

Κεφάλαιο 1 - ΕΙΣΑΓΩΓΗ

Σύνοψη

Στο κεφάλαιο αυτό, περιγράφονται τα κυριότερα στάδια μιας δειγματοληπτικής έρευνας και δίνεται μια εισαγωγή της Θεωρίας Δειγματοληψίας. Οι δειγματοληπτικές έρευνες βασίζονται σε δεδομένα που καταγράφονται για ένα υποσύνολο του πληθυσμού, το δείγμα (*sample*), και έχουν ως στόχο την εξαγωγή συμπερασμάτων που αφορούν άγνωστες παραμέτρους του πληθυσμού. Τόσο η χρήση, όσο και η θεματολογία τους, είναι ευρύτατη. Οι μέθοδοι δειγματοληψίας χωρίζονται σε μεθόδους πιθανότητας (*probability sampling*) και μεθόδους μη-πιθανότητας (*non-probability sampling*). Για τα δείγματα που επιλέγονται σύμφωνα με μια μέθοδο πιθανότητας, η εξαγωγή συμπερασμάτων και η επέκτασή τους στον πληθυσμό είναι εφικτή κάνοντας χρήση της Στατιστικής συμπερασματολογίας. Για τα δεδομένα που συλλέγονται σύμφωνα με ένα δείγμα μη-πιθανότητας, είναι εφικτό να εφαρμοστεί περιγραφική στατιστική ώστε να παρουσιαστούν τα αποτελέσματα της έρευνας, αλλά δεν μπορεί να γίνει επέκταση των αποτελεσμάτων για τον πληθυσμό. Συνοπτικά, η Θεωρία Δειγματοληψίας (*Sampling Theory*) είναι το πεδίο της στατιστικής που περιλαμβάνει τη μεθοδολογία (α) για την επιλογή ενός δείγματος από έναν μεγαλύτερο πληθυσμό και (β) για την εξαγωγή αξιόπιστων συμπερασμάτων για τα ερωτήματα της έρευνας που θα έχουν ισχύ για τον πληθυσμό. Οι πιο γνωστές από τις μεθόδους δειγματοληψίας που βασίζονται στις πιθανότητες είναι η απλή τυχαία (*simple random*), η συστηματική (*systematic*), η στρωματοποιημένη (*stratified*) και η δειγματοληψία κατά ομάδες (*cluster*).

1.1. Δειγματοληπτική Έρευνα

Με την ανάπτυξη των υπολογιστών, του διαδικτύου (*internet*) και των μέσων μαζικής ενημέρωσης και επικοινωνίας, η πρόσβαση σε δεδομένα είναι πολύ πιο εύκολη απ'ότι παλιότερα. Τα δεδομένα αυτά προέρχονται ή αφορούν ζητήματα γύρω από κάθε τομέα της προσωπικής, κοινωνικής ή επαγγελματικής ζωής μας. Ο μεγάλος όγκος των δεδομένων καθιστά αναγκαία την έκφρασή τους με ποσοτικά στοιχεία και γραφήματα, με σκοπό την καλύτερη οργάνωση, παρουσίαση και κατανόησή τους, και την εξαγωγή συμπερασμάτων. Η ανάγκη για την ύπαρξη ποσοτικών συμπερασμάτων για διάφορα θέματα της καθημερινότητας είναι το αντικείμενο των δειγματοληπτικών ερευνών.

Χαρακτηριστικό παράδειγμα δειγματοληπτικών ερευνών είναι οι δημοσκοπήσεις (*opinion polls*), όπου επιλέγεται ένα δείγμα από ένα σύνολο ανθρώπων και όσοι το αποτελούν καλούνται να απαντήσουν εκφράζοντας την προσωπική τους άποψη ή προτίμηση γύρω από ένα θέμα, π.χ. εάν είναι ικανοποιημένοι από το σύστημα υγείας, σε τι βαθμό προτιμούν ένα καταναλωτικό προϊόν, ποιο πολιτικό κόμμα τους εκφράζει περισσότερο κτλ. Οι δημοσκοπήσεις χρησιμοποιούνται με μεγάλη συχνότητα σε έρευνες μαρκετινγκ, διαφήμισης και έρευνες πολιτικού χαρακτήρα.

Οι δειγματοληπτικές έρευνες μπορεί να αφορούν και πληθυσμούς που δεν έχουν ως μέλη τους ανθρώπους. Π.χ. μια έρευνα που έχει ως στόχο να απαντήσει στο ερώτημα πόσα στρέμματα καλλιεργήσιμης γης μιας χώρας καλλιεργούνται με σιτάρι, ή ποιο ποσοστό των προϊόντων από μια γραμμή παραγωγής είναι ελαττωματικά.

Άλλα παραδείγματα δειγματοληπτικών ερευνών είναι οι έρευνες που διεξάγονται σε επίπεδο χωρών ή μικρότερων διοικητικών μονάδων, με σκοπό την εκτίμηση π.χ. του ποσοστού ανεργίας των κατοίκων και πώς αυτό μεταβάλλεται ανά μήνα, του ποσοστού των κατοίκων που ολοκληρώνουν την εκπαίδευση σε κάθε εκπαιδευτική βαθμίδα (βασική, υποχρεωτική, ανώτερη, ανώτατη), του μεγέθους του πληθωρισμού, του ποσοστού των νοικοκυριών της χώρας που δεν έχουν κεντρική θέρμανση κτλ.

Οι δειγματοληπτικές έρευνες που θα μας απασχολήσουν στη συνέχεια αφορούν έρευνες που έχουν ως στόχο τη μελέτη του πληθυσμού σε μια συγκεκριμένη χρονική στιγμή (χωρίς διάρκεια) και από τις οποίες ενδιαφερόμαστε να βγάλουμε συμπεράσματα για άγνωστες ποσότητες του πληθυσμού. Οι έρευνες αυτές λέγονται *cross-sectional* και ιδανικά αποτελούν ένα στιγμιότυπο (*snapshot*) του πληθυσμού τη στιγμή της έρευνας.

Αντίθετα, οι *cohort* δειγματοληπτικές έρευνες είναι έρευνες οι οποίες στοχεύουν στη μελέτη ενός χαρακτηριστικού του πληθυσμού στη διάρκεια του χρόνου, και κατά πόσο ή σε ποιο βαθμό επηρεάζεται το υπό μελέτη χαρακτηριστικό από άλλους παράγοντες.

Η ανάγκη για γρήγορη εξαγωγή αποτελεσμάτων και για σύγκριση αυτών σε τακτά χρονικά διαστήματα καθιστά τη χρήση των δειγματοληπτικών ερευνών επιβεβλημένη σε σχέση με την εναλλακτική επιλογή που είναι η απογραφή του πληθυσμού. Στην παράγραφο που ακολουθεί, δίνονται αναλυτικά οι ορισμοί του πληθυσμού, της απογραφής και του δείγματος (βλ. [Groves et al, 2009](#)).

1.2. Πληθυσμός – Δείγμα

Όλα τα παραδείγματα που δόθηκαν στην προηγούμενη παράγραφο έχουν ως κοινό γνώρισμα ότι αναφέρονται σε ένα σύνολο με συγκεκριμένο πλήθος μελών, π.χ. οι κάτοικοι μιας χώρας, το σύνολο των νοικοκυριών της χώρας κτλ. Το σύνολο των μελών που είναι πεπερασμένου, και όχι άπειρου, πλήθους ονομάζεται *πεπερασμένος πληθυσμός (finite population)*. Σε όλα τα κεφάλαια που ακολουθούν, υποθέτουμε ότι ο πληθυσμός είναι πεπερασμένος.

Ορισμός 1.1

Αντικειμενικός πληθυσμός (target population) είναι το σύνολο των μελών για το οποίο ενδιαφερόμαστε να εξαγάγουμε κάποια συμπεράσματα.

Παράδειγμα 1.1

Για τη δειγματοληπτική έρευνα που στοχεύει στον υπολογισμό του ποσοστού ανεργίας των κατοίκων μιας χώρας, ο αντικειμενικός πληθυσμός είναι όλοι οι κάτοικοι της χώρας (άνω των 18 ετών, οι οποίοι μπορούν να εργαστούν).

Ορισμός 1.2

Η έρευνα που βασίζεται στη μελέτη όλων των στοιχείων ενός πληθυσμού, ονομάζεται *απογραφή (census)*.

Στην περίπτωση της απογραφής, το ερώτημα (ή τα ερωτήματα) της έρευνας τίθεται σε κάθε μέλος του πληθυσμού. Τα στοιχεία που θα συγκεντρωθούν με το τέλος της έρευνας χρησιμοποιούνται για τον υπολογισμό συγκεντρωτικών ποσοτήτων (π.χ. μέση τιμή) του πληθυσμού. Τα δεδομένα μιας απογραφής προσφέρουν τη δυνατότητα έγκυρων αποτελεσμάτων για τα ερωτήματα της έρευνας, δηλαδή αποτελεσμάτων που βασίζονται στην πλήρη πληροφορία για τον πληθυσμό και δεν εμπεριέχουν ‘στατιστικό λάθος’ (statistical error). Η έκφραση ‘στατιστικό λάθος’ ή ‘αβεβαιότητα’, που αναφέρεται συχνά σε δημοσιεύσεις αποτελεσμάτων δειγματοληπτικών ερευνών, θα διευκρινιστεί επακριβώς στη συνέχεια του κεφαλαίου. Παρόλο το πλεονέκτημα της απογραφής ως προς την ακρίβεια των αποτελεσμάτων, είναι προφανές ότι η διεξαγωγή μιας τέτοιας διαδικασίας είναι χρονοβόρα υπόθεση, με μεγάλο κόστος σε σχεδιασμό, οργάνωση και υλοποίηση. Επίσης, λόγω αυτών των χαρακτηριστικών, η απογραφή δεν είναι εφικτό να πραγματοποιείται σε πυκνά χρονικά διαστήματα, και κατά συνέπεια δεν είμαστε σε θέση να διαπιστώνουμε κατά πόσον οι ποσότητες του πληθυσμού που μας ενδιαφέρουν μεταβάλλονται στην πάροδο του χρόνου. Π.χ. εάν ενδιαφερόμαστε για το ποσοστό ανεργίας σε μια μεγάλη πόλη ή χώρα, ο υπολογισμός του βάσει απογραφής κάθε 2 μήνες για παράδειγμα, θα ήταν ανέφικτος. Επίσης, ένα άλλο πρακτικό πρόβλημα είναι ότι για την υλοποίηση μιας απογραφής είναι απαραίτητο να υπάρχει μια λίστα που περιλαμβάνει ένα προς ένα όλα τα μέλη του πληθυσμού.

Εναλλακτικά της απογραφής, η έρευνα βασίζεται σε ένα υποσύνολο του πληθυσμού που λέγεται *δείγμα (sample)* και κάνουμε λόγο για *δειγματοληπτική έρευνα (sampling survey)*. Κατά τη διεξαγωγή μιας δειγματοληπτικής έρευνας, τα μέλη του πληθυσμού που αντιστοιχούν στο δείγμα εντοπίζονται και καλούνται να απαντήσουν στο ερώτημα ή τα ερωτήματα της έρευνας. Τα στοιχεία που συλλέγονται κατά τη διεξαγωγή μιας δειγματοληπτικής έρευνας χρησιμοποιούνται και αυτά, όπως και της απογραφής, για τον υπολογισμό συγκεντρωτικών ποσοτήτων, όχι όμως του πληθυσμού, αλλά των αντίστοιχων ποσοτήτων στο δείγμα. Η διατύπωση οποιουδήποτε συμπεράσματος για ποσότητες του πληθυσμού, με βάση τις αντίστοιχες του δείγματος, απαιτεί την ανάπτυξη κατάλληλης μεθοδολογίας. Λόγω της επιλογής μέρους μόνο του πληθυσμού,

τα αποτελέσματα μιας δειγματοληπτικής έρευνας εμπεριέχουν μια ‘τυχειότητα’ ή ‘αβεβαιότητα’, σε αντίθεση με την απογραφή.

Στα προφανή πλεονεκτήματα της δειγματοληπτικής έρευνας περιλαμβάνονται το μικρό κόστος, η σύντομη διάρκεια της συλλογής δεδομένων και η επαναληψιμότητα της διαδικασίας, δηλ. τα τρία στοιχεία που αποτελούν τα βασικά μειονεκτήματα της απογραφής. Γενικά, κατά τον σχεδιασμό και τη διεξαγωγή μιας δειγματοληπτικής έρευνας, γίνεται προσπάθεια να επιτευχθεί μείωση του κόστους και του χρόνου διεξαγωγής της έρευνας, με ταυτόχρονο έλεγχο της αβεβαιότητας που θα εμπεριέχεται στα αποτελέσματα. Η ισορροπία μεταξύ αυτών των δύο συνιστωσών εξαρτάται από πολλούς παράγοντες και θα είναι ένα από τα αντικείμενα που θα μας απασχολήσουν σε επόμενα κεφάλαια.

Τέλος, μπορούμε να προσθέσουμε ένα ακόμα πλεονέκτημα της δειγματοληπτικής έρευνας. Επειδή ο αριθμός των συμμετεχόντων στην έρευνα είναι μικρότερος σε σύγκριση με την απογραφή, ο όγκος των δεδομένων που έχουμε να χειριστούμε είναι μικρότερος, με αποτέλεσμα να ελαχιστοποιούνται σφάλματα τα οποία σχετίζονται με την καταγραφή, την κωδικοποίηση και την επεξεργασία των δεδομένων, και στα οποία είναι πιθανόν να υποπέσουμε. Ένα άλλο πρακτικό πλεονέκτημα της δειγματοληπτικής έρευνας έναντι της απογραφής, είναι ότι, σε ορισμένες περιπτώσεις, η καταγραφή μιας μέτρησης για μια μονάδα του πληθυσμού συνεπάγεται αλλοίωση ή καταστροφή της μονάδας. Παράδειγμα, η διάρκεια ζωής ενός ηλεκτρικού λαμπτήρα από μια γραμμή παραγωγής λαμπτήρων, ή ο έλεγχος της ικανοποιητικής ανάπτυξης των ριζών ενός φυτού που καλλιεργείται σε ένα φυτώριο. Στις περιπτώσεις αυτές, η δειγματοληψία αποτελεί μοναδική λύση και όχι απλώς εναλλακτική της απογραφής.

Απ’ την άλλη πλευρά, ενώ ο όγκος των δεδομένων για μια δειγματοληψία είναι μικρότερος σε σύγκριση με την απογραφή, η υιοθέτηση της δειγματοληπτικής έρευνας ως επιλογής είναι πιο απαιτητική σε ό,τι αφορά το μαθηματικό και στατιστικό υπόβαθρο που είναι απαραίτητο προκειμένου (i) να επιλεγούν οι μονάδες του πληθυσμού που θα αποτελούν το δείγμα και (ii) να εξαχθούν αξιόπιστα αποτελέσματα για τα ερωτήματα της έρευνας, που θα έχουν ισχύ για τον πληθυσμό. Τα δύο τελευταία αυτά χαρακτηριστικά της δειγματοληπτικής έρευνας αποτελούν και το αντικείμενο της *Θεωρίας Δειγματοληψίας (Sampling Theory)*.

Δίνουμε στη συνέχεια έναν ακόμα ορισμό για τον πληθυσμό εκτός του αντικειμενικού, ο οποίος συνδέεται με την επιλογή της δειγματοληπτικής έρευνας έναντι της απογραφής για την εξαγωγή των αποτελεσμάτων.

Ορισμός 1.3

Υπό μελέτη πληθυσμός ή δειγματοληπτικό πλαίσιο (sampling frame) είναι το σύνολο των μελών στα οποία θα βασιστούμε προκειμένου να υλοποιήσουμε τη δειγματοληπτική έρευνα.

Παράδειγμα 1.2

Για τη δειγματοληπτική έρευνα του Παραδείγματος [1.1](#), που στοχεύει στον υπολογισμό του ποσοστού ανεργίας των κατοίκων μιας χώρας, εάν η έρευνα διεξαχθεί επιλέγοντας τυχαία τηλεφωνικούς αριθμούς από τους τηλεφωνικούς καταλόγους της χώρας, τότε ο υπό μελέτη πληθυσμός είναι όλοι οι κάτοικοι της χώρας (άνω των 18 ετών, οι οποίοι μπορούν να εργαστούν) και οι οποίοι έχουν τηλέφωνο.

Παράδειγμα 1.3

Εάν ενδιαφερόμαστε για τον υπολογισμό του ποσοστού των παιδιών μιας πόλης, ηλικίας 6 ετών, τα οποία δεν έχουν κάνει τα βασικά εμβόλια, τότε:

- (i) Στην περίπτωση που η έρευνα γίνει εντοπίζοντας πρώτα όλα τα παιδιά ηλικίας 6 ετών της πόλης, τα οποία είναι εγγεγραμμένα στα δημοτολόγια της πόλης, και επιλέγοντας ένα δείγμα, τότε ο *υπό μελέτη πληθυσμός* είναι τα παιδιά 6 ετών της πόλης που είναι εγγεγραμμένα στα δημοτολόγια.
- (ii) Εάν η έρευνα διεξαχθεί επιλέγοντας ένα υποσύνολο Δημοτικών Σχολείων της πόλης, και τα παιδιά της Πρώτης Τάξης των Σχολείων αυτών συμπεριληφθούν στο δείγμα, ο *υπό μελέτη πληθυσμός* είναι τα παιδιά 6 ετών της πόλης που φοιτούν σε ένα από τα Δημοτικά Σχολεία.

Ο αντικειμενικός πληθυσμός για το παράδειγμά μας είναι τα παιδιά της πόλης ηλικίας 6 ετών.

Γενικότερα, ενώ ο αντικειμενικός πληθυσμός είναι μονοσήμαντα ορισμένος, ο υπό μελέτη πληθυσμός ενδέχεται να έχει πολλούς ορισμούς, ακόμα κι αν αναφερόμαστε στην ίδια έρευνα. Αυτό συμβαίνει γιατί ο υπό μελέτη πληθυσμός συνδέεται με τον τρόπο υλοποίησης της έρευνας και κατά συνέπεια θα διαφέρει, δηλ. θα ορίζεται διαφορετικά, ανάλογα με τη μέθοδο επιλογής του δείγματος και με τη μέθοδο πρόσβασης στα μέλη του συνόλου για την επικοινωνία και την καταγραφή της μέτρησης.

Ο υπό μελέτη πληθυσμός είναι στη γενική περίπτωση ένα υποσύνολο του αντικειμενικού πληθυσμού. Στο Παράδειγμα 1.2, ο υπό μελέτη πληθυσμός δεν συμπεριλαμβάνει τους κατοίκους της χώρας που δεν έχουν τηλέφωνο. Στο Παράδειγμα 1.3(ii), ο υπό μελέτη πληθυσμός δεν συμπεριλαμβάνει τα παιδιά της πόλης που δεν φοιτούν σε κάποιο από τα σχολεία. Προφανώς, όσο πιο κοντά είναι ο υπό μελέτη πληθυσμός στον αντικειμενικό πληθυσμό, τόσο καλύτερο είναι το δειγματοληπτικό πλαίσιο, γιατί αποφεύγονται τα σφάλματα μη-κάλυψης (*non-coverage error*).

Ο αντικειμενικός πληθυσμός πολλές φορές αναφέρεται και ως *θεωρητικός πληθυσμός* (*theoretical population*), γιατί είναι το σύνολο των μελών στο οποίο θεωρητικά στοχεύουμε. Αντίθετα, ο υπό μελέτη πληθυσμός, όπως είδαμε, είναι ο πληθυσμός που χρησιμοποιούμε στην πράξη.

1.3. Δείγματα πιθανότητας και δείγματα μη-πιθανότητας

Ανάλογα με τον μηχανισμό επιλογής των μονάδων του πληθυσμού στο δείγμα, ο οποίος ονομάζεται και *δειγματοληπτικό σχέδιο* (*sampling design*), τα δείγματα χωρίζονται αρχικά σε δύο μεγάλες κατηγορίες. Τα *δείγματα πιθανότητας* (*probability samples*) και τα *δείγματα μη-πιθανότητας* (*non-probability samples*). Αντίστοιχα, η δειγματοληψία ονομάζεται δειγματοληψία πιθανότητας και δειγματοληψία μη-πιθανότητας.

1.3.1 Δείγματα Πιθανότητας

Ένα δείγμα λέγεται δείγμα πιθανότητας όταν η κάθε μονάδα του πληθυσμού έχει μια πιθανότητα, συγκεκριμένη και μη-μηδενική, να συμπεριληφθεί στο δείγμα. Η πιθανότητα αυτή είναι προκαθορισμένη πριν από την επιλογή του δείγματος.

Συνεπώς, σύμφωνα με τα δείγματα πιθανότητας, η μέθοδος δειγματοληψίας δεν αποκλείει κάποιες μονάδες του πληθυσμού από το ενδεχόμενο να είναι μέρη του δείγματος. Επιπλέον, η προκαθορισμένη, και κατά συνέπεια γνωστή, πιθανότητα επιλογής της κάθε μονάδας του πληθυσμού συνεπάγεται ή εγγυάται ότι στη διαδικασία επιλογής του δείγματος υπεισέρχεται ο παράγοντας της τυχαιότητας.

Παράδειγμα 1.4

Έστω ότι ενδιαφερόμαστε για τη γνώμη (θετική ή αρνητική) 200 δημοτών για μια απόφαση του δημοτικού συμβουλίου της πόλης τους.

Εάν επιλέξουμε τυχαία 200 άτομα από το δημοτολόγιο της πόλης και επικοινωνήσουμε στη συνέχεια μαζί τους, π.χ. μέσω ταχυδρομείου, τότε το δείγμα είναι ένα δείγμα πιθανότητας. Το κάθε μέλος του υπό μελέτη πληθυσμού, ο κάθε δημότης στην προκειμένη περίπτωση, έχει μια προκαθορισμένη πιθανότητα να ανήκει στο δείγμα. Ειδικότερα, για το παράδειγμά μας, οι πιθανότητες επιλογής είναι ίσες μεταξύ των μελών του πληθυσμού, και συγκεκριμένα ίσες με 1 προς το πλήθος των εγγεγραμμένων δημοτών.

Παράδειγμα 1.5

Έστω ότι το θέμα της έρευνας αφορά την ίδρυση παιδικών σταθμών, και συνεπώς ενδιαφέρει περισσότερο τους γονείς που έχουν παιδιά σε προσχολική ηλικία. Θα μπορούσαμε να επιλέξουμε 150 άτομα από τη συγκεκριμένη αυτή κατηγορία και 50 άτομα από όλους τους υπόλοιπους. Το δείγμα αυτό παραμένει ένα δείγμα πιθανότητας, όπου οι πιθανότητες είναι άνισες μεταξύ των δύο κατηγοριών των δημοτών της πόλης, αλλά είναι γνωστές και προκαθορισμένες. Πιο συγκεκριμένα, οι πιθανότητες εδώ προσδιορίζονται από τα πλήθη, έστω N_A και N_B , των δημοτών που ανήκουν στην πρώτη και τη δεύτερη κατηγορία, αντίστοιχα. Η

πιθανότητα που θα έχει ο κάθε δημότης με παιδί σε προσχολική ηλικία να επιλεγεί για το δείγμα θα είναι $1/N_A$ ενώ η πιθανότητα για τους δημότες χωρίς παιδιά αυτής της ηλικίας θα είναι $1/N_B$.

Και στα δύο παραδείγματα υπάρχει ο παράγοντας της τυχαιότητας.

Εάν, προκειμένου να διεξάγουμε την ίδια έρευνα με αυτή του Παραδείγματος [1.4](#), αποφασίζαμε να συμπληρώσουμε το δείγμα των 200 δημοτών ρωτώντας τους πρώτους 200 που θα περάσουν από ένα συγκεκριμένο σημείο ενός πολυσύχναστου δρόμου, το δείγμα δεν είναι δείγμα πιθανότητας. Αυτό συμβαίνει γιατί σύμφωνα με την παραπάνω διαδικασία επιλογής, ο κάθε δημότης δεν έχει μια θετική, μη-μηδενική πιθανότητα επιλογής. Οι δημότες που δεν θα περάσουν από το σημείο αυτό του δρόμου θα έχουν μηδενική πιθανότητα επιλογής.

Οι πιο γνωστές μέθοδοι δειγματοληψίας που ανήκουν στην κατηγορία των δειγματοληψιών πιθανότητας είναι οι:

- Απλή τυχαία,
- Στρωματοποιημένη,
- Συστηματική,
- Δείγματα με άνισες πιθανότητες - Δειγματοληψία ανάλογα του μεγέθους,
- Κατά ομάδες ή σε πολλά στάδια,
- Συνδυασμός των παραπάνω.

Στα κεφάλαια που ακολουθούν, γίνεται μελέτη της κάθε μιας από τις παραπάνω μεθόδους, αναλυτικά.

Το μεγάλο πλεονέκτημα των δειγμάτων πιθανότητας είναι ότι επειδή η διαδικασία επιλογής του δείγματος διέπεται από τους νόμους των πιθανοτήτων, είναι εφικτή η επέκταση των όποιων συμπερασμάτων από το δείγμα στον πληθυσμό, με τη βοήθεια της στατιστικής συμπερασματολογίας. Ορισμένα στοιχεία της περιοχής της στατιστικής συμπερασματολογίας, θα δώσουμε στην παράγραφο 1.5.

1.3.2 Δείγματα Μη-Πιθανότητας

Τα δείγματα μη-πιθανότητας είναι στον αντίποδα των δειγμάτων πιθανότητας. Η μέθοδος επιλογής των μονάδων του δείγματος δεν διέπεται από τους νόμους της πιθανότητας, αλλά βασίζεται σε κριτήρια όπως η ευκολία, η εύκολη πρόσβαση, η διαθεσιμότητα, ο σύντομος χρόνος συλλογής των δεδομένων κτλ. Τα κριτήρια αυτά δεν εξασφαλίζουν μια θετική και προκαθορισμένη πιθανότητα επιλογής στο δείγμα για το κάθε μέλος του πληθυσμού. Αντίθετα η επιλογή ή μη των μελών του πληθυσμού στο δείγμα γίνεται με βεβαιότητα.

Τα βασικότερα είδη δειγματοληψίας που ανήκουν στην κατηγορία των δειγμάτων μη-πιθανότητας είναι:

Δείγματα Ευκολίας (accessibility or convenience samples)

Είναι τα δείγματα όπου οι δειγματοληπτικές μονάδες επιλέγονται από τον πληθυσμό με κριτήριο την ευκολία, και όχι την τυχαιότητα ή την επιδίωξη της αντιπροσωπευτικότητας του πληθυσμού.

Ένα παράδειγμα δείγματος ευκολίας είναι όταν το δείγμα συλλέγεται μέσω μιας εφαρμογής στο διαδίκτυο όπου όποιος θέλει να συμμετέχει στην έρευνα και να απαντήσει στην ερώτηση μπορεί να το κάνει μόνος του. Επίσης, στο Παράδειγμα [1.4](#), το δείγμα που συλλέγεται ρωτώντας τους 200 πρώτους που θα περάσουν από το σταθερό σημείο όπου βρίσκεται ο συνεντευκτής, είναι ένα δείγμα ευκολίας. Και στις δύο περιπτώσεις, υπάρχει έντονος ο κίνδυνος των μεροληπτικών αποτελεσμάτων. Στη μεν πρώτη περίπτωση, η συμμετοχή στην έρευνα γίνεται με πρωτοβουλία του μέλους του πληθυσμού, και όχι με τυχαιότητα, άρα η συμμετοχή ή όχι εξαρτάται από το αν το άτομο έχει θετική ή όχι άποψη για το θέμα της έρευνας. Επίσης, αποκλείονται όλοι όσοι, είτε δεν έχουν πρόσβαση στο διαδίκτυο, είτε δεν είδαν τη συγκεκριμένη εφαρμογή της έρευνας.

Στη δεύτερη περίπτωση, η μεροληψία είναι υπερκτική γιατί, για παράδειγμα, αποκλείονται τα μέλη του πληθυσμού που λόγω ηλικίας, ή επειδή εργάζονται, δεν διέρχονται από τον συγκεκριμένο δρόμο την ώρα της έρευνας.

Γενικότερα, στα δείγματα ευκολίας δεν είναι εφικτό να εξαχθούν αποτελέσματα τα οποία στη συνέχεια θα γενικευτούν για τον πληθυσμό, γιατί το δείγμα δεν είναι αντιπροσωπευτικό. Η ίδια αυτή παρατήρηση ισχύει για όλα τα δείγματα μη-πιθανότητας συνολικά.

Δείγματα Κρίσης (Judgmental samples)

Στα δείγματα κρίσης ο ερευνητής επιλέγει τις μονάδες του πληθυσμού με βάση την προσωπική του κρίση, ή την εμπειρία του από προηγούμενες έρευνες με παρόμοιο θέμα στο ίδιο σύνολο πληθυσμού. Για παράδειγμα, σε δημοσκοπήσεις με στόχο πολιτικές έρευνες, αποτελέσματα εκλογών κτλ., ο ερευνητής μπορεί να επιλέξει το δείγμα του συμπεριλαμβάνοντας με βεβαιότητα περιοχές του πληθυσμού που έχουν ιδιαίτερα χαρακτηριστικά, π.χ. στις πιο πρόσφατες εκλογές, τα εκλογικά αποτελέσματα των περιοχών αυτών ήταν πολύ κοντά στα τελικά αποτελέσματα όλης της επικράτειας. Οι περιοχές αυτές αποκαλούνται «περιοχές βαρόμετρο» και επιλέγονται με βεβαιότητα στο δείγμα, γιατί βάσει της εμπειρίας από προηγούμενες εκλογές θεωρούνται αντιπροσωπευτικές.

Δείγματα Χιονοστιβάδας (Snowball sampling)

Στα δείγματα χιονοστιβάδας, το δείγμα γίνεται προσβάσιμο στον ερευνητή μέσω ενός μικρού αρχικού συνόλου δείγματος που είναι διαθέσιμο σε εκείνον. Η κάθε μιά δειγματοληπτική μονάδα του αρχικού δείγματος προσφέρει τα στοιχεία, και άρα την πρόσβαση, σε ένα σύνολο από άλλα μέλη του πληθυσμού, τα οποία συμπεριλαμβάνονται στο δείγμα, και τα οποία με τη σειρά τους προσφέρουν πρόσβαση σε ένα άλλο σύνολο κοκ. Ο στατιστικός αναλυτής, συνεπώς, αποκτά το δείγμα μέσω των αρχικών εκπροσώπων, χωρίς προσπάθεια εντοπισμού και χωρίς ανάγκη να διαθέτει στοιχεία για τον πληθυσμό (λίστα μελών, πλήθος κτλ).

Δείγματα με προκαθορισμένα ποσοστά (Quota sampling)

Σύμφωνα με αυτή τη μέθοδο δειγματοληψίας, ο στατιστικός αναλυτής συμπεριλαμβάνει στο δείγμα του μονάδες του πληθυσμού έτσι, ώστε το τελικό δείγμα να έχει εκπροσώπους από κάθε κατηγορία του πληθυσμού, και μάλιστα με αναλογία ίση με εκείνη που ισχύει για τον πληθυσμό. Οι κατηγορίες ορίζονται συνήθως με βάση ένα δημογραφικό κριτήριο, π.χ φύλο, ηλικιακές ομάδες κτλ.

Έστω, για παράδειγμα, ότι ενδιαφερόμαστε να συλλέξουμε ένα δείγμα 100 φοιτητών ενός τμήματος ΑΕΙ, και ότι θεωρούμε το φύλο των φοιτητών σημαντικό παράγοντα για την έρευνα. Αν είναι γνωστό ότι 40% των φοιτητών είναι γυναίκες και 60% είναι άντρες, και το δείγμα συλλεγεί επιλέγοντας τις πρώτες 40 φοιτήτριες του τμήματος που θα εντοπίσουμε σε μια επίσκεψή μας στο προαύλιο της Σχολής, και αντίστοιχα τους πρώτους 60 άντρες φοιτητές, τότε το δείγμα θα είναι ένα δείγμα με προκαθορισμένα ποσοστά.

Η δειγματοληψία με προκαθορισμένα ποσοστά έχει ομοιότητες με τη στρωματοποιημένη, και ειδικότερα την αναλογική στρωματοποιημένη, αλλά δεν αποτελεί δειγματοληψία πιθανότητας όπως η στρωματοποιημένη. Μια περιγραφή και σύγκριση της δειγματοληψίας με προκαθορισμένα ποσοστά με τη στρωματοποιημένη δίνεται στην παράγραφο 4.4.

Για περαιτέρω ανάλυση των δειγματοληπτικών σχεδίων μη-πιθανότητας, παραπέμπουμε μεταξύ άλλων στο κεφάλαιο 2 του βιβλίου [Henry](#) (1990).

Στα κεφάλαια που ακολουθούν, θα ασχοληθούμε αποκλειστικά με τα δείγματα πιθανότητας. Τα δείγματα μη-πιθανότητας δεν επιτρέπουν γενίκευση για τον πληθυσμό, ούτε την ανάπτυξη της στατιστικής μεθοδολογίας η οποία παρέχει ακριβή στοιχεία και ιδιότητες για τα αποτελέσματα του δείγματος, με τη βοήθεια των οποίων μπορούμε στη συνέχεια να εξαγάγουμε ασφαλέστερα συμπεράσματα για τον πληθυσμό. Στην παράγραφο που

ακολουθεί, θα δοθεί ο απαραίτητος μαθηματικός συμβολισμός. Στη συνέχεια, δίνουμε μια συνοπτική επισκόπηση των κυριότερων στοιχείων από τη Στατιστική συμπερασματολογία.

1.4. Συμβολισμός

Υποθέτουμε ότι ο πληθυσμός για τον οποίο ενδιαφερόμαστε να εξαγάγουμε συμπεράσματα με τη βοήθεια ενός δείγματος είναι πεπερασμένος, και έστω ότι N είναι το πλήθος των μελών του πληθυσμού. Το πλήθος αυτό ονομάζεται και *μέγεθος του πληθυσμού (population size)*. Ανάλογα, συμβολίζουμε με n και το ονομάζουμε *μέγεθος δείγματος (sample size)* το πλήθος των μονάδων του πληθυσμού που επιλέγονται για το δείγμα.

Ορισμός 1.4

Χαρακτηριστικό ενός πληθυσμού (population characteristic) ονομάζεται το ερώτημα για τον πληθυσμό, στο οποίο μας ενδιαφέρει να δώσουμε μια απάντηση.

1.4.1 Πληθυσμιακά μεγέθη

Το χαρακτηριστικό του πληθυσμού συμβολίζεται συνήθως με ένα κεφαλαίο γράμμα, έστω Y . Το χαρακτηριστικό Y λαμβάνει, για κάθε μέλος του πληθυσμού, μια συγκεκριμένη τιμή, την απάντηση στο ερώτημα της έρευνας. Για παράδειγμα, εάν το χαρακτηριστικό του πληθυσμού για το οποίο επιθυμούμε να εξαγάγουμε συμπεράσματα είναι το εισόδημα, υπάρχει μια τιμή που αντιστοιχεί στο εισόδημα για το κάθε μέλος του πληθυσμού. Το χαρακτηριστικό Y μπορεί να θεωρηθεί ως μια *τυχαία μεταβλητή (τ.μ.) (random variable, (rv))* με δυνατές τιμές τις δυνατές τιμές του χαρακτηριστικού για τα μέλη του πληθυσμού.

Η τιμή του χαρακτηριστικού Y για το i μέλος του πληθυσμού συμβολίζεται με Y_i . Επίσης, το διάνυσμα με στοιχεία τα Y_i για όλα τα δυνατά μέλη του πληθυσμού, $\mathbf{U} = \{Y_1, Y_2, \dots, Y_N\}$, ονομάζεται διάνυσμα του πληθυσμού (population vector) για το χαρακτηριστικό Y . Είναι φανερό ότι το διάνυσμα \mathbf{U} είναι γνωστό μόνο μετά από μια απογραφή.

Αν Y είναι το υπό μελέτη χαρακτηριστικό του πληθυσμού, τότε οι ποσότητες για τις οποίες πιο συχνά ενδιαφερόμαστε να εξαγάγουμε συμπεράσματα είναι συνήθως:

- η *μέση τιμή (mean value)* του χαρακτηριστικού για τον πληθυσμό. Όταν ο υπό μελέτη πληθυσμός είναι πεπερασμένος, η μέση τιμή ορίζεται ως το άθροισμα των μετρήσεων του Y για όλα τα μέλη του πληθυσμού διαιρούμενο με το πλήθος τους. Εάν συμβολίσουμε τη μέση τιμή του Y για τον πληθυσμό με \bar{Y} , τότε:

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$$

- το *σύνολο (population total)* των τιμών του χαρακτηριστικού για τον πληθυσμό, είναι το άθροισμα των τιμών του χαρακτηριστικού Y_i για όλα τα μέλη του πληθυσμού και συμβολίζεται με Y_T . Θα είναι:

$$Y_T = \sum_{i=1}^N Y_i$$

Προκύπτει άμεσα από τους παραπάνω ορισμούς ότι $Y_T = N\bar{Y}$. Οι ποσότητες \bar{Y} και Y_T έχουν νόημα για ένα συνεχές χαρακτηριστικό Y ή, ισοδύναμα, για μια συνεχή τυχαία μεταβλητή. Για παράδειγμα, εάν Y είναι το εισόδημα των κατοίκων μιας χώρας, τότε ενδιαφερόμαστε να εκτιμήσουμε το μέσο εισόδημα των κατοίκων, δηλαδή το \bar{Y} ή το συνολικό εισόδημα των κατοίκων Y_T . Ανάλογα, εάν Y είναι η κατανάλωση νερού (σε λίτρα)

των νοικοκυριών μιας πόλης για έναν μήνα, τότε \bar{Y} είναι η μέση κατανάλωση νερού σε ένα μήνα για τα νοικοκυριά της πόλης και Y_T η συνολική κατανάλωση νερού (σε λίτρα) της πόλης για ένα μήνα.

- το ποσοστό (*percentage*) ενός χαρακτηριστικού στον πληθυσμό. Εάν το υπό μελέτη χαρακτηριστικό αντιστοιχεί σε μια κατηγορική με 2 ή περισσότερα επίπεδα, τότε το ενδιαφέρον της έρευνας επικεντρώνεται στην εξαγωγή συμπερασμάτων σχετικά με το ποσοστό των μελών του πληθυσμού που ανήκουν σε ένα από τα δυνατά επίπεδα του χαρακτηριστικού. Για παράδειγμα, εάν το χαρακτηριστικό Y αντιστοιχεί στο τύπο καπνιστή και λαμβάνει 3 δυνατές τιμές: 'μη-καπνιστής', 'μέτρια καπνιστής', 'έντονα καπνιστής', τότε ως θέμα μιας δειγματοληπτικής έρευνας μπορεί να είναι η εκτίμηση του ποσοστού των μελών του πληθυσμού που ανήκουν στην κατηγορία 'έντονα καπνιστής'.

Το ποσοστό συμβολίζεται με P και ορίζεται ως:

$$P = \frac{A}{N}$$

όπου A είναι το πλήθος των μελών του πληθυσμού που ανήκουν στην υπό μελέτη κατηγορία.

- Η ποσότητα A , δηλαδή το πλήθος των μελών του πληθυσμού που ανήκουν στην υπό μελέτη κατηγορία. Προφανώς ισχύει:

$$A = NP$$

Μια επιπλέον πολύ σημαντική ποσότητα του πληθυσμού είναι η διασπορά των τιμών του πληθυσμού γύρω από τη μέση τιμή για το χαρακτηριστικό, η οποία συμβολίζεται με S^2 και ορίζεται ως:

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2$$

Η διασπορά (*variance*) S^2 δίνει πληροφορία σχετικά με την ετερογένεια ή ομοιογένεια των μετρήσεων του χαρακτηριστικού στο σύνολο του πληθυσμού. Όσο μεγαλύτερη η διασπορά S^2 για έναν πληθυσμό, τόσο πιο ετερογενής είναι ο πληθυσμός. Όπως θα δούμε στη συνέχεια, ο ρόλος της διασποράς του πληθυσμού για το χαρακτηριστικό παίζει ουσιαστικό ρόλο σχεδόν σε κάθε στάδιο της έρευνας: στον σχεδιασμό και την επιλογή της δειγματοληπτικής μεθόδου, στον καθορισμό του μεγέθους του δείγματος, και βέβαια στην ανάλυση των δεδομένων και τον υπολογισμό των ιδιοτήτων των εκτιμητών.

Σύμφωνα με τον ορισμό των ποσοτήτων \bar{Y} , Y_T , P , A και S^2 είναι φανερό ότι οι ποσότητες αυτές είναι γνωστές, μόνο αν είναι γνωστές οι τιμές του χαρακτηριστικού για όλα τα μέλη του πληθυσμού. Για να τονιστεί το γεγονός ότι οι παραπάνω ποσότητες αναφέρονται στο σύνολο του πληθυσμού ονομάζονται και *πληθυσμιακές ποσότητες* ή *πληθυσμιακά μεγέθη* ή *παράμετροι του πληθυσμού* (*population quantities, population parameters*), πχ πληθυσμιακή μέση τιμή, πληθυσμιακό σύνολο κτλ. Γενικά, οι πληθυσμιακές ποσότητες είναι σταθερές ποσότητες αφού αναφέρονται σε απογραφικά στοιχεία και δεν εμπεριέχουν τυχαιότητα, είναι όμως συνήθως άγνωστες, και η εκτίμησή τους αποτελεί τον σκοπό της διενέργειας της έρευνας.

1.4.2 Δειγματικές ποσότητες

Ανάλογα με τον συμβολισμό και τους ορισμούς των πληθυσμιακών ποσοτήτων, μπορούν να οριστούν οι αντίστοιχες *δειγματικές ποσότητες* (*sample quantities*). Έστω ότι το μέγεθος του δείγματος είναι n . Για τα δείγματα πιθανότητας, οι n παρατηρήσεις του δείγματος είναι τυχαίες μεταβλητές, γιατί εμπεριέχουν τον παράγοντα της τυχαιότητας.

Στη βιβλιογραφία, έχει επικρατήσει να συμβολίζουμε το δείγμα ως s . Για το διάνυσμα των παρατηρήσεων του δείγματος, χρησιμοποιούνται δύο εναλλακτικοί συμβολισμοί:

$$(\alpha) s = \{y_1, y_2, \dots, y_n\}$$

Δηλαδή, διατηρούμε το ίδιο γράμμα που έχουμε δώσει για το χαρακτηριστικό του πληθυσμού, αλλά με μικρούς χαρακτήρες ώστε να διαχωρίζονται οι δειγματικές από τις πληθυσμιακές ποσότητες, και:

$$(\beta) s = \{X_1, X_2, \dots, X_n\}$$

Στον συμβολισμό αυτό, χρησιμοποιούμε ένα άλλο κεφαλαίο γράμμα για τις μετρήσεις του δείγματος σε σχέση με τις μετρήσεις του πληθυσμού.

Στα επόμενα κεφάλαια, για να τονίσουμε το γεγονός ότι οι μετρήσεις του δείγματος είναι τυχαίες μεταβλητές, τις οποίες συνήθως συμβολίζουμε με κεφαλαία γράμματα, υιοθετούμε τον (β) συμβολισμό για το δείγμα.

Η τυχαία μεταβλητή X_i ($i = 1, 2, \dots, n$) είναι η μεταβλητή που καταγράφει την i -οστή μέτρηση του δείγματος. Οι δυνατές τιμές της είναι το σύνολο $\{Y_1, Y_2, \dots, Y_N\}$, π.χ. $X_3 = Y_{20}$, σημαίνει ότι ως 3η μέτρηση του δείγματος επιλέχθηκε η 20ή μέτρηση του πληθυσμού. Προφανώς, σε κάθε επανάληψη του δείγματος, η τιμή του πληθυσμού που επιλέγεται για την πλήρωση της i θέσης του δείγματος μπορεί να διαφέρει.

Αν $\{X_1, X_2, \dots, X_n\}$ είναι ένα δείγμα μεγέθους n , τότε ορίζεται:

- η *δειγματική μέση τιμή* (*sample mean value*) η οποία συμβολίζεται με \bar{X} και υπολογίζεται ως:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

- η *δειγματική διασπορά* (*sample variance*) των τιμών του δείγματος η οποία συμβολίζεται με s^2 . Η δειγματική διασπορά ορίζεται ανάλογα με την πληθυσμιακή S^2 . Αναλυτικά:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Η διασπορά s^2 υπολογίζεται αριθμητικά μετά τη συλλογή και καταγραφή των μετρήσεων του δείγματος και η πληροφορία που προσφέρει είναι ανάλογη εκείνης της S^2 . Όσο μεγαλύτερη είναι η s^2 , τόσο μεγαλύτερη ετερογένεια παρατηρείται στις μετρήσεις του δείγματος, και αντίστροφα.

Οι τυχαίες μεταβλητές του δείγματος X_i έχουν χαρακτηριστικά τα οποία σχετίζονται τόσο με τον ίδιο τον πληθυσμό από τον οποίο έχει επιλεγεί το δείγμα, όσο και με τον τρόπο δειγματοληψίας. Η εξάρτηση της X_i από τον τρόπο δειγματοληψίας θα γίνει περισσότερο κατανοητή στην παράγραφο που ακολουθεί.

1.5. Στοιχεία από τη Στατιστική Συμπερασματολογία

Στην παρούσα παράγραφο, δίνονται επιγραμματικά στοιχεία από τη στατιστική συμπερασματολογία, με έμφαση στη διαφοροποίηση των συνήθων εννοιών όταν οι παρατηρήσεις του δείγματος προέρχονται από έναν πεπερασμένο πληθυσμό.

Υποθέτουμε ένα δείγμα πιθανότητας $s = \{X_1, X_2, \dots, X_n\}$ το οποίο έχει επιλεγεί από έναν πεπερασμένο πληθυσμό μεγέθους N . Έστω επίσης θ μια πληθυσμιακή ποσότητα ή παράμετρος (π.χ. \bar{Y} , Y_T κτλ.) η οποία είναι άγνωστη, και για την οποία ενδιαφερόμαστε να εξαγάγουμε συμπεράσματα μέσω του δείγματος.

Ορισμός 1.5

Εκτιμητής (*estimator*) $\hat{\theta}$ του θ ονομάζεται μια συνάρτηση των τυχαίων μεταβλητών του δείγματος, που χρησιμοποιείται με σκοπό την εκτίμηση του θ .

Σύμφωνα με τον ορισμό, θα είναι $\hat{\theta} = \hat{\theta}(s) = \hat{\theta}(X_1, X_2, \dots, X_n)$ και, κατά συνέπεια: (α) ο εκτιμητής είναι επίσης μια τυχαία μεταβλητή ως συνάρτηση τυχαίων μεταβλητών και (β) είναι εφικτό να υπολογίσουμε την αριθμητική τιμή του αμέσως μετά τη διεξαγωγή της έρευνας, όταν θα είναι γνωστές οι αριθμητικές τιμές των X_i ($i = 1, 2, \dots, n$).

Ένας όρος γενικότερος του εκτιμητή είναι η *στατιστική συνάρτηση* (*σ.σ.*) ή *στατιστικό* (*statistical function* ή *statistic*). Στατιστικό ονομάζεται μια συνάρτηση των τ.μ. του δείγματος. Σύμφωνα με τον Ορισμό 1.5, ένας εκτιμητής είναι στατιστική συνάρτηση. Τα περισσότερα από τα συμπεράσματα στη θεωρία δειγματοληψίας βασίζονται στη *δειγματική κατανομή* (*sampling distribution*). Ως δειγματική κατανομή μιας στατιστικής συνάρτησης ορίζεται η κατανομή των τιμών της στατιστικής συνάρτησης που προκύπτουν, εάν γίνει εξάντληση όλων των δυνατών δειγμάτων σύμφωνα με τον τρόπο που ακολουθείται για την επιλογή του δείγματος.

Η *αναμενόμενη τιμή* (*expected value*) μιας σ.σ. $t = t(X_1, X_2, \dots, X_n) = t(s)$ υπολογίζεται λαμβάνοντας υπόψη τις δυνατές τιμές της συνάρτησης $t = t(s)$, δηλαδή για όλα τα δυνατά δείγματα s και τις αντίστοιχες πιθανότητες πραγματοποίησης του κάθε δείγματος, σύμφωνα με τον τρόπο δειγματοληψίας. Θα είναι συνεπώς:

$$E(t) = \sum_s \pi(s)t(s) \quad (1.1)$$

όπου ο δείκτης s του αθροίσματος λαμβάνει όλες τις δυνατές τιμές του δείγματος, $t(s)$ είναι η τιμή της σ.σ. t υπολογισμένη για το δείγμα s , και $\pi(s)$ η πιθανότητα επιλογής του s .

Ο ορισμός του εκτιμητή είναι αρκετά γενικός, με αποτέλεσμα να επιτρέπει οποιαδήποτε συνάρτηση (οποιασδήποτε μορφής) να είναι θεωρητικά ένας εκτιμητής της ποσότητας που μας ενδιαφέρει. Υπάρχει συνεπώς ανάγκη για αξιολόγηση και σύγκριση των εκτιμητών μεταξύ τους. Η αξιολόγηση γίνεται με βάση μια σειρά κριτηρίων, ορισμένα εκ των οποίων αποτελούν ταυτόχρονα και ιδιότητες των εκτιμητών. Παραθέτουμε στη συνέχεια τα πιο σημαντικά από αυτά.

Ορισμός 1.6

Αμεροληψία (*unbiasedness*) Ένας εκτιμητής $\hat{\theta}$ θα λέγεται αμερόληπτος εκτιμητής της παραμέτρου θ εάν η αναμενόμενη τιμή του ισούται με θ , δηλ. $E(\hat{\theta}) = \theta$.

Η αναμενόμενη τιμή του εκτιμητή, $E(\hat{\theta})$, υπολογίζεται σύμφωνα με τη σχέση (1.1). Ας δούμε την εφαρμογή του ορισμού της αναμενόμενης τιμής και της ιδιότητας της αμεροληψίας ενός εκτιμητή μέσα από ένα παράδειγμα. Υποθέτουμε ένα μικρό θεωρητικό παράδειγμα πληθυσμού, όπου για χάρη κατανόησης οι τιμές του πληθυσμού είναι γνωστές.

Παράδειγμα 1.6

Σε μια μικρή πόλη που έχει 5 δημοτικά σχολεία, επιθυμούμε να εκτιμήσουμε τον μέσο αριθμό μαθητών ανά σχολείο, λαμβάνοντας ένα δείγμα 2 σχολείων. Έστω ότι τα στοιχεία των συνολικών αριθμών των μαθητών στα σχολεία είναι γνωστά για τον πληθυσμό:

α/α σχολείου	1	2	3	4	5
Συνολικός αριθμός μαθητών	59	28	90	44	36

Η παράμετρος του πληθυσμού που ενδιαφερόμαστε να εκτιμήσουμε είναι η μέση τιμή \bar{Y} του αριθμού μαθητών στα σχολεία, έστω θ , δηλ. $\bar{Y} = \theta$. Επειδή ο πληθυσμός είναι γνωστός, είναι $\theta = \frac{257}{5} = 51.4$.

Εάν η επιλογή του δείγματος γίνει επιλέγοντας τυχαία 2 σχολεία από τα 5 χωρίς επαναποθέτηση, τότε τα δυνατά δείγματα που μπορούν να προκύψουν είναι όσοι οι συνδυασμοί 5 ανά 2, δηλαδή $\binom{5}{2} = \frac{5!}{2!3!} = 10$. Επειδή οι πιθανότητες επιλογής κατά την τυχαία επιλογή είναι ίσες, το καθένα από τα 10 δείγματα έχει ίση πιθανότητα επιλογής και κατά συνέπεια ίση με $\frac{1}{10}$. Θεωρούμε στη συνέχεια ως εκτιμητή $\hat{\theta}$ του ποσοστού θ , τη δειγματική μέση τιμή, $\hat{\theta} = \bar{X}$. Ο πίνακας που ακολουθεί δίνει ένα προς ένα όλα τα δυνατά δείγματα μεγέθους 2, καθώς και την εκτίμηση $\hat{\theta}$ που λαμβάνεται από το κάθε δείγμα χωριστά.

Δείγμα s_i	Πιθανότητα επιλογής $\pi(s_i)$	Εκτιμητής $\hat{\theta}(s_i)$
{1,2}	$\frac{1}{10}$	$\frac{59 + 28}{2} = 43.5$
{1,3}	$\frac{1}{10}$	$\frac{59 + 90}{2} = 74.5$
{1,4}	$\frac{1}{10}$	$\frac{59 + 44}{2} = 51.5$
{1,5}	$\frac{1}{10}$	$\frac{59 + 36}{2} = 47.5$
{2,3}	$\frac{1}{10}$	$\frac{28 + 90}{2} = 59$
{2,4}	$\frac{1}{10}$	$\frac{28 + 44}{2} = 36$
{2,5}	$\frac{1}{10}$	$\frac{28 + 36}{2} = 32$
{3,4}	$\frac{1}{10}$	$\frac{90 + 44}{2} = 67$
{3,5}	$\frac{1}{10}$	$\frac{90 + 36}{2} = 63$
{4,5}	$\frac{1}{10}$	$\frac{44 + 36}{2} = 40$

Από την εξάντληση των δειγμάτων μεγέθους 2, διαπιστώνουμε ότι η εκτίμηση του μέσου αριθμού μαθητών ανά σχολείο (το οποίο στην προκειμένη περίπτωση γνωρίζουμε ότι είναι 51.4) λαμβάνει 10 διαφορετικές τιμές, με εύρος από 32 έως 74.5. Η δειγματική κατανομή του εκτιμητή είναι οι 10 αυτές διαφορετικές τιμές, ενώ η πιθανότητα πραγματοποίησης της κάθε τιμής ισούται με την πιθανότητα επιλογής του αντίστοιχου δείγματος. Το στατιστικό σφάλμα που εμπεριέχεται στις δειγματοληπτικές έρευνες συνδέεται με το γεγονός ότι ο εκτιμητής είναι μια τυχαία μεταβλητή και λαμβάνει τόσες δυνατές τιμές, όσες και το πλήθος των δυνατών δειγμάτων. Αντίθετα, η πληθυσμιακή ποσότητα $\theta = 51.4$ είναι σταθερή.

Χρησιμοποιώντας τις δύο τελευταίες στήλες του πίνακα και εφαρμόζοντας τον ορισμό της αναμενόμενης τιμής (1.1), θα έχουμε:

$$\begin{aligned}
E(\hat{\theta}) &= \sum_{i=1}^{10} \pi(s_i) \hat{\theta}(s_i) = \sum_{i=1}^{10} \frac{1}{10} \hat{\theta}(s_i) \\
&= \frac{1}{10} (43.5 + 74.5 + 51.5 + 47.5 + 59 + 36 + 32 + 67 + 63 + 40) = 51.4
\end{aligned}$$

Άρα ο εκτιμητής $\hat{\theta}$ που υιοθετήσαμε είναι ένας αμερόληπτος εκτιμητής του μέσου αριθμού των μαθητών ανά σχολείο.

Διαπιστώνουμε ότι για την εύρεση της μέσης τιμής του εκτιμητή στο παράδειγμα, λάβαμε υπόψη μας (α) όλα τα δυνατά δείγματα μεγέθους που μπορούν να επιλεγούν και (β) τον τρόπο δειγματοληψίας, δηλ. τις πιθανότητες επιλογής του καθενός από τα δυνατά δείγματα. Συνεπώς, η αναμενόμενη τιμή του ίδιου εκτιμητή $\hat{\theta} = \bar{X}$, θα είναι ενδεχομένως διαφορετική, εάν υιοθετήσουμε ένα εναλλακτικό δειγματοληπτικό σχέδιο, π.χ. με άνισες πιθανότητες για τα 10 δυνατά δείγματα.

Το σύνολο των δυνατών δειγμάτων που προκύπτουν με μια μέθοδο δειγματοληψίας ονομάζεται δειγματοληπτικός χώρος (sampling space) και συμβολίζεται συνήθως με \mathcal{S} . Για το παράδειγμά μας, ο δειγματοληπτικός χώρος είναι:

$$\mathcal{S} = \{\{1,2\}, \{1,3\}, \{1,3\}, \dots, \{4,5\}\}$$

Ο αριθμός των στοιχείων του \mathcal{S} ταυτίζεται με το πλήθος των δυνατών δειγμάτων που προκύπτουν σύμφωνα με τον τρόπο δειγματοληψίας που ακολουθούμε στην έρευνα. Εάν για το Παράδειγμα 1.5 θεωρούμε ως τρόπο δειγματοληψίας: επιλογή με τυχαίο τρόπο ενός σχολείου μεταξύ αυτών με α/α 1,2 και 3 και ενός σχολείου μεταξύ των 4 και 5, ο δειγματικός χώρος θα είναι

$$\mathcal{S} = \{\{1,4\}, \{1,5\}, \{2,4\}, \{2,5\}, \{3,4\}, \{4,5\}\}$$

που αποτελείται από $\binom{3}{1} \binom{2}{1} = 6$ δείγματα.

Ορισμός 1.7

Το ποσό μεροληψίας (*bias*) ενός εκτιμητή $\hat{\theta}$, συμβολίζεται ως $\text{bias}(\hat{\theta})$ και δίνεται από τη σχέση $\text{bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$.

Εάν $\text{bias}(\hat{\theta}) > 0$, αυτό σημαίνει ότι ο εκτιμητής παρουσιάζει θετική μεροληψία, δηλ. για το δοθέν δειγματοληπτικό σχέδιο αναμένουμε η δειγματική κατανομή του $\hat{\theta}$ να έχει μέση τιμή μεγαλύτερη του θ και, κατά συνέπεια, ο εκτιμητής να υπερ-εκτιμά την παράμετρο. Αντίστοιχα, αν $\text{bias}(\hat{\theta}) < 0$, έχουμε αρνητική μεροληψία ή ισοδύναμα ο εκτιμητής υπο-εκτιμά την παράμετρο. Προφανώς, αν $\text{bias}(\hat{\theta}) = 0$ ο εκτιμητής είναι αμερόληπτος.

Ένα άλλο κριτήριο σύγκρισης εκτιμητών είναι η *ακρίβεια* (*accuracy*). Η ακρίβεια ως ιδιότητα ενός εκτιμητή δίνει ένα μέτρο της συγκέντρωσης, ή, αντίθετα, της απόκλισης που παρουσιάζουν μεταξύ τους οι δυνατές τιμές του εκτιμητή. Όσο πιο πυκνά είναι οι δυνατές τιμές του εκτιμητή, όσο δηλαδή μεγαλύτερη είναι η συγκέντρωση, τόσο μεγαλύτερη είναι η ακρίβεια του εκτιμητή.

Ένα μέτρο της ακρίβειας του εκτιμητή είναι το μέσο τετραγωνικό σφάλμα, (*mean square error*) που συμβολίζεται με MSE.

Ορισμός 1.8

Το μέσο τετραγωνικό σφάλμα ενός εκτιμητή $\hat{\theta}$ ορίζεται ως η αναμενόμενη τιμή της τετραγωνικής απόκλισης του εκτιμητή από την προς εκτίμηση ποσότητα:

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 \quad (1.2)$$

Η αναμενόμενη τιμή στον ορισμό του MSE υπολογίζεται όπως και στη σχέση (1.1), δηλαδή:

$$MSE(\hat{\theta}) = \sum_{s \in \mathcal{S}} \pi(s) (\hat{\theta}(s) - \theta)^2$$

Ακριβής (*accurate*) είναι ένας εκτιμητής με μικρό μέσο τετραγωνικό σφάλμα.

Το MSE σχετίζεται με το γνωστό μέτρο της διακύμανσης ενός εκτιμητή. Υπενθυμίζουμε ότι η *διακύμανση* (*variance*) ενός εκτιμητή ορίζεται από τη σχέση:

$$\text{Var}(\hat{\theta}) = E(\hat{\theta} - E(\hat{\theta}))^2 \quad (1.3)$$

Από τις σχέσεις (1.2) και (1.3) εύκολα προκύπτει ότι εάν $E(\hat{\theta}) = \theta$, δηλαδή εάν ο εκτιμητής $\hat{\theta}$ είναι αμερόληπτος, τότε $MSE(\hat{\theta}) = \text{Var}(\hat{\theta})$.

Γενικότερα, αποδεικνύεται ότι η σχέση που συνδέει το MSE ενός εκτιμητή και τη διακύμανσή του είναι

$$MSE(\hat{\theta}) = \text{Var}(\hat{\theta}) + \text{bias}^2(\hat{\theta}) \quad (1.4)$$

Μεταξύ όλων των αμερόληπτων εκτιμητών, ο εκτιμητής με την ελάχιστη διακύμανση λέγεται *αποτελεσματικός* (*efficient*). Αν $\hat{\theta}_1$ και $\hat{\theta}_2$ είναι δύο αμερόληπτοι εκτιμητές με $\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2)$ τότε ο εκτιμητής $\hat{\theta}_1$ λέγεται σχετικά αποτελεσματικός.

Ορισμός 1.9

Τυπικό σφάλμα (*standard error*) του εκτιμητή $\hat{\theta}$, συμβολικά $se(\hat{\theta})$, ονομάζεται η θετική τετραγωνική ρίζα της διακύμανσης του εκτιμητή, δηλ.

$$se(\hat{\theta}) = \sqrt{\text{Var}(\hat{\theta})}$$

Για τον υπολογισμό της αναμενόμενης τιμής, του μέσου τετραγωνικού σφάλματος ή της διακύμανσης ενός εκτιμητή λαμβάνεται υπόψη ο δειγματικός χώρος και οι πιθανότητες επιλογής των δειγμάτων. Συνεπώς, όλες οι παραπάνω ποσότητες εξαρτώνται από το δειγματοληπτικό σχέδιο που υιοθετήθηκε στην έρευνα.

Τέλος, ως υπενθύμιση, ο Πίνακας 1.1 συγκεντρώνει ορισμένες από τις βασικές ιδιότητες της αναμενόμενης τιμής, της διακύμανσης και ποσοτήτων που ορίζονται με τη βοήθεια αυτών, όπως η *συμμεταβλητότητα* (*covariance*) και η *συσχέτιση* (*correlation*).

Τα σύμβολα X, Y, Z αντιπροσωπεύουν τυχαίες μεταβλητές, ενώ τα a, b, c πραγματικούς αριθμούς.

Όνομασία	Ιδιότητες
Αναμενόμενη Τιμή	<ol style="list-style-type: none"> 1. $E(a) = a$ 2. $E(aX + bY) = aE(X) + bE(Y)$
Διακύμανση	<ol style="list-style-type: none"> 1. $\text{Var}(a) = 0$ 2. $\text{Var}(aX) = a^2 \text{Var}(X)$ 3. Αν X, Y είναι ασυσχέτιστες: $\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y)$

<p>Συμμεταβλητότητα</p> $\text{Cov}(X, Y) = E[(X - \bar{X})(Y - \bar{Y})] = E(XY) - E(X)E(Y)$	<ol style="list-style-type: none"> 1. Αν $\text{Cov}(X, Y) = 0$, τότε X, Y ασυσχέτιστες 2. $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$ 3. $\text{Cov}(aX, Y) = a \text{Cov}(X, Y)$ 4. $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$ 5. $\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2\text{Cov}(X, Y)$
<p>Συσχέτιση</p> $\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$	$-1 \leq \rho(X, Y) \leq 1$

Πίνακας 2.1 Βασικές ιδιότητες κυριότερων στατιστικών ποσοτήτων

1.6. Κύρια στοιχεία μιας δειγματοληπτικής έρευνας

Στην παράγραφο 1.1 είδαμε ορισμένα παραδείγματα δειγματοληπτικών ερευνών. Έχοντας δώσει ενδιάμεσα το θεωρητικό υπόβαθρο και τη σχετική ορολογία, είμαστε σε θέση να δώσουμε σε πιο συγκεκριμένη και αναλυτική μορφή τα στοιχεία από τα οποία αποτελείται μια δειγματοληπτική έρευνα. Για μεγαλύτερη ανάπτυξη και εξειδίκευση ανάλογα με το θέμα της έρευνας, παραπέμπουμε στους [Henry](#) (1990), [Blair](#), Czajka and Blair (2013), [Fink](#) (2013) και [Floyd](#) and Fowel (2001).

- *Το Δειγματοληπτικό σχέδιο (sampling design).*

Είναι ο τρόπος επιλογής των μονάδων στο δείγμα. Το δειγματοληπτικό σχέδιο επιλέγεται μετά από συνεργασία (i) του στατιστικού επιστήμονα, (ii) εκείνου που αναθέτει την έρευνα και καθορίζει τα ερωτήματά της και (iii) τους ανθρώπους που θα διεξαγάγουν τη δειγματοληψία, των συνεντευκτών (interviewers) όπως λέγονται.

Για την επιλογή του δειγματοληπτικού σχεδίου τονίζουμε τον σημαντικό, αν και όχι προφανή, ρόλο εκείνου που αναθέτει την έρευνα. Ο ρόλος του είναι σημαντικός, επειδή θέτει τα ερωτήματα και συνεπώς καθορίζει ποιες μεταβλητές πρέπει να καταγραφούν. Επίσης, προσδιορίζει τους χρηματικούς πόρους που είναι διαθέσιμοι για την έρευνα και θέτει τις απαιτήσεις ή προδιαγραφές της έρευνας ως προς την επιθυμητή ακρίβεια των εκτιμητών που θα παραχθούν. Για παράδειγμα, εάν το πρόβλημα είναι η εκτίμηση ενός ποσοστού, μπορεί να επιθυμεί η έρευνα να διεξαχθεί έτσι ώστε ο εκτιμητής του ποσοστού να έχει σφάλμα που δεν ξεπερνά το 2%. Όλα τα παραπάνω συνυπολογίζονται από τον στατιστικό αναλυτή, προκειμένου να επιλέξει το δειγματοληπτικό σχέδιο ή τον συνδυασμό των δειγματοληπτικών σχεδίων και τη μέθοδο εκτίμησης των πληθυσμιακών παραμέτρων.

Ο ρόλος του συνεντευκτή στη διαδικασία σχεδιασμού της έρευνας είναι συμπληρωματικός και βοηθητικός. Δίνει πληροφορίες για την εκτίμηση του χρόνου και της δυσκολίας της συλλογής των δεδομένων. Τα στοιχεία αυτά συνεκτιμώνται από τον στατιστικό αναλυτή.

- *Οι μετρήσεις της έρευνας (survey measurements).*

Μετά την ανάθεση της έρευνας από κάποιο φορέα/άτομο, ένα ειδικευμένο προσωπικό επιλέγει τις μεταβλητές που θα συμπεριληφθούν και τη μονάδα μέτρησης της καθεμιάς. Το ίδιο αυτό προσωπικό, που συχνά απαρτίζεται από ψυχολόγους και κοινωνιολόγους, συντάσσει στη συνέχεια το *ερωτηματολόγιο (questionnaire)*.

Η σύνταξη του ερωτηματολογίου είναι σημαντική και πολλές φορές πολύπλοκη. Ο τελικός στόχος ενός καλά σχεδιασμένου ερωτηματολογίου είναι να προσφέρει με το πέρας της έρευνας ένα αξιόπιστο σύνολο δεδομένων, το οποίο στη συνέχεια θα αναλύσει ο στατιστικός αναλυτής.

Υπάρχουν αρκετοί κανόνες σύμφωνα με τους οποίους σχεδιάζεται ένα καλό ερωτηματολόγιο (βλ. [Floyd](#) and Fowler, 2001). Εν συντομία, η σύνταξη ενός ερωτηματολογίου γίνεται επιδιώκοντας (α) να έχει συγκεκριμένο στόχο (focus), (β) συντομία και (γ) απλότητα. Τα τρία αυτά στοιχεία εξασφαλίζουν ελαχιστοποίηση των σφαλμάτων λόγω κακής διατύπωσης των ερωτήσεων που έχει

ως αποτέλεσμα λανθασμένη κατανόηση και καταγραφή της μέτρησης, απαλοιφή της μεροληψίας που προέρχεται από κατευθυνόμενες ερωτήσεις, και μετρήσεις που είναι προϊόν είτε κούρασης είτε τυχαίας επιλογής απαντήσεων.

- *Διεξαγωγή της έρευνας (survey operations)*

Κατά τη διεξαγωγή της έρευνας γίνεται προσπάθεια να μην υπάρξουν παρεκκλίσεις από τον αρχικό σχεδιασμό σε κανένα από τα ενδιάμεσα στάδια. Αυτό είναι σημαντικό, γιατί η αξιοπιστία των δεδομένων και των αποτελεσμάτων εξασφαλίζεται μόνο εφόσον η πραγματοποίηση της έρευνας ακολουθεί τις υποθέσεις της στατιστικής θεωρίας βάσει της οποίας ισχύουν οι ιδιότητες των εκτιμητών. Για παράδειγμα, εάν η καταγραφή της μέτρησης σε μια ερώτηση γίνεται συστηματικά με λανθασμένο τρόπο από τον συνεντευκτή, ο εκτιμητής που θα υπολογιστεί για τη μεταβλητή που αντιστοιχεί στη συγκεκριμένη ερώτηση θα είναι μεροληπτικός, ακόμα και αν θεωρητικά, ως στατιστική συνάρτηση, ο εκτιμητής έχει την ιδιότητα της αμεροληψίας.

Ο αριθμός των μελών της ομάδας διεξαγωγής της έρευνας εξαρτάται από τους σκοπούς, την έκταση και τον προϋπολογισμό της έρευνας.

Τέλος, σημαντικό ρόλο κατά τη διεξαγωγή της έρευνας στην αποφυγή και πρόληψη προβλημάτων τα οποία ενδέχεται να οδηγήσουν σε απόκλιση από τον σχεδιασμό, παίζει η *πιλοτική έρευνα (pilot study)*.

Πιλοτική είναι μια έρευνα η οποία διεξάγεται πριν από την κύρια έρευνα. Ακολουθεί τον σχεδιασμό της κύριας έρευνας σε κάθε στάδιο (δειγματοληπτικό σχέδιο, ερωτηματολόγιο, καταγραφή, ανάλυση), αλλά είναι μικρότερης κλίμακας, δηλ. επιλέγεται ένα αρκετά μικρότερο δείγμα από εκείνο της κύριας. Η πιλοτική έρευνα είναι απαραίτητη κυρίως σε νέες έρευνες, και η σκοπιμότητά της είναι πολύπλευρη. Σε αρχική φάση, ελέγχεται το ερωτηματολόγιο και εντοπίζονται τυχόν παρερμηνείες ή ερωτήσεις που δεν είναι κατανοητές. Επίσης, γίνεται μια εκτίμηση του χρόνου της διεξαγωγής της έρευνας και ένας έλεγχος για την καλή εκπαίδευση και προετοιμασία των συνεντευκτών. Τέλος, και σημαντικότερο, λαμβάνονται κάποιες πρώτες εκτιμήσεις για τις άγνωστες παραμέτρους του πληθυσμού, οι οποίες θα χρησιμεύσουν στην υλοποίηση της κύριας έρευνας. Για παράδειγμα, όπως θα δούμε στο Κεφάλαιο 2, για τον προσδιορισμό του απαιτούμενου μεγέθους του δείγματος ώστε να πληρούνται ορισμένες προδιαγραφές, είναι απαραίτητο να διαθέτουμε στοιχεία για τον πληθυσμό, όπως η διακύμανση του υπό μελέτη χαρακτηριστικού. Εάν τα στοιχεία αυτά δεν παρέχονται από μια πρόσφατη απογραφή ή από μια προγενέστερη έρευνα, τότε η πιλοτική είναι αναγκαία.

- *Στατιστική Ανάλυση (Statistical Analysis)*

Μετά τη διεξαγωγή της έρευνας και την καταγραφή των μετρήσεων, είναι διαθέσιμο ένα σύνολο δεδομένων. Η ανάλυση των δεδομένων θα γίνει από τον στατιστικό ερευνητή εφαρμόζοντας τα αποτελέσματα της Θεωρίας Δειγματοληψίας και λαμβάνοντας υπόψη το δειγματοληπτικό σχέδιο που υιοθετήθηκε κατά την έρευνα. Τα κύρια σημεία της ανάλυσης είναι ο υπολογισμός των εκτιμητών των αγνώστων ποσοτήτων του πληθυσμού βάσει του δείγματος, ο υπολογισμός των σφαλμάτων της εκτίμησης και η κατασκευή διαστημάτων εμπιστοσύνης.

Σε σύνθετα δειγματοληπτικά σχέδια, όπου γίνεται συνδυασμός αρκετών επιμέρους δειγματοληπτικών σχεδίων, ο υπολογισμός των τυπικών σφαλμάτων των εκτιμητών γίνεται με τη βοήθεια προσεγγιστικών μεθόδων.

Τέλος, τα αποτελέσματα της ανάλυσης καταγράφονται και παρουσιάζονται.

Βιβλιογραφικές Αναφορές

[Blair](#), J., Czaja, R. F. & Blair, E. (2013). *Designing Surveys: A Guide to Decisions and Procedures*. 3rd Edition. Sage Publications.

[Fink](#), A. (2013). *How to Conduct Surveys: A Step-by-Step Guide*. 5th Edition. Sage Publications.

[Floyd](#), J. & Fowler, Jr. (2001). *Survey Research Methods (Applied Social Research Methods)*. 3rd Edition. Sage Publications.

[Groves](#), R.M. Floyd, J., Fowler, Jr., Couper, M.P., Lepkowski, J.M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology*. 2nd Edition, Wiley-Blackwell.

[Henry](#), G. T. (1990). *Practical Sampling*, Sage Publications (CA).