

## Μέθοδοι Συγκέντρωσης Δεδομένων

Οι διάφορες μέθοδοι συγκέντρωσης στατιστικών στοιχείων μπορούν να συνοψισθούν σε δύο μεγάλες ομάδες :τις απογραφές και τις δειγματοληπτικές έρευνες.

Η διαδικασία με την οποία συλλέγονται όλες οι παρατηρήσεις του πληθυσμού σε μία συγκεκριμένη χρονική στιγμή ονομάζεται **απογραφή**. Η απογραφή δεν είναι δυνατή παρά μόνον όταν ο πληθυσμός είναι καλά ορισμένος, και πεπερασμένος. Σε μια τέτοια περίπτωση η γνώση μας για τον πληθυσμό είναι καθολική. Σύμφωνα με το αντικείμενο που ενδιαφερόμαστε, οι απογραφές παίρνουν και διαφορετικό χαρακτήρα όπως :

- Οι Γεωργικές απογραφές, στις οποίες συγκεντρώνονται πληροφορίες για τις εκτάσεις που καλλιεργούνται, το είδος της γεωργικής παραγωγής, τον αριθμό των γεωργικών μηχανημάτων, το είδος των λιπασμάτων κ.λ.π.
- Οι Δημογραφικές απογραφές, στις οποίες συλλέγονται στοιχεία σχετικά με το φύλο, την ηλικία, το επάγγελμα κ.λ.π.
- Οι Οικονομικές απογραφές, όπου συγκεντρώνονται στοιχεία σχετικά με την οικονομική κατάσταση (δαπάνες, εισοδήματα, ...) των στατιστικών μονάδων που απογράφονται.
- Οι Βιομηχανικές απογραφές, κατά τις οποίες δεν περιοριζόμαστε στην απλή καταγραφή των υπάρχοντων βιομηχανικών καταστημάτων, αλλά συγκεντρώνουμε πληροφορίες που αφορούν τον κλάδο της οικονομικής τους δραστηριότητας, τον αριθμό απασχολούμενων, το επίπεδο μηχανοργάνωσης κ.λ.π.

Είναι δυνατόν ακόμη και όταν ο πληθυσμός είναι πεπερασμένος, να μην προτιμηθεί η μέθοδος της απογραφής. Αυτό οφείλεται στα μειονεκτήματα των απογραφών τα οποία είναι:

α) Το μεγάλο κόστος. Για να πραγματοποιηθεί μία απογραφή χρειάζεται ειδική προεργασία καθώς και μεγάλος αριθμός εκπαιδευμένων απογραφέων (τα άτομα που θα συγκεντρώνουν τα στοιχεία). Γι' αυτή την προεργασία αλλά και την ίδια ημέρα της απογραφής απαιτούνται μεγάλες δαπάνες. Αυτός είναι και ο λόγος που η απογραφή πληθυσμού γίνεται κάθε δεκαετία, ενώ η οικονομικότερη απογραφή βιομηχανιών και βιοτεχνιών κάθε πενταετία.

β) Τα εξειδικευμένα άτομα. Για να πραγματοποιηθεί μια απογραφή είναι απαραίτητο να υπάρχει ένας μεγάλος αριθμός ειδικευμένων ατόμων, των απογραφέων, ώστε να αποφευχθούν τα προσωπικά σφάλματα. Επειδή όμως πολλές φορές δεν είναι δυνατό, λόγω της έκτασης των απογραφών, να διενεργούνται μόνο από εξειδικευμένα άτομα, των οποίων η αυστηρή και πλήρης εκπαίδευση αποικλείει προσωπικά σφάλματα, η χρήση μη ειδικευμένων απογραφέων έχει ως συνέπεια τη συγκέντρωση πληροφοριών που λόγω των προσωπικών σφαλμάτων είναι δυνατό να δώσουν απατηλή εικόνα της διάρθρωσης του πληθυσμού.

γ) Επικαιρότητα αποτελεσμάτων. Ο μεγάλος αριθμός τόσο των πληροφοριών όσο και των ατόμων που αποτελούν τον πληθυσμό στις απογραφές δεν επιτρέπουν να έχουμε τη δυνατότητα να δημοσιεύσουμε σύντομα τα αποτελέσματα. Έτσι πολλές φορές, παρά τη μηχανογραφική επεξεργασία των στοιχείων, τα αποτελέσματα χάνουν την επικαιρότητα τους και περιορίζονται στην ιστορική τους μόνο αξία.

Τέλος σημειώνεται ότι σε ορισμένες περιπτώσεις η παρατήρηση ενός χαρακτηριστικού μίας στατιστικής μονάδας συνεπάγεται καταστροφή της. Για

παράδειγμα η εξέταση της διάρκειας ζωής μιας παρτίδας ηλεκτρικών λαμπτήρων μπορεί να γίνει μόνον δειγματοληπτικά. Έτσι, ενώ θεωρητικά είναι επιθυμητό να έχουμε κάθε μέτρηση του πληθυσμού, στην πράξη θα έχουμε μόνον ένα, συνήθως μικρό, μέρος αυτών των μετρήσεων.

**Δειγματοληψία** είναι η απογραφή ορισμένων συγκεκριμένων χαρακτηριστικών ενός τμήματος του πληθυσμού. Το τμήμα του πληθυσμού που απογράφεται ονομάζεται **δείγμα**.

Σκοπός των δειγματοληπτικών ερευνών είναι να προσδιορίσουμε όσο γίνεται ακριβέστερα ιδιότητες του πληθυσμού, μελετώντας τα στοιχεία του δείγματος. Οι εκτιμήσεις για τις ιδιότητες του πληθυσμού, σε αντίθεση με ότι συμβαίνει στις απογραφές, δεν είναι ακριβείς, αλλά αντίθετως αποτελούν προσεγγίσεις που περιέχουν κάποιο σφάλμα. Το σύνολο των τεχνικών με τις οποίες εξάγονται συμπεράσματα για τον πληθυσμό χρησιμοποιώντας τις δειγματικές πληροφορίες αναφέρεται ως στατιστική συμπερασματολογία. Η διαδικασία αυτή κατά την οποία η γνώση για το δείγμα γενικεύεται στον πληθυσμό ονομάζεται **επαγωγική στατιστική**. Η επαγωγική στατιστική χρησιμοποιείται στις εμπειρικές επιστήμες όπου η διαδικασία παραγωγής νέας γνώσης συνίσταται στην εκτέλεση ενός πειράματος και στη γενίκευση των συμπερασμάτων από το πείραμα αυτό σε όλα τα όμοια πειράματα που θα μπορούσαν να γίνουν.

Η γενίκευση των πληροφοριών του δείγματος σε ολόκληρο τον πληθυσμό συνεπάγεται αβεβαιότητα η οποία μπορεί να μετρηθεί υπό την προϋπόθεση ότι το δείγμα είναι τυχαίο. Το τυχαίο δείγμα δεν είναι αυτό που έχει προκύψει στην τύχη αλλά αντίθετα, έχει επιλεγεί με καλά ορισμένες αρχές της θεωρίας πιθανοτήτων. Η τυχαία επιλογή των μονάδων του δείγματος από τον πληθυσμό σημαίνει ότι δεν υπεισέρχεται ο υποκειμενικός παράγοντας στην επιλογή ή όχι μιας στατιστικής μονάδας και κάθε μονάδα του πληθυσμού έχει γνωστή πιθανότητα να επιλεγεί σαν μονάδα του δείγματος.

**Τυχαίες δειγματοληψίες** είναι εκείνες, που κατά τη δειγματοληπτική συγκέντρωση των ατόμων του δείγματος, όλα τα άτομα του πληθυσμού έχουν την ίδια πιθανότητα τυχαίας επιλογής. Για να εφαρμόσουμε αυτές τις μεθόδους θα πρέπει:

- Να έχουμε στη διάθεση μας όλες τις μονάδες του πληθυσμού που θέλουμε δειγματοληπτικά να μελετήσουμε (σε μορφή κατάστασης ή σε μορφή αρχείου ηλεκτρονικού υπολογιστή).
- Να αριθμηθούν όλες οι μονάδες του πληθυσμού ώστε επιλέγοντας τυχαία τόσους αριθμούς όσο το επιθυμητό μέγεθος του δείγματος, να γνωρίζουμε σε ποιες μονάδες αντιστοιχούν αυτοί οι αριθμοί.

Με αυτόν τον τρόπο έχουμε μια συστηματική καταγραφή των μονάδων του πληθυσμού. Το πλαίσιο διευκολύνει σημαντικά την οργάνωση της δειγματοληψίας. Ανάλογα με τον τρόπο που επιτυγχάνεται η τυχαιότητα οι τυχαίες δειγματοληψίες διακρίνονται σε:

- Απλές τυχαίες δειγματοληψίες.
- Συστηματικές δειγματοληψίες.
- Τυχαίες δειγματοληψίες κατά στρώματα.

Όταν κάθε στατιστική μονάδα του πληθυσμού που μελετούμε έχει την ίδια πιθανότητα να επιλεγεί σαν μονάδα του δείγματος, τότε έχουμε **απλή τυχαία δειγματοληψία**

Η πιθανότητα επιλογής μιας μονάδας του πληθυσμού στο δείγμα πρέπει να είναι ανεξάρτητη από την επιλογή κάποιας άλλης μονάδας. Έτσι, η επιλογή  $n$  στατιστικών μονάδων από πληθυσμό μεγέθους  $N$  θα πρέπει να είναι ίδια με την εξαγωγή  $n$  σφαιρών από μια κάλπη που περιέχει  $N$  σφαίρες αριθμημένες από το 1 ως το  $N$ .

Η απλή τυχαία δειγματοληψία μπορεί να πραγματοποιηθεί και με τη χρήση των πινάκων τυχαίων αριθμών. Οι τυχαίοι αριθμοί δημιουργούνται με ειδικά προγράμματα Η/Υ και είναι τέτοιοι ώστε ανεξάρτητα από το που ξεκινάμε, ανεξάρτητα από την κατεύθυνση στην οποία κινούμαστε, τα 10 ψηφία 0, 1, 2, ..., 8, 9 εμφανίζονται τυχαία με ίση πιθανότητα.

Πολλές φορές, παρά το ότι οι στατιστικές μονάδες του πληθυσμού είναι αριθμημένες ή διατεταγμένες έτσι ώστε να θεωρούνται αριθμημένες, δεν είναι εύκολο να επιτύχουμε αντιπροσωπευτικό δείγμα εφαρμόζοντας την απλή τυχαία δειγματοληψία. Εάν π.χ. θέλουμε να επιλέξουμε δείγμα 100 πελατών από τους 1000 ενός ξενοδοχείου, που είναι αριθμημένοι με τη σειρά άφιξης τους σ' αυτό από την 1η Ιανουαρίου μέχρι την 31η Δεκεμβρίου και εφαρμόσουμε την απλή τυχαία δειγματοληψία, δεν είμαστε σίγουροι για την κατανομή των ατόμων του δείγματος μέσα στο χρόνο. Σε τέτοιες περιπτώσεις επιλέγουμε τυχαία την πρώτη μονάδα του δείγματος και στη συνέχεια μέχρι να συμπληρωθούν οι  $n$  μονάδες, επιλέγουμε αυτές που ο αριθμός τους συμπίπτει με τους όρους μιας αριθμητικής προόδου, με λόγο της αριθμητικής προόδου  $N/n$  (όπου  $N$  το πλήθος των ατόμων του πληθυσμού και  $n$  το μέγεθος του δείγματος).

Μια διαδικασία της μορφής αυτής ονομάζεται **συστηματική δειγματοληψία** και το δείγμα συστηματικό δείγμα. Σε ένα τέτοιο σχήμα δειγματοληψίας κάθε στοιχείο του πληθυσμού έχει την ίδια πιθανότητα να περιληφθεί στο δείγμα όμως κάθε δυνατό δείγμα ίσου μεγέθους δεν έχει την ίδια πιθανότητα να επιλεγεί. Έτσι π.χ. το 20ο και το 21ο στοιχείο έχουν μηδενική πιθανότητα να περιληφθούν στο ίδιο δείγμα. Η συστηματική δειγματοληψία είναι χρήσιμη όχι μόνον για καταγεγραμμένα σε ένα πλαίσιο στοιχεία αλλά και για παραγόμενα εν σειρά προϊόντα, ή, για μονάδες διατεταγμένες στο χώρο. Έτσι, π.χ. για να ελέγξουμε την ποιότητα ενός προϊόντος που παράγεται μαζικά μπορούμε να παίρνουμε για δειγματοληπτικό έλεγχο κάθε 30ο παραγόμενο προϊόν ή ένα προϊόν ανά ένα τέταρτο της ώρας από τη σειρά παραγωγής.

Η συστηματική δειγματοληψία μπορεί να δημιουργήσει μεροληψία στις ακόλουθες περιπτώσεις:

1. Όταν τα στοιχεία του πληθυσμού είναι διατεταγμένα σύμφωνα με κάποιο κριτήριο οπότε το σημείο εκκίνησης επιδρά συστηματικά στο αποτέλεσμα. Έτσι, π.χ. αν τα άτομα είναι διατεταγμένα σύμφωνα με το εισόδημα τους το δείγμα που θα ξεκινήσει από το 1ο άτομο και θα περιλαμβάνει το 51ο, 101ο κοκ. θα δώσει χαμηλότερη εκτίμηση του μέσου εισοδήματος από το δείγμα που θα ξεκινήσει από το 50ο. Στην περίπτωση αυτή η συστηματική δειγματοληψία θα πρέπει να αρχίζει από την αρχική μεσαία τιμή ή θα πρέπει να επιχειρείται άλλο σχήμα δειγματοληψίας.
2. Όταν το πλαίσιο έχει περιοδικότητα υποπολλαπλάσια του μεγέθους του δείγματος. Έτσι π.χ. αν θελήσουμε να πάρουμε τυχαίο δείγμα 30 ημερών με συστηματική δειγματοληψία τότε θα καταλήξουμε να παρατηρούμε την ίδια μέρα κάθε μήνα.

Όταν ο πληθυσμός μπορεί να διαιρεθεί σε στρώματα ομοιογενή ως προς το χαρακτηριστικό που μας ενδιαφέρει τότε μπορούμε να πάρουμε ένα τυχαίο δείγμα

από κάθε στρώμα, οπότε στο συνολικό δείγμα που προκύπτει αντιπροσωπεύονται όλα τα στρώματα. Η μέθοδος αυτή ονομάζεται **στρωματοποιημένη δειγματοληψία**.

Η στρωματοποίηση μπορεί να γίνει κατά φύλο, κατά ηλικία, κατά επάγγελμα, κατά το οικογενειακό εισόδημα κτλ. Η στρωματοποιημένη δειγματοληψία συνιστάται ιδιαίτερα στις περιπτώσεις που υπάρχει διαφορετικό πλαίσιο για κάθε στρώμα. Τέτοιες είναι οι περιπτώσεις των τμημάτων ενός πανεπιστημίου, των σχολείων μιας εκπαιδευτικής περιφέρειας ή των εργοστασίων μιας παραγωγικής επιχείρησης.

Να σημειωθεί ότι όταν οι διαφορές μεταξύ των στρωμάτων είναι μικρές συγκρινόμενες με τις διαφορές μέσα στα στρώματα τότε η στρωματοποιημένη δειγματοληψία δεν αναμένεται να μας δώσει ακριβέστερη εικόνα του πληθυσμού από την απλή τυχαία δειγματοληψία.

Κατά τη διεξαγωγή των στατιστικών μελετών επιδιώκουμε να μειώνουμε στο ελάχιστο τις δυσκολίες που συναντάμε κατά τη συγκέντρωση των στατιστικών δεδομένων και να οδηγούμαστε σε όσο το δυνατό πιο αξιόπιστα συμπεράσματα. Για να μειωθούν αυτές οι δυσκολίες και να αυξηθεί η αξιοπιστία των συμπερασμάτων θα πρέπει να λαμβάνονται ορισμένες προφυλάξεις. Στο στάδιο της δειγματοληψίας θα πρέπει να ληφθούν υπόψη τα παρακάτω σφάλματα στην επιλογή του δείγματος:

- Το αυτοεπιλεγόμενο δείγμα έχει την τάση να περιλαμβάνει κυρίως άτομα με φανατισμό ή τουλάχιστον με πολύ ισχυρή γνώμη για το ερευνηόμενο θέμα. Έτσι π.χ. σε τηλεοπτικές εκπομπές λόγου, οι τηλεθεατές έχουν την δυνατότητα να «ψηφίζουν» για ένα θέμα, συνήθως με ένα «ναι» ή με ένα «όχι» τηλεφωνώντας με δική τους χρέωση. Είναι προφανές ότι ένα τέτοιο δείγμα περιλαμβάνει μόνον αυτούς που έχουν ισχυρή άποψη και συγχρόνως δεν υπολογίζουν το οικονομικό κόστος του τηλεφωνήματος.
- Δειγματοληψία σε ορισμένο σημείο του δρόμου και σε ορισμένη ώρα της ημέρας αποκλείει όλους εκείνους, οι οποίοι τη συγκεκριμένη ώρα έχουν λόγους να μην βρίσκονται στο δρόμο, είτε διότι εργάζονται, είτε διότι έχουν μειωμένη κινητικότητα, είτε επειδή το καταναλωτικό προφίλ τους δεν ταιριάζει με αυτό του συγκεκριμένου δρόμου.
- Ομοίως δειγματοληψία οικιών από τηλεφώνου 8-2 το πρωί έχει ως αποτέλεσμα μόνο οι οικογένειες με ένα τουλάχιστον μη εργαζόμενο μέλος να αντιπροσωπεύονται στο δείγμα.