

ΚΕΦΑΛΑΙΟ 1

ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ

1.1 Γενικότητες

Συχνά βλέπουμε όλοι μας στον ημερήσιο και περιοδικό τύπο τα αποτελέσματα δειγματοληπτικών ερευνών που αναφέρονται στην παραγωγή ή κατανάλωση διαφόρων προϊόντων (π.χ. συνέπειες του καπνίσματος για την υγεία), στον προσδιορισμό των αποφάσεων ή επιθυμιών της κοινής γνώμης για πολιτικά θέματα, τρέχοντα κοινωνικο-οικονομικά προβλήματα ή τη στάση τους απέναντι σε διάφορα περιβαλλοντικά γεγονότα κ.λπ.

Είναι όμως εξίσου συχνό το φαινόμενο η τεχνική και οι διαδικασίες της δειγματοληπτικής έρευνας να εφαρμόζονται με ένα χονδροειδή και ερασιτεχνικό τρόπο. Δίνεται έτσι η εντύπωση στον καθένα ότι είναι εύκολο να αναλάβει τη διεξαγωγή και ανάλυση μιας δειγματοληπτικής έρευνας παραβλέποντας τα διάφορα βήματα που απαιτούνται για μια προσεκτικά διενεργούμενη δειγματοληπτική έρευνα. Το να πάρουμε μερικά τηλεφωνήματα ή να βγούμε στο δρόμο μ' ένα μικρόφωνο παίρνοντας συνεντεύξεις από διάφορα άτομα, είναι προφανές ότι, το μόνο που δεν μπορούμε να κάνουμε είναι να παρουσιάσουμε τα αποτελέσματά τους σαν μαρτυρίες έρευνας.

Ένας τρόπος για να πάρουμε τις απαραίτητες πληροφορίες που χρειαζόμαστε από κάποιο πληθυσμό είναι η εξέταση όλων των μονάδων του πληθυσμού. Π.χ. για να βρούμε τη μέση επίδοση των μαθητών μιας τάξης σε κάποιο μάθημα χρειάζεται η βαθμολογία όλων των μαθητών. Σε μία περιοχή που υπάρχει κάποιο σπουδαίο περιβαλλοντικό γεγονός, όπως π.χ. ένας αρχαιολογικός χώρος, ένα σπήλαιο, μία λίμνη, ένας δρυμός κ.λπ. θα μας ενδιέφερε η γνώμη όλων των κατοίκων για τη γενικότερη τουριστική αξιοποίηση του γεγονότος αυτού με όλα τα δυνατά ίσως επακόλουθα που θα μπορούσαν να επιδράσουν στα ήθη και έθιμα της περιοχής. Επίσης η βιομηχανική αξιοποίηση μιας περιοχής θα είχε

σαν επακόλουθο τη μόλυνση του περιβάλλοντος, ρίχνοντας, για παράδειγμα, τα βιομηχανικά απόβλητα σε ένα παρακείμενο ποταμό.

Η μέθοδος αυτή συλλογής πληροφοριών, η εξέταση δηλαδή όλων των μονάδων που απαρτίζουν τον πληθυσμό που μας ενδιαφέρει, καλείται **ολική απογραφή** (census). Για παράδειγμα, η Στατιστική Υπηρεσία της χώρας μας (ΕΣΥΕ) κάνει κάθε 10 χρόνια απογραφή του πληθυσμού, η οποία αποτελεί κύρια πηγή δεδομένων δημογραφικού οικονομικού, εμπορικού και βιομηχανικού χαρακτήρα. Η τελευταία απογραφή έγινε το 2001.

Σε πολλές όμως περιπτώσεις η εξέταση όλων των μονάδων του πληθυσμού είναι δύσκολη ή ακόμα και αδύνατη όταν κυρίως συνεπάγεται καταστροφή των εξεταζόμενων αντικειμένων (ηλεκτρικές λυχνίες, εκρηκτικοί μηχανισμοί, κ.λπ.). Επίσης, ο γιατρός για να υπολογίσει την αποτελεσματικότητα ενός νέου φαρμάκου στην καταπολέμηση μιας ασθένειας είναι αδύνατο να περιμένει να δοκιμαστεί το φάρμακο σε όλα τα άτομα που πάσχουν από τη συγκεκριμένη ασθένεια. Έτσι, λόγω χρονικών και κυρίως κοστολογικών παραγόντων, ολοένα και περισσότερο επιβάλλεται η διενέργεια δειγματοληπτικών ερευνών σε μία προσπάθεια να συλλεγούν οι αναγκαίες πληροφορίες από κάποια μικρή ομάδα ή υποσύνολο του πληθυσμού κατά τέτοιο τρόπο ώστε τα αποτελέσματα που προκύπτουν από την εξέταση της ομάδας αυτής να είναι **αντιπροσωπευτικά** για τον μελετώμενο πληθυσμό. Ο ερευνητής κάνει τις παρατηρήσεις του σ' αυτήν την ομάδα και μετά γενικεύει τα αποτελέσματα για ολόκληρο τον πληθυσμό. Αυτή η ομάδα ή υποσύνολο του πληθυσμού καλείται **δείγμα** (sample).

Οι αρχές και οι μέθοδοι για τη συλλογή και ανάλυση δεδομένων από πεπερασμένους πληθυσμούς είναι ένας κλάδος της Στατιστικής γνωστός σαν "Μέθοδοι Δειγματοληπτικών Ερευνών" (Sample Survey Methods) ή γενικότερα σαν **Δειγματοληψία** (Sampling). Η ανάλυση δειγματοληπτικών ερευνών είναι το κυριότερο εργαλείο έρευνας σε ένα μεγάλο φάσμα εφαρμογών, όπως στη Βιομηχανία, Εκπαίδευση, Ιατρική, Ψυχολογία, Κοινωνιολογία, Γεωπονία κ.λπ. Οι δειγματοληπτικές έρευνες των ανθρωπίνων πληθυσμών (συχνά αναφερόμενες και σαν δημοσκοπήσεις) αποτελούν σπουδαίες πηγές βασικής γνώσης των κοινωνικών επιστημών. Οικονομολόγοι, ψυχολόγοι, πολιτικοί επιστήμονες και κοινωνιολόγοι μελετούν ποικίλα θέματα όπως πρότυπα εσόδων-εξόδων των οικογενειών, την επίδραση της επαγγελματικής απασχόλησης των γυναικών στην οικογενειακή ζωή, τις συγκοινωνιακές και ταξιδιωτικές συνήθειες των κατοίκων μιας πόλης, τις προτιμήσεις των ψηφοφόρων για τους υποψήφιους και τις θέσεις αυτών. Η Εθνική Στατιστική Υπηρεσία κάθε χώρας διενεργεί σε τακτά χρονικά διαστήματα δειγματοληπτικές έρευνες για να πάρει πληροφορίες για τον πληθωρισμό, την απασχόληση και την ανεργία στη χώρα. Ανάλογα με τα αποτελέσματα διαμορφώνεται και η κυβερνητική πολιτική στα θέματα αυτά.

1.2 Έννοια του πληθυσμού και του δείγματος

Μολονότι οι πληθυσμοί μπορεί να κυμαίνονται από μικροί και πεπερασμένοι μέχρι άπειροι και υποθετικοί, σε μία δειγματοληπτική έρευνα ενδιαφερόμαστε κυρίως για πεπερασμένους πληθυσμούς οι οποίοι πρέπει να ορίζονται πλήρως, δηλαδή να μπορούν να καταγραφούν και να μετρηθούν. Ο **ακριβής προσδιορισμός** του πληθυσμού για τον οποίο ενδιαφερόμαστε είναι το **πρώτο βήμα** σε μία δειγματοληπτική έρευνα. Ο όρος πληθυσμός χρησιμοποιείται για να περιγράψει οποιαδήποτε ομάδα ατόμων, αντικειμένων ή παρατηρήσεων για τα οποία συμβαίνει να ενδιαφερόμαστε. Παρακάτω θα δώσουμε τις έννοιες του πληθυσμού και του δείγματος όπως αυτές χρησιμοποιούνται στις δειγματοληπτικές έρευνες, ενώ στην επόμενη παράγραφο θα δώσουμε περισσότερα στοιχεία για τον πλήρη καθορισμό του πληθυσμού.

Απλό στοιχείο (elementary unit or element) ορίζεται κάθε μονάδα ενός συνόλου στην οποία γίνεται μία διαδικασία μέτρησης ή παρατήρησης κάποιας ιδιότητας και η καταγραφή των αποτελεσμάτων. Για παράδειγμα, από τους μαθητές ενός σχολείου που θέλουμε να εκτιμήσουμε κάποιο χαρακτηριστικό όπως βάρος, ύψος, επίδοση, προτίμηση για κάποιο άθλημα, κ.λπ., θα πάρουμε σαν στοιχείο τον μαθητή.

Πληθυσμός (population) ή ολότητα ορίζεται ένα βασικό σύνολο στοιχείων που πρόκειται να μελετηθεί ως προς μία ή περισσότερες χαρακτηριστικές ιδιότητες. Είναι το σύνολο δηλαδή των απλών στοιχείων για τα οποία μας ενδιαφέρει να βγάλουμε κάποια συμπεράσματα.

Δειγματοληπτικές μονάδες (sampling units or enumeration units) είναι μη επικαλυπτόμενες συλλογές απλών στοιχείων, όπως οι τάξεις διαφόρων σχολείων, τα νοικοκυριά μιας περιοχής ορίζουν δειγματοληπτικές μονάδες.

Για παράδειγμα, εάν θέλουμε να διεξάγουμε μια δειγματοληπτική έρευνα για να εκτιμήσουμε το σύνολο των ατόμων (ή το ποσοστό) που διαμένουν στην Αθήνα και δεν έχουν επισκεφτεί ποτέ οδοντίατρο, τότε ο πληθυσμός αποτελείται από όλα τα άτομα που διαμένουν στην Αθήνα και απλή μονάδα (element) είναι κάθε άτομο που διαμένει στην Αθήνα. Δειγματοληπτική μονάδα μπορεί να είναι όμως το νοικοκυριό, καθόσο είναι πιο εύκολο να έχουμε ένα κατάλογο, μία λίστα (βλ. παρακάτω) νοικοκυριών παρά ένα κατάλογο ατόμων. Η δειγματοληπτική μονάδα πρέπει να ορίζεται πλήρως, με σαφήνεια και χωρίς παρερμηνείες ώστε να είναι δυνατή η γενίκευση των συμπερασμάτων που θα προκύψουν από τη δειγματοληψία (δείγμα) σ' ολόκληρο τον ερευνώμενο πληθυσμό. Οι ερευνητές - συνεντεύκτες που θα διεξαγάγουν την έρευνα πρέπει να είναι πλήρως ενημερωμένοι με τον ορισμό των μονάδων αυτών. Είναι προφανής, για παρά-

δειγμα, η διαφορά μεταξύ “οικογένειας” και “νοικοκυριού” (βλ. παρακάτω).

Παρατήρηση. Η δειγματοληπτική μονάδα (όπως για παράδειγμα ένα νοικοκυριό), μπορεί να μην συμπίπτει με το απλό στοιχείο που στην προκειμένη περίπτωση μπορεί να είναι όλα τα άτομα του νοικοκυριού (με ή χωρίς κάποιους περιορισμούς, όπως εργαζόμενοι, άνω των 18 ετών, κ.λπ.). Εάν κάθε δειγματοληπτική μονάδα περιέχει ένα και μόνο ένα στοιχείο από τον πληθυσμό, τότε μία δειγματοληπτική μονάδα και ένα στοιχείο του πληθυσμού είναι τα ίδια (μονοπρόσωπο νοικοκυριό).

Οι πληθυσμοί διακρίνονται σε *άπειρους* ή *πεπερασμένους*. Τα στοιχεία του πληθυσμού μπορεί να είναι ανθρώπινες οντότητες, αλλά μπορεί επίσης να είναι και οποιοδήποτε άλλο σύνολο αντικειμένων ή τιμών που προέκυψε από πειράματα ή παρατηρήσεις. Το πείραμα ή οι παρατηρήσεις μπορούν να επαναληφθούν, θεωρητικά τουλάχιστο, άπειρες φορές. Το σύνολο των τιμών που προκύπτει μετά από άπειρες εκτελέσεις ενός πειράματος ή γενικότερα το μη πεπερασμένο σύνολο στοιχείων για τα οποία ενδιαφερόμαστε να εξετάσουμε μία ή περισσότερες ιδιότητες ονομάζεται *άπειρος πληθυσμός*. Διαφορετικά, όταν δηλαδή το πλήθος των απλών στοιχείων που αποτελούν τον πληθυσμό είναι πεπερασμένο, τότε ο πληθυσμός αυτός καλείται *πεπερασμένος*.

Οι πληθυσμοί διακρίνονται επίσης σε *ομογενείς* ή *ετερογενείς*, ανάλογα της διαφοροποίησης των τιμών που λαμβάνει κάθε ένα στοιχείο από αυτά. Η διάκριση αυτή του πληθυσμού είναι πολύ χρήσιμη σε ορισμένες τεχνικές δειγματοληψίας όπως θα δούμε παρακάτω, π.χ. στη στρωματοποιημένη και στην κατά συστάδες δειγματοληψία. Τέλος, ανάλογα της σταθερότητας ή της μεταβλητότητας των τιμών των στοιχείων σε συνάρτηση με το χρόνο, οι πληθυσμοί διακρίνονται σε *στατικούς* ή *δυναμικούς*.

Τα χαρακτηριστικά ως προς τα οποία εξετάζουμε ένα πληθυσμό λέγονται *μεταβλητές* (variables) και συμβολίζονται με τα γράμματα y, x, z κ.λπ. Οι δυνατές τιμές που μπορεί να πάρει μία μεταβλητή λέγονται τιμές της μεταβλητής. Για παράδειγμα, οι δυνατές τιμές της μεταβλητής “φύλο” είναι αγόρι ή κορίτσι, της μεταβλητής “αριθμός παιδιών” σε μία οικογένεια είναι $0, 1, 2, \dots$. Τις μεταβλητές τις διακρίνουμε σε ποσοτικές και σε ποιοτικές ή κατηγορικές.

Οι *ποσοτικές* μεταβλητές διακρίνονται ανάλογα με το πλήθος των τιμών που μπορούν να πάρουν σε *διακριτές* και σε *συνεχείς*. *Διακριτή* θεωρείται η μεταβλητή που λαμβάνει ένα σύνολο τιμών στις οποίες μπορούν να αντιστοιχισθούν, ένα προς ένα, στοιχείο του συνόλου των φυσικών αριθμών. Αντίθετα, μια μεταβλητή χαρακτηρίζεται σαν *συνεχής* εάν παίρνει, θεωρητικά τουλάχιστον, κάθε τιμή στο διάστημα (α, β) , όπου $-\infty < \alpha < \beta < +\infty$.

Τυπικά παραδείγματα διακριτών μεταβλητών είναι π.χ. ο αριθμός των ατόμων των νοικοκυριών, ο αριθμός των ασθενών που καταφθάνουν στη μονάδα

του χρόνου σ' ένα νοσοκομείο που εφημερεύει, αριθμός μαθημάτων που οφείλει ένας φοιτητής πέραν του 8ου εξαμήνου, οι "πόντοι" που παίρνει ένας παίχτης σε μία παρτίδα σκάκι (όπως 1 για νίκη, 0,5 για ισοπαλία και 0 για ήττα). Αντίθετα, η ηλικία, το βάρος, ο χρόνος που χρειάζονται οι φοιτητές να απαντήσουν στα θέματα δειγματοληψίας στις εξετάσεις, αποτελούν τυπικά παραδείγματα συνεχών μεταβολών.

Εκτός από τις ποσοτικές μεταβλητές που αναφέραμε πιο πάνω, πολλές φορές έχουμε μεταβλητές για τις οποίες δεν είναι δυνατό να ληφθούν πλήρως καθοριστικές τιμές, αλλά μόνο ενδεικτικές ή συγκριτικές. Αυτές είναι οι λεγόμενες **ποιοτικές** μεταβλητές. Τέτοιες μεταβλητές είναι π.χ. η οικονομική κατάσταση και η υγεία των ανθρώπων που μπορεί να χαρακτηριστεί σαν κακή, μέτρια ή καλή, καθώς επίσης και το ενδιαφέρον των καθηγητών μέσης εκπαίδευσης για επιμορφωτικά σεμινάρια, που μπορεί να χαρακτηριστεί σαν υψηλό, μέτριο, χαμηλό ή μηδαμινό.

Δειγματοληπτικό πλαίσιο. Στην απλούστερή του μορφή το δειγματοληπτικό πλαίσιο (frame) είναι μία **λίστα** από στοιχεία (elements) που καλύπτουν τον ερευνώμενο πληθυσμό (πληθυσμό στόχο). Μπορεί να αποτελείται από μία **φυσική λίστα**, όπως ένας κατάλογος, μία εξαγωγή (output) από υπολογιστή, μία μαγνητική ταινία, ένας τηλεφωνικός κατάλογος ή ένα σύνολο καρτών. Στις περιπτώσεις αυτές έχουμε το λεγόμενο πλαίσιο-λίστα (list frame). Μπορεί επίσης να είναι ένα οικοδομικό σχέδιο πόλεως ή αεροφωτογραφία μίας περιοχής ή γενικότερα μία γεωγραφική περιοχή, οπότε έχουμε το λεγόμενο area frame. Τέλος, μπορεί να αποτελείται από μία **εννοιολογική** λίστα όπως, για παράδειγμα, όλα τα αυτοκίνητα (γενικότερα, τροχοφόρα) που πάρκαραν σε κάποιο μέρος μεταξύ 8 π.μ. και 8 μ.μ. κατά το μήνα Οκτώβριο.

Ο προσδιορισμός του πλαισίου παίζει σημαντικό ρόλο στο σχεδιασμό μιας δειγματοληπτικής έρευνας. Καθορίζει πόσο καλά καλύπτεται ο πληθυσμός, επηρεάζει τη μέθοδο συλλογής της πληροφορίας και επιδρά στην αποτελεσματικότητα με την οποία λαμβάνεται ένα δείγμα. Συνεπώς το δειγματοληπτικό πλαίσιο αποτελεί βασική προϋπόθεση για την επιτυχία οποιασδήποτε δειγματοληπτικής έρευνας. Πρέπει να περιέχει όλο το δειγματοληπτούμενο πληθυσμό έτσι ώστε κάθε δειγματοληπτική μονάδα να έχει κάποια ευκαιρία επιλογής κατά τη δειγματοληψία. Σε αντίθετη περίπτωση, τα συμπεράσματα της έρευνας δεν μπορούν αξιόπιστα να γενικευθούν επαγωγικά στο σύνολο του ερευνώμενου πληθυσμού.

Για παράδειγμα, ας υποθέσουμε ότι μας ενδιαφέρει η γνώμη των εν ενεργεία ιατρών όλης της χώρας για το ΕΣΥ. Το πλαίσιο-λίστα που θα μπορούσε ίσως να χρησιμοποιηθεί για τη συγκεκριμένη έρευνα είναι ο κατάλογος των μελών του Πανελληνίου Ιατρικού Συλλόγου. Διαπιστώνεται τότε, ότι τα προβλήματα που μπορεί να προκύψουν είναι:

- Ο κατάλογος δεν θα περιέχει όλους τους γιατρούς, κυρίως τους νέους που δεν έχουν ακόμη εγγραφεί.
- Μερικοί από τους ήδη εγγεγραμμένους δυνατόν να έχουν εγκαταλείψει το επάγγελμα, να έχουν συνταξιοδοτηθεί, να έχουν πεθάνει κ.λπ.
- Σε περίπτωση πολλαπλών καταλόγων μπορεί να υπάρξει επικάλυψη.

Σε μία τέτοια περίπτωση μπορεί ένας ερευνητής να αποφασίσει:

- την απόρριψη της συγκεκριμένης λίστας και τη δημιουργία νέας εξ αρχής
- να τη χρησιμοποιήσει όπως έχει, αδιαφορώντας για τις επιπτώσεις που θα προκύψουν
- την προσαρμογή της υπάρχουσας λίστας βάσει των νεότερων δεδομένων (updating), αν και αυτό είναι συνήθως αδύνατο να γίνει.

Παραδείγματα δειγματοληπτικών πλαισίων είναι το αρχείο των οικοδομικών τετραγώνων της ερευνώμενης περιοχής όπως δίνονται από την ΕΣΥΕ, ο τηλεφωνικός κατάλογος του ΟΤΕ, τα γυμνάσια-λύκεια όπως δίνονται από το Υπουργείο Παιδείας, ο κατάλογος των νοσοκομείων ή/και κλινικών όπως δίνονται από το Υπουργείο Υγείας, κ.λπ. Οι κατάλογοι αυτοί που χρησιμοποιούνται για τις δειγματοληπτικές έρευνες πρέπει να ενημερώνονται πριν από κάθε έρευνα που διεξάγεται.

Δείγμα είναι μια συλλογή δειγματοληπτικών μονάδων από το πλαίσιο. Τόσο το μέγεθος του δείγματος όσο και ο τρόπος επιλογής του αποτελούν βασική προϋπόθεση για την επιτυχία μιας δειγματοληπτικής έρευνας, όπως θα δούμε και παρακάτω στην Παράγραφο 1.4.

Ο αντικειμενικός σκοπός της έρευνας είναι η **εκτίμηση** ορισμένων παραμέτρων του πληθυσμού από την πληροφορία που περιέχεται στο δείγμα. Το ενδιαφέρον συγκεντρώνεται κυρίως στην εξέταση των παρακάτω παραμέτρων.

- α) **Πληθυσμιακό ολικό:** Όπως, για παράδειγμα, το συνολικό εισόδημα των κατοίκων μιας πόλης, η συνολική καλλιεργήσιμη έκταση μιας περιοχής, ο συνολικός αριθμός επισκεπτών ενός αρχαιολογικού χώρου σε δεδομένη χρονική διάρκεια.
- β) **Πληθυσμιακός μέσος:** Το μέσο βάρος ή η μέση επίδοση των μαθητών ενός σχολείου, τα μέσα (μηνιαία) έξοδα (κατά κεφαλή) των μελών των νοικοκυριών μιας περιοχής που ξοδεύονται για διασκέδαση, ο μέσος ημερήσιος αριθμός επισκεπτών ενός αρχαιολογικού χώρου σε μία δεδομένη χρονική διάρκεια (π.χ. Ιούνιος - Σεπτέμβριος).

γ) **Πληθυσμιακό ποσοστό:** Το ποσοστό (αναλογία) επισκεπτών ενός αρχαιολογικού χώρου που είναι κάτω των 30 ετών (ή που είναι Έλληνες), το ποσοστό των φοιτητών που καπνίζουν, το ποσοστό των κατοίκων μιας περιοχής που είναι υπέρ κάποιας υπουργικής απόφασης (υπέρ κάποιου νομοσχεδίου), το ποσοστό των ασθενών που επισκέφθηκαν ένα νοσοκομείο κατά τη διάρκεια του προηγούμενου έτους για κάποια πάθηση και δήλωσαν κάποιο κληρονομικό παρελθόν, ποσοστό των παιδιών ηλικίας 10-12 ετών που πάσχουν από τερηδόνα, κ.λπ.

δ) **Λόγος** δύο χαρακτηριστικών, όπως η αναλογία διδασκόντων προς διδασκόμενους, συνολικός χρόνος αφιέρωσης για τις δραστηριότητες κάποιου προγράμματος προς το συνολικό διαθέσιμο χρόνο σε διάφορες δραστηριότητες, συνολικός αριθμός πραγματοποιηθέντων ημερομισθίων σε μία βιομηχανία προς το συνολικό αριθμό των υπαλλήλων που εργάζονται στη βιομηχανία αυτή, κ.λπ.

Η γνώση των διαφόρων ειδών τεχνικής της έρευνας δεν είναι αρκετή από μόνη της. Αυτή η γνώση πρέπει να σχετίζεται με τη γνώση των βασικών κοινωνικών επιστημών, τη γνώση των στατιστικών διαδικασιών και τη γνώση της περιοχής και του συγκεκριμένου αντικειμένου της έρευνας. Γι' αυτό απαραίτητη είναι πολλές φορές η συνεργασία πολλών ειδικών έτσι ώστε ο καθένας να συμβάλει περισσότερο στον τομέα ευθύνης του.

Στην έρευνα, τα πρώτα στάδια - τα στάδια προγραμματισμού - είναι τα δυσκολότερα και τα πιο αποφασιστικά. Τα διάφορα στάδια μιας έρευνας είναι:

- 1) Ο προσδιορισμός και ο καθορισμός του προβλήματος.
- 2) Η μελέτη σχετικών προηγούμενων εργασιών.
- 3) Η απόφαση για τις κατάλληλες τεχνικές.
- 4) Η εκλογή και ο καθορισμός του δείγματος.
- 5) Η συλλογή δεδομένων.
- 6) Η επεξεργασία, η ανάλυση και η επεξήγηση των αποτελεσμάτων.
- 7) Η συγγραφή της έκθεσης.

Δια μέσου των τεσσάρων πρώτων σταδίων η στρατηγική είναι να σχεδιαστεί μία κατάσταση, η οποία θα παραγάγει σχετική μαρτυρία, που να επαληθεύει ή όχι μίαν υπόθεση ή να δώσει απαντήσεις σε μία ειδική ερώτηση. Αυτά τα 4 στάδια χρειάζονται τουλάχιστον το 1/3 του διαθέσιμου χρόνου για όλη την έρευνα.

Στα επόμενα δύο στάδια (5 & 6) αφιερώνεται το 1/3 επίσης του διαθέσιμου χρόνου και το τελικό στάδιο (7) θα χρειαστεί τον υπόλοιπο χρόνο.

Παρατηρούμε λοιπόν ότι η πραγματοποίηση μιας δειγματοληπτικής έρευνας περιλαμβάνει αρκετά ξεχωριστά στάδια, τα οποία είναι εξίσου σημαντικά έτσι ώστε η έρευνά μας να είναι αποτελεσματική. Συνήθως στις περισσότερες μελέτες δίνεται έμφαση στη στατιστική ανάλυση και παρουσίαση των αποτελεσμάτων. Εξίσου σημαντικό όμως ρόλο έχουν και ο καθορισμός του προβλήματος, ο καθορισμός των στοιχείων που επιθυμούμε να συλλέξουμε, η επιλογή της κατάλληλης μεθόδου συλλογής των στοιχείων και ο καθορισμός της τεχνικής και του δείγματος, η επιλογή της κατάλληλης μεθόδου συλλογής των στοιχείων και, τέλος, η σωστή οργάνωση και εμφάνιση του ερωτηματολογίου.

1.3 Καθορισμός του πληθυσμού

Σε δειγματοληπτικές έρευνες, είναι αρκετές οι περιπτώσεις που εμφανίζεται ο πληθυσμός σαν το επιλεγόμενο δείγμα. Για παράδειγμα, μας ενδιαφέρει να δούμε πως διασκεδάζουν οι νέοι, και για το λόγο αυτό επιλέγουμε ορισμένους μαθητές από τα γυμνάσια και λύκεια της Αττικής. Με τον τρόπο αυτό ταυτίζουμε τον πληθυσμό των “νέων” (πρέπει να καθοριστεί από που μέχρι που) με τους μαθητές των γυμνασίων και λυκείων. Τι γίνεται με τους υπόλοιπους “νέους” που δεν είναι μαθητές;

Το πρώτο βήμα για τον καθορισμό του πληθυσμού είναι να αποφασιστεί αν αυτός είναι ένας πληθυσμός ατόμων, οικογενειών (νοικοκυριών), επιχειρήσεων ή οτιδήποτε άλλο.

Η πηγή από την οποία θα συλλεγούν τα δεδομένα δεν είναι απαραίτητο να είναι ταυτοτική με τον ορισμό του πληθυσμού. Έστω για παράδειγμα ότι μας ενδιαφέρει να εκτιμήσουμε τον αριθμό των καθηγητών των λυκείων μιας περιοχής που απεργούν σε μία συγκεκριμένη μέρα. Τότε ο πληθυσμός είναι αυτός των καθηγητών των λυκείων. Το ερωτηματολόγιο όμως (για τη συλλογή της πληροφορίας) μπορεί να απευθύνεται στους λυκειάρχες των λυκείων της περιοχής που εξετάζουμε. Απλούστερα, ένα μέλος οικογένειας μπορεί να αναφερθεί στα οικογενειακά έσοδα ή τον αριθμό των μελών των οικογενειών. Όταν όμως η μελέτη αναφέρεται π.χ. σε στάση των ατόμων έναντι, ας πούμε, κάποιου κοινωνικού φαινομένου ή για πολιτική τοποθέτηση, τότε οι πληροφορίες πρέπει να παρθούν από κάθε άτομο χωριστά.

Διαφαίνεται έτσι η αναγκαιότητα του ξεκάθਾਰου προσδιορισμού του πληθυσμού. Τις περισσότερες φορές μεταξύ της εκλογής των ατόμων και οικογενειών σαν ολότητας επικρατεί μία σύγχυση. Θα πρέπει οι μελέτες συμπεριφοράς καταναλωτών, χρήσης μέντιουμ, δραστηριοτήτων ελεύθερων ωρών, να χρησιμο-

ποιούν πληθυσμούς ατόμων ή οικογενειών (νοικοκυριών);

Η απόφαση είναι δύσκολη και θα εξαρτηθεί από τους ειδικούς σκοπούς της έρευνας. Πρέπει όμως, ότι και να χρησιμοποιηθεί, να ξεκαθαριστεί κατά το σχεδιασμό της μελέτης. Από το είδος του πληθυσμού ο ερευνητής θα μπορέσει να σχεδιάσει καλύτερα τις ερωτήσεις, να αποφασίσει για την έκταση του ερωτηματολογίου, το αναγνωστικό επίπεδο των ερωτήσεων, το είδος των οδηγιών που θα δοθούν. Στη συνέχεια, πρέπει να ληφθούν διάφορες επί πλέον αποφάσεις σχετικά με τις μονάδες που θα συμπεριληφθούν στον πληθυσμό. Για το λόγο αυτό έχουμε τα παρακάτω κριτήρια:

α) Γεωγραφικά: Μια έρευνα μπορεί να αναφέρεται σε μία συγκεκριμένη περιοχή (πόλη, νομό) ή σ' ολόκληρη την επικράτεια. Όταν, για παράδειγμα, μας ενδιαφέρει να εξετάσουμε τη στάση των κατοίκων μιας περιοχής για τις επιπτώσεις που θα έχει η δημιουργία ενός νέου εργοστασίου (π.χ. λιπασμάτων) στην περιοχή, θα πρέπει να ορίσουμε πλήρως τα όρια της εν λόγω περιοχής. Όταν μας ενδιαφέρει να εξετάσουμε τα προβλήματα θορύβου που δημιουργούνται στους κατοίκους "γύρω" από το αεροδρόμιο, πρέπει να ξεκαθαριστεί μέχρι που φτάνει το "γύρω" από το αεροδρόμιο.

β) Ηλικία: Ανάλογα με τη σχεδιαζόμενη μελέτη πρέπει να καθορίζεται και η ηλικία των ατόμων του πληθυσμού. Μπορεί να έχουμε μόνο κατώτερο όριο ηλικίας, όπως π.χ. για τους ψηφοφόρους είναι τα 18. Για εργασιακές μελέτες είναι συνήθως τα 16. Μία μελέτη μπορεί να ενδιαφέρεται για την τερηδόνα παιδιών 10-12 ετών ή 13-15 ετών κ.λπ. Άνω όριο ηλικίας μπορεί να τεθεί στην περίπτωση για παράδειγμα που η μελέτη έχει να κάνει με γυναίκες με ηλικία ικανή για τεκνοποίηση.

γ) Άλλες δημογραφικές μεταβλητές: Το φύλο, η οικογενειακή κατάσταση και η εκπαίδευση είναι άλλες μεταβλητές που μερικές φορές χρησιμοποιούνται για να ορίσουν ένα πληθυσμό. Ενώ ως προς το φύλο δεν υπάρχει γενικά πρόβλημα καθορισμού, οι άλλες μεταβλητές πρέπει να ορίζονται με προσοχή.

δ) Ατομικές μεταβλητές: Διάφορες ατομικές μεταβλητές, όπως η εθνικότητα, υπηκοότητα, το εκλογικό δικαίωμα κ.λπ. είναι συνήθως σημαντικές μεταβλητές για τις μελέτες εκλογικής συμπεριφοράς, ενώ αντίθετα για μελέτες της κοινής γνώμης δεν λαμβάνονται κατά κανόνα και τόσο πολύ υπόψη.

ε) Οικογενειακές μεταβλητές: Στις περιπτώσεις που βασικός σκοπός της έρευνας είναι η συγκέντρωση διαφόρων στοιχείων για τα άτομα των οικογενειών μιας περιοχής, πρέπει πρώτα να οριστεί η έννοια της οικογένειας ή συνήθως η έννοια του νοικοκυριού. Όπως χρησιμοποιείται στις περισσότερες μελέτες σαν νοικοκυριό ορίζεται (ΕΣΥΕ):

i) Δύο ή περισσότερα άτομα, ανεξάρτητα αν υπάρχει ή όχι συγγένεια μεταξύ τους, τα οποία κατοικούν μαζί σε μία κατοικία, προμηθεύονται συνήθως από κοινού τα απαραίτητα για τη συντήρησή τους και τρώνε κατά κανόνα μαζί (**πολυπρόσωπο νοικοκυριό**).

ii) Ένα άτομο που κατοικεί μόνο του ή κατοικεί μαζί με άλλα άτομα αλλά δεν προμηθεύεται από κοινού με τα άλλα άτομα τα απαραίτητα για τη συντήρησή του, ούτε τρώει μαζί τους (**μονοπρόσωπο νοικοκυριό**).

Ένα δεύτερο στοιχείο που πρέπει να ληφθεί υπόψη κατά το σχεδιασμό μιας έρευνας είναι κατά πόσο θα θεωρηθούν **σαν μέλη του νοικοκυριού** τα άτομα εκείνα, που κατά τη διάρκεια της έρευνας, απουσιάζουν προσωρινά από την κατοικία τους. Κατά την ΕΣΥΕ, **μέλη του νοικοκυριού θεωρούνται** και τα άτομα εκείνα που απουσιάζουν για λόγους τουρισμού, επίσκεψης συγγενών και φίλων, εργασίας (σιδηροδρομικοί, οδηγοί και εισπράκτορες υπηρεσιακών λεωφορείων, πλανόδιοι έμποροι, ναυτικοί και αλιείς που βρίσκονται μέσα ή έξω από τα χωρικά ύδατα της Χώρας για οσοδήποτε χρονικό διάστημα, κράτησης σε κρατητήρια ή εισαγωγής σε κλινικές και νοσοκομεία (εκτός από σανατόρια, ψυχιατρεία και άσυλα ανιάτων).

Αντίθετα, **δεν θεωρούνται μέλη του νοικοκυριού** τα άτομα που διαμένουν σε άλλη πόλη για σπουδές ή εργασία, οι φιλοξενούμενοι για χρονικό διάστημα λιγότερο του εξαμήνου, οι παραδουλεύτρες, όσοι υπηρετούν τη θητεία τους στις Ένοπλες Δυνάμεις, όσοι βρίσκονται στο εξωτερικό για σπουδές ή εργασία, οι φυλακισμένοι (ανεξάρτητα από το χρόνο φυλάκισης).

Ο ορισμός της κατοικίας πρέπει επίσης να διευκρινιστεί. **Κατοικία** θεωρείται ένας χώρος που κτίστηκε ή μετατράπηκε προς το σκοπό να χρησιμοποιηθεί για κατοίκηση. Έτσι, κινητά σπίτια, τρέιλερς, σκηνές ή βάρκες θεωρούνται ότι αποτελούν κατοικία (νοικοκυριών) αν πράγματι χρησιμοποιούνται γι' αυτό το σκοπό κατά το χρόνο της έρευνας. Αντίθετα, εξαιρούνται αν χρησιμοποιούνται μόνο για επιχειρήσεις ή για διακοπές.

στ) Διάφορα χαρακτηριστικά: Ανάλογα με τον αντικειμενικό σκοπό κάθε δειγματοληπτικής έρευνας κρίνεται απαραίτητη η ακριβής περιγραφή εκείνων των χαρακτηριστικών που η έρευνα προτίθεται να εξετάσει.

Έστω ότι θέλουμε να εξετάσουμε κάποιο χαρακτηριστικό για τους μαθητές που φοιτούν σε λύκειο, όπως π.χ. τα κριτήριά τους για την εκλογή των 5-μελών και 15-μελών συμβουλίων τους, πόσοι από αυτούς (ποσοστό) δεν ενδιαφέρονται για πανεπιστημιακές σπουδές αλλά να βρουν κάποια δουλειά. Είναι προφανές εδώ ότι πρέπει να επιμείνουμε περισσότερο στον καθορισμό του συγκεκριμένου πληθυσμού των μαθητών λυκείων. Πρέπει να ξεκαθαρίσουμε αν εννοούμε δημόσια ή/και ιδιωτικά σχολεία. Επίσης σχολεία για παιδιά με ειδικές ανάγκες, πειραματικά ή πρότυπα, ειδικά λύκεια όπως π.χ. μουσικά, διάφορες σχολές δη-

μόσιες ή ιδιωτικές που μπορούν να θεωρηθούν σαν “ισοδύναμες” με τα λύκεια, Τεχνικά Επαγγελματικά Λύκεια (ΤΕΕ), κ.λπ.

Ένα δεύτερο παράδειγμα, όχι τόσο απλό όπως το προηγούμενο, είναι ο σχεδιασμός μιας δειγματοληπτικής έρευνας με αντικειμενικό σκοπό τη συγκέντρωση των απαραίτητων στοιχείων για τη μελέτη της διαρθρώσεως της **απασχόλησης** των μελών των νοικοκυριών όλης της Χώρας. Για το παράδειγμα αυτό θα πρέπει να οριστεί πρώτα η ηλικία των μελών των νοικοκυριών για τα οποία θα συγκεντρωθούν στοιχεία σχετικά με την απασχόληση. Συνήθως, τα στοιχεία της απασχόλησης αναφέρονται σε μέλη νοικοκυριών **ηλικίας 14 ετών και άνω**. Στη συνέχεια, πρέπει να οριστεί ποιοί θεωρούνται σαν **απασχολούμενοι** και ποιοί θεωρούνται σαν **άνεργοι**.

Σύμφωνα με ανάλογες δειγματοληπτικές έρευνες απασχόλησης που διενεργούνται από την ΕΣΥΕ και καλύπτουν τα μέλη των νοικοκυριών όλων των περιοχών (αστικών, ημιαστικών και αγροτικών) της Χώρας, έχουμε τους παρακάτω ορισμούς.

A) Σαν απασχολούμενοι θεωρούνται όλα τα άτομα άνω των 14 ετών οι οποίοι:

- α) Κατά το “χρονικό διάστημα αναφοράς” (συνήθως σαν “περίοδος αναφοράς” αναφέρεται η τελευταία προ της ημέρας της έρευνας ημερολογιακή εβδομάδα, οπότε έχουμε την “εβδομάδα αναφοράς”) έχουν ή είχαν **κανονική εργασία**, δηλαδή αμειβόμενη εργασία η οποία είναι ή προβλέπεται να είναι διάρκειας τουλάχιστο τεσσάρων εβδομάδων και εβδομαδιαίας απασχόλησης τουλάχιστο 14 ωρών.
- β) Απουσιάζουν κατά την εβδομάδα αναφοράς από την “κανονική τους εργασία” λόγω ασθένειας, άδειας αναψυχής, ταξιδιού, καφικιών συνθηκών, επισκευής μηχανημάτων ή καταστήματος, εποχικότητας, κ.λπ.
- γ) Εργάστηκαν κατά την “εβδομάδα αναφοράς” **τουλάχιστο μία ώρα**, έστω και αν δεν έχουν “κανονική εργασία”. Εξαιρέση αποτελούν όσοι βοηθούν σε οικογενειακή επιχείρηση χωρίς αμοιβή και έχουν εργαστεί κατά την “εβδομάδα αναφοράς” λιγότερες από 14 ώρες.

Σαν εργάσιμος χρόνος θεωρείται ο χρόνος (σε ώρες) που πράγματι έχει διατεθεί σε οποιαδήποτε μορφή εργασίας. Υπολογίζονται δηλαδή οι ώρες εργασίας που πραγματοποιήθηκαν στα πλαίσια του κανονικού ωραρίου, οι υπερωρίες με ή χωρίς αμοιβή και ο χρόνος μικρών διαλειμάτων για ξεκούραση στο χώρο εργασίας.

B) Σαν άνεργοι θεωρούνται τα άτομα ηλικίας 14 ετών και άνω που δεν εργάστηκαν καθόλου κατά το χρόνο αναφοράς, ούτε είχαν κανονική εργασία (και

απουσίαζαν για τους λόγους που αναφέρθηκαν πιο πάνω) και τα οποία **ζητούσαν** εργασία σαν μισθωτοί (με μισθό ή ημερομίσθιο).

Τα άτομα που χαρακτηρίζονται “σαν απασχολούμενοι” ή “σαν άνεργοι” σύμφωνα με τους παραπάνω ορισμούς, αποτελούν το **εργατικό δυναμικό** της Χώρας.

Οι άνεργοι διακρίνονται σε δύο κατηγορίες:

- i) Σ’ αυτούς που δεν εργάστηκαν ποτέ στο παρελθόν και αποτελούν τους “**νέο υς**” άνεργους, και
- ii) σ’ αυτούς που εργάστηκαν κατά το παρελθόν σε μία κανονική ή όχι εργασία.

Παράδειγμα 1.1 Ας υποθέσουμε ότι ενδιαφερόμαστε να διεξάγουμε μια έρευνα σε μία πόλη για να μελετήσουμε στοιχεία σχετικά με την υγειονομική ασφάλιση των κατοίκων.

Για τον λόγο αυτό πρέπει πρώτα να γίνει μια περιγραφή της μελέτης που θέλουμε να πραγματοποιήσουμε, να θέσουμε κάποια πιθανά ερωτήματα, ποια θα είναι τα δεδομένα που είναι απαραίτητα για να απαντηθούν τα ερωτήματα που έχουμε θέσει, ποια είναι η γνώση που θα μας προσφέρουν αυτά τα δεδομένα.

Καθορισμός στοιχείων που επιθυμούμε να συλλέξουμε:

- α) Εκτίμηση ποσοστού ατόμων που έχουν ασφάλεια υγείας.
 - Τα άτομα έχουν ή όχι υγειονομική ασφάλιση.
 - Η ασφάλιση, αν υπάρχει, είναι κρατική ή/και ιδιωτική.
 - Ποια μέλη της οικογένειας καλύπτονται ή δεν καλύπτονται από την ασφάλιση.
 - Αν υπάρχει ιδιωτική ασφάλιση αυτή είναι οικογενειακή ή ατομική.
- β) Σε τι διαφέρει το παραπάνω ποσοστό με βάση συγκεκριμένα πληθυσμιακά χαρακτηριστικά, π.χ.
 - φύλο
 - ηλικία
 - φυλή της “κεφαλής” της οικογένειας
 - εισόδημα
 - αριθμός ατόμων οικογένειας (δομή οικογένειας)
- γ) Έξοδα για υγειονομική ασφάλιση.

- Ετήσιο ποσό ασφάλισης για οικογενειακή ή ατομική ασφάλεια.
- δ) Νοσηλεία των ατόμων κατά τη διάρκεια του χρόνου αναφοράς (έτος, εξάμηνο, μήνας κ.λπ.)
- Διάρκεια νοσηλείας
 - Είδος νοσηλείας
 - Έξοδα νοσηλείας.

Συμπερασματικά, τα απαραίτητα στοιχεία για τη συγκεκριμένη μελέτη είναι:

- Ηλικία και φύλο όλων των ατόμων σε κάθε οικογένεια
- “Κεφαλή” της οικογένειας και σχέση του κάθε μέλους με την “κεφαλή” της οικογένειας
- Φυλή της “κεφαλής” της οικογένειας
- Εισόδημα κάθε ατόμου της οικογένειας
- Υγειονομική ασφάλιση ή όχι και αν αυτή είναι οικογενειακή ή όχι, δημόσια ή ιδιωτική
- Έξοδα για υγειονομική κάλυψη
- Διάρκεια, είδος και έξοδα νοσηλείας στο χρόνο αναφοράς.

1.4 Εκλογή δείγματος

Όταν η εκλογή ενός δείγματος από τον πληθυσμό γίνει έτσι ώστε να έχει ακριβώς τα ίδια χαρακτηριστικά με αυτόν, τότε τα αποτελέσματα που θα ληφθούν από το δείγμα θα είναι, εντός ορισμένων ορίων, τα ίδια με τα αποτελέσματα που θα λαμβάνονταν από ολόκληρο τον πληθυσμό. Αυτά τα “ορισμένα όρια” καθορίζονται από κριτήρια στατιστικής σημαντικότητας. Η πληροφορία που περιέχεται στο εκλεγόμενο δείγμα και θα χρησιμοποιηθεί στην εκτίμηση των παραμέτρων του πληθυσμού βρίσκεται υπό τον έλεγχο του ερευνητή. Εξαρτάται από τον αριθμό των δειγματοληπτικών μονάδων που θα συμπεριλάβει το δείγμα και το οποίο καλείται **μέγεθος δείγματος**. Η διαδικασία που χρησιμοποιείται για την επιλογή του δείγματος, όπως και το μέγεθος δείγματος που θα εκλεγεί είναι σημαντικά στοιχεία για να έχουμε **αντιπροσωπευτικό** δείγμα του υπό εξέταση πληθυσμού. Αυξάνοντας το μέγεθος δείγματος ασφαλώς θα έχουμε και καλύτερη ακρίβεια στην εκτίμηση των χαρακτηριστικών του πληθυσμού. Από την άλλη πλευρά όμως το κόστος για τη δειγματοληψία, την επεξεργασία και παρουσίαση των αποτελεσμάτων αυξάνει. Πολύ μεγάλο δείγμα συνεπάγεται

σπατάλη χρήματος και κόπου. Πολύ μικρό δείγμα είναι δυνατό να οδηγήσει σε μεροληπτικές εκτιμήσεις. Συνοπτικά, οι διάφοροι παράγοντες που επηρεάζουν την επιλογή του δείγματος είναι:

- α) Μέγεθος πληθυσμού.
- β) Στατιστικό σφάλμα - επίπεδο ακριβείας.
- γ) Στάθμη εμπιστοσύνης.
- δ) Μεταβλητότητα του διερευνούμενου πληθυσμού.
- ε) Διαθέσιμο κόστος της έρευνας.
- στ) Μέθοδος επιλογής δείγματος.

Για τον προσδιορισμό του **μεγέθους δείγματος** (sample size) ή δίνεται κάποιο **περιθώριο σφάλματος** (d) απόκλισης της εκτιμήτριας από το εκτιμώμενο καθώς επίσης και η **αξιοπιστία ή εμπιστοσύνη** (confidence coefficient) $1 - \alpha$ με την οποία θα γίνει η εκτίμηση.

Εάν θ είναι κάποια παράμετρος του πληθυσμού και $\hat{\theta}$ μία εκτιμήτρια της θ , τότε το περιθώριο σφάλματος d ορίζεται έτσι ώστε

$$P[|\hat{\theta} - \theta| \leq d] \geq 1 - \alpha.$$

Συνήθως σαν **επίπεδο σημαντικότητας** α εκλέγεται $\alpha = 0,05$ ή $\alpha = 0,01$ οπότε έχουμε συντελεστή εμπιστοσύνης 95% ή 99%, αντίστοιχα. Πρέπει όμως να σημειωθεί ότι μία “προσεκτική” επιλογή μικρού δείγματος είναι δυνατό να δώσει πολύ καλύτερα αποτελέσματα από ένα μεγαλύτερο δείγμα που δεν έχει εκλεγεί κατάλληλα.

Ενδεικτικό εδώ είναι το παράδειγμα των προεδρικών εκλογών στις ΗΠΑ το 1936, όπου υπήρξε πλήρης αποτυχία της πρόγνωσης των εκλογικών αποτελεσμάτων. Η δειγματοληψία που διενήργησε η εταιρεία “Literary Digest” στηρίχτηκε στην εκλογή 10.000.000 ατόμων του πληθυσμού από λίστα-πλαίσιο των τηλεφωνικών καταλόγων των διαφόρων πόλεων και από τηλεφωνικούς καταλόγους ιδιοκτητών αυτοκινήτων. Στο ερωτηματολόγιο που ταχυδρομήθηκε απάντησε περίπου το 1/4 των επιλεγέντων (2.400.000 άτομα) και από την ανάλυση που έγινε φάνηκε ότι το εκλογικό σώμα θα έδινε μεγάλη πλειοψηφία στους Ρεπουβλικάνους με υποψήφιο πρόεδρο τον Landon. Η μέρα των εκλογών όμως επιφύλαξε μία δυσάρεστη έκπληξη τόσο για τους Ρεπουβλικάνους όσο και για τους στατιστικούς αναλυτές της δημοσκοπήσης: λιγότερο από 40% των ψηφοφόρων προτίμησαν το Ρεπουβλικανικό κόμμα, και ο Δημοκρατικός υποψήφιος Ρούσβελτ εκλέχτηκε με την ιστορική πλειοψηφία 60,7%, αντίθεση με το 40,9%

που πρόβλεψε η δημοσκόπηση της “Literary Digest”. Το λάθος ασφαλώς της πρόγνωσης (20 περίπου ποσοστιαίες μονάδες) οφείλετο στο γεγονός ότι οι τηλεφωνικοί κατάλογοι, εκείνη την εποχή τουλάχιστο, δεν περιλάμβαναν τις κατώτερες κοινωνικο-οικονομικές τάξεις. Αντίθετα το δημοσκοπικό γραφείο του G. Gallup χρησιμοποιώντας δείγμα 50.000 ατόμων πρόβλεψε το σωστό αποτέλεσμα. Καταλήγουμε λοιπόν στο συμπέρασμα ότι μια προσεκτική επιλογή μικρότερου δείγματος είναι δυνατόν να δώσει καλύτερα αποτελέσματα από ένα μεγαλύτερο δείγμα που δεν έχει εκλεγεί κατάλληλα.

Παρόλο που οι “επιτυχίες” των προεκλογικών δημοσκοπήσεων για προεδρικές, κοινοβουλευτικές ή δημοτικές εκλογές είναι πάμπολες, δεν λείπουν και οι μεγάλες αποτυχίες αυτών. Με τη διεξαγωγή μιας δειγματοληψίας αναζητείται η πρόθεση ψήφου του ερωτώμενου. Αυτός όμως άλλα μπορεί να λείει στην έρευνα και άλλα μπορεί να πράττει μπροστά στην κάλπη. Ένα λάθος των ερευνητών στην μη σωστή εκτίμηση του μεγέθους του δείγματος, τον προσδιορισμό του πληθυσμού, στην ανταπόκριση των ερωτωμένων, στην εκτίμηση του ποσοστού αποχής από τις εκλογές αλλά και αρκετά άλλα λάθη μπορούν εύκολα να οδηγήσουν σε παταγώδη αποτυχία των δημοσκοπήσεων. Εκτός της “κλασσικής” περίπτωσης που αναφέραμε προηγούμενα με τις εκλογές του 1936 στις ΗΠΑ, έχουμε αρκετές και σημαντικές άλλες αποτυχίες, όχι μόνο στη χώρα μας όπως τις ζήσαμε τελευταία, αλλά και σε χώρες που επί σειρά ετών χρησιμοποιούν τις δημοσκοπήσεις ως εργαλείο πολιτικής. Μεταξύ αυτών αναφέρουμε τις δημοσκοπήσεις για:

- Τις προεδρικές εκλογές στις ΗΠΑ το 1948. Τότε ο Χάρι Τρούμαν φαινόταν να χάνει τις εκλογές και τα στελέχη του Δημοκρατικού Κόμματος του ζητούσαν επιμόνως να παραιτηθεί. Τα τρία μεγάλα ινστιτούτα Γκάλοπ, Γκρόσλεϊ και Ρόπερ προανέγγελλαν τη νίκη του Ρεπουμπλικανού υποψήφιου Τόμας Νιούι. Στις εκλογές του Νοεμβρίου 1948 ο Χάρι Τρούμαν εξελέγη με διαφορά 114 εκλεκτορικών ψήφων.
- Στις βουλευτικές εκλογές στη Βρετανία το 1970 το Ινστιτούτο Γκάλοπ είχε υπερεκτιμήσει τα ποσοστά των Εργατικών δίνοντας στον υποψήφιο του Εργατικού Κόμματος Χάρολντ Ουίλσον προβάδισμα 9,4 μονάδων, ο οποίος όμως τελικά έχασε τις εκλογές. Αυτό αποδόθηκε στο μεγάλο ποσοστό αποχής 28,5%, ενώ οι δημοσκοπήσεις το εκτιμούσαν κάτω του 8,5%.
- Στις εκλογές του 1992 στη Βρετανία οι δημοσκοπήσεις έδειχναν ισοδυναμία μεταξύ των Εργατικών και Συντηρητικών τελικά όμως οι Συντηρητικοί κέρδισαν με 8 μονάδες διαφορά.

- Στις κοινοβουλευτικές εκλογές στη Γαλλία το 1997 οι δημοσκοπήσεις προέβλεπαν νίκη του κεντροδεξιού συνασπισμού, τελικά όμως τις εκλογές πήρε η Αριστερά.
- Στις κοινοβουλευτικές εκλογές στη χώρα μας το 2000 ακόμη και τα “exit polls” έδιναν νίκη της Νέας Δημοκρατίας τελικά όμως κέρδισε το ΠΑ.ΣΟ.Κ.

1.5 Σφάλματα δειγματοληψίας

Το κυριότερο μειονέκτημα μίας δειγματοληπτικής μεθόδου έναντι ολικής απογραφής του πληθυσμού είναι ότι οι εκτιμήσεις των χαρακτηριστικών του πληθυσμού από ένα μέρος αυτού έχουν τα λεγόμενα “σφάλματα δειγματοληψίας”. Με κατάλληλο όμως μηχανισμό για την επιλογή των μονάδων του δείγματος είναι δυνατός ο περιορισμός των σφαλμάτων αυτών και, το σπουδαιότερο, η πιθανοθεωρητική εκτίμηση του μεγέθους αυτών.

Για την αξιολόγηση της ακρίβειας μίας δειγματοληπτικής έρευνας και για λόγους ευκολίας τα σφάλματα ταξινομούνται σε δύο κατηγορίες:

1. Δειγματοληπτικά σφάλματα.
2. Μη δειγματοληπτικά σφάλματα.

1.5.1 Δειγματοληπτικά σφάλματα

Τα σφάλματα αυτά μπορεί να προκύψουν από τη μη σωστή επιλογή της κατάλληλης μεθόδου δειγματοληψίας, και την επιλογή ακατάλληλου δείγματος, όσον αφορά τη μεταβλητότητα και το μέγεθος του δείγματος. Συνεπώς τα σφάλματα αυτά μπορούν να οδηγήσουν σε εκτιμήσεις με μεγάλα διαστήματα εμπιστοσύνης.

Μία καλή δειγματοληπτική έρευνα πρέπει να περιλαμβάνει τον υπολογισμό των δειγματοληπτικών σφαλμάτων. Οι πληροφορίες για τα δειγματοληπτικά σφάλματα θα πρέπει να είναι άμεσα διαθέσιμες σε όσους προτίθενται να χρησιμοποιήσουν τα στατιστικά αποτελέσματα. Δηλαδή θα πρέπει να αναφέρεται τόσο η δειγματοληπτική μέθοδος, όσο και το σφάλμα της εκτίμησης μίας παραμέτρου.

1.5.2 Μη δειγματοληπτικά σφάλματα

Ο όρος “μη δειγματοληπτικό σφάλμα” προέρχεται από το γεγονός ότι αυτός ο τύπος σφάλματος μπορεί να συμβεί σε οποιαδήποτε έρευνα, είτε αυτή γίνεται με τη μέθοδο της απογραφής, είτε με τη μέθοδο της δειγματοληψίας.

Δυστυχώς, σε αντίθεση με τα δειγματοληπτικά σφάλματα, δεν υπάρχει μία απλή και άμεση μέθοδος για την εκτίμηση του μεγέθους των μη δειγματοληπτικών σφαλμάτων. Τα μη δειγματοληπτικά σφάλματα μπορούν να ταξινομηθούν σε δύο κατηγορίες:

α) **Μη συστηματικά** (ή σφάλματα τυχαίου τύπου)

Είναι τα σφάλματα των οποίων οι επιδράσεις προσεγγιστικά αλληλοαναιρούνται όταν χρησιμοποιούνται αρκετά μεγάλου μεγέθους δείγματα.

β) **Συστηματικά** (ή μεροληψίες)

Είναι τα σφάλματα τα οποία οδηγούν την τιμή της δειγματικής συνάρτησης πάντοτε προς τα πάνω ή πάντοτε προς τα κάτω από την τιμή της παραμέτρου και γι' αυτό δεν μπορούν να αλληλοεξουδετερωθούν, π.χ. μία ζυγαριά ζυγίζει τα πάντα κατά (έστω) 50gr βαρύτερα.

Κύριες αιτίες των μη δειγματοληπτικών σφαλμάτων είναι:

Λάθος επιλογή δείγματος που οφείλεται στον ελλιπή προσδιορισμό του πληθυσμού και σε ακαταλληλότητα της λίστας.

Σφάλματα συνέντευξης. Αυτά οφείλονται σε

1. Έλλειψη συνεντεύξεων. Οι πληροφορίες αποκτώνται μόνο από ένα μέρος του δείγματος.
2. Σφάλματα συνεντευκτών. Οι συνεντεύκτες μπορεί να διαβάσουν λανθασμένα ερώτηση ή να διαστρέψουν τις απαντήσεις με δικά τους λόγια.

Σφάλματα ερωτηματολογίου. Όπως

1. Χρησιμοποίηση λέξεων που δεν είναι κατανοητές.
2. Πολύπλοκα ερωτήματα που δημιουργούν κούραση στους ερωτώμενους.
3. Λάθος σειρά ερωτήσεων. Αφορά τις ερωτήσεις που αλληλοκαλύπτονται και αυτές που αν τοποθετηθούν ενδιάμεσα προκαλούν άσχημη διάθεση στον απαντητή.
4. Υπερφορτωμένες ερωτήσεις. Για λόγους εξοικονόμησης χρόνου ρωτάμε 2 ή περισσότερες ερωτήσεις σε μία.
5. Αναφορά σε ερωτήσεις που δεν χρησιμεύουν. Είναι αυτές που είτε η απάντησή τους δεν έχει σχέση με την έρευνά μας, είτε η απάντησή τους είναι δεδομένη (κοινωνικά αποδεκτές).

Σφάλματα απαντητή. Αυτά οφείλονται σε

1. Απόκριψη αλήθειας. Από φόβο ή καχυποψία δε λέγεται η αλήθεια.

2. Άγνοια. Οι ερωτώμενοι δεν γνωρίζουν την πληροφορία που τους ζητείται.

3. Άρνηση απάντησης.

Σφάλματα επεξεργασίας. Οφείλονται στην:

1. Κωδικοποίηση.

2. Πληκτρολόγηση.

Σφάλματα από το κόστος της έρευνας

Λόγω μεγάλου κόστους ξεφεύγουμε από την αρχική σχεδίαση της έρευνας (π.χ. ελαττώνουμε το μέγεθος δείγματος).

Το άθροισμα των μη δειγματοληπτικών και δειγματοληπτικών σφαλμάτων είναι το συνολικό σφάλμα μίας δειγματοληπτικής έρευνας. Βασικό λοιπόν πρόβλημα της δειγματοληψίας είναι να ελαχιστοποιήσουμε το συνολικό σφάλμα.

1) Για τα δειγματοληπτικά σφάλματα

Εδώ είναι σαφές ότι όσο μεγαλύτερο είναι το δείγμα, τόσο τα δειγματοληπτικά σφάλματα τείνουν να εξουδετερωθούν.

2) Για τα μη δειγματοληπτικά σφάλματα

Η προηγούμενη αναφορά στις αιτίες των μη δειγματοληπτικών σφαλμάτων αποκαλύπτει ότι θα μπορούσαν να μειωθούν σημαντικά ως εξής:

- Ακριβής καθορισμός του πληθυσμού.
- Εκλογή κατάλληλης λίστας.
- Πλήρης εκπαίδευση των συνεντευκτών.
- Κατασκευή του ερωτηματολογίου, έτσι που να προκαλεί το ενδιαφέρον του απαντητή και να τον βοηθάει να απαντήσει σωστά.
- Προεξέταση του ερωτηματολογίου με δοκιμαστική έρευνα (pilot study).
- Ακριβής καθορισμός των χαρακτηριστικών που πρόκειται να εκτιμηθούν.
- Ακριβής εκτίμηση του κόστους της έρευνας.

Πρέπει να σημειώσουμε ότι η μείωση των μη δειγματοληπτικών σφαλμάτων απαιτεί μικρά δείγματα, γιατί τα περισσότερα από τα παραπάνω προληπτικά μέτρα που αναφέραμε είναι δαπανηρά. Για παράδειγμα η προσπάθεια για εκπαίδευση προσοντούχων συνεντευκτών είναι συνήθως χρονοβόρα και δαπανηρή.

Οπωσδήποτε όμως ένα μικρό δείγμα τείνει να έχει ένα μεγαλύτερο δειγματοληπτικό σφάλμα. Συνεπώς εκείνοι που είναι υπεύθυνοι για τις δειγματοληπτικές έρευνες πρέπει να πάρουν σημαντικές αποφάσεις, όσον αφορά την αξιοποίηση των περιορισμένων ανθρωπίνων και οικονομικών δυνατοτήτων κατά τέτοιο τρόπο, ώστε να προσετοιμάσουν την εξασφάλιση της υψηλότερης δυνατής αξιοπιστίας των δειγματοληπτικών αποτελεσμάτων.

1.6 Σχεδιασμός ερωτηματολογίου

Το ερωτηματολόγιο είναι ένα βασικό εργαλείο στα προβλήματα έρευνας. Ο καλός σχεδιασμός των ερωτηματολογίων αποτελεί τη σωστή βάση για μία πετυχημένη έρευνα. Δεν πρέπει να ξεχνάμε ότι οι δειγματοληπτικές έρευνες σε ανθρώπινους πληθυσμούς (δημοσκοπήσεις) **δείχνουν αυτό που δηλώνουν οι ερωτώμενοι και όχι αυτό που πράγματι κάνουν**. Αν ένα ερωτηματολόγιο δεν είναι καλά σχεδιασμένο έτσι ώστε να αντλεί τη σωστή πληροφορία, σίγουρα ο αντικειμενικός σκοπός της έρευνας θα έχει αποτύχει. Στην περίπτωση που το ερωτηματολόγιο είναι ακατάλληλο, οι πληροφορίες θα είναι ατελείς, άσχετες ή διφορούμενες και καμιά εκ των υστέρων ανάλυση ή ερμηνεία δεν μπορεί να διορθώσει τα πράγματα ώστε να βγουν τα σωστά αποτελέσματα.

Πρέπει να τονιστεί ότι το ερωτηματολόγιο είναι απλώς ένα εργαλείο, και όχι πάντα το καλύτερο. Υπάρχουν περιπτώσεις που οι πληροφορίες είναι καλύτερα να αναζητηθούν σε αρχεία, στατιστικές ή να ληφθούν με άμεσες παρατηρήσεις ή μετρήσεις. Άλλωστε και τα ερωτώμενα άτομα αρκετές φορές δεν δίνουν τις σωστές πληροφορίες, όπως π.χ. όταν

- δεν γνωρίζουν τίποτα σχετικό με το αντικείμενο της έρευνας
- δεν θυμούνται την πληροφορία (όταν κυρίως αναφερόμαστε στο παρελθόν)
- πρόκειται για προσωπικές ερωτήσεις στις οποίες οι απαντήσεις είναι συχνά ψευδείς
- δεν κατάλαβαν σωστά την ερώτηση επειδή περιείχε αόριστες λέξεις
- οι ερωτήσεις είναι βαρετές και ο απαντητής απαντά στην τύχη και χωρίς προσοχή.

Έτσι, ο σχεδιασμός του ερωτηματολογίου απαιτεί προσοχή για να αποφευχθούν σοβαρά λάθη. Στη συνέχεια, θα εξετάσουμε τις ιδιότητες ενός καλού ερωτηματολογίου και τον τρόπο καταρτισμού του.

1.6.1 Διαδοχικά βήματα για τη σύνταξη ενός ερωτηματολογίου

Η διαδικασία σχεδιασμού και οργάνωσης του ερωτηματολογίου χωρίζεται στις παρακάτω τρεις φάσεις:

A) Προετοιμασία: Στη φάση αυτή πρέπει:

- 1) Να κατανοήσουμε πλήρως τους στόχους της έρευνας.
- 2) Να συζητήσουμε με τον ενδιαφερόμενο (για τη διεξαγωγή της συγκεκριμένης έρευνας) οτιδήποτε σχετικό με την έρευνα.

- 3) Να ενημερωθούμε πλήρως γύρω από το θέμα της έρευνας ανατρέχοντας σ' όλες τις σχετικές πηγές.
- 4) Να συνταχθεί μία σειρά ερωτήσεων, οι απαντήσεις των οποίων θα δώσουν τις απαιτούμενες πληροφορίες.
- 5) Η τοποθέτηση των ερωτήσεων σε μία λογική σειρά που θα τις κάνει όσο το δυνατό πιο ενδιαφέρουσες και αποτελεσματικές.
- 6) Να φροντίσουμε έτσι ώστε το ερωτηματολόγιο να βοηθάει στην επεξεργασία των αποτελεσμάτων με τη χρήση ηλεκτρονικού υπολογιστή.
- 7) Να μην επιτραπεί η τελική ανάμειξη του ενδιαφερόμενου (πελάτη) στη διαμόρφωση του ερωτηματολογίου.

B) Βασικές αρχές:

Πρέπει να ληφθούν υπόψη γενικές βασικές αρχές, τόσο για τα χαρακτηριστικά του ερωτηματολογίου όσο και στη διαμόρφωση των ερωτήσεων.

B1) Χαρακτηριστικά ενός ερωτηματολογίου

Ένα ερωτηματολόγιο πρέπει:

- α) Να ανταποκρίνεται στους στόχους της έρευνας. Οι ερευνητές δεν πρέπει να παρασύρονται από τις δικές τους προκαταλήψεις ή τα δικά τους ενδιαφέροντα ώστε να δίνουν υπερβολική σημασία σε μία λιγότερο σημαντική ερώτηση και να παραμελούν κάποια πιο σημαντική.
- β) Να είναι όσο το δυνατό σύντομο και απλό ενώ ταυτόχρονα να καλύπτει όλες τις σχετικές πληροφορίες.
- γ) Να εξασφαλίζει την ανταπόκριση του ερωτώμενου.
- δ) Να διευκολύνει τους ερωτώμενους να δίνουν τις πληροφορίες που έχουν.
- ε) Να αποφεύγει την παροχή άσχετων πληροφοριών από τον ερωτώμενο, καθώς και τις παρεκκλίσεις από το θέμα.
- στ) Να περιέχει ερωτήσεις-κλειδιά για τον έλεγχο της αξιοπιστίας των απαντήσεων.
- ζ) Να διευκολύνει την εκτίμηση, ανάλυση και ερμηνεία αποτελεσμάτων.
- η) Να μην υπερβαίνει τον διαθέσιμο χρόνο.

B2) Διαμόρφωση των ερωτήσεων

Μερικοί βασικοί κανόνες για τη διαμόρφωση των ερωτήσεων είναι οι εξής:

- α) Πρέπει να αποφεύγονται οι μεγάλοι μήκους ερωτήσεις. Μία σειρά από σύντομες ερωτήσεις σε διάφορες απόψεις του θέματος είναι καλύτερη από μία μακροσκελή ερώτηση που συνήθως είναι και πολύπλοκη.
- β) Οι ερωτήσεις πρέπει να είναι απλές. Να μην περιέχουν δυσνόητες λέξεις ή λέξεις που στο ευρύ κοινό μπορεί να έχουν διαφορετική σημασία.
- γ) Ο ερωτώμενος να μπορεί να απαντήσει από τη δική του πείρα και γνώση.
- δ) Να αποκλείονται οι περιττές ερωτήσεις.
- ε) Επιβάλλεται μεγάλη προσοχή στις ερωτήσεις που χρειάζονται αριθμητική ή ποσοτική απάντηση. Π.χ. σε μία ερώτηση.

- Πόσα σαπούνια χρησιμοποιείτε στο σπίτι σας κάθε μήνα;

οπωσδήποτε θα υπάρξει πρόβλημα γιατί, όπως είναι γνωστό, υπάρχουν διάφορου μεγέθους σαπούνια.

- στ) Οι ερωτήσεις πρέπει να επιδέχονται απαντήσεις που να μπορούν να σημειωθούν (απαντηθούν) γρήγορα και σωστά.
- ζ) Σε ορισμένες περιπτώσεις προτιμάται να τονίζεται κάποια άποψη προκειμένου ο απαντητής να ενθαρρύνεται να δώσει τη σωστή απάντηση. Όπως στις περιπτώσεις που η απάντηση του ερωτώμενου μπορεί να θεωρηθεί ότι είναι κατώτερη ή όχι κοινωνικά αποδεκτή.
- η) Αντίθετα με την προηγούμενη περίπτωση, οι ερωτήσεις δεν πρέπει να φορτώνονται γιατί ο απαντητής επηρεάζεται. Όπως για παράδειγμα

- Πολλοί ειδικοί πιστεύουν ότι ... αλλά άλλοι διαφωνούν. Ποια είναι η δική σας γνώμη;

Σε μία τέτοια περίπτωση υπάρχει το ενδεχόμενο ο απαντητής να συμφωνήσει με τους "ειδικούς".

- θ) Να μην περιλαμβάνονται ερωτήσεις που ζητούν ταυτόχρονα δύο διαφορετικά θέματα. Όπως για παράδειγμα στην ερώτηση

- Είστε ευχαριστημένοι με το μισθό και το ωράριο της δουλειάς σας;

μπορεί κάποιος να είναι ευχαριστημένος με το ωράριο όχι όμως και το μισθό του!

- Προκειμένου να προβεί η κυβέρνηση στην εκτροπή του Αχελώου, πρέπει κατά τη γνώμη σας τα απαραίτητα έξοδα να καλυφθούν από ειδική φορολογία ή να γίνει περικοπή άλλων δαπανών γεωργικής ανάπτυξης;

Μπορεί κάποιος να μην είναι υπέρ της εκτροπής του Αχελώου, οπότε οποιαδήποτε απάντηση περιττεύει.

- ι) Ιδιαίτερη προσοχή απαιτείται σε ερωτήσεις διπλής άρνησης, όπως
- Εγκρίνετε ή δεν εγκρίνετε να απαγορευτεί η κυκλοφορία των ΤΑΞΙ στους λεωφορειοδρόμους;

Μία τέτοια ερώτηση σίγουρα θα δημιουργήσει σύγχυση στον ερωτώμενο.

- ια) Πρέπει να αντιμετωπίζονται προσεκτικά οι ερωτήσεις που ενοχλούν, π.χ. προσωπικές.

Μία μέθοδος για αποφυγή τέτοιου είδους ερωτήσεων είναι η μέθοδος Warner που περιγράφεται παρακάτω μ' ένα παράδειγμα.

Έστω ότι η ερώτηση που ενδιαφέρει τον ερευνητή είναι:

A: "Κάνετε ειλικρινή δήλωση στην εφορία;"

Η μέθοδος Warner περιλαμβάνει τα βήματα:

- α) Ο ερωτώμενος ρίχνει ένα νόμισμα και δεν ανακοινώνει το αποτέλεσμα (K ή I) στον ερευνητή.
- β) Στη συνέχεια, εκλέγει με τυχαίο τρόπο ένα ψηφίο από τα 0, 1, 2, ..., 9 και δεν λέει στον ερευνητή τι ακριβώς ψηφίο διάλεξε.
- γ) Ο ερευνητής του υποβάλλει δύο ερωτήσεις, την παραπάνω ερώτηση A και την
B: "Το αποτέλεσμα της ρίψης του νομίσματος είναι K ;"
- δ) Ο ερωτώμενος, με υπόδειξη του ερευνητή, θα απαντήσει καταφατικά ή αρνητικά (NAI ή OXI) σε μία μόνο ερώτηση (την οποία φυσικά δεν θα ξέρει ο ερευνητής). Θα απαντήσει στην ερώτηση A αν το ψηφίο που διάλεξε είναι 0, 1, 2, 3, διαφορετικά θα απαντήσει στην ερώτηση B.

Η διαδικασία αυτή επαναλαμβάνεται για όλα τα άτομα του δείγματος.

Στη συνέχεια, το ποσοστό των ατόμων του πληθυσμού που απαντούν NAI στην ερώτηση A, βρίσκεται χρησιμοποιώντας το θεώρημα της ολικής πιθανότητας.

Ορίζουμε τα ενδεχόμενα

N : Ο ερωτώμενος απάντησε ΝΑΙ
 A : “ “ “ στην ερώτηση Α
 B : “ “ “ στην ερώτηση Β

Τότε

$$P(N) = P(N|A)P(A) + P(N|B)P(B).$$

Έτσι, αν έχουμε π.χ. συνολικό ποσοστό απαντήσεων με “ΝΑΙ” 60%, λαμβάνουμε (δεδομένου ότι $P(A) = 4/10$, $P(B) = 6/10$ και $P(N|B) = 1/2$)

$$\frac{6}{10} = P(N|A)\frac{4}{10} + \frac{1}{2}\frac{6}{10}$$

από την οποία προκύπτει η

$$P(N|A) = \frac{3}{4} = 75\%$$

δηλ. το ποσοστό των ατόμων στον πληθυσμό που κάνουν ειλικρινείς δηλώσεις είναι 75%.

Γ) Δοκιμαστική έρευνα

Μια δοκιμαστική έρευνα πρέπει να γίνεται στο 5%-10% του τελικού δείγματος για να διαπιστώνεται αν

- α) το ερωτηματολόγιο δημιουργεί ή όχι αρνητικές αντιδράσεις
- β) κρατάει το ενδιαφέρον μέχρι τέλους
- γ) η σειρά των ερωτήσεων είναι σωστή
- δ) ο χρόνος της συνέντευξης είναι λογικός
- ε) δεν υπάρχουν δυσνόητες έννοιες.

Επίσης ο προέλεγχος μπορεί

- στ) να βοηθήσει τον συνεντευχτή στην εξοικώωσή του με το ερωτηματολόγιο
- ζ) να δώσει τις απαραίτητες πληροφορίες για τη σύνταξη των ερωτήσεων πολλαπλής εκλογής στην τελική τους μορφή. Οι απαντήσεις που θα δοθούν κατά τη διάρκεια του προελέγχου δίνουν μία εικόνα του είδους των απαντήσεων που θα πρέπει να αναμένονται σε ωρισμένες ερωτήσεις.

- η) Τέλος, μπορούμε να πούμε ότι γίνεται εξέταση των αποτελεσμάτων της παραπάνω δοκιμής, από άποψη ευχέρειας ταξινόμησης και ποιότητας των συγκεντρωμένων πληροφοριών. Αν μετά τον προέλεγχο γίνουν μεγάλες αλλαγές και αναμόρφωση του ερωτηματολογίου, πρέπει να γίνει ξανά δεύτερος έλεγχος (pilot study) πριν την τελική υποβολή του.

Κατ' αυτόν τον τρόπο μπορούμε στη συνέχεια να περιορίσουμε πολύ τα μη δειγματοληπτικά σφάλματα.

1.6.2 Τύποι ερωτήσεων

Υπάρχουν διάφοροι τύποι ερωτήσεων που χρησιμοποιούνται στον καταρτισμό ενός ερωτηματολογίου.

α) Απλής εκλογής. Δίνουν στον ερωτώμενο τη δυνατότητα μόνο δύο απαντήσεων, συνήθως ΝΑΙ ή ΟΧΙ, όπως π.χ.

- Έχετε κάνει μεταπτυχιακές σπουδές;

ΝΑΙ ☐

ΟΧΙ ☐

β) Πολλαπλής εκλογής. Ο ερωτώμενος μπορεί να διαλέξει μεταξύ ενός αριθμού απαντήσεων. Π.χ.

- Για ποιό λόγο αποφασίσατε να κάνετε μεταπτυχιακές σπουδές;

Επιστημονική εξέλιξη ☐

Επαγγελματική αποκατάσταση ☐

Κοινωνικό γόητρο ☐

Επίδομα σπουδών ☐

Επιστημονικό ενδιαφέρον ☐

γ) Ανοιχτές ερωτήσεις. Επιτρέπουν στον απαντητή να δώσει μία απάντηση κατά τη δική του κρίση. Π.χ.

- Πώς νομίζετε ότι πρέπει να γίνονται οι εξετάσεις στο μάθημα της Δειγματοληψίας;

δ) Συνδυασμός κλειστών ερωτήσεων με ανοικτή. Σε μερικές ερωτήσεις μπορεί να έχουμε συνδυασμό περίπτωσης πολλαπλής επιλογής με δυνατότητα στο τέλος ανοικτής απάντησης. Για παράδειγμα, στις παραπάνω απαντήσεις για

τις μεταπτυχιακές σπουδές μπορεί να προστεθεί στο τέλος και η δυνατότητα απάντησης

Κάτι άλλο. Προσδιορίστε

ε) Ερωτήσεις σημαντικότητας. Χρησιμοποιώντας για την καταγραφή της κλιμάκωσης της γνώμης που επικρατεί για κάποιο γεγονός, καταλληλότητα ενός μηχανήματος κ.λπ. Όπως για παράδειγμα,

- Πώς κρίνετε το νέο νομοσχέδιο για τα φορολογικά μέτρα;

- Πολύ καλό ☐
 Καλό ☐
 Μέτριο ☐
 Κακό ☐
 Πολύ κακό ☐

Η τυποποίηση στις ερωτήσεις, εκτός του ότι είναι αρκετά κατανοητή από τους απαντητές, μπορεί επί πλέον να χρησιμοποιηθεί ταυτόχρονα σε πολλές ερωτήσεις. Π.χ. χρησιμοποιώντας τους βαθμούς 1, 2, 3 (με το 1 να δείχνει ελάχιστο και το 3 μέγιστο) απαντήστε στα παρακάτω ερωτήματα:

- Πώς κρίνετε το μάθημα της Δειγματοληψίας από άποψη:

α. Επαγγελματικής χρησιμότητας	1	2	3
β. Ποιότητα διδασκαλίας	1	2	3
γ. Συμμετοχής φοιτητών	1	2	3

Γενικά μπορεί να ειπωθεί ότι:

- i) Για προσωπικές συνεντεύξεις και ερωτηματολόγια που αποστέλλονται ταχυδρομικά μπορούν να χρησιμοποιηθούν όλοι οι τύποι των ερωτήσεων.
- ii) Για τηλεφωνικές συνεντεύξεις είναι προτιμότερες οι ερωτήσεις απλής εκλογής. Μπορούν επίσης να συνδυαστούν και με μερικές ανοικτές ερωτήσεις.

Γενικά οι κλειστές ερωτήσεις είναι πιο αξιόπιστες και μπορούν να χρησιμοποιηθούν εύκολα σε στατιστικές τεχνικές. Είναι εύκολο να απαντηθούν, να κωδικοποιηθούν και να αναλυθούν. Από την άλλη πλευρά, οι ανοικτές ερωτήσεις είναι χρήσιμες για να καταγραφούν σκέψεις και στην ανάλυση σχολίων και σπάνιων απαντήσεων. Υπάρχουν περιπτώσεις που οι ερωτώμενοι θέλουν να απαντήσουν με δικά τους λόγια.

Είναι προφανές ότι τόσο οι κλειστές όσο και οι ανοικτές ερωτήσεις έχουν, ανάλογα με το αντικείμενο που εξετάζεται, τα πλεονεκτήματα και τα μειονεκτήματά τους. Έτσι, για παράδειγμα, σε μία ανοικτή ερώτηση

- Ποιό είναι κατά τη γνώμη σας το κυριότερο πρόβλημα που αντιμετωπίζει η περιοχή σας;

και στην αντίστοιχη κλειστή ερώτηση

- Ποιό από τα παρακάτω προβλήματα είναι κατά τη γνώμη σας το κυριότερο που αντιμετωπίζει η περιοχή σας;

(Αναφέρονται στο τέλος ωρισμένα από τα προβλήματα που συνήθως παρουσιάζονται όπως, κυκλοφοριακό, παιδικοί σταθμοί, χώροι πρασίνου κ.λπ.).

Απαντώντας κάποιος στην ανοικτή ερώτηση μπορεί να μην αναφέρει κάποιο από τα προβλήματα που αναφέρονται στην κλειστή ερώτηση, παρόλο που θα το θεωρούσε σαν το σημαντικότερο, ίσως γιατί δεν θα το σκέφτηκε εκείνη τη στιγμή. Έτσι δεν μαθαίνουμε την πραγματική του γνώμη. Αυτό προφανώς αποτελεί μειονέκτημα των ανοικτών ερωτήσεων. Από την άλλη πλευρά όμως, απαντώντας κάποιος στην κλειστή ερώτηση, του αφαιρείται η δυνατότητα να πει το κυριότερο κατά τη γνώμη του πρόβλημα της περιοχής, αν αυτό δεν συμπεριλαμβάνεται στον κατάλογο. Αυτό προφανώς αποτελεί μειονέκτημα των κλειστών ερωτήσεων.

Πλεονέκτημα των κλειστών ερωτήσεων είναι το γεγονός ότι ο απαντητής βλέποντας τον κατάλογο των προβλημάτων διαλέγει κάποιο που το θεωρεί σημαντικό ενώ ίσως δεν θα το σκεφτότανε σε μία ανοικτή ερώτηση. Η κωδικοποίηση στις κλειστές ερωτήσεις (για την παραπέρα ανάλυση) είναι ευκολότερη παρά στις ανοικτές ερωτήσεις, όπου οι απαντήσεις μπορεί να ξεφεύγουν από το θέμα και να δίνουν απαντήσεις χωρίς καμμία πληροφορία. Για παράδειγμα, στην ερώτηση

- Πότε μετακομίσατε στην τωρινή σας κατοικία;

μπορεί κάποιος να απαντήσει

- πριν 10 χρόνια, · από τότε που παντρεύτηκα
- πριν λίγο καιρό, · όταν ήταν τα παιδιά μικρά.

Μερικές περιπτώσεις που μπορούν να χρησιμοποιηθούν οι ανοικτές ερωτήσεις είναι:

- i) όταν υπάρχουν πολλές κατηγορίες απαντήσεων που δεν μπορούν να προβλεφθούν
- ii) όταν απαιτείται η ανεπηρέαστη απάντηση του ερωτώμενου
- iii) όταν ενδιαφερόμαστε γενικότερα για την τοποθέτηση του απαντητή σε κάποιο πολιτικό ή κοινωνικό φαινόμενο

- iv) σε μια σειρά από κλειστές ερωτήσεις ακολουθεί κάποια ανοικτή για να δοθεί έτσι η αίσθηση στον ερωτώμενο ότι δίνεται και σ' αυτόν η ευκαιρία να εκφράσει ελεύθερα την άποψή του.

1.6.3 Ταξινόμηση και παρουσίαση των ερωτήσεων

Κάθε ερωτηματολόγιο έχει συνήθως μια εισαγωγή με την οποία εξετάζεται αν ο ερωτώμενος θέλει να συμμετάσχει στην έρευνα ή όχι. Στην εισαγωγή πρέπει να δηλώνεται για ποιόν γίνεται η έρευνα και ποιό είναι το θέμα της. Αυτές οι πληροφορίες πρέπει να δίνονται με όσο το δυνατόν πιο σύντομο τρόπο, έτσι ώστε γρήγορα να υποβάλλεται η πρώτη ερώτηση. Όταν υποβληθεί η πρώτη ερώτηση και δοθεί η απάντηση, ο ερωτώμενος συνήθως δένεται και σπάνια αρνείται να συνεχίσει.

Ένα ερωτηματολόγιο, για παράδειγμα, της εταιρείας XXXXX στις σφυγμομετρήσεις της κοινής γνώμης για την ακροαματικότητα και αναγνωσιμότητα, συνοδευόταν από τον παρακάτω πρόλογο:

Καλημέρα σας / Καλησπέρα σας,

Είμαι εκπρόσωπος της εταιρείας XXXXX που κάνει σφυγμομετρήσεις της κοινής γνώμης για διάφορα θέματα.

Αυτές τις ημέρες ρωτάμε τους ανθρώπους για τις συνήθειες που έχουν σχετικά με τις εφημερίδες, το ραδιόφωνο, την τηλεόραση, τα περιοδικά και άλλα θέματα της καθημερινής ζωής.

Το νοικοκυριό σας έχει επιλεγεί με επιστημονικές μεθόδους, μεταξύ πολλών άλλων νοικοκυριών, και η βοήθειά σας, θα μας είναι πολύτιμη.

Θέλουμε να σας βεβαιώσουμε ότι οι απαντήσεις σας θα μείνουν εντελώς ανώνυμες και ό,τι μας πείτε θα χρησιμοποιηθεί αποκλειστικά και μόνο για στατιστικούς πίνακες.

Ευχαριστώ.

Μερικές φορές όμως, αναφέροντας το όνομα της εταιρείας για την οποία γίνεται η έρευνα μπορεί να επηρεάσει τις απαντήσεις. Είναι λάθος να πούμε

- Κάνω μία έρευνα για την οδοντόκρεμα YYYYY. Εσείς τι οδοντόκρεμα χρησιμοποιείτε;

Επίσης, το να ονομάσουμε συγκεκριμένο περιοδικό ή ασφαλιστική εταιρεία, μπορεί να θεωρηθεί από τον ερωτώμενο σαν κάποια προσπάθεια πώλησης.

Πολλές φορές δεν πρέπει να αναφέρεται ούτε το ακριβές θέμα της έρευνας. Π.χ. δεν είναι σωστό να πούμε ότι:

- Κάνουμε μία έρευνα για τα μαζικά μέσα συγκοινωνίας

και μετά να ρωτήσουμε για τα προβλήματα της περιοχής. Ούτε μπορούμε να πούμε π.χ. ότι η έρευνα γίνεται με θέμα την πυρηνική ιατρική γιατί λίγα άτομα αισθάνονται ικανά να ανταποκριθούν σε μία τέτοια συνέντευξη. Συνήθως είναι καλύτερα να περιγράφουμε το σκοπό της έρευνας με γενικές λέξεις, όπως **κοινωνικά προβλήματα, οικογενειακή ζωή κ.λπ.**

Σε αρκετές περιπτώσεις δεν είναι αναγκαίο να προσδιορίσουμε το χρόνο της συνέντευξης ή να πούμε ότι οι απαντήσεις είναι εμπιστευτικές. Όμως, ο συνεντευχτής πρέπει να είναι έτοιμος να απαντήσει αν ο ερωτώμενος το ζητήσει.

Η πρώτη ερώτηση θα πρέπει να είναι σχετικά εύκολη και όχι απειλητική. Δεν ξεκινάμε ποτέ με προσωπικές ερωτήσεις με τις οποίες ο ερωτώμενος θα προβληματίζεται να απαντήσει. Οι δύσκολες και ευαίσθητες ερωτήσεις γίνονται όταν ο ερωτώμενος έχει προσαρμοστεί στο κλίμα της συνέντευξης.

Δημογραφικές ερωτήσεις, όπως ηλικία, εισόδημα, επάγγελμα και μόρφωση συνήθως τοποθετούνται στο τέλος του ερωτηματολογίου γιατί πολλοί θεωρούν αυτά τα θέματα προσωπικά.

Καλό είναι να προετοιμάζουμε τον ερωτώμενο σε περιπτώσεις αλλαγής των θεμάτων. Όπως για παράδειγμα,

- Και τώρα μερικές ερωτήσεις σχετικά με

Η ροή των ερωτήσεων δεν πρέπει να είναι μονότονη. Για τους λόγους που αναφέρθηκαν προηγούμενα πρέπει να βάζουμε μερικές ανοικτές ερωτήσεις σε διάφορα σημεία του ερωτηματολογίου. Επίσης για τον έλεγχο της αξιοπιστίας του ερωτώμενου κάποιες ερωτήσεις (οι ίδιες ή παρόμοιες) τοποθετούνται σε διαφορετικά σημεία του ερωτηματολογίου, οπότε έτσι ελέγχεται αν ο ερωτώμενος απάντησε το ίδιο στις ερωτήσεις αυτές (βλ. § 1.8 παρακάτω).

Τέλος, αναφέρουμε ότι και η εκτύπωση, εμφάνιση και όλη γενικά η παρουσίαση του ερωτηματολογίου επηρεάζουν το ποσοστό απαντήσεων και το επίπεδό τους.

Από τα παραπάνω φαίνεται ότι η τελική διαμόρφωση του ερωτηματολογίου προκύπτει σαν αποτέλεσμα σύνθετης και πολύ προσεκτικής εργασίας και συνεπώς δεν μπορούμε να χρησιμοποιούμε έτοιμους τύπους ερωτηματολογίων που έγιναν στο παρελθόν για κάποια έστω παραπλήσια έρευνα και να τους εφαρμόζουμε απόλυτα στη συγκεκριμένη κάθε φορά έρευνα που σχεδιάζουμε.

1.7 Συλλογή πληροφοριών

Στην παράγραφο αυτή αναφέρουμε ορισμένα στοιχεία για τις τεχνικές συλλογής των δεδομένων που απαιτεί μία δειγματοληπτική έρευνα.

Οι πληροφορίες συλλέγονται συνήθως με τη βοήθεια τυποποιημένων ερωτήσεων. Οι ερωτήσεις είναι κοινές για όλα τα άτομα του πληθυσμού και έτσι κάθε εξεταζόμενο άτομο απαντά στις ίδιες ακριβώς ερωτήσεις. Οι πληροφορίες που συγκεντρώνονται είναι εμπιστευτικές και σε καμμία περίπτωση δεν πρέπει να αποκαλύπτονται. Στην Ελλάδα το απόρρητο των στοιχείων των δειγματοληπτικών ερευνών προστατεύεται από το άρθρο 40 του Ν.Δ. 3627/56 και οι παραβάτες τιμωρούνται με αυστηρές ποινές.

Για τη συγκέντρωση των απαραίτητων πληροφοριών υπάρχουν αρκετές μέθοδοι. Πολλές από αυτές όμως δεν έχουν πρακτική αξία κυρίως λόγω του υψηλού κόστους που απαιτούν ή είναι εξαιρετικά πολύπλοκες. Οι κυριότεροι τρόποι συλλογής πληροφοριών που χρησιμοποιούν οι ερευνητές, κυρίως όταν γίνεται χρήση ερωτηματολογίων είναι

- Ταχυδρομικά (έρευνα μέσω ταχυδρομείου)
- Τηλεφωνικά (τηλεφωνική συνέντευξη)
- Προσωπική συνέντευξη
- Συνδυασμός των παραπάνω
- Αυτόματα.

Κάθε μία από τις μεθόδους αυτές συλλογής των πληροφοριών παρουσιάζει πλεονεκτήματα έναντι των άλλων, αλλά ταυτόχρονα έχει και τα μειονεκτήματά της.

α) Ταχυδρομική συλλογή. Η μέθοδος αυτή έχει σχετικά μικρό κόστος, και αυτό ακριβώς αποτελεί και το κυριότερο πλεονέκτημά της. Με χαμηλό κόστος μπορούν να συλλεγούν πληροφορίες από ένα μεγάλο αριθμό ατόμων. Η μέθοδος αυτή είναι επίσης αποτελεσματική όταν η έρευνα αναφέρεται στα μέλη ειδικών κατηγοριών ατόμων. Στα άτομα αυτά υπάρχει κατά κανόνα ένα αυξημένο ενδιαφέρον των ερωτώμενων για την έρευνα. Για παράδειγμα ανάμεσα στα μέλη ενός επαγγέλματος υπάρχει εντονότερο ενδιαφέρον όταν κυρίως η έρευνα αφορά το μέλλον του επαγγέλματός τους. Στην περίπτωση αυτή το ποσοστό απαντήσεων είναι πολύ μεγαλύτερο παρά σε άλλες περιπτώσεις, φτάνοντας συχνά μέχρι και 70%.

Όταν δεν έχουμε ειδικές κατηγορίες ατόμων όπως προηγούμενα, η μέθοδος αυτή παρουσιάζει διάφορα μειονεκτήματα. Το κυριότερο μειονέκτημα συνίσταται στο πολύ μικρό ποσοστό απαντήσεων, πολλές φορές μάλιστα γύρω στο 20%. Κατ' άλλους ερευνητές - κυρίως σε έρευνες εμπορίου - το ποσοστό απαντήσεων δεν ξεπερνά το 5-10%. Ένας τρόπος αντιμετώπισης του χαμηλού ποσοστού

απαντήσεων είναι η αποστολή δεύτερης (ή και τρίτης) υπενθυμιτικής επιστολής, η επισύναψη απαντητικού φακέλλου με πληρωμένα ταχυδρομικά τέλη ή ακόμα και η προσφορά κάποιας “χρηματικής αποζημίωσης” (πολλές φορές συμμετοχή σε κλήρωση κάποιου δώρου). Όλα αυτά όμως αυξάνουν το κόστος της έρευνας. Είναι ενδιαφέρον εδώ να σημειώσουμε ότι σε δειγματοληπτική έρευνα που έγινε για τους πτυχιούχους του Μαθηματικού Τμήματος παρόλο που όλα τα άτομα του δείγματος ενημερώθηκαν τηλεφωνικά για την έρευνα και τους σκοπούς της και σχεδόν όλοι μας διαβεβαίωσαν ότι θα αποστείλουν απαντημένο το ερωτηματολόγιο που τους είχε αποσταλεί, τελικά μόνο το 40% το έκανε.

Άλλα μειονεκτήματα της μεθόδου αυτής είναι η δυσκολία εκλογής αντιπροσωπευτικού δείγματος και τα σφάλματα που θα προκύψουν από πιθανή μη σωστή συμπλήρωση των ερωτήσεων που θα οφείλεται σε άγνοια, παρανόηση κ.λπ.

β) Τηλεφωνικά. Η τηλεφωνική συνέντευξη είναι μία αποτελεσματική μέθοδος συλλογής ορισμένων δεδομένων, έχει και αυτή μικρό σχετικά κόστος, είναι απλή και γρήγορη. Η μέθοδος αυτή συνίσταται κυρίως όταν απευθυνόμαστε σε πολυάσχολους διευθυντές επιχειρήσεων και οργανισμών ή σε άλλα άτομα που δεν είναι διαθέσιμα για μία προσωπική συνέντευξη ή που ένα ερωτηματολόγιο, το οποίο θα στελλότανε ταχυδρομικά θα ξεχνιότανε, στην καλύτερή του περίπτωση, κλεισμένο μέσα σε κάποιο συρτάρι.

Οι τηλεφωνικές συνεντεύξεις χρησιμοποιούνται κυρίως όταν χρειάζεται απλή καταγραφή στατιστικών στοιχείων, όπως π.χ. πόσοι υπάλληλοι απήργησαν, ακροαματικότητα ραδιοφωνικών ή τηλεοπτικών προγραμμάτων, κ.λπ.

Από την άλλη πλευρά, και η μέθοδος αυτή συλλογής των πληροφοριών παρουσιάζει διάφορα μειονεκτήματα. Το ερωτηματολόγιο πρέπει να είναι **σύντομο** και απλό, υπάρχει ο κίνδυνος παρανόησης των ερωτήσεων εκ μέρους του απαντητή ή της καταγραφής των απαντήσεων από την πλευρά του συνεντευκτή. Πολλές πληροφορίες σπάνια δίνονται τηλεφωνικά πράγμα που περιορίζει το βάθος της έρευνας. Επίσης ένα κύριο μειονέκτημα της μεθόδου αυτής είναι ότι πολλά άτομα δεν διαθέτουν τηλέφωνο (δεν είναι γνωστό, κρυφός αριθμός) και έτσι οδηγούμαστε σε λανθασμένα συμπεράσματα (βλ. παράδειγμα των προεδρικών εκλογών στις ΗΠΑ το 1936).

Όπως θα δούμε και παρακάτω η τηλεφωνική επικοινωνία είναι απαραίτητη και:

- στην επιλογή καταλλήλων ατόμων για προσωπικές συνεντεύξεις
- στον έλεγχο πληροφοριών που συγκεντρώθηκαν είτε από προσωπικές συνεντεύξεις, είτε από έρευνες με το ταχυδρομείο (quality control)

γ) Προσωπική συνέντευξη. Η μέθοδος αυτή της συλλογής της απαραίτητης πληροφορίας για μία δειγματοληπτική έρευνα είναι συνήθως η πιο αξιόπιστη. Είναι ουσιώδης σε περιπτώσεις που πρόκειται να συλλεγούν πολύπλοκες πληροφορίες και όταν η έρευνα απαιτεί ακρίβεια και έλεγχο σε βάθος. Με τη μέθοδο αυτή είναι εύκολο στον συνεντευχτή να παρατηρεί τις αντιδράσεις του συνομιλητή του και τις εκφράσεις του προσώπου του, σχηματίζοντας έτσι σωστότερη γνώμη για το τι ακριβώς πιστεύει. Χρειάζεται όμως άρτια καταρτισμένο συνεντευχτή και με μεγάλη εμπειρία ο οποίος θα έχει εκπαιδευτεί και δοκιμαστεί κατάλληλα πριν αρχίσει την επιτόπια έρευνα.

Άλλα πλεονεκτήματα της μεθόδου αυτής είναι:

- δεν θα έχουμε τόσο μεγάλο ποσοστό αναπάντητων ερωτήσεων (ερωτηματολογίων) όσο στις άλλες μεθόδους (τηλεφωνικά, ταχυδρομικά).
- δίνονται οι απαραίτητες διευκρινίσεις σε οποιαδήποτε απορία του απαντητή ή του ζητείται ξανά τι ακριβώς επιδιώκει η ερώτηση όταν διαφανεί ότι βρίσκεται εκτός θέματος
- μπορούν να επιδειχθούν στον ερωτώμενο διάφορα δείγματα, σχεδιαγράμματα, κατάλογοι αντικειμένων για τα οποία ενδιαφέρεται η έρευνα, κ.λπ.
- παρέχεται πληρέστερη ενημέρωση του απαντητή σχετικά με τις ερωτήσεις.

Αντίθετα, το κυριότερο μειονέκτημα της μεθόδου αυτής συλλογής της πληροφορίας είναι το υψηλό κόστος που απαιτείται, ιδίως, όταν πρόκειται να ληφθούν προσωπικές συνεντεύξεις από άτομα που βρίσκονται σε μακρινές αποστάσεις, οπότε πρέπει να συμπεριληφθούν στα έξοδα και τα εισιτήρια, οι διανυκτερεύσεις και οπωσδήποτε το πρόβλημα του χρόνου. Επίσης το υψηλό κόστος για την εκπαίδευση του κατάλληλου προσωπικού, γιατί αλλιώς, όπως είδαμε και παραπάνω, θα έχουμε μεταφορά των απαντήσεων του ερωτώμενου παραποιημένες βάσει προσωπικών προκαταλήψεων ή και δυσχέρειας σωστής καταγραφής εκ μέρους του συνεντευχτή.

δ) Συνδυασμός των παραπάνω. Μπορούμε π.χ. να τηλεφωνήσουμε στα άτομα που δεν έστειλαν ταχυδρομικά το ερωτηματολόγιό τους, ή επίσης να κλείσουμε ραντεβού για προσωπική συνέντευξη.

ε) Αυτόματα. Η μέθοδος αυτή εφαρμόζεται σε περιπτώσεις που δεν έχουμε ερωτηματολόγιο, αλλά μας ενδιαφέρουν κάποιες μετρήσεις, όπως π.χ. κίνηση οχημάτων σ' ένα δρόμο σε διάφορες χρονικές περιόδους, καταμέτρηση ακροαματικότητας κάποιας τηλεοπτικής εκπομπής με κάποιο “μετρητή” που ενσωματώνεται στην τηλεόραση, καταγραφή στατιστικών κυρίως στοιχείων από έτοιμους καταλόγους (όπως για παράδειγμα, από καταστάσεις της Γραμματείας μπορούμε να δούμε ή να εκτιμήσουμε το ποσοστό των φοιτητών που περνούν όλα τα μαθήματά τους από την πρώτη περίοδο), καταγραφή αριθμού ατόμων που επισκέπτονται ένα μουσείο, κ.λπ.

Για τις μεθόδους μάρκετινγκ (marketing) ο ερευνητής μπορεί να συγκεντρώσει πληροφορίες από διαφημίσεις στον τύπο και στα άλλα μέσα ενημέρωσης, από εμπορικές εκθέσεις, κ.λπ.

Παράδειγμα 1.1 (συνέχεια). Κάθε μία από τις παραπάνω μεθόδους συλλογής των δεδομένων έχει συγκεκριμένες επιταγές γεγονότος που θα πρέπει να λαμβάνονται υπόψη κατά το σχεδιασμό της δειγματοληπτικής μελέτης. Για το παράδειγμα της υγειονομικής ασφάλισης είναι δύσκολο να βρούμε όλα τα στοιχεία ασφάλισης που χρειαζόμαστε από ήδη υπάρχουσες λίστες και πρωτόκολλα. Έτσι είναι απαραίτητο να σχεδιαστεί ένα ερωτηματολόγιο που πρέπει να συμπληρωθεί. Η αποστολή του ερωτηματολογίου ταχυδρομικά έχει μεν μικρό κόστος αλλά η λίστα των επιθυμητών προς συλλογή στοιχείων περιέχει προσωπικά δεδομένα που αφορούν σε εισόδημα και σε επίπεδο υγειονομικής ασφάλισης, πιθανή ασθένεια κ.ά. Για το λόγο αυτό η προσωπική συνέντευξη είναι ίσως πιο ασφαλής, θα είναι όμως και πιο ακριβή.

Σημαντικός επίσης είναι και ο καθορισμός των ατόμων που θα μπορούν να συμμετέχουν στη μελέτη και να δίνουν πληροφορίες που αφορούν είτε τον εαυτό τους είτε και για τα άλλα άτομα της οικογένειας. Για το παραπάνω παράδειγμα ο “αρχηγός” της οικογένειας θα μπορούσε να δώσει στοιχεία - απαντήσεις για όλα τα μέλη της οικογένειας. Ασφαλώς και θα γνωρίζει τη σχέση των μελών του νοικοκυριού, αν έχουν ή όχι υγειονομική ασφάλιση αν έχουν νοσηλευτεί κατά τη διάρκεια του χρόνου αναφοράς. Αν όμως ζητούσαμε στην έρευνα και το βαθμό ικανοποίησης των ατόμων από τις προσφερόμενες υπηρεσίες κατά τη διάρκεια νοσηλείας τους, ή αν υπάρχουν ερωτήματα όπως η αίσθηση του πόνου, τότε οι απαντήσεις σ’ αυτά τα ερωτήματα είναι εντελώς υποκειμενικές και θα έπρεπε να ρωτηθούν όλα τα μέλη του νοικοκυριού - οικογένειας.

1.8 Έλεγχος ποιότητας

Σε κάθε ένα από τα βήματα σχεδιασμού και διεξαγωγής μιας δειγματοληπτικής έρευνας πρέπει να γίνεται και ο σχετικός έλεγχος αν όλα έχουν γίνει σωστά ή αν έχουν γίνει παραλείψεις, υπερβολές ή λάθη οπότε πρέπει να διορθωθούν. Πρέπει να ελεγχθεί αν καθορίστηκε σωστά ο πληθυσμός, το πλαίσιο, το απλό στοιχείο, η δειγματοληπτική μονάδα, τα χαρακτηριστικά που ενδιαφέρουν τον ερευνητή. Μέσω της πιλοτικής έρευνας, πέραν της συλλογής προκαταρκτικής πληροφορίας για την προ-εκτίμηση κάποιων παραμέτρων, μας δίνεται η δυνατότητα ελέγχου του ερωτηματολογίου, αν δηλαδή σε κάποιες ερωτήσεις παρουσιάζονται προβλήματα κατανόησης ή δυσκολίας να αναμορφωθεί. Τα ερωτηματολόγια που τελικά θα χρησιμοποιηθούν πρέπει να είναι αμερόληπτα, αξιόπιστα και έγκυρα (βλ. § 1.9).

Μετά τη συλλογή των δεδομένων, πριν την ανάλυσή τους πρέπει να γίνουν διάφοροι έλεγχοι ποιότητας των δεδομένων αυτών. Για παράδειγμα, στα ερωτηματολόγια που είναι επώνυμα ή κωδικοποιημένα πρέπει να γίνεται σχετικός έλεγχος αν πράγματι το ερωτηματολόγιο συμπληρώθηκε από το συγκεκριμένο άτομο και όχι από τον συνεντεύκτη που μπορεί να το συμπληρώσει σίτι του (δυστυχώς γίνονται και αυτά!). Όπως αναφέραμε σ' ένα ερωτηματολόγιο υπάρχουν και κάποιες ερωτήσεις - δικλίδες για τον έλεγχο της αξιοπιστίας του ερωτώμενου. Αν διαπιστωθεί πράγματι κάτι τέτοιο, τότε το άτομο αυτό (ερωτηματολόγιο) αφαιρείται από το δείγμα.

Τέλος, ιδιαίτερος έλεγχος πρέπει να γίνεται στην κωδικοποίηση των απαντήσεων και αν αυτές μεταφέρθηκαν σωστά στον υπολογιστή όπου θα γίνει η επεξεργασία των δεδομένων. Με διάφορα υποπρογράμματα μπορούν να εντοπιστούν κάποιες ακραίες τιμές οπότε είτε διορθώνονται αν όντως έχουν μεταφερθεί λάθος είτε αφαιρούνται από το δείγμα. Για παράδειγμα, δεν μπορεί ένας μαθητής να δηλώνει ότι διαβάζει καθημερινά 54 ώρες!

1.9 Αμεροληψία, αξιοπιστία και εγκυρότητα

Η αμεροληψία (unbias), η αξιοπιστία (reliability) και η εγκυρότητα (validity) αποτελούν τα βασικά ποιοτικά επιστημονικά κριτήρια τα οποία θα πρέπει να ικανοποιεί κάθε ερωτηματολόγιο, συνέντευξη, παρατήρηση ή τεστ.

Εάν για παράδειγμα, σε επαναλαμβανόμενες δοκιμές η μέση τιμή των εκτιμήσεων διαφέρει από την ποσότητα που προτιθέμεθα να εκτιμήσουμε, τότε λέμε ότι η εκτιμήτριά μας είναι μεροληπτική (biased). Αν θ είναι μία προς εκτίμηση παράμετρος και $\hat{\theta}$ μια εκτιμήτρια, τότε η μεροληψία (bias) της $\hat{\theta}$, $b(\hat{\theta})$, εκφράζεται από τη διαφορά $E(\hat{\theta}) - \theta$. Δηλαδή μία εκτιμήτρια είναι μεροληπτική εάν η αναμενόμενη τιμή της διαφέρει από την πραγματική τιμή. Εάν η μεροληψία μιας εκτιμήτριας $\hat{\theta}$ είναι ίση με μηδέν, $b(\hat{\theta}) = E(\hat{\theta}) - \theta = 0$, τότε καλείται αμερόληπτη εκτιμήτρια (unbiased estimator). Με άλλα λόγια η $\hat{\theta}$ είναι αμερόληπτη εκτιμήτρια (α.ε.) μιας παραμέτρου θ εάν η μέση τιμή της δειγματικής κατανομής της $\hat{\theta}$ είναι ίση με θ , $E(\hat{\theta}) = \theta$.

Ένας από τους βασικούς στόχους μιας δειγματοληπτικής έρευνας, όπως εξάλλου και στη Στατιστική συμπερασματολογία γενικότερα, είναι ο προσδιορισμός μιας εκτιμήτριας $\hat{\theta}$ μιας παραμέτρου θ έτσι ώστε αυτή να είναι α.ε. αλλά ταυτόχρονα να έχει όσο το δυνατό μικρότερη διασπορά, δηλαδή τη μέση τιμή των τετραγώνων των αποκλίσεων των τιμών της $\hat{\theta}$ από τη μέση τιμή της δειγματικής κατανομής της εκτίμησης

$$V(\hat{\theta}) = E(\hat{\theta} - E(\hat{\theta}))^2.$$

Συνήθως όμως αντί της διασποράς, $V(\hat{\theta})$, χρησιμοποιούμε το **μέσο τετραγωνικό σφάλμα** (mean square error) μιας εκτιμήτριας το οποίο ορίζεται ως η μέση τιμή των τετραγώνων των αποκλίσεων των τιμών της $\hat{\theta}$ από την πραγματική τιμή της προς εκτίμηση παραμέτρου θ , δηλαδή

$$MT\Sigma(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

για την οποία αποδεικνύεται ότι ισχύει

$$MT\Sigma(\hat{\theta}) = V(\hat{\theta}) + b^2(\hat{\theta}).$$

Προφανώς όταν η $\hat{\theta}$ είναι αμερόληπτη ($b(\hat{\theta}) = 0$) το μέσο μετρωγωνικό σφάλμα ισούται με τη διασπορά.

Με τον όρο αξιοπιστία καθορίζεται ο βαθμός εξασφάλισης του κατά πόσο το λαμβανόμενο αποτέλεσμα από την εκτέλεση ενός test ή τη συμπλήρωση ενός ερωτηματολογίου αντιστοιχεί στην πραγματικότητα. Αναφερόμαστε δηλαδή σε κάποιο μέτρο που μας εκφράζει πόσο κοντά είναι κάθε παρατήρηση (μέτρηση) στην αντίστοιχη μέση τιμή σε ένα επαναλαμβανόμενο αριθμό δοκιμών.

Οι μετρήσεις αξιοπιστίας χωρίζονται γενικά σε δύο κατηγορίες: μετρήσεις ευστάθειας (stability measures) και ισοδυναμίας (equivalence measures).

Η απάντηση ενός ατόμου σε ένα θέμα, ή σύνολο θεμάτων μπορεί να διαφέρει από καιρού σε καιρό εξαιτίας διαφόρων αιτιών (αναποφάσιστος - κούραση - αρρώστεια κ.λπ.). Όλα αυτά αποτελούν σφάλματα της μέτρησης και συνεπώς επηρεάζουν την αξιοπιστία των θεμάτων. Η αξιοπιστία μιας μέτρησης εκτιμάται με τη συσχέτιση αυτής με την πάροδο του χρόνου (τώρα - παλαιότερα) και ονομάζεται μέτρηση ευστάθειας ή εξέταση - επανεξέταση αξιοπιστίας (test - retest). Όταν για παράδειγμα περιγράφουμε ένα ερωτηματολόγιο ως “αξιόπιστο” εννοούμε ότι μπορούμε να στηριχθούμε σε αυτό για να έχουμε τις “ίδιες” απαντήσεις όταν ερωτηθούν τα “ίδια” άτομα σε διαφορετικά μέρη ή άλλα άτομα στο “ίδιο” μέρος. Δηλαδή εάν η συμπλήρωση ενός ερωτηματολογίου γινότανε “κάτω από παρόμοιες συνθήκες” αναμένονται “ταυτόσημα αποτελέσματα”. Ενώ ιστορικά η αξιοπιστία εκτιμάται κυρίως χρησιμοποιώντας μετρήσεις ευστάθειας, μια διαφορετική προσέγγιση δίνεται μέσω της χρήσης μετρήσεων ισοδυναμίας. Παράλληλοι όροι επεξεργάζονται την ίδια χρονική στιγμή και συσχετίζονται για να εκτιμήσουν την αξιοπιστία σε αυτού του είδους τις μετρήσεις.

Καταστρώνοντας λοιπόν ένα δειγματοληπτικό σχέδιο επιδιώκουμε να έχουμε τελικά αξιόπιστες εκτιμήσεις για τα εκτιμώμενα πληθυσμιακά χαρακτηριστικά. Εάν υποθέσουμε ότι δεν υπάρχουν σφάλματα μέτρησης στην έρευνα, τότε η αξιοπιστία μιας εκτιμήτριας εκφράζεται συναρτήσει της διασποράς της εκτιμήτριας. **Όσο μικρότερη είναι η διασπορά της εκτιμήτριας τόσο μεγαλύτερη θα είναι η αξιοπιστία της εκτίμησης.**

Τέλος, η εγκυρότητα (validity) ενός τέστ ή ερωτηματολογίου είναι ένα μέτρο που δείχνει τον βαθμό με τον οποίο μετριέται κάτι κατά τη διάρκεια της έρευνας, δηλαδή πόσο κοντά είναι η μέτρηση στην τιμή - στόχο. Αν μετράει ότι επροτίθετο να μετρήσει. Η εγκυρότητα δηλαδή μιας εκτιμήτριας είναι συνυφασμένη με την μεροληψία της εκτιμήτριας. Υποθέτοντας πάλι ότι δεν υπάρχουν σφάλματα μέτρησης στην έρευνα, **όσο μικρότερη είναι η μεροληψία τόσο μεγαλύτερη θα είναι και η εγκυρότητα της εκτίμησης.**

Συνδυάζοντας την αξιοπιστία και την εγκυρότητα έχουμε την **ακρίβεια** (accuracy) μιας εκτιμήτριας, η οποία θα είναι τώρα συνυφασμένη με το μέσο τετραγωνικό σφάλμα. **Όσο μικρότερο είναι το μέσο τετραγωνικό σφάλμα τόσο μεγαλύτερη θα είναι η ακρίβεια της εκτίμησης.**

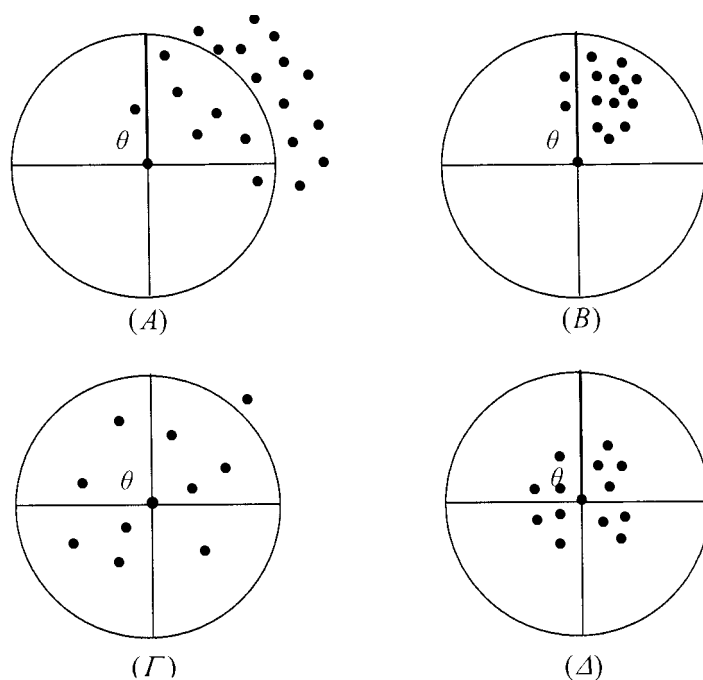
Παράδειγμα 1.2 Ας υποθέσουμε ότι με 4 διαφορετικά όπλα (A, B, Γ, Δ) “στοχεύουμε” ένα συγκεκριμένο αριθμό φορών κατά ενός “στόχου” θ , με τα παρακάτω αποτελέσματα.

Παρατηρούμε ότι με το όπλο (Δ) έχουμε βολές πολύ κοντά στο στόχο, επομένως μπορεί να θεωρηθεί ότι έχει μεγάλη αξιοπιστία, μικρή μεροληψία, (ισοδύναμα “μεγάλη” αμεροληψία) και υψηλή ακρίβεια (εγκυρότητα). Με το όπλο (B) έχουμε περίπου την ίδια εικόνα σημείων αλλά εκτός στόχου. Επομένως το όπλο (B) έχει μεγάλη αξιοπιστία, μεγάλη μεροληψία και μικρή εγκυρότητα. Οι “εικόνες” των όπλων A και Γ είναι περίπου οι ίδιες, του Γ όμως είναι γύρω από τον στόχο. Επομένως το όπλο (A) έχει μικρή αξιοπιστία, μεγάλη μεροληψία και χαμηλή εγκυρότητα, ενώ το όπλο (Γ) έχει μικρή αξιοπιστία, μικρή μεροληψία και μεγάλη εγκυρότητα.

Προφανώς στα ίδια συμπεράσματα θα καταλήγαμε για την αμεροληψία, αξιοπιστία ή εγκυρότητα τεσσάρων ερωτηματολογίων (tests) αντί “όπλων”, όπου θ θα είναι η πραγματική τιμή (πληθυσμιακή τιμή) του χαρακτηριστικού που θέλουμε να εκτιμήσουμε και με τελίτσες ($\bullet$) είναι οι αντίστοιχες εκτιμώμενες τιμές σε διαδοχικές επαναλήψεις - δοκιμές.

Περίληπτικά, οι σχέσεις μεταξύ μεροληψίας - εγκυρότητας και διασποράς - αξιοπιστίας για τις 4 περιπτώσεις δίνονται στον παρακάτω πίνακα.

Περίπτωση	Μεροληψία	Διασπορά	Αξιοπιστία	Εγκυρότητα
A	Μεγάλη	Μεγάλη	Μικρή	Μικρή
B	Μεγάλη	Μικρή	Μεγάλη	Μικρή
Γ	Μικρή	Μεγάλη	Μικρή	Μεγάλη
Δ	Μικρή	Μικρή	Μεγάλη	Μεγάλη



1.10 Δειγματοληψία και Στατιστική - Σύντομη Ανασκόπηση

Οι δημοσκοπήσεις και γενικότερα οι δειγματοληπτικές έρευνες είναι αλληλένδετες με τη στατιστική θεωρία. Έχουν αναπτυχθεί και εξελιχθεί με τις θεωρητικές και πρακτικές έρευνες των κορυφαίων στατιστικών. Στην ανάπτυξη της δειγματοληπτικής θεωρίας και πρακτικής συνέβαλαν οι κορυφαίοι επιστήμονες που έπαιζαν πρωτεύοντα ρόλο και στην ανάπτυξη της στατιστικής θεωρίας. Όπως αναφέρουν οι Hansen and Madow (1976) “οι καλές δειγματοληπτικές έρευνες είναι **ακριβές** και συχνά απαιτούν χρόνο”. Εάν, όμως, για το σχεδιασμό και τη διεξαγωγή μιας δειγματοληπτικής έρευνας ληφθούν όλες εκείνες οι σύνθετες αποφάσεις, όπως για

- τον καθορισμό του υπό διερεύνηση πληθυσμού
- την ύπαρξη δειγματοληπτικού σχεδίου
- τον προσδιορισμό του μεγέθους δείγματος
- το πρόβλημα των μη αποκρίσεων

- το πρόβλημα των ελλειπουσών τιμών
- τον υπολογισμό δειγματοληπτικών σφαλμάτων
- την αντιμετώπιση των μη δειγματοληπτικών σφαλμάτων

τότε οι δειγματοληπτικές έρευνες θα είναι και **ακριβείς**.

Η θεωρία των δειγματοληπτικών ερευνών είναι μαθηματική και αποτελεί μέρος της θεωρητικής μαθηματικής στατιστικής. Η πρακτική της όμως, δηλαδή ο σχεδιασμός και η εφαρμογή της, αναπτύσσονται ως αναπόσπαστο μέρος της συνολικής ερευνητικής διαδικασίας. Παρά τη θυελλώδη διαμάχη στα μέσα της δεκαετίας του 1950 σχετικά με τη θεμελίωση της συμπερασματολογίας στη δειγματοληψία στο πλαίσιο της μαθηματικής στατιστικής, για τη θεωρητική ανάπτυξη και εξέλιξη της η δειγματοληπτική θεωρία βασίστηκε στην πρακτική της. Σ' αυτό συνέβαλαν καθοριστικά αρκετοί κορυφαίοι στατιστικοί οι οποίοι είχαν την ικανότητα να διατυπώσουν και να αναπτύξουν ένα θεωρητικό πλαίσιο για την πρακτική της, βασιζόμενοι στην προσωπική ερευνητική τους εμπειρία.

Όπως προκύπτει από υπάρχουσες αναφορές υπήρχε μεγάλη διάσταση απόψεων μεταξύ των "φανατικών" υποστηρικτών της ολικής απογραφής και της δειγματοληψίας. Όπως επίσης και στη συνέχεια για την "αντιπροσωπευτική δειγματοληψία" και τη "δειγματοληψία πιθανοτήτων".

Πρώτα δείγματα χρήσης δειγματοληπτικών τεχνικών έχουμε από τους:

- Caspar Neumann, ευαγγελιστής εφημέριος και μανιώδης ερασιτέχνης δημογράφος. Χρησιμοποίησε καταλόγους γεννήσεων και κηδεών το 1687 για να αντικρούσει τους ισχυρισμούς των αστρολόγων ότι ορισμένες ηλικίες στη ζωή των ανθρώπων είναι ιδιαίτερα επικίνδυνες από την επίδραση της σελήνης στην υγεία. Για το σκοπό αυτό αποτάθηκε στον Leibniz το 1689 και στη συνέχεια η Royal Society ανέθεσε στον Edmond Halley ο οποίος το 1693 δημοσίευσε την εργασία του περί "θνησιμότητας της ανθρωπότητας" βασισμένη στις παρατηρήσεις του Neumann.
- Η "μέθοδος του Laplace" ο οποίος πρότεινε το 1786 και εφάρμοσε το 1802 τον προσδιορισμό των γεννήσεων στη Γαλλία βασιζόμενος σε απογραφή μερικών προσεκτικά επιλεγμένων κοινοτήτων της Γαλλίας και όχι όλης της χώρας.
- Kiaer, A. N. (1838-1919). Σ' αυτόν οφείλεται ο όρος "αντιπροσωπευτική δειγματοληψία" στην ομιλία του στο ISI στη Βέρνη το 1895 η οποία αντιμετώπισε στη συνέχεια πολλές κριτικές και αντιδράσεις για τη χρησιμότητά της.

- Bowley, A. L. (1869-1957). Καθηγητής στο LSE ήταν ο πρώτος που έδωσε το 1906 μια θεωρία συμπερασματολογίας για τις δειγματοληπτικές έρευνες, στον προεδρικό του λόγο προς το Τμήμα Οικονομικής Επιστήμης και Στατιστικής της Βρετανικής Εταιρείας για την Προαγωγή της Επιστήμης.
- Jerzy Neyman (1894-1981). Στις 19 Ιουνίου του 1934 παρουσίασε στο Royal Statistical Society το θεμελιώδες άρθρο του “Περί των δύο διαφορετικών πλευρών της αντιπροσωπευτικής μεθόδου: η μέθοδος της στρωματοποιημένης δειγματοληψίας και η μέθοδος της σκόπιμης επιλογής” όπου αξιολογεί συγκριτικά τις δύο μεθόδους βάσει θεωρητικών και πρακτικών τους ιδιοτήτων. Για πρώτη φορά ορίζεται η έννοια των διαστημάτων εμπιστοσύνης. Ενώ μέχρι τότε η τυχαία επιλογή απαιτούσε ίσες πιθανότητες επιλογής, ο Neyman ήταν ο πρώτος που πρότεινε ότι οι πιθανότητες επιλογής των διαφόρων μονάδων στο δείγμα μπορεί να είναι και άνισες. Έτσι υπολόγισε τον βέλτιστο καταμερισμό για τη στρωματοποιημένη δειγματοληψία, υποστηρίζοντας όμως ότι στις περισσότερες περιπτώσεις πρέπει να χρησιμοποιείται αναλογικός καταμερισμός (Neyman allocation). Στη συνέχεια, σε σειρά άρθρων του αναπτύσσονται η κατά συστάδες και η διπλή δειγματοληψία. Παρουσίασε ισχυρά επιχειρήματα για την απόρριψη της σκόπιμης δειγματοληψίας.
- Fisher, Sir R. A. (1890-1962). Ενώ τα παραδείγματα των Kiaer, Bowley και Neyman αναφέρονται στη μελέτη μεγάλων πληθυσμών ατόμων, στη δεκαετία του 1920 ο Fisher απασχολείται με δειγματοληπτικές έρευνες σε πληθυσμούς αγροκτημάτων, γεωργικών μονάδων, πειραματικών καλλιεργειών κ.ά.
- Το έργο και οι ιδέες του Fisher οδήγησαν στη διεξαγωγή πολλών πειραμάτων από τον ίδιο και στη συνέχεια από τους Yates, Cochran και Wishart στον Πειραματικό Σταθμό του Rothamsted. Έτσι η ανάπτυξη και η εφαρμογή της δειγματοληπτικής θεωρίας διατηρήθηκαν σε υψηλά επίπεδα σ’ όλη τη διάρκεια του πολέμου.

Ενώ αυτά συνέβαιναν στην Αγγλία και γενικότερα στην Ευρώπη, παράλληλη ήταν και η ανάπτυξη της δειγματοληπτικής θεωρίας και στις ΗΠΑ. Μεγάλη όμως επίδραση στους στατιστικούς των ΗΠΑ είχε η πρόσκληση του Neyman από τον Deming το 1937 στη Ουάσινγκτον και οι διαλέξεις που έδωσε και στη συνέχεια ως καθηγητής και διευθυντής του Στατιστικού Εργαστηρίου στο Πανεπιστήμιο Berkeley της Καλιφόρνιας.

Τέλος σημαντική ήταν και η συμβολή του Ινδικού Στατιστικού Ινστιτούτου που ιδρύθηκε το 1931 από τον Mahalanobis. Αυτός εισήγαγε τα πανομοιότυπα (replicated) δείγματα για τις εκτιμήσεις διασποράς και τη μέτρηση των

δειγματοληπτικών σφαλμάτων. Αποτελέσματα αυτών των εργασιών του δημοσιεύτηκαν το 1946 στο JRSS.

Η ανάπτυξη και η εφαρμογή των δειγματοληπτικών τεχνικών συνεχίστηκε στα ίδια βήματα μέχρι τις αρχές της δεκαετίας του 1960. Στο διάστημα αυτό γράφτηκαν και τα σπουδαιότερα βιβλία δειγματοληψίας που παραμένουν μέχρι και σήμερα βασικά και απαραίτητα εγχειρίδια για τη διδασκαλία της δειγματοληπτικής θεωρίας.

Στα μέσα της δεκαετίας του 1960 ξέσπασε μια θυελλώδης θεωρητική διαμάχη με αντικείμενο τη θεμελίωση της δειγματοληπτικής θεωρίας. Το αποτέλεσμα ήταν να τεθεί μια διαχωριστική γραμμή ανάμεσα στους στατιστικούς που ανέπτυσαν θεωρία από τη θεωρία και τους στατιστικούς που ανέπτυσαν θεωρία από την πρακτική της, δηλαδή ανάμεσα στους μαθηματικούς στατιστικούς και τους θεωρητικούς στατιστικούς.

Το 1994, ο Smith στην προσπάθεια του να διερευνήσει την πιθανότητα συμφιλίωσης των δύο προσεγγίσεων κατέληγε: “Το συνολικό μου συμπέρασμα είναι ότι δεν υπάρχει μοναδικό παράδειγμα για τη στατιστική συμπερασματολογία και ότι διαφορετικές κατηγορίες προβλημάτων απαιτούν διαφορετικές λύσεις. Αντί να αναζητούμε την ενότητα, πρέπει να επικεντρωθούμε στον προσδιορισμό των διαφορών και να απολαύσουμε την ποικιλία του αντικειμένου μας. Η πλήρης συμφιλίωση ούτε εφικτή είναι ούτε επιθυμητή. Ζήτη η διαφορά”.

Ενώ στην Ευρώπη, Ασία και Αμερική η ανάπτυξη της δειγματοληψίας και η διεξαγωγή δειγματοληπτικών ερευνών συνεχιζόταν σταθερά, η δραστηριότητα που αναπτύχθηκε στην Ελλάδα για τη διεξαγωγή ερευνών μεγάλης έκτασης ήταν περιορισμένη. Η Γενική Στατιστική Υπηρεσία Ελλάδος (ΓΣΥΕ) ήταν σχεδόν αποδιοργανωμένη μέχρι το 1953 οπότε ιδρύθηκε η ΕΣΥΕ. Στα πανεπιστήμια όχι μόνο η θεωρία δειγματοληψίας δεν διδασκότανε αλλά ούτε και η στατιστική συμπερασματολογία γενικότερα.

Η Συμφωνία της Βάρκιζας στις 12/2/1945 προέβλεπε τη διεξαγωγή δημοψηφίσματος για την προκήρυξη εκλογών και την αποστολή παρατηρητών από τις συμμαχικές δυνάμεις για τον έλεγχο της γνησιότητας της λαϊκής ετυμηγορίας. Έτσι συγκροτήθηκαν οι δύο Συμμαχικές Αποστολές:

AMFOGE 1 (Allied Mission for Observing Greek Elections), και

AMFOGE 2 (Allied Mission to Observe the Revision of Greek Electoral Lists).

Η πρώτη αποστολή συγκροτήθηκε από Αμερικανούς, Βρετανούς και Γάλλους παρατηρητές για να παρακολουθήσουν τις εκλογές της 31ης Μαρτίου 1946. Σκοπός της αποστολής ήταν η διενέργεια δειγματοληπτικών ερευνών και η συλλογή πληροφοριών για τις συνθήκες προετοιμασίας και διεξαγωγής των εκλογών. Η σπουδαιότητα που έδινε η επιτροπή για τον καθορισμό του πληθυ-

σμού, την πληρότητα των εκλογικών καταλόγων και εν γένει τη διεξαγωγή των ερευνών φαίνεται και από τη συμμετοχή στην αποστολή αυτή των κορυφαίων στατιστικών Raymond Jessen, William Edwards Deming, Oscar Kempthorne, Joseph Daly και Jerzy Neyman (παρόλο που ο τελευταίος “αποπέμφθηκε λόγω ανυπακοής”).

Η δεύτερη αποστολή ήρθε στην Ελλάδα το καλοκαίρι του 1946, αποτελείτο από τους ίδιους παραπάνω στατιστικούς εκτός του Daly που αντικαταστάθηκε από τον Blythe. Ο Neyman παρόλο που προτάθηκε και για τη δεύτερη αποστολή δεν αποδέχθηκε. Σκοπός της AMFOGE 2 ήταν να εξεταστούν η επάρκεια, η ακρίβεια και η πληρότητα των εγγραφών στους εκλογικούς καταλόγους που ετοιμαζόνταν για το δημοψήφισμα της 1ης Σεπτεμβρίου.

Από την Ελληνική πλευρά συμμετείχαν οι Ιωάννης Μιχαλόπουλος, τ. διευθυντής της ΓΣΥΕ και Διευθυντής του Υπουργείου Εθνικής Οικονομίας, ο Δημήτριος Καλιτσούνακης, καθηγητής Πολιτικής Οικονομίας στην ΑΣΟΕΕ και ο Άγγελος Αγγελόπουλος, καθηγητής Δημόσιας Οικονομίας στο Πανεπιστήμιο Αθηνών.

Αποτελέσματα των δύο συμμαχικών αποστολών δημοσιεύθηκαν σε αρκετά περιοδικά και βιβλία (American Sociological Review, JASA και στα βιβλία δειγματοληψίας των Jessen, Yates, Kish, κ.ά.) και αποτέλεσαν πράγματι σημαντική συμβολή στην ανάπτυξη της δειγματοληπτικής έρευνας.

Εκτενής ιστορική και κριτική παρουσίαση της δειγματοληπτικής θεωρίας βρίσκεται στα εξαιρετικά άρθρα κορυφαίων στατιστικών στα καλύτερα στατιστικά περιοδικά JRSS Ser A, JASA, The Annals of Math. Statistics, κ.ά. Για πλήρη αναφορά και για εκτενέστερη βιβλιογραφία παραπέμπουμε στο βιβλίο της Καίτης Μιχαλοπούλου “Στην Αυτοκρατορία των Ενδείξεων - η Ιστορία της Δειγματοληπτικής Πρακτικής στην Ελλάδα”, εκδόσεις Παπαζήση, 2004.

1.11 Ασκήσεις

1. Αναφέρατε περιληπτικά γιατί είναι απαραίτητο να προηγηθεί μια βιβλιογραφική έρευνα για τους ορισμούς που θα χρησιμοποιηθούν σε μία δειγματοληπτική έρευνα.
2. Ερωτηματολόγια που συμπληρώνονται από τους ερωτώμενους πρέπει να ελέγχονται πριν τη χρήση τους για την εγκυρότητα και την αξιοπιστία τους. Σχολιάστε περιληπτικά τι σημαίνει αυτό.
3. Δώστε μερικά παραδείγματα δειγματοληψίας χωρίς πιθανότητα (nonrandom sampling), δειγματοληψίας ευκολίας (convenient sampling), δειγματοληψίας σκοπιμότητας (purposive sampling), εθελοντικής δειγματο-

ληψίας (voluntary sampling), δειγματοληψίας ποσοστώσεων (quota sampling)) και δειγματοληψίας κατά πιθανότητα (probability sampling).

4. Γιατί είναι απαραίτητος ένας προέλεγχος ενός ερωτηματολογίου;
5. Ενδιαφέρεστε να κάνετε μία έρευνα αγοράς για την απορρόφηση των πτυχιούχων Μαθηματικών “στους ευρύτερους” χώρους υγείας. Δώστε τα διάφορα βήματα που θα ακολουθήσετε για το σχεδιασμό της δειγματοληψίας που θα κάνετε.
6. Εξηγείστε αναλυτικά με τι ασχολείται η δειγματοληψία και ποια είναι η χρησιμότητά της.
7. Περιγράψετε τις διάφορες μεθόδους συλλογής δεδομένων για μία δειγματοληπτική έρευνα αναφέροντας και τα πλεονεκτήματα - μειονεκτήματά τους.
8. α) Για τη διεξαγωγή μιας δειγματοληπτικής έρευνας δώστε τους ορισμούς των παρακάτω εννοιών:
 i) πληθυσμός, ii) πλαίσιο, iii) δείγμα, iv) στοιχειώδης μονάδα, v) δειγματοληπτική μονάδα, vi) μεταβλητή (χαρακτηριστικό), vii) δειγματοληπτικά σφάλματα, viii) μη δειγματοληπτικά σφάλματα.
 β) Ενδιαφέρεστε να σχεδιάσετε μια δειγματοληπτική έρευνα για να εκτιμήσετε το μέσο αριθμό ωρών που αφιερώνονται καθημερινά για γυμναστική από τους ενηλίκους μιας περιοχής. Ποιες μπορεί να είναι οι έννοιες (i) - (vi) του παραπάνω ερωτήματος (α);
9. Ας υποθέσουμε ότι ένας εκπαιδευτικός φορέας (π.χ. το Υπουργείο Παιδείας) σας αναθέτει να διεξάγετε μια δειγματοληπτική έρευνα για να διαπιστώσετε εάν η διαφορά των 3 μονάδων μεταξύ προφορικού και γραπτού βαθμού στις εξετάσεις της Β' και Γ' Λυκείου σύμφωνα με την εκπαιδευτική μεταρρύθμιση είναι ένα αντικειμενικό μέτρο και ότι δεν ευνοεί σημαντικά κάποιους μαθητές έναντι άλλων μαθητών.
 Σχεδιάστε τη δειγματοληπτική έρευνα που θα κάνετε περιγράφοντας σε συντομία τα διάφορα βήματα, όπως για παράδειγμα: καθορισμός πλαισίου, επιλογή μεθόδου, μέγεθος δείγματος, διαμόρφωση ερωτηματολογίου, κ.λπ.
10. Αυτές τις μέρες οι δημοσκοπήσεις και τα περί δημοσκοπήσεων βρίσκονται στην ημερήσια διάταξη. Πως σχολιάζετε εσείς τα περί της αξίας ή μη των δημοσκοπήσεων;

11. Αναφέρατε μερικά πλεονεκτήματα και μειονεκτήματα της δειγματοληψίας και της απογραφής.
12. Ποια από τα παρακάτω αποτελούν πλεονεκτήματα ή μειονεκτήματα της δειγματοληψίας;
 - α) Οικονομία χρόνου
 - β) Χαμηλό κόστος
 - γ) Μεγαλύτερη ακρίβεια
 - δ) Μελέτη ενός σπάνιου χαρακτηριστικού
 - ε) Μεγαλύτερη ευχέρεια εφαρμογής
 - στ) Άγνωστος πληθυσμός.
13. Διακρίνετε μεταξύ τυχαίας (random sampling) και μη τυχαίας (nonrandom sampling) δειγματοληψίας. Δώστε μερικά παραδείγματα.
14. Για να αυξήσει το βαθμό αξιοπιστίας των αποτελεσμάτων μιας έρευνας τί πρέπει να προσέξει ιδιαίτερα ο ερευνητής;
15. Από τις παρακάτω ερωτήσεις ποιές θα μπορούσαν να διαμορφωθούν καλύτερα;
 - Πώς θα περιγράφατε την κατάσταση της υγείας σας;(Τον τελευταίο μήνα πώς θα περιγράφατε την κατάσταση της υγείας σας;)
 - Συμφωνείτε να γίνεται αξιολόγηση του διδακτικού προσωπικού από τους φοιτητές και το Υπουργείο Παιδείας;
(Όχι δύο ερωτήσεις σε μία).
 - Πόσες φορές τον προηγούμενο χρόνο είχατε επισκεφτεί ή μιλήσει με γιατρό για την υγεία σας;

(Το χρονικό διάστημα για το οποίο ενδιαφερόμαστε είναι αρκετά μεγάλο και ασφαλώς ο/η ερωτώμενος/η δεν μπορεί να θυμάται επακριβώς. Σε τέτοιες περιπτώσεις ο χρόνος αναφοράς είναι μία βδομάδα ή ένας μήνας. Χρησιμοποιούμε σαν μονάδα αναφοράς το χρόνο όταν ρωτάμε για σπάνια αλλά σημαντικά γεγονότα, όπως η γέννηση ενός παιδιού, η αγορά ενός σπιτιού κ.λπ.

 - Επίσκεψη σημαίνει στο ιατρείο, νοσοκομείο, σπίτι του;
 - Συμπεριλαμβάνονται οι τηλεφωνικές συμβουλές;

- Συμπεριλαμβάνονται και οι ασκούμενοι γιατροί, οι ψυχίατροι, αν η επίσκεψη αφορούσε π.χ. ένα εμβόλιο που έγινε από τη νοσοκόμα;

- Πήρατε πρωϊνό χθές;

(Πρέπει να διευκρινιστεί μέχρι ποιά ώρα θεωρείται ότι κάποιος παίρνει “πρωϊνό”. Συνήθως “πρωϊνό” είναι ένα γεύμα που τρώγεται πριν τις 10 π.μ.

- Αν κάποιος ήπια ένα σκέτο καφέ θεωρείται ότι πήρε πρωϊνό;

Καλύτερα εδώ είναι να ρωτηθεί τι κατανάλωσαν πριν τη 10 π.μ. και η κωδικοποίηση να γίνει μετά από τον ερευνητή, ή να δηλωθεί μία λίστα όπως αυγό, κρέας, γάλα, καφές, χυμός, ψημένο ψωμί, σιτηρά, φρούτα κ.λπ.).

- Ποιο τηλεοπτικό σταθμό παρακολουθείτε περισσότερο;

(Υπάρχει ασάφεια ως προς το “περισσότερο”. Περισσότερη ώρα ή περισσότερες φορές;)

16. Ποια από τις παρακάτω ερωτήσεις είναι κατά τη γνώμη σας καταλληλότερη;

- 1α: Πότε πηγαίνετε συνήθως στη δουλειά σας;
- 1β: Τι ώρα συνήθως φεύγετε από το σπίτι για τη δουλειά σας;
- 2α: Εάν έχετε ένα ή περισσότερα αυτοκίνητα παρακαλώ καταγράψτε το έτος κυκλοφορίας και τη μάρκα καθενός ξεκινώντας από το νεότερο.
- 2β: Παρακαλώ καταγράψτε το έτος κυκλοφορίας και τη μάρκα καθενός από τα αυτοκίνητά σας.
- 3α: Πως αντιδράτε όταν θέλετε κάτι και ο/η σύντροφός σας θέλει κάτι άλλο και οι δύο ενδιαφέρεστε πάρα πολύ για την προτίμησή σας;
- 3β: Πως ρυθμίζετε τις διαφωνίες σας με τον/την σύντρόφό σας όταν και οι δύο ενδιαφέρεστε πάρα πολύ για την προτίμησή σας;
- 4α: Δεν βλέπετε κάποιο κίνδυνο για τη δουλειά σας με το νέο νομοσχέδιο;
- 4β: Βλέπετε κανένα κίνδυνο για τη δουλειά σας με το νέο νομοσχέδιο;

17. Με ποιά από τα παρακάτω ασχολείται **λιγότερο** μια δειγματοληπτική έρευνα;

- α. Εξαγωγή περιγραφικών μέτρων
 - β. Εξαγωγή έγκυρων και αξιόπιστων εκτιμήσεων
 - γ. Περιγραφή διαφόρων χαρακτηριστικών του πληθυσμού
 - δ. Έλεγχος υποθέσεων για διάφορες παραμέτρους του πληθυσμού.
18. Ποιο από τα παρακάτω περιγράφει **καλύτερα** εκτιμήσεις που λαμβάνονται από μία απογραφή;
- α. Υπάρχουν τόσο δειγματοληπτικά όσο και μη-δειγματοληπτικά σφάλματα.
 - β. Υπάρχουν δειγματοληπτικά αλλά όχι μη-δειγματοληπτικά σφάλματα.
 - γ. Υπάρχουν μη-δειγματοληπτικά αλλά όχι δειγματοληπτικά σφάλματα.
 - δ. Δεν υπάρχουν ούτε δειγματοληπτικά ούτε μη-δειγματοληπτικά σφάλματα.
19. Αναφέρατε περιληπτικά τα σχετικά πλεονεκτήματα της χρήσης προσωπικών συνεντεύξεων, τηλεφωνικών συνεντεύξεων και ταχυδρομικής αποστολής ερωτηματολογίων ως μεθόδους συλλογής δεδομένων για κάθε μία από τις παρακάτω περιπτώσεις:
- α. Ένας υπεύθυνος τηλεοπτικού καναλιού ενδιαφέρεται να εκτιμήσει το ποσοστό των τηλεθεατών στη χώρα που παρακολουθούν το κανάλι σε μία δεδομένη χρονική περίοδο.
 - β. Ο εκδότης μιας εφημερίδας ενδιαφέρεται να ερευνήσει τη στάση του κοινού έναντι του είδους των νέων που καλύπτονται από την εφημερίδα του.
 - γ. Ο δήμαρχος μιας πόλης ενδιαφέρεται να προσδιορίσει τις αντιδράσεις των κατοίκων μιας περιοχής σχετικά με τις μονοδρομήσεις των δρόμων της περιοχής που σχεδιάζονται να γίνουν.
 - δ. Το Υπουργείο Παιδείας ενδιαφέρεται να εκτιμήσει τις αντιδράσεις των φοιτητών σε ένα επικείμενο νομοσχέδιο για την παιδεία.
20. Εξηγείστε περιληπτικά τι είναι παράμετρος και τι εκτιμήτρια μιας παραμέτρου. Ποιες ιδιότητες πρέπει να έχει μια “καλή” εκτιμήτρια και τι σημαίνει καθεμία από τις ιδιότητες που θα αναφέρετε;

21. Σημειώστε ένα Σ για τη ΣΩΣΤΗ ή ένα Λ για τη ΛΑΘΟΣ για καθεμιά από τις παρακάτω προτάσεις.

Α/Α	ΠΡΟΤΑΣΗ	ΣΩΣΤΗ (Σ) ΛΑΘΟΣ (Λ)
1	Ο αντικειμενικός σκοπός μιας δειγματοληπτικής έρευνας είναι η εκτίμηση ορισμένων παραμέτρων του πληθυσμού από την πληροφορία που περιέχεται στο δείγμα.	
2	Για τους σκοπούς μιας δειγματοληπτικής έρευνας πρέπει να παίρνουμε πάντα μεγάλο δείγμα	
3	Τα δειγματοληπτικά σφάλματα οδηγούν σε μεγάλα διαστήματα εμπιστοσύνης.	
4	Η δειγματοληπτική μονάδα δεν είναι τίποτε άλλο από το απλό στοιχείο ενός πληθυσμού.	
5	Η μείωση των μη δειγματοληπτικών σφαλμάτων απαιτεί μεγάλα δείγματα.	
6	Μια δοκιμαστική έρευνα πρέπει να γίνεται στο 5-10% του τελικού δείγματος για να διαπιστώνεται αν κρατά το ενδιαφέρον μέχρι τέλους.	
7	Η υγεία των ανθρώπων κατατάσσεται στις διακριτές μεταβλητές	
8	Η εκλογή μικρού δείγματος υστερεί πάντα σε σχέση με ένα μεγαλύτερο δείγμα.	
9	Το δειγματοληπτικό πλαίσιο καθορίζει πόσο καλά καλύπτεται ο πληθυσμός, επηρεάζει τη μέθοδο συλλογής της πληροφορίας και επιδρά στην αποτελεσματικότητα με την οποία λαμβάνεται ένα δείγμα.	
10	Το επίπεδο σημαντικότητας α και το σφάλμα d είναι αντιστρόφως ανάλογα με το μέγεθος δείγματος n .	
11	Ο ακριβής προσδιορισμός του δειγματοληπτικού πλαισίου περιορίζει τα μη δειγματοληπτικά σφάλματα	
12	Όσο μεγαλύτερη είναι η διασπορά του πληθυσμού τόσο μικρότερο μέγεθος δείγματος θα πάρουμε.	
13	Με μια μέθοδο δειγματοληψίας A επιλέγουμε δείγμα μεγέθους n_1 και με μια μέθοδο δειγματοληψίας B επιλέγουμε δείγμα μεγέθους n_2 , κατά πολύ μεγαλύτερο από το n_1 . Τότε η μέθοδος B θα μας δώσει σίγουρα καλύτερα αποτελέσματα από την A .	
14	Οι δημογραφικές ερωτήσεις όπως ηλικία, εισόδημα, επάγγελμα κ.λπ. συνήθως τοποθετούνται στην αρχή του ερωτηματολογίου.	
15	Ένα πλεονέκτημα των ανοικτών ερωτήσεων είναι ότι ο ερωτώμενος μπορεί να δώσει μια απάντηση σύμφωνα με τη δική του κρίση.	
16	Η μελέτη σχετικών προηγούμενων εργασιών είναι απαραίτητο στάδιο μιας δειγματοληπτικής έρευνας.	
17	Η ολική απογραφή πλεονεκτεί πάντα μιας δειγματοληψίας.	

A/A	ΠΡΟΤΑΣΗ	ΣΩΣΤΗ (Σ) ΛΑΘΟΣ (Λ)
18	Ένα ερωτηματολόγιο με μεγάλη αξιοπιστία είναι πάντοτε προτιμότερο κάποιου άλλου.	
19	Είναι προτιμότερο να χρησιμοποιούνται ερωτηματολόγια για τα οποία έχει ελεγχθεί η εγκυρότητα και η αξιοπιστία τους	
20	Απαραίτητο συστατικό μιας δειγματοληπτικής έρευνας είναι ο ποιοτικός έλεγχος σ' όλα τα στάδιά της.	
21	Η αξιοπιστία είναι αντιστρόφως ανάλογη της διασποράς	
22	Η εγκυρότητα είναι ανάλογη της μεροληψίας	
23	Διεξάγοντας μια πιλοτική έρευνα προσδοκούμε στην ελάττωση των μη-δειγματοληπτικών σφαλμάτων.	
24	Στην απογραφή έχουμε μη-δειγματοληπτικά σφάλματα αλλά όχι δειγματοληπτικά	
25	Στην απογραφή έχουμε δειγματοληπτικά σφάλματα αλλά όχι μη-δειγματοληπτικά	
26	Σε μία δειγματοληψία εκτός από το δειγματοληπτικό σχέδιο, μεταξύ των (α) στατιστικών αναφορών, (β) διαδικασιών εκτίμησης και (γ) κόστη εκτιμήσεων μας ενδιαφέρει κυρίως το (β).	
27	Τα μη-συστηματικά σφάλματα ονομάζονται και μεροληψίες	
28	Τα σφάλματα τυχαίου τύπου ονομάζονται και συστηματικά σφάλματα.	
29	Η δυσκολία εκλογής αντιπροσωπευτικού δείγματος είναι ένα από τα μειονεκτήματα της ταχυδρομικής συλλογής των δεδομένων.	
30	Για να αυξήσουμε το ποσοστό απαντήσεων μπορούμε να αναφέρουμε στα άτομα που επιλέγονται στο δείγμα ότι θα συμμετέχουν σε κλήρωση κάποιου δώρου.	