

## ΠΡΟΛΟΓΟΣ

Η συνεχώς επεκτεινόμενη χρήση της Στατιστικής σχεδόν σε όλα τα επιστημονικά πεδία, επιβάλλει την κατανόηση των βασικών τουλάχιστον στατιστικών εννοιών και μεθόδων από έναν μεγάλο αριθμό φοιτητών και σπουδαστών της Τριτοβάθμιας Εκπαίδευσης αλλά και επιστημόνων που οι ειδικότητές τους έχουν ως βασικό εργαλείο τη Στατιστική.

Μικρή συμβολή προς αυτή την κατεύθυνση ελπίζουμε ότι αποτελεί αυτό το βιβλίο της Στατιστικής Μεθοδολογίας, που ενσωματώνει τους ήδη υπάρχοντες δύο τόμους με τον ίδιο τίτλο και περιλαμβάνει ουσιαστικές προσθήκες και βελτιώσεις. Απευθύνεται σε αναγνώστες που έρχονται για πρώτη φορά σε επαφή με τη Στατιστική, αλλά έχουν ικανοποιητικές γνώσεις Άλγεβρας, ενώ για ορισμένα κεφάλαια χρειάζονται επιπλέον γνώσεις Μαθηματικής Ανάλυσης.

Η ύλη αποτελείται από τέσσερα μέρη. Στο πρώτο μέρος αναπτύσσονται οι μέθοδοι της Περιγραφικής Στατιστικής και η θεωρία Αριθμοδεικτών, που είναι ιδιαίτερα χρήσιμη στις οικονομικές επιστήμες. Στο δεύτερο μέρος, μετά από μία σύντομη αναφορά στη Συνδυαστική Ανάλυση, αναπτύσσεται η Θεωρία Πιθανοτήτων και οι Διακριτές και συνεχείς τυχαίες μεταβλητές και διμεταβλητές. Το τρίτο μέρος περιλαμβάνει τις αρχές της Τυχαίας δειγματοληψίας, την Εκτιμητική και τους Ελέγχους υποθέσεων, Παραμετρικούς, που περιλαμβάνουν και όσους γίνονται με Ανάλυση της διακύμανσης, και Μη παραμετρικούς. Στο τέταρτο μέρος αναπτύσσονται η Απλή και Πολλαπλή παλινδρόμηση.

Ακολουθεί μία σειρά επιλεγμένων ασκήσεων με ορισμένες υποδείξεις και τις απαντήσεις τους, οπότε δίνεται η δυνατότητα στον αναγνώστη να επαληθεύσει άμεσα τη λύση τους.

Το βιβλίο ολοκληρώνεται με μία σύντομη αναφορά στο στατιστικό πακέτο Minitab, με τους αναγκαίους Στατιστικούς Πίνακες και με τη σχετική βιβλιογραφία.

Στο σημείο αυτό, θέλω να ευχαριστήσω θερμά τον αγαπητό φίλο και συνάδελφο κο Ανδρέα Ξενάκη για τις ιδιαίτερα εύστοχες υποδείξεις του και την αγαπητή φίλη και συνάδελφο κα Θέμιδα Μίνογλου, των Εκδόσεων «ΚΡΙΤΙΚΗ», που ανέλαβε την έκδοση αυτού του βιβλίου.

Ιανουάριος 2010

Π. ΖΑΪΡΗΣ

## ΕΙΣΑΓΩΓΗ

Το φυσικό, κοινωνικό και οικονομικό περιβάλλον μεταβάλλεται με ολοένα αυξανόμενους ρυθμούς, όχι πάντα προς την επιθυμητή κατεύθυνση. Για να μπορούμε να παρακολουθούμε –όσο είναι δυνατόν– τις μεταβολές του, να κάνουμε διορθωτικές κινήσεις και να προβλέπουμε τη μελλοντική του μορφή, είναι αναγκαία τόσο η συχνή καταγραφή των μεγεθών που καθορίζουν την εξέλιξή του, όσο και η παρακολούθηση των διαχρονικών μεταβολών τους. Για τους ίδιους ακριβώς λόγους παρακολουθείται η διαχρονική εξέλιξη των αποτελεσμάτων κάθε μορφής επαγγελματικής δραστηριότητας.

Το έργο της καταγραφής, της διαχρονικής παρακολούθησης και της πρόβλεψης αναλαμβάνει η επιστήμη της Στατιστικής, αφήνοντας την άσκηση πολιτικής σε άλλες επιστήμες. Ο όρος Στατιστική προέρχεται κατ' άλλους από το λατινικό «status», που υπονοεί τη γενική κατάσταση του πληθυσμού, και κατ' άλλους από το αρχαίο ελληνικό ρήμα «στατίζω», που σημαίνει «τοποθετώ – ταξινομώ».

Οι πρώτες ενδείξεις διενέργειας στατιστικών ερευνών ανάγονται στους αρχαίους χρόνους, με προσπάθειες καταγραφής πληθυσμών για στρατιωτικούς ή φορολογικούς λόγους, που συνεχίστηκαν μέχρι τον 16ο αιώνα. Ουσιαστικά, όμως, η Στατιστική θεμελιώθηκε τον 17ο αιώνα με υπολογισμό απλών περιγραφικών μέτρων, όπως μέσοι όροι και ποσοστά. Ο υπολογισμός τους αποτέλεσε την απαρχή της Περιγραφικής Στατιστικής. Η ανάπτυξη της Στατιστικής άρχισε τον 19ο αιώνα, με τη θεμελίωση της Θεωρίας Πιθανοτήτων, που κύριος σκοπός της ήταν να υποστηριχθούν θεωρίες σχετικές με τα αποτελέσματα ορισμένων τυχερών παιχνιδιών. Η Θεωρία Πιθανοτήτων διατυπώνει θεωρητικά υποδείγματα (μοντέλα) που επιδιώκουν να μελετήσουν εξελίξεις με αβέβαιο χαρακτήρα.

Σήμερα, στατιστικές έρευνες διενεργούνται σε πεπερασμένους, ή όχι, πληθυσμούς για να ερευνηθεί μια μεγάλη ποικιλία χαρακτηριστικών τους. Ειδικότερα:

- Οι κοινωνικές έρευνες αναφέρονται σε ανθρώπινους πληθυσμούς και μελετούν μία μεγάλη ποικιλία καταστάσεων και δραστηριοτήτων της κοινωνικής και οικονομικής ζωής τους. Πιο συγκεκριμένα, μελετούν δημογραφικά χαρακτηριστικά όπως φύλο, ηλικία, καταγωγή, οικογενειακή κατάσταση, φυλή, φυσικά χαρακτηριστικά όπως διαστάσεις, βάρος, μορφή και κοινωνικοοικονομικά χαρακτηριστικά όπως εισοδήματα, δαπάνες, ανεργία, μετανάστευση, υγεία, συνθήκες κατοικίας, αναψυχή, εγκληματικότητα, γεννητικότητα, προσδόκιμο ζωής, χρήση μέσων συγκοινωνίας, συνθήκες διαβίωσης της τρίτης ηλικίας κ.λπ.
- Οι δημοσκοπήσεις και σφυγμομετρήσεις, πρωτεργάτης των οποίων θεωρείται ο Gallup, ελέγχουν τις απόψεις του πληθυσμού σε συγκεκριμένα κοινωνικά, οικονομικά και πολιτικά θέματα.
- Οι έρευνες αγοράς προσδιορίζουν τις προτιμήσεις των καταναλωτών, την αποτελεσματικότητα της διαφήμισης, την αναγνωσιμότητα, ακροαματικότητα και τηλεθέαση των μέσων μαζικής ενημέρωσης.
- Ειδική κατηγορία ερευνών, με καθαρά οικονομικό χαρακτήρα, οι οποίες προσδιορίζουν το ύψος της παραγωγής, την παραγωγικότητα, τις τιμές, την κατανάλωση, την αποδοτικότητα των εργαζομένων, κάνουν ποιοτικούς ελέγχους προϊόντων κ.λπ.
- Επίσης διενεργούνται γεωργικές, κτηνοτροφικές, θαλάσσιες, κλιματολογικές κ.λπ. έρευνες, όπως και μια σειρά εξειδικευμένων ερευνών στα πλαίσια συγκεκριμένων επιστημών, όπως της ιατρικής, της βιολογίας, της ψυχολογίας κ.λπ.

Τις τελευταίες δεκαετίες παρατηρείται συνεχώς αυξανόμενη χρήση της Στατιστικής σε έναν μεγάλο αριθμό επιστημών όπως η φυσική, η χημεία, η ιατρική, η οικονομία, η βιολογία και η ψυχολογία. Η εξέλιξη αυτή οδήγησε σε αλματώδη ανάπτυξη των βασικών κλάδων της που είναι η Περιγραφική Στατιστική, η Θεωρία Πιθανοτήτων και η Επαγωγική Στατιστική, τόσο σε επίπεδο θεωρίας όσο και σε επίπεδο εφαρμογών. Παράλληλα αναπτύχθηκε μια σειρά νέων κλάδων της Στατιστικής, όπως η Οικονομετρία, που, μέσω εμπειρικών δεδομένων, παρακολουθεί την εξέλιξη οικονομικών κυρίως μεταβλητών, οι Στοχαστικές Διαδικασίες, η Δημογραφία, η Θεωρία Πολυμεταβλητών, ο Σχεδιασμός Πειραμάτων, οι Τεχνικές Δειγματοληψίας, ο Ποιοτικός Έλεγχος

χος, η Ανάλυση Χρονολογικών Σειρών, η Θεωρία Αποφάσεων κ.λπ. Η ταχύτατη εξέλιξη και διάδοση των ηλεκτρονικών υπολογιστών οδήγησε στη δημιουργία ολοκληρωμένων πακέτων επεξεργασίας στατιστικών στοιχείων, με τα οποία η ανάλυσή τους έγινε σχετικά απλή διαδικασία.

## ΚΕΦΑΛΑΙΟ 1

# ΤΑ ΣΤΑΤΙΣΤΙΚΑ ΣΤΟΙΧΕΙΑ ΚΑΙ ΟΙ ΜΕΘΟΔΟΙ ΣΥΛΛΟΓΗΣ ΤΟΥΣ

### 1.1. Είδη των στατιστικών στοιχείων

Η Στατιστική ασχολείται με τη συλλογή, παρουσίαση και ανάλυση των «στατιστικών στοιχείων», δηλαδή πληροφοριών που συλλέγονται για να μελετηθούν συγκεκριμένα χαρακτηριστικά μιας πολυπληθούς ομάδας μελών, που λέγεται «πληθυσμός». Για παράδειγμα, από τον πληθυσμό των νοικοκυριών μιας πόλης συλλέγονται στατιστικά στοιχεία για να μελετηθεί ο αριθμός μελών τους, το εισόδημά τους, ο τόπος κατοικίας τους, το αν έχουν internet κ.λπ. Οι τιμές κάθε χαρακτηριστικού ενός πληθυσμού αποτελούν ένα «στατιστικό πληθυσμό» ή «πληθυσμό τιμών», και κάποιες μπορούν να επαναλαμβάνονται. Έτσι, οι αριθμοί μελών 3, 1, 4, 4, 5, 2, 2, ... ενός πληθυσμού νοικοκυριών, είναι ένας πληθυσμός τιμών. Σημειώνεται ότι ως πληθυσμός τιμών μπορεί να θεωρηθεί και κάθε διαχρονική εξέλιξη, όπως οι καθημερινές τιμές μιας μετοχής για μία τριακονταετία αλλά και κάθε συνεχές, όπως το νερό μιας λίμνης.

Η συλλογή στατιστικών στοιχείων για να ερευνηθούν διάφορα χαρακτηριστικά ενός πληθυσμού γίνεται με ειδικές στατιστικές έρευνες και η παρουσίασή τους γίνεται με ειδικούς πίνακες και διαγράμματα. Η ανάλυσή τους, δηλαδή η εύρεση αντιπροσωπευτικών τιμών που τα εκφράζουν συνοπτικά, επιτρέπει τη διατύπωση συμπερασμάτων και προβλέψεων τόσο για τα ίδια τα ερευνώμενα χαρακτηριστικά, όσο και για τις υπάρχουσες σχέσεις μεταξύ τους μέσα στον πληθυσμό.

Τα στατιστικά στοιχεία που συλλέγονται για κάποιο χαρακτηριστικό ενός πληθυσμού  $N$  μελών, εκφράζονται από μία μεταβλητή  $X$  με  $k$  ( $\leq N$ ) διαφορετικές μεταξύ τους τιμές. Έτσι, οι αριθμοί μελών 3, 2, 3, 2, 2, 3, 3, 5, 1, 1 του πληθυσμού των  $N = 10$  νοικοκυριών μιας πολυκα-

τοικίας, εκφράζονται από τη μεταβλητή  $X$ , που περιλαμβάνει  $k = 4$  τιμές, τις 1, 2, 3, 5. Η μεταβλητή αυτή μπορεί να είναι «ποσοτική» ή «ποιοτική».

Μία μεταβλητή θα είναι ποσοτική, αν μετριέται με μία αριθμητική κλίμακα. Θα λέγεται «διακριτή», αν το πλήθος των τιμών της είναι πεπερασμένο ή το πολύ αριθμήσιμο, όπως ο αριθμός μελών των νοικοκυριών ή το σύνολο των άρτιων αριθμών, και «συνεχής», αν μπορεί να πάρει οποιαδήποτε τιμή μέσα σε ένα διάστημα (που είναι μη αριθμήσιμο σύνολο τιμών), όπως κάθε μήκος, εμβαδόν ή όγκος, κάθε χρονική περίοδος κ.λπ. Σημειώνεται ότι κάθε χρηματικό μέγεθος θεωρείται επίσης συνεχής μεταβλητή.

Η κλίμακα μέτρησης των ποσοτικών μεταβλητών μπορεί να είναι:

- (α) «Κλίμακα λόγου»: Είναι η πληρέστερη κλίμακα μέτρησης και χρησιμοποιείται στις περισσότερες ποσοτικές μεταβλητές. Οι τιμές της εκφράζουν διάταξη (μικρότερο-μεγαλύτερο) και η τιμή 0 σημαίνει έλλειψη της ποσότητας που μετριέται. Ίσες διαφορές μεταξύ τιμών σημαίνουν ίσες διαφορές μεταξύ των ποσοτήτων που μετρούν, και ο λόγος δύο τιμών δίνει τη σχέση μεταξύ των αντίστοιχων μεγεθών.
- (β) «Κλίμακα διαστημάτων»: Οι τιμές της καθορίζονται αυθαίρετα. Εκφράζουν διάταξη, αλλά η τιμή 0 δεν υπονοεί έλλειψη της ποσότητας που μετριέται. Ίσες διαφορές μεταξύ τιμών σημαίνουν ίσες διαφορές μεταξύ των ποσοτήτων που μετρούν, αλλά ο λόγος δύο τιμών δεν δίνει τη σχέση μεταξύ των αντίστοιχων μεγεθών. Κλασικά παραδείγματα κλίμακας διαστημάτων είναι οι κλίμακες θερμοκρασίας και βαθμολογίας επιδόσεων. Έτσι, η θερμοκρασία  $30^\circ$  δεν σημαίνει διπλάσια ποσότητα θερμότητας από τη θερμοκρασία  $15^\circ$ , ενώ η θερμοκρασία  $0^\circ$  δεν υπονοεί έλλειψη θερμότητας.

Μία μεταβλητή θα είναι ποιοτική, αν κατατάσσει τα μέλη του πληθυσμού σε προκαθορισμένες ποιοτικές ή γεωγραφικές κατηγορίες, όπως φύλο, μόρφωση, καταγωγή, ή μοιράζει σε προκαθορισμένες κατηγορίες τη συνολική τιμή ενός χαρακτηριστικού, όπως οι συνολικές δαπάνες ενός νοικοκυριού που μοιράζονται σε δαπάνες κατοικίας, διατροφής, ένδυσης κ.λπ. Μια ποιοτική μεταβλητή θα λέγεται:

- (i) «Ιεραρχική» ή «διατάξιμη», αν οι κατηγορίες της ιεραρχούνται συνήθως προς το καλύτερο, όπως η μόρφωση, η υπηρεσιακή κατάσταση, η υγεία κ.λπ. Οι ποιοτικές ιεραρχικές μεταβλητές συνήθως ποσοτικοποιούνται για διευκόλυνσή μας, αν βαθμολογήσουμε τις κατηγορίες τους με «τάξεις» 1, 2, 3..., συνήθως από την ασθενέστερη προς την ισχυρότερη, χωρίς όμως ίσες διαφορές μεταξύ δύο τάξεων να υπονοούν ίσες διαφορές μεταξύ των καταστάσεων που εκφράζουν. Για παράδειγμα, στην ποιοτική ιεραρχική μεταβλητή «Μόρφωση», η κατηγορία «Όχι υποχρεωτική εκπαίδευση» έχει τάξη 1, η κατηγορία «Υποχρεωτική εκπαίδευση» έχει τάξη 2, η κατηγορία «Απολυτήριο Λυκείου» έχει τάξη 3 κ.λπ.
- (ii) «Κατηγορική» ή «μη διατάξιμη», αν οι κατηγορίες της δεν ιεραρχούνται, όπως φύλο, καταγωγή κ.λπ.

Σημειώνεται, τέλος, ότι ποσοτικά ή ποιοτικά στατιστικά στοιχεία θα λέγονται «διαστρωματικά» αν αναφέρονται σε μία συγκεκριμένη χρονική περίοδο και «χρονολογικά» αν αναφέρονται στη διαχρονική εξέλιξη ενός χαρακτηριστικού.

## 1.2. Συλλογή των στατιστικών στοιχείων

Η διαδικασία της συλλογής των στατιστικών στοιχείων στα πλαίσια μιας στατιστικής έρευνας είναι ιδιαίτερα σημαντική, γιατί απ' αυτήν εξαρτάται η ποιότητα των στοιχείων, άρα και η αξιοπιστία των αποτελεσμάτων της έρευνας.

### 1.2.1. Μέθοδοι συλλογής στοιχείων

Η συλλογή των στατιστικών στοιχείων, για να ερευνηθεί ένα χαρακτηριστικό του πληθυσμού, μπορεί να γίνει ή με απογραφή, ή με δειγματοληψία από τον ίδιο τον πληθυσμό, ή από υπάρχοντα αρχεία. Ειδικότερα:

1. «Απογραφή» είναι η συλλογή των στοιχείων από ολόκληρο τον πληθυσμό. Η ανάλυσή τους (εύρεση αντιπροσωπευτικών τιμών τους) και η εξαγωγή συμπερασμάτων γίνεται από την Περιγραφική Στατιστική.



2. «Δειγματοληψία» είναι η επιλογή ενός μικρού μέρους του πληθυσμού, που λέγεται δείγμα, και η συλλογή των στοιχείων γίνεται μόνο από τα μέλη του δείγματος. Ειδικά αν το δείγμα είναι τυχαίο, δηλαδή αν όλα τα μέλη του πληθυσμού έχουν την ίδια δυνατότητα να επιλεγούν στο δείγμα, μπορούμε με κατάλληλους χειρισμούς να προσεγγίσουμε ικανοποιητικά την πραγματική τιμή του χαρακτηριστικού στον πληθυσμό. Η ανάλυση των δειγματικών στοιχείων γίνεται από την Περιγραφική Στατιστική, ενώ η γενίκευση των αποτελεσμάτων που συνάγονται από το δείγμα, σε ολόκληρο τον πληθυσμό, γίνεται από την Επαγωγική Στατιστική.

Η επιλογή ενός τυχαίου δείγματος από έναν πληθυσμό μπορεί να γίνει ή με δειγματοληψία μελών ή με δειγματοληψία ομάδων.

Κατά τη δειγματοληψία μελών, τα μέλη του τυχαίου δείγματος επιλέγονται απευθείας από τον πληθυσμό. Αν όμως ο πληθυσμός είναι μεγάλος και γεωγραφικά διεσπαρμένος, οι μετακινήσεις των ερευνητών που θα τα ερευνήσουν κοστίζουν πολύ σε χρόνο και χρήμα. Σ' αυτή την περίπτωση, μοιράζουμε τον πληθυσμό σε ομάδες γειτονικών μελών, επιλέγουμε ένα τυχαίο δείγμα ομάδων και, στην απλούστερη περίπτωση, ερευνούμε όλα τα μέλη τους. Για παράδειγμα, αν θέλουμε ένα τυχαίο δείγμα 30 νοικοκυριών από τον πληθυσμό των 300 νοικοκυριών ενός χωριού, επιλέγουμε εύκολα 30 κατοικίες και ερευνούμε τα νοικοκυριά τους. Η λήψη στοιχείων θα είναι πολύ σύντομη επειδή οι μετακινήσεις μας από κατοικία σε κατοικία είναι πολύ μικρές. Αν όμως θέλουμε δείγμα περίπου 1000 νοικοκυριών της Θεσσαλίας, τα νοικοκυριά που θα επιλεγούν θα είναι πιθανότατα σε διαφορετικές πόλεις και χωριά, οπότε οι μετακινήσεις θα κοστίσουν πολύ σε χρόνο και χρήμα. Γι' αυτό, θεωρούμε όλα τα νοικοκυριά της Θεσσαλίας ομαδοποιημένα στα οικοδομικά τετράγωνα των κατοικιών τους και γνωρίζοντας ότι έχουν κατά μέσον όρο από 50 νοικοκυριά, επιλέγουμε ένα δείγμα 20 τετραγώνων και ερευνούμε όλα τα περίπου 1000 νοικοκυριά τους. Η επίσκεψη σε 20 τετράγωνα, ορισμένα από τα οποία πιθανόν να βρίσκονται στην ίδια πόλη, είναι πολύ πιο σύντομη από την επίσκεψη 1000 μεμονωμένων νοικοκυριών σε ολόκληρη τη Θεσσαλία. Η δειγματοληψία ομάδων αποτελεί αντικείμενο ενός νεότερου κλάδου της Στατιστικής, που λέγεται «Τεχνικές Δειγματοληπτικών Ερευνών» και δεν θα μας απασχολήσει στα επόμενα.

Τα «δειγματοληπτικά σχέδια» της τυχαίας δειγματοληψίας, δηλαδή οι

μέθοδοι με τις οποίες επιλέγεται ένα τυχαίο δείγμα μελών του πληθυσμού, είναι:

**(α)** «Απλή τυχαία δειγματοληψία»: Από αριθμημένο πληθυσμό  $N$  μελών επιλέγεται ένα δείγμα  $n$  μελών (ή ομάδων) του, έτσι ώστε κάθε μέλος του πληθυσμού να έχει την ίδια δυνατότητα να επιλεγεί στο δείγμα. Η απλή τυχαία δειγματοληψία γίνεται με κλήρωση (αν ο πληθυσμός είναι μικρός) ή από πίνακα τυχαίων αριθμών.

Ο πίνακας τυχαίων αριθμών περιέχει πολυψήφιους αριθμούς (συνήθως με 35 ή 40 ψηφία), χωρισμένους σε πενταψήφια τμήματα που γράφονται το ένα κάτω από το άλλο. Κάθε ψηφίο αυτών των αριθμών επιλέγεται με τυχαίο τρόπο από τα ψηφία 0, 1, 2, ..., 9. Ενδεικτικό είναι το παρακάτω μέρος ενός πίνακα τυχαίων αριθμών, η χρήση του οποίου γίνεται κατανοητή με την επόμενη εφαρμογή.

.....

.....

25209	64846	25256	00839	.....
34231	74654	38125	45537	.....
31231	58390	86610	65282	.....
67673	30980	25855	30918	.....
62413	39808	02913	98976	.....
74831	74553	30801	76782	.....

.....

.....

**ΕΦΑΡΜΟΓΗ 1.1.** Από αριθμημένο πληθυσμό  $N = 82$  μελών να επιλεγεί ένα απλό τυχαίο δείγμα  $n = 4$  μελών, επιλέγοντας στην τύχη να χρησιμοποιήσουμε τις πρώτες στήλες του τρίτου πενταψήφιου τμήματος του παραπάνω πίνακα.

*Λύση*

Επειδή ο  $N = 82$  είναι διψήφιος, θεωρούμε το αρχικό διψήφιο τμήμα των πενταψήφιων της τρίτης στήλης με τους αριθμούς 25, 38, 86, 25, 02 και 30. Στο απλό τυχαίο δείγμα μας θα πάρουμε το 25ο μέλος του πληθυσμού, κατόπιν το 38ο, παραλείπουμε τον αριθμό 86 γιατί  $N = 82$ , άρα δεν υπάρχει 86ο μέλος του πληθυσμού, παραλείπουμε και τον αριθμό 25 γιατί τον έχουμε ήδη πάρει και συνεχίζουμε παίρνοντας ακόμη το 2ο και το 30ο μέλος του πληθυσμού.

**(β)** «Συστηματική δειγματοληψία». Η επιλογή ενός συστηματικού δείγματος  $n$  μελών από αριθμημένο πληθυσμό  $N$  μελών γίνεται ως εξής: Υπολογίζουμε το πηλίκον  $\delta = N/n$ , που λέγεται βήμα, και στην απλούστερη περίπτωση είναι ακέραιο. Βρίσκουμε έναν τυχαίο αριθμό  $k$  από το 1 μέχρι και το  $\delta$  και παίρνουμε στο δείγμα τα  $n$  μέλη του πληθυσμού που έχουν αριθμηση

$$k, k + \delta, k + 2\delta, \dots, k + (n - 1)\delta.$$

**ΕΦΑΡΜΟΓΗ 1.2.** Από αριθμημένο πληθυσμό  $N = 1000$  μελών να επιλεγεί το συστηματικό δείγμα  $n = 50$  μελών, με τυχαίο  $k = 8$ .

*Λύση*

Βρίσκουμε το πηλίκον  $\delta = 1000/50 = 20$  και έστω τον τυχαίο αριθμό  $k = 8$ . Το δείγμα θα περιλάβει το 8ο, το  $(8 + 20 =)28ο$ , το 48ο, ... και το 988ο μέλος του αριθμημένου πληθυσμού.

**(γ)** «Στρωματοποιημένη δειγματοληψία»: Διενεργείται αν ο πληθυσμός μπορεί να μοιραστεί σε ομοιογενή στρώματα, δηλαδή σε κατηγορίες όπου οι τιμές κάθε κατηγορίας, ως προς το χαρακτηριστικό που ερευνάται, πλησιάζουν πολύ μεταξύ τους. Έτσι, σε έρευνα εισπράξεων των καταστημάτων μιας πόλης, με βάση την πληροφόρησή μας για τις εισπράξεις τους, τα στρωματοποιούμε σε μικρά, μεσαία και μεγάλα. Επιλέγουμε ένα απλό τυχαίο ή συστηματικό δείγμα από κάθε στρώμα και εκτιμούμε το χαρακτηριστικό μας για το στρώμα αυτό. Κατόπιν, συνθέτουμε κατάλληλα αυτές τις εκτιμήσεις και βρίσκουμε την εκτίμηση για ολόκληρο τον πληθυσμό.

**3.** Η συλλογή στοιχείων μπορεί ακόμα, όπως είπαμε, να γίνει απογραφικά ή δειγματοληπτικά από αρχεία που έχουν ήδη καταρτιστεί από συγκεκριμένους φορείς, οι οποίοι τα έχουν συλλέξει είτε για ερευνητικούς σκοπούς είτε απλώς για τη δημιουργία Τράπεζας Πληροφοριών. Πρέπει όμως τα στοιχεία τους να είναι πρόσφατα και χωρίς διπλοεγγραφές μελών του πληθυσμού (οι Τ/Φ κατάλογοι έχουν διπλοεγγραφές, γιατί ορισμένα μέλη του πληθυσμού μπορούν να έχουν περισσότερες από μία Τ/Φ γραμμές). Τέτοιοι φορείς είναι η Στατιστική Υπηρεσία κάθε χώρας, η οποία οργανώνει το Στατιστικό Σύστημα της χώρας της, Υπουργεία, Δημόσιες και Δημοτικές Υπηρεσίες, Τράπεζες, Κοινωνικά Ιδρύματα, Επαγγελματικοί Σύλλογοι κ.λπ.

Σημειώνεται, τέλος, ότι η συλλογή στοιχείων μπορεί να γίνει και από μη αριθμημένους πληθυσμούς, όπως από τα αυτοκίνητα που περνούν σε συγκεκριμένη χρονική περίοδο από μία διασταύρωση.

### 1.2.2. Σφάλματα

Σε μία στατιστική έρευνα, η τιμή  $x$  που θα βρεθεί για ένα χαρακτηριστικό του πληθυσμού, σχεδόν πάντα διαφέρει από την πραγματική του τιμή  $X$ , δηλαδή περιέχει σφάλμα  $\Sigma$ , και γι' αυτό λέγεται εκτίμηση της πραγματικής τιμής  $X$ .

Το σφάλμα  $\Sigma$  εκφράζεται ή ως «απόλυτο σφάλμα»

$$\Sigma = |X - x|$$

ή ως «σχετικό σφάλμα»

$$\Sigma\% = \frac{|X - x|}{X} \cdot 100\%,$$

που είναι ανεξάρτητο μονάδων.

Οι τιμές των σφαλμάτων προφανώς δεν υπολογίζονται, επειδή η πραγματική τιμή  $X$  στον πληθυσμό είναι άγνωστη. Με κατάλληλες όμως μεθόδους προσδιορίζουμε το μέγιστο αναμενόμενο απόλυτο σφάλμα  $a$  ή το μέγιστο σχετικό σφάλμα  $\varepsilon\%$  της εκτίμησης  $x$  και κατασκευάζουμε ένα διάστημα μέσα στο οποίο περιμένουμε να βρίσκεται η πραγματική τιμή  $X$  του πληθυσμού. Ειδικότερα:

Αν  $|X - x| \leq a$ , τότε  $-a \leq X - x \leq a$ , άρα  $x - a \leq X \leq x + a$ .

Επίσης, αν

$$\frac{|X - x|}{x} \cdot 100 \leq \varepsilon\% \rightarrow |X - x| \leq \frac{\varepsilon \cdot x}{100}, \text{ άρα } x - \frac{\varepsilon \cdot x}{100} \leq X \leq x + \frac{\varepsilon \cdot x}{100}.$$

**ΕΦΑΡΜΟΓΗ 1.3.** Αν το μέγεθος του πληθυσμού μιας πόλης βρέθηκε  $x = 30000$ , με σφάλμα (i)  $a = 1000$  και (ii)  $\varepsilon\% = 2\%$ , σε ποιο διάστημα αναμένεται να περιλαμβάνεται το πραγματικό του μέγεθος  $X$ ;

*Λύση*

$$(i) 30000 - 1000 \leq X \leq 30000 + 1000 \rightarrow 29000 \leq X \leq 31000$$

$$(ii) 30000 - \frac{2 \cdot 30000}{100} \leq X \leq 30000 + \frac{2 \cdot 30000}{100} \rightarrow 29400 \leq X \leq 30600.$$

Το σφάλμα  $\Sigma$ , που υπεισέρχεται κατά την εκτίμηση ενός χαρακτηριστικού στον πληθυσμό, μπορεί να είναι:

**(α)** «Δειγματοληπτικό σφάλμα»: Εμφανίζεται μόνο σε δειγματοληψία και οφείλεται στο ότι διαφορετικά δείγματα δίνουν διαφορετικές εκτιμήσεις του χαρακτηριστικού στον πληθυσμό. Αν η δειγματοληψία είναι τυχαία, με κατάλληλες μεθόδους εκτιμάται το μέγεθός του και είναι δυνατή η μείωσή του σε αποδεκτά επίπεδα.

**(β)** «Μη δειγματοληπτικό σφάλμα»: Εμφανίζεται και σε απογραφή και σε δειγματοληψία, το μέγεθός του δεν μπορεί να προσδιοριστεί και οφείλεται:

- Σε μη κάλυψη ολόκληρου του πληθυσμού. Ενδεικτικά, σε έρευνα φοιτητών στο αμφιθέατρο, οι εργαζόμενοι φοιτητές απουσιάζουν.
- Σε μη ανταπόκριση ορισμένων ερευνώμενων, οι οποίοι ή δεν εντοπίζονται ή αρνούνται να απαντήσουν.
- Σε εσφαλμένα στοιχεία από άγνοια ή ψευδείς δηλώσεις των ερευνώμενων.
- Σε απειρία του ερευνητή. Για παράδειγμα, σε έρευνα μισθών, ο άπειρος ερευνητής δεν διευκρινίζει αν ζητείται ο καθαρός ή ο ακαθάριστος μισθός. Άρα, αν κάποιος εργαζόμενος δώσει τον καθαρό και οι υπόλοιποι τον ακαθάριστο μισθό, ο μέσος τους δεν εκφράζει τίποτε.

Σημειώνεται, τέλος, ότι αν οι λαμβανόμενες τιμές αναμένεται από την εμπειρία μας να είναι συνήθως μεγαλύτερες (ή συνήθως μικρότερες) από τις πραγματικές, το μη δειγματοληπτικό σφάλμα λέγεται «συστηματικό». Για παράδειγμα, οι ηλικίες που δηλώνουν οι περισσότερες κυρίες είναι συνήθως μικρότερες από τις πραγματικές. Αν όχι, το σφάλμα λέγεται «μη συστηματικό».

### 1.2.3. Ερωτηματολόγια

Τα συλλεγόμενα στοιχεία από ολόκληρο τον πληθυσμό ή από ένα δείγμα του καταγράφονται σε ειδικά ερωτηματολόγια που καταρτίζονται με ιδιαίτερη προσοχή.

Οι βασικές αρχές στις οποίες στηρίζεται η κατάρτιση ενός ερωτηματολογίου είναι:

- Το ερωτηματολόγιο καταρτίζεται από εξειδικευμένο στατιστικό σε συνεργασία με ειδικό πάνω στο αντικείμενο της έρευνας.
- Τα ερωτήματα πρέπει να είναι τα απολύτως απαραίτητα, για να μην κουράζουν ερευνώμενο και ερευνητή αλλά και για να αποφεύγεται μη ανταπόκριση λόγω μεγάλου ερωτηματολογίου.
- Τα ερωτήματα πρέπει να είναι σαφή, με τις απαραίτητες διευκρινίσεις. Για παράδειγμα, σε έρευνα μισθών πρέπει να διευκρινίζεται αν πρόκειται για καθαρούς ή ακαθάριστους μισθούς.
- Τα ερωτήματα πρέπει να είναι διατυπωμένα έτσι ώστε να μην κατευθύνουν τους ερευνώμενους σε μια συγκεκριμένη απάντηση. Το ερώτημα, «Δεν νομίζετε ότι οι ποινές που επιβάλλουν τα δικαστήρια είναι πολύ επεικεείς;» κατευθύνει τον ερευνώμενο να απαντήσει καταφατικά.
- Σε ερωτήματα άποψης, οι απαντήσεις συνήθως είναι:  
ΔΙΑΦΩΝΩ ΡΙΖΙΚΑ – ΔΙΑΦΩΝΩ – ΔΕΝ ΕΧΩ ΓΝΩΜΗ – ΣΥΜΦΩΝΩ – ΣΥΜΦΩΝΩ ΑΠΟΛΥΤΑ και ποσοτικοποιούνται σε κλίμακα με αντίστοιχες τιμές  $-2, -1, 0, 1, 2$ .
- Ευαίσθητα ερωτήματα, σχετικά με ψυχικές ασθένειες, σεξουαλική ζωή, εγκληματικότητα, χρήση ναρκωτικών κ.λπ., πρέπει να υποβάλλονται με ιδιαίτερη προσοχή, στο τέλος του ερωτηματολογίου. Διαφορετικά, μπορεί ο ερευνώμενος να μην ανταποκριθεί σε όλα τα επόμενα ερωτήματα.

Τα ερωτηματολόγια συμπληρώνονται ή από τον ερευνητή με προσωπική συνέντευξη στο χώρο του ερευνώμενου, ή τηλεφωνικώς, ή παραδίδονται στον ερευνώμενο για συμπλήρωση και άμεση επιστροφή, ή ακόμα αποστέλλονται ταχυδρομικώς ή με e-mail για να συμπληρωθούν από τον ερευνώμενο και να επιστραφούν με τον ίδιο τρόπο.

### 1.3. Πλεονεκτήματα της τυχαίας δειγματοληψίας

Η δειγματοληπτική έρευνα με τυχαία δειγματοληψία πλεονεκτεί της απογραφικής, γιατί γίνεται ταχύτερα και τα αποτελέσματά της δίνουν την τρέχουσα πραγματικότητα, πράγμα που δύσκολα επιτυγχάνει η

απογραφή λόγω συλλογής και επεξεργασίας μεγάλου όγκου στοιχείων. Γίνεται με μικρότερο κόστος αλλά και με λιγότερο προσωπικό που είναι ειδικά εκπαιδευμένο, πράγμα που εξασφαλίζει πολύ μικρά μη δειγματοληπτικά σφάλματα. Πράγματι, στην απογραφή χρησιμοποιούνται πολλοί και όχι έμπειροι ερευνητές, οπότε το μέγεθος των μη δειγματοληπτικών σφαλμάτων μπορεί να είναι πολύ μεγάλο. Επίσης, στην τυχαία δειγματοληψία το δειγματοληπτικό σφάλμα μπορεί να εκτιμηθεί προσεγγιστικά και με κατάλληλες μεθόδους να περιοριστεί σε αποδεκτό επίπεδο. Άρα η τυχαία δειγματοληψία είναι πιο αξιόπιστη από την απογραφή. Σημειώνεται, τέλος, ότι η δειγματοληψία είναι η μοναδική μέθοδος συλλογής στοιχείων αν η απογραφή είναι αδύνατη, όπως σε έρευνα για τη μόλυνση της θάλασσας ή σε εξέταση αίματος ενός ατόμου ή αν με απογραφή καταστραφεί ολόκληρη η παραγωγή, όπως σε έρευνα για τη διάρκεια λειτουργίας ενός τύπου λαμπών.