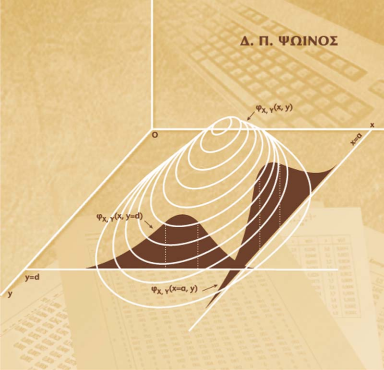
Δ. Π. ΨΩΙΝΟΣ



ΣΤΑΤΙΣΤΙΚΗ

ΘΕΣΣΑΛΟΝΙΚΗ 1999

ΕΚΔΟΣΕΙΣ ΖΗΤΗ

Κάθε γνήσιο αντίτυπο φέρει την υπογραφή του συγγραφέα

ISBN 960-431-561-7

Φωτοστοιχειοθεσία
Εκτύπωση

Π. ΖΗΤΗ & Σια ΟΕ
18ο χλμ Θεσ/νίκης-Περαίας
Τ.Θ. 171 • Νέοι Επιβάτες Θεσσαλονίκης • Τ.Κ. 570 19
Τηλ.: 03920-72.222 (5 γραμ.) - Fax: 03920-72.229
e-mail: info@ziti.gr

Βιβλιοπωλείο

ΕΚΔΟΣΕΙΣ **ΖΗΤΗ**
Αρμενοπούλου 27 • 546 35 Θεσσαλονίκη
Τηλ. 0310-203.720, Fax 0310-211.305
e-mail: sales@ziti.gr

www.ziti.gr

Πρόλογος

Μετά την εξάντληση της προηγούμενης έκδοσης, που χρησιμοποιήθηκε αρκετά χρόνια ως διδακτικό σύγγραμμα από τους φοιτητές των τμημάτων Μηχανολόγων και Χημικών Μηχανικών της Πολυτεχνικής Σχολής του Α.Π.Θ., θεώρησα ότι ήταν ανάγκη να ξαναγραφεί το βιβλίο. Με την ευκαιρία αυτή ήταν φυσικό να συμπληρωθούν και να αναθεωρηθούν διάφορα θεματικά στοιχεία, που είχε φανεί, από τη μακρόχρονη χρήση του βιβλίου, ότι χρειάζονταν βελτιώσεις. Αυτά έγιναν χωρίς ουσιαστικά να αλλάξει η διάρθωση της ύλης. Έτσι, και πάλι, στο πρώτο κεφάλαιο παρουσιάζεται η περιγραφική στατιστική, στο δεύτερο περιλαμβάνονται στοιχεία από τη θεωρία πιθανοτήτων και στο τρίτο αναπτύσσονται οι βασικές στατιστικές κατανομές. Στα επόμενα τρία κεφάλαια εξετάζονται τα θέματα των εκτιμήσεων, του ελέγχου των υποθέσεων και της παλινδρόμησης και συσχέτισης αντίστοιχα. Τέλος, προσέθεσα το έβδομο κεφάλαιο, το οποίο είναι μια εισαγωγή στο ιδιαίτερα σημαντικό θέμα της ανάλυσης μεταβλητότητας. Συμπληρωματικά, σε δυο παραρτήματα δίνονται ορισμένα στοιχεία από τη θεωρία συνόλων και μια σειρά από στατιστικούς πίνακες, που είναι εντελώς απαραίτητοι κατά τη μελέτη του κειμένου.

Κατά τη δόμηση τόσο της συνολικής ύλης, όσο και της ύλης κάθε κεφαλαίου επιδιώχθηκε βασικά η πιο καλή κάλυψη των αναγκών των φοιτητών και των διπλωματούχων Μηχανικών σε γνώσεις στατιστικής ανάλυσης. Έτσι, καταβλήθηκε ιδιαίτερη προσπάθεια να γίνεται φανερό ότι οι μέθοδοι της στατιστικής ανάλυσης δεν αποτελούν απλές αριθμητικές επεξεργασίες, αλλά μια θεμελιωμένη και πειθαρχημένη επιστημονική ενότητα, που πρέπει μάλιστα να χρησιμοποιείται μετά από αρκετή κατανόηση. Παράλληλα, δε χρησιμοποιήθηκαν προχωρημένες μαθηματι-

κές αναλύσεις εκεί όπου από τη μια δε θα προσέφεραν τίποτα σ' εκείνον που ενδιαφέρεται μόνο να χρησιμοποιήσει τη Στατιστική, και από την άλλη δεν οδηγούσαν σε λογικές ασυνέχειες κατά την παρουσίαση των διαφόρων θεμάτων, που θα ήταν σε βάρος της εύκολης κατανόησής τους. Επιδιώχθηκε, δηλαδή, το περιεχόμενο να μην είναι ούτε πολύ θεωρητικό, ούτε όμως πρακτικό, γεγονός που εκφράζει τη γενική άποψή μου ότι χρήσιμη θεωρία σημαίνει πράξη και πράξη που στερείται θεωρητικής θεμελίωσης είναι τουλάχιστο ελλιπής.

Η οργάνωση όλων των κεφαλαίων είναι ίδια. Συγκεκριμένα, στην αρχή περιγράφεται το πρόβλημα, ή τα προβλήματα, που πρέπει να λυθεί· στη συνέχεια εξετάζεται λεπτομερώς ο τρόπος με τον οποίον μπορούμε να το λύσουμε και τέλος δίνονται χαρακτηριστικά παραδείγματα. Παράλληλα, προσέθεσα ορισμένες εφαρμογές, που έχουν προκύψει από την ανάλυση πρακτικών προβλημάτων, με τις οποίες πιστεύω ότι θα βοηθηθεί ο αναγνώστης να κατανοήσει τη φύση των προβλημάτων που μπορεί να αντιμετωπίσει με τις μεθόδους της Στατιστικής· κάτι που δεν μπορεί να γίνει με τις ασκήσεις, που διευκολύνουν την κατανόηση της θεωρίας, αλλά δεν αποκαλύπτουν προβληματικές καταστάσεις, που μπορούν να αντιμετωπισθούν με τις στατιστικές αναλύσεις. Το σύνολο σχεδόν των παραδειγμάτων και των εφαρμογών συνδέεται με πραγματικά προβλήματα. Το ίδιο συμβαίνει και με τις ασκήσεις που ακολουθούν κάθε θεωρητική ενότητα. Σ' όλες τις ασκήσεις δίνονται οι απαντήσεις. Βιβλιογραφικές παραπομπές στο εσωτερικό του κειμένου γίνονται μόνο στις περιπτώσεις που δεν αποδεικνύεται κάτι που χρησιμοποιείται. Βέβαια, ο σχεδιασμός του βιβλίου έχει γίνει έτσι ώστε να είναι πλήρες με την έννοια ότι δε χρειάζεται να ανατρέξει ο αναγνώστης σε συμπληρωματικά συγγράμματα.

Με βάση τα παραπάνω ελπίζεται ότι ο μελετητής του βιβλίου θα κατανοήσει εύκολα τη θεωρία, θα αντιληφθεί τη χρησιμότητα της εφαρμογής των μεθόδων της Στατιστικής στους διάφορους τομείς ενασχόλησης των Μηχανικών και θα αποκτήσει την ικανότητα να τη χρησιμοποιεί για να λύνει πραγματικά προβλήματα.

Το κείμενο αυτό εξελίχθηκε στην σημερινή του μορφή με σταδιακές βελτιώσεις, όπου χρειαζόταν, στην παρουσίαση της ύλης και τον προσανατολισμό του περιεχομένου του στις ανάγκες τόσο των φοιτητών των Τμημάτων Μηχανολόγων και Χημικών Μηχανικών, όσο και των διπλωματούχων Μηχανικών των αντιστοίχων ειδικοτήτων. Σ' αυτή τη μακρόχρονη διαδικασία είναι φυσικό να έχω βοηθηθεί τόσο από το προσωπικό *του Τομέα Βιομηχανικής Διοίκησης, όσο και από τους φοιτητές μου με διάφορα σχόλια και παρατηρήσεις. Όλους τους ευχαριστώ θερμά. Επώνυμα όμως θέλω να ευχαριστήσω τη Μηχανολόγο Μηχανικό κ. Παρασκευή Καπετανοπούλου για την πολύτιμη βοήθειά της στην επεξεργασία των δοκιμίων καθώς και τις εκδόσεις Ζήτη που επιμελήθηκαν και αυτή την έκδοση.*

Θεσσαλονίκη, Σεπτέμβριος 1999
ΔΗΜΗΤΡΙΟΣ Π. ΨΩΙΝΟΣ

Περιεχόμενα

6. ΠΑΛΙΝΔΡΌΜΗΣΗ - ΣΥΣΧΈΤΙΣΗ

7. ΑΝΆΛΥΣΗ ΜΕΤΑΒΛΗΤΌΤΗΤΑς

ΠΑΡΑΡΤΗΜΑΤΑ

Εισαγωγή

Η βεβαιότητα στη ζωή είναι η εξαίρεση· ενώ ο κανόνας είναι η αβεβαιότητα, το ενδεχόμενο, το πιθανό. Γι' αυτό ο άνθρωπος διαφόρφωσε από πολύ παλιά την έννοια του περισσότερο ή λιγότερο πιθανού, δηλαδή την έννοια της πιθανότητας.

Μολονότι όμως ο άνθρωπος απασχολήθηκε με τις πιθανότητες από τους αρχαίους χρόνους, οι μαθηματικοί ενδιαφέρθηκαν ιδιαίτερα γι' αυτές στα μέσα του 17ου αιώνα και συγκεκριμένα το 1654. Τότε, ο ερασιτέχνης μαθηματικός και παίχτης τυχερών παιχνιδιών Chevalier de Meré έθεσε μια σειρά από ερωτήσεις σχετικές με τα τυχερά παιχνίδια στο Blaise Pascal (1623-1662). Ο Pascal στην προσπάθειά του να απαντήσει στις ερωτήσεις αυτές, και λύνοντας διάφορα σχετικά προβλήματα σε συνεργασία με τον Pierre de Fermat (1601-1665) δημιούργησε τις πρώτες βάσεις της Θεωρίας Πιθανοτήτων. Το ξεκίνημα όμως που προκάλεσαν ο Pascal και ο Fermat το ολοκλήρωσε ο Jacques Bernoulli (1654-1705), που ήταν ο πρώτος από μια οικογένεια με εννιά μεγάλους μαθηματικούς, γιατί κατέληξε σε θεμελιώδη συμπεράσματα. Ταυτόχρονα, ο ίδιος απέδειξε τη χρησιμότητα των πιθανοτήτων σε κοινωνικά και οικονομικά θέματα. Όλοι οι μεταγενέστεροι μελετητές των πιθανοτήτων θεμελίωσαν την εργασία τους στα ευρήματα του Bernoulli. Ανάμεσα στους πιο γνωστούς είναι οι Abraham de Moivre (1667-1754), Rierre Simon marquis de Laplace (1749-1827) και Karl Friedrich Gauss (1777-1855). Ο καθένας από αυτούς ανέπτυξε χωριστά την ιδέα της κανονικής καμπύλης. Πάνω σ' αυτή θεμελιώνεται η θεωρία σφαλμάτων που παίζει σημαντικό ρόλο στη Στατιστική. Ανεξάρτητα όμως από αυτή τη συγκεκριμένη συμβολή τους, το έργο αυτών των μεγάλων μαθηματικών καθιέρωσε την επιστημονική αυτοτέλεια της Θεωρίας Πιθανοτήτων.

Παράλληλα με τη Θεωρία Πιθανοτήτων αλλά ανεξάρτητα, αναπτύχθηκε η Στατιστική*. Ο σκοπός της Στατιστικής από παλιά –έχουν βρεθεί στοιχεία που αναφέρονται σε απογραφή που έγινε στην Ιουδαία το 2030 π.Χ.– και μέχρι το 17ο αιώνα ήταν η περιγραφή των πιο σημαντικών πράξεων του Κράτους [Corning (1600-1681) και μαθητές του]. Η άποψη αυτή εκφράζει το γεγονός ότι τη Στατιστική τη χρησιμοποιούσαν αποκλειστικά μέχρι τότε οι κρατικές υπηρεσίες για να συγκεντρώσουν και να επεξεργαστούν αριθμητικά δεδομένα σχετικά με τον πληθυσμό και τα οικονομικά του κράτους.

Στο 18ο αιώνα ο Deparcieux (1703-1768) και άλλοι πλούτισαν τη Στατιστική με ιδέες και μεθόδους της Θεωρίας Πιθανοτήτων και βοήθησαν έτσι αποφασιστικά στην ανάπτυξή τους. Ο Adolph Quetelet (1796-1874) διαμόρφωσε τη γνώση που υπήρχε μέχρι τότε στη Στατιστική σε εργαλείο έρευνας των κοινωνικών φαινομένων. Ο τελευταίος από τους πρωτοπόρους στη Στατιστική, όπως αυτή έχει διαμορφωθεί σήμερα, είναι ο Sir Francis Galton (1822-1911). Η μεγάλη συβμολή του Galton βρίσκεται στην εργασία του «Natural Inheritance», που δημοσιεύτηκε το 1889. Η εργασία αυτή ήταν η κινητήρια δύναμη για τον Karl Pearson, «τον πατέρα της Στατιστικής», όπως είναι γνωστός σήμερα, για να κινήσει τη Στατιστική σε νέους επαναστατικούς δρόμους σκέψης. Ο Karl Pearson θεμελίωσε την πρώτη από τις τέσσερεις μεγάλες σύγχρονες τάσεις της Στατιστικής, που ακόμη και σήμερα βρίσκεται σε εκπληκτική πρόοδο. Η συμβολή όμως του Pearson δεν μπορεί να δοθεί σε μερικές γραμμές, αφού από όλους ομολογείται ότι δεν πέρασε χρόνος από το 1890 μέχρι το θάνατό του το 1936 χωρίς να σημειώσει ουσιαστική συμβολή στη Στατιστική. Το πιο σημαντικό έργο του Pearson είναι οι τεχνικές που ασχολούνται με τα μεγάλα δείγματα, η κατανομή X^2 και η κατασκευή και η δημοσίευση μιας σειράς από στατιστικούς πίνακες, που είναι ιδιαίτερα χρήσιμοι και σήμερα σε οποιονδήποτε ασχολείται με τη Στατιστική. Τη δεύτερη τάση, που ασχολείται με την έρευνα και την ανάπτυξη μεθόδων για τα μικρά δείγματα, την οφείλουμε στους R.A. Fisher και W.S. Gosset. Από αυτή την τάση άρχισε η σχεδίαση των πειραμάτων, ο έλεγχος των υποθέσεων, η πολλαπλή συσχέτιση και η ανάλυση της μεταβλητότητας και συμμεταβλητότητας. Στο χρονικό διάστημα ανάμεσα στους δύο παγκόσμιους πολέμους, ο Jerzy Neyman και ο Egon Pearson ξεκίνησαν την τρίτη τάση της σύγχρονης Στατιστικής. Σ’ αυτή την τάση δόθηκε έμφαση στον

* Η καθημερινή χρήση του όρου «στατιστική» έχει δύο έννοιες. Η πρώτη είναι συνώνυμη με τα αριθμητικά δεδομένα. Μιλάμε π.χ. για τη στατιστική των τροχαίων ατυχημάτων στην Ελλάδα το 1998. Η δεύτερη έννοια του όρου «στατιστική» αναφέρεται στις αρχές και μεθόδους, που έχουν αναπτυχθεί για να επεξεργαζόμαστε αριθμητικά δεδομένα. Νομίζω ότι είναι προτιμότερο να περιορίζουμε τη χρήση του όρου στη δεύτερη έννοιά του. Ενώ για την πρώτη να χρησιμοποιούμε τον όρο «αριθμητικά δεδομένα».

έλεγχο των υποθέσεων. Η τέταρτη τάση διαμορφώθηκε από το Abraham Wald και κινήθηκε προς τη θεωρία των αποφάσεων. Οι εργασίες των ερευνητών που αναφέραμε, αλλά και πολλών άλλων, ανέπτυξαν τη Στατιστική, που θεμελιώνεται στους νόμους της Θεωρίας Πιθανοτήτων και που μπορούμε να την ορίσουμε ως το σύνολο των αρχών και μεθόδων με τις οποίες συλλέγουμε, παριστάνουμε, αναλύουμε και ερμηνεύουμε αριθμητικά δεδομένα, τόσο από την άποψη του θεωρητικού της περιεχομένου, όσο και των εφαρμογών της.

Η ανάπτυξη της Στατιστικής είναι τόση, ώστε καθένα από τα θέματα που αναφέραμε σχετικά με τις τέσσερις σύγχρονες τάσεις, όπως λ.χ. η ανάλυση μεταβλητότητας, ο έλεγχος υποθέσεων, κτλ. να καλύπτει ολόκληρους τόμους στη σχετική βιβλιογραφία. Όλες αυτές οι τάσεις οδήγησαν στην ωρίμανση της Στατιστικής των Συμπερασμάτων, που εκφράζει την πραγματική αξία της Στατιστική. Με αυτή έχουμε αποκτήσει τη δυνατότητα να συλλέγουμε κατάλληλα αριθμητικά δεδομένα και από αυτά να συναγάγουμε συμπεράσματα σε σχέση με τη λειτουργία και συμπεριφορά ενός συστήματος (παραγωγική διαδικασία, εργοστάσιο, οργάνωση κτλ.) από το οποίο προέρχονται τα αριθμητικά δεδομένα.

Η τεράστια ανάπτυξη της Θεωρίας Πιθανοτήτων και της Στατιστικής τόσο από άποψη θεωρίας, όσο και εφαρμογών τους έχει κάνει τη γνώση τους απαραίτητη σε κάθε επιστήμονα. Δεν πρόκειται να αποπειραθούμε να αιτιολογήσουμε αυτό τον ισχυρισμό στην εισαγωγή ενός διδακτικού κειμένου, γιατί έτσι θα υποτιμούσαμε αναγκαστικά τη σπουδαιότητα τόση της Θεωρίας Πιθανοτήτων όσο και τη Στατιστικής. Σπουδαιότητα που ο αναγνώστης θα αναγνωρίσει βαθμιαία καθώς θα δημιουργεί την εικόνα των δυνατοτήτων της καθεμιάς. Αυτό όμως που μπορούμε να πούμε από τώρα είναι ότι η Στατιστική και η Θεωρία Πιθανοτήτων βοήθησαν πολύ το σύγχρονο άνθρωπο να κατανοήσει τόσο τον μακρόκοσμο, όσο και τον μικρόκοσμο μέσα στον οποίο ζει και δραστηριοποιείται.

1 Εισαγωγή στη στατιστική ανάλυση

1.1 Γενικά

Συχνά χρειάζεται να μελετήσουμε ένα μέγεθος ή μια ιδιότητα που χαρακτηρίζει ένα σύνολο από άτομα ή αντικείμενα. Μπορεί π.χ. να θέλουμε να μελετήσουμε το ύψος των φοιτητών ενός Πανεπιστημίου ή τα ελαττωματικά προϊόντα που περιέχονται στην ημερήσια παραγωγή ενός εργοστασίου. Ένα τέτοιο σύνολο το λέμε *πληθυσμό*. Ανάλογα με το πλήθος των μελών του ο πληθυσμός μπορεί να είναι πεπερασμένος ή άπειρος. Ο πληθυσμός λ.χ. των συσκευών που παράγονται από ένα εργοστάσιο σ' ένα χρόνο είναι πεπερασμένος, ενώ ο πληθυσμός που δημιουργείται από τα αποτελέσματα που παρουσιάζονται όταν ρίχνουμε συνέχεια ένα νόμισμα είναι θεωρητικά άπειρος. Για να μελετήσουμε όμως έναν πληθυσμό, μπορεί είτε να μην μπορούμε, είτε να μη μας συμφέρει από άποψη προσπάθειας και χρόνου, να συλλέξουμε αριθμητικά –στατιστικά– στοιχεία μετρώντας ή παρατηρώντας κάθε στοιχείο από το σύνολο.

Εφόσον δεν μπορούμε, ή δεν πρέπει, να ασχοληθούμε με κάθε στοιχείο του πληθυσμού, καταφεύγουμε σ' ένα μέρος αυτού που το λέμε *δείγμα*. Το δείγμα, υπό ορισμένες προϋποθέσεις, μπορούμε να θεωρούμε ότι αντιπροσωπεύει τον πληθυσμό από τον οποίον προέρχεται. Αυτό σημαίνει ότι μπορούμε να εξάγουμε συμπεράσματα για τον πληθυσμό από το δείγμα. Το γεγονός ότι μπορούμε να θεωρούμε τα αριθμητικά στοιχεία του δείγματος ως αντιπροσωπευτικά των στοιχείων του πληθυσμού και να εξάγουμε συμπεράσματα για τον πληθυσμό από το δείγμα αποτελεί την ουσία της στατιστικής ανάλυσης.

Αν το χαρακτηριστικό του πληθυσμού που θέλουμε να μελετήσουμε παίρνει αριθμητικές τιμές που διαφέρουν μεταξύ τους από τυχαία αίτια, δηλαδή από αίτια που δεν μπορούν να ελεγχθούν από τον άνθρωπο, το λέμε *στατιστικό* ή *στοχαστικό μέγεθος* και το συμβολίζουμε γενικά με μια μεταβλητή x, y, ..., κτλ.

Τα στατιστικά μεγέθη τα διακρίνουμε σε *συνεχή* ή *ασυνεχή* ανάλογα με το είδος των τιμών που μπορούν να πάρουν. Παραδείγματος χάρη το μήκος, το βάρος, ο χρόνος κτλ., είναι συνεχή στατιστικά μεγέθη, ενώ ο αριθμός των βροχερών ημερών την εβδομάδα, ο αριθμός των γεννήσεων σε ορισμένο αριθμό οικογενειών το χρόνο κτλ. είναι ασυνεχή στατιστικά μεγέθη. Επειδή κάθε στατιστική έρευνα αρχίζει με τη συλλογή, την κατάλληλη ταξινόμηση και τη γραφική παράσταση των αριθμητικών τιμών ενός στατιστικού μεγέθους αρχίζουμε αμέσως να εξετάζουμε τις μεθόδους με τις οποίες μπορούμε να κάνουμε αυτές τις επεξεργασίες.

1.2. Συλλογή στατιστικών στοιχείων

Για να συλλέξουμε στατιστικά στοιχεία, δηλαδή αριθμητικές τιμές ενός στατιστικού μεγέθους, χρησιμοποιούμε δύο μεθόδους: την *απογραφή* και τη *δειγματοληψία*. Την απογραφή τη χρησιμοποιούμε κάθε φορά που είναι ανάγκη να μετρήσουμε ή να παρατηρήσουμε κάθε μονάδα ενός πληθυσμού. Τη δειγματοληψία τη χρησιμοποιούμε όταν είναι αρκετό να πάρουμε ένα μικρό τμήμα από τον πληθυσμό.

Τις αριθμητικές τιμές ενός στατιστικού μεγέθους, ανεξάρτητα από το αν κάνουμε δειγματοληψία ή απογραφή, μπορούμε να τις συλλέξουμε με πολλούς τρόπους. Τον τρόπο που τελικά θα χρησιμοποιήσουμε τον επιλέγουμε παίρνοντας υπόψη μας τη φύση του στατιστικού μεγέθους. Παραδείγματος χάρη, αν το στατιστικό μέγεθος που μας απασχολεί είναι η διάμετρος ενός άξονα που κατασκευάζεται, ο κατάλληλος τρόπος να συλλέξουμε τα αναγκαία στατιστικά στοιχεία είναι να μετρήσουμε τις διαμέτρους των αξόνων· αν όμως το μέγεθος είναι η βαθμολογία των φοιτητών σ' ένα ορισμένο μάθημα, ο καλύτερος τρόπος είναι να βρούμε μία κατάσταση βαθμολογίας τους στο συγκεκριμένο μάθημα. Αν ακόμη το στατιστικό μέγεθος είναι οι επισκέπτες που μπαίνουν καθημερινά σε μία έκθεση, σ' ένα μουσείο κτλ., τη συλλογή των στοιχείων μπορούμε να την κάνουμε τοποθετώντας έναν αυτόματο μετρητή στην είσοδο.

Από τα παραπάνω παραδείγματα φαίνεται καθαρά ότι αυτός που χρησιμοποιεί τη στατιστική σα μέσο έρευνας πρέπει να διδαχθεί περισσότερο από την πείρα, και λιγότερο από ένα σύντομο διδακτικό κείμενο, τους τρόπους με τους οποίους μπορεί να συγκεντρώνει τα στατιστικά στοιχεία που χρειάζεται κάθε φορά.

1.3. Ταξινόμηση στατιστικών στοιχείων

Στον πίνακα 1.1 έχουμε καταχωρήσει τους προαγωγικούς βαθμούς 100 φοιτητών. Οι βαθμοί μεταξύ τους δεν έχουν καμία σχέση επειδή τους αντιγράψαμε από μια ονομαστική κατάσταση των φοιτητών.

Πίνακας 1.1. Προαγωγικοί βαθμοί 100 φοιτητών.

5,55	6,05	7,05	6,90	6,25	6,75	6,15	7,00	6,60	5,25
8,05	7,05	6,20	7,35	7,85	6,90	7,25	5,90	7,15	6,40
6,70	5,75	6,70	5,85	6,75	6,50	5,40	7,95	6,80	6,65
6,10	7,10	7,30	6,15	7,65	6,90	7,60	6,10	7,30	6,35
7,20	8,10	6,25	7,15	6,00	6,65	6,20	7,65	6,90	5,90
5,60	7,40	7,05	5,85	7,60	6,80	7,40	7,70	7,80	6,00
6,80	6,80	5,90	7,55	6,30	5,70	6,75	7,20	8,40	7,45
6,85	6,60	7,55	6,85	6,55	7,70	7,10	5,85	7,75	6,05
6,40	7,50	6,65	5,60	7,35	6,85	6,55	7,75	6,80	7,25
6,60	6,75	6,85	6,60	7,00	6,65	7,70	6,85	6,70	6,50

Αν υποθέσουμε τώρα ότι θέλουμε να βρούμε το μικρότερο βαθμό, μπορούμε αναμφισβήτητα να το κάνουμε χωρίς μεγάλη δυσκολία. Αν όμως θέλουμε να διαπιστώσουμε κατά πόσο η βαθμολογία των φοιτητών στο σύνολό της είναι ικανοποιητική, ή αν υπάρχουν σημαντικές αποκλίσεις στην απόδοση των φοιτητών, είναι δύσκολο αν όχι αδύνατο, να δώσουμε απάντηση με μία απλή επισκόπηση του πίνακα. Συνεπώς οι βαθμοί, όπως τους έχουμε γράψει στον πίνακα, μας δίνουν ελάχιστες πληροφορίες σε σχέση με τη βαθμολογία των 100 φοιτητών. Κι εκείνες όμως που μας δίνουν χρειάζεται να τις βγάλουμε με δυσκολία.

Αν υποθέσουμε όμως ότι τους βαθμούς τους είχαμε κατατάξει με βάση το μέγεθός τους, θα ήταν ιδιαίτερα εύκολο να βρούμε το μικρότερο και το μεγαλύτερο βαθμό και να διαπιστώσουμε έτσι τα όρια μέσα στα οποία βρίσκεται το σύνολο των βαθμών. Επίσης, θα μπορούσαμε να παρατηρήσουμε αν υπάρχει κάποια συγκέντρωση βαθμών σ' ένα ορισμένο αριθμό κι ακόμη να αποκτήσουμε μια πρώτη εικόνα του τρόπου σύμφωνα με τον οποίο κατανέμονται οι βαθμοί ανάμεσα στο μικρότερο και στο μεγαλύτερο.

Για να ταξινομήσουμε όμως τις τιμές ενός στατιστικού μεγέθους που μας απασχολεί –στη συγκεκριμένη περίπτωση η βαθμολογία των φοιτητών– με βάση το μέγεθός τους, αντιμετωπίζουμε κάποια δυσκολία. Η δυσκολία αυτή γίνεται μεγαλύτερη όταν το πλήθος των τιμών που πρέπει να ταξινομηθούν είναι μεγάλο. Αλλά κι αν το κάνουμε αυτό, τελικά ούτε μπορούμε να βρούμε τις πληροφορίες που χρειαζόμαστε σε σχέση με το στατιστικό μέγεθος, όπως θα διαπιστώσουμε παρακάτω, ούτε μας διευκολύνει ο τρόπος αυτός να προχωρήσουμε στη μαθηματική ανάλυση με την οποία θα αποκτήσουμε τις πληροφορίες που χρειαζόμαστε. Χρειαζόμαστε συνεπώς κάποια άλλη διαδικασία για να επεξεργαζόμαστε τα διάφορα στατιστικά στοιχεία. Αυτή τη διαδικασία θα εξετάσουμε στην παρακάτω παράγραφο.

1.4. Κατανομές συχνότητας

1.4.1. Παράσταση με τη βοήθεια πίνακα

Ας δούμε τώρα τον τρόπο με τον οποίον πρέπει να ταξινομούμε και γενικά να οργανώνουμε τις τιμές ενός στατιστικού μεγέθους για να διευκολυνόμαστε στην ανάλυσή τους. Μολονότι με αυτή την οργάνωση των τιμών χάνουμε ορισμένες λεπτομέρειες, που μας δίνει η οργάνωση των στοιχείων κατά τάξη μεγέθους, τελικά μας διευκολύνει να έχουμε μια συνοπτική εικόνα των στοιχείων και να υπολογίσουμε ορισμένες χαρακτηριστικές τιμές τους, που μας είναι εντελώς απαραίτητες για να εξετάσουμε συνολικά ένα στατιστικό μέγεθος.

◆ **Εκλογή αριθμού κλάσεων.** Στην αρχή βρίσκουμε την περιοχή των τιμών, δηλαδή τη διαφορά της μικρότερης από τη μεγαλύτερη τιμή και τη διαιρούμε σ' ένα κατάλληλο αριθμό τμημάτων, που τα λέμε *κλάσεις*. Στον πίνακα 1.1 ο μικρότερος αριθμός είναι ο 5,25 και ο μεγαλύτερος ο 8,40, δηλαδή η περιοχή των τιμών μέσα στην οποία κατανέμονται οι βαθμοί έχει πλάτος 3,15. Αυθαίρετα θεωρούμε ότι αυτή έχει πλάτος 3,50. Δεχόμενοι ως περιοχή των τιμών το 3,50 αντί του 3,15, για λόγους απλούστευσης της διαδικασίας, δεν εισάγουμε λάθος αφού στο πρόσθετο τμήμα δεν έχουμε τιμές του στατιστικού μεγέθους που μελετούμε. Την περιοχή τη διαιρούμε σε 7 κλάσεις. Δυστυχώς, δεν υπάρχει γενικός κανόνας που να μπορούμε να ακολουθούμε κάθε φορά που χρειάζεται να καθορίσουμε τον αριθμό, ή το πλάτος, των κλάσεων. Οι κλάσεις στις οποίες διαιρούμε την περιοχή των τιμών ενός στατιστικού μεγέθους δεν είναι απαραίτητο να έχουν το ίδιο πλάτος. Υπάρχουν μάλιστα περιπτώσεις στις οποίες μπορεί να είναι καλύτερα να χρησιμοποιούνται κλάσεις με διαφορετικά πλάτη. Αυτό εξαρτάται από το τι ακριβώς θέλουμε να περιγράψουμε κάθε φορά.

Οι σκέψεις με τις οποίες προσπαθούμε να οδηγηθούμε στο σωστό αριθμό κλάσεων, με ίσα ή άνισα πλάτη, είναι οι εξής: Αν δημιουργήσουμε πάρα πολλές κλάσεις κι αν το πλήθος των τιμών είναι μικρό, σε πολλές από τις κλάσεις μπορεί είτε να μην εμφανιστούν τιμές, είτε να εμφανιστούν αναλογικά ελάχιστες. Αυτό όμως μπορεί να αλλοιώνει την εικόνα της πραγματικής τους κατανομής. Αν αντίθετα δημιουργήσουμε λίγες κλάσεις κι αν το πλήθος των τιμών είναι μεγάλο, η εικόνα της κατανομής των τιμών εντός της περιοχής των κλάσεων μπορεί επίσης να είναι αλλοιωμένη σε σχέση με την πραγματικότητα. Έτσι, ο αριθμός των κλάσεων ο οποίος τελικά πρέπει να χρησιμοποιηθεί εξαρτάται από το πλήθος των τιμών. Όσο μεγαλύτερο είναι το πλήθος των τιμών τόσο μεγαλύτερο αριθμό κλάσεων πρέπει να χρησιμοποιούμε. Από την πείρα έχει προκύψει ότι σπάνια πρέπει να χρησιμοποιούνται λιγότερες από 6 κλάσεις και περισσότερες από 16.

Στη βιβλιογραφία συναντούμε τον τύπο του Sturges, που δίνει τον αριθμό των κλάσεων K σα συνάρτηση του πλήθους των τιμών ενός στατιστικού μεγέθους[1]. Ο τύπος αυτός είναι ο εξής:

$$K = 1+3{,}3 \text{ λογ } N, \qquad (1.1)$$

όπου N είναι το πλήθος των τιμών. Δυστυχώς όμως, το πρόβλημα της εκλογής του αριθμού των κλάσεων δε λύνεται με αυτό τον τύπο γιατί από τη μια απαιτεί το πλάτος τους να είναι ίσο, και από την άλλη μας οδηγεί συχνά σε μικρό αριθμό κλάσεων. Γι’ αυτό πρέπει να τον χρησιμοποιούμε με προσοχή και μόνο συμπληρωματικά με τις σκέψεις που εκθέσαμε, παραπάνω.

Αφού εκλέξουμε τον αριθμό των κλάσεων, βρίσκουμε το πλάτος κάθε κλάσης. Στην περίπτωση του παραδείγματος των βαθμών το πλάτος κάθε κλάσης είναι προφανώς 0,50.

◆ **Όρια κλάσης.** Όπως είπαμε θεωρήσαμε αυθαίρετα ως πλάτος της περιοχής τιμών το 3,50 και το διαιρέσαμε σε 7 κλάσεις. Οι κλάσεις λοιπόν μπορεί να είναι οι 5,0-5,5, 5,5-6,0 κτλ. Έτσι όμως, το 5,5 είναι ταυτόχρονα το άνω όριο της πρώτης και το κάτω όριο της δεύτερης κλάσης. Επειδή αυτό μας δημιουργεί συχνά δυσκολίες χρησιμοποιούμε το συμβολισμό που βλέπουμε στον πίνακα 1.2. Δηλαδή, την πρώτη κλάση την συμβολίζουμε ως 5,0-5,4 και όχι ως 5,0-5,5 και τη δεύτερη ως 5,5-5,9 κ.ο.κ. Έτσι, το ανώτερο όριο της πρώτης κλάσης το βρίσκουμε αν στον αριθμό του συμβολισμού της κλάσης προσθέσουμε τον πρώτο της επόμενης κλάσης και το άθροισμα το διαιρέσουμε με δύο. Δηλαδή,

$$\frac{5{,}4 + 5{,}5}{2} = 5{,}45$$

Επειδή όμως το πλάτος των κλάσεων είναι 0,50, το κατώτερο όριο της πρώτης κλάσης είναι το 5,45-0,50 = 4,95. Ενώ λοιπόν οι κλάσεις στις οποίες κατα-

1. Βλ. Παππά, Α.Ι., Εφηρμοσμένη Στατιστική, Αθήνα 1960, σελ. 31.

λήγουμε στο παράδειγμά μας είναι οι 4,95-5,45, 5,45-5,95 κτλ. στον πίνακα 1.2 χρησιμοποιούμε το συμβολισμό 5,0-5,4, 5,5-5,9 κ.ο.κ. για να τις εκφράσουμε.

Πίνακας 1.2. Κατανομή συχνότητας προαγωγικών βαθμών 100 φοιτητών.

Κλάσεις βαθμών	**Αριθμός φοιτητών**
5,0 - 5,4	2
5,5 - 5,9	11
6,0 - 6,4	16
6,5 - 6,9	34
7,0 - 7,4	20
7,5 - 7,9	14
8,0 - 8,4	3
Σύνολο	**100**

Αφού εκλέξουμε τον αριθμό των κλάσεων και καθορίσουμε τα όριά τους, βρίσκουμε τον αριθμό των τιμών που ανήκουν σε κάθε κλάση. Τον αριθμό αυτό τον λέμε συχνότητα της κλάσης και τον συμβολίζουμε με το f. Έτσι, η συχνότητα π.χ. της κλάσης 6,0-6,4 του παραδείγματος των βαθμών είναι 16.

Στην περίπτωση που συμπίπτουν πολλές τιμές με όρια κλάσεων, είναι πιο σωστό να θεωρούμε ότι οι μισές από αυτές ανήκουν στην κατώτερη κλάση και οι υπόλοιπες στην ανώτερη.

Αφού προσδιορίσουμε και τη συχνότητα κάθε κλάσης έχουμε δημιουργήσει την κατανομή συχνότητας, όπως φαίνεται στον πίνακα 1.2 για το παράδειγμα των βαθμών.

Έχοντας αποσαφηνίσει παραπάνω τις βασικές έννοιες μπορούμε τώρα να προχωρήσουμε σε ορισμένους γενικούς ορισμούς.

Κατανομή συχνότητας λέμε το σύνολο των ζευγών των τιμών (x_1, f_1), (x_2, f_2), ..., (x_n, f_n), όπου $f_1, f_2, ..., f_n$ είναι οι συχνότητες με τις οποίες εμφανίζονται οι τιμές $x_1, x_2, ..., x_n$ ενός στατιστικού μεγέθους. Στην περίπτωση που πρόκειται περί κατανομής συχνότητας ενός συνεχούς στατιστικού μεγέθους, ως x_1, x_2, ..., x_n θεωρούμε τα μέσα των κλάσεων.

Σχετική συχνότητα μιας κλάσης λέμε το λόγο της συχνότητας της κλάσης προς το άθροισμα των συχνοτήτων όλων των κλάσεων. Έτσι η σχετική συχνότητα π.χ. της κλάσης 5,0-5,4 του παραδείγματος των βαθμών είναι 0,02. Τη σχετική συχνότητα συχνά την εκφράζουμε και ως ποσοστό. Είναι φανερό ότι το άθροισμα των σχετικών συχνοτήτων όλων των κλάσεων είναι ίσο με τη μονάδα. Με τον ίδιο τρόπο με τον οποίο σχηματίσαμε την κατανομή συχνότητας των κλάσεων, μπορούμε να σχηματίσουμε και την κατανομή της σχετικής συχνότητας. Στη συγκεκριμένη περίπτωση αυτή είναι εντελώς φανερή, επειδή το πλήθος των τιμών είναι 100.

Αθροιστική συχνότητα μιας κατανομής μέχρι το ανώτερο όριο μιας κλάσης λέμε το άθροισμα των συχνοτήτων όλων των κλάσεων μέχρι αυτό το όριο. Στον πίνακα 1.3 παρουσιάζουμε την αθροιστική συχνότητα των βαθμών του πίνακα 1.1.

Σχετική αθροιστική συχνότητα μιας κατανομής μέχρι το ανώτερο όριο μιας κλάσης λέμε το λόγο της αθροιστικής συχνότητας μέχρι το όριο αυτό προς το άθροισμα των συχνοτήτων όλων των κλάσεων.

Πίνακας 1.3. Αθροιστική συχνότητα προαγωγικών βαθμών 100 φοιτητών.

Βαθμοί	Αριθμός φοιτητών
4,95	0
5,45	2
5,95	13
6,45	29
6,95	63
7,45	83
7,95	97
8,45	100

Η σχετική αθροιστική συχνότητα μέχρι το ανώτερο όριο μιας κλάσης της αθροιστικής συχνότητας του πίνακα 1.3 είναι ολοφάνερη, επειδή το πλήθος των τιμών του παραδείγματος είναι συμπτωματικά 100.

● **Παράδειγμα 1.1:** Στον παρακάτω πίνακα δίνονται οι μετρήσεις των παραχθέντων προϊόντων σε μια βιομηχανία ανά εργατοώρα σε 60 εργάσιμες ημέρες.

59	77	89	52	73	77	63	73	57	56
56	67	69	70	73	62	65	65	92	71
41	49	75	75	96	79	53	55	61	61
83	81	81	45	73	94	58	67	67	69
69	71	69	77	77	81	87	65	67	91
71	63	51	73	83	89	93	57	60	59

Ζητείται να καθοριστούν τα όρια των κλάσεων της κατανομής συχνότητας των τιμών αυτών.

Κατ' αρχήν εύκολα διαπιστώνουμε ότι η μικρότερη τιμή στον πίνακα είναι η 41 και η μεγαλύτερη η 96. Επειδή το πλήθος των τιμών είναι σχετικά μικρό, θα χρησιμοποιήσουμε μικρό αριθμό κλάσεων, έστω 6.

Έτσι, έχουμε πλάτος κλάσεων $\frac{96\text{-}41}{6} = 9{,}2$. Για λόγους απλότητας στους υπολογισμούς μας επιλέγουμε πλάτος κλάσεων ίσο με 10. Στη συνέχεια επιλέγουμε το προσωρινό κατώτερο όριο της πρώτης κλάσης. Αυτό μπορεί να είναι οποιοσδήποτε αριθμός μικρότερος ή ίσος με το 41. Επιλέγουμε τον αριθμό 40 ως το κατώτερο όριο της πρώτης κλάσης. Συνεπώς, 50 είναι το κατώτερο όριο της δεύτερης κλάσης κ.ο.κ. Άρα, οι κλάσεις είναι οι 40-49, 50-59, 60-69, 70-79, 80-89 και 90-99. Το ανώτερο όριο της πρώτης κλάσης είναι το 49,5 οπότε το κατώτερο όριό της το 39,5. Τελικά λοιπόν τα πραγματικά όρια των κλάσεων που μπορούμε να χρησιμοποιήσουμε είναι τα 39,5-49,5-59,5, κ.ο.κ. Και βέβαια, ο συμβολισμός τους είναι ο 40-49, 50-59, κτλ.

1.4.2. Γραφική παράσταση

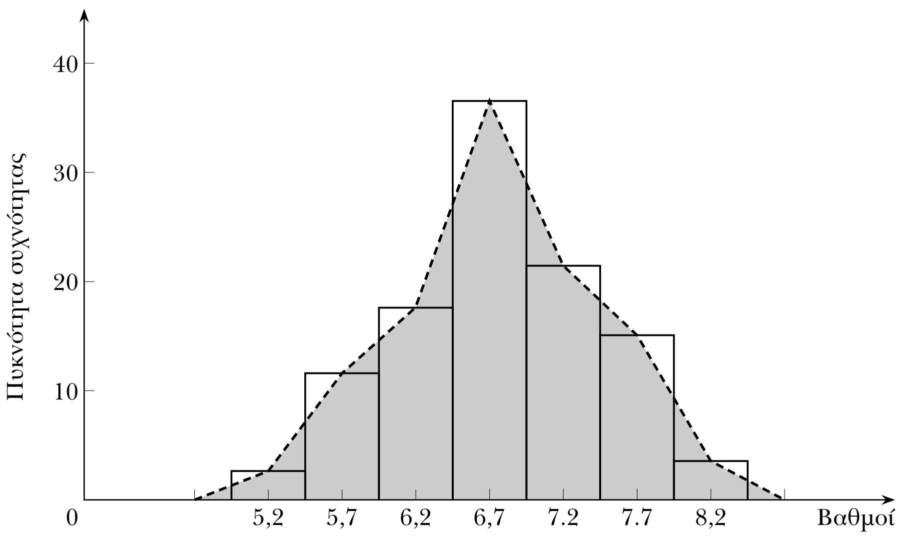
◆ **Περίπτωση συνεχούς μεγέθους.** Στην προηγούμενη παράγραφο εξετάσαμε τον τρόπο με τον οποίο μπορούμε να παρουσιάσουμε τις κατανομές συχνότητας με τη βοήθεια ενός πίνακα. Στη συνέχεια ας δούμε πώς μπορούμε να παραστήσουμε μία κατανομή συχνότητας γραφικά.

Θεωρούμε δύο ορθογωνικούς άξονες. Στον οριζόντιο άξονα παίρνουμε τις τιμές του στατιστικού μεγέθους, ενώ στον κατακόρυφο τις συχνότητες, εννοείται με κατάλληλες κλίμακες. Για κάθε κλάση –ανεξαρτήτως του πλάτους της– δημιουργούμε ένα ορθογώνιο ως εξής: α) τη βάση του ορθογωνίου την παίρνουμε ίση με το πλάτος της κλάσης και την τοποθετούμε στον οριζόντιο άξονα έτσι ώστε το μέσο της να συμπίπτει με το μέσο της κλάσης, και β) το ύψος του ορθογωνίου το επιλέγουμε έτσι ώστε το εμβαδόν του να είναι ανάλογο προς τη συχνότητα της κλάσης. Για να εκφράζει το εμβαδό συχνότητα πρέπει να ισχύει η σχέση

$$\mathrm{E_i} = \Pi_\mathrm{i} \cdot \frac{\mathrm{f_i}}{\Pi_\mathrm{i}} = \mathrm{f_i}\,, \qquad (1.2)$$

όπου $\mathrm{E_i}$, Π_i και $\mathrm{f_i}$ είναι το εμβαδόν του ορθογωνίου, το πλάτος της βάσης του και η συχνότητα της κλάσης i αντίστοιχα. Δηλαδή, το ύψος του ορθογωνίου πρέπει πάντοτε να το εκφράζουμε από άποψη διαστάσεων σε συχνότητα ανά μονάδα του στατιστικού μεγέθους· επομένως σε *πυκνότητα συχνότητας*. Στην περίπτωση που οι κλάσεις έχουν ίδιο πλάτος, τα ύψη των ορθογωνίων είναι ανάλογα με τις συχνότητες των κλάσεων, και συνηθίζουμε να τα παίρνουμε αριθμητικά ίσα με τις συχνότητες των κλάσεων, όπως έχουμε κάνει στο σχήμα 1.1.

Το σχήμα που δημιουργούμε με τον τρόπο που περιγράψαμε το λέμε *ιστόγραμμα*. Στο σχήμα 1.1 φαίνεται το ιστόγραμμα της κατανομής συχνότητας των βαθμών του πίνακα 1.2. Αν συμπεριλάβουμε στο ιστόγραμμα και τις δύο ακραίες κλάσεις που έχουν μηδενική συχνότητα σαν ορθογώνια με ύψος ίσο με το μηδέν και ενώσουμε τα μέσα των πάνω πλευρών όλων των ορθογωνίων με ευθύγραμμα τμήματα, δημιουργούμε το *πολύγωνο συχνότητας*. Με αυτό τον τρόπο, το άθροισμα της επιφάνειας των ορθογωνίων είναι ίσο με την επιφάνεια που περικλείεται ανάμεσα στην πολυγωνική γραμμή και τον οριζόντιο άξονα. Αυτό ισχύει όταν τα πλάτη των κλάσεων είναι ίσα. Όταν όμως δεν είναι ίσα, οι επιφάνειες που προσθέτουμε και αφαιρούμε αντιστοιχούν σε ίσες συχνότητες.



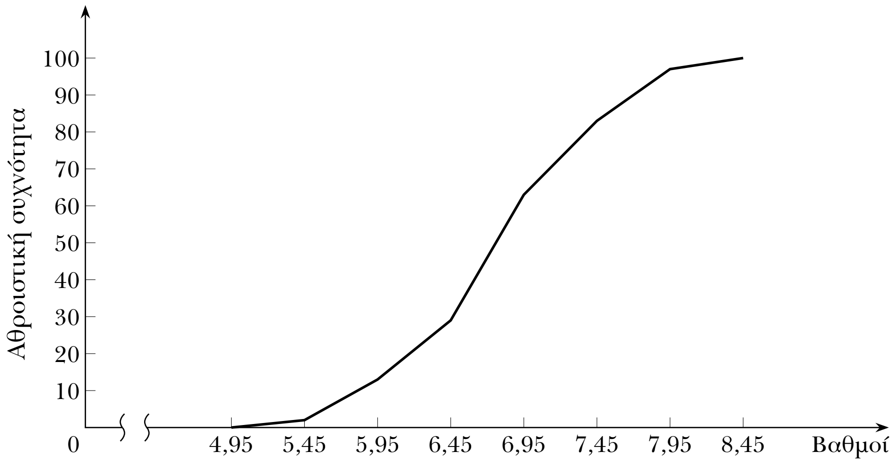
Σχήμα 1.1. Ιστόγραμμα και πολύγωνο συχνότητας προαγωγικών βαθμών 100 φοιτητών.

Όπως σχηματίσαμε το ιστόγραμμα και το πολύγωνο συχνότητας μπορούμε να σχηματίσουμε και τα αντίστοιχα της σχετικής συχνότητας, αρκεί αντί της κλίμακας πυκνότητας συχνότητας να χρησιμοποιήσουμε κλίμακα πυκνότητας σχετικής συχνότητας. Σ' αυτή την περίπτωση το συνολικό εμβαδό του ιστογράμματος είναι πάντοτε ίσο με τη μονάδα. Κι αυτό γιατί είναι το άθροισμα των εμβαδών όλων των ορθογωνίων που καθένα τους είναι ίσο με τη σχετική συχνότητα της αντίστοιχης κλάσης.

Για να κατασκευάσουμε το ιστόγραμμα και το πολύγωνο της σχετικής συχνότητας στην περίπτωση που έχουμε ήδη κατασκευάσει το ιστόγραμμα και το πολύγωνο συχνότητας, χρειάζεται να αλλάξουμε μόνο την κλίμακα πυκνότητας συχνότητας σε κλίμακα πυκνότητας σχετικής συχνότητας. Τα διαγράμματα παραμένουν τα ίδια.

Την αθροιστική συχνότητα την παριστάνουμε γραφικά με το αθροιστικό διάγραμμα των συχνοτήτων. Το διάγραμμα αυτό μας δίνει, για κάθε τιμή, τη συχνότητα των τιμών των στοιχείων μέχρι την τιμή αυτή. Για να κατασκευάσουμε το αθροιστικό διάγραμμα, στα σημεία του οριζόντιου άξονα, που αντιστοιχούν στα ανώτερα όρια των κλάσεων, παίρνουμε τεταγμένες ίσες με την αθροιστική συχνότητα μέχρι του συγκεκριμένου ορίου. Στη συνέχεια ενώνουμε τα άκρα των τεταγμένων με ευθείες. Έτσι, το διάγραμμα μας δίνει με ικανοποιητική προσέγγιση την αθροιστική συχνότητα και για ενδιάμεσες τιμές των κλάσεων. Η προσέγγιση βέβαια είναι τόσο καλύτερη, όσο μικρότερο είναι το πλάτος των κλάσεων και όσο οι τιμές κατανέμονται ομοιόμορφα μέσα στην κλάση.

Στο σχήμα 1.2 παρουσιάζουμε το αθροιστικό διάγραμμα συχνοτήτων της αθροιστικής συχνότητας των βαθμών των 100 φοιτητών του πίνακα 1.3.



Σχήμα 1.2. Διάγραμμα αθροιστικής συχνότητας προαγωγικών βαθμών 100 φοιτητών.

Διάγραμμα αθροιστικής συχνότητας μπορούμε να κατασκευάσουμε και από το ιστόγραμμα των συχνοτήτων. Γι’ αυτό αρκεί οι τεταγμένες του αθροιστικού διαγράμματος σε κάποια τιμή να εκφράζουν το εμβαδό του ιστογράμματος μέχρι αυτή τη τιμή.

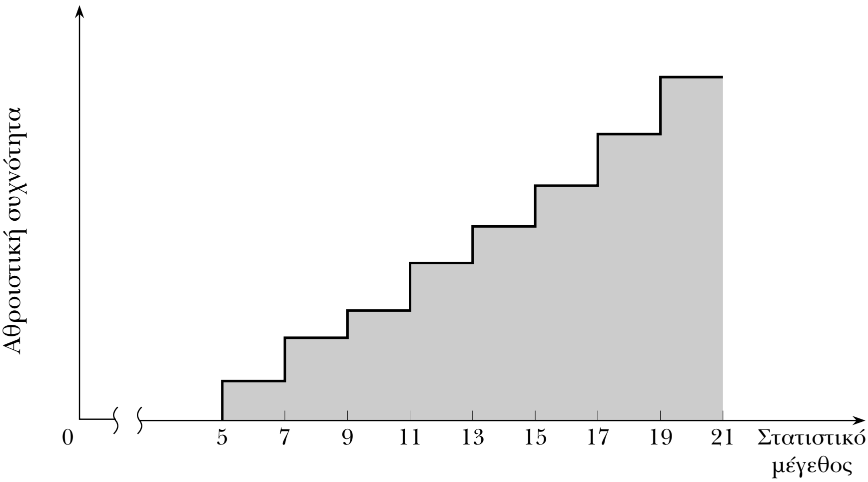
Το διάγραμμα της σχετικής αθροιστικής συχνότητας μπορούμε να το κατασκευάσουμε όπως κατασκευάσαμε το διάγραμμα της αθροιστικής συχνότητας.

Υπάρχουν εξάλλου περιπτώσεις που χρησιμοποιούμε την αθροιστική συχνότητα από του κατώτερου ορίου μιας κλάσης και πέρα αντί της αθροιστικής συχνότητας μέχρι του ανώτερου ορίου μιας κλάσης. Το αθροιστικό διάγραμμα σ' αυτή την περίπτωση μπορούμε να το κατασκευάσουμε ακριβώς όπως περιγράψαμε πρωτύτερα.

◆ **Περίπτωση ασυνεχούς μεγέθους.** Στην παράγραφο 1.1 ορίσαμε ως ασυνεχές στατιστικό μέγεθος το μέγεθος που παίρνει μόνο διακεκριμένες τιμές σε μία περιοχή τιμών.

Το γεγονός ότι το στατιστικό μέγεθος παίρνει μόνο διακεκριμένες τιμές δεν πρέπει να μας δημιουργεί την εντύπωση ότι η γραφική παράσταση της κατανομής της συχνότητας ενός τέτοιου μεγέθους δεν μπορεί παρά να είναι μόνο ορισμένες κάθετες γραμμές στα σημεία που αντιστοιχούν στις τιμές του και να έχουν μήκος ίσο με την συχνότητα εμφανίσεως της αντίστοιχης τιμής του. Στην πραγματικότητα μπορούμε να κατασκευάσουμε ιστόγραμμα και τιμών ασυνεχούς μεγέθους, και το εμβαδόν του να είναι ανάλογο με την συχνότητα, όπως και στην περίπτωση των τιμών των συνεχών μεγεθών. Αυτό το πετυχαίνουμε αν τις διακεκριμένες τιμές τις οποίες παίρνει το ασυνεχές μέγεθος τις θεωρήσουμε ότι έχουν την έννοια των μέσων των κλάσεων των τιμών των συνεχών μεγεθών.

Είναι φανερό ότι με αυτή τη θεώρηση, τις τιμές του ασυνεχούς μεγέθους μπορούμε να τις χρησιμοποιήσουμε σαν τιμές ενός συνεχούς μεγέθους, με την παρακάτω μόνο διαφορά. Για να δημιουργήσουμε το αθροιστικό διάγραμμα των τιμών ενός συνεχούς μεγέθους είπαμε ότι συνδέουμε τα άκρα των τεταγμένων στα ανώτερα όρια των κλάσεων με ευθείες γραμμές, και θεωρούμε ότι το διάγραμμα που δημιουργείται με αυτόν τον τρόπο δίνει με ικανοποιητική προσέγγιση την αθροιστική συχνότητα και για τις ενδιάμεσες τιμές των κλάσεων. Βέβαια, η υπόθεση αυτή δεν μπορεί να γίνει με τις τιμές των ασυνεχών μεγεθών. Γι' αυτό το διάγραμμα της αθροιστικής συχνότητας ή σχετικής συχνότητας ενός ασυνεχούς μεγέθους πρέπει να το δημιουργούμε ως εξής: στα ανώτερα όρια των κλάσεων δημιουργούμε τεταγμένες ίσες με την αθροιστική συχνότητα, ή σχετική συχνότητα, μέχρι εκείνου του ορίου. Επειδή η συχνότητα αυτή διατηρείται σταθερή μέχρι του ανώτερου ορίου της θεωρούμενης κλάσης φέρουμε ένα τμήμα γραμμής παράλληλο προς τον άξονα των x μέχρι αυτού του ορίου. Και συνεχίζουμε κατ' αυτόν τον τρόπο. Το αποτέλεσμα είναι ένα κλιμακωτό διάγραμμα της μορφής του σχήματος 1.3.



Σχήμα 1.3. Διάγραμμα αθροιστικής συχνότητας ασυνεχούς στατιστικού μεγέθους.

1.4.3. Κατανομές συχνότητας διsδιάστατων στατιστικών μεγεθών

Στις προηγούμενες παραγράφους ασχοληθήκαμε με τις κατανομές συχνότητας ενός στατιστικού μεγέθους και τον τρόπο με τον οποίο τις εκφράζουμε είτε με τη βοήθεια ενός πίνακα, είτε με γραφικές παραστάσεις. Κι όλα αυτά βέβαια τα κάνουμε για να μελετήσουμε κάποιο χαρακτηριστικό κάποιου στατιστικού πληθυσμού. Ο οποιοσδήποτε όμως στατιστικός πληθυσμός δεν είναι απαραίτητο να έχει μόνο ένα χαρακτηριστικό. Μπορεί να έχει και περισσότερα από ένα και να μας ενδιαφέρει όχι μόνο η μελέτη ενός από αυτά. Αν έχει μεν περισσότερα, αλλά μας ενδιαφέρει να μελετήσουμε μόνο ένα χαρακτηριστικό, τότε κάνουμε όσα εξηγήσαμε στις παραπάνω παραγράφους. Τα ίδια εξάλλου μπορούμε να κάνουμε όταν μας ενδιαφέρει να μελετήσουμε περισσότερα από ένα χαρακτηριστικά ενός στατιστικού πληθυσμού, που όμως είναι ανεξάρτητα μεταξύ τους.

Συχνά όμως χρειάζεται να μελετήσουμε κατά πόσο δύο ή περισσότερα χαρακτηριστικά ενός στατιστικού πληθυσμού είναι ανεξάρτητα μεταξύ τους. Κι αν δεν είναι ανεξάρτητα να προσδιορίσουμε την εξάρτηση που υπάρχει μεταξύ τους.

◆ **Διάγραμμα διασποράς.** Για να αναγνωρίσουμε γραφικά κατά πόσο υπάρχει εξάρτηση μεταξύ δύο μεγεθών σχεδιάζουμε το *διάγραμμα διασποράς*.

Για να σχεδιάσουμε ένα διάγραμμα διασποράς, όταν έχουμε ένα σύνολο δεδομένων δύο στατιστικών μεγεθών, συμβολίζουμε το ένα με Χ και το άλλο με Υ. Κάθε ζεύγος τιμών των Χ και Υ το αντιστοιχούμε σ’ ένα σημείο στη γραφική μας απεικόνιση.

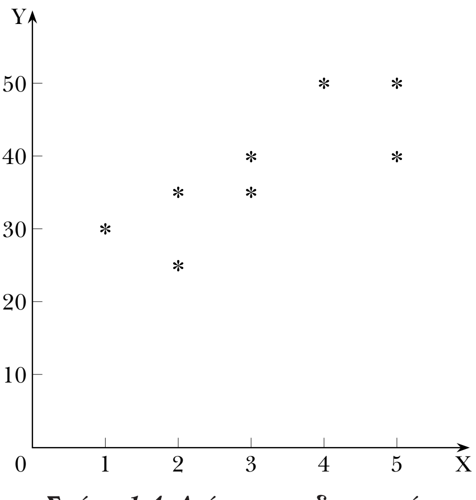
Έστω λοιπόν ότι έχουμε το παρακάτω παράδειγμα.

Μια μικρή επιχείρηση είχε σχετικά ομοιόμορφο όγκο πωλήσεων τα προηγούμενα χρόνια. Το τρέχον έτος αποφάσισε να μεταβάλει τη δαπάνη διαφήμισης από μήνα σε μήνα για να διαπιστώσει αν αυτό έχει σημαντική επίδραση στον όγκο πωλήσεών της. Για να βοηθηθεί στην εκτίμηση του αποτελέσματος που έχει η διαφήμιση στον όγκο των πωλήσεων συνέλεξε τα στοιχεία που σημειώνονται στον πίνακα 1.4.

Πίνακας 1.4. Δεδομένα παραδείγματος.

Μήνας	**Διαφημιστική δαπάνη (X)** *(εκατομμύρια δραχμές)*	**Αξία πωλήσεων (Y)** *(εκατομμύρια δραχμές)*
1	1	30
2	3	40
3	5	40
4	4	50
5	2	35
6	5	50
7	3	35
8	2	25

Αν σχεδιάσουμε σ' ένα διάγραμμα, στο οποίο Y είναι ο κατακόρυφος άξονας και X ο οριζόντιος, τα 8 ζεύγη τιμών, έχουμε την εικόνα που φαίνεται στο σχήμα 1.4. Αυτό είναι το διάγραμμα διασποράς των ζευγών των τιμών των X και Y.



Σχήμα 1.4. Διάγραμμα διασποράς.

Από τη μορφή του διαγράμματος διασποράς προκύπτουν δύο πληροφορίες αναφορικά προς τη σχέση των δύο μεγεθών. Κατ' αρχήν, παρατηρούμε ότι ο όγκος των πωλήσεων αυξάνει γενικά καθώς αυξάνει η διαφημιστική δαπάνη. Οσάκις, δύο μεγέθη μεταβάλλονται όπως αυτά του παραδείγματός μας, δηλαδή οι τιμές τους τείνουν να αυξάνονται μαζί και να μειώνονται μαζί λέμε ότι υπάρχει *θετική σχέση* μεταξύ των δύο μεγεθών.

Η δεύτερη παρατήρηση είναι ότι μοιάζει η σχέση μεταξύ των δύο μεγεθών να είναι *γραμμική*. Αν και τα οκτώ σημεία δεν βρίσκονται σ' ευθεία γραμμή, μπορούμε να φαντασθούμε ότι σχεδιάζουμε μια ευθεία γραμμή ανάμεσα στα σημεία του διαγράμματος διασποράς η οποία προσεγγίζει ικανοποιητικά τη θετική σχέση μεταξύ των δύο μεγεθών. Επ' αυτού του θέματος επανερχόμαστε αναλυτικά στο έκτο κεφάλαιο.

Στις περιπτώσεις που μας απασχολούν περισσότερα στατιστικά μεγέθη ενός στατιστικού πληθυσμού μπορούμε να γενικεύσουμε όσα είπαμε για ένα μέγεθος και να χειρισθούμε τα σχετικά προβλήματα που δημιουργούνται όταν έχουμε δύο ή περισσότερα μεγέθη. Εδώ θα περιοριστούμε για καθαρά πρακτικούς λόγους μόνο στην περίπτωση που έχουμε δύο μεγέθη στην οποία εξάλλου μπορούμε να έχουμε γραφικές απεικονίσεις.

Είναι φανερό ότι μπορούμε να έχουμε δύο μεγέθη (δισδιάστατο μέγεθος) είτε συνεχή είτε ασυνεχή ανάλογα με τις τιμές που μπορούν να πάρουν. Μπορούμε επίσης την κατανομή συχνότητας να την παραστήσουμε είτε με τη μορφή ενός πίνακα είτε με ιστογράμματα.

◆ **Παράσταση με τη βοήθεια πίνακα.** Κατ' αρχήν σχηματίζουμε τις κλάσεις παίρνοντας υπόψη μας όσα είπαμε στην παράγραφο 1.4.1. Έχοντας τις κλάσεις για κάθε στατιστικό μέγεθος μπορούμε να σχηματίσουμε τους *πίνακες διπλής εισόδου*.

Στους πίνακες αυτούς καταχωρούμε τις συχνότητες ή τις σχετικές συχνότητες $f(x_i, y_j) = f_{ij}$ των τιμών του στατιστικού μεγέθους X που ανήκουν στην κλάση (α_{i-1}, α_i) και των τιμών του μεγέθους Y που ανήκουν στην κλάση (β_{j-1}, β_j).

Στο δεξιό περιθώριο του πίνακα καταχωρούμε τα αθροίσματα των συχνοτήτων

$$f_{i\bullet} = \sum_{y_j} f(x_i, y_j), \tag{1.3}$$

ενώ στο κάτω περιθώριο τα αθροίσματα

$$f_{\bullet j} = \sum_{x_i} f(x_i, y_j) \tag{1.4}$$

Πίνακας 1.5. Κατανομή συχνότητας δισδιάστατου στατιστικού μεγέθους.

X \ Y		(β_0, β_1)	(β_1, β_2) …	(β_{j-1}, β_j) …	$(\beta_{\lambda-1}, \beta_\lambda)$	**Σύνολο**
		y_1	y_2	y_j	y_λ	
(α_0, α_1)	x_1	$f(x_1, y_1)$	$f(x_1, y_2)$	$f(x_1, y_j)$	$f(x_1, y_\lambda)$	$f_{1.}$
(α_1, α_2)	x_2	$f(x_2, y_1)$	$f(x_2, y_2)$	$f(x_2, y_j)$	$f(x_2, y_\lambda)$	$f_{2.}$
⋮	⋮	⋮	⋮	⋮	⋮	
(α_{i-1}, α_i)	x_i	$f(x_i, y_1)$	$f(x_i, y_2)$	$f(x_i, y_j)$	$f(x_i, y_\lambda)$	$f_{i.}$
⋮	⋮	⋮	⋮	⋮	⋮	
(α_{k-1}, α_k)	x_k	$f(x_k, y_1)$	$f(x_k, y_2)$	$f(x_k, y_j)$	$f(x_k, y_\lambda)$	$f_{k.}$
Σύνολο		$f_{.1}$	$f_{.2}$	$f_{.j}$	$f_{.\lambda}$	$f_{..}$

Τις συχνότητες αυτές τις λέμε *περιθωριακές κατανομές συχνότητας*. Ο χαρακτηρισμός τους ως περιθωριακές γίνεται απλά και μόνο γιατί τις καταγράφουμε στα περιθώρια του πίνακα.

◆ **Αθροιστική συχνότητα.** Πολλές φορές, για πρακτικούς λόγους, χρειάζεται να υπολογίσουμε για τις δισδιάστατες μεταβλητές τις αθροιστικές κατανομές συχνοτήτων που τις βρίσκουμε όπως ακριβώς βρήκαμε και τις αθροιστικές των μονοδιάστατων μεγεθών.

Έτσι, τις περιθωριακές αθροιστικές συχνότητες μέχρι τις τιμές x_i και y_j αντίστοιχα τις προσδιορίζουμε από τις σχέσεις

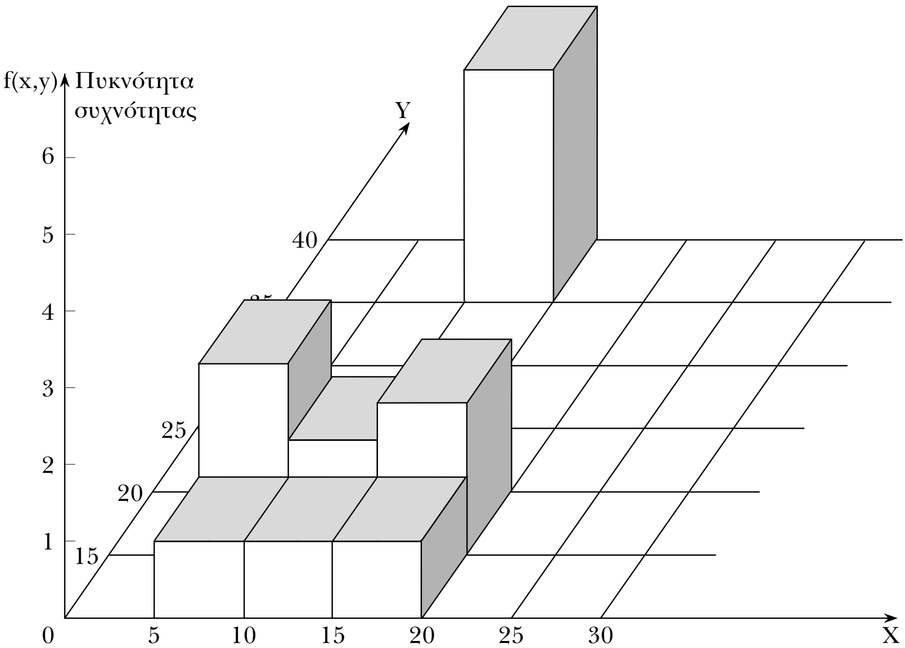
$$F_{i.} = \sum_{x_i \le t_i} \sum_{y_j} f(x_i, y_j), \tag{1.5}$$

$$F_{.j} = \sum_{x_i} \sum_{y_j \le s_j} f(x_i, y_j), \tag{1.6}$$

και τη συνολική αθροιστική συχνότητα μέχρι τις τιμές x_i και y_j από τη

$$F_{ij} = \sum_{x_i \le t_i} \sum_{y_j \le s_j} f(x_i, y_j) \tag{1.7}$$

◆ **Γραφική παράσταση.** Στο σχήμα 1.5 φαίνεται ένα ιστόγραμμα συχνότητας ενός δισδιάστατου στατιστικού μεγέθους, που ο τρόπος σχηματισμού του είναι εντελώς ανάλογος με εκείνο που χρησιμοποιούμε για τα μονοδιάστατα στατιστικά μεγέθη. Δηλαδή, έχουμε πυκνότητα συχνότητας, κτλ. αλλά είναι συχνότητα ανά μονάδα επιφάνειας του δισδιάστατου μεγέθους κτλ. Και βέβαια ισχύουν αναλογικά όλα όσα είπαμε για τα ιστογράμματα συχνοτήτων ή σχετικών συχνοτήτων, για τα διαγράμματα των αθροιστικών συχνοτήτων, ή σχετικών αθροιστικών κτλ. στα μονοδιάστατα μεγέθη και για τα δισδιάστατα μεγέθη.



Σχήμα 1.5. Γραφική παράσταση κατανομής συχνότητας δισδιάστατου στατιστικού μεγέθους.

Παράδειγμα 1.2: Δίνονται τα δεδομένα του πίνακα 1.4 και θέλουμε να σχεδιάσουμε το ιστόγραμμα της κατανομής της διαφημιστικής δαπάνης και της αξίας πωλήσεων.

Κατ' αρχήν από τα στοιχεία του πίνακα 1.4 μπορούμε να σχηματίσουμε τις κατανομές των συχνοτήτων των X και Y, όπως φαίνεται στους πίνακες 1.6 και 1.7.

Στη συνέχεια μπορούμε να σχηματίσουμε την κατανομή συχνότητας των X και Y, όπως φαίνεται στον πίνακα 1.8. Τις κλάσεις τόσο στους πίνακες 1.6 και 1.7, όσο και στον 1.8 τις σημειώνουμε με τον συμβολισμό τους, όπως τον εξηγήσαμε στην παράγραφο 1.4.1, ενώ οι πραγματικές κλάσεις είναι 0-1, 1-2, 2-3, κτλ. και 00-10, 10-20, 20-30, 30-40, κ.ο.κ.

Πίνακας 1.6. Κατανομή συχνότητας διαφημιστικής δαπάνης (Χ).

Κλάσεις (εκ. δρχ.)	**Συχνότητες**
0,1-0,9	0
1,1-1,9	2
2,1-2,9	4
3,1-3,9	0
4,1-4.9	2
5,1-5,9	0
Σύνολο	**8**

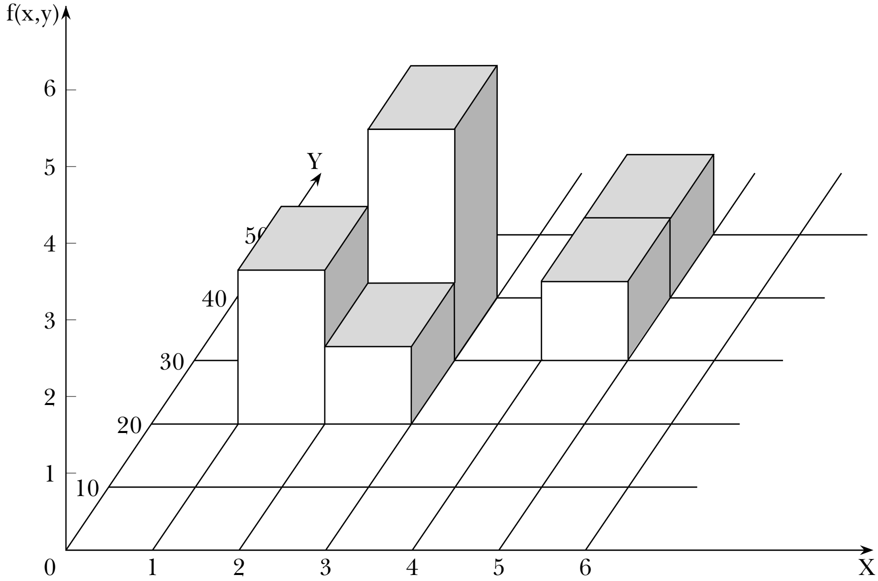
Πίνακας 1.7. Κατανομή συχνότητας αξίας πωλήσεων (Υ).

Κλάσεις (εκ. δρχ.)	**Συχνότητες**
1-9	0
11-19	0
21-29	3
31-39	4
41-49	1
Σύνολο	**8**

Πίνακας 1.8. Δισδιάστατη κατανομή συχνότητας των Χ και Υ.

X \ Y	**0,1-0,9**	**1,1-1,9**	**2,1-2,9**	**3,1-3,9**	**4,1-4,9**	**5,1-5,9**	**f(Y)**
1-9							**0**
11-19							**0**
21-29		2	1				**3**
31-39			3		1		**4**
41-49					1		**1**
f(X)	**0**	**2**	**4**	**0**	**2**	**0**	**8**

Από τη δισδιάστατη κατανομή συχνότητας μπορούμε εύκολα να παρατηρήσουμε ότι 2 φορές αξία πωλήσεων από 20 έως 30 εκ. επετεύχθει με διαφημιστική δαπάνη από 1 έως 2 εκ., κ.ο.κ.



Σχήμα 1.6. Γραφική παράσταση κατανομής δισδιάστατου μεγέθους.

1.5. Μορφές κατανομών συχνότητας

Είναι αξιοσημείωτο ότι οι γραφικές παραστάσεις των τιμών των στατιστικών μεγεθών μπορούν να ομαδοποιηθούν στις παρακάτω μορφές ιστογραμμάτων: συμμετρικό, ασύμμετρο, σχήματος J, σχήματος U, και άλλα.

Ας δούμε στη συνέχεια ορισμένα παραδείγματα στατιστικών μεγεθών που τα ιστογράμματα των κατανομών τους ανήκουν σε κάποια από τις παραπάνω μορφές:

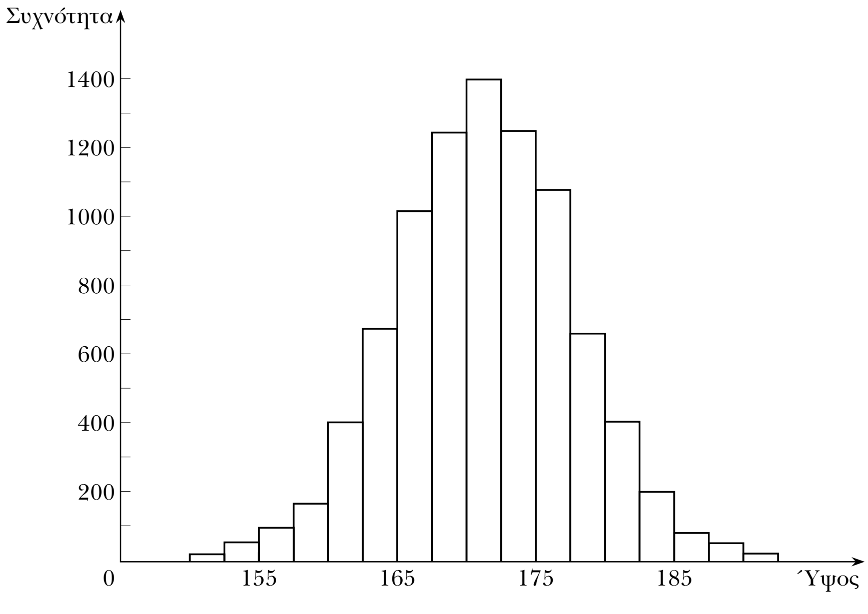
◆ **Κατανομή που ακολουθούν τα ύψη των ανδρών**

Στον πίνακα 1.9 σημειώνουμε την κατανομή συχνότητας του ύψους 8668 ανδρών.

Το ιστόγραμμα της κατανομής φαίνεται στο σχήμα 1.7. Είναι φανερό ότι αυτό είναι *συμμετρικό*.

Πίνακας 1.9. Παράδειγμα συμμετρικής κατανομής.

Ύψος (cm)	**Αριθμός ανδρών**	**Ύψος (cm)**	**Αριθμός ανδρών**
145	2	171,5	1230
147,5	4	175	1063
150	14	177,5	646
152,5	41	180	392
155	83	182,5	202
157,5	169	185	79
160	394	187,5	52
162,5	669	190	16
165	990	192,5	5
167,5	1223	195	2
170	1392		



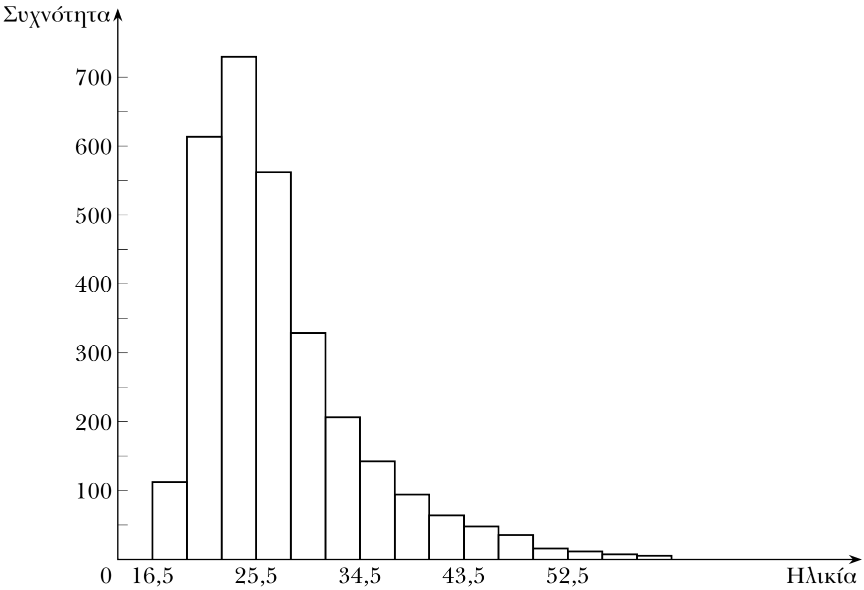
Σχήμα 1.7. Ιστόγραμμα των υψών των ανδρών.

Κατανομή που ακολουθούν οι ηλικίες γάμου των ανδρών

Στον πίνακα 1.10 φαίνεται η κατανομή συχνότητας της ηλικίας γάμου των ανδρών. Η μορφή του ιστογράμματος (βλ. σχήμα 1.8) είναι *ασυμμετρική*.

Πίνακας 1.10. Παράδειγμα ασυμμετρικής κατανομής

Ηλικία ανδρών	Αριθμός γάμων	Ηλικία ανδρών	Αριθμός γάμων
16,5	3	49,5	36
19,5	110	52,5	21
22,5	610	55,5	16
25,5	730	58,5	11
28,5	565	61,5	8
31,5	334	64,5	6
34,5	205	67,5	5
37,5	142	70,5	3
40,5	93	73,5	2
43,5	62	76,5	1
46,5	47		



Σχήμα 1.8. Ιστόγραμμα ηλικιών γάμου ανδρών.

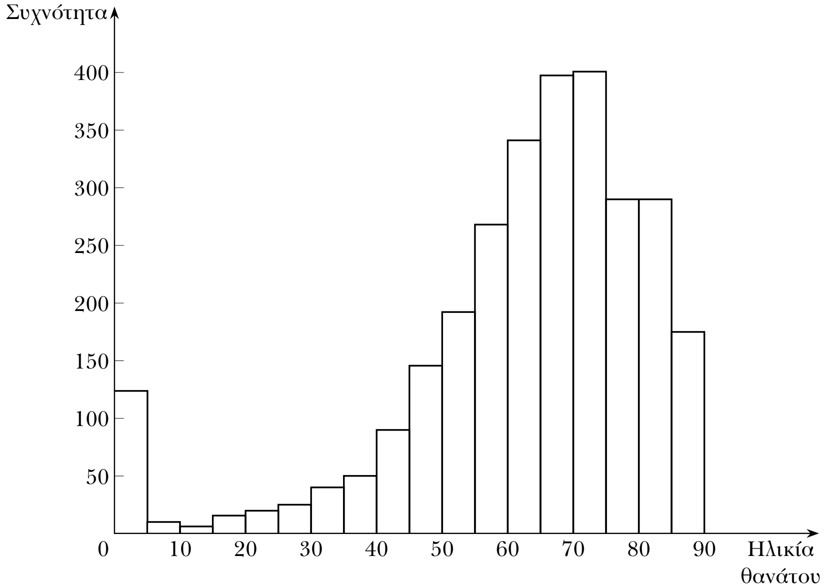
Τη μορφή της κατανομής του σχήματος 1.8 τη λέμε ασυμμετρική με θετική λοξότητα, γιατί το μεγαλύτερο άκρο της είναι προς τη θετική κατεύθυνση του άξονα των x. Με τη λοξότητα θα ασχοληθούμε στην παράγραφο 1.7.4.

◆ Κατανομή θνησιμότητας

Την κατανομή του σχήματος 1.9, που ανήκει στις διάφορες μορφές, μπορούμε να την θεωρήσουμε σαν επαλληλία δύο κατανομών. Η μια είναι μία κα-

Πίνακας 1.11. Παράδειγμα κατανομής.

Ηλικία	Αριθμός θανάτων	Ηλικία	Αριθμός θανάτων
0	122	45	90
5	10	50	145
10	6	55	192
15	11	60	268
20	16	65	342
25	19	70	396
30	24	75	401
35	40	80	290
40	51	85 και άνω	175



Σχήμα 1.9. Ιστόγραμμα θνησιμότητας.

τανομή σχήματος αντίστροφης J, που απεικονίζει τη βρεφική θνησιμότητα, και η άλλη είναι μια ασύμμετρη με αρνητική λοξότητα που απεικονίζει τη φυσιολογική θνησιμότητα.

Η κατανομή αυτή έχει ιδιαίτερη σημασία γιατί αποτελεί τη βάση τόσο για τις εκτιμήσεις που γίνονται προκειμένου να σχεδιαστούν διάφορα σχήματα ασφάλειας ζωής, όσο και για τη μελέτη διαφόρων προβλημάτων συντήρησης και αντικατάστασης των τεχνολογικών συστημάτων. Η εξήγηση για τη δεύτερη χρήση της βρίσκεται στο γεγονός ότι οι ηλικίες των τεχνολογικών εξαρτημάτων ή και ολόκληρων συσκευών ακολουθούν αυτή την κατανομή, αν τις εκφρασουμε σε ώρες λειτουργίας τους ή σε άλλες κατάλληλες μονάδες μέτρησης της ηλικίας στην οποία παθαίνουν βλάβες*.

1.6 Χαρακτηριστικές τιμές θέσης και διασποράς

1.6.1. Γενικά

Μέχρι τώρα εξετάσαμε πώς σχηματίζουμε και απεικονίζουμε γραφικά μια κατανομή συχνότητας. Μολονότι αυτό είναι χρήσιμο και αναγκαίο στάδιο στην επεξεργασία των τιμών ενός δείγματος μας δημιουργεί ουσιαστικά μόνο τη δυνατότητα να κάνουμε υποκειμενικές συγκρίσεις σε σχέση με τις κατανομές συχνότητας.

Χρειάζεται λοιπόν να προχωρήσουμε την επεξεργασία πιο πέρα για να καταλήξουμε σε χρήσιμα συμπεράσματα. Το επόμενο βήμα που κάνουμε είναι η συμπύκνωση των τιμών του δείγματος έτσι ώστε να δημιουργείται η δυνατότητα π.χ. για κάποιες συγκρίσεις ανάμεσα στον πληθυσμό από τον οποίον προέρχεται το δείγμα και κάποιον άλλον πληθυσμό. Τη συμπύκνωση των πληροφοριών που περιέχει το δείγμα την πετυχαίνουμε με τον υπολογισμό ορισμένων χαρακτηριστικών τιμών, που δεν είναι τίποτε άλλο από αριθμητικές τιμές που εκφράζουν συγκεκριμένα χαρακτηριστικά των κατανομών.

Τις χαρακτηριστικές τιμές τις διακρίνουμε σε χαρακτηριστικές τιμές θέσης και διασποράς. Πέρα από αυτές τις δύο κατηγορίες συχνά χρησιμοποιούμε και τις χαρακτηριστικές τιμές ανώτερης τάξης με τις οποίες θα αχοληθούμε στην παράγραφο 1.7.

Οι χαρακτηριστικές τιμές θέσης καθορίζουν τη θέση της κατανομής σε σχέση με τις τιμές που παίρνει το αντίστοιχο στατιστικό μέγεθος. Η πιο συνήθης από τις χαρακτηριστικές τιμές θέσης είναι η μέση τιμή, που είναι μια τιμή περί

---

* Βλ. Jardine A.K.S., Operational Research in Maintenance, Manchester University Press Manchester, 1970, p. 29.

την οποίαν συγκεντρώνεται ένα μεγάλο μέρος των τιμών τόσο του δείγματος, όσο και του πληθυσμού από τον οποίο προέρχεται αυτό.

1.6.2. Χαρακτηριστικές τιμές θέσης

Οι χαρακτηριστικές τιμές θέσης είναι οι παρακάτω:

- η μέση τιμή,
- η μεσαία τιμή,
- η συχνότερη τιμή,
- η γεωμετρική μέση τιμή,
- η αρμονική μέση τιμή.

◆ **Μέση τιμή.** Η μέση τιμή, $\bar{x}$, των τιμών $x_1, x_2, ..., x_N$ ενός δείγματος είναι ο αριθμητικός μέσος όρος τους, και υπολογίζεται από τη σχέση:

$$\bar{x} = \frac{x_1 + x_2 + ... + x_N}{N} = \frac{\sum_{i=1}^{N} x_i}{N} . \qquad (1.8)$$

Αν καθεμιά από τις τιμές $x_1, x_2, ..., x_n$ παρουσιάζεται με συχνότητα $f_1, f_2, ..., f_n$ αντίστοιχα, τη μέση τιμή τους την υπολογίζουμε από τη σχέση:

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + ... + f_nx_n}{f_1 + f_2 + ... + f_n} = \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i} .$$

Επειδή όμως ισχύει η σχέση

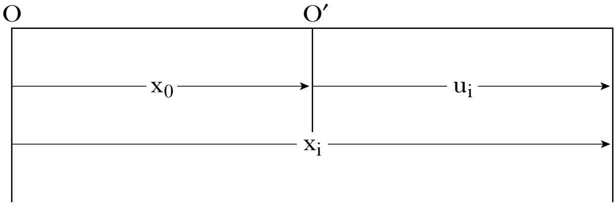
$$N = \sum_{i=1}^{n} f_i ,$$

έχουμε

$$\bar{x} = \frac{\sum_{i=1}^{n} f_ix_i}{N} . \qquad (1.9)$$

Ο υπολογισμός της μέσης τιμής από τις παραπάνω σχέσεις δεν είναι ευχερής, όταν κυρίως υπάρχουν πολλές τιμές. Γι' αυτό χρησιμοποιούμε συνήθως μία ταχύτερη διαδικασία που διευκολύνει τους υπολογισμούς, ενώ εξακολουθεί να είναι συνεπής με το μαθηματικό ορισμό της μέσης τιμής.

Συγκεκριμένα, θεωρούμε ως αρχή μέτρησης των τιμών του δείγματος την αρχή O' αντί της O, που είναι η πραγματική αρχή μέτρησής τους. Με άλλα λόγια κάνουμε μια μετατόπιση της αρχής κατά x_0 όπως φαίνεται στο σχήμα 1.10.



Σχήμα 1.10. Υπολογισμός μέσης τιμής.

Οπότε έχουμε

$$x_i = x_0 + u_i \,.$$

Αν αντικαταστήσουμε τις τιμές x_i με το $x_0 + u_i$ στη σχέση (1.9), παίρνουμε

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_0}{N} + \frac{\sum_{i=1}^{n} f_i u_i}{N} = x_0 + \frac{\sum_{i=1}^{n} f_i u_i}{N} \,, \tag{1.10}$$

και

$$\bar{x} = x_0 + \bar{u} \,. \tag{1.11}$$

Για να απλουστεύσουμε πραγματικά τη διαδικασία υπολογισμών –η χρησιμότητα αυτής της απλούστευσης θα γίνει φανερή παρακάτω– πρέπει να εκλέξουμε κατάλληλα την τιμή του x_0.

Πριν εξετάσουμε και τις άλλες χαρακτηριστικές τιμές θέσης, είναι σκόπιμο να σχολιάσουμε ορισμένες ιδιότητες της μέσης τιμής:

- Η μέση τιμή συνοψίζει την τάση μιας σειράς μετρήσεων, ή γενικά μιας κατανομής συχνότητας, και είναι η χαρακτηριστική τιμή θέσης που χρησιμοποιείται περισσότερο από όλες τις άλλες.
- Η μέση τιμή είναι η τιμή ως προς την οποία το άθροισμα των αποκλίσεων όλων των τιμών από αυτή είναι ίσο με το 0. Δηλαδή, αν α_i είναι η απόκλιση της x_i από τη $\bar{x}$, τότε

$$\sum_{i=1}^{N} \alpha_i = 0 \,.$$

Αυτό προκύπτει εύκολα ως εξής

$$\sum_{i=1}^{N} \alpha_i = \sum_{i=1}^{N} (x_i - \bar{x}) = \sum_{i=1}^{N} x_i - N\bar{x} = \sum_{i=1}^{N} x_i - N\left(\frac{\sum_{i=1}^{N} x_i}{N}\right) = 0 \,.$$