

# 2

## ΣΥΛΛΟΓΗ ΣΤΑΤΙΣΤΙΚΩΝ ΣΤΟΙΧΕΙΩΝ

### 2.1. ΣΤΑΤΙΣΤΙΚΑ ΣΤΟΙΧΕΙΑ

Για να χρησιμοποιήσουμε τις μεθόδους της Στατιστικής, απαιτούνται αριθμητικά δεδομένα, που τα ονομάζουμε *Στατιστικά Στοιχεία*. Το πρόβλημα που έχουμε να λύσουμε αναφέρεται σε μια ομάδα ατόμων ή αντικειμένων, που θέλουμε να μελετήσουμε ή να βγάλουμε ορισμένα συμπεράσματα. Η ομάδα αυτή έχει ορισμένα ιδιαίτερα χαρακτηριστικά, π.χ. για κάθε άνθρωπο έχουμε γένος, ηλικία, βάρος, ύψος, χρώμα ματιών, ευφυΐα, εκπαίδευση, εισόδημα, κ.ά. Για ένα αντικείμενο, π.χ. αυτοκίνητο, έχουμε μάρκα, χρόνο κατασκευής, ιπποδύναμη, χρώμα, χιλιόμετρα που έχει κάνει κ.ά.

Τα χαρακτηριστικά κάθε ατόμου ή αντικειμένου μπορούν να χωριστούν σε ποιοτικά (γένος ή χρώμα) και σε ποσοτικά (ύψος, ιπποδύναμη). Τα ποσοτικά χαρακτηριστικά μπορούν να εκφραστούν με ακέραιους ή δεκαδικούς αριθμούς. Τα ποιοτικά με λέξεις ή με κάποια αριθμητική αντιστοιχία π.χ. άνδρας = 1, γυναίκα = 2, η οποία προκύπτει κατόπιν καποιας κωδικοποίησης.

Όταν μελετούμε στατιστικά ένα πρόβλημα, μπορούμε να αναφερθούμε σε ένα ή περισσότερα χαρακτηριστικά (ιδιότητες) μιας ομάδας ατόμων ή αντικειμένων.

**Κάθε ένα χαρακτηριστικό, όταν μετρηθεί ή εξετασθεί για κάθε άτομο ή αντικείμενο, θα δώσει μία ένδειξη (ποιοτική ή ποσοτική) που μετατρέπεται σε έναν αριθμό. Το σύνολο των αριθμών, που αντιστοιχούν στο ίδιο χαρακτηριστικό της ομάδας που μελετούμε, αποτελεί έναν στατιστικό πληθυσμό όπως η βαθμολογία των φοιτητών μιας Σχολής, το εισόδημα των κατοίκων μιας χώρας, κ.ά.**

Ένας **πληθυσμός** ανάλογα με το μέγεθός του μπορεί να είναι πολύ μεγάλος (άπειρος), μεγάλος ή μικρός. Ανάλογα με τον σκοπό της μελέτης που κάνουμε και τα μέσα που διαθέτουμε, συνήθως τα οικονομικά, και το μέγεθος του πληθυσμού, παίρνουμε ολόκληρο τον πληθυσμό ή ένα μέρος του (δηλ. ένα υποσύνολο), που το ονομάζουμε **δείγμα**.

Ο τρόπος που διαλέγουμε ένα δείγμα, το μέγεθος του δείγματος κ.ά., αποτελούν το περιεχόμενο του κλάδου της Στατιστικής, που λέγεται **δειγματοληψία**.

Όποιος και αν είναι ο τρόπος επιλογής του δείγματος, το σπουδαιότερο στοιχείο πρέπει να είναι η αμεροληψία. Ένα δείγμα λέγεται αμερόληπτο, όταν κάθε στοιχείο του πληθυσμού έχει τις ίδιες δυνατότητες (ή πιθανότητες) με όλα τα άλλα στοιχεία του πληθυσμού να αποτελέσει στοιχείο του δείγματος. Αν κάνουμε μια κοινωνική έρευνα, δεν πρέπει να πάρουμε τα ονόματα των ατόμων, που θα ρωτήσουμε, από τον τηλεφωνικό κατάλογο, γιατί θα υπάρξουν άτομα, τα οποία δεν έχουν τηλέφωνο, άρα δεν θα έχουν τη δυνατότητα να μπουν στο δείγμα και το δείγμα θα είναι μεροληπτικό.

## 2.2. ΤΡΟΠΟΙ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ ΚΑΙ ΜΕΓΕΘΟΣ ΔΕΙΓΜΑΤΟΣ

Ένα από τα σπουδαιότερα προβλήματα τα οποία παρουσιάζονται σε κάθε δειγματοληψία είναι και ο προσδιορισμός του μεγέθους του δείγματος. Οι παράγοντες οι οποίοι επηρεάζουν το μέγεθος του δείγματος, το οποίο φυσικά είναι τόσο καλύτερο όσο είναι μεγαλύτερο, είναι το κόστος για τη λήψη των στατιστικών δεδομένων, η ακρίβεια με την οποία ζητούμε τα αποτελέσματα και ο τρόπος δειγματοληψίας.

Οι πιο συνηθισμένοι τρόποι δειγματοληψίας είναι οι εξής:

1. **Απλή δειγματοληψία.** Παίρνουμε από όλο τον πληθυσμό με μέγεθος  $N$ , ένα δείγμα με μέγεθος  $n$  χρησιμοποιώντας τυχαίους αριθμούς ή ακόμη και κλήρωση.
2. **Κατευθυνόμενη δειγματοληψία.** Παίρνουμε ένα δείγμα με μέγεθος  $n$  από έναν πληθυσμό μεγέθους  $N$ , αλλά τα άτομα ή τα αντικείμενα, των οποίων η μέτρηση του χαρακτηριστικού το οποίο εξετάζουμε, θα μπει στο δείγμα πρέπει να πληρούν ορισμένες προϋποθέσεις ή κριτήρια. Δηλαδή θα γίνεται επιλογή αυτών οι οποίοι θα δώσουν τιμές για το δείγμα.

3. **Στρωματοποιημένη δειγματοληψία.** Ο πληθυσμός με βάση ορισμένο κριτήριο διαχωρίζεται σε  $K$  στρώματα ή  $K$  ομάδες. Από κάθε ομάδα  $i$  στρώμα ή  $i$  ομάδα παίρνουμε  $n_i$  μετρήσεις όπου τα  $n_i$  υπολογίζονται από τις σχέσεις:

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \dots = \frac{n_i}{N_i} = \dots = \frac{n_k}{N_k} = \frac{n}{N}$$

όπου  $N_i$  το μέγεθος κάθε στρώματος ή ομάδας,  $n$  το μέγεθος του δείγματος και  $N$  το μέγεθος του πληθυσμού.

4. **Γεωγραφική δειγματοληψία.** Η περιοχή στην οποία θα γίνει η δειγματοληψία χωρίζεται σε μικρότερες περιοχές. Οι μικρότερες περιοχές σε ακόμη μικρότερες κ.ο.κ. Ο τρόπος αυτός έχει πολλές ομοιότητες με τη στρωματοποιημένη δειγματοληψία π.χ. αν θέλουμε να πάρουμε ένα δείγμα στο οποίο να έχουμε αντιπροσώπευση από όλη την Ελλάδα, χωρίζουμε τη χώρα σε περιοχές. Από κάθε περιοχή διαλέγουμε με τυχαίο τρόπο μία ή περισσότερες πόλεις. Από κάθε πόλη διαλέγουμε με τυχαίο πάλι τρόπο ορισμένους δρόμους και από κάθε δρόμο ορισμένα κτίρια. Φυσικά ο αριθμός των δρόμων, κτιρίων κλπ. έχει σχέση με το μέγεθος της πόλης για να έχουμε τελικά μέγεθος δείγματος  $n_i$  από κάθε πόλη ανάλογο με τον πληθυσμό της.

Για όλους τους τρόπους δειγματοληψίας το κόστος ισούται με  $C = c \cdot n$ , όπου  $c$  το κόστος για τη λήψη μιας μέτρησης ή ενός ερωτηματολογίου. Το μέγεθος  $n$  υπολογίζεται και είναι ανεξάρτητο από το μέγεθος του πληθυσμού  $N$  με βάση το σφάλμα που δεχόμαστε στην εξαγωγή του συμπεράσματος, π.χ. αν διαθέτουμε 1.000.000 δρχ. και για κάθε ερωτηματολόγιο έχουμε κόστος 2000 δρχ. τότε φυσικά μπορούμε να πάρουμε δείγμα με μέγεθος  $n = 1.000.000/2000 = 50$ . Το μέγεθος του απαραίτητου δείγματος για να έχουμε σφάλμα  $\varepsilon\%$  θα υπολογιστεί από τους τύπους στην παράγραφο 8.13. Με τους ίδιους τύπους θα μπορούμε να υπολογίσουμε το σφάλμα  $\varepsilon$  αν τα χρήματα τα οποία διαθέτουμε για την δειγματοληψία επαρκούν για να πάρουμε δείγμα με ορισμένο μέγεθος.

### 2.3. ΠΗΓΕΣ ΣΤΑΤΙΣΤΙΚΩΝ ΣΤΟΙΧΕΙΩΝ

Τα στατιστικά στοιχεία τα συγκεντρώνουμε από διάφορες πηγές, που μπορούμε να τις χωρίσουμε σε πρωτογενείς ή δευτερογενείς.

Στις πρωτογενείς πηγές ανήκουν η απογραφή και η δειγματοληψία. Η απογραφή γίνεται κάθε ορισμένο χρονικό διάστημα (π.χ. 10 χρόνια), όταν πρόκειται για τα στοιχεία των κατοίκων ενός κράτους ή για συγκεκριμένο τομέα, π.χ. αγροτική, βιομηχανική απογραφή. Η δειγματοληψία, δηλαδή η λήψη ενός δείγματος για το οποίο έχουμε από πριν καθορίσει τον τρόπο που θα το πάρουμε και το μέγεθός του, γίνεται, όταν θέλουμε να βγάλουμε ένα συμπέρασμα ή να υπολογίσουμε το μέγεθος ενός χαρακτηριστικού, που μας ενδιαφέρει.

Τόσο στην απογραφή, όσο και στη δειγματοληψία συμπληρώνεται ένα στατιστικό δελτίο για το άτομο που ρωτιέται ή για το αντικείμενο που μας ενδιαφέρει (π.χ. στατιστικό δελτίο εγγραφής φοιτητών), που το λέμε ερωτηματολόγιο. Η κατασκευή του ερωτηματολογίου απαιτεί ομάδα επιστημόνων με ειδικές γνώσεις, γιατί πάνω σ' αυτό καθώς και στην εκλογή των ατόμων που ρωπιούνται, στηρίζεται η αξιοπιστία των συμπερασμάτων της εφαρμογής των στατιστικών μεθόδων.

Το ερωτηματολόγιο μπορεί να έχει πολλές ερωτήσεις και να συμπληρωθεί από το ίδιο το άτομο το οποίο ερωτάται ή από ειδικευμένο προσωπικό και οι ερωτήσεις πρέπει να είναι τέτοιας μορφής, ώστε να αποφεύγονται αναληθείς απαντήσεις.

Στις δευτερογενείς πηγές περιλαμβάνονται διάφορες υπηρεσίες, δημόσιες ή ιδιωτικές, Τράπεζες, δήμοι, κοινότητες, στις οποίες η συγκέντρωση των στατιστικών στοιχείων είναι κατά κάποιο τρόπο υποχρεωτική ή έμμεση. Οι πηγές αυτές δίδουν τα απαραίτητα στατιστικά στοιχεία, όταν χρειάζονται. Η Δημογραφία αποτελεί ένα σοβαρό κλάδο της Στατιστικής, που ασχολείται με τα στατιστικά στοιχεία των κατοίκων κάθε χώρας, τα οποία αναφέρονται στη γέννηση, οικογενειακή κατάσταση, θάνατο και γενικά στις διάφορες ληξιαρχικές πράξεις.

### 2.4. ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ - ΚΩΔΙΚΟΠΟΙΗΣΗ

Αν η συλλογή των στατιστικών στοιχείων γίνει με τη χρήση ερωτηματολογίου, πρέπει να φροντίσουμε να πετύχει ο σκοπός της συλλογής τους. Οι ερωτήσεις πρέπει να είναι απλές και να έχουν μια λογική σειρά. Οι απα-

νήσεις πρέπει να αποτελούνται από τις λέξεις ΝΑΙ ή ΟΧΙ, ένα σημείο (X) ή κενό, ή το πολύ από έναν αριθμό. Γενικά οι ερωτήσεις πρέπει να περιέχουν όλες τις δυνατές απαντήσεις και να είναι τέτοιες, ώστε να επιτρέπουν την σύνταξη πινάκων στατιστικών στοιχείων.

Μετά τη συμπλήρωση των ερωτηματολογίων, για να μετατραπούν τα στατιστικά στοιχεία σε αριθμητικά, απαιτείται κάποια κωδικοποίηση, δηλαδή η αμφιμονοσήμαντη (ένα προς ένα) αντιστοίχιση των ενδείξεων κάθε χαρακτηριστικού με αριθμούς.

Π.χ. Αν το χαρακτηριστικό που μελετούμε είναι οι γραμματικές γνώσεις του ατόμου που ρωτήθηκε, μπορούμε να έχουμε την εξής κωδικοποίηση:

0	=	αγράμματος
1	=	απόφοιτος Δημοτικού
2	=	απόφοιτος Γυμνασίου
3	=	απόφοιτος Λυκείου
4	=	πτυχιούχος Ανωτέρας Σχολής
5	=	πτυχιούχος Ανωτάτης Σχολής
6	=	μεταπτυχιακό δίπλωμα Master
7	=	Διδακτορικό δίπλωμα Ph.D.

Η κωδικοποίηση ενός ερωτηματολογίου πρέπει να έχει μελετηθεί κατά την σύνταξή του.

Η επεξεργασία των κωδικοποιημένων στοιχείων μέχρι το 1950 γινόταν με διάφορες μηχανές διάτρητων δελτίων, που τις σχεδίασε το 1890 ο Dr. Hollerith, για να επεξεργαστεί τα στατιστικά στοιχεία της απογραφής των Η.Π.Α. Ο ίδιος επινόησε και το διάτρητο δελτίο, πάνω στο οποίο με κωδική μορφή αποτυπώνονται με διατρήσεις οι χαρακτήρες (ψηφία, γράμματα, σύμβολα), που χρησιμοποιούνται για την παρουσίαση των στατιστικών στοιχείων. Οι μηχανές αυτές, που λέγονται κλασικές μηχανές, χρησιμοποιήθηκαν από όλες τις Στατιστικές Υπηρεσίες. Από το 1950, για την επεξεργασία των Στατιστικών στοιχείων, χρησιμοποιούνται αντί για τις κλασικές μηχανές, ηλεκτρονικοί υπολογιστές. Ο πρώτος Η/Υ κατασκευάστηκε από την εταιρεία UNIVAC για τη Στατιστική Υπηρεσία των ΗΠΑ, το 1950 και χρησιμοποιήθηκε για την επεξεργασία των στατιστικών στοιχείων της απογραφής του 1950. Σήμερα όλες οι Στατιστικές Υπηρεσίες καθώς και οι ιδιωτικές εταιρείες επεξεργάζονται τα στατιστικά στοιχεία διερεύνησης της γνώμης ή διαφόρων απογραφών χρησιμοποιούν τις καλύτερες μεθόδους και τους πλέον σύγχρονους Η/Υ.