τα 292 τραίνα που λειτουργούν καθημερινά, παρουσιάζοντας με χαρακτηριστικό τρόπο τις αιχμές λειτουργίας το πρωί και το βράδυ. Το διάγραμμα αυτό περιέχει 619 αριθμούς. Αντίθετα, ένα συνηθισμένο χρονοδιάγραμμα δρομολογίων που εμφανίζεται στα συνήθη χρονοδιαγράμματα δρομολογίων τραίνων άλλων χωρών περιέχει πολύ περισσότερα στοιχεία. Για παράδειγμα, το χρονοδιάγραμμα με τις ίδιες ακριβώς πληροφορίες αν εμφανιζόταν με το συνήθη τρόπο θα είχε τη μορφή του πίνακα 2.3.2.

Πίνακας 2.3.2

Δρομολόγια τραίνων

5.06	7.17	8.28	9.31	10.40	11.57	13.12	14.28
5.18	7.23	8.30	9.33	10.45	11.59	13.17	14.32
5.31	7.26	8.32	9.41	10.49	12.05	13.19	14.37
5.40	7.30	8.38	9.43	10.54	12.08	13.25	14.39
5.46	7.35	8.40	9.50	10.57	12.12	13.28	14.45
5.58	7.38	8.42	9.53	11.00	12.17	13.32	14.48
6.04	7.40	8.50	9.57	11.05	12.19	13.37	14.52
6.12	7.45	8.52	10.01	11.08	12.25	13.39	14.57
6.18	7.47	8.54	10.03	11.12	12.28	13.45	14.59
6.21	7.49	9.00	10.07	11.17	12.32	13.48	15.05
6.30	7.54	9.02	10.11	11.19	12.37	13.52	15.08
6.38	7.56	9.04	10.12	11.25	12.39	13.57	15.12
6.41	7.58	9.10	10.17	11.28	12.45	13.59	15.17
6.49	8.03	9.12	10.20	11.32	12.48	14.05	15.19
6.55	8.06	9.14	10.22	11.37	12.52	14.08	15.25
6.59	8.09	9.20	10.26	11.39	12.57	14.12	15.28
7.03	8.18	9.22	10.29	11.45	12.59	14.17	15.32
7.08	8.20	9.24	10.34	11.48	13.05	14.19	15.37
7.14	8.22	9.29	10.37	11.52	13.08	14.25	15.39

Πίνακας 2.3.2 (συνέχεια)

Δρομολόγια τραίνων

15.45	16.52	17.53	18.45	19.40	20.39	21.51	23.36
15.48	16.59	17.55	18.48	19.43	20.41	21.58	23.47
15.52	17.01	17.57	18.53	19.45	20.46	22.01	23.54
15.57	17.04	18.01	18.55	19.47	20.50	22.09	24.03
15.59	17.10	18.03	18.57	19.51	20.52	22.11	24.15
16.05	17.12	18.05	19.01	19.53	20.58	22.17	24.21
16.08	17.14	18.07	19.04	19.55	21.01	22.21	24.23
16.06	17.19	18.13	19.06	20.00	21.06	22.29	
16.16	17.22	18.15	19.08	20.02	21.09	22.32	
16.18	17.24	18.17	19.01	20.04	21.11	22.39	
16.21	17.26	18.21	19.01	20.10	21.18	22.44	
16.27	17.30	18.23	19.04	20.12	21.21	22.51	
16.29	17.32	18.25	19.06	20.14	21.26	22.53	
16.32	17.34	18.28	19.08	20.19	21.29	22.59	
16.38	17.36	18.33	19.13	20.21	21.31	23.04	
16.40	17.40	18.35	19.15	20.23	21.38	23.10	
16.42	17.43	18.37	19.17	20.30	21.41	23.14	
16.48	17.45	18.41	19.20	20.32	21.46	23.21	
16.50	17.47	18.43	19.23	20.34	21.50	23.30	

Αυτή η μορφή παρουσίασης περιέχει 1396 αριθμούς και τελείες. Επομένως, το διάγραμμα μίσχου-φύλλου, για τα δρομολόγια των τραίνων αυτών περιέχει 777 λιγότερους χαρακτήρες. Το πιο σημαντικό όμως είναι ότι δίνει μια πολύ καλύτερη εικόνα για τα δρομολόγια των τραίνων και τη συχνότητά τους, από αυτή που δίνει ο δεύτερος και συνήθης πίνακας.

Σε ένα διάγραμμα μίσχου-φύλλου όμως είναι δυνατόν να αποτυπωθούν ακόμα περισσότερες πληροφορίες. Αυτό γίνεται όταν τοποθετηθούν αριθμοί και από τις δύο πλευρές του κεντρικού μίσχου.

Ενα τέτοιο παράδειγμα αποτελεί ο πίνακας 2.3.3 που ακολουθεί. Ο πίνακας αυτός εμφανίζει τα δρομολόγια των τραίνων προς πολλές κατευθύνσεις από τον συγκεκριμένο σταθμό. Συγκεκριμένα αναφέρεται στις αναχωρήσεις τραίνων από τις πλατφόρμες 7 και 8 του σταθμού στο αριστερό μέρος του διαγράμματος και από τις πλατφόρμες 5 και 6 του σταθμού στο δεξί μέρος του διαγράμματος. Τα δύο μικρά βέλη αριστερά δείχνουν πώς οι αριθμοί αλλάζουν κατεύθυνση γύρω από μία γωνία όταν δεν υπάρχει χώρος να φιλοξενηθούν οι χρόνοι αναχώρησης τις πρωινές ώρες αιχμής στο διαθέσιμο πλαίσιο. (Ο πίνακας αυτός αναφέρεται στη γραμμή Tokaido στον ίδιο σταθμό της Ιαπωνίας, δηλαδή το σταθμό Yokohama και περιλαμβάνεται στο ίδιο βιβλίο πληροφοριών όπως και ο προηγούμενος στη σελίδα 72).

Πίνακας 2.3.3

Αναχωρήσεις τραίνων από διάφορες πλατφόρμες

横浜駅発東海道線

7・8番(品川・新橋・東京方面)	時	5・6番線(小田原・熱海・静岡方面)
59 09 4 54 25	5	25 47 4 52
56 51 47 38 34 20 09 04	6	14 31 54
52 49 45 42 38 35 31 27 21 16 12 04 00	7	07 16 32 36 41 51
47 43 39 35 31 27 24 20 17 13 10 06 03	8	03 13 25 27 34 41 48 58
52 57 40 36 28 14 05	9	05 12 25 27 35 46 59
58 56 47 37 32 24 19 10 06	10	10 19 25 34 49 55
56 43 34 30 21 16 08 05	11	04 19 25 29 39 49
59 41 36 23 09	12	04 19 25 34 49 55
56 45 35 28 10	13	07 19 34 45 55
49 36 31 26 09	14	04 19 34 43 50 55
51 40 31 20 09 04	15	03 19 32 36 49
56 51 40 32 30 19 05	16	01 23 35 49 59
56 44 35 33 19 12	17	03 14 17 23 29 34 42 48 55
56 51 36 25 06	18	01 07 13 20 26 29 36 39 47 57
59 53 49 41 30 13 08	19	04 12 15 22 24 29 39 49 59
53 37 24 09	20	09 18 29 39 42 50
48 34 18 05	21	03 13 24 27 34 45
43 22 06	22	04 19 34 49
51 13	23	06 13 18 44 54
07	0	23

特 L ○ △ ● ▲ ◆ 無印 品
寝台特急 特急 急行 土休日運転 休日運休 土休日運休 運転日に注意 東京行 品川行

特 L ○ △
寝台特急 特急 急行 土休日運転

● ▲ ◆ 沼 小 静 平 伊
休日運休 土休日運休 運転日に注意 沼津行 小田原行 静岡行 平塚行 伊東行

国 国府津行

下 修 兵 御
伊豆急下田行 修善寺行 浜松行 御殿場行

Μερικές φορές ο τρόπος αυτός παρουσίασης του διαγράμματος μίσχου-φύλλου ονομάζεται *διάγραμμα μίσχου-φύλλου διπλής όψης (back to back stem and leaf plot).* (Φυσικά, οι επιβάτες των ιαπωνικών τραίνων χρησιμοποιούν για δεκαετίες τώρα αυτές τις πληροφορίες χωρίς να έχουν ιδέα για το όνομα αυτό!)

2.4 Χρήση του MINITAB και του SAS για την κατασκευή διαγραμμάτων Μίσχου- Φύλλου

Χρήση του MINITAB

Οταν τα δεδομένα ενός προβλήματος έχουν εισαχθεί στον υπολογιστή είναι πολύ εύκολο να κατασκευασθεί το διάγραμμα μίσχου - φύλλου. Αυτό μπορεί να γίνει με την εντολή STEM-AND-LEAF, το όνομα της στήλης (column) που περιέχει τα δεδομένα και την υποεντολή INCREMENT. Η υποεντολή INCREMENT χρησιμοποιείται για να καθορίσει την επιθυμητή απόσταση μεταξύ των τιμών των μίσχων.

Για παράδειγμα, προκειμένου να κατασκευάσουμε το διάγραμμα μίσχου-φύλλου για τα δεδομένα των βαθμών των 20 φοιτητών δίνουμε τις εξής εντολές:

```
MTB>  SET C1
DATA> 93  83  86  83  56  63  64  73  78  81
      62  88  54  72  74  87  78  61  63  89
DATA> END
MTB>  STEM-AND-LEAF C1
SUBC> INCREMENT = 10
```

Το αποτέλεσμα των εντολών αυτών και η απάντηση του Minitab δίνονται στο σχήμα που ακολουθεί.

```
  2  5  46
  7  6  12334
(5)  7  23488
  8  8  1336789
  1  9  3
```

Οπως παρατηρούμε, το Minitab έχει διατάξει τις τιμές των φύλλων σε κάθε μίσχο σε αύξουσα σειρά.

Η πρώτη στήλη του Minitab αναφέρεται στις αθροιστικές συχνότητες για κάθε μίσχο ξεκινώντας από τους ακραίους μίσχους με κατεύθυνση τον κεντρικό μίσχο. Δηλαδή για κάθε μίσχο η στήλη αυτή δηλώνει τον αριθμό των παρατηρήσεων που αντιστοιχούν σε αυτόν και σε όλους τους προηγούμενους όπως κινούμεθα από τους ακραίους μίσχους προς τον κεντρικό μίσχο. Εξαίρεση αποτελεί ο κεντρικός μίσχος για τον οποίο δηλώνεται η απόλυτη συχνότητα σε παρένθεση. Ο κεντρικός μίσχος είναι ο μίσχος που περιέχει την διάμεσο του δείγματος. (Οπως θα δούμε αργότερα, η διάμεσος ενός δείγματος είναι η τιμή που χωρίζει τα δεδομένα του δείγματος περίπου στη μέση όταν αυτά τοποθετηθούν με σειρά τάξης μεγέθους). Ετσι, στο παράδειγμά μας, υπάρχουν 2 βαθμοί για τον πρώτο μίσχο, 7 βαθμοί για τους δύο πρώτους μίσχους και 5 βαθμοί στον κεντρικό μίσχο. Ξεκινώντας τώρα από τον ανώτερο (τελευταίο) μίσχο, έχουμε 1 βαθμό για αυτόν και 8 βαθμούς για τους μίσχους με αριθμό 8 και πάνω.

Οπως βλέπουμε, το διάγραμμα μίσχου-φύλλου αφ'ενός μεν μας δίνει την κατανομή των δεδομένων, αφ'ετέρου δε τα οργανώνει με τρόπο που να είναι διαθέσιμα για παρά πέρα ανάλυση.

1. Ενδειξη της κατανομής. Παρατηρώντας το διάγραμμα μίσχου και φύλλου του παραδείγματος βλέπουμε ότι οι βαθμοί κατανέμονται ομοιόμορφα μεταξύ των χαμηλών τιμών του 60 και των υψηλών τιμών του 80. Μόνο τρεις βαθμοί βρίσκονται έξω από την κλίμακα αυτή: ένας στα 90 και δύο στα 50.

2. Οργάνωση των δεδομένων. Το διάγραμμα μίσχου-φύλλου παρουσιάζει την κατανομή των δεδομένων χωρίς να χάνονται λεπτομέρειες από τα δεδομένα. Οποιαδήποτε στατιστική συνάρτηση μπορεί να υπολογιστεί από ένα διάγραμμα μίσχου-φύλλου δοθέντος ότι όλα τα δεδομένα δίνονται στο διάγραμμα.

Γενική παρατήρηση: Στο Minitab οι εντολές καθορίζονται πλήρως με τα τέσσερα μόνο πρώτα γράμματα της αντίστοιχης εντολής. Ετσι, για παράδειγμα, οι εντολές για το προηγούμενο παράδειγμά μας, δοθέντος ότι τα δεδομένα είχαν ήδη εισαχθεί στη στήλη C1, θα ήταν

```
MTB>   STEM C1;
SUBC>  INDR 10.
```

Χρήση του SAS

Προκειμένου να χρησιμοποιήσουμε το πακέτο SAS για να κατασκευάσουμε το διάγραμμα μίσχου-φύλλου, χρησιμοποιούμε την εντολή

```
PROC UNIVARIATE PLOT;
```

Εκτός από το διάγραμμα μίσχου-φύλλου η εντολή αυτή αυτόματα παρουσιάζει και αποτελέσματα που αναφέρονται σε διάφορα περιγραφικά μέτρα (ροπές, τεταρτημόρια κ.λ.π) όπως επίσης σε ακραίες τιμές των δεδομένων και διάγραμμα πλαισίου-απολήξεων (box plot). (Οι έννοιες αυτές θα συζητηθούν σε επόμενα κεφάλαια).

Για παράδειγμα, προκειμένου να δημιουργήσουμε ένα διάγραμμα μίσχου-φύλλου των δεδομένων του πίνακα 2.5.2 που αναφέρονται στη διάρκεια των υπεραστικών συνδιαλέξεων και, υποθέτοντας ότι τα δεδομένα έχουν τοποθετηθεί στη μεταβλητή X, χρησιμοποιούμε τις εξής εντολές:

```
PROC UNIVARIATE PLOT;
VAR X;
```

Η απάντηση του προγράμματος δίνει διάφορα άλλα στοιχεία στα οποία ανφερθήκαμε προηγουμένως, όπως επίσης το διάγραμμα μίσχου-φύλλου που μας ενδιαφέρει.

Stem	Leaf
22	12
21	01224
20	038
19	1123567789
18	0144577
17	236

2.5 Κατανομές Συχνότητας και Ιστογράμματα (Frequency Distributions and Histograms)

Ενας εναλλακτικός τρόπος, ίσως περισσότερο γνωστός, παρουσίασης των δεδομένων είναι η κατασκευή της κατανομής συχνότητας των δεδομένων (frequency distribution) και του αντιστοίχου ιστογράμματος (histogram).

Για να κατασκευάσουμε την κατανομή συχνότητας διαιρούμε το διάστημα που καλύπτουν οι διαθέσιμες τιμές των δεδομένων σε μια σειρά από *υποδιαστήματα (κατηγορίες, κλάσεις, τάξεις)*. Στη συνέχεια προσδιορίζουμε τον αριθμό των δεδομένων που περιέχονται σε κάθε μια κλάση. Για παράδειγμα, ας θεωρήσουμε τους 20 βαθμούς φοιτητών στο διαγώνισμα Στατιστικής. Οπως είδαμε, οι βαθμοί αυτοί βρίσκονται στο διάστημα από 54 ως 93. Εστω ότι διαιρούμε το διάστημα αυτό στις εξής κλάσεις:

50-59, 60-69, 70-79, 80-89, 90-99.

Μια **συχνότητα (frequency)** είναι ο αριθμός των τιμών δεδομένων που βρίσκονται σε μια δεδομένη κλάση. Δοθέντος ότι αυτές οι κλάσεις συμπίπτουν με το διάγραμμα μίσχου-φύλλου του σχήματος 2.3.1, ο αριθμός των τιμών των δεδομένων σε κάθε μια κλάση έχει ήδη καθορισθεί. Επομένως οι συχνότητες είναι 2, 5, 5, 7 και 1 αντίστοιχα για τις ορισθείσες κλάσεις.

Μερικές φορές είναι περισσότερο χρήσιμο να ξέρουμε το ποσοστό *(proportion)* των τιμών των δεδομένων που βρίσκεται σε κάθε μία κλάση αντί αυτού καθεαυτού του αριθμού. Μια **σχετική συχνότητα (relative frequency)** είναι το ποσοστό όλων των τιμών των δεδομένων που βρίσκονται σε μια δεδομένη κλάση. Οι συχνότητες και οι σχετικές συχνότητες για τους βαθμούς των 20 φοιτητών στο μάθημα της Στατιστικής δίνονται στον πίνακα 2.5.1.

Πίνακας 2.5.1

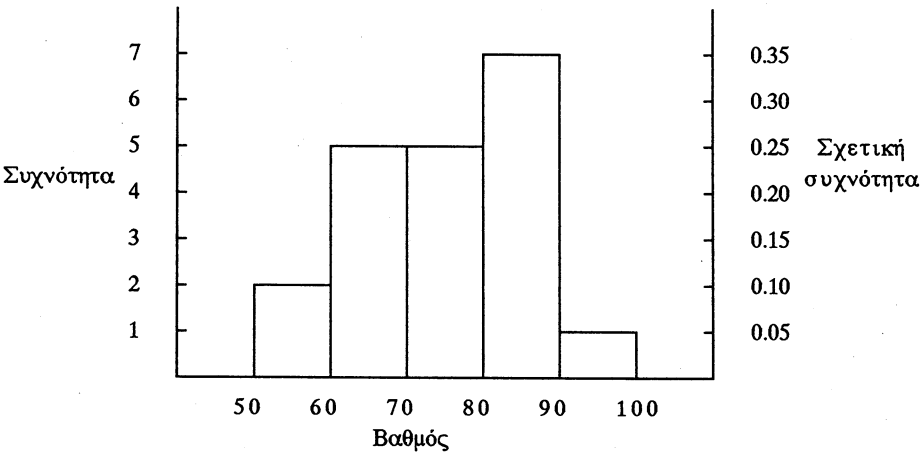
Συχνότητες και Σχετικές Συχνότητες των βαθμών των 20 φοιτητών στο διαγώνισμα Στατιστικής

Κλάση	Συχνότητα	Σχετική Συχνότητα
50-59	2	2/20 = 0.10
60-69	5	5/20 = 0.25
70-79	5	5/20 = 0.25
80-89	7	7/20 = 0.35
90-99	1	1/20 = 0.05
Σύνολο	20	1.00

Παρ'ότι η κατανομή συχνότητας ενός συνόλου δεδομένων παρουσιάζεται με ικανοποιητικό τρόπο με τη μορφή πίνακα, σε πολλές περιπτώσεις προτιμούμε να παρουσιάσουμε τη συχνότητα ή τη σχετική συχνότητα με τη μορφή μιας γραφικής παράστασης σε σχέση με τα διαστήματα των κλάσεων.

Μια τέτοια γραφική παράσταση ονομάζεται **ιστόγραμμα**.

Το ιστόγραμμα των συχνοτήτων και των σχετικών συχνοτήτων των βαθμών των εξετάσεων των 20 φοιτητών του πίνακα 2.5.5 δίνονται στο σχήμα 2.5.1.



Σχήμα 2.5.1
Ιστόγραμμα των βαθμών των 20 φοιτητών

Παρατηρούμε ότι τόσο η συχνότητα όσο και η σχετική συχνότητα για κάθε κλάση βρίσκονται στον κατακόρυφο άξονα ενώ οι κλάσεις βρίσκονται στον οριζόντιο άξονα. Επομένως, τόσο η συχνότητα όσο και η σχετική συχνότητα μπορούν να εκφρασθούν στο ίδιο ιστόγραμμα. Ο λόγος που πολλές φορές παρουσιάζεται μία προτίμηση στο ιστόγραμμα της σχετικής συχνότητας είναι ότι το εύρος των τιμών στον κατακόρυφο άξονα είναι πάντοτε καθορισμένο μεταξύ 0 και 1.

Παράδειγμα: Ως ένα άλλο παράδειγμα ας θεωρήσουμε τον αριθμό των υπεραστικών τηλεφωνημάτων που γίνονται από την κεντρική διοίκηση ενός επαρχιακού Πανεπιστημίου. Το Πανεπιστημίου αυτό ενδιαφέρεται να περιορίσει το κόστος που προέρχεται από υπεραστικά τηλεφωνήματα. Υπάρχουν διάφοροι τρόποι περιορισμού αυτού του κόστους. Ενας από αυτούς αναφέρεται στον περιορισμό των σύντομων τηλεφωνημάτων που γίνονται από το Πανεπιστήμιο προς το Υπουργείο Παιδείας για διάφορα ερωτήματα. Η Πρυτανεία του Πανεπιστημίου συνέστησε στους διοικητικούς υπαλλήλους να συγκεντρώνουν τα ερωτήματα προς το Υπουργείο έτσι ώστε να περιοριστούν τα σύντομα τηλεφωνήματα και να γίνονται τα ερωτήματα συνολικά σε ώρες που το κόστος της υπεραστικής τηλεφωνικής επικοινωνίας είναι χαμηλότερο.

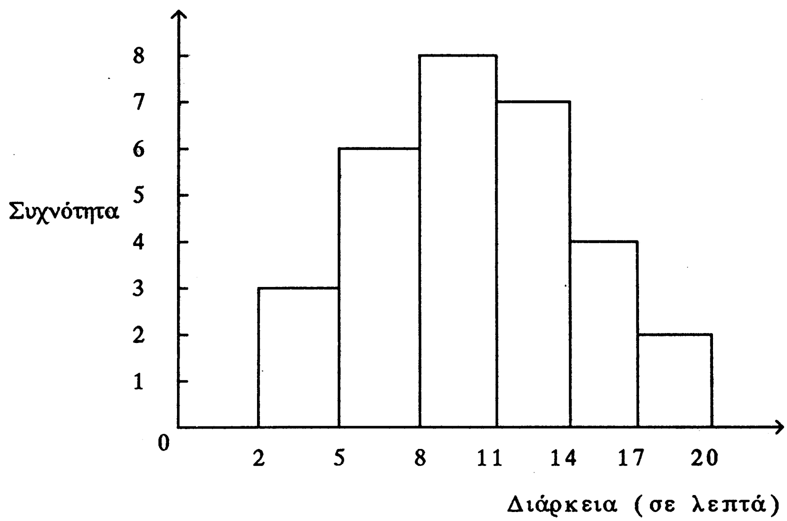
Στην προσπάθειά του αυτή, το Πανεπιστήμιο συγκέντρωσε ένα τυχαίο δείγμα 30 υπεραστικών τηλεφωνημάτων που έγιναν σε μια δεδομένη εβδομάδα. Οι χρόνοι που διήρκεσαν τα τηλεφωνήματα αυτά καταγράφονται στον πίνακα 2.5.2.

Πίνακας 2.5.2

Διάρκεια Υπεραστικών Συνδιαλέξεων (σε λεπτά)

11.8	3.6	16.6	13.5	4.8	8.3
8.9	9.1	7.7	2.3	12.1	6.1
10.2	8.0	11.4	6.8	9.6	19.5
15.3	12.3	8.5	15.9	18.7	11.7
6.2	11.2	10.4	7.2	5.5	14.5

Το ιστόγραμμα για τα δεδομένα αυτά φαίνεται στο σχήμα 2.5.2 που ακολουθεί.



Σχήμα 2.5.2

Ιστόγραμμα Διάρκειας υπεραστικών Συνδιαλέξεων

Οπως παρατηρούμε τόσο στο ιστόγραμμα αυτό όσο και στο ιστόγραμμα του προηγουμένου παραδείγματος, το εμβαδόν κάθε παραλληλογράμου συχνότητας είναι ανάλογο της συχνότητας αυτής.

Ο πίνακας 2.5.3 μας δίνει την κατανομή της σχετικής συχνότητας της διάρκειας των υπεραστικών τηλεφωνημάτων.

Πίνακας 2.5.3

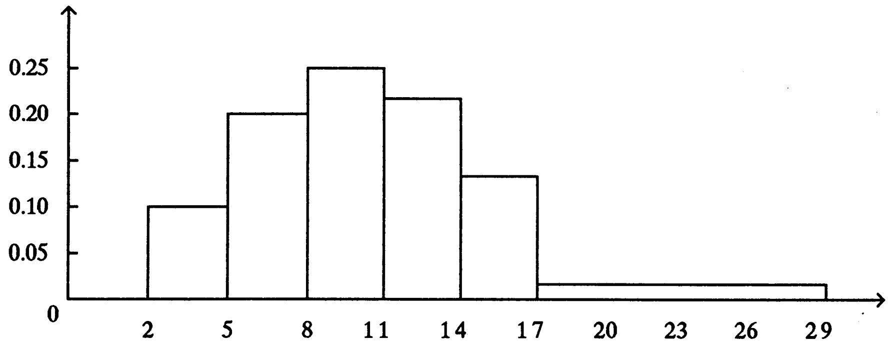
Κατανομή Σχετικής Συχνότητας
Διάρκειας Υπεραστικών Τηλεφωνημάτων

Ορια Κλάσεων	**Σχετική Συχνότητα**
2 μέχρι 5	3/30 = 0.10
5 μέχρι 8	6/30 = 0.20
8 μέχρι 11	8/30 = 0.267
11 μέχρι 14	7/30 = 0.233
14 μέχρι 17	4/30 = 0.133
17 μέχρι 20	2/30 = 0.067

Σημείωση: Οι τιμές που συμπίπτουν με το πάνω όριο μιας κλάσης δεν περιλαμβάνονται στην κλάση αυτή.

Οπως είπαμε και προηγουμένως, εν γένει, επιλέγουμε κλάσεις ίσης έκτασης (εύρους). Υπάρχουν όμως περιπτώσεις που είμαστε υποχρεωμένοι να χρησιμοποιήσουμε και κλάσεις διαφορετικού εύρους ώστε να αποφύγουμε να χρησιμοποιούμε πολλές κλάσεις με πολύ μικρό αριθμό συχνοτήτων. Για παράδειγμα, ας υποθέσουμε ότι στα δεδομένα του πίνακα 2.5.3 το 6.7% των υπεραστικών τηλεφωνημάτων είχαν διάρκεια μεταξύ 17 και 29 λεπτών και όχι μεταξύ 17 και 20 λεπτών όπως διαπιστώθηκε. Στην περίπτωση αυτή θα ήταν καλύτερα να χρησιμοποιήσουμε το ιστόγραμμα σχετικής συχνότητας που δίνεται στο σχήμα 2.5.3 όπου οι ακραίες προς τα πάνω κλάσεις έχουν συγχωνευθεί.

Είναι σημαντικό να παρατηρήσουμε ότι το ύψος του ορθογωνίου, που αντιστοιχεί σε μία σειρά από κλάσεις που έχουν συμπτυχθεί, πρέπει να προσαρμοσθεί με τρόπο ώστε το εμβαδόν του ορθογωνίου να είναι ανάλογο της σχετικής συχνότητας όλων των μετρήσεων που βρίσκονται μεταξύ 17 και 29.



Σχήμα 2.5.3
Ιστόγραμμα Σχετικής Συχνότητας με ανισομήκεις κλάσεις

Οι μετρήσεις και στα δύο προηγούμενα παραδείγματα βρίσκονται μέσα σε ένα σαφώς καθορισμένο πλαίσιο. Υπάρχουν όμως περιπτώσεις που εμφανίζονται μετρήσεις με πολύ ακραίες τιμές στο ένα ή στο άλλο άκρο των παρατηρήσεων. Οταν, για παράδειγμα, έχουμε μετρήσεις που να αναφέρονται σε εισόδημα οικογενειών είναι δυνατόν να έχουμε ορισμένες παρατηρήσεις με πολύ υψηλές τιμές. Σε τέτοιες περιπτώσεις χρησιμοποιούμε **κλάσεις ανοικτού εύρους (open-ended classes)** ώστε να παίρνουμε υπόψη μας τέτοιες ακραίες μετρήσεις.

Τα βασικά σημεία τα οποία θα πρέπει να προσέχουμε όταν κατασκευάζουμε κατανομές συχνότητας είναι τα εξής:

1. Οι κλάσεις που επιλέγουμε θα πρέπει να είναι τέτοιες που να δίνουν τη σωστή εικόνα της κατανομής των δεδομένων. Ο καθορισμός του αριθμού των κλάσεων είναι εν γένει αυθαίρετος. Συνήθως χρησιμοποιούμε από πέντε ως είκοσι κλάσεις. Οσο μεγαλύτερη

ποσότητα δεδομένων είναι διαθέσιμη τόσο περισσότερες κλάσεις θα πρέπει να χρησιμοποιούνται. Αυτό γιατί εάν ο αριθμός των κλάσεων που θα χρησιμοποιηθούν είναι πολύ μικρός είναι ενδεχόμενο να αποκρυβούν σημαντικά χαρακτηριστικά των δεδομένων με την ομαδοποίησή τους. Από το άλλο μέρος, εάν ο αριθμός των κλάσεων είναι μεγάλος σε σχέση με τα δεδομένα θα έχουμε πολλές κλάσεις που θα είναι ή κενές ή με μικρό αριθμό παρατηρήσεων και η κατανομή που θα εμφανίζουν θα δίνει μια όχι ικανοποιητική περιγραφή των δεδομένων.

2. Αφού δούμε το εύρος των τιμών του δείγματος θα πρέπει να καθορίσουμε το εύρος κάθε κλάσης (class width). Ως ένα γενικό κανόνα για το εύρος της κλάσης διαιρούμε τη διαφορά της μικρότερης από τη μεγαλύτερη μέτρηση με τον επιθυμητό αριθμό των κλάσεων που θέλουμε να χρησιμοποιήσουμε. Κατά κανόνα θα πρέπει να προσπαθούμε να έχουμε κλάσεις ίσου εύρους εκτός και αν υπάρχουν λόγοι όπως αυτοί που αναφέρθηκαν στο παράδειγμα.

3. Καθορισμός των ορίων των κλάσεων: Είναι προφανές ότι τα όρια αυτά θα πρέπει να καθορίζονται με τρόπο ώστε οι μετρήσεις να κατανέμονται σε μια μόνο από τις δυνατές κατηγορίες.

2.6 Χρήση του MINITAB και του SAS για την κατασκευή ιστογραμμάτω

Χρήση του MINITAB

Για την κατασκευή ιστογραμμάτων με τη χρήση του Minitab απαιτείται η εντολή HISTOGRAM μαζί με τις υποεντολές INCREMENT και START. Η υποεντολή START (αρχή) καθορίζει το μεσαίο σημείο της πρώτης κλάσης ενώ η υποεντολή INCREMENT (προσαύξηση) καθορίζει την απόσταση μεταξύ των μέσων σημείων (midpoints) των κλάσεων. Θα πρέπει να επισημάνουμε και πάλι ότι η τελευταία εντολή τελειώνει με τελεία (.).

Για να κατασκευάσουμε το ιστόγραμμα των βαθμών των 20 φοιτητών στο μάθημα της Στατιστικής, γράφουμε τις εξής εντολές:

```
MTB> histogram C1;
SUBC> increment = 10;
SUBC> start = 45.
```

Η απάντηση του Minitab, με δεδομένο ότι έχουμε τοποθετήσει τα δεδομένα στη στήλη 1 με το όνομα grades είναι η εξής:

```
Histogram of grades N = 20
Midpoint          Count
  45.0              0
  55.0              2   **
  65.0              5   *****
  75.0              5   *****
  85.0              7   *******
  95.0              1   *
```

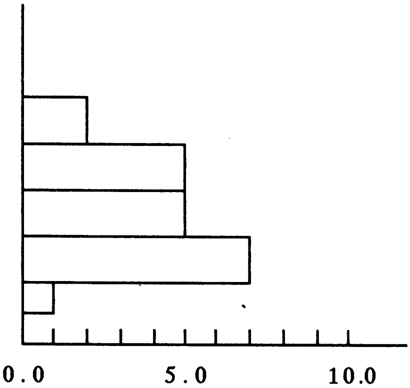
Οπως παρατηρούμε, το Minitab δίνει το ιστόγραμμα τοποθετημένο με πλάγιο τρόπο.

Σημείωση: Εάν θέλουμε να κατασκευάσουμε ένα ιστόγραμμα με τη μορφή ορθογωνίων, χρησιμοποιούμε την εντολή GHISTOGRAM αντί της εντολής HISTOGRAM.

Με την εντολή GHISTOGRAM για το παράδειγμα με τους βαθμούς των 20 φοιτητών, θα πάρουμε ως απάντηση το σχήμα 2.6.1 που ακολουθεί.

grades N = 20

Midpoint	Count
45.0	0
55.0	2
65.0	5
75.0	5
85.0	7
95.0	1



Σχήμα 2.6.1

Στο ίδιο αποτέλεσμα μπορούμε να καταλήξουμε χρησιμοποιώντας την εντολή

HISTOGRAM C1 K1 K2

όπου C1 είναι η στήλη 1 των παρατηρήσεων που έχουν ήδη αποθηκευθεί, K1 είναι το ενδιάμεσο σημείο την πρώτης κλάσης και K2 είναι το εύρος των κλάσεων.

Για παράδειγμα, για να φτάξουμε το ιστόγραμμα των δεδομένων του πίνακα 2.5.2 της διάρκειας των υπεραστικών συνδιαλέξεων γράφουμε τις εξής εντολές:

HISTOGRAM C1 K1 K2

SET C1

11.8	8.9	10.2	15.3	6.2
3.6	9.1	8.0	12.3	11.2
16.6	7.7	11.4	8.5	10.4
13.5	2.3	6.8	15.9	7.2
4.8	12.1	9.6	18.7	5.5
8.3	6.1	19.5	11.7	14.5

END

HISTOGRAM C1 3.5 3

Οι εντολές αυτές δίνουν το ακόλουθο αποτέλεσμα.

```
Histogram of C1 N = 30
Midpoint          Count
  3.50              3  * * *
  6.50              6  * * * * * *
  9.50              8  * * * * * * * *
 12.50              7  * * * * * * *
 15.50              4  * * * *
 18.50              2  * *
```

Σημείωση: Με παρόμοιες εντολές μπορούμε να κατασκευάσουμε ένα διάγραμμα σημείων (dotplot). Η εντολή είναι

MTB> dotplot C1

Η εντολή αυτή κατασκευάζει ένα διάγραμμα σημείων για τους 20 βαθμούς που έχουμε στη μεταβλητή C1.

Και στην περίπτωση αυτή μπορούμε να χρησιμοποιήσουμε τις υποεντολές INCREMENT και START αν θέλουμε να κάνουμε κάποιες μεταβολές. Στην περίπτωση αυτή η εντολή INCREMENT καθορίζει την επιθυμητή απόσταση μεταξύ δύο σημείων (που απεικονίζονται με το σύμβολο +), ενώ η εντολή START καθορίζει το πρώτο σημείο (τον πρώτο σταυρό).

Χρήση του SAS

Προκειμένου να κατασκευάσουμε ένα ιστόγραμμα (το οποίο στο SAS λέγεται frequency bar chart) και όταν τα δεδομένα των μετρήσεων πρέπει να ομαδοποιηθούν σε κλάσεις σύμφωνα με τις τιμές μιας μεταβλητής X γράφουμε τις εντολές

PROC CHART;
VBAR X/AXIS=a TO b BY c MIDPOINTS=ML TO MU BY W;

Οι τιμές στον κατακόρυφο άξονα (vertical axis) (VBAR) έχουν εύρος από το a έως το b με αύξηση μεγέθους c. Τα μεσαία σημεία των κλάσεων στη μεταβλητή X έχουν εύρος από το ML έως το MU (midpoint lower to midpoint upper) και εύρος κλάσης W (width).

Για παράδειγμα, προκειμένου να κατασκευάσουμε το ιστόγραμμα των δεδομένων του πίνακα 2.5.2 που αναφέρονται στη διάρκεια υπεραστικών συνδιαλέξεων υποθέτωντας ότι τα δεδομένα έχουν τοποθετηθεί ως μεταβλητή X (σύμφωνα με όσα έχουμε πει προηγούμενα) θα δώσουμε τις εξής εντολές

```
PROC CHART;
VBAR X/AXIS=0 TO 8 BY 1 MIDPOINTS=3.5 TO 18.5 BY 3;
```

Οι εντολές αυτές θα οδηγήσουν στην εξής απάντηση.

```
   |
 8 -                *****
   |                *****
 7 -                *****   *****
   |                *****   *****
 6 -        *****   *****   *****
   |        *****   *****   *****
 5 -        *****   *****   *****
   |        *****   *****   *****
 4 -        *****   *****   *****   *****
   |        *****   *****   *****   *****
 3 -  ***** *****   *****   *****   *****
   |  ***** *****   *****   *****   *****
 2 -  ***** *****   *****   *****   *****   *****
   |  ***** *****   *****   *****   *****   *****
 1 -  ***** *****   *****   *****   *****   *****
   |  ***** *****   *****   *****   *****   *****
   ---------------------------------------------------
       3.5   6.5     9.5    12.5    15.5    18.5
```

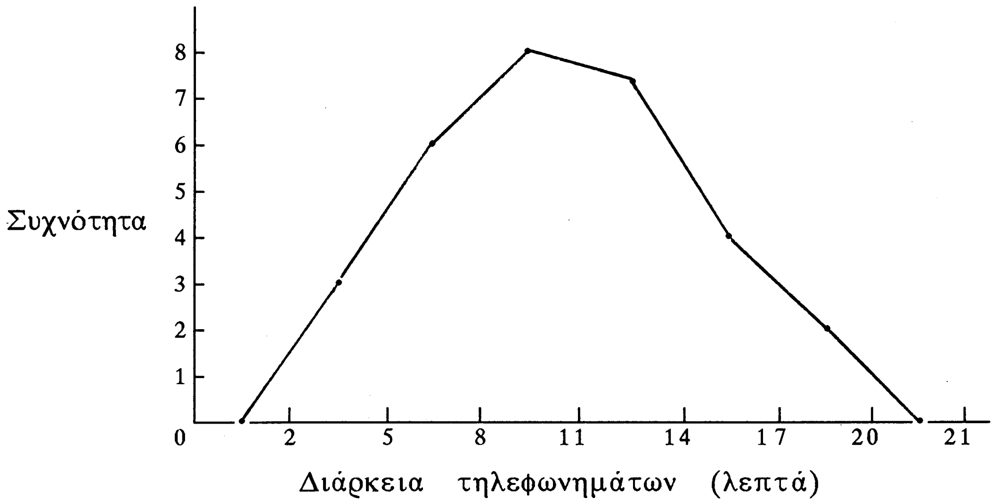
Σχήμα 2.6.2

Σημείωση: Εχουμε κάνει την παραδοχή ότι η κλάση από το 6.5 μέχρι το 8.0 δεν περιλαμβάνει την τιμή 8.0, ενώ, σύμφωνα με τη λογική του SAS, η κλάση αυτή περιλαμβάνει την τιμή 8.0. Επομένως, προκειμένου το SAS να μας δώσει ακριβώς το ίδιο ιστόγραμμα όπως αυτό που θα κάναμε με το χέρι θα πρέπει να χρησιμοποιήσουμε ως ενδιάμεσα σημεία των κλάσεων τα 3.49 ως 18.49.

2.7 Πολύγωνα συχνότητας (Frequency Polygons)

Ενας άλλος τρόπος παρουσίασης κατανομών συχνότητας με γράφημα είναι το **πολύγωνο συχνότητας (frequency polygon)**. Το πολύγωνο συχνότητας κατασκευάζεται ως εξής: Σημειώνουμε τη συχνότητα κάθε κλάσης πάνω από το μέσο σημείο της κλάσης και στη συνέχεια ενώνουμε τα διαδοχικά σημεία που προκύπτουν με ευθύγραμμα τμήματα. Συνήθως το πολύγωνο αυτό "κλείνει" με την θεώρηση μιας πρόσθετης θεωρητικής κλάσης (με συχνότητα 0) σε κάθε ένα από τα άκρα του διαστήματος των τιμών της κατανομής και την προσθήκη, στη συνέχεια, στην τεθλασμένη γραμμή δύο ευθυγράμμων τμημάτων που συνδέουν τα άκρα της τεθλασμένης με τα ενδιάμεσα σημεία των κλάσεων αυτών.

Το πολύγωνο συχνότητας για τη διάρκεια των υπεραστικών τηλεφωνημάτων του πίνακα 2.5.2. δίνεται στο σχήμα 2.7.1 που ακολουθεί.



Σχήμα 2.7.1

Πολύγωνο Συχνότητας της διάρκειας υπεραστικών τηλεφωνημάτων

Τα πολύγωνα συχνότητας είναι χρήσιμα για να μας δίνουν μια γενική ιδέα για τη μορφή της κατανομής.

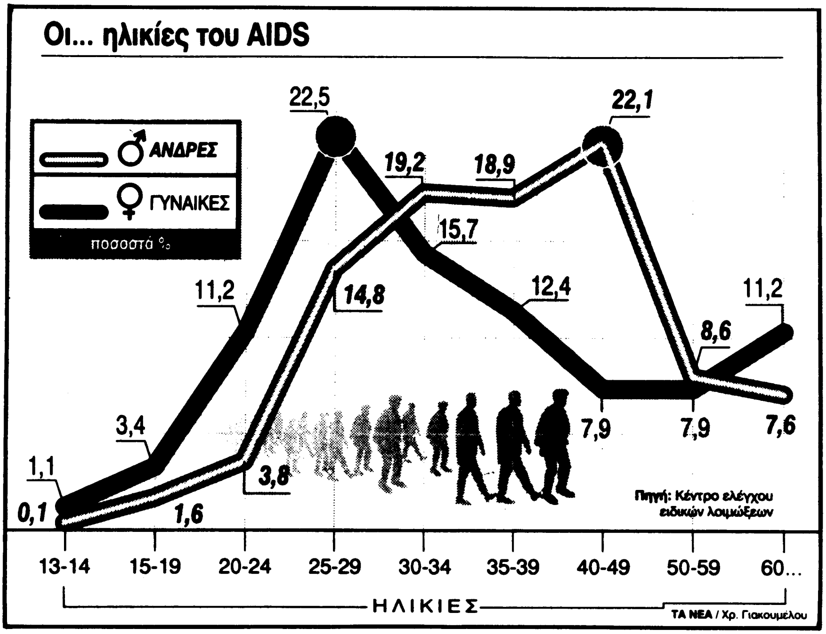
Οπως είδαμε στα ιστογράμματα, αντί για συχνότητες μιας κατανομής δεδομένων μπορούμε να μετρήσουμε τις σχετικές συχνότητες. Με τον ίδιο τρόπο μπορούμε να κατασκευάσουμε ένα **πολύγωνο σχετικής συχνότητας (relative frequency polygon)**. Τα πολύγωνα σχετικής συχνότητας επιτρέπουν μια οπτική σύγκριση δύο κατανομών με την τοποθέτηση ενός πολυγώνου σχετικής συχνότητας πάνω σε ένα άλλο.

Για παράδειγμα, το πολύγωνο Α (συνεχής καμπύλη) στο σχήμα 2.7.2 που ακολουθεί είναι το πολύγωνο σχετικής συχνότητας της διάρκειας των τηλεφωνημάτων του πίνακα 2.5.2. Το πολύγωνο Β (διακεκομμένη καμπύλη) θα μπορούσε να παριστάνει, για παράδειγμα την κατανομή της διάρκειας των τηλεφωνημάτων πριν εφαρμοσθεί το νέο πρόγραμμα περιορισμού των εξόδων. Η διαφορά στη μορφή των δύο κατανομών προκύπτει με προφανή τρόπο από το γράφημα αυτό.

Σχήμα 2.7.2

Σύγκριση Πολυγώνων Σχετικών Συχνοτήτων

Ενα δεύτερο παράδειγμα αποτελεί το πολύγωνο συχνότητας που δημοσίευσε η εφημερίδα ΤΑ ΝΕΑ στις 4 Νοεμβρίου 1993 (σχήμα 2.7.2α) και απεικονίζει τις κατανομές της σχετικής συχνότητας κρουσμάτων AIDS ανάλογα με τις ηλικίες για άνδρες και γυναίκες για το σύνολο των κρουσμάτων που εμφανίσθηκαν στην Ελλάδα τους 9 πρώτους μήνες του 1993.



Σχήμα 2.7.2α

Κατανομή σχετικών συχνοτήτων κρουσμάτων AIDS κατά ηλικία και φύλλο (ΤΑ ΝΕΑ, 4 Νοεμβρίου 1993)

Υπάρχουν μερικές περιπτώσεις που διευκολύνει καλύτερα να κατασκευάσουμε το **πολύγωνο αθροιστικής σχετικής συχνότητας (cumulative relative frequency polygon)** το οποίο συνήθως ονομάζεται **ακιδωτό (ogive)**.

Η αθροιστική σχετική συχνότητα μιας συγκεκριμένης κλάσης είναι το ποσοστό των μετρήσεων που είναι μικρότερες από το πάνω άκρο της κλάσης αυτής. Επομένως, η αθροιστική σχετική συχνότητα μιας κλάσης βρίσκεται με άθροιση των σχετικών συχνοτήτων της κλάσης αυτής και των σχετικών συχνοτήτων όλων των προηγουμένων κλάσεων. Ο πίνακας 2.7.1 που ακολουθεί δίνει την αθροιστική σχετική συχνότητα της διάρκειας των τηλεφωνημάτων του πίνακα 2.5.2 όπως αυτές προκύπτουν από την κατανομή της σχετικής συχνότητας του πίνακα 2.5.3.

Πίνακας 2.7.1

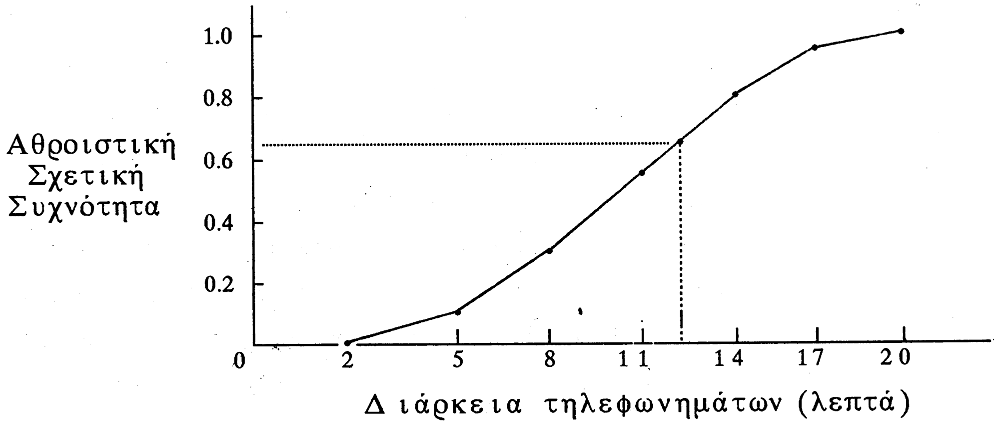
Κατανομή Αθροιστικής Σχετικής Συχνότητας της Διάρκειας των Τηλεφωνημάτων

Ορια κλάσεων	**Αθροιστική Σχετική Συχνότητα**
2 έως 5	3/30 = 0.10
5 έως 8	9/30 = 0.30
8 έως 11	17/30 = 0.567
11 έως 14	24/30 = 0.80
14 έως 17	28/30 = 0.933
17 έως 20	30/30 = 1.00

Από την αθροιστική σχετική συχνότητα του πίνακα 2.7.1 παρατηρούμε ότι, για παράδειγμα, το 30% των τηλεφωνημάτων είχαν διάρκεια μικρότερη των 8 λεπτών. Επειδή δε το 80% των τηλεφωνημάτων ήταν συντομότερα των 14 λεπτών, μπορούμε να συμπεράνουμε ότι το 20% των τηλεφωνημάτων υπερέβησαν τα 14 λεπτά. Αυτού του είδους οι πληροφορίες μπορούν να ληφθούν από το γράφημα της αθροιστικής σχετικής συχνότητας, δηλαδή το ακιδωτό.

Η κατασκευή του ακιδωτού διαφέρει από αυτή του πολυγώνου της σχετικής συχνότητας: Η αθροιστική σχετική συχνότητα κάθε κλάσης σημειώνεται πάνω από το άνω άκρο της αντίστοιχης κλάσης και τα διαδοχικά σημεία που παριστάνουν τις αθροιστικές σχετικές συχνότητες ενώνονται στη συνέχεια με ευθύγραμμα τμήματα. Το ακιδωτό "δένεται" στον οριζόντιο άξονα με ένα ευθύγραμμο τμήμα που συνδέει το κάτω άκρο της χαμηλότερης κλάσης με το σημείο που παριστάνει την αθροιστική σχετική συχνότητά της.

Το ακιδωτό που αντιστοιχεί στην κατανομή αθροιστικής συχνότητας του πίνακα 2.7.1 δίνεται στο σχήμα 2.7.3.



Σχήμα 2.7.3

Ακιδωτό (Ogive) Διάρκειας Τηλεφωνημάτων

Από το ακιδωτό μιας αθροιστικής σχετικής συχνότητας είναι δυνατόν να προσδιορισθεί κατά προσέγγιση το ποσοστό των μετρήσεων που είναι μικρότερες από οποιαδήποτε τιμή του οριζόντιου άξονα. Ετσι από το ακιδωτό του σχήματος 2.7.3 μπορούμε να εκτιμήσουμε ότι το ποσοστό των τηλεφωνημάτων που ήταν συντομότερα από 12 λεπτά είναι περίπου 0.65.

2.8 Κυκλικά Διαγράμματα, Ραβδογράμματα, Εικονογράμματα και Γραμμογραφήματα (Pie Charts, Bar Charts and Line Charts)

Στην ενότητα αυτή θα αναφερθούμε σε γραφικές παραστάσεις που χρησιμοποιούνται κυρίως για να αποτυπώσουν γραφικά *ποιοτικά δεδομένα.* (Αυτό βέβαια δεν αποκλείει τη χρησιμοποίησή τους και για ποσοτικά δεδομένα). Η μεγάλη αύξηση του αριθμού των μικρουπολογιστών που είναι διαθέσιμοι σε εταιρείες και επιχειρήσεις έχει επιτρέψει σε μια μεγάλη κατηγορία ανθρώπων να παρουσιάζουν με γραφικές μεθόδους πολλά από τα δεδομένα που έχουν διαθέσιμα. Παρ'ότι αυτού του είδους οι γραφικές παραστάσεις είναι αναρίθμητες, θα παρουσιάσουμε εκείνες που χρησιμοποιούνται πιο συχνά.

Οπως είδαμε στις προηγούμενες ενότητες, εάν τα δεδομένα είναι ποσοτικά και προέρχονται από ένα δεδομένο πληθυσμό, οι περιγραφικές μέθοδοι που περιγράψαμε είναι χρήσιμες και κατάλληλες. Αυτό που οι μέθοδοι εκείνες ουσιαστικά κάνουν είναι να χωρίζουν τα δεδομένα σε κατηγορίες και να απαριθμούν τις μετρήσεις που αντιστοιχούν σε κάθε κατηγορία. Οι κατηγορίες ορίζονται με ένα μάλλον αυθαίρετο τρόπο με αντικειμένικο σκοπό να περιγράψουν με κάποιο τρόπο πώς κατανέμονται τα δεδομένα.

Στην περίπτωση των ποιοτικών δεδομένων, η ταξινόμηση σε κατηγορίες μπορεί να γίνει με φυσικό και λιγότερο αυθαίρετο τρόπο.

Κυκλικά Διαγράμματα (Pie Charts)

Το **κυκλικό διάγραμμα (pie chart)** είναι απλά ένας κύκλος που υποδιαιρείται σε κυκλικούς τομείς ("φέτες") **(slices)** που αντιπροσωπεύουν διάφορες κατηγορίες.

Για παράδειγμα, μια αθλητική εφημερίδα ρώτησε τη γνώμη 250 ατόμων ειδικών γύρω από τα ποδοσφαιρικά όσο αφορά το ποιά ομάδα θα κερδίσει το ποδοσφαιρικό πρωτάθλημα Α' Εθνικής κατηγορίας την επόμενη χρονιά. Οι απαντήσεις που δόθηκαν καταγράφονται στον πίνακα 2.8.1.

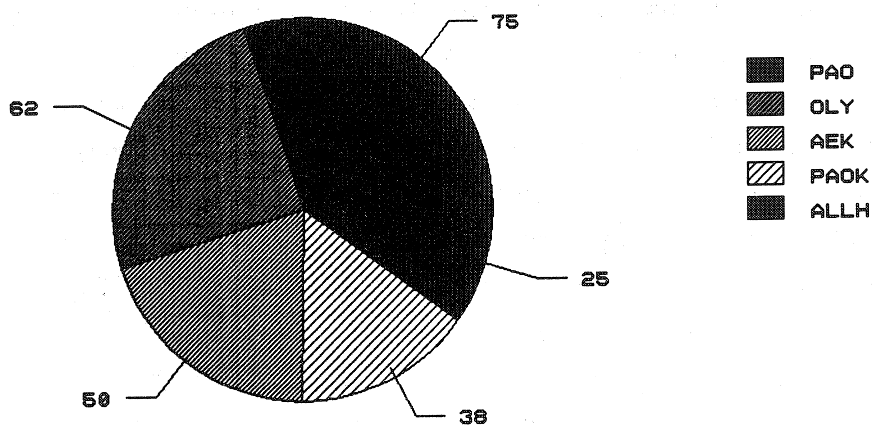
Τα δεδομένα στο παράδειγμα αυτό είναι ποιοτικά. Η εφημερίδα θεωρεί ότι θα ενδιαφέρει περισσότερο τους αναγνώστες της να έχουν τα ποσοστά παρά τους απόλυτους αριθμούς κάθε κατηγορίας ιδιαίτερα γιατί ενδιαφέρεται να τους δώσει τη δυνατότητα να κάνουν συγκρίσεις με παρόμοια ρεπορτάζ προηγουμένων χρόνων. (Ενας πρόσθετος λόγος για κάτι τέτοιο είναι ότι επιτρέπει τη σύγκριση γνωμών που προέρχονται από διαφορετικούς αριθμούς ερωτηθέντων).

Πίνακας 2.8.1

Πιθανή Πρωταθλήτρια Ποδοσφαίρου

Ομάδα	Αριθμός Υποστηρικτών	Ποσοστό Υποστηρικτών
ΠΑΟ	75	30%
Ολυμπιακός	62	25%
ΑΕΚ	50	20%
ΠΑΟΚ	38	15%
Αλλη	25	10%
Σύνολο	250	100%

Το κυκλικό διάγραμμα για τα δεδομένα του πίνακα 2.8.1 δίνεται στη συνέχεια.



Σχήμα 2.8.1

Πιθανή Πρωταθλήτρια Ποδοσφαίρου

Θα πρέπει να επισημανθεί ότι το εμβαδόν κάθε "φέτας" είναι ανάλογο του ποσοστού που αντιστοιχεί στη συγκεκριμένη κατηγορία. Δοθέντος ότι ο κύκλος έχει 360 μοίρες, κάθε 1% των παρατηρήσεων θα πρέπει να αντιστοιχεί σε $(0.01)(360)=3.6^\circ$. Για παράδειγμα, η γωνία που αντιστοιχεί στους ερωτηθέντες που θεωρούν ότι πρωταθλητής θα είναι ο Παναθηναικός, θα έχει άνοιγμα $30(3.6)=108^\circ$.

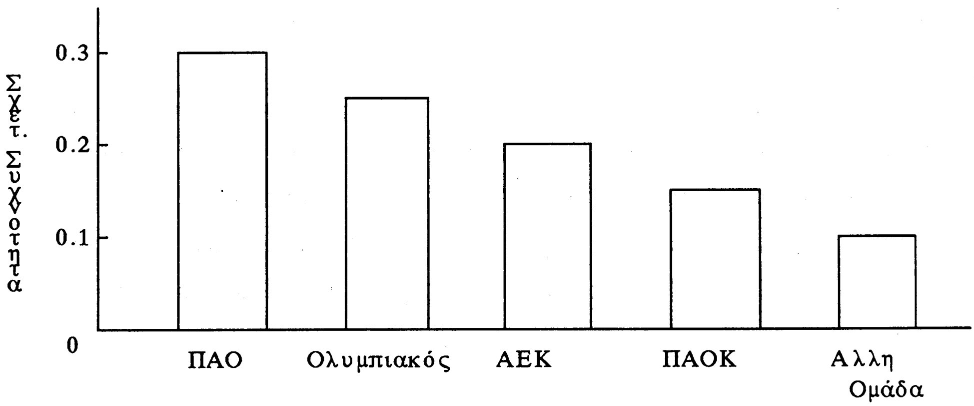
Τα κυκλικά διαγράμματα είναι αποτελεσματικά στις περιπτώσεις που αντικειμενικός σκοπός είναι να παρουσιαστούν οι συνιστώσες μιας ολότητας με τρόπο που αναδεικνύει τα σχετικά τους μεγέθη. Επειδή ακριβώς είναι πολύ παραστατικά, τα κυκλικά διαγράμματα χρησιμοποιούνται από εφημερίδες και περιοδικά για διάφορους λόγους και κυρίως όταν θέλουν να παρουσιάσουν οικονομικά μεγέθη. Τα κυκλικά διαγράμματα μπορούν να χρησιμοποιηθούν επίσης προκειμένου να συγκριθούν δύο διαφορετικές αναλύσεις ενός μεγέθους (που, ενδεχομένως, αντιστοιχούν σε διαφορετικές χρονικές στιγμές).

Συχνά, για μια πιο εντυπωσιακή παρουσίαση, οι κυκλικοί τομείς του διαγράμματος εμφανίζονται σε κάποια απόσταση από από το κέντρο του κύκλου.

Ραβδογράμματα (Bar charts)

Τα **ραβδογράμματα (bar charts)** παρέχουν έναν εναλλακτικό τρόπο γραφικής παρουσίασης ποιοτικών δεδομένων όπως αυτά του προηγουμένου παραδείγματος. Ενα ραβδόγραμμα παριστά τη συχνότητα (ή τη σχετική συχνότητα) σε κάθε κατηγορία σαν ένα ορθογώνιο που είναι κάθετο στον οριζόντιο άξονα. Το ύψος κάθε ορθογωνίου (ράβδου) είναι ανάλογο της συχνότητας (ή της σχετικής συχνότητας) της αντίστοιχης κατηγορίας. Δοθέντος ότι τα ραβδογράμματα αντιστοιχούν σε κατηγορίες ή σημεία μάλλον παρά σε κλάσεις διαστημάτων (όπως τα ορθογώνια ενός ιστογράμματος), το εύρος των ορθογωνίων μπορεί να είναι αυθαίρετο παρ'ότι θα πρέπει να είναι το ίδιο για όλα τα ορθογώνια (για όλες τις συχνότητες). Σε μερικές περιπτώσεις, προκειμένου το διάγραμμα να είναι περισσότερο σαφές, αφήνουμε κάποιο χώρο μεταξύ των ράβδων (των ορθογωνίων).

Μια σημαντική διαφορά των ραβδογραμμάτων από τα ιστογράμματα είναι ότι δεν υπάρχει φυσική σειρά με την οποία θα πρέπει να τοποθετηθούν οι κατηγορίες των ποιοτικών δεδομένων. Παρ'όλα αυτά ένας συχνό χρησιμοποιούμενος κανόνας, για προφανείς λόγους, είναι να διατάσσουμε τις κατηγορίες ανάλογα με τις συχνότητες που αντιστοιχούν σε κάθε μια από αυτές. Ετσι, η κατηγορία με την υψηλότερη συχνότητα συνήθως τοποθετείται πρώτη (ή τελευταία), εκείνη που έχει την αμέσως μικρότερη συχνότητα δεύτερη (ή προτελευταία) κ.ο.κ. Το σχήμα 2.8.2 που ακολουθεί δίνει το ραβδόγραμμα των προτιμήσεων για τον επόμενο πρωταθλητή.



Σχήμα 2.8.2
Ραβδόγραμμα Προτίμησης για τον Επόμενο Πρωταθλητή

Ενα ραβδόγραμμα το οποίο απεικονίζει την κατανομή συχνότητας ποσοτικών δεδομένων με τις κατηγορίες διατεταγμένες ανάλογα με τη συχνότητα ονομάζεται **διάγραμμα Pareto (Pareto diagram)**.

Πολλές φορές αντί να χρησιμοποιούμε μια διαφορετική ράβδο (ορθογώνιο) για κάθε κατηγορία προτιμούμε να εμφανίζουμε όλες τις κατηγορίες σε ένα μόνο ορθογώνιο. Προτιμούμε, δηλαδή, να στοιβάζουμε ή να συσσωρεύουμε τις αντίστοιχες συχνότητες σε μια ράβδο. Ενα τέτοιο γράφημα λέγεται **ραβδόγραμμα συνιστωσών (component bar chart)**. Το ορθογώνιο χωρίζεται σε τμήματα (συνιστώσες) των οποίων το ύψος είναι ανάλογο με τη συχνότητα (ή τη σχετική συχνότητα) της κατηγορίας που αντιπροσωπεύει. Τα ραβδογράμματα συνιστωσών αποτελούν μια καλή εναλλακτική λύση όταν θέλουμε να συγκρίνουμε δύο αναλύσεις του αυτού μεγέθους (εναλλακτική λύση των δύο διαφορετικών κυκλικών διαγραμμάτων).

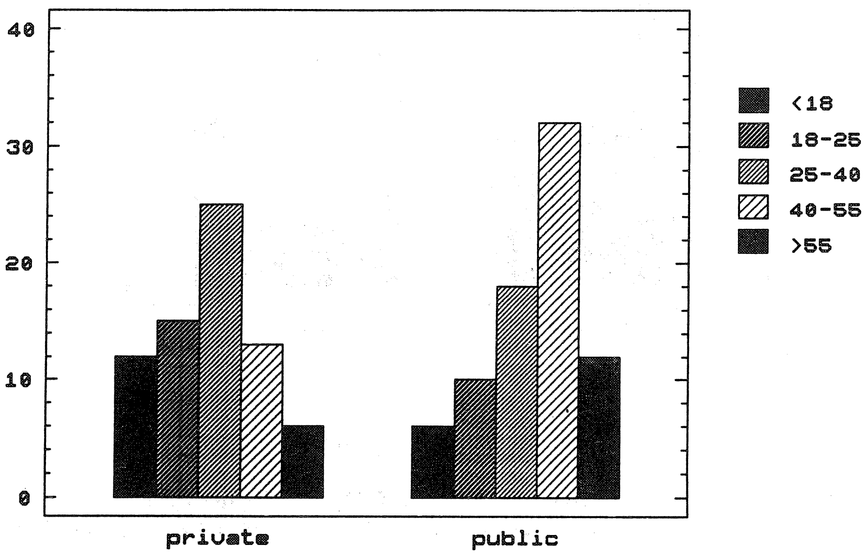
Τα διαγράμματα αυτά επίσης δίνουν τη δυνατότητα στον αναγνώστη να διαπιστώσει το μέγεθος των μεταβολών για κάθε κατηγορία καλύτερα από ότι στα κυκλικά διαγράμματα.

Τα ραβδογράμματα προσφέρουν επίσης τη δυνατότητα για συγκρίσεις συχνοτήτων μεταξύ διαφορετικών υποομάδων των δεδομένων. Για παράδειγμα, ας υποθέσουμε ότι έχουμε ένα δείγμα στοιχείων για τον τρόπο με τον οποίο πηγαίνουν στη δουλειά τους υπάλληλοι γραφείων κατά ηλικία, όπως φαίνεται στον πίνακα που ακολουθεί.

Πίνακας 2.8.2

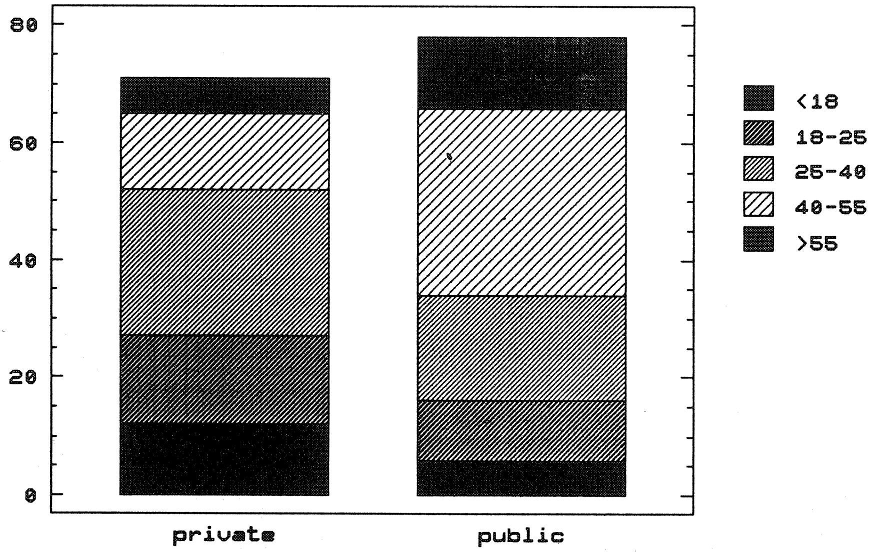
Κατηγορία ηλικίας	Τρόπος Μετακίνησης Ιδιωτικό μέσο	Δημόσιο μέσο
Κάτω των 18	12	6
18 - 25	15	10
25 - 40	25	18
40 - 45	13	32
Πάνω από 55	6	12

Τα ραβδογράμματα 2.8.3α και 2.8.3β μας δίνουν μια εικόνα του τρόπου με τον οποίο τα δύο είδη μεταφορικών μέσων χρησιμοποιούνται από τις διάφορες κατηγορίες ηλικιών.



Σχήμα 2.8.3α

Ραβδόγραμμα κατανομής συχνότητας χρήσης των διαφόρων μέσων μεταφοράς κατά ηλικία (ομαδοποιημένες (clustered) συχνότητες)

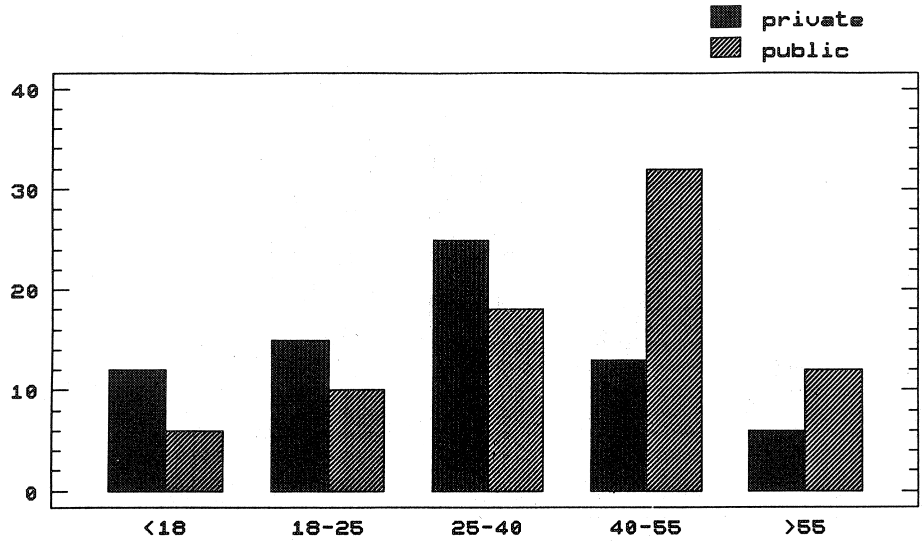


Σχήμα 2.8.3β

Ραβδόγραμμα κατανομής συχνότητας χρήσης των διαφόρων μέσων μεταφοράςκατά ηλικία (συσσωρευμένες (stacked) συχνότητες)

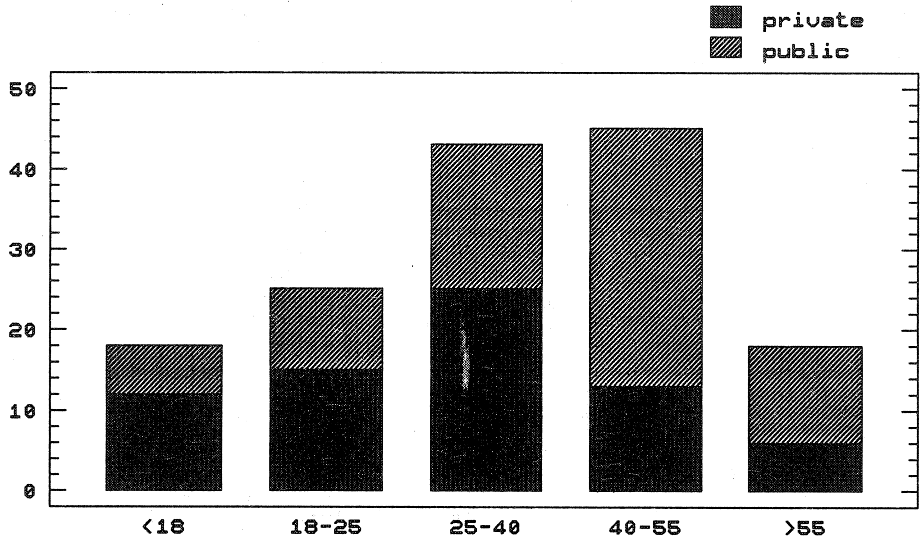
Στην παρουσίαση αυτή, το μέσο είναι η κύρια μεταβλητή και η κατηγορία της ηλικίας η δευτερεύουσα μεταβλητή.

Μπορεί κανείς να αντιστρέψει τους ρόλους των δύο αυτών μεταβλητών και να θεωρήσει την ηλικία ως κύρια μεταβλητή και το μέσο ως δευτερεύουσα. Αυτό οδηγεί στα ραβδογράμματα 2.8.3γ και 2.8.3δ που τώρα θα αναφέρονται στη συχνότητα με την οποία οι διάφορες κατηγορίες ηλικίας χρησιμοποιούν τα δύο είδη μεταφορικών μέσων.



Σχήμα 2.8.3γ

Ραβδόγραμμα κατανομής συχνότητας των διαφόρων ηλικιών κατά χρησιμοποιούμενο μέσο μεταφοράς (ομαδοποιημένες (clustered) συχνότητες)



Σχήμα 2.8.3δ

Ραβδόγραμμα κατανομής συχνότητας των διαφόρων ηλικιών κατά χρησιμοποιούμενο μέσο μεταφοράς (συσσωρευμένες (stacked) συχνότητες)

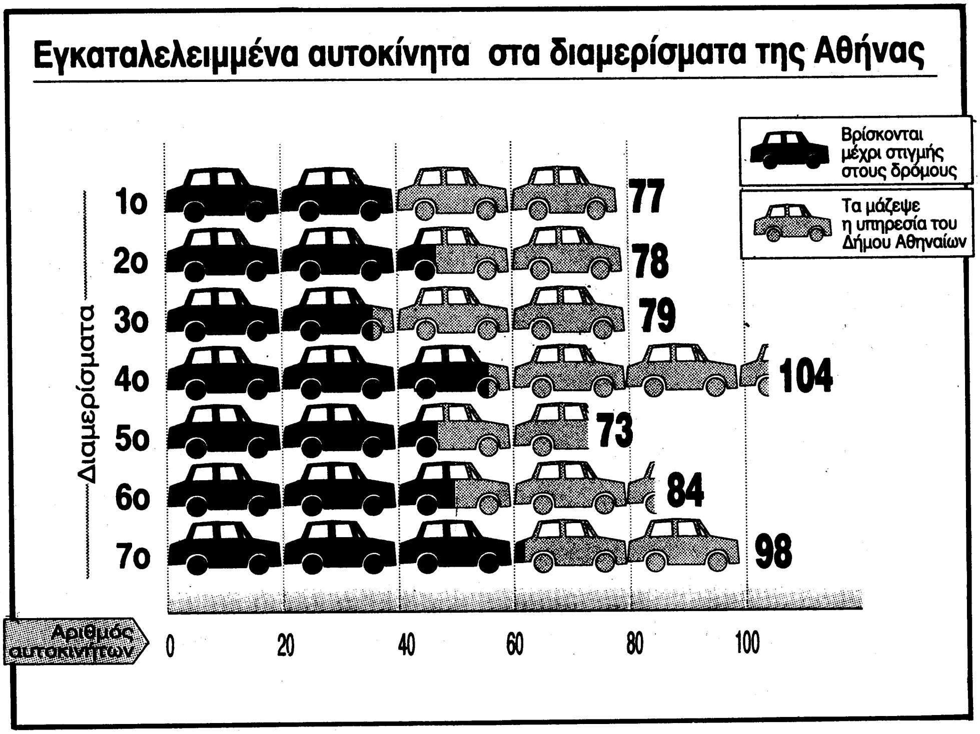
Σημείωση: Δοθέντος ότι τόσο το κυκλικό διάγραμμα όσο και το ραβδόγραμμα χρησιμοποιούνται για να παρουσιάσουν με γραφικό τρόπο ποιοτικά δεδομένα, γενιέται το ερώτημα ποια από τις δύο μεθόδους είναι καταλληλότερη. Η απάντηση είναι ότι αυτό εξαρτάται από το πού θέλουμε να δώσουμε έμφαση.

Εικονογράμματα (Pictograms)

Τα **εικονογράμματα (pictograms)** είναι διαγράμματα εικόνων τα οποία στηρίζονται στη λογική των ραβδογραμμάτων, όπου όμως οι ράβδοι αντικαθίστανται με εικόνες αντικειμένων (όπως αυτοκινήτων, σάκων, χρημάτων, ζώων, ανθρώπων) για να δώσουν μια πιο παραστατική εικόνα της κατάστασης.

Χαρακτηριστικό παράδειγμα αποτελεί το εικονόγραμμα του σχήματος 2.8.4 που δημοσιεύθηκε στην εφημερίδα ΤΑ ΝΕΑ (Πέμπτη 4 Νοεμβρίου 1993, σελ.22) σε σχέση με ρεπορτάζ που έγινε για τα εγκαταλελειμμένα αυτοκίνητα που υπάρχουν στους δρόμους της Αθήνας. Η σχετική έρευνα, όπως αναφέρει η εφημερίδα, έγινε στα εξής διαμερίσματα της πρωτεύουσας:

1ο Ομόνοια, Κουκάκι, Κεραμεικός, Σύνταγμα, Μοναστηράκι, Κολωνάκι, Εξάρχεια, Νεάπολη, Λ.Στρέφη, Μεταξουργείο (μέρος)
2ο Παγκράτι, Μέτς, Αγ. Αρτέμιος, Αγ. Ιωάννης Βουλιαγμένης, Αγ.Γεώργιος, Κυνοσάργους, Νέος Κόσμος, Δουργούτι, Αγ.Σώστης
3ο Ανω και Κάτω Πετράλωνα, Θησείο, Βοτανικός, Μέρος Μεταξουργείου, Κεραμεικός, Ρουφ
4ο Ακαδημ. Πλάτωνος, Παλαμήδι, Κολωνός, Σεπόλια, Θυμαράκια, Κάτω Πατήσια, Λόφος Σκουζέ, Κολοκυνθού
5ο Προμπονάς, Ριζούπολη, Λαμπρινή, Αλυσίδα, Ανω Πατήσια, Κυπριάδου, Κολιάτσου
6ο Κυψέλη, Ανιάτων, Λεβίδι, Αγγελοπούλου, Πολύγωνο
7ο Αμπελόκηποι, Γκύζη, Πεδίον Αρεως, Αγ. Θωμάς, Ν. Φιλοθέη, Ελληνορώσων



Σχήμα 2.8.4

Εγκαταλελειμμένα αυτοκίνητα στα διαμερίσματα της Αθήνας (ΤΑ ΝΕΑ 4 Νοεμβρίου 1993)

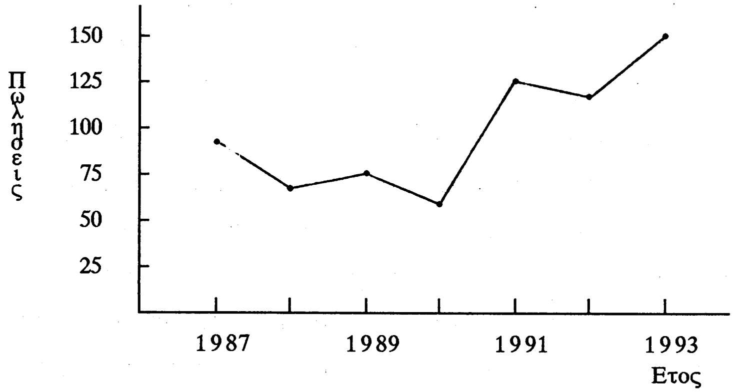
Γραμμογραφήματα (Line Charts)

Ενα **γραμμογράφημα (line chart)** κατασκευάζεται ως εξής: Σημειώνουμε την συχνότητα μιας κατηγορίας πάνω από το σημείο του οριζόντιου άξονα το οποίο αντιπροσωπεύει την κατηγορία αυτή και, στη συνέχεια, ενώνουμε τα σημεία αυτά με ευθύγραμμα τμήματα.

Επειδή το γραμμογράφημα συνήθως χρησιμοποιείται όταν οι κατηγορίες είναι σημεία στο χρόνο (χρονικές στιγμές) ονομάζεται επίσης και **γράφημα χρονολογικής σειράς** (time-series chart).

Τα γραμμογραφήματα έχουν το πλεονέκτημα ότι απεικονίζουν καθαρά την διαχρονική τάση των συνιστωσών της χρονολογικής σειράς.

Το σχήμα 2.8.5 για παράδειγμα απεικονίζει τις πωλήσεις μιας εταιρείας (σε εκατομμύρια δραχμές) για τα έτη 1987 έως 1993.



Σχήμα 2.8.5

Εν γένει, μπορεί να λεχθεί ότι τα κυκλικά διαγράμματα, τα ραβδογράμματα και τα γραμμογραφήματα χρησιμοποιούνται εκτεταμένα σε μελέτες και αναφορές που συντάσσονται από επιχειρήσεις, κυβερνήσεις και τα μέσα ενημέρωσης.

Υπάρχουν πάρα πολλές διαφορετικές μορφές γραφικών παραστάσεων των δεδομένων που έχουν ως στόχο να δώσουν μια συνοπτική παρουσίαση των δεδομένων. Στην ενότητα αυτή αναφερθήκαμε στις πιο διαδεδομένες από αυτές.

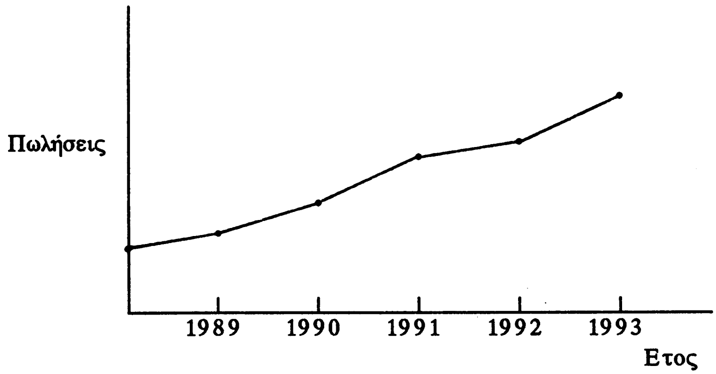
2.9 Παραποιήσεις με Γραφικές Παραστάσεις

Οι γραφικές παραστάσεις έχουν διεισδύσει στις εφημερίδες, στα περιοδικά, στις επιχειρήσεις, στις οικονομικές αναλύσεις, στα σεμινάρια κ.λ.π., εξαιτίας κυρίως της μεγάλης διάδοσης των υπολογιστών και αντιστοίχων πακέτων που επιτρέπουν την αποθήκευση, ανάσυρση, επεξεργασία και σύνοψη μεγάλης ποσότητας δεδομένων. Θα πρέπει λοιπόν να είμαστε σε θέση να αξιολογούμε τις πληροφορίες που παρουσιάζονται μέσω γραφικών τεχνικών. Δεν θα πρέπει να ξεχνάμε σε τελευταία ανάλυση ότι οι γραφικές τεχνικές απλώς δημιουργούν μία οπτική εντύπωση, γεγονός που αφήνει περιθώρια για παραποιήσεις στην παρουσίαση αυτή.

Θα πρέπει ίσως εδώ να λεχθεί ότι τέτοιες παραποιήσεις είναι τόσο εύκολες και έχουν γίνει τόσο συνηθισμένες ώστε το 1992 το Καναδικό Ινστιτούτο Ορκωτών Λογιστών (Canadian Institute of Chartered Accountants) υποχρεώθηκε να αρχίσει να δίνει βιβλία και να βάζει περιορισμούς σε γραφικές παραστάσεις που αναφέρονται σε οικονομικά δεδομένα μετά από μία μελέτη εκατοντάδων ετησίων αναφορών μεγάλων εταιρειών. Στη μελέτη αυτή βρέθηκε ότι 8% των απολογισμών περιέχουν τουλάχιστον ένα παραπλανητικό γράφημα που παρουσιάζει τα δεδομένα ωραιοποιώντας ή καλύπτοντας άσχημες καταστάσεις.

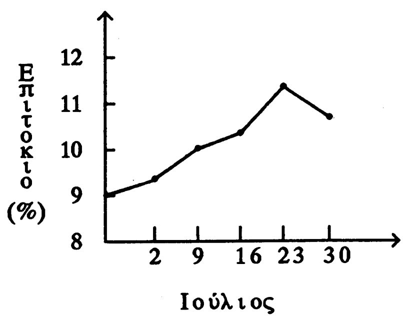
Θα πρέπει βέβαια να σημειωθεί ότι παραποιημένες παραστάσεις δεν γίνονται μόνο ηθελημένα αλλά και από ανθρώπους που, λόγω απειρίας, δεν έχουν την ικανότητα να δώσουν την ακριβή εικόνα. Μερικές μέθοδοι παραποίησης στοιχείων με γραφικές παραστάσεις είναι οι εξής:

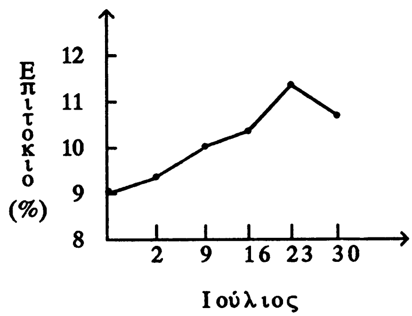
- Παρουσίαση γραφημάτων χωρίς κλίμακα σε έναν από τους δύο άξονες. Για παράδειγμα, το γράφημα της χρονολογικής σειράς των πωλήσεων μιας εταιρείας όπως δίνεται στο σχήμα 2.9.2 ενδέχεται να απεικονίζει ρυθμό ανάπτυξης 100% ή 1% για τα τελευταία πέντε χρόνια ανάλογα με την κλίμακα του κατακόρυφου άξονα. Τέτοια γραφήματα δεν θα πρέπει να λαμβάνονται υπόψη.



Σχήμα 2.9.2
Γράφημα χωρίς κατακόρυφη κλίμακα

- *Η λεζάντα η οποία παρουσιάζει ένα γράφημα.* Θα πρέπει να αποφεύγουμε να επηρεαζόμαστε από τη λεζάντα γιατί πολλές φορές έχει πρόθεση να οδηγήσει σε λάθος συμπεράσματα. Ετσι, για παράδειγμα, η εντύπωσή μας για την τάση των επιτοκίων θα είναι ίσως διαφορετική ανάλογα με το αν η λεζάντα είναι αυτή του σχήματος 2.9.3(α) ή του σχήματος 2.9.3(β). Θα πρέπει να προσπαθούμε να ερμηνεύουμε οι ίδιοι το γράφημα.



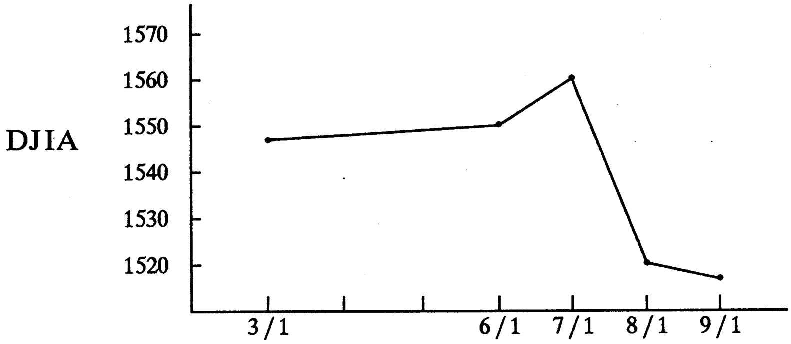


Σχήμα 2.9.3

(α) Τα επιτόκια άρχισαν τελικά να μειώνονται

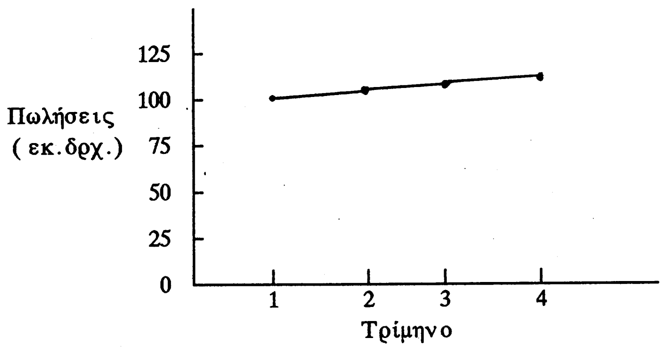
(β) Την τελευταία εβδομάδα παρατηρήθηκε προσωρινή κάμψη στην ανοδική τάση των επιτοκίων

- Λανθασμένη εικόνα και εντύπωση δίνεται επίσης αν σε ένα γράφημα εμφανίζονται μόνο απόλυτες μεταβολές στις τιμές και όχι ποσοστιαίες μεταβολές. Μια μεταβολή 100 δραχμών σε μια μετοχή αξίας 200 δραχμών είναι, σχετικά, χειρότερη από μια μεταβολή 100 δραχμών στην τιμή μιας μετοχής αξίας 10.000 δραχμών. Ενα παράδειγμα τέτοιας περίπτωσης εμφανίστηκε στις 9 Ιανουαρίου 1986 στις εφημερίδες της Βόρειας Αμερικής. Την ημέρα εκείνη οι εφημερίδες παρουσίασαν γραφικές παραστάσεις παρόμοιες με αυτές του σχήματος 2.9.4 και ανέφεραν ότι η χρηματιστηριακή αγορά, όπως αυτή μετριέται με τον δείκτη Dow Jones (Dow Jones Industrial Average (DJIA)), είχε υποστεί τη μεγαλύτερη ημερήσια πτώση σε σχέση με την προηγούμενη μέρα. Η πτώση ήταν 39 μονάδες μεγαλύτερη ακόμα και από τις απώλειες της "Μαύρης Τρίτης" (Black Tuesday) της 28ης Οκτωβρίου 1929. Παρ'ότι οι απώλειες ήταν πράγματι μεγάλες, πολλές αναλύσεις των μέσων ενημέρωσης παρέλειψαν να σημειώσουν ότι το επίπεδο του DJIA το 1986 ήταν πολύ υψηλότερο από 1929. Μια πιο καλή εικόνα της κατάστασης θα μπορούσε να δοθεί αν αναφερόταν ότι οι απώλειες στις 8 Ιανουαρίου 1986 αντιπροσώπευαν μια πτώση 2.5% ενώ η πτώση το 1929 ήταν 12.8%. (Για τους ενδιαφερόμενους μπορούμε να αναφέρουμε ότι η χρηματαγορά κέρδισε 12% μέσα σε δυο μήνες από την ημερομηνία αυτή της ιστορικής πτώσης και 40% μέσα σε ένα χρόνο. Η χειρότερη πτώση μιας ημέρας για όλη την ιστορία ήταν 22% και συνέβη στις 19 Οκτωβρίου 1929).



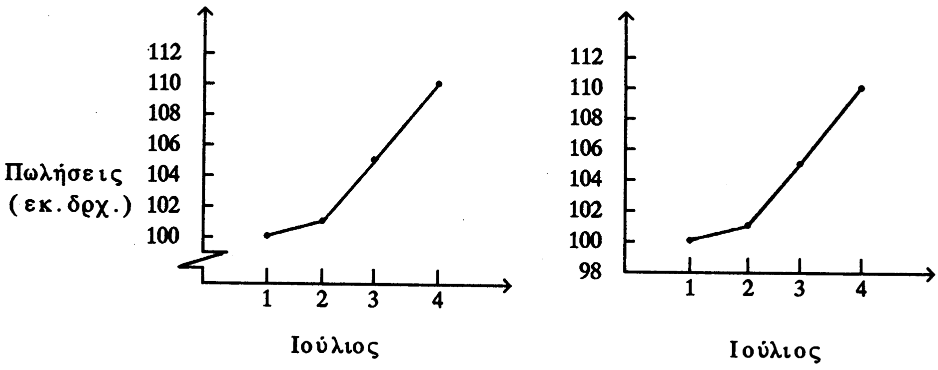
Σχήμα 2.9.4
Πτώση του DJIA το 1986

- Αλλος τρόπος με τον οποίο μπορεί να δοθεί μια παραποιημένη εικόνα των δεδομένων είναι η *"επιμήκυνση" του κάθετου άξονα,* μια τεχνική που γίνεται με αλλαγή της κλίμακας στον κάθετο άξονα έτσι ώστε κάθε μονάδα να παρουσιάζεται με μεγαλύτερο ύψος από ότι προηγουμένως. Αυτό κυρίως γίνεται με γραφήματα που παρουσιάζουν τη δραστηριότητα επιχειρήσεων που θέλουν να εμφανίσουν πωλήσεις ή παραγωγή μεγαλύτερη από την πραγματική. Με τον τρόπο αυτό (με την αύξηση, δηλαδή της κλίμακας στον κατακόρυφο άξονα) η κλίση του γραφήματος εμφανίζεται οπτικά (αλλά όχι αριθμητικά) μεγαλύτερη.



Σχήμα 2.9.5
Τριμηνιαίες πωλήσεις προηγουμένου έτους

Τυπικό παράδειγμα αποτελεί το γράφημα του σχήματος 2.9.5 που απεικονίζει την ανάπτυξη των τριμηνιαίων πωλήσεων μιας εταιρείας για το προηγούμενο έτος από 100 εκατομμύρια δραχμές σε 110 εκατομμύρια δραχμές. Αυτή η κατά 10% αύξηση των τριμηνιαίων πωλήσεων μπορεί να εμφανισθεί (οπτικά) περισσότερο δραματική με την μέθοδο της "επιμήκυνσης" του κατακόρυφου άξονα όπως φαίνεται στο γράφημα 2.9.6. Για να φιλοξενηθεί στον διαθέσιμο χώρο η επιμηκυμένη κλίμακα, υιοθετείται συχνά η θλάση (break) του κατακόρυφου άξονα, δηλαδή η διακοπή της συνέχειάς του όπως στο σχήμα 2.9.6(α), ή και η περικοπή του (truncation), δηλαδή η μεταφορά της αρχής του σε ένα σημείο μεγαλύτερο από το 0, όπως στο σχήμα 2.9.6(β). Με οποιονδήποτε από αυτούς τους τρόπους επιτυγχάνεται να αρχίζει η κατακόρυφη κλίμακα από ένα σημείο μεγαλύτερο του 0.

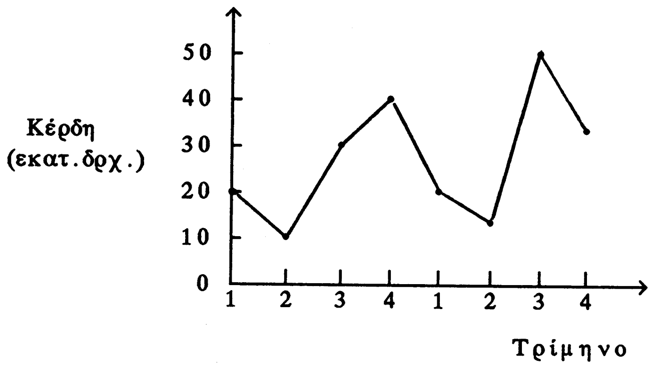


Σχήμα 2.9.6

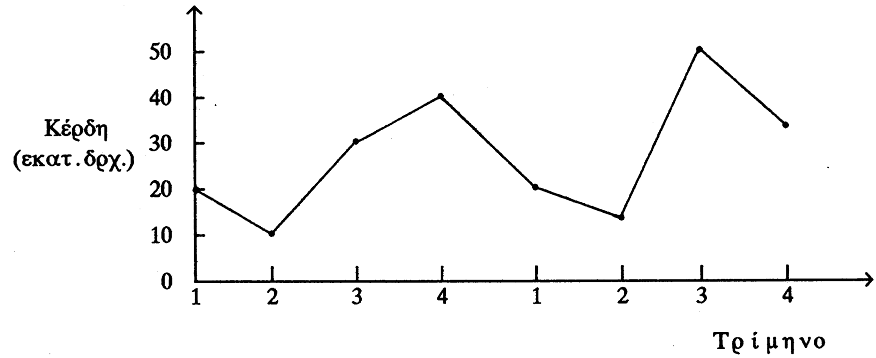
(α) Θλάση του κατακόρυφου άξονα **(β) Περικοπή του κατακόρυφου άξονα**

Η εντύπωση ότι η κλίση είναι πιο απότομη από ότι στην πραγματικότητα μπορεί επίσης να επιτευχθεί με "συρρίκνωση" του οριζοντίου άξονα οπότε τα σημεία στον άξονα αυτό πλησιάζουν το ένα στο άλλο.

Ακριβώς το αντίθετο αποτέλεσμα επιτυγχάνεται με την "επιμήκυνση" του οριζοντίου άξονα έτσι ώστε η απόσταση μεταξύ των σημείων στον άξονα αυτό να μεγαλώσει και η κλίση και οι τάσεις να εμφανίζονται λιγότερο απότομες. Αυτή η παραπλανητική εικόνα χρησιμοποιείται κυρίως από αυτούς οι οποίοι θέλουν να παρουσιάσουν ότι τα δεδομένα που έχουν εμφανίζουν μια εποχιακή σταθερότητα (και όχι μεγάλες μεταβολές). Ετσι το γράφημα του σχήματος 2.9.7(α) που απεικονίζει τα κέρδη μιας εταιρείας δείχνει σημαντικές ταλαντώσεις των κερδών και προς τα πάνω και προς τα κάτω από τρίμηνο σε τρίμηνο. Ομως η εταιρεία θα μπορούσε να δώσει την εντύπωση μιας ανεκτής σταθερότητας στα κέρδη από τρίμηνο σε τρίμηνο με την "επιμήκυνση" του οριζόντιου άξονα όπως φαίνεται στο γράφημα του σχήματος 2.9.7(β).

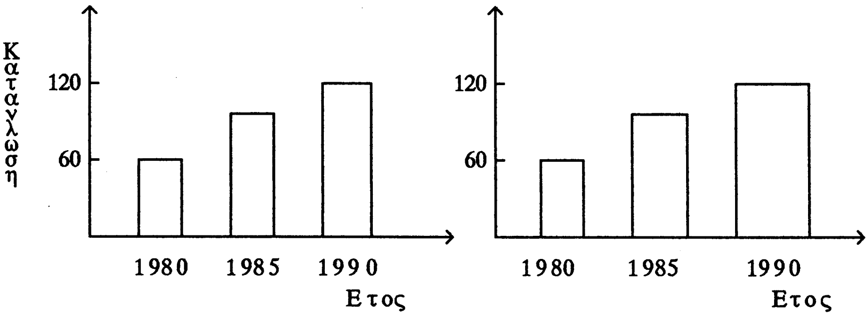


Σχήμα 2.9.7
(α) Συρρικνωμένος οριζόντιος άξονας



Σχήμα 2.9.7
(β) Επιμηκυνθείς οριζόντιος άξονας

Παρόμοιες παραποιήσεις μπορούν να πραγματοποιηθούν με τα ραβδογράμματα και πάλι με την "επιμήκυνση" ή συρρίκνωση του κατακόρυφου ή του οριζόντιου άξονα. Μια άλλη μέθοδος εμφάνισης παραποιημένης εικόνας με ραβδογράμματα είναι η κατασκευή των ορθογωνίων να γίνεται έτσι ώστε το εύρος τους να είναι ανάλογο με το ύψος τους. Με τον τρόπο αυτό η οποιαδήποτε διαφορά στις συχνότητες εμφανίζεται πολλαπλάσια από ότι πραγματικά είναι. Η μορφή αυτή της παραποίησης παρουσιάζεται στο σχήμα 2.9.8 όπου, για παράδειγμα, η μέση μηνιαία κατανάλωση για φαγητό οικογενειών για τρία συγκεκριμένα χρόνια (σε χιλιάδες δραχμές) απεικονίζεται ορθά με το γράφημα 2.9.8(α) να έχει διπλασιασθεί στο τέλος της δεκαετίας 1980-1990. Αντίθετα, στο σχήμα 2.9.8(β) η απεικόνιση της κατάστασης είναι παραπλανητική αφού οι ράβδοι με βάση ανάλογη του ύψους τους δίνουν την εντύπωση τετραπλασιασμού της κατανάλωσης.



Σχήμα 2.9.8

(α) Σωστό ραβδόγραμμα

(β) Αυξανόμενο εύρος ράβδου για την επίτευξη διαστρεβλωμένης εικόνας

Παρόμοιες παραποιήσεις μπορούν να γίνουν με εικονογράμματα.

Και στις περιπτώσεις αυτές εμφανίζεται συχνά η κατάσταση να μεγενθύνεται η διάσταση μιας εικόνας, δηλαδή το εύρος της μαζί με το ύψος της.

Ολα τα προηγούμενα παραδείγματα οδηγούν στο συμπέρασμα ότι προκειμένου να αντιληφθούμε καλά την εικόνα που παρουσιάζουν δεδομένα που εμφανίζονται με τη μορφή γραφικών παραστάσεων θα πρέπει να δώσουμε πολύ μεγαλύτερη σημασία στις αριθμητικές τιμές από ότι στο γράφημα αυτό καθεαυτό. Θα πρέπει τέλος να προσέχουμε ιδιαίτερα τις κλίμακες στους άξονες, και να αγνοούμε τα γραφήματα χωρίς κλίμακες στους άξονες.

ΑΣΚΗΣΕΙΣ

2.1 Ο αναλυτής εργασιών μιας τράπεζας κατέγραψε τον αριθμό των συναλλαγών (καταθέσεις και αναλήψεις πελατών) καθημερινά για μια περίοδο 10 εβδομάδων. Τα δεδομένα ήταν ώς εξής: (σε κάθε γραμμή έχουμε μια ημέρα της εβδομάδας από τη Δευτέρα έως την Παρασκευή)

ΔΕ	ΤΡ	ΤΕ	ΠΕ	ΠΑ
64	96	75	105	169
67	104	74	73	202
70	116	89	112	230
68	95	121	83	168
55	109	99	94	157
68	90	105	78	179
64	89	105	89	219
71	87	116	82	132
55	78	87	90	119
74	83	73	75	148

α) Να κατασκευασθεί το διάγραμμα σημείων για τα δεδομένα αυτά.

β) Ποια είναι τα συμπεράσματα που συνάγετε από το διάγραμμα αυτό των σημείων όσο αφορά τη διαδικασία συναλλαγών;

γ) Ποιο μέρος του διαγράμματος αυτού έχει περισσότερη έννοια; Εξηγείστε.

δ) Προσδιορίστε όσο το δυνατόν περισσότερες πηγές μεταβλητότητας στα δεδομένα αυτά με βάση τις γνώσεις σας για τις συναλλαγές τραπεζών.

ε) Καθορίστε τη διεργασία ή τον πληθυσμό για τα οποία μπορεί να εξαχθεί συμπερασματολογία με βάση τα δεδομένα αυτά.

στ)Με βάση τα αποτελέσματα στα προηγούμενα ερωτήματα, σε τι προτάσεις καταλήγετε σε σχέση με τις καθημερινές ανάγκες σε προσωπικό;

2.2 Στη συνέχεια της προηγούμενης μελέτης, ο αναλυτής της τράπεζας κατέγραψε το χρόνο (σε λεπτά) που απαιτείται για να συμπληρωθεί μια συναλλαγή από ένα τυχαίο δείγμα 50 πελατών στον τελευταίο μήνα. Τα δεδομένα είναι ως εξής:

2.3	0.2	2.9	0.4	2.8	2.4	4.4	5.8	2.8	3.3
3.3	9.7	2.5	5.6	9.5	1.8	4.7	0.7	6.2	1.2
7.8	0.8	0.9	0.4	1.3	3.1	3.7	7.2	1.6	1.9
2.4	4.6	3.8	1.5	2.7	0.4	1.3	1.1	5.5	3.4
4.2	1.2	0.5	6.8	5.2	6.3	7.6	1.4	0.5	1.4

α) Να κατασκευάσετε ένα διάγραμμα μίσχου-φύλλου για τα δεδομένα αυτά.

β) Να κατασκευάσετε ένα ιστόγραμμα για τα δεδομένα αυτά.

γ) Να περιγράψετε την κατανομή των χρόνων συναλλαγής.

δ) Να προσδιορίσετε τη διεργασία ή τον πληθυσμό στην οποία θα μπορεί να εφαρμοσθεί η όποια συμπερασματολογία προκύψει από τα δεδομένα αυτά.

2.3 Οι μέσες μηνιαίες πωλήσεις για τον τελευταίο χρόνο (σε εκατοντάδες χιλιάδες δραχμές) των 20 πρατηρίων μιας βιομηχανίας είναι οι εξής:

40.2	29.3	35.6	88.2	42.9
26.9	28.7	99.8	35.6	37.8
44.2	32.3	55.2	50.6	25.2
31.7	36.8	45.2	25.1	39.7

α) Να οργανωθούν τα δεδομένα αυτά με την κατασκευή ενός διαγράμματος μίσχου-φύλλου.

β) Να κατασκευασθεί ένα ιστόγραμμα.

γ) Με βάση τα αποτελέσματα στις ερωτήσεις (α) και (β) θα μπορούσε κανείς να οδηγηθεί στο συμπέρασμα ότι κάποια πρατήρια αποδίδουν καλύτερα από άλλα; Να δικαιολογήσετε την απάντησή σας.

2.4 Σε σχέση με την προηγούμενη άσκηση

α) Να καθορίσετε όλες τις πηγές μεταβλητότητας των δεδομένων που νομίζετε ότι υπάρχουν και να καθορίσετε για κάθε μια από αυτές αν προέρχεται από κοινή αιτία ή ειδική αιτία.

β) Ποιες από αυτές τις αιτίες μεταβλητότητας ελέγχονται από τον διευθυντή του πρατηρίου και ποιες είναι πέρα από τον έλεγχό του;

γ) Πώς θα χρησιμοποιούσατε τις απαντήσεις στα προηγούμενα δύο ερωτήματα για να καθορίσετε αν ο αριθμός των πωλήσεων θα μπορούσε να χρησιμοποιηθεί για να χαρακτηρισθεί η απόδοση των πρατηρίων;

2.5 Σε μια εταιρεία που κατασκευάζει μπαταρίες επελέγησαν τυχαία 32 μπαταρίες και υποβλήθηκαν σε έναν έλεγχο διάρκειας ζωής. Τα δεδομένα που ακολουθούν αναφέρονται στο χρόνο ζωής (σε ώρες) για τις 32 αυτές μπαταρίες:

52.5	62.7	58.9	65.7	49.3	62.8	48.3	52.9
58.9	57.3	60.4	59.6	58.1	55.3	54.9	63.4
62.3	64.4	52.7	54.9	48.8	54.6	64.2	57.2
56.8	53.1	58.7	61.6	63.3	51.7	59.5	56.8

α) Να καθορίσετε τη διεργασία ή τον πληθυσμό για τον οποίο μπορεί να εξαχθεί συμπερασματολογία με βάση τα δεδομένα αυτά.

β) Να κατασκευάσετε ένα διάγραμμα μίσχου-φύλλου και να το χρησιμοποιήσετε για να περιγράψετε την κατανομή του χρόνου ζωής των μπαταριών.

γ) Να κατασκευάσετε ένα ιστόγραμμα.

δ) Με βάση την εργασία που κάνετε για τις ερωτήσεις (β) και (γ) να περιγράψετε την ποσότητα της μεταβλητότητας στο χρόνο ζωής των μπαταριών που δημιουργεί η διαδικασία αυτή.

2.6 Οι βαθμοί σε ένα διαγώνισμα Στατιστικής ήταν οι εξής:

75	66	77	66	64	73	91	65	59	86
61	86	61	58	70	77	80	58	94	78
62	79	83	54	52	45	82	48	67	55

α) Να κατασκευάσετε ένα διάγραμμα μίσχου-φύλλου για τα δεδομένα αυτά.

β) Να κατασκευάσετε την κατανομή συχνότητας για τα δεδομένα χρησιμοποιώντας έξι κλάσεις.

γ) Να κατασκευάσετε το ιστόγραμμα σχετικής συχνότητας.

δ) Να βρείτε την αθροιστική κατανομή σχετικής συχνότητας.

ε) Ποιο είναι το ποσοστό των βαθμών που είναι μικρότεροι από 70; Ποιο είναι το ποσοστό αυτών που είναι μεγαλύτεροι από 70;

2.7 Ενας μεσίτης ενδιαφέρεται να μελετήσει την κατανομή των τιμών διαμερισμάτων που πουλήθηκαν τον τελευταίο χρόνο σε μια ακριβή περιοχή που θέλει να ανοίξει γραφείο. Με τυχαία δειγματοληψία συγκέντρωσε τις εξής τιμές διαμερισμάτων: (χιλιάδες ανά m^2)

274	429	229	435	260
222	292	419	242	202
235	215	390	359	409

375	209	265	440	365
319	338	414	249	279

α) Να κατασκευασθεί ένα διάγραμμα μίσχου-φύλλου για τις τιμές αυτές.

β) Να κατασκευασθεί η κατανομή συχνότητας για τις τιμές με τη χρήση πέντε κλάσεων και την τιμή 200 ως τη μικρότερη τιμή για την πρώτη κλάση.

γ) Να κατασκευασθεί το ιστόγραμμα σχετικής συχνότητας για το δείγμα αυτό.

δ) Να κατασκευασθεί ένα ακιδωτό για τις τιμές αυτές.

ε) Να βρεθεί το ποσοστό των τιμών που είναι μικρότερες από 300.000.

2.8 Ο φύλακας που κάθεται στην κεντρική είσοδο ενός κτιρίου για να περάσει την ώρα του αποφάσισε να μετρήσει τον αριθμό των ανθρώπων που εισέρχονται στο κτίριο το πρωί σε μια περίοδο 30 λεπτών. Για την καταμέτρηση χρησιμοποίησε ένα σύστημα όπου όλοι οι χρόνοι άφιξης στρογγυλοποιούνταν στο λεπτό που άρχιζε, π.χ. η ώρα άφιξης ατόμων που έφτασαν στο διάστημα των 59 δευτερολέπτων από τις 7:45 έως τις 7 και 45'και 59'' καταγράφηκε ως 7:45. Στη συγκεκριμένη περίπτωση τα δεδομένα που κατέγραψε δίνονται στον πίνακα που ακολουθεί.

α) Να κατασκευασθεί η κατανομή σχετικής συχνότητας για τους χρόνους αύξησης με τη χρήση έξι κλάσεων.

β) Να κατασκευασθεί η κατανομή αθροιστικής σχετικής συχνότητας για τους χρόνους άφιξης.

Ωρα	Αριθμός αφίξεων	Ωρα	Αριθμός αφίξεων	Ωρα	Αριθμός αφίξεων
7 : 45	1	7 : 55	15	8 : 05	0
7 : 46	4	7 : 56	21	8 : 06	0
7 : 47	2	7 : 57	9	8 : 07	4
7 : 48	0	7 : 58	12	8 : 08	0
7 : 49	4	7 : 59	6	8 : 09	1
7 : 50	7	8 : 00	4	8 : 10	1
7 : 51	10	8 : 01	10	8 : 11	0
7 : 52	8	8 : 02	2	8 : 12	0
7 : 53	4	8 : 03	0	8 : 13	1
7 : 54	9	8 : 04	3	8 : 14	0

γ) Ας υποθέσουμε ότι όλοι οι εργαζόμενοι στο κτίριο αυτό θα πρέπει να αρχίσουν να εργάζονται στις 8:00 ακριβώς. Ποιο είναι το ποσοστό των εργαζομένων του δείγματος που έφτασαν στη εργασία τους καθυστερημένοι;

δ) Ποιες είναι οι υποθέσεις που πρέπει να κάνει κανείς εάν θέλει να χρησιμοποιήσει τα δειγματικά αυτά δεδομένα για να εκτιμήσει το ποσοστό των εργαζομένων που φθάνουν καθυστερημένοι σε κάποια τυχαία μέρα;

2.9 Οι μηνιαίες πωλήσεις ενός καταστήματος (σε εκατοντάδες χιλιάδες δραχμές) τον τελευταίο χρόνο δίνονται στον πίνακα που ακολουθεί.

α) Να κατασκευασθεί ένα ραβδόγραμμα σχετικής συχνότητας για τα δεδομένα αυτά.

β) Να κατασκευασθεί το αντίστοιχο γραμμοδιάγραμμα για τα δεδομένα αυτά.

Μήνας	Πωλήσεις	Μήνας	Πωλήσεις
Ιανουάριος	65	Ιούλιος	88
Φεβρουάριος	61	Αύγουστος	93
Μάρτιος	70	Σεπτέμβριος	91
Απρίλιος	74	Οκτώβριος	78
Μάιος	72	Νοέμβριος	68
Ιούνιος	80	Δεκέμβριος	84

2.10 Το τεύχος της 6ης Φεβρουαρίου 1989 του περιοδικού *TIME* είχε ένα άρθρο με τίτλο "η άλλη πλευρά του ανταγωνισμού των εξοπλισμών" ("The other arms race"). Στο άρθρο αυτό παρέχονταν στοιχεία για τις αγορές όπλων από τους αμερικανούς πολίτες. Για να δώσει μεγαλύτερη έμφαση στα στοιχεία αυτά, το περιοδικό χρησιμοποίησε ένα ραβδόγραμμα προκειμένου να συγκρίνει τον αριθμό των ατόμων που δολοφονήθηκαν με μικρά όπλα (handguns). Να κατασκευάσετε το ραβδόγραμμα αυτό από τα δεδομένα που ακολουθούν.

Ανθρωποι που σκοτώθηκαν από όπλα (handguns)

Χώρα	Πληθυσμός (σε εκατ.)	Αριθμός
Καναδάς	25	5
Αγγλία	57	8
Ιαπωνία	121	46
Η.Π.Α.	239	8092

2.11 Η πολυεθνική εταιρεία NESTLÈ είχε για το 1987 πωλήσεις σε παγκόσμιο επίπεδο 23 δισεκατομμυρίων δολαρίων. Η κατανομή του ποσοστού των πωλήσεων ανάλογα με τη γεωγραφική περιοχή δίνεται στον πίνακα που ακολουθεί.

Πωλήσεις της NESTLÉ το 1987

Περιοχή	Πωλήσεις (ποσοστό)
Ευρώπη	43%
Βόρεια Αμερική	29%
Ασία	13%
Λατινική Αμερική και Καραιβική	10%
Αφρική	3%
Αυστραλία, Νέα Ζηλανδία και νησιά του Ειρηνικού	2%

α) Υποδείξτε πώς οι πληροφορίες αυτές μπορούν να παρουσιαστούν σε κάποιον ενδιαφερόμενο με τη χρησιμοποίηση κάποιας γραφικής τεχνικής.

β) Να εξηγήσετε γιατί διαλέξατε τη συγκεκριμένη τεχνική.

2.12 Προκειμένου να μελετήσει την ποιότητα των προιόντων που παράγει, ο κατασκευαστής των εξαρτημάτων ενός μηχανήματος καταγράφει τον αριθμό των ελαττωματικών εξαρτημάτων που αναγκάστηκε να απορρίψει τις τελευταίες 25 μέρες. Τα αποτελέσματα ήταν τα εξής:

21	8	17	22	19
18	19	14	17	11
6	21	25	19	9
12	16	16	10	29
24	6	21	20	25

α) Να κατασκευασθεί η κατανομή συχνότητας για τα στοιχεία αυτά. Να χρησιμοποιηθούν πέντε κλάσεις με κατώτερο άκρο της πρώτης κλάσης τα πέντε στοιχεία.

β) Να κατασκευασθεί το ιστόγραμμα σχετικής συχνότητας για τα δεδομένα.

γ) Ποια η σχέση μεταξύ των εμβαδών του ιστογράμματος και της σχετικής συχνότητας των παρατηρήσεων;

2.13 Οι ηλικίες των 25 διδασκόντων σε ένα πανεπιστημιακό τμήμα είναι οι εξής:

50	64	32	55	41
44	24	46	58	47
36	52	54	44	66
47	59	51	61	57
49	28	42	38	45

α) Να κατασκευασθεί ένα διάγραμμα μίσχου-φύλλου για τις ηλικίες αυτές.

β) Να κατασκευασθεί η κατανομή συχνοτήτων για τα δεδομένα χρησιμοποιώντας πέντε κλάσεις με την τιμή 20 ως το κατώτερο άκρο της πρώτης κλάσης.

γ) Να κατασκευασθεί το ιστόγραμμμα σχετικής συχνότητας για τα δεδομένα χρησιμοποιώντας πέντε κλάσεις και την τιμή 20 ως το κατώτερο άκρο της πρώτης κλάσης.

δ) Να κατασκευασθεί ένα ακιδωτό για τα δεδομένα αυτά.

ε) Ποιο είναι το ποσοστό της συνολικής επιφάνειας κάτω από το ιστόγραμμα της ερώτησης (β) μεταξύ των τιμών 20 και 40;