# ΚΕΦ. 2

## ΤΡΟΠΟΙ ΠΑΡΟΥΣΙΑΣΗΣ ΣΤΑΤΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Υπάρχουν 2 βασικοί τρόποι παρουσίασης των στατιστικών δεδομένων:

1. Οι πίνακες
2. Τα διαγράμματα

## 1. ΠΙΝΑΚΕΣ

Όταν μια ποσοτική μεταβλητή παίρνει περισσότερες από 30 τιμές, τότε για την καλύτερη επεξεργασία τους τις πινακοποιούμε σε ομάδες που ονομάζονται κλάσεις ή διαστήματα (τάξεις) έτσι ώστε κάθε τιμή να ανήκει σε μία μόνο κλάση. Για την κατασκευή ενός τέτοιου πίνακα ακολουθούμε τα εξής βήματα:

1.1. Επιλέγουμε τον αριθμό των κλάσεων

Εδώ μπορούμε να επιλέξουμε αυθαίρετα όσες κλάσεις θέλουμε (ανάλογα και με το εύρος των τιμών που διαθέτουμε).
Εμπειρικά το πλήθος των κλάσεων δεν θα πρέπει να υπερβαίνει τις 18. Ένας αυτόματος τρόπος επιλογής του πλήθους των κλάσεων είναι μέσω του τύπου Sturges

$\boxed{q = 1 + 3{,}32 \cdot \log_{10}(n)}$ (1) όπου $q$ το πλήθος των κλάσεων και $n$ το μέγεθος του δείγματος. Έτσι προκύπτει ο ακόλουθος πίνακας:

| $n$ | $q$ |
|---|---|
| $0 \le n \le 19$ | 5 |
| $20 \le n \le 49$ | 6 |
| $50 \le n \le 99$ | 7 |
| $100 \le n \le 199$ | 8 |
| $200 \le n \le 399$ | 9 |
| $400 \le n \le 699$ | 10 |
| $700 \le n \le 999$ | 11 |
| $n \geq 1000$ | 12 |

1.2. Υπολογίζουμε το πλάτος κάθε κλάσης

Το πλάτος κάθε κλάσης μπορεί να ορισθεί αυθαίρετα. Ωστόσο αν όλες οι κλάσεις έχουν το ίδιο πλάτος, αυτό διευκολύνει τις στατιστικές αναλύσεις. Για τον λόγο αυτό χρησιμο-

Ποιούμε τον τύπο

$$\boxed{\text{πλάτος κλάσης } c = \frac{\text{εύρος τιμών } R}{\text{αριθμός κλάσεων } q}} \quad (2)$$

όπου εύρος τιμών $R = X_{max} - X_{min} =$ μεγαλύτερη − μικρότερη παρατήρηση (3)

Αν ο c προκύψει δεκαδικός αριθμός, τότε στρογγυλοποιούμε πάντα προς τα πάνω.

1.3. Καθορίζουμε τα όρια κάθε κλάσης.

Ο καθορισμός των ορίων πρέπει να γίνει έτσι ώστε και να καλύπτονται όλες οι τιμές και να μην υπάρχουν επικαλύψεις. Ξεκινάμε με τη μικρότερη παρατήρηση ή λίγο πιο κάτω από αυτή (δηλ. στρογγυλοποιούμε προς τα κάτω) και με βήμα c καλύπτουμε όλες τις παρατηρήσεις. Το άνω όριο κάθε κλάσης δεν ανήκει στην κλάση αλλά στην αμέσως επόμενη. Επιπλέον η μικρότερη τιμή θα ανήκει στην πρώτη κλάση, η μεγαλύτερη τιμή στην τελευταία κλάση και κάθε ενδιάμεση τιμή σε μία και μόνο κλάση.

Παράδ. Στο παράδειγμα των 190 ημερήσιων μισθών της Α-Ω Χρηματιστηριακής έχουμε ότι $R = x_{min} - x_{max} = 97.200 - 11.340 = 85.860$ και $q = 16$ κλάσεις. Άρα $c = \frac{R}{q} = \frac{85.860}{16}$ $= 5.366,25$, οπότε μπορούμε να θεωρήσουμε π.χ. ότι $c = 6.000$ καθώς θα ομαδοποιήσουμε μισθούς. Αν θεωρήσουμε $x_L = 11.000$, τότε έχουμε την ομαδοποίηση (σε χιλ. €): 11-17, 17-23, 23-29, 29-35, 35-41, 41-47, 47-53, 53-59, 59-65, 65-71, 71-77, 77-83, 83-89, 89-95, 95-101, 101-107. Παρατηρούμε ότι ο μικρότερος μισθός ανήκει στην πρώτη κλάση όμως ο μεγαλύτερος μισθός δεν ανήκει στην τελευταία κλάση. Το ίδιο πρόβλημα έχουμε αν θεωρήσουμε $x_L = 10.000$ ή $9.000$ ή $8.000$. Το πρόβλημα διορθώνεται αν θεωρήσουμε $x_L = 7.000$ ή ακόμα ότι $x_L = 6.000$ όπως στο βιβλίο. Βέβαια θα μπορούσαμε να είχαμε θεωρήσει ότι π.χ. $c = 10.000$ οπότε η ομαδοποίηση θα ήταν διαφορετική. Γενικά σκόπιμο είναι να ακολουθούμε τους τύπους (1) και (2) στη διαδικασία της ομαδοποίησης εκτός κι αν ορίζεται κάτι άλλο.

Στη συνέχεια με διαλογή φτιάχνουμε (5

την στήλη των (απόλυτων) συχνοτήτων.
Η συχνότητα κάθε κλάσης $v_i$ ισούται με το πλήθος των παρατηρήσεων που ανήκουν σ' αυτή. Το άθροισμα των συχνοτήτων, που μπορεί να εμφανίζεται στην τελευταία γραμμή του πίνακα, πρέπει να ισούται με το μέγεθος $n$ του δείγματος. Μερικές φορές, για λόγους που θα γίνουν φανεροί στα επόμενα, αντί να λέμε "η συχνότητα $v_i$ της κλάσης $i$" λέμε "η συχνότητα $v_i$ της κεντρικής της τιμής $x_i$" όπου $x_i$ η κεντρική ή αντιπροσωπευτική τιμή που ορίζεται ως ο μέσος όρος των άκρων της κλάσης $i$.

Στην επόμενη στήλη υπολογίζουμε τις σχετικές συχνότητες. Η σχετική συχνότητα $f_i$ της κλάσης $i$ ορίζεται ως το πηλίκο $\frac{v_i}{n}$ και εκφράζεται και ως ποσοστό %.
Το άθροισμα όλων των σχετικών συχνοτήτων πρέπει να είναι 1 ή 100%

Η επόμενη στήλη δίνει τις αθροιστικές

συχνότητες. Ως αθροιστική συχνότητα (6)
$N_i$ της κλάσης $i$ ορίζεται το άθροισμα
$N_i = v_1 + v_2 + \ldots + v_i$ και δίνει το πλήθος
των παρατηρήσεων που υπάρχουν συνολικά μέχρι το άνω όριο της κλάσης $i$.
Η αθροιστική συχνότητα της τελευταίας κλάσης ισούται με το μέγεθος του δείγματος.

Η τελευταία στήλη δίνει τις σχετικές αθροιστικές συχνότητες. Η σχετική αθροιστική συχνότητα $F_i$ της κλάσης $i$ ορίζεται ως το άθροισμα $F_i = f_1 + f_2 + \ldots + f_i$ και δίνει το ποσοστό των παρατηρήσεων που υπάρχουν συνολικά μέχρι το άνω όριο της κλάσης $i$. Η σχετική αθροιστική συχνότητα της τελευταίας κλάσης πρέπει να ισούται με 1 ή 100%.

Το αποτέλεσμα είναι ένας πίνακας όπως αυτός στη σελ. 34 του βιβλίου (πίνακας 2.3) που λέγεται πίνακας κατανομής συχνοτήτων ή απλούστερα πίνακας συχνοτήτων (της μεταβλητής "ημερήσιοι μισθοί").

Πίνακες συχνοτήτων κατασκευάζονται (7)
και για ποιοτικές μεταβλητές μόνο που περιέχουν τις στήλες των συχνοτήτων και των σχετικών συχνοτήτων (δες πίνακα 2.4 σελ. 36 βιβλίου)

## 2. ΔΙΑΓΡΑΜΜΑΤΑ

### 2.1 Ιστόγραμμα

Για την κατασκευή του ιστογράμματος συχνοτήτων θεωρούμε ένα σύστημα ορθογωνίων αξόνων στον οριζόντιο άξονα του οποίου σημειώνουμε τα όρια των κλάσεων ή/και τις κεντρικές τιμές τους (σε όποια κλίμακα θέλουμε). Στη συνέχεια κατασκευάζουμε ορθογώνια παραλληλόγραμμα που έχουν βάσεις τα διαστήματα των κλάσεων και ύψος (στον κάθετο άξονα) ανάλογο της συχνότητας κάθε κλάσης. Σύμφωνα με ορισμένους συγγραφείς το εμβαδό κάθε ορθογωνίου, και όχι το ύψος του, πρέπει να είναι ανάλογο της συχνότητας κάθε κλάσης. (δες διάγραμμα 2.1 σελ. 37 βιβλίο)
Αντί για συχνότητες, στον κάθετο άξονα,

μπορούμε να έχουμε σχετικές συχνότητες, αθροιστικές συχνότητες, σχετικές αθροιστικές συχνότητες, οπότε προκύπτουν αντίστοιχα τα <u>ιστογράμματα σχετικών συχνοτήτων/ αθροιστικών συχνοτήτων / σχετικών αθροιστικών συχνοτήτων</u> (δες πχ. διάγραμμα 2.2 σελ. 38 βιβλίου)

Από το ιστόγραμμα συχνοτήτων μπορούμε να κατασκευάσουμε το <u>πολύγωνο συχνοτήτων</u>. Θεωρούμε δύο υποθετικές κλάσεις στην αρχή και στο τέλος αντίστοιχα, ίδιου πλάτους με τις υπόλοιπες και ύψους 0 (δηλ. μηδενικής συχνότητας) και στη συνέχεια ενώνουμε τα μέσα των άνω βάσεων των ορθογωνίων με ευθύγραμμα τμήματα

Το πολύγωνο συχνοτήτων μπορεί να κατασκευασθεί ξεχωριστά και ως <u>γραμμογράφημα</u>. Δηλ. παίρνουμε τις κεντρικές τιμές των κλάσεων, των υποθετικών κλάσεων συμπεριλαμβανομένων, και τις τοποθετούμε σ' έναν οριζόντιο άξονα. Στον κατακόρυφο άξονα τοποθετούμε τις αντίστοιχες συχνότητες και ενώνουμε τα

σημεία που δημιουργούνται με ευθύγραμμα (9)
τμήματα. Το εμβαδό του χωρίου μεταξύ
του πολυγώνου συχνοτήτων και του οριζόντιου
άξονα ισούται με το άθροισμα των εμβαδών
των ορθογωνίων παραλληλογράμμων. Αν
έχουμε κατασκευάσει κάθε ορθογώνιο παραλ-
ληλόγραμμο έτσι ώστε το εμβαδό του να
ισούται με τη συχνότητα της αντίστοιχης
κλάσης, τότε το άθροισμα των εμβαδών
τους ισούται με το μέγεθος του δείγματος.
Όσο το πλάτος των κλάσεων μικραίνει δηλ.
οι κλάσεις γίνονται λεπτότερες, τόσο το πολύ-
γωνο συχνοτήτων τείνει να πάρει τη μορφή
μιας καμπύλης που ονομάζεται καμπύλη
συχνοτήτων. Η καμπύλη συχνοτήτων μας δίνει
μια πρώτη ένδειξη για το είδος της κατανομής.
Ανάλογα μπορούμε να κατασκευάσουμε
το πολύγωνο σχετικών συχνοτήτων. Αυτό
είναι ενδεικτικό της διασποράς ή μεταβλητό-
τητας των τιμών. Το εμβαδό του χωρίου
μεταξύ του οριζόντιου άξονα και του πολυ-
γώνου σχετικών συχνοτήτων ισούται με 1 (ή
100%) αν έχουμε κατασκευάσει νωρίτερα
κάθε ορθογώνιο παρ/μο έτσι ώστε το εμβαδό
του να ισούται με την σχετική συχνότητα της

αντίστοιχης κλάσης (δες π.χ. διάγραμμα 2.2 σελ. 38 και διάγραμμα 2.3 σελ. 39 βιβλίο).

Από το ιστόγραμμα αθροιστικών συχνοτήτων κατασκευάζεται η αθροιστική πολυγωνική γραμμή. Ενώνουμε, με ευθύγραμμα τμήματα, το αριστερό άκρο της κάτω βάσης του πρώτου ορθογωνίου και τα δεξιά άκρα των άνω βάσεων των υπόλοιπων ορθογωνίων. Μια σημαντική χρήση της αθροιστικής πολυγωνικής γραμμής είναι η εξής: από ένα σημείο του οριζόντιου άξονα (π.χ. το 21 στο παράδειγμά μας) φέρνουμε κάθετη και προσδιορίζουμε το σημείο τομής της με την αθροιστική πολυγωνική γραμμή. Από το σημείο τομής φέρνουμε κάθετη στον κατακόρυφο άξονα οπότε και προσδιορίζουμε την τεταγμένη του σημείου τομής. Ο αριθμός αυτός αντιπροσωπεύει το πλήθος των εργαζομένων που αμείβονται ετησίως με μέχρι 21.000 €. Η διαδικασία αυτή ονομάζεται (γραμμική) παρεμβολή.

Ανάλογα, από το ιστόγραμμα σχετικών αθροιστικών συχνοτήτων κατασκευάζουμε την

Ποσοστιαία αθροιστική πολυγωνική γραμμή ή Πολύγωνο σχετικών αθροιστικών συχνοτήτων. Αν π.χ. από το σημείο 21 του οριζόντιου άξονα φέρουμε κάθετη, προσδιορίσουμε το σημείο τομής της με την ποσοστιαία αθροιστική πολυγωνική γραμμή, και μετά φέρνοντας κάθετη στον κατακόρυφο άξονα προσδιορίσουμε την τεταγμένη του σημείου τομής, τότε θα έχουμε προσδιορίσει το % ποσοστό των εργαζομένων που αμείβονται ετησίως με μέχρι 21.000 € (δες π.χ. διάγραμμα 2.4 σελ. 40 βιβλίου).

## 2.2 Κυκλικό διάγραμμα

Το κυκλικό διάγραμμα ή διάγραμμα πίτας χρησιμοποιείται κυρίως για ποιοτικές μεταβλητές. Κάθε κατηγορία αντιστοιχεί σε ένα κομμάτι (δηλ. κυκλικό τομέα) και το μήκος του (σε μοίρες °) είναι ανάλογο της σχετικής συχνότητας της αντίστοιχης κατηγορίας. Π.χ. στο παράδειγμά μας: Ασφάλεια $\rightarrow \frac{4}{100} \cdot 360^\circ = 14,4^\circ$

Πτυχιούχοι → $\frac{11}{100} \cdot 360^\circ = 39,6^\circ$

Πτυχιούχοι με MBA → $\frac{8}{100} \cdot 360^\circ = 28,8^\circ$

Αναλυτές → $\frac{2}{100} \cdot 360^\circ = 7,2^\circ$

Προϊστάμενοι → $\frac{3}{100} \cdot 360^\circ = 10,8^\circ$

Υπάλληλοι → $\frac{41}{100} \cdot 360^\circ = 147,6^\circ$

Εκπαιδευόμενοι → $\frac{31}{100} \cdot 360^\circ = 111,6^\circ$

Σύνολο → $14,4^\circ + 39,6^\circ + 28,8^\circ + 7,2^\circ + 10,8^\circ + 147,6^\circ + 111,6^\circ = 360^\circ$

(δες. διάγραμμα 2.5 σελ. 41 βιβλίο)

Το κυκλικό διάγραμμα μπορεί να χρησιμοποιηθεί και για ποσοτικές διακριτές μεταβλητές.

## 2.3 Ραβδόγραμμα

Τα ραβδογράμματα ή ακιδωτά διαγράμματα χρησιμοποιούνται στη γραφική απεικόνιση συχνοτήτων και σχετικών συχνοτήτων (ποσοστιαία ακιδωτά διαγράμματα) ποιοτικών μεταβλητών. Μπορούν όμως να χρησιμοποιηθούν και για ποσοτικές μεταβλητές π.χ. σε πίνακες διπλής εισόδου και είναι ιδιαίτερα παραστατικά στις περιπτώσεις που εξετάζουμε μεταβολή μεγεθών σε σχέση με τον χρόνο. Τα ακιδωτά διαγράμματα διακρίνονται σε: <u>απλά</u> (δες π.χ. διάγραμμα 2.7 σελ. 42), σε <u>σωρευτικά</u> ή <u>διαγράμματα συνιστωσών</u> όπου η κάθε ράβδος χωρίζεται σε μερικά μεγέθη ή συνιστώσες ανάλογα με τη συχνότητα ή τη σχετική συχνότητα της κάθε υποκατηγορίας (δες π.χ. διάγραμμα 2.8 σελ. 43, διάγραμμα 2.10 σελ. 45), σε <u>πολλαπλά</u> όπου κάθε μερικό μέγεθος ή υποκατηγορία αντιστοιχεί σε ξεχωριστή ράβδο (δες π.χ. διάγραμμα 2.9 σελ. 44)

Το κύριο χαρακτηριστικό που διαφοροποιεί το ραβδόγραμμα από το ιστόγραμμα είναι ότι οι ράβδοι κάθε κατηγορίας στο πρώτο δεν εφάπτονται μεταξύ τους όμως γίνεται στο δεύτερο με τις ράβδους κάθε κλάσης.

## 2.4 Χρονόγραμμα

Τα <u>χρονογράμματα</u> ή <u>χρονοδιαγράμματα</u> είναι γραμμογραφήματα που απεικονίζουν την εξέλιξη μεγεθών στη διάρκεια του χρόνου (π.χ. μήνας, τρίμηνο, έτος κ.λπ.). Ο κάθετος άξονας μπορεί να αρχίζει από το 0, μπορεί όμως να καλύπτει μόνο το εύρος τιμών του μεγέθους ή των μεγεθών που εξετάζουμε (δες π.χ. διάγραμμα 2.11 σελ. 45, διάγραμμα 2.12 σελ. 46).