

1^ο ΚΕΦΑΛΑΙΟ

ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

1.1 Το περιεχόμενο της περιγραφικής στατιστικής

Με τις στατιστικές μεθόδους επιδιώκεται αφενός η συνοπτική αλλά εμπεριστατωμένη παρουσίαση των ευρημάτων μιας μελέτης (**περιγραφική στατιστική**) και αφετέρου η συναγωγή συμπερασμάτων που βασίζονται στα ευρήματα αυτά (**συμπερασματολογική στατιστική**). Η γνώση της περιγραφικής στατιστικής αποτελεί επιπροσθέτως απαραίτητη υποδομή για την κατανόηση των εννοιών της συμπερασματολογικής στατιστικής.

1.2 Διάκριση των μεταβλητών

Οι μεταβλητές (μεγέθη, χαρακτηριστικά, variables) βασικά διακρίνονται σε **ποιοτικές** (qualitative or categorical) και **ποσοτικές** (quantitative or numerical). Στις γενικές αυτές κατηγορίες υπάρχουν και άλλες υποδιαιρέσεις.

Ποιοτικές μεταβλητές είναι εκείνες που δεν επιδέχονται αριθμητικές μετρήσεις, αλλά περιγράφονται οι κατηγορίες στις οποίες ταξινομούνται οι παρατηρήσεις, όπως π.χ. το φύλο, το χρώμα της ίδιας, η παρουσία ή απουσία ορισμένου παθολογικού χαρακτηριστικού, η επιβίωση (ή όχι) μετά από ορισμένο χρονικό διάστημα των πασχόντων από μια νόσο κ.ο.κ.

Η απλούστερη μορφή ποιοτικών παρατηρήσεων είναι εκείνη με δύο μόνο κατηγορίες, όπως για το φύλο είναι: άνδρας – γυναίκα, για το κάπνισμα μπορεί να είναι: καπνιστής – μη καπνιστής κ.λ.π. Τα δεδομένα αυτά ονομάζονται **διαδικτικά** (binary), ή **διχοτομικά**, ή ναι/όχι ή 0 – 1. Άλλες ποιοτικές μεταβλητές έχουν περισσότερες από δύο κατηγορίες, όπως η ομάδα αίματος: A, B, AB, O, η οικογενειακή κατάσταση: ανύπαντρος-η/παντρεμένος-η/χωρισμένος-η/χήρος-α. Στο παράδειγμα της ομάδας αίματος προφανώς η κατηγοροποίηση δεν είναι διατάξιμη, σε αντίθεση με το αποτέλεσμα μιας νόσου, όπου η ταξινόμηση είναι δυνατό να γίνει με τη σειρά ίαση, βελτίωση, στασι-

2 ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

μότητα, επιδείνωση, θάνατος. Οι μεταβλητές των οποίων οι κατηγορίες ιεραρχούνται ονομάζονται **διαβαθμιζόμενες ή διατάξιμες** (ordinal) και μπορεί να μετατραπούν σε δεδομένα με δύο κατηγορίες, όταν απαιτείται απλούστερη παρουσίαση ή ανάλυση.

Ποσοτικές μεταβλητές είναι εκείνες οι οποίες επιδέχονται αριθμητικές μετρήσεις, όπως τα φυσικά μεγέθη (ανάστημα, βάρος), τα βιολογικά μεγέθη (χοληστερόλη, σάκχαρο) κ.λ.π.

Θα πρέπει να τονιστεί ότι οι ποσοτικές μεταβλητές μπορεί να μετατραπούν σε διατάξιμες και οι διατάξιμες σε ποιοτικές, ενώ το αντίθετο δεν είναι κατά βάση δυνατό. Έτσι, οι τιμές της χοληστερόλης του ορού είναι δυνατό να διακριθούν σε πολύ υψηλές, υψηλές και φυσιολογικές (διατάξιμες παρατηρήσεις), ή απλά σε αυξημένες και φυσιολογικές (ποιοτικές παρατηρήσεις).

Οι ποσοτικές μεταβλητές χωρίζονται σε **συνεχείς** και **ασυνεχείς**.

Ασυνεχείς λέγονται εκείνες που είναι δυνατό να λάβουν μόνο ορισμένες αριθμητικές τιμές. Σε ορισμένες περιπτώσεις αυτές θυμίζουν τα διατάξιμα χαρακτηριστικά, αλλά διαφέρουν από αυτά επειδή οι διάφορες τιμές που παίρνουν δεν ιεραρχούνται απλώς, αλλά είναι γνωστό και το σχετικό τους μέγεθος. Έτσι, ενώ στο παράδειγμα του διατάξιμου χαρακτηριστικού της προηγούμενης παραγράφου είναι γνωστό ότι η βελτίωση είναι καλύτερη από τη στασιμότητα και η στασιμότητα καλύτερη από την επιδείνωση, δεν είναι δυνατό να λεχθεί αν η απόσταση των 2 κατηγοριών είναι η ίδια. Αντίθετα, αν το μετρούμενο μέγεθος είναι ασυνεχές ποσοτικό, π.χ. ο αριθμός των μαθητών στις τάξεις του λυκείου, είναι προφανές ότι μια τάξη με 20 μαθητές είναι 2πλάσια σε μέγεθος από μια τάξη με 10 μαθητές κ.ο.κ. και η διαφορά 5 μαθητών είναι πάντα η ίδια σε όλο το εύρος των παρατηρούμενων τιμών.

Οι **συνεχείς** ποσοτικές μεταβλητές μπορούν να πάρουν θεωρητικά όλες τις τιμές των πραγματικών αριθμών, τουλάχιστον σ' ένα διάστημα. Στην πράξη, βέβαια, συνήθως μετρούνται με μια καθορισμένη εκ των προτέρω ακρίβεια.

1.3 Ποιοτικές μεταβλητές

Μετά τη συλλογή των ποιοτικών ή των διατάξιμων παρατηρήσεων ενδείκνυται η κατάταξή τους σε ομάδες με κοινό χαρακτηριστικό ή κοινό συνδυασμό χαρακτηριστικών. Έτσι, οι μαθητές μιας τάξης θα μπορούσαν να διαχωριστούν σε αγόρια και σε κορίτσια. Η κατάταξη αυτή αποτελεί την **κατανομή συχνοτήτων** (frequency distribution). Χρήσιμη είναι επίσης η παράθεση των αποτελεσμάτων με μορφή εκατοστιαίων ποσοστών. Σ' αυτό δεν συντρέχουν θεωρητικοί λόγοι, αλλά η εκτίμηση των δεδομένων είναι ευχερέστερη, όταν αυτά παρέχονται σε ποσοστιαία έκφραση. Παρόλα αυτά τονίζεται, ότι η παράθεση των αριθμών, στους οποίους βασίζεται ο υπολογισμός των ποσο-

στών, είναι απαραίτητη σε κάθε περίπτωση, δεδομένου ότι π.χ. μεγαλύτερη αξία έχει το 75% που βασίζεται σε 1000 παρατηρήσεις, από το 75% που βασίζεται σε 100. Θα πρέπει να σημειωθεί ότι μια ταξινόμηση ποιοτικού ή διατάξιμου χαρακτηριστικού θεωρείται επιτυχής, όταν κάθε παρατήρηση μπορεί να περιληφθεί σε μια και μόνο μια από τις ταξινομητικές κατηγορίες. Έτσι, η ταξινόμηση των ατόμων μιας ομάδας σε υγιή, πάσχοντα από καρκίνο του πνεύμονα και πάσχοντα από πνευμονικό εμφύσημα είναι ανεπιτυχής, γιατί ορισμένα άτομα είναι δυνατόν (σπανίως) να πάσχουν τόσο από καρκίνο του πνεύμονα, όσο κι από πνευμονικό εμφύσημα. Ομοίως, η ταξινόμηση των ανθρώπων σε έγγαμους και άγαμους είναι ανεπιτυχής, γιατί ορισμένα άτομα ενδέχεται να μην ανήκουν σε οποιαδήποτε από τις παραπάνω κατηγορίες (π.χ. οι διαζευγμένοι). Γενικώς, είναι σκόπιμο, σε κάθε ταξινόμηση να περιλαμβάνεται μια κατηγορία παρατηρήσεων που καλύπτει χαρακτηριστικά άλλα ή άγνωστα, ώστε να εξασφαλίζεται η προϋπόθεση της πληρότητας.

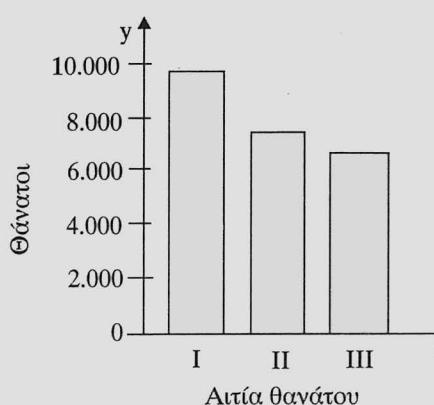
1.4 Γραφικές παραστάσεις κατανομών συχνοτήτων ποιοτικών μεταβλητών

Η περισσότερο εύχρηστη μέθοδος γραφικής παράστασης κατανομής συχνοτήτων ποιοτικών και διατάξιμων χαρακτηριστικών είναι το ιστόγραμμα. Σ' αυτό, ο μεν οριζόντιος άξονας αναφέρεται στο εξεταζόμενο ποιοτικό χαρακτηριστικό, ο δε κάθετος στον αριθμό των παρατηρήσεων, που έγιναν στην κάθε κατηγορία του ποιοτικού χαρακτηριστικού. Το πλήθος των παρατηρήσεων αναφέρεται συνήθως ως **συχνότητα** (frequency).

Στα ιστογράμματα ποιοτικών παρατηρήσεων η συχνότητα εκφράζεται με μόνη την κλίμακα του κάθετου άξονα. Συνεπώς, η σχηματογράφηση γίνεται με ορθογώνια παραλληλόγραμμα-στήλες, γι' αυτό και το ιστόγραμμα καλεί-

ΣΧΗΜΑ 1.1

Θάνατοι από νεοπλασίες (I), αγγειακές βλάβες του κεντρικού νευρικού συστήματος (II) και νόσους της καρδιάς (III) σ' ένα πληθυσμό σε μια ορισμένη χρονική περίοδο. Ο κάθετος άξονας αρχίζει από το μηδέν.



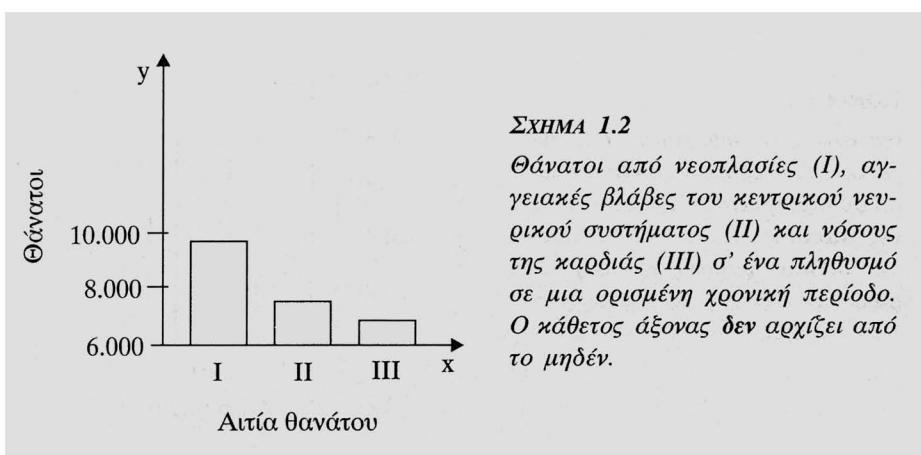
ται και **διάγραμμα στηλών**. Οι στήλες έχουν ορισμένο ύψος, ανάλογα με τη συχνότητα, αλλά το εύρος της βάσης των στηλών και η μεταξύ τους απόσταση δεν έχουν οποιαδήποτε σημασία (Σχήμα 1.1). Επίσης, η τοποθέτηση των στηλών στον οριζόντιο άξονα μπορεί να γίνει με οποιαδήποτε σειρά. Έτσι στο σχήμα 1.1 τα χαρακτηριστικά II και III μπορεί να υποστούν αμοιβαία μετάθεση, εκτός αν αναφέρονται σε διατάξιμο χαρακτηριστικό, οπότε η διάταξη στον οριζόντιο άξονα θα πρέπει να αντιστοιχεί στην υπάρχουσα φυσική διαβάθμιση. Για παράδειγμα, στην εξέλιξη ή έκβαση μιας νόσου η σειρά είναι: ίαση, βελτίωση, στασιμότητα, επιδείνωση, θάνατος – ή η αριθμώς αντίστροφη.

Στα ιστογράμματα, η κλίμακα του οριζόντιου άξονα μπορεί να αρχίζει από οποιαδήποτε κατηγορία ποιοτικού χαρακτηριστικού. Αντίθετα, η κλίμακα του κάθετου άξονα θα πρέπει κατά το δυνατό να αρχίζει από το μηδέν. Η μη τήρηση του κανόνα αυτού ενδέχεται να οδηγήσει σε παραπλανητικές εντυπώσεις, οι οποίες φαίνονται με τη συγκριτική εκτίμηση των σχημάτων 1.1 και 1.2.

1.5 Ποσοτικές μεταβλητές

Οι ποσοτικές μεταβλητές μπορεί να διακριθούν σε **ασυνεχείς** (discrete) και **συνεχείς** (continuous). Ασυνεχείς είναι εκείνες, που από τη φύση τους, δεν μπορεί να πάρουν κάθε αριθμητική τιμή, αλλά μόνο ορισμένες. Έτσι, οι μικροβιακές αποικίες σ' ένα στερεό θρεπτικό υλικό είναι υποχρεωτικά ένας ακέραιος θετικός αριθμός (συμπεριλαμβανομένου και του μηδενός), και ποτέ δεκαδικός ή αρνητικός. Το ίδιο ισχύει και για τα τερηδονισμένα δόντια ή τον αριθμό των καταγμάτων σ' ένα συγκεκριμένο άτομο.

Οι συνεχείς μεταβλητές, εξάλλου, μπορεί να πάρουν οποιεδήποτε αριθμητικές τιμές, παρόλο που αυτές περιορίζονται συνήθως μεταξύ ενός ανώτε-



ρου κι ενός κατώτερου ορίου. Έτσι, το ανάστημα ενός ανθρώπου δεν ξεπερνάει τα 2,3 μ. ούτε είναι μικρότερο των 40 εκ., μέσα όμως στα παραπάνω κατά προσέγγιση όρια μπορεί να πάρει οποιαδήποτε τιμή. Ως συνεχή μεγέθη εκτιμώνται, για παράδειγμα, οι διάφορες χημικές ουσίες στα βιολογικά υγρά (αύμα, ούρα, κ.λ.π.).

1.6 Η κατανομή συχνοτήτων για ποσοτικές μεταβλητές

Μετά τη συλλογή των ποσοτικών παρατηρήσεων, ακολουθεί συνήθως η διαμόρφωση της κατανομής συχνοτήτων, δηλαδή ο υπολογισμός του πλήθους (συχνότητας) των παρατηρήσεων, οι οποίες διαπιστώθηκαν στις διάφορες τιμές του μετρούμενου μεγέθους. Για να διευκρινιστεί η έννοια της κατανομής συχνοτήτων δίνεται το παρακάτω παράδειγμα: 400 οικογένειες, δλες με τρία παιδιά, ωρτήθηκαν για τον αριθμό των αγοριών τους. Τα αποτελέσματα συνοψίζονται στον πίνακα 1.1.

Στο παράδειγμα αυτό το μετρούμενο μέγεθος είναι ο αριθμός των αγοριών στην οικογένεια, δηλαδή μέγεθος ασυνεχές που επιδέχεται μόνο τέσσερις αριθμητικές τιμές (0 ή 1 ή 2 ή 3). Στην κάθε μια από τις τιμές αυτές βρέθηκαν οι συχνότητες που αναγράφονται αντίστοιχα.

Μπορεί το μετρούμενο μέγεθος να είναι ασυνεχές, αλλά να παίρνει πολλές αριθμητικές τιμές όπως, π.χ. ο αριθμός των πυοσφαιρίων που βρίσκονται σ' ένα οπτικό πεδίο κατά τη μικροσκοπική εξέταση του ιζήματος των ούρων που είναι δυνατό να μετρηθούν άλλοτε μεν 0 ή 1 ή 2 πυοσφαίρια, άλλοτε δε 40 ή 50. Είναι ευνόητο, ότι μια κατανομή συχνοτήτων που εκτείνεται σε τόσο ευρεία κλίμακα είναι κατά κανόνα δύσχρηστη. Η δυσχέρεια είναι δυνατό να ξεπεραστεί με την κατάταξη των παρατηρήσεων σε ομάδες

ΠΙΝΑΚΑΣ 1.1

Κατανομή συχνοτήτων 400 οικογενειών με τρία παιδιά, ανάλογα με τον αριθμό των αγοριών

Αγόρια	Συχνότητα
0	52
1	133
2	146
3	69
Σύνολο	400