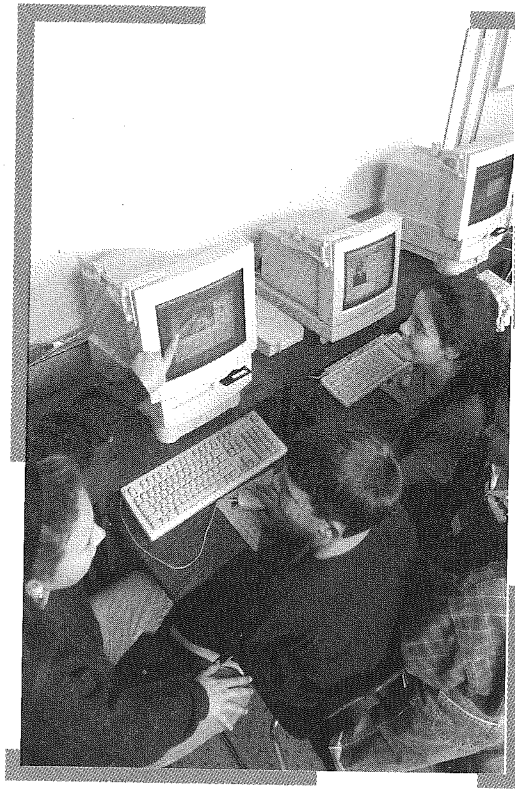
Ο Δημήτρης Καραγεώργος είναι μαθηματικός με διδακτορικές σπουδές στην Αγγλία. Υπηρέτησε τριάντα πέντε χρόνια στη Δευτεροβάθμια Εκπαίδευση ως καθηγητής, γυμνασιάρχης, λυκειάρχης, σχολικός σύμβουλος, μόνιμος πάρεδρος και σύμβουλος του Παιδαγωγικού Ινστιτούτου.
Σήμερα διδάσκει ως επίκουρος καθηγητής στο Πανεπιστήμιο Αθηνών (τομέας Παιδαγωγικής του τμήματος Φ.Π.Ψ.) τα μαθήματα «Διδακτική των θετικών επιστημών» και «Μεθοδολογία έρευνας στις επιστήμες της αγωγής».
Στη μακρόχρονη εκπαιδευτική του θητεία δίδαξε το μάθημα των μαθηματικών σε όλες τις τάξεις του Γυμνασίου και Λυκείου, ασχολήθηκε με τη σύνταξη αναλυτικών προγραμμάτων, τη συγγραφή διδακτικών εγχειριδίων, που διδάσκονται στη Δευτεροβάθμια Εκπαίδευση, την παραγωγή υποστηρικτικού διδακτικού υλικού για τα μαθηματικά και την επιμόρφωση των εκπαιδευτικών.
Έχει γράψει εννιά βιβλία μαθηματικών και έχει δημοσιεύσει σε ελληνικά και ξένα περιοδικά περισσότερες από πενήντα εργασίες που αναφέρονται στα προηγούμενα αντικείμενα και στον εκπαιδευτικό σχεδιασμό.
Ένα από τα θέματα που τον απασχόλησαν την τελευταία εικοσαετία ήταν και η εκπαιδευτική έρευνα.
Ένα μέρος της προσπάθειάς του αυτής καταγράφεται σε αυτό το βιβλίο.

Δημήτρης Λ. Καραγεώργος
ΕΠΙΚΟΥΡΟΣ ΚΑΘΗΓΗΤΗΣ ΠΑΝΕΠΙΣΤΗΜΙΟΥ ΑΘΗΝΩΝ
Επιτ. ΣΥΜΒΟΥΛΟΣ ΤΟΥ ΠΑΙΔΑΓΩΓΙΚΟΥ ΙΝΣΤΙΤΟΥΤΟΥ

ΜΕΘΟΔΟΛΟΓΙΑ ΕΡΕΥΝΑΣ
στις επιστήμες της αγωγής
ΜΙΑ ΔΙΔΑΚΤΙΚΗ ΠΡΟΣΕΓΓΙΣΗ



Σαββάλας
ΕΚΔΟΣΕΙΣ

3.4 Το μέγεθος του δείγματος

Είπαμε στην § 3.2 ότι η αντιπροσωπευτικότητα του δείγματος καθορίζεται από το μέγεθός του και από τον τρόπο επιλογής του.

Τίθεται απλά το ερώτημα: Πόσο μεγάλο πρέπει να είναι ένα δείγμα για να θεωρείται ικανοποιητικό; Στο ερώτημα αυτό δεν υπάρχει σαφής απάντηση. Όπως είπαμε στην ίδια παράγραφο, έχει διαπιστωθεί από δειγματοληπτικές έρευνες ότι όσο μεγαλύτερο είναι το μέγεθος του δείγματος, τόσο πιο αξιόπιστα είναι τα συμπεράσματα που εξάγονται από το δείγμα, αρκεί να έχει ληφθεί με τον ενδεδειγμένο τρόπο, δηλαδή να διασφαλίζεται η τυχαιότητα της επιλογής.

Το μέγεθος του δείγματος είναι συνάρτηση πολλών παραγόντων, όπως:

α) Της **ανομοιογένειας του πληθυσμού**. Είναι κατανοητό ότι όσο πιο ανομοιογενής είναι ο πληθυσμός ως προς το χαρακτηριστικό που μελετάμε στα άτομα, τόσο μεγαλύτερο πρέπει να είναι το δείγμα για να καλύψουμε όλες τις περιπτώσεις.

β) Του **βαθμού ακρίβειας** που προκαθορίζουμε να έχουμε στην επέκταση των ευρημάτων σε όλο τον πληθυσμό. Όσο μικρότερη είναι η οριζόμενη από εμάς ανεκτή διαφορά μεταξύ στατιστικών του δείγματος και των αντίστοιχων παραμέτρων του πληθυσμού τόσο μεγαλύτερο πρέπει να είναι το μέγεθος του δείγματος.

γ) Της **σχέσης** που υπάρχει **μεταξύ του μεγέθους του δείγματος ν και του τυπικού σφάλματος**. Αυτή τη σχέση θα τη δούμε στις επόμενες παραγράφους και θα επισημάνουμε ότι τα μεγέθη αυτά είναι αντιστρόφως ανάλογα και μάλιστα το τυπικό σφάλμα είναι αντιστρόφως ανάλογο του $\sqrt{\nu}$.

Όπως θα δούμε και παρακάτω, το μέγεθος του δείγματος πρέπει να φροντίζουμε να είναι μεγαλύτερο του 30, δηλαδή $\nu > 30$, γιατί τότε η δειγματική κατανομή του στατιστικού προσεγγίζει την κανονική κατανομή.

Εκτός από το **τυπικό σφάλμα** που είναι στατιστικό σφάλμα, έχουμε και το **σφάλμα μεροληψίας**, το οποίο οφείλεται στην μεροληπτική και όχι στην τυχαία επιλογή του δείγματος.

Όταν επιλέγεται ένα ενιαίο δείγμα από τον πληθυσμό, τα ποσοστά του δείγματος δίνονται έτοιμα από τους στατιστικολόγους σε πίνακες μαζί με το αντίστοιχο ποσοστό σφάλματος (Κασσωτάκης Μ., 1978).

Παραθέτουμε στη συνέχεια δύο πίνακες όπου ο πρώτος (πίνακας 3.1) δείχνει το μέγεθος που πρέπει να έχει το δείγμα με κίνδυνο σφάλματος 2% σε 99 από τα 100 δείγματα και ο δεύτερος (πίνακας 3.2) το μέγεθος του δείγματος με σφάλμα 1% σε 997 από τα 1000 δείγματα. Θεωρούμε ότι τα δείγματα είναι τυχαία.

Πίνακας 3.1

Αριθμός στοιχείων πληθυσμού (N)	Αριθμός στοιχείων δείγματος (ν)	Εκατοστιαία αναλογία $\left(\frac{\nu}{N}\cdot 100\right)$
200	110	55
500	190	38
1000	260	26
2000	320	16
5000	380	7,6
10.000	420	4,2
20.000	440	2,2
50.000	450	0,9
100.000	500	0,5

Πίνακας 3.2

Αριθμός στοιχείων πληθυσμού (N)	Αριθμός στοιχείων δείγματος (ν)	Εκατοστιαία αναλογία $\left(\frac{\nu}{N}\cdot 100\right)$
200	180	90
500	360	72
1000	560	56
2000	800	40
5000	1000	20
10.000	1200	12
20.000	1400	7
50.000	1500	3
100.000	2000	2

Όσον αφορά το πλήθος ν του δείγματος θα πρέπει να αναφέρουμε την περίπτωση όπου το δείγμα είναι δύο κατηγοριών, για παράδειγμα μαθητές από αστικές και αγροτικές περιοχές, οπότε σε αυτή την περίπτωση χρησιμοποιείται ο επόμενος τύπος για τον προσδιορισμό του. Ισχύει δηλαδή ότι $\nu = \left(\frac{Z}{E}\right)^2 pq$, όπου:

ν	το μέγεθος του δείγματος,
Z	η z-τιμή, που αντιστοιχεί σε προκαθορισμένο επίπεδο εμπιστοσύνης,
E	το σφάλμα δειγματοληψίας,
p	το ποσοστό υποκειμένων της μιας κατηγορίας,
q	το ποσοστό υποκειμένων της άλλης κατηγορίας και
$p + q = 1$	από τις ιδιότητες των πιθανοτήτων.

Για παράδειγμα, αν:

α) το επίπεδο εμπιστοσύνης είναι 99%, τότε $Z = 2{,}58$, δηλαδή το δείγμα θα είναι αντιπροσωπευτικό του πληθυσμού με πιθανότητα 99%,

β) $E = 0{,}05$ το σφάλμα, δηλαδή η διαφορά την οποία αναμένουμε να έχει η μέση τιμή $\overline{X}$ του δείγματος από την αντίστοιχη μ του πληθυσμού,

γ) $p = 0{,}4$ το ποσοστό της μιας από τις δύο κατηγορίες και

δ) $q = 0{,}6$ το ποσοστό της άλλης κατηγορίας, τότε

$$\nu = \left(\frac{2{,}58}{0{,}05}\right)^2 \cdot 0{,}4 \cdot 0{,}6 = 639 \text{ άτομα}$$

Επομένως με δείγμα 639 ατόμων δεν θα έχουμε σφάλμα δειγματοληψίας μεγαλύτερο ή μικρότερο του 5%.

Το τυπικό σφάλμα $\sigma_{\overline{X}}$, όπως θα δούμε αμέσως παρακάτω, πολλές φορές υπολογίζεται από τον τύπο $\sigma_{\overline{X}} = \sqrt{\frac{pq}{\nu}}$ ή $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{\nu}}$, όπου σ η τυπική απόκλιση της κατανομής του πληθυσμού. Όταν δεν γνωρίζουμε την τυπική απόκλιση σ του πληθυσμού, χρησιμοποιούμε την τυπική απόκλιση s του δείγματος. Έτσι έχουμε το τυπικό σφάλμα $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{\nu}}$ ή $\sigma_{\overline{X}} = \frac{s}{\sqrt{\nu}}$, οπότε η z-τιμή είναι

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{\nu}}} \tag{1}$$

και αφού ορίσαμε ως σφάλμα δειγματοληψίας $E = \overline{X} - \mu$, από τη σχέση (1) έχουμε:

$$E = \overline{X} - \mu = Z \cdot \frac{\sigma}{\sqrt{\nu}} \quad \text{ή} \quad E = \overline{X} - \mu = Z \cdot \frac{s}{\sqrt{\nu}}$$

Δηλαδή τα όρια εμπιστοσύνης του μέσου όρου μ του πληθυσμού θα είναι:

$$\text{Ανώτατο όριο:} \quad L_1 = \overline{X} + Z \cdot \frac{s}{\sqrt{\nu}} \quad \text{και}$$

$$\text{Κατώτατο όριο:} \quad L_2 = \overline{X} - Z \cdot \frac{s}{\sqrt{\nu}}$$

Οι αντίστοιχοι τύποι για τον υπολογισμό των ορίων εμπιστοσύνης ενός ποσοστού είναι:

$$L_1 = \overline{p} + z \cdot \sqrt{\frac{pq}{\nu}} \quad \text{και} \quad L_2 = \overline{p} - z \cdot \sqrt{\frac{pq}{\nu}}$$

Για παράδειγμα, αν βρέθηκε ότι το ποσοστό υπέρ μιας θέσης, σε μια έρευνα με δείγμα μεγέθους $\nu = 400$, είναι $p = 40\%$, τότε το τυπικό σφάλμα του ποσοστού θα είναι

$$\sigma_{\overline{p}} = \sqrt{\frac{pq}{\nu}} = \sqrt{\frac{40 \cdot 60}{400}} \simeq 2{,}45$$

οπότε το διάστημα εμπιστοσύνης θα είναι

$$(L_2, L_1) = \left(\overline{p} - z \cdot \sqrt{\frac{pq}{\nu}},\ \overline{p} + z \cdot \sqrt{\frac{pq}{\nu}}\right) = (35{,}2,\ 44{,}8)$$

με πιθανότητα 95%. Ο αριθμός 1,96 είναι η z-τιμή που αντιστοιχεί σε πιθανότητα 95%, δηλαδή σε α-επίπεδο 5%.

3.5 Δειγματική κατανομή

Υπενθυμίζουμε ότι ένα από τα πιο σοβαρά προβλήματα της Επαγωγικής Στατιστικής είναι να προσδιορίσουμε τις παραμέτρους της κατανομής ενός πληθυσμού (π.χ. τη μέση τιμή μ, την τυπική απόκλιση σ κ.λπ.) από τις αντίστοιχες παραμέτρους (τις οποίες λέμε στατιστικά) δειγμάτων που παίρνουμε, με τους τρόπους που περιγράψαμε προηγουμένως, από τον πληθυσμό. Δηλαδή να εκτιμήσουμε τα άγνωστα μ και σ από τα γνωστά $\overline{x}$ και s. Τα στατιστικά

$$\overline{x} = \frac{1}{\nu}\sum x_i \quad \text{και} \quad \sigma = \sqrt{\frac{1}{\nu}\sum (x_i - \overline{x})^2}$$

είναι συναρτήσεις των δεδομένων x_i κάθε δείγματος.

Αν από έναν πληθυσμό μεγέθους N πάρουμε δύο δείγματα μεγέθους ν, τότε για το καθένα θα έχουμε έναν μέσο όρο $\overline{x}$ και μια τυπική απόκλιση s. Σπάνια οι μέσες τιμές των δειγμάτων και οι τυπικές αποκλίσεις συμπίπτουν.