

ΕΙΣΑΓΩΓΗ ΣΤΙΣ ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΥΣ

«Οι κυνικοί λένε ότι υπάρχουν ψέματα, δαμονισμένα ψέματα και η Στατιστική. Και καθημερινά βομβαρδιζόμαστε με Στατιστικά μεγέθη, από τη συναλλαγματική ισοτιμία της στερλίνας ή τους τελευταίους δείκτες ανεργίας μέχρι τους μέσους όρους απόδοσης αθλητών. Όμως πόσο πολύ μπορούμε να εμπιστευόμαστε τη Στατιστική;»

(Ελεύθερη Μετάφραση από το New Scientist, Inside Science, Ειδικό Ένθετο, αριθμός 61, Μάιος 1993)

Η απάντηση στο ερώτημα αυτό είναι συνάρτηση των γνώσεων μας στην ανάλυση δεδομένων, με τη χρήση της Στατιστικής. Η στατιστική ανάλυση έχει σχεδιαστεί με τέτοιο τρόπο, ώστε να μας επιτρέπει να εξάγουμε λογικά συμπεράσματα, χρησιμοποιώντας πολλές φορές ατελή ή ακόμα και ανακριβή δεδομένα. Η ανάλυση των δεδομένων απαιτεί πρώτα τη συλλογή όλων των σχετικών πληροφοριών. Όμως, η συγκεκριμένη διαδικασία δεν είναι καθόλου εύκολη. Στην πραγματικότητα, η συλλογή όλων των σχετικών πληροφοριών ίσως είναι ανέφικτη. Επιπλέον, το βασικό πρόβλημα είναι ότι τα σύνολα δεδομένων σπανίως εμφανίζονται με ομαλό και τακτοποιημένο τρόπο. Δεν παρουσιάζουν, αυτό που θα γίνει σαφές αργότερα, κανονικότητα. Καθώς λοιπόν ο πραγματικός κόσμος δεν αναμένεται να συμπεριφέρεται με έναν ακριβή και καθορισμένο τρόπο, η στατιστική ανάλυση αποσκοπεί στη δημιουργία πληροφόρησης που να είναι συνεπής και χρήσιμη και να μας επιτρέπει να προβούμε σε ακριβή συμπεράσματα και να εξετάσουμε την αξιοπιστία τους.

Υπάρχουν πολλά στατιστικά συμπεράσματα, τα οποία παρατηρούμε καθημερινά στην τηλεόραση, στα περιοδικά, στις εφημερίδες και σε άλλες πηγές πληροφόρησης. Για παράδειγμα, παρακολουθούμε τα στατιστικά ενός αγώνα καλαθοσφαίρισης ή ποδοσφαίρου, με τα οποία αναλύεται η απόδοση των παικτών, και των ομάδων. Επίσης, παρακολουθούμε τις δημοσκοπήσεις που αφορούν τη δημοτικότητα των πολιτικών και των κομμάτων.

Μπορούμε όμως να σκεφτούμε παρόμοια παραδείγματα, όπου η χρησιμοποίηση στατιστικών τεχνικών να μας βοηθά να παρουσιάσουμε το πώς σκέφτονται και συμπεριφέρονται οι άνθρωποι; Το αντικείμενο της κοινωνικής έρευνας είναι να κατανοήσουμε, και αν είναι εφικτό να κατασκευάσουμε, ένα υπόδειγμα, το οποίο να ερμηνεύει τον πραγματικό κόσμο. Έτσι, χρειαζόμαστε πληροφορίες με τη μορφή δεδομένων. Μετά τη

συλλογή των πληροφοριών, η Στατιστική μας βοηθά να αναλύσουμε και να ερμηνεύσουμε κάθε πιθανώς υπαρκτή σχέση μεταξύ των συλλεχθέντων δεδομένων. Άρα:

Η Στατιστική είναι η επιστήμη, που αναφέρεται στη συλλογή, την παρουσίαση, την ανάλυση και τη χρησιμοποίηση αριθμητικών πληροφοριών, με σκοπό να κατανοήσουμε τα ποιοτικά και ποσοτικά χαρακτηριστικά ενός οικονομικού, κοινωνικού, ή βιολογικού συστήματος.

Χρησιμοποίηση των πληροφοριών σημαίνει ότι αντλούμε συμπεράσματα για την ερμηνεία του συνόλου των δεδομένων. Συχνά, τέτοιου είδους συμπεράσματα, μπορούν να πάρουν τη μορφή προβλέψεων.

Θα μπορούσε επίσης να ειπωθεί ότι:

Η Στατιστική είναι μία μαθηματική γλώσσα, η οποία με εργαλείο τη θεωρία των πιθανοτήτων μας καθιστά ικανούς να συμπεράνουμε για διάφορα θέματα, εντός κάποιων λογικών ορίων σφαλμάτων, από τη χρησιμοποίηση πληροφοριών με τη μορφή δεδομένων.

Η Στατιστική χωρίζεται σε Περιγραφική (Descriptive) και Επαγωγική (Inferential ή Inductive). Η **Περιγραφική Στατιστική** περιγράφει τα βασικά χαρακτηριστικά ενός συνόλου δεδομένων, δίχως να εξάγει συμπεράσματα για τον πληθυσμό. Έτσι, η Περιγραφική Στατιστική στοχεύει στην οργάνωση και στην παρουσίαση πληροφοριών μέσω ενός άμεσου και απλού τρόπου. Για παράδειγμα, παρουσιάζουμε την ετήσια ποσότητα νεροποντής σε κάποια περιοχή, την ανεργία, τον πληθωρισμό κ.λπ. Ποσοστά, δείκτες, αναλογίες, συχρότητες, διαγράμματα, πίνακες και γραφήματα αποτελούν μερικούς τρόπους με τους οποίους οργανώνουμε, συνοψίζουμε και περιγράφουμε τα δεδομένα.

Ειδικότερα, τα δεδομένα είναι ένα σύνολο αριθμητικών τιμών οι οποίες σχετίζονται με το υπό εξέταση φαινόμενο. Η Στατιστική μας βοηθά να μετατρέψουμε έναν τεράστιο όγκο πληροφοριών σε μια εύκολα κατανοητή και ταυτόχρονα περιεκτική διατύπωση. Όταν ένας ερευνητής ρωτά κάποια άτομα, παράγεται ένα σύνολο πιθανών απαντήσεων. Αν ρωτήσουμε 100 άτομα για τη μέση δαπάνη ρουχισμού θα λάβουμε πιθανώς 100 διαφορετικές απαντήσεις. Αυτό το σύνολο των απαντήσεων ορίζεται ως μια **μεταβλητή**. Κάθε απάντηση αντιπροσωπεύει μια συγκεκριμένη τιμή.

Η **Επαγωγική Στατιστική** χρησιμοποιεί τις πληροφορίες που αντλούνται από δείγματα, ώστε να προβλέψει τα βασικά χαρακτηριστικά του πληθυσμού. Ο όρος **πληθυσμός** αναφέρεται στη συλλογή όλων των στοιχείων τα οποία υπάρχουν για τη μεταβλητή που περιγράφουμε. Δηλαδή, ο πληθυσμός είναι το σύνολο των στοιχείων που μελετώνται ή το σύνολο των παρατηρήσεων που επικεντρώνουν το ενδιαφέρον του ερευνητή. Από την άλλη πλευρά, ένα **δείγμα** είναι ένα επιλεγμένο τμήμα του πληθυ-

σμού, ή ένα υποσύνολο των παρατηρήσεων αυτών, το οποίο επιλέγεται τυχαία από τον πληθυσμό. Μερικές φορές η συγκέντρωση πληροφοριών πραγματοποιείται σε ολόκληρο τον πληθυσμό, διαδικασία που είναι γνωστή ως *απογραφή πληθυσμού* (census). Όμως, επειδή η μελέτη όλων των στοιχείων του πληθυσμού είναι μια δαπανηρή και χρονοβόρα διαδικασία, χρησιμοποιούμε αντιπροσωπευτικά δείγματα από τον πληθυσμό, τον οποίο ενδιαφερόμαστε να μελετήσουμε. Για παράδειγμα, σε μια δημοσκόπηση, το εκλογικό αποτέλεσμα του συνολικού εκλογικού σώματος μπορεί να προβλεφθεί από ένα τυχαίο δείγμα ατόμων.

1.1 Κατηγορίες δεδομένων

Τα **δεδομένα** είναι ένας στατιστικός όρος για τις πληροφορίες που σχετίζονται με την υπό εξέταση μεταβλητή. Είναι γεγονότα και αριθμοί, που συλλέγουμε, παρουσιάζουμε, αναλύουμε και ερμηνεύουμε. Τα ακατέργαστα δεδομένα είναι αυτά που συλλέγουμε στην αρχική τους μορφή. Μπορούμε να υποστηρίξουμε ότι τα δεδομένα αποτελούν την πρώτη ύλη για τη Στατιστική. Υπάρχουν τέσσερις κλίμακες μέτρησης, οι οποίες καθορίζουν το σύνολο των πληροφοριών που περιλαμβάνονται στα δεδομένα:

- α) *Ονομαστική ή κατηγορική κλίμακα (Nominal ή categorical scale)*: Δεδομένα σε τέτοια κλίμακα αποτελούνται από τον αριθμό των παρατηρήσεων, οι οποίες σχετίζονται με συγκεκριμένες κατηγορίες (για παράδειγμα, μη αριθμητικές, όπως φύλο, οικογενειακή κατάσταση, επάγγελμα κ.λπ., ή αριθμητικές, όπως ο αριθμός ISBN ενός βιβλίου κ.λπ.). Η ονομαστική κλίμακα χρησιμοποιεί μόνον τους αριθμούς που αναγνωρίζουν κατηγορίες. Για παράδειγμα, αν είχαμε 2 κατηγορίες, τότε θα τις αριθμούσαμε ως 1 και 2. Οι αριθμητικές τιμές της ονομαστικής κλίμακας μέτρησης (1 και 2) δεν δείχνουν κάποια ποσότητα ή σύνολο, αλλά δείχνουν τα ονόματα των κατηγοριών.
- β) *Τακτική κλίμακα (Ordinal scale)*: Δεδομένα σε τέτοια κλίμακα αποτελούν οι παρατηρήσεις, οι οποίες κατατάσσονται ή διατάσσονται ώστε να δείξουν πρώτο επίπεδο, δεύτερο επίπεδο, τρίτο επίπεδο κ.λπ. Παραδείγματα τακτικής κλίμακας μπορεί να έχουμε όταν οι πόλεις κατατάσσονται σύμφωνα με τον αριθμό των υπαρχόντων μουσείων τους ή με το επίπεδο της εξυπηρέτησης που προσφέρεται στα εστιατόριά τους (άψογο, καλό, μέσο, κακό), κ.λπ.
- γ) *Διαστημική κλίμακα (Interval scale)*: Εδώ τα δεδομένα αποτελούνται από τις παρατηρήσεις, οι οποίες μετρώνται σε κλίμακα η οποία είναι κατασκευασμένη με τέτοιο τρόπο ώστε όμοιες αποστάσεις μεταξύ των αριθμών να αντιστοιχούν σε όμοιες αποστάσεις μεταξύ των τιμών του χαρακτηριστικού που μετράται, ενώ το 0 επιλέγεται αυθαίρετα. Για παράδειγμα, οι βαθμοί θερμοκρασίας Fahrenheit ή η βαθμολογία του τεστ της γνώσης της Αγγλικής γλώσσας TOEFL (450, 500, 700 κ.λπ.). Τα δεδομένα είναι πάντοτε αριθμητικά. Για παράδειγμα, οι κλίμακες θερμοκρασίας Fahrenheit και Celsius είναι κλίμακες μέτρησης διαστημάτων. Παρότι για την κλίμακα Fahrenheit οι 32 βαθμοί ορίζονται ως το σημείο πήξης του νερού, για την

κλίμακα Celsius το αντίστοιχο σημείο ορίζεται ως 0 βαθμοί. Αναφορικά, και για τις 2 κλίμακες θερμοκρασίας μπορούμε να επισημάνουμε ότι οι 25 βαθμοί είναι κατά 5 βαθμούς θερμότεροι από τους 20 και οι 35 βαθμοί είναι κατά 10 βαθμούς θερμότεροι από τους 25. Όμως, το σύνολο της θερμοκρασίας που συνιστά μια μεταβολή σε καταγεγραμμένη θερμοκρασία, είναι διαφορετικό για τις δύο κλίμακες.

- δ) *Αναλογική κλίμακα (Ratio scale)*: Εδώ έχουμε παρατηρήσεις οι οποίες μετρώνται σε κλίμακα, της οποίας η κατασκευή είναι τέτοια, ώστε όμοιες αποστάσεις μεταξύ αριθμών να αντιστοιχούν σε όμοιες αποστάσεις μεταξύ των τιμών της μετρημένης μεταβλητής, ενώ το 0 αντιπροσωπεύει την απουσία του χαρακτηριστικού που μελετάται. Δηλαδή, η μέτρηση της τυχαίας μεταβλητής μπορεί να συγκριθεί με την αναλογική κλίμακα, διότι υπάρχει ένα πραγματικά μηδενικό σημείο. Για παράδειγμα, οι ηλικίες των ανθρώπων αποτελούν δεδομένα αριθμοδεικτών και το 0 αντιπροσωπεύει τη στιγμή της γέννησης. Το συγκεκριμένο είδος δεδομένων είναι πάντοτε αριθμητικό (για παράδειγμα, ύψος, σωματικό βάρος, τιμή προϊόντος κ.λπ.).

Μια επιπρόσθετη διάκριση για τα δύο τελευταία είδη δεδομένων είναι τα συνεχή ή τα διακριτά δεδομένα. Τα **διακριτά δεδομένα** (discrete data) είναι παρατηρήσεις, των οποίων οι τιμές περιορίζονται σε συγκεκριμένα σημεία ενός φάσματος τιμών. Μια μεταβλητή είναι διακριτή όταν λαμβάνει συγκεκριμένες τιμές. Τα **συνεχή δεδομένα** (continuous data) είναι παρατηρήσεις οι οποίες ενδέχεται να περιλαμβάνουν οποιαδήποτε τιμή ενός φάσματος τιμών. Μια συνεχής μεταβλητή ενδέχεται να λάβει οποιαδήποτε αριθμητική τιμή (για παράδειγμα το ύψος ή το σωματικό βάρος ενός ανθρώπου και γενικά κάθε μη ακέραια μετρήσιμη τιμή).

Παράδειγμα 1.1:

- 1) Οι ημερομηνίες γέννησης των φοιτητών κάποιου τμήματος ενός πανεπιστημίου είναι δεδομένα διαστημικής κλίμακας. Το μηδέν δε δηλώνει την αρχή του χρόνου, αλλά την ημερομηνία γέννησης του Χριστού. Οι ημερομηνίες γέννησης είναι διακριτά δεδομένα με ελάχιστη μονάδα μέτρησης την ημέρα (συνήθως).
- 2) Οι μέσοι όροι των βαθμών είναι δεδομένα αναλογικής κλίμακας. Ο μηδενικός μέσος όρος υποδηλώνει την απουσία κάποιου βαθμού διάκρισης (π.χ. άριστα). Εδώ έχουμε συνεχή δεδομένα.
- 3) Τα μεγέθη των φορεμάτων των κοριτσιών σε μια σχολική αίθουσα είναι δεδομένα διαστημικής κλίμακας. Μια μηδενική τιμή είναι, το σημείο εκκίνησης για τις μετρήσεις. Τα μεγέθη των φορεμάτων είναι διακριτά δεδομένα, τα οποία εκφράζονται σε ακέραιες ή μισές μονάδες.
- 4) Ο αριθμός των ψήφων για κάθε υποψήφιο είναι δεδομένα αναλογικής κλίμακας, αν οι υποψήφιοι ληφθούν ως παρατηρήσεις. Αν κάθε ψηφοφόρος ληφθεί ως μια παρατήρηση, τότε τα δεδομένα είναι ονομαστικής κλίμακας.

- 5) Η κατάταξη των υποψηφίων είναι δεδομένα τακτικής κλίμακας.
- 6) Η λίστα των ηττημένων και των νικητών είναι ονομαστικά δεδομένα.
- 7) Μερικά παραδείγματα συνεχών μεταβλητών είναι το σωματικό βάρος, το ύψος, η ηλικία, η ταχύτητα ενός κινητού σώματος και η θερμοκρασία, ενώ παραδείγματα διακριτών μεταβλητών είναι η έκβαση της ρίψης ενός ζαριού (1, 2, 3, 4, 5, 6), ή της ρίψης ενός νομίσματος (κεφαλή, γράμματα).

Αφού συλλεχθούν τα δεδομένα, η Στατιστική μας επιτρέπει να τα περιγράψουμε με σκοπό να τα γνωρίσουμε καλύτερα. Τα πιο χρήσιμα χαρακτηριστικά για την περιγραφή των δεδομένων είναι τα μέτρα θέσης (όπως ο μέσος), τα μέτρα διασποράς (όπως η διακύμανση ή η τυπική απόκλιση) και η μορφή της κατανομής των δεδομένων (όπως η κανονική κατανομή). Σε μια σωστή και πλήρη ανάλυση ενδεχομένως να χρειάζονται περισσότερα από ένα από τα παραπάνω χαρακτηριστικά. Μια *παράμετρος* είναι ένα περιγραφικό χαρακτηριστικό (όπως ο μέσος) ενός πληθυσμού και συμβολίζεται με ελληνικά γράμματα και ένα στατιστικό είναι ένα περιγραφικό χαρακτηριστικό ενός δείγματος και συμβολίζεται με αγγλικά γράμματα. Στην Επαγωγική Στατιστική χρησιμοποιούμε στατιστικές (από δείγματα) για να συμπεράνουμε για τις πληθυσμιακές παραμέτρους. Μια μεταβλητή είναι ένα ενδιαφέρον χαρακτηριστικό για τις οντότητες (entities), για τις οποίες συλλέγονται τα δεδομένα.

Οι στατιστικές μέθοδοι χρησιμοποιούνται για να εκτιμηθούν οι πληθυσμιακές παράμετροι κάτω από συγκεκριμένες υποθέσεις για τις παραμέτρους αυτές. Αν δεν υπάρχει καμία απαιτούμενη υπόθεση για τις πληθυσμιακές παραμέτρους, τότε ενδεχομένως να ασχολούμαστε με *μη παραμετρική (απαραμετρική) Στατιστική*. Κάποιοι άλλοι σημαντικοί ορισμοί για τα δεδομένα είναι τα ποιοτικά (qualitative) και τα ποσοτικά (quantitative) δεδομένα. Τα ποιοτικά είναι συλλεχθέντα δεδομένα σε ονομαστική ή τακτική κλίμακα (nominal ή ordinal scale), ενώ τα ποσοτικά είναι αποκτηθέντα δεδομένα σε διαστημική ή αναλογική κλίμακα.

Παράδειγμα 1.2:

Ας φανταστούμε μια απλή έρευνα με ερωτηματολόγια, της οποίας κάποια από τα αποτελέσματα παρουσιάζονται στον παρακάτω πίνακα.

	Περιοχή	Επάγγελμα	Ηλικία (έτη)	Εισόδημα (€)	Δαπάνες (€)
1	Αττική	Καθηγητής	36	25.000	15.000
2	Κυκλάδες	Τεχνίτης	24	18.000	9.000
3	Θεσσαλία	Τραπεζικός	49	29.000	20.000

Στο συγκεκριμένο πίνακα κάθε γραμμή (οριζόντια) και κάθε στήλη (κάθετη) είναι μια διαφορετική - μεμονωμένη μεταβλητή. Αν κοιτάξουμε τις γραμμές θα προσδιορίσουμε τα γενικά χαρακτηριστικά του υπό εξέταση ατόμου, ενώ αν κοιτάξουμε τις στήλες, θα παρατηρήσουμε τις τιμές που παίρνει η κάθε μεταβλητή. Η επιγραφή της γραμμής μάς δείχνει το όνομα της μεταβλητής. Στον ίδιο πίνακα, το επάγγελμα και η περιοχή είναι ονομαστικές ή κατηγορικές μεταβλητές, ενώ το εισόδημα, η δαπάνη και η ηλικία είναι αριθμητικά δεδομένα.

1.2 Πηγές δεδομένων

Υπάρχουν διάφορες πηγές από όπου μπορούμε να αντλήσουμε δημοσιευμένα στοιχεία. Οι κυβερνήσεις συλλέγουν πολλά δεδομένα μέσω διαφόρων υπηρεσιών και υπουργείων. Η μεγαλύτερη πηγή δημοσιευμένων στοιχείων είναι η Ελληνική Στατιστική Αρχή (ΕΛ.ΣΤΑΤ), η οποία συλλέγει τα αποτελέσματα αρκετών ερευνών για ποικίλα θέματα και δημοσιεύει διάφορες εκδόσεις όπως το Μηνιαίο Στατιστικό Δελτίο, κ.ά. Η ΕΛ.ΣΤΑΤ διεξάγει ανά δεκαετία απογραφές για τον πληθυσμό και τις κατοικίες της χώρας και δημοσιεύει περιλήψεις των συγκεκριμένων δεδομένων. Επίσης, πληροφορίες μπορούν να συλλεχθούν από τα Τμήματα Μελετών και Ανάλυσης των διαφόρων Τραπεζών, όπως τη Διεύθυνση Οικονομικής Ανάλυσης και Μελετών της Τράπεζας Ελλάδος κ.λπ.

Άλλες πηγές πληροφοριών αποτελούν τα λογιστικά βιβλία των επιχειρήσεων, και των κατασκευαστικών εταιρειών, τα ερωτηματολόγια έρευνας αγοράς, οι βάσεις δεδομένων των πανεπιστημίων και των διαφόρων ερευνητικών ιδρυμάτων, όπως το Ίδρυμα Οικονομικών και Βιομηχανικών Ερευνών (ΙΟΒΕ) και το Κέντρο Προγραμματισμού και Οικονομικών Ερευνών (ΚΕΠΕ), οι διεθνείς οργανισμοί, όπως η Παγκόσμια Τράπεζα (World Bank), ο Οργανισμός Οικονομικής Συνεργασίας και Ανάπτυξης (OECD) κ.λπ., και τα Βιομηχανικά και Εμπορικά Επιμελητήρια. Συχνά η προσωπική επικοινωνία βοηθά στην εύρεση των απαιτούμενων για την ανάλυση δεδομένων.

1.3 Μορφές δεδομένων

Τα δεδομένα συνήθως παρουσιάζονται στις ακόλουθες μορφές:

1. *Δεδομένα Χρονολογικών Σειρών (time series data)*, τα οποία ερμηνεύουν τη μεταβολή μιας μεταβλητής κατά τη διάρκεια του χρόνου. Δηλαδή, δείχνουν μετρήσεις σε διάφορα σημεία μέσα στο χρόνο (έτη, τρίμηνα ή μήνες, ημέρες). Για παράδειγμα, ο πίνακας 1.1 παρουσιάζει τα ετήσια δεδομένα για το Ακαθάριστο Εγχώριο Προϊόν (ΑΕΠ) σε αγοραίες τιμές της χώρας και για την περίοδο 1997-2018 (Πηγή στοιχείων η Ελληνική Στατιστική Αρχή).

Πίνακας 1.1: Παράδειγμα Χρονολογικής Σειράς

Έτη	ΑΕΠ	Έτη	ΑΕΠ
1997	10.760	2008	21.845
1998	11.684	2009	21.386
1999	12.432	2010	20.324
2000	13.071	2011	18.643
2001	14.011	2012	17.311
2002	14.994	2013	16.475
2003	16.371	2014	16.402
2004	17.683	2015	16.381
2005	18.134	2016	16.378
2006	19.769	2017	16.736
2007	21.061	2018	16.908

2. *Διαστρωματικά στοιχεία (cross sectional data)*, είναι τα δεδομένα τα οποία αντλούνται από έναν πληθυσμό σε δεδομένο χρονικό σημείο. Για παράδειγμα, ο πίνακας 1.2 παρουσιάζει το ετήσιο διαθέσιμο εισόδημα (Y_d) σε € για ένα τυχαίο δείγμα 10 νοικοκυριών το 2019.

Πίνακας 1.2: Παράδειγμα διαστρωματικών δεδομένων

Μεταβλητές	Οικογένεια									
	1	2	3	4	5	6	7	8	9	10
Y_d	31.456	26.611	25.111	29.022	10.139	31.004	28.831	20.091	32.533	24.116

3. *Χρονικά (ή δυναμικά) διαστρωματικά στοιχεία (Panel data)*, είναι ένας συνδυασμός των παραπάνω δύο κατηγοριών, αφού αναφέρονται σε δεδομένα που αποτελούνται από μια χρονολογική σειρά για κάποιες διαστρωματικές μονάδες που μελετώνται. Μπορούμε να ισχυριστούμε ότι έχουμε επαναλαμβανόμενα διαστρωματικά δεδομένα, τα οποία συλλέγονται από τις ίδιες πηγές για κάποιο συγκεκριμένο χρονικό διάστημα. Το συγκεκριμένο είδος δεδομένων ενδεχομένως να αναφέρεται σε χώρες, περιοχές, επιχειρήσεις, νοικοκυριά, κ.λπ. Μπορεί να παραχθεί λαμβάνοντας παρατηρήσεις συγκεκριμένων χρονολογικών σειρών για μια ποικιλία διαστρωματικών οντοτήτων. Ας θεωρήσουμε ενδεικτικά, και μόνο ως ένα απλοποιημένο υποθετικό παράδειγμα, δεδομένα για τέσσερις (μόνο) χώρες και τη χρονική περίοδο 2014-2019 αναφορικά με τις μεταβλητές εκπομπές αέριων ρύπων

και κατά κεφαλήν Ακαθάριστο Εγχώριο Προϊόν (ΑΕΠ/κ) σε αντίστοιχες μονάδες μέτρησης (Πίνακας 1.3). Αυτά τα στοιχεία μπορούν να θεωρηθούν ως χρονικά διαστρωματικά στοιχεία. Αν είχαμε για τις χώρες αυτές μία μεγαλύτερη χρονική περίοδο, έστω 1980-2019, τότε θα μπορούσαμε να μιλήσουμε για δυναμικά διαστρωματικά στοιχεία.

Πίνακας 1.3: Παράδειγμα χρονικών διαστρωματικών στοιχείων

	Έτη	Εκπομπές αέριων ρύπων	ΑΕΠ/κ
Χώρα 1	2014	7,73	2.685
	2015	7,93	2.716
	2016	7,55	2.145
	2017	8,17	2.691
	2018	8,74	3.194
	2019	8,14	3.114
Χώρα 2	2014	8,11	2.932
	2015	8,32	3.075
	2016	8,35	3.045
	2017	8,92	3.414
	2018	9,27	3.397
	2019	10,9	3.588
Χώρα 3	2014	8,76	3.055
	2015	9,23	3.609
	2016	8,11	4.429
	2017	8,29	4.719
	2018	10,4	4.077
	2019	17,2	5.119
Χώρα 4	2014	16,5	5.287
	2015	29,9	5.277
	2016	44,7	5.579
	2017	56,6	5.741
	2018	50,7	5.578
	2019	49,9	5.436

ΑΣΚΗΣΕΙΣ

1. Δώστε μία σύντομη περιγραφή των όρων Περιγραφική και Επαγωγική Στατιστική, καθώς και των όρων πληθυσμός και δείγμα. Διακρίνετε επίσης την παράμετρο και το στατιστικό.
2. Αναγνωρίστε το είδος των δεδομένων στις παρακάτω περιπτώσεις. Αν τα δεδομένα είναι αναλογικής ή διαστημικής κλίμακας προσδιορίστε αν είναι διακριτά ή συνεχή.
 - i. Οι ημερομηνίες γέννησης των φοιτητών σε ένα έτος σπουδών
 - ii. Τους βαθμούς του τεστ ευφυΐας IQ σε μία ομάδα ατόμων
 - iii. Τα μεγέθη παντελονιών σε ένα κατάστημα ενδυμάτων
 - iv. Ο μέσος όρος των βαθμών των φοιτητών του δεύτερου έτους
 - v. Ο αριθμός των φοιτητών του δεύτερου έτους των οποίων τα επίθετα αρχίζουν με κάθε γράμμα της αλφαβήτου
 - vi. Ο αριθμός των γραμμάτων σε κάθε όνομα φοιτητή
 - vii. Οι απαντήσεις πελατών σε ερώτηση σχετικά με την ποιότητα των προσφερόμενων υπηρεσιών σε κάποιο εστιατόριο με κωδικοποίηση
 - 1 για πολλή χαμηλή ποιότητα,
 - 2 για χαμηλή ποιότητα,
 - 3 για μέτρια ποιότητα,
 - 4 για καλή ποιότητα,
 - 5 για πολλή καλή ποιότητα.
3. Μια έρευνα με ερωτηματολόγια αφορά τους λόγους επίσκεψης ενός πάρκου και περιλαμβάνει ερωτήσεις σχετικές με κοινωνικοοικονομικά χαρακτηριστικά των ερωτηθέντων, όπως:
 - i. Ηλικία
 - ii. Φύλο
 - iii. Ετήσιο εισόδημα
 - iv. Μορφωτικό επίπεδο
 - v. Οικογενειακή κατάσταση
 - vi. Αριθμός παιδιών
 - vii. Συχνότητα επισκέψεων στο πάρκο
 - viii. Λόγοι επίσκεψης

Περιγράψτε τις μεταβλητές αυτές ως προς την κλίμακα μέτρησης και χαρακτηρίστε κάθε μεταβλητή ως ποιοτική ή ποσοτική.

4. Συζητήστε και συγκρίνετε τις διάφορες κλίμακες μέτρησης.
5. Χαρακτηρίστε τον τύπο των δεδομένων μίας εκλογικής αναμέτρησης
 - (α) Αριθμός ψήφων κάθε υποψηφίου
 - (β) Λίστα εκλεγμένων και μη εκλεγμένων υποψηφίων
 - (γ) Τελική κατάταξη υποψηφίων
6. Αναφέρατε παραδείγματα σχετικά με διαστρωματικά στοιχεία, χρονολογικές σειρές και χρονικά διαστρωματικά στοιχεία.
7. Χαρακτηρίστε τις ακόλουθες μεταβλητές ως ποσοτικές ή ποιοτικές και τις ποσοτικές ως διακριτές ή συνεχείς:
 - i. Αριθμός παιδιών σε μία οικογένεια (0,1,2,3,...)
 - ii. Αριθμός δωματίων ενός οικήματος (1,2,3,4,...)
 - iii. Χρώμα ματιών (μαύρα, γαλανά, πράσινα, καστανά)
 - iv. Ένδειξη ρίψης ζαριού (1,2,3,4,5,6)
 - v. Ένδειξη ρίψης νομίσματος ("πρόσωπο", "γραμματα")
 - vi. Ύψος ατόμων σε εκατοστά (150, 151,..., 178, 179, ...)
 - vii. Κατάσταση υγείας (άριστη, πολλή καλή, καλή, μέτρια, κακή, πολλή κακή)
 - viii. Βάρος σε κιλά (50, 51,...70, 71, ...)
 - ix. Σχήμα προσώπου (στρογγυλό, ωοειδές, με γωνίες)