

Συγκοινωνιακές Έρευνες

Οι Συγκοινωνιακές Έρευνες βασίζονται στη συλλογή στοιχείων αντιπροσωπευτικών του πληθυσμού.

Δειγματοληψία:

Κατά τη δειγματοληψία παρουσιάζονται δύο τύποι σφάλματος:

(I) Σφάλμα που οφείλεται στη τυχαιότητα (δείγμα από τον εξεταζόμενο πληθυσμό): δεν επηρεάζει τις μέσες τιμές των εκτιμήσεων των παραμέτρων, αλλά μόνο τη διασπορά (δηλ. το επίπεδο εμπιστοσύνης). Το σφάλμα αυτό **συσχετίζεται** με το μέγεθος του δείγματος

(II) Σφάλμα μεροληψίας (σφάλμα στον καθορισμό του πληθυσμού, σφάλμα μεθοδολογίας δειγματοληψίας). Αυτός ο τύπος σφάλματος, **επηρεάζει** τις εκτιμώμενες τιμές των παραμέτρων

Μέγεθος δείγματος:

-Εξαρτάται από τον συμβιβασμό (συγκερασμό) κόστους δειγματοληψίας και επιθυμητής ακρίβειας

-Επηρεάζει την εκτίμηση των τιμών των παραμέτρων μέσω των στατιστικών τύπων που εφαρμόζονται.

-Το μέγεθος του πληθυσμού δεν επηρεάζει το μέγεθος του δείγματος εκτός εξαιρέσεων (πολύ μικρών πληθυσμών)

-Βασίζεται στο Κεντρικό Οριακό Θεώρημα (ΚΟΘ) που ορίζει ότι οι εκτιμήσεις των μέσων τιμών του δείγματος τείνουν να ακολουθήσουν την κανονική κατανομή καθώς το μέγεθος του δείγματος (n) αυξάνει (δηλ. λαμβάνει τιμές >30). Ισχύει και για μικρότερα δείγματα αν μπορεί να υποθεθεί ότι ο πληθυσμός ακολουθεί κανονική κατανομή.

Έστω πληθυσμός μεγέθους N του οποίου κάποιο χαρακτηριστικό έχει μέση τιμή μ και διασπορά σ^2 . Σύμφωνα με το ΚΟΘ η βέλτιστη εκτίμηση της μέσης τιμής είναι $\mu = \bar{x}$ και της διασποράς σ^2 η S^2 (η διασπορά του δείγματος). Σε αυτή τη περίπτωση, το τυπικό σφάλμα της μέσης τιμής του δείγματος ισούται με

$$s(\bar{x}) = \sqrt{(N - n)S^2 / nN}.$$

Για μεγάλους πληθυσμούς και μικρά δείγματα το $(N - n)/N \rightarrow 1$ και επομένως

$$s(\bar{x}) = S/\sqrt{n}$$

Για τον υπολογισμό του n επιλύεται η παρακάτω σχέση:

$$n = S^2 / s(\bar{x})^2$$

Η πιο πάνω διαδικασία παρουσιάζει δύο προβλήματα: την εκτίμηση της διασποράς S^2 του δείγματος και την επιλογή ενός αποδεκτού τυπικού σφάλματος για τη μέση τιμή.

Το πρώτο είναι προφανές και οδηγεί στο συμπέρασμα ότι αφού πρέπει να υπολογιστεί αφού παρθεί το δείγμα, θα πρέπει να υπολογιστεί από κάποια άλλη πηγή.

Το δεύτερο συναρτάται με το επιδιωκόμενο **επίπεδο εμπιστοσύνης** του μέσου το δείγματος ως εκτίμηση του πραγματικού μέσου του πληθυσμού. Συνήθως δεν καθορίζεται κάποιο συγκεκριμένο τυπικό σφάλμα αλλά ένα **διάστημα** περί τον μέσο για κάποιο επίπεδο εμπιστοσύνης.

Καταρχήν πρέπει να καθοριστεί ένα επίπεδο εμπιστοσύνης για το διάστημα που αναφέρεται στην αποδοχή σφάλματος στον υπολογισμό του μέσου όρου, π.χ. ένα τυπικό επίπεδο 95% συνεπάγεται την αποδοχή σφάλματος στο 5% των περιπτώσεων.

Επίσης, απαιτείται ο καθορισμός των ορίων του επιπέδου εμπιστοσύνης περί τον μέσον, είτε σε απόλυτες ή σε σχετικές τιμές, π.χ. ως ποσοστό της τιμής του μέσου. Γι' αυτόν τον λόγο, το μέγεθος του δείγματος εκφράζεται συνήθως ως συνάρτηση του αναμενόμενου συντελεστή απόκλισης ($\Sigma A = \sigma/\mu$).

Για παράδειγμα, ένα υποτεθεί ότι ο πληθυσμός ακολουθεί κανονική κατανομή και ληφθεί επίπεδο εμπιστοσύνης ίσο με 95%, αυτό θα συνεπάγεται ότι γίνονται αποδεκτές τιμές έως $1,96s(\bar{x})$ αφού η περιοχή $\mu \pm 1,96\sigma$ (κάτω από την καμπύλη της συνάρτησης πιθανότητας) περιέχει το 95% της πιθανότητας της κανονικής κατανομής. Αν παράλληλα καθοριστεί και ένα ποσοστό σφάλματος στην εκτίμηση του μέσου όρου, π.χ. ίσο με 10%, τότε το διάστημα καθορίζεται ως $\mu \pm 0,1\mu$ και ισχύει

$$s(\bar{x}) = 0,1\mu/1,96 = 0,051\mu$$

και με αντικατάσταση στην προηγούμενη σχέση εκτίμησης του δείγματος προκύπτει ότι

$$n = \left(\frac{S}{0,051\mu}\right)^2 = 384\Sigma A^2$$

Συγκοινωνιακές μελέτες

Η εκτίμηση του μεγέθους δείγματος στις συγκοινωνιακές έρευνες απαιτεί ακριβή γνώση των μεταβλητών που πρέπει να εκτιμηθούν, των αντίστοιχων συντελεστών απόκλισης, της απαιτούμενης ακρίβειας εκτίμησης και του επιπέδου εμπιστοσύνης.

Ο πρώτος παράγων είναι πολύ σημαντικός και συνήθως περιβάλλεται από ασάφεια στην ακριβή επιδίωξη της έρευνας. Αν π.χ. απαιτείται μόνο η εκτίμηση των συντελεστών παραγωγής μετακινήσεων, τότε δείγματα μεγέθους 1000 απογραφομένων εγγυάται **επίπεδα εμπιστοσύνης** της τάξης του 90% για **διαστήματα (ποσοστά) σφάλματος** εκτίμησης ίσα με 5%.

Η παραπάνω εκτίμηση μεγέθους δείγματος δυσμενοποιείται στη περίπτωση των ερευνών Προέλευσης – Προορισμού ιδιαίτερα δε στη περίπτωση που επιδιώκεται ‘αντιπροσωπευτικότητα’ για κάθε στοιχείο του μητρώου των μετακινήσεων. Για παράδειγμα, αν το κάθε στοιχείο του μητρώου έχει περί τις 1000 μετακινήσεις, τότε αποδεικνύεται ότι ένα δείγμα της τάξης λίγο μεγαλύτερου του 4% (δηλ. περίπου 4,3%) εγγυάται σφάλματα μικρότερα του 25% με επίπεδο εμπιστοσύνης 90%. Εάν όμως το πλήθος των μετακινήσεων ανάμεσα στις ζώνες της περιοχής μελέτης είναι της τάξης των 20-30, τότε για να επιτευχθεί το ίδιο επίπεδο ακρίβειας θα απαιτείτο μέχρι και δείγμα 100% (το σύνολο δηλαδή του πληθυσμού!).

Το πρόβλημα απλουστεύεται αν υπάρχουν στοιχεία για τον συντελεστή απόκλισης. Για τον λόγο αυτόν, θα μπορούσε να αντληθεί σχετική εμπειρία από την πληθώρα προηγούμενων συγκοινωνιακών μελετών που διεξήγαν παρόμοιες έρευνες Προέλευσης – Προορισμού.

Αν οι τρεις καθοριστικοί παράγοντες (ΣA , ποσοστό σφάλματος και επίπεδο εμπιστοσύνης) είναι γνωστοί, τότε μπορεί να εκτιμηθεί το μέγεθος του δείγματος με βάση τον παρακάτω τύπο:

$$n = \frac{\Sigma A^2 Z_{\alpha}^2}{E^2}$$

όπου ΣA όπως ορίστηκε προηγούμενα, Z_{α} η τιμή της τυπικής κανονικής κατανομής για επίπεδο εμπιστοσύνης α και E το ποσοστό σφάλματος (που προσδιορίζει το επιδιωκόμενο επίπεδο ακρίβειας εκτίμησης).

Παράδειγμα 1: Να εκτιμηθεί το μέγεθος του δείγματος που απαιτείται προκειμένου να εκτιμηθεί ο αριθμός των μετακινήσεων ανά νοικοκυριό, με $\Sigma A=1,0$, ποσοστό σφάλματος 5% και επίπεδο εμπιστοσύνης 95%:

$$n = \frac{(1,0)^2 (1,645)^2}{(0,05)^2} = 1082$$

Παράδειγμα 2: Αν στο παραπάνω παράδειγμα αυξηθεί το ΣA κατά 10%, δηλ. γίνει 1,10, το μέγεθος δείγματος αυξάνεται κατά 21%:

$$n = \frac{(1,1)^2 (1,645)^2}{(0,05)^2} = 1310$$