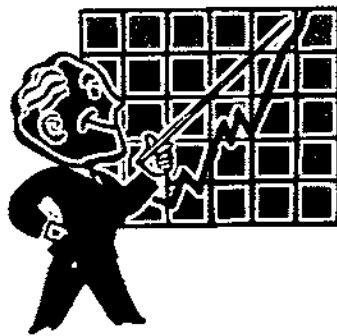




ΤΕΙ ΠΑΤΡΩΝ
ΤΜΗΜΑ: ΛΟΓΙΣΤΙΚΗΣ
Πτυχιακή εργασία

ΘΕΜΑ:

ΜΕΘΟΔΟΛΟΓΙΑ ΔΙΕΞΑΓΩΓΗΣ ΣΤΑΤΙΣΤΙΚΩΝ ΕΡΕΥΝΩΝ



Καθηγήτρια:

Κα ΒΑΘΗ ΠΑΝΑΓΙΩΤΑ

Σπουδαστές:

ΔΗΜΑ ΙΩΑΝΝΑ

ΔΕΥΤΟΣ ΔΗΜΟΣΘΕΝΗΣ



ΠΑΤΡΑ 2002

ΠΕΡΙΕΧΟΜΕΝΑ

	Σελ.
Εισαγωγή	3
Κεφάλαιο 1 ^ο	4
Κεφάλαιο 2 ^ο	49
Κεφάλαιο 3 ^ο	55
Κεφάλαιο 4 ^ο	64
Κεφάλαιο 5 ^ο	82
Βιβλιογραφία	92

ΕΙΣΑΓΩΓΗ

Η εργασία μας αποτελείται από πέντε (5) κεφάλαια, στα οποία διαπραγματευόμαστε τη μεθοδολογία διεξαγωγής ερευνών.

Στο 1^ο κεφάλαιο αναφέρουμε την Μεθοδολογία σχεδιασμού των ερευνών και συγκεκριμένα, καταγράφουμε σε γενικές γραμμές τα στάδια διενέργειας μιας έρευνας, την επιλογή δείγματος, την μέθοδο συγκέντρωσης του απαιτούμενου στατιστικού υλικού και τις πηγές συλλογής πληροφοριών, την διατύπωση των ερωτήσεων, το χρονοδιάγραμμα και τον προϋπολογισμό της έρευνας και την επεξεργασία των στατιστικών στοιχείων.

Στο 2^ο κεφάλαιο αναφέρουμε τις Βασικές Αρχές της Δειγματοληψίας καθώς η επιλογή του δείγματος αποτελεί το σπουδαιότερο πρόβλημα στο στάδιο του σχεδιασμού της έρευνας.

Έτσι στο 3^ο κεφάλαιο αναλύουμε τις Δειγματοληπτικές Μεθόδους αναφέροντας πλεονεκτήματα, μειονεκτήματα και παραδείγματα για κάθε μία από αυτές.

Στο 4^ο κεφάλαιο με τίτλο συλλογή στατιστικών στοιχείων ασχολούμαστε με τους τρόπους παρουσίασης των στοιχείων μιας έρευνας.

Τέλος, στο 5^ο κεφάλαιο αναφερόμαστε στα φράγματα και παρουσιάζουμε την γενική μέθοδο κατασκευής διαστημάτων εμπιστοσύνης.

ΚΕΦΑΛΑΙΟ 1

ΜΕΘΟΔΟΛΟΓΙΑ ΣΧΕΔΙΑΣΜΟΥ ΤΩΝ ΕΡΕΥΝΩΝ

1. ΣΤΑΔΙΑ ΔΙΕΝΕΡΓΕΙΑΣ ΜΙΑΣ ΕΡΕΥΝΑΣ

Για να διενεργήσουμε μια επιστημονική έρευνα πρέπει να ακολουθήσουμε βασικά στάδια προκειμένου να έχουμε αξιόπιστα αποτελέσματα.

Τα στάδια αυτά είναι κοινά σε κάθε έρευνα, ελάχιστα διαφέρουν και είναι:

- 1,1 Σχεδιασμός της έρευνας
- 1,2 Επιλογή του δείγματος
- 1,3 Επιλογή της μεθόδου συγκέντρωσης του απαιτούμενου στατιστικού υλικού
- 1,4 Σύνταξη ερωτηματολογίου
- 1,5 Προετοιμασία του προσωπικού
- 1,6 Χρονοδιάγραμμα της έρευνας
- 1,7 Προϋπολογισμός της έρευνας
- 1,8 Δοκιμαστική έρευνα
- 1,9 Στατιστική ανάλυση και ερμηνεία των συγκεντρωθέντων εμπειρικών δεδομένων
- 1,10 Παρουσίαση της ερευνητικής μελέτης

1,1 Σχεδιασμός και προγραμματισμός της έρευνας

Το πρώτο και βασικό στάδιο για την διενέργεια μίας έρευνας είναι ο σχεδιασμός και προγραμματισμός της, που θα καθορίσει την στρατηγική που θα ακολουθήσει η ερευνητική ομάδα για την υλοποίηση της έρευνας. Καταρχήν η επιστημονική ομάδα που θα ασχοληθεί με την έρευνα θα

πρέπει να είναι σχετική με το θέμα και να περιλαμβάνει και έναν στατιστικό με εμπειρία στην διεξαγωγή ερευνών.

Η ομάδα αυτή θα πρέπει να προσδιορίσει τη φύση του προβλήματος καθώς και να διατυπώσει τον αντικειμενικό σκοπό της. Ακόμα θα πρέπει να μελετήσει παρόμοιες έρευνες που έχουν γίνει στο παρελθόν αλλά και την βιβλιογραφία που υπάρχει.

Στο σχεδιασμό μιας έρευνας, βασική θέση έχει το χρονοδιάγραμμα από τη στιγμή της ανάληψης της έρευνας μέχρι τη δημοσίευση των αποτελεσμάτων της, ο προϋπολογισμός της, η σύνταξη των ερωτηματολογίων, η μέθοδος συλλογής των στοιχείων και η δοκιμαστική έρευνα.

Το στάδιο του σχεδιασμού είναι βασικό για την σωστή έκβαση μιας έρευνας γιατί αν δεν έχει μελετηθεί σωστά, μπορεί αργότερα στο στάδιο της υλοποίησης και της στατιστικής ανάλυσης να φανεί ότι ο σχεδιασμός ήταν ελλιπής ή ανεφάρμοστος.

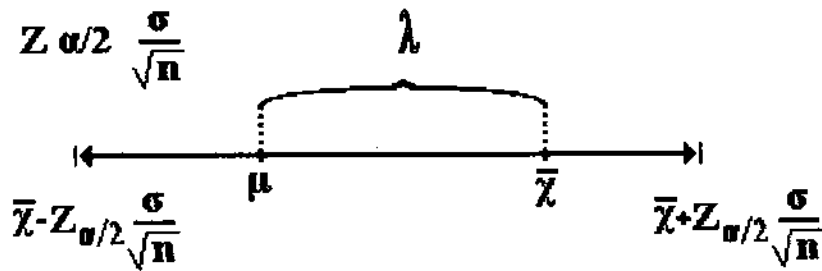
1,2 Επιλογή δείγματος

Ο προσδιορισμός του μεγέθους του δείγματος αποτελεί το σπουδαιότερο πρόβλημα στο στάδιο του σχεδιασμού της έρευνας.

Το μέγεθος του δείγματος εξαρτάται από τον σκοπό της έρευνας, από το πόσο ακριβής θέλουμε να είναι η δειγματική εκτίμηση ή από το σφάλμα δειγματοληψίας. Το σφάλμα δειγματοληψίας επηρεάζεται από το μέγεθος του δείγματος, από την ομοιογένεια και την κατανομή του πληθυσμού και το βαθμό εμπιστοσύνης. Όσο μεγαλύτερο είναι το μέγεθος του δείγματος και ομοιογενές τόσο καλύτερα αποτελέσματα έχουμε.

Ο ερευνητής μπορεί να προσδιορίζει, η εκτίμησή του να βρίσκεται μέσα σε κάποια όρια με πιθανότητα συνήθως 95% ή 99% και με

δειγματοληπτικό σφάλμα λ το οποίο δεν μπορεί να είναι με συντελεστή εμπιστοσύνης $(1-\alpha)100\%$ μεγαλύτερο της ποσότητας



Αν $x=\mu$ τότε το σφάλμα θα είναι $\lambda=0$. Αν ο μ βρίσκεται στο ανώτερο όριο του διαστήματος εμπιστοσύνης $(\bar{x}+Z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$ θα έχουμε:

$$\lambda=Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \text{ (σχ. 1.1.)}$$

λύνοντας ως προς n θα έχουμε

$$\frac{Z_{\alpha/2}^2 \sigma^2}{\lambda^2}$$

Η τιμή του μέσου μ μπορεί να βρίσκεται σε οποιαδήποτε θέση στο διάστημα αυτό. Δηλ. αν θέλουμε η εκτίμηση x του μέσου αριθμητικού μ να διαφέρει της πραγματικής τιμής του μ λιγότερο της τιμής λ και με συντελεστή εμπιστοσύνης $1-\alpha$

$$P[|x-\mu|<\lambda] = 1-\alpha$$

Τότε θα πάρουμε δείγμα μεγέθους:

$$\frac{Z_{\alpha/2}^2 \sigma^2}{\lambda^2}$$

Η σχέση αυτή προσδιορίζει το μέγιστο μέγεθος με το οποίο εξασφαλίζεται πλάτος διαστήματος λ για επίπεδο εμπιστοσύνης $1-\alpha$.

Για την τιμή της διακύμανσης σ^2 του πληθυσμού, ο ερευνητής μπορεί να έχει ενδείξεις ή πληροφορίες από προηγούμενες έρευνες.

Το δειγματοληπτικό σφάλμα λ για δεδομένη τιμή του βαθμού εμπιστοσύνης είναι τόσο μεγαλύτερο, όσο μεγαλύτερη είναι η διασπορά του σ^2 .

Παράδειγμα 1.1

Επιθυμούμε να εκτιμήσουμε το μέσο βάρος των παραγόμενων πεπονιών σε μια πειραματική καλλιέργεια. Ποιο πρέπει να είναι το απαιτούμενο μέγεθος του τυχαίου δείγματος, ώστε με πιθανότητα 99% το δειγματοληπτικό σφάλμα να μην είναι μικρότερο του ενός κιλού, όταν η διακύμανση είναι $\sigma^2=4$;

Λύση:

$$1-\alpha=0,99, \alpha=0,01, z_{\alpha/2}=2,58, \lambda=1$$

Θα έχουμε:

$$1 = 2,58 \frac{4}{\sqrt{n}} \Rightarrow n = 2,58^2 \frac{2^2}{1} = 26,6 \cong 27$$

Παράδειγμα 1.2

Μια εταιρεία ερευνών, που επιθυμεί να βρει το μέσο ετήσιο οικογενειακό εισόδημα των κατοίκων μιας πόλης, αποφασίζει να πάρει ένα δείγμα από οικογένειες που είναι εγγεγραμμένες σε ορισμένους καταλόγους. Ποιο μέγεθος δείγματος θα πάρει, ώστε το σφάλμα εκτίμησης $|\bar{x} - \mu|$ να μην υπερβαίνει τις 10 μονάδες, δεδομένου ότι $\sigma=50$, το δε επίπεδο σημαντικότητας είναι 5%;

Λύση:

Το μέγεθος του δείγματος δίνεται από την παρακάτω σχέση:

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{\lambda^2} = \frac{(1,96)^2 (50)^2}{10^2} = 96$$

Στην παραπάνω περίπτωση το σφάλμα $\lambda = 10 = |\bar{x} - \mu|$ δεν μπορεί να υπερβεί, με βαθμό εμπιστοσύνης $(1-\alpha)100\%$, ένα προσδιορισμένο επίπεδο λ , όταν το μέγεθος του δείγματος είναι 96 μονάδες και δίνεται από τη σχέση:

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{\lambda^2}$$

Προσδιορισμός μεγέθους δείγματος για εκτίμηση ποσοστών

α) Αν θέλουμε η εκτίμηση \hat{p} του ποσοστού P να διαφέρει της πραγματικής τιμής P λιγότερο ενός αριθμού λ με πιθανότητα $1-\alpha$ δηλ.

$$\Pr (|\hat{p}-P| \leq \lambda=1-\alpha$$

τότε θα πάρουμε δείγμα μεγέθους

$$n = \frac{Z_{\alpha/2}^2}{\lambda^2} \cdot P(1-P)$$

όπου P το ποσοστό του πληθυσμού που συνήθως είναι άγνωστο αλλά μπορούμε να πάρουμε ενδείξεις ή πληροφορίες από προηγούμενες έρευνες ή να πάρουμε $P=0,50$

β) Αν θέλουμε η εκτίμηση \hat{p} του ποσοστού P να διαφέρει της πραγματικής τιμής P λιγότερο από ένα ποσοστό κ της άγνωστης παραμέτρου P , με πιθανότητα $1-\alpha$ δηλ.

$$\Pr (|\hat{p}-P| < \kappa P=1-\alpha$$

Τότε θα πάρουμε δείγμα μεγέθους

$$n = \frac{Z_{\alpha/2}^2}{\kappa^2} \cdot P(1-P)$$

Παράδειγμα 1.3

Έστω ότι θέλουμε να πάρουμε ένα τυχαίο δείγμα για να εκτιμήσουμε το ποσοστό των κατοίκων μιας πόλης που παρακολουθούν ένα πρόγραμμα στην τηλεόραση. Η ζητούμενη ακρίβεια θα είναι ίση με $\pm 0,05$ και το επίπεδο εμπιστοσύνης 95%. Ποιο μέγεθος δείγματος πρέπει να πάρουμε;

α) Αν δεν γνωρίζουμε το ποσοστό των ατόμων που παρακολουθούν το πρόγραμμα;

β) Όταν γνωρίζουμε ότι το ποσοστό των ατόμων που έχουν παρακολουθήσει παρόμοια προγράμματα στο παρελθόν ήταν 25%

Λύση :

α)

$$k = |\hat{p} - P| = 0,05$$

$$n = Z_{\alpha/2}^2 \cdot \frac{P(1-P)}{K^2}$$

Επειδή στην πρώτη περίπτωση δε γνωρίζουμε P του πληθυσμού, θα χρησιμοποιήσουμε την τιμή $P = \frac{1}{2}$ ώστε το γινόμενο $P(1-P)$ να πάρει την πιο μεγάλη τιμή, για να έχουμε μια καλή εκτίμηση.

Η μέγιστη τιμή του γινομένου $P(1-P)$ είναι 0,25 και προκύπτει όταν $P=0,5$ και $(1-P)=0,5$ (γιατί $P+Q=1$).

Επομένως θα έχουμε:

$$n = Z_{\alpha/2}^2 \cdot \frac{P(1-P)}{K^2} = (1,96)^2 \frac{(\frac{1}{2})(\frac{1}{2})}{(0,05)^2} = 359$$

β) Όταν το P είναι γνωστό, δηλαδή $P=0,25 \Rightarrow 1 - P = 0,75$, θα έχουμε:

$$n = \frac{(1,96)^2 (0,25)(0,75)}{(0,05)^2} = 288$$

Παράδειγμα 1.4

Μια βιομηχανία παράγει προϊόντα ενός τύπου X και το ποσοστό των ελαττωματικών προϊόντων της είναι άγνωστο. Για την εκτίμηση του αγνώστου αυτού ποσοστού η εταιρεία παίρνει ένα δείγμα κατά τέτοιο τρόπο, ώστε το ποσοστό των ελαττωματικών του πληθυσμού να μη

διαφέρει περισσότερο από 2% του αντίστοιχου ποσοστού του πληθυσμού, σε επίπεδο εμπιστοσύνης 95%. Αν υποθέσουμε ότι από ένα δοκιμαστικό δείγμα μιας περιορισμένης έκτασης προέκυψε ποσοστό 12% να βρεθεί το μέγεθος του δείματος.

Λύση :

$$κ = |p - P| = 2\%$$

$$n = Z_{\alpha/2} \frac{P(1-P)}{κ^2} = \frac{(1,96)^2 (0,12)(0,88)}{(0,02)^2} = 1,014$$

Προσδιορισμός του μεγέθους του δείγματος στην πράξη

Στις προηγούμενες παραγράφους προσδιορίζουμε το ιδανικό μέγεθος δείγματος. Στην πράξη όμως παράγοντες όπως τα χρηματικά διαθέσιμα και ο χρόνος για την πραγματοποίηση μιας έρευνας επιβάλλουν περιορισμούς στο μέγεθος του δείγματος.

Στο κόστος της δειγματοληψίας ανήκουν τα γενικά έξοδα (π.χ. διάφορα υλικά, αμοιβές ερευνητών) το κόστος που αντιστοιχεί σε κάθε συνέντευξη κ.τ.λ. άρα η σχέση που συνδέει το μέγεθος του δείγματος και το κόστος της έρευνας θα είναι

$$K = \Gamma + n \cdot \lambda$$

Ένα άλλο πρόβλημα που εμφανίζεται στην πράξη είναι ότι σε ένα δείγμα μελετούμε περισσότερα από ένα χαρακτηριστικά και δεν είναι δυνατόν να προσδιορίζουμε το μέγεθος του δείγματος για κάθε χαρακτηριστικό ξεχωριστά. Εδώ μεγάλη σημασία έχει η πείρα του ερευνητή που σχεδιάζει την επιλογή του δείγματος.

Βλέπουμε λοιπόν ότι ο προσδιορισμός και η επιλογή του μεγέθους του δείγματος δεν είναι απλό πράγμα αλλά απαιτείται συνδυασμός επιστήμης και εμπειρίας στο σχεδιασμό της δειγματοληπτικής έρευνας.

Παράδειγμα 1.5

Προκειμένου να μελετήσουμε το μηνιαίο εισόδημα των κατοίκων μιας πόλης αποφασίζουμε να πάρουμε ένα τυχαίο δείγμα.

Ερωτάται:

- α) Ποιο μέγεθος δείγματος θα πρέπει να πάρουμε, ώστε το δειγματοληπτικό σφάλμα να μην υπερβαίνει τις ± 1.000 δραχμές, δεδομένου ότι $\sigma = 18.000$, σε επίπεδο σημαντικότητας 5%;
- β) Ποιο πρέπει να είναι το συνολικό κόστος της έρευνας, αν τα γενικά έξοδα – έστω ότι είναι σταθερά – ανέρχονται σε 5.000.000 δρχ. και τα έξοδα για τη συμπλήρωση κάθε ερωτηματολογίου που αντιστοιχεί σε κάθε μονάδα δείγματος ανέρχονται σε 500 δραχμές;
- γ) Αν όμως διαθέτουμε για την έρευνα 6.000.000 δρχ., ποιο πρέπει να είναι το δειγματοληπτικό σφάλμα;

Λύση:

- α) Το μέγεθος του δείγματος θα είναι:

$$n = z_{\alpha/2}^2 \frac{\sigma^2}{k^2} = 1,96^2 \frac{18.000^2}{1000^2} = 1244$$

- β) $K = \Gamma + n \cdot \lambda = 5.000.000 + 1.244 \cdot 500 = 5.622.000$

Άρα, το συνολικό κόστος θα είναι 5.622.000 δρχ.

- γ) Αν διαθέσουμε για την έρευνα 6.000.000 δρχ., θα έχουμε:

$$K = \Gamma + n \cdot \lambda \Rightarrow 6.000.000 = 5.000.000 + 500 \cdot n \Rightarrow n = 2.000$$

Το δειγματοληπτικό σφάλμα θα είναι:

$$k = \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = \pm 1,96 \frac{18.000}{\sqrt{2.000}} = \pm 789$$

Δεδομένου ότι το δείγμα έχει αυξηθεί, το δειγματοληπτικό σφάλμα έχει μειωθεί.

1,3 Επιλογή της μεθόδου συγκέντρωσης του απαιτούμενου στατιστικού υλικού και πηγές συλλογής πληροφοριών

Το τρίτο και πιο βασικό στάδιο για την μελέτη ενός φαινομένου είναι η συγκέντρωση των στατιστικών πληροφοριών που αφορούν το πρόβλημα που θέλουμε να εξετάσουμε.

Σ' αυτό το στάδιο η συγκέντρωση των πληροφοριών και των στοιχείων πρέπει να γίνει με ιδιαίτερη προσοχή γιατί απ' αυτά θα εξαρτηθεί η αξία των συμπερασμάτων. Αν τα στοιχεία που θα

συλλέξουμε είναι λανθασμένα ή ψεύτικα, τότε και τα συμπεράσματα θα είναι αναληθή.

Μπορούν να συγκεντρωθούν τα στοιχεία αυτά από πηγές όπως π.χ. από διάφορα κέντρα και ινστιτούτα ερευνών, από δημόσιους και ιδιωτικούς οργανισμούς, από τράπεζες, Διεθνείς οργανισμούς κ.τ.λ.

Στη χώρα μας η μεγαλύτερη πηγή για παροχή στατιστικών στοιχείων είναι η Εθνική Στατιστική Υπηρεσία Ελλάδας (ΕΣΥΕ).

Για να αποφασίσουμε πια μέθοδος συλλογής στοιχείων και πληροφοριών θα ακολουθήσουμε, δηλ. αν θα εφαρμοστεί η καθολική έρευνα (απογραφή) ή η δειγματοληψία ή η μέθοδος των συνεχών εγγραφών θα εξαρτηθεί από το σκοπό της έρευνας, το μέγεθος των μονάδων πληθυσμού, το κόστος, το χρόνο υλοποίησης της έρευνας και την ακρίβεια των αποτελεσμάτων της.

Η απογραφή συνίσταται στην συγκέντρωση στοιχείων απ' όλες τις στατιστικές μονάδες του πληθυσμού που θέλουμε να μελετήσουμε.

Παρουσιάζει όμως τα ακόλουθα μειονεκτήματα:

- α) Απαιτεί μεγάλο κόστος
- β) Καθυστέρηση στην δημοσίευση των αποτελεσμάτων λόγω του όγκου των πληροφοριών και των στατιστικών μονάδων.
- γ) Η απογραφή δεν γίνεται από ειδικευμένο προσωπικό έτσι στην συμπλήρωση των ερωτηματολογίων μπορεί να υπάρχουν σφάλματα.

Αντίθετα με τη γενική απογραφή, αν ο πληθυσμός που θέλουμε να μελετήσουμε από την άποψη ορισμένων ιδιοτήτων αποτελείται από μεγάλο πλήθος στατιστικών μονάδων και αν η μελέτη των ιδιοτήτων καταστρέφει τις μονάδες του πληθυσμού που εξετάζουμε, τότε η γενική απογραφή είναι πρακτικά αδύνατη και οικονομικά και χρονικά ασύμφορη.

Στις περιπτώσεις αυτές εφαρμόζουμε την μέθοδο της δειγματοληψίας, όπου εξετάζουμε μόνο ένα δείγμα το οποίο επιλέγουμε κατά τέτοιο τρόπο ώστε οι πληροφορίες, οι εκτιμήσεις και τα συμπεράσματα που θα βγάλουμε να έχει ισχύ για το σύνολο του πληθυσμού στο οποίο ανήκει το δείγμα.

Η χρησιμοποίηση της μεθόδου της δειγματοληψίας έχει σαφώς περισσότερα πλεονεκτήματα απ' ότι η γενική απογραφή.

α) Πλεονεκτήματα της δειγματοληψίας

Τα σημαντικότερα πλεονεκτήματα της δειγματοληψίας είναι:

I. Μεγαλύτερη ταχύτητα πληροφοριών

Η συλλογή των στοιχείων που χρειάζεται για τη μελέτη των χαρακτηριστικών ιδιοτήτων του πληθυσμού, μπορεί να γίνει πιο γρήγορα με τη δειγματοληψία απ' ότι με την απογραφή.

II. Μεγαλύτερη ακρίβεια

Με τις δειγματοληπτικές έρευνες έχουμε την δυνατότητα όταν το δείγμα είναι μικρό να αφιερωθεί περισσότερος χρόνος και μεγαλύτερη προσοχή στις συνεντεύξεις για την συμπλήρωση του ερωτηματολογίου. Ακόμα η εκπαίδευση και επίβλεψη των απογραφών γίνεται καλύτερα έτσι ώστε η ακρίβεια των πληροφοριών να είναι μεγαλύτερη. Βλέπουμε ότι η δειγματοληπτική έρευνα μπορεί να δώσει αποτελέσματα με μεγαλύτερη ακρίβεια απ' ότι η απογραφή.

III. Μεγαλύτερη ευχέρεια εφαρμογής

Η δειγματοληπτική έρευνα εφαρμόζεται στις περιπτώσεις που η γενική απογραφή είναι δυνατή, αλλά παράλογη. Π.χ. αν ένας γιατρός θέλει να προσδιορίσει τα ερυθρά αιμοσφαίρια ενός ασθενούς, δεν θα πάρει όλη την ποσότητα του αίματος του ασθενούς, αλλά μόνο λίγα γραμμάρια (δείγμα).

Στις περιπτώσεις αυτές το πρόβλημα δεν είναι αν θα ακολουθήσουμε την μέθοδο της καθολικής απογραφής ή τη

δειγματοληπτική μέθοδο, αλλά αν θα μείνουμε χωρίς πληροφορίες ή θα τις πάρουμε από ένα δείγμα, αφού η μέθοδος της απογραφής είναι παράλογη να εφαρμοστεί.

IV. Χαμηλό κόστος

Κύριος και αντικειμενικός σκοπός κάθε δειγματοληπτικής έρευνας είναι η λήψη μιας πληροφορίας με τη μεγαλύτερη δυνατή ακρίβεια και με το ελάχιστο κόστος.

V. Αδύνατη η εφαρμογή της γενικής απογραφής

Η δειγματοληπτική έρευνα μπορεί να εφαρμοστεί σε εκείνες τις περιπτώσεις που η καθολική έρευνα είναι αδύνατη.

B) Μειονεκτήματα της δειγματοληψίας

Η μέθοδος της δειγματοληψίας παρουσιάζει και μερικά μειονεκτήματα, όπως:

Αν οι μονάδες του πληθυσμού που ερευνούμε εμφανίζονται πολύ σπάνια, τότε το δείγμα που θα πρέπει να μελετήσουμε θα είναι πολύ μεγάλο αν θέλουμε να έχουμε αξιόπιστα αποτελέσματα.

Ο σχεδιασμός και η εκτέλεση της δειγματοληψίας χρειάζονται ιδιαίτερη προσοχή και θα πρέπει να ακολουθηθεί αυστηρά η θεωρητική διαδικασία που επιβάλλεται για την επιλογή του δείγματος και η στατιστική ανάλυση των αποτελεσμάτων της δειγματοληψίας.

Διάφοροι παράγοντες όπως η κακή σχεδίαση και εκτέλεση της δειγματοληψίας, η μη κατάλληλη μέθοδος διενέργειας της δειγματοληψίας και τα ανεπαρκή δεδομένα, οδηγούν σε μερική αποτυχία της έρευνας.

Άλλο βασικό μειονέκτημα της δειγματοληψίας είναι τα δειγματοληπτικά σφάλματα, δηλ. η διαφορά μεταξύ των αποτελεσμάτων που προκύπτει από την τυχαία επιλογή των δειγματοληπτικών μονάδων και των αποτελεσμάτων που προκύπτουν από την καθολική έρευνα.

Συνεχείς εγγραφές στατιστικών στοιχείων

Με τη μέθοδο αυτή συλλέγουμε και καταχωρούμε συνεχώς τα στατιστικά στοιχεία σε ειδικά βιβλία ή έντυπα μόλις αυτά εμφανιστούν. Π.χ. οι καταχωρήσεις σε ειδικά βιβλία στο ληξιαρχείο των γεννήσεων, θανάτων και γάμων, η τουριστική κίνηση, οι εξαγωγές και εισαγωγές, τα μετεωρολογικά στοιχεία κ.τ.λ.

Οι τρόποι συλλογής των στατιστικών πληροφοριών, τόσο στην περίπτωση της καθολικής απογραφής όσο και στην περίπτωση της δειγματοληψίας είναι: η παρατήρηση, η προσωπική συνέντευξη και η ταχυδρομική αποστολή του ερωτηματολογίου.

α) Παρατήρηση

Η μέθοδος αυτή χρησιμοποιείται περισσότερο στις φυσικές επιστήμες, ενώ αντίθετα στις στατιστικές έρευνες είναι λιγότερο διαδεδομένη π.χ. για την έρευνα συμπεριφοράς του πληθυσμού μιας χώρας, για τη μελέτη υποανάπτυκτων κοινωνιών, κ.τ.λ.

β) Προσωπική συνέντευξη

Η μέθοδος αυτή είναι η καλύτερη για την συλλογή στατιστικών στοιχείων, αλλά και η πιο διαδεδομένη στις δειγματοληπτικές έρευνες.

Η συμπλήρωση του ερωτηματολογίου στην μέθοδο αυτή γίνεται από εκπαιδευμένα άτομα που λέγονται ερευνητές ή ερευνήτριες.

Το έργο των ερευνητών συνίσταται:

- i) Στον εντοπισμό των μονάδων του δείγματος
- ii) Στην πραγματοποίηση των συνεντεύξεων
- iii) Στον τρόπο προβολής των ερωτήσεων
- iv) Στην καταγραφή των απαντήσεων.

Η καταγραφή φαίνεται απλή εργασία, αλλά παρουσιάζονται λάθη που οφείλονται κυρίως στο ότι το έργο είναι κουραστικό ενώ ο τρόπος συμπλήρωσης του ερωτηματολογίου πολλές φορές είναι πολύπλοκος.

Η μέθοδος της προσωπικής συνέντευξης έχει τα εξής πλεονεκτήματα:

I) Τα στοιχεία που συγκεντρώνονται είναι ποιοτικά καλύτερα γιατί δίνονται διευκρινίσεις στους ερευνώμενους έτσι ώστε να δίνουν τις σωστές απαντήσεις.

II) Το ποσοστό ανταπόκρισης των ερευνομένων είναι μεγαλύτερο.

III) Τα στοιχεία μπορούν να συγκεντρωθούν απευθείας με κομπιούτερ.

Το μεγάλο κόστος είναι το μειονέκτημα της μεθόδου αυτής.

γ) Ταχυδρομική αποστολή του ερωτηματολογίου

Η μέθοδος αυτή είχε μεγάλη εφαρμογή παλαιότερα, αλλά σήμερα έχει περιοριστεί αρκετά, εφαρμόζεται σε ορισμένες περιπτώσεις που είναι δύσκολη η συλλογή πληροφοριών όπως όταν ο πληθυσμός είναι διασκορπισμένος σε μεγάλες εκτάσεις, ο διαθέσιμος χρόνος και οι χρηματικοί πόροι είναι περιορισμένοι.

Τα πλεονεκτήματα της μεθόδου αυτής είναι το μικρό κόστος, η ταχεία διεξαγωγή της έρευνας, η αποφυγή σφαλμάτων του ερευνητή και η καλύτερη συγκέντρωση εμπιστευτικών πληροφοριών.

Τα μειονεκτήματα της μεθόδου αυτής είναι πρώτον ότι μόνο ένα μικρό ποσοστό απαντά και δεύτερον τα δειγματοληπτικά σφάλματα είναι σημαντικά.

Ενημερωτική επιστολή

Η ενημερωτική επιστολή παίζει σημαντικό ρόλο στην επιτυχία μιας δειγματοληπτικής έρευνας. Σαν σκοπό έχει να προετοιμάσει τα

ερευνώμενα άτομα έτσι ώστε να συνεργαστούν με προθυμία και ειλικρίνεια με τον ερευνητή.

Με αυτόν τον τρόπο τα στοιχεία που συγκεντρώνουν οι ερευνητές είναι πιο έγκυρα ποιοτικά και το ποσοστό ανταπόκρισης των ερευνώμενων μεγαλύτερο.

Με την ενημερωτική επιστολή πρέπει:

α) Να γίνεται περιγραφή της έρευνας και αναφορά ότι η επιλογή των ερευνωμένων έγινε με τυχαίο τρόπο

β) Να περιγράφεται ο σκοπός της έρευνας

γ) Να επισημαίνεται ότι οι πληροφορίες που θα συγκεντρωθούν είναι εμπιστευτικές και ότι δημοσιεύονται ομαδοποιημένα στοιχεία, ώστε να μην αποκαλύπτονται ατομικά χαρακτηριστικά.

1,4 Σύνταξη ερωτηματολογίου

Η σύνταξη του ερωτηματολογίου αποτελεί το τέταρτο στάδιο διεξαγωγής μιας έρευνας. Το ερωτηματολόγιο είναι ένα έντυπο όπου καταχωρούνται οι πληροφορίες που παίρνουμε από τους ερωτώμενους, αποτελεί μέσο συλλογής των στατιστικών πληροφοριών.

Το ερωτηματολόγιο πρέπει να συνταχθεί με μεγάλη προσοχή γιατί η επιτυχία μιας έρευνας θα εξαρτηθεί από τα συγκεντρωθέντα στοιχεία που περιέχονται στο ερωτηματολόγιο.

Ένα ερωτηματολόγιο για να είναι αποτελεσματικό θα πρέπει:

- 1) Να είναι σύντομο.** Ένα μεγάλο ερωτηματολόγιο απομακρύνει τους ερευνητές και ερευνώμενους. Το κόστος έρευνας είναι μεγαλύτερη ενώ λόγω των αρνήσεων η ποιότητα μειώνεται.
- 2) Να είναι εύκολο στην απάντηση και να μπορεί εύκολα να ταχυδρομηθεί.**
- 3) Να έχει σχεδιαστεί καλά.**

- 4) Τα ερωτήματα να έχουν λογική ακολουθία Η συνέντευξη διεξάγεται περισσότερο ομαλά αν κάθε ερώτηση οδηγεί φυσιολογικά στην απάντηση.
- 5) Να είναι δυνατή η επεξεργασία των στοιχείων .
- 6) Να έχει κατάλληλο σχηματικό μέγεθος.

Όσον αφορά το περιεχόμενο των ερωτήσεων θα πρέπει να έχουμε υπ' όψιν μας τα εξής:

1. Την ικανότητα του ατόμου που θα ερωτηθεί.

Αν δηλαδή ο πληθυσμός που θα ερωτηθεί έχει τις κατάλληλες γνώσεις για να απαντήσει. Πολλές φορές τα άτομα δεν δηλώνουν άγνοια για ψυχολογικούς λόγους, με αποτέλεσμα οι απαντήσεις να μην ανταποκρίνονται στην πραγματικότητα.

2. Τη θέληση του ερωτώμενου.

Συνήθως αποφεύγονται ερωτήσεις που το περιεχόμενό τους προξενούν κοινωνική αντίδραση.

α) Διατύπωση των ερωτήσεων

Υπάρχουν ορισμένες καθοδηγητικές αρχές σχετικά με τη διατύπωση των ερωτήσεων. Μερικές τέτοιες αρχές είναι:

I. Απλότητα της γλώσσας. Η γλώσσα του ερωτηματολογίου πρέπει να είναι απλή και κατανοητή γιατί ένα μεγάλο ποσοστό του πληθυσμού κατέχει στοιχειώδη εκπαίδευση, ενώ μερικοί ίσως να είναι αναλφάβητοι.

II. Σαφήνεια. Τα ερωτήματα πρέπει να είναι απλά διατυπωμένα, χωρίς ασάφειες.

III. Αποφυγή ερωτήσεων που οδηγούν σε αποκρύψεις.

IV. Αποφυγή διατύπωσης που προκαλεί ψυχολογικές αντιδράσεις. Οι ερωτήσεις πρέπει να είναι κατά τέτοιο τρόπο διατυπωμένες ώστε να μην προσβάλλουν τον ερωτώμενο.

V. Αποφυγή κατευθυνόμενων ερωτήσεων. Η κατευθυνόμενη ερώτηση σκοπό έχει να οδηγήσει τον ερωτώμενο σε επιθυμητή απάντηση.

IV. Προκωδικογράφιση. Για τα ερωτήματα τα οποία χρειάζεται περιορισμένη απάντηση πρέπει να γίνεται προκωδικογράφιση.

Με την προκωδικογράφιση των απαντήσεων διευκολύνεται σημαντικά η επεξεργασία του ερωτηματολογίου.

Τέλος, θα πρέπει κάτω από κάθε ερώτημα να δίνονται οδηγίες, οι οποίες να διευκολύνουν τον ερευνητή να μεταβιβάσει στον ερευνώμενο τι ακριβώς ζητείται μέσω των ερωτημάτων.

ΠΑΡΑΔΕΙΓΜΑ 1^ο

Αν μας αναθέσουν μια έρευνα για τη μακροζωία των Ελλήνων, να καταρτιστεί το αναγκαίο ερωτηματολόγιο και να περιγραφεί ο τρόπος συλλογής του δείγματος αν ο ερευνώμενος πληθυσμός είναι ηλικίας 70 ετών και άνω:

1) το ερωτηματολόγιο από τα μέλη της ερευνητικής ομάδας μπορεί να συνταχθεί ως εξής:

ΔΕΛΤΙΟ ΕΡΕΥΝΑΣ ΓΙΑ ΤΗ ΜΑΚΡΟΖΩΙΑ

Νομός:.....

Επαρχία:.....

Δήμος ή Κοινότητα:.....

Οικισμός:.....

Οδός και αριθμός:.....

A. ΓΕΝΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ

1. Ονοματεπώνυμο

2. Φύλλο: Άρρεν Θήλυ

3. Έτος γεννήσεως

4. Τόπος γεννήσεως :

Δήμος ή κοινότητα.....

Οικισμός

Επαρχία

6. Οικογενειακή κατάσταση :

άγαμος

έγγαμος

χήρος

διαζευμένος(η)

7. Πόσα παιδιά έχετε γεννήσει μέχρι σήμερα;

(μόνο για γυναίκες).....

Από αυτά σήμερα ζούν.....

8. Πόσα παιδιά έχει γεννήσει η μητέρα σας;

Από αυτά σήμερα ζουν

9. Ζει σήμερα η μητέρα σας;

Ναι Όχι

Αν ναι, πόσων ετών είναι;.....

Αν όχι, σε ποια ηλικία πέθανε;.....

10. Ζει σήμερα ο πατέρας σας;

Ναι Όχι

Αν ναι, πόσων ετών είναι;.....

Αν όχι, σε ποια ηλικία πέθανε;.....

11. Επίπεδο εκπαίδευσης:

(να γραφεί ο ανώτερος τίτλος σπουδών)

πτυχίο ΑΕΙ

Πτυχίο ΤΕΙ ή Ανωτ. Σχολής

Απολυτήριο Λυκείου

Απολυτήριο Δημοτικού

Αναλφάβητος

12. Εργάζεστε συνήθως;

Ναι Όχι

Αν ναι να γραφεί το είδος εργασίας

.....

Αν όχι ποια είναι η ασχολία σας;

.....

Β. ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΚΑΤΟΙΚΙΑΣ ΚΑΙ ΕΙΣΟΔΗΜΑΤΟΣ

13. Η κατοικία στην οποία στεγάζετε είναι:

Ιδιόκτητη

Με ενοίκιο

Συλλογική κατοικία (γηροκομείο κ.τ.λ)

Άλλη περίπτωση (να γραφεί)

14. Αν διαμένετε σε ιδιόκτητη κατοικία ή με ενοίκιο, μένετε:

- Μόνος
- Με τη σύζυγο ή τον σύζυγο
- Με συγγενικά πρόσωπα
- Με άλλα άτομα

15. Ποιες είναι οι πηγές συντηρήσεως;

- Εργασία
- Σύνταξη
- Εισόδημα από περιουσία
- Επίδομα ή βοηθήματα
- Άλλη περίπτωση (να γραφεί)

16. Σε ποια εισοδηματική τάξη ανήκετε;

(Λαμβάνεται υπόψη το καθαρό ετήσιο εισόδημα)

- μέχρι 300.000 δρχ.
- 300.000 – 600.000 δρχ.
- 600.000 – 900.000 δρχ.
- 900.000 – 1.200.000 δρχ.
- 1.200.000 – 1.500.000 δρχ.
- 1.500.000 και άνω δρχ.

Γ. ΔΙΑΤΡΟΦΗ

17. Ποια από τα παρακάτω γεύματα παίρνετε συνήθως;

- Πρωινό
- Μεσημεριανό
- Βραδινό
- Άλλη περίπτωση (να γραφεί)

18. Αν παίρνετε πρωινό, ποια από τα παρακάτω είδη συνήθως παίρνετε;

- Γάλα
- Αυγά
- Βούτυρο
- Μέλι
- Μαρμελάδα
- Τσάι
- Ψωμί
- Μπικον
- Κέικ
- Φρυγανιές
- Καφέ
- Λοιπά (να γραφεί)

19. Ποια από τα παρακάτω είδη διατροφής τρώτε και σε ποια συχνότητα (σημειώνετε X στον αντίστοιχο χώρο του πίνακα)

Είδος Διατροφής	Καθημερινά	Κάθε 2ημέρες	Κάθε 3ημέρες	Κάθε 4ημέρες	Κάθε ...ημέρες	Καθόλου
Κρέας						
Ψάρι						
Όσπρια						
Τυρί						
Γιαούρτη						
Αυγά						
Λαδερά						
Χόρτα						
Σαλάτες						

Γλυκά						
Λοιπά (να γραφεί)						

20. Αν κάποια από τα ανωτέρω είδη διατροφής δεν τρώτε καθόλου, να γραφεί το είδος και ο λόγος:

α).....

β).....

γ).....

δ).....

21. Πως προτιμάτε το μαγείρεμα των φαγητών;

Αλμυρά

Με λίγο αλάτι

Με πολύ σάλτσα

Βραστά

Τηγανητά

Αλλιώς (να γραφεί).....

22. Ποια φαγητά προτιμούσατε νέος;

α).....

β).....

γ).....

23. Σήμερα έχετε αλλάξει τρόπο διατροφής για κάποιο λόγο υγείας;

ΝΑΙ ΟΧΙ

24. Όταν βρίσκεστε στο τραπέζι για φαγητό τρώτε γρήγορα ή αργά;

Γρήγορα

Αργά

25. Όταν παίρνετε το γεύμα σας παραχορταίνετε πάντα ή μόνο όταν σας αρέσει πολύ κάποιο φαγητό ή ποτέ;

Παραχορταίνω πάντα

Παραχορταίνω σπάνια

Δεν παραχορταίνω ποτέ

26. Μετά το φαγητό ξαπλώνετε;

ΝΑΙ ΟΧΙ

27. Ποια από τα παρακάτω είδη αναψυκτικών ή οινοπνευματωδών ποτών πίνετε και σε ποια ποσότητα και συχνότητα; (σημειώστε X στον αντίστοιχο χώρο του πίνακα).

Είδος αναψυκτικών ή οινοπνευματ ωδών ποτών	Καθημερινά	Σπανίως	Καθόλου
	1 ποτήρι 2 ποτήρια 3 ποτήρια ... ποτήρια		
Πορτοκαλάδα			
Λεμονάδα			
Κόκα – κόλα			
Κρασί			
Ούζο			
Τσίπουρο			
Ουίσκι			
Λοιπά (να γραφεί)			

Δ. ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΥΓΕΙΑΣ

28. Έχετε κάποιο πρόβλημα υγείας;

ΝΑΙ ΟΧΙ

Αν ΝΑΙ, ποιο ή ποια από τα παρακάτω:

- Πίεση
- Αναπνευστικά
- Στομάχι
- Καρδιά
- Αρθρικά
- Όραση
- Ακοή
- Μνήμη
- Άλλη περίπτωση (να γραφει)

29. Ποιο από τα παραπάνω προβλήματα υγείας είχατε από νέος;

- α).....
- β).....
- γ).....

30. Κάνατε κάποιες καταχρήσεις όταν είσαστε νέος;

ΝΑΙ ΟΧΙ

Αν ΝΑΙ ποιες;

- α).....
- β).....
- γ).....

31. ποιες ώρες κοιμάστε συνήθως;

32. Ποια ώρα ξυπνάτε το πρωί;

33. Ποια ώρα κοιμάστε το βράδυ συνήθως;

34. Κάθε πότε πηγαίνετε στο γιατρό για εξετάσεις;

Τακτικά

Μόνο όταν έχω κάποιο πρόβλημα

Άλλη περίπτωση (να γραφεί)

35. Ποιο είναι το βάρος σας;

36. Ποιο είναι το ύψος σας; (σε εκατοστά)

37. Καπνίζετε σήμερα;

ΝΑΙ

ΟΧΙ

Αν ΝΑΙ πόσα τσιγάρα καπνίζετε την ημέρα;

Αν ΟΧΙ για ποιο λόγο δεν καπνίζετε;

38. Καπνίζατε σε νεότερη ηλικία

ΝΑΙ

ΟΧΙ

Αν ΝΑΙ πόσα τσιγάρα καπνίζατε την ημέρα;

E. ΣΥΜΠΛΗΡΩΜΑΤΙΚΑ ΣΤΟΙΧΕΙΑ

39. Είστε γενικά ευχαριστημένος από τη ζωή σας;

ΝΑΙ

ΟΧΙ

Αν ΟΧΙ για ποιο λόγο;.....

40. Πως περνάτε συνήθως την ημέρα σας όταν δεν εργάζεστε;

- Στην τηλεόραση
- Στο καφενείο
- Με τους φίλους
- Άθληση
- Άλλη περίπτωση (να γραφεί)

41. Περιγράψτε μας ποιες κοινωνικές παροχές θα θέλατε από το Κράτος κατά σειρά σπουδαιότητας;

- Ικανοποιητική σύνταξη
- Οίκους ευγηρίας
- Υγειονομική περίθαλψη
- Άλλη περίπτωση (να γραφεί)

42. Θα θέλατε να μας πείτε λίγα λόγια γενικά για την κοινωνία, για την οικογένεια για τη συμπεριφορά των παιδιών προς τους ηλικιωμένους γονείς, για το Κράτος, που θα είναι χρήσιμα για τη βελτίωση του βιοτικού σας επιπέδου;

.....

.....

.....

.....

Όνοματεπώνυμο ερευνητή

Όνομα.....

Επώνυμο.....

Τηλ.....

Ημερομηνία

Η διαδικασία για την επιλογή του δείγματος για τη μελέτη της μακροζωίας στη χώρα μας θα γίνει ως εξής:

Επιλογή δείγματος

Έχοντας υπόψη το βαθμό ανομοιογένειας που παρουσιάζουν τα χαρακτηριστικά που θα μελετηθούν το σκοπό της έρευνας και το κόστος της, η επιλογή του δείγματος θα γίνει τέτοιο κατά τρόπο ώστε το δείγμα να είναι αντιπροσωπευτικό, δηλαδή να μπορεί να γίνει αναγωγή του δείγματος στο σύνολο του ερευνώμενου πληθυσμού με αξιοπιστία. Για το σκοπό αυτό ο ερευνώμενος πληθυσμός θα χωριστεί σε τρία στρώματα αστικός, ημιαστικός και αγροτικός πληθυσμός. Στη συνέχεια θα γίνει επιλογή δείγματος μέσω από κάθε στρώμα ως εξής:

α) Αστικός πληθυσμός

Λαμβάνοντας υπόψη το κόστος της έρευνας και ότι ο τρόπος ζωής των ατόμων ηλικίας 70 ετών και άνω δεν διαφέρει μεταξύ των αστικών κέντρων της χώρας, η έρευνα θα περιοριστεί στην περιφέρεια πρωτεύουσας. Αυτό δεν επηρεάζει την αντιπροσωπευτικότητα του δείγματος, δεδομένου ότι ο αστικός πληθυσμός της περιφέρειας πρωτεύουσας αποτελεί το 54% του ερευνώμενου πληθυσμού, όπως φαίνεται στον πίνακα 3.1.

Η επιλογή του δείγματος θα γίνει σε δύο στάδια. Κατά το πρώτο στάδιο θα επιλεγούν με τυχαίο τρόπο οι μονάδες επιφανείας (δηλ. ένας αριθμός οικοδομικών τετραγώνων, περίπου 200, από χάρτες κλίμακας 1:500). Στη συνέχεια, κατά το δεύτερο στάδιο, θα επιλεγεί με τυχαίο τρόπο ένας αριθμός νοικοκυριών που έχουν άτομα ηλικίας 70 ετών και άνω, από κατάλογο-πλαίσιο που θα καταρτίσει ο ερευνητής κατά οικοδομικό τετράγωνο του δείγματος. Στον κατάλογο-πλαίσιο θα καταχωρηθούν όλα τα νοικοκυριά που διαμένουν στη μονάδα επιφανείας (οικοδομικό τετράγωνο δείγματος) και έχουν άτομα ηλικίας 70 ετών και άνω. Η επιλογή των νοικοκυριών θα γίνει από τον ερευνητή με συστηματική δειγματοληψία.

Σε κάθε οικοδομικό τετράγωνο θα ερευνηθούν έξι (6) νοικοκυριά, επομένως το διάστημα δειγματοληψίας θα είναι:

$$\lambda = \frac{\text{Αριθμός νοικοκυριών στο οικοδομικό τετράγωνο}}{6}$$

β) Ημιαστικός πληθυσμός

Ο ερευνώμενος πληθυσμός στις ημιαστικές περιοχές δεν διαφέρει σημαντικά ως προς τον τρόπο ζωής των ατόμων ηλικίας 70 ετών και άνω. Επίσης, από τον πίνακα 1 φαίνεται ότι τα άτομα ηλικίας 70 ετών και άνω που διαμένουν στις ημιαστικές περιοχές αποτελούν μόνο το 11,6% του πληθυσμού.

Κατόπιν των ανωτέρω και λαμβάνοντας υπόψη το κόστος και την αξιοπιστία των αποτελεσμάτων, η έρευνα θα περιοριστεί σε 9 ημιαστικές πόλεις της χώρας (μία σε κάθε γεωγραφικό διαμέρισμα).

Η επιλογή του δείγματος θα γίνει σε τρία στάδια. Κατά το πρώτο στάδιο θα επιλεγούν με τυχαίο τρόπο 9 ημιαστικές πόλεις μια σε κάθε γεωγραφικό διαμέρισμα, ώστε να αντιπροσωπεύονται όλες οι περιοχές της χώρας. Κατά το δεύτερο στάδιο θα επιλεγούν με τυχαίο τρόπο οι μονάδες επιφάνειας (οικοδομικά) που θα ανέρχονται σε 10 περίπου. Στη συνέχεια κατά το τρίτο στάδιο, θα επιλεγεί με τυχαίο τρόπο ένας αριθμός νοικοκυριών που έχουν άτομα ηλικίας 70 ετών και άνω, από κατάλογο – πλαίσιο που θα καταρτίσει ο ερευνητής για το σύνολο των 10 οικοδομικών τετραγώνων. Στον κατάλογο – πλαίσιο θα καταχωρηθούν όλα τα νοικοκυριά που διαμένουν στα οικοδομικά τετράγωνα του δείγματος. Η επιλογή των νοικοκυριών θα γίνει από τον ερευνητή με συστηματική δειγματοληψία. Στο σύνολο των 10 οικοδομικών τετραγώνων θα ερευνηθούν 20 νοικοκυριά, επομένως το διάστημα δειγματοληψίας θα είναι:

$$\Lambda = \frac{\text{Αριθμός νοικοκυριών στο οικοδομικό τετράγωνο του δείγματος}}{20}$$

γ) Αγροτικός πληθυσμός

Ο ερευνώμενος πληθυσμός στις αγροτικές περιοχές πασουςιάζει ανομοιογένεια ως προς τον τρόπο ζωής των ατόμων ηλικίας 70 ετών και άνω μεταξύ των γεωγραφικών περιοχών. Επίσης, ο αγροτικός πληθυσμός αποτελεί το 42% του πληθυσμού. Για να είναι το δείγμα αντιπροσωπευτικό, η έρευνα θα καλύψει όλους τους Νομούς της Χώρας. Από κάθε Νομό θα επιλέγουν τρεις Κοινότητες μια ορεινή μια ημιορεινή και μια πεδινή.

Η επιλογή το στάδιο θα επιλέγουν με τυχαίο τρόπο τρεις Κοινότητες σε κάθε Νομό με διάκριση σε ορεινές ημιορεινές και πεδινές, ώστε να αντιπροσωπεύονται όλες οι γεωγραφικές περιοχές που αποτελούν βασικό παράγοντα διαφοροποίησης στον τρόπο ζωής των ατόμων ηλικίας 70 ετών και άνω, από κατάλογο – πλαίσιο που θα καταρτίσει ο ερευνητής για κάθε κοινότητα του δείγματος που περιέχεται στον πίνακα. Στον κατάλογο – πλαίσιο θα καταχωρηθούν όλα τα νοικοκυριά που διαμένουν στην κοινότητα και έχουν άτομα ηλικίας άνω των 70 ετών. Η επιλογή των νοικοκυριών θα γίνει με συστηματική δειγματοληψία από το σύνολο των νοικοκυριών που καταχωρήθηκαν στον κατάλογο – πλαίσιο κάθε κοινότητας θα ερευνηθούν 10 νοικοκυριά επομένως το διάστημα δειγματοληψίας θα είναι:

$$\lambda = \frac{\text{Αριθμός νοικοκυριών στην κοινότητα δείγματος}}{10}$$

Αστικός, ημιαστικός και αγροτικός πληθυσμός ηλικίας 70 ετών και άνω κατά γεωγραφικό διαμέρισμα Απογραφή 1981

Γεωγραφικό διαμέρισμα	Αστικός	Ημιαστικός	Αγροτικός
Σύνολο Ελλάδας	374.664	94.335	340.483
Περιφ. Πρωτευούσης	204.531		
Λοιπή Στερεά Ελλάς και Εύβοια	19.428	27.882	50.010
Πελοπόννησος	25.204	12.475	75.611
Ιόνιοι νήσοι	3.026	2.773	17.608
Ήπειρος	5.014	1.757	23.045
Θεσσαλία	18.849	8.436	30.024
Μακεδονία	67.541	23.430	67.855
Θράκη	7.567	3.707	13.423
Νήσοι Αιγαίου	11.806	8.813	29.176
Κρήτη	11.698	5.062	33.731

Πηγή: Γ.Γ. ΕΣΥΕ

Πίνακας 3.2
Ημιαστικές πόλεις δείγματος
κατά γεωγραφικό διαμέρισμα

Γεωγραφικό διαμέρισμα	Πόλεις	Πληθυσμος
Λοιπή Στερεά Ελλάς και Εύβοια	Αλιάρτος	3502
Πελοπόννησος	Λεωνιδίου	3557
Ιόνιοι νήσοι	Ιθάκη	2037
Ήπειρος	Κόνιτσα	2.849
Θεσσαλία	Ελασσών	6.527
Μακεδονία	Λαγκαδάς	5.890
Θράκη	Σουφλίου	5.043
Νήσοι Αιγαίου	Νάξος	3.735
Κρήτη	Μοίραι	3.501

Πίνακας 3.3
Κοινότητες δείγματος κατά νομό με διάκριση αυτών
σε ορεινές, ημιορεινές, πεδινές και πληθυσμός αυτών

	Ορεινές		Ημιορεινές		Πεδινές	
1. ΣΤΕΡΕΑ ΕΛΛΑΔΑ						
Απλωσακαρνανία	Σταθά	503	Χρυσοβέργιου	482	Αγ. Ηλία	263
Αττικής (Υπόλοιπο)	Άνω Φαναριου	391	Διονύσου	556	Αγ. Κων/νου	428
Βοιωτίας	Αγ. Άνας	570	Υψηλάντου	584	Θουρίου	230
Ευβοίας	Μύλων	166	Ανδίων	249	Παλιούρα	679
Ευρυτανίας*	Αγ. Τριάδας	477	Ανιάδας	142	Βράχας	233
Φθιώτιδος	Τριλόφου	182	Παλαμά	368	Αγραπηδιάς	275
Φωκίδος	Δροσάτου	184	Μαραθιά	382	Καστρακίου	455
* Όλες ορεινές						
2. ΠΕΛΟΠΟΝΗΣΟΣ						
Αργολίδος	Κρυονερίου	476	Μυκηνών	440	Φούρνων	371
Αρκαδίας	Αετοράχης	202	Βούτση	180	Καριταίνης	304
Αχαΐας	Κρήνης	278	Γρηγόρη	171	Κούμαρη	222
Ηλείας	Αστρά	381	Δάφνης	536	Ελαιώνος	509
Κορινθίας	Γαλατά	137	Ελληνικού	90	Σουλίου	464
Λακωνίας	Αγριάνων	132	Βασιλικής	85	Ελαίας	215
Μεσσηνίας	Μηλέας	208	Τρικόρφου	511	Άμμου	269
3. ΝΗΣΙΑ ΙΟΝΙΟΥ						
Ζακύνθου	Αγ. Μαρίνης	152	Αγαλά	310	Βουγιόπου	344
Κερκύρας	Ζυγού	255	Δροσάτου	233	Ξανθάτων	246
Κεφαλληνίας	Λεύκης	97	Κιονίου	265	Κεραμειών	141
Λευκάδος	Αγ. Ηλίας	236	Καλάμου	454	Καριωτών	310

Ορεινές		Ημιορεινές		Πεδινές	
4. Η ΠΕΙΡΟΣ					
Αρτας	Αντιμοριακής	373	Κεντριού	380	Καρωνιοίσις 21α
Θεσπρωτικές	Μαυραουδίου	283	Καταβολίρας	348	Παλαιχωρίου 311
Κορινθίας	Αστροαγγελών	287	Γραμμένου	268	Πεταπλου 355
Πρωτέρας	Αρβανός	188	Βράχου	397	Καστρίου 407
5. ΘΕΣΣΑΛΙΑ					
Καρδίτσης	Αιδαίνου	324	Αγ. Δημητρίου	175	Αυτίτισης 210
Λοκράς	Άρτας	329	Ιπείας	390	Αγ. Γεωργίου 228
Μαγνησίας	Ανοβρας	66	Μεταχίου	136	Ακρίνου 463
Τρικαλών	Αιπλοχωρίου	213	Παναγίτου	416	Ζηλευτής 634
6. ΜΑΚΕΔΟΝΙΑ					
Γρεβενών	Κυθρήρας	269	Εξάρχου	183	Κυβούτου 577
Αρτας	Αχλαδιάς	157	Νικηφόρου	561	Περαχωράς 252
Ημαθίας	Κοιμωτικής	210	Γεωργιανών	273	Κυψέλης 536
Πενταπόλεως	Λαδοίου	284	Σταυκηνάς	238	Ελευσίνης 492
Κοζάνης	Χαλκίτου	244	Κρυονέριου	560	Κερκονής 534
Καστοριάς	Βράχου	136	Μηλίτου	368	Αιγής 209
Κιλίκης	Κήρης	494	Φινυαί	197	Αμαράντων 599
Κοζάνης	Διόφρου	117	Κημιθίου	151	Πεπονής 215
Πέλλης	Θηραπεύτας	397	Δροσερίου	615	Βρυτών 514
Παργών	Βίρας	503	Μαυχοχωρίου	511	Αρών 412
Σερρών	Αγρινής	534	Τριάδος	352	Επεαμίλων 316
Φλώρινας	Αγ. Γεωργίου	237	Πρωτής	137	Τριπαταμού 371
Κιθωνίδας	Οκτωπόδος	519	Πετραλώνων	438	Ριζών 524
7. ΘΡΑΚΙΑ					
Έβρου	Αιουμής	67	Κερακής	258	Πυλαίας 421
Έβρου	Γαλάτης	175	Δαφνίου	385	Μυγκού 600
Ροδόπης	Παντοπόρου	650	Κερυδίας	260	Πελαγίας 139
8. ΝΗΣΙΑ ΑΙΓΑΙΟΥ					
Αρδεκανησίου	Μονολίθου	394	Πλατανίων	251	Αρχονίας 270
Κυκλάδων	Μέσης	110	Μέσης	260	Εγγόρων 53
Λέσβου	Περπημίας	500	Αερώνος	320	Αγ. Δημητρίου 621
Σάμου	Αρεθούσης	216	Ευεθίου	656	Μυλων 286
9. ΚΡΗΤΗ					
Ηρακλείου	Βαχού	21	Διονυσίου	573	Αρωρών 287
Λασιθίου	Μετλέρας	180	Φωκηνής	438	Ασκού 189
Ρεθύμνης	Δρωσκάλ	93	Ασπινής	187	Κερωνής 224
Χανίων	Ελάτης	249	Αλεξάνδρου	263	Μοδίου 213

1,5 Πρόσληψη προσωπικού και προετοιμασία αυτού για συλλογή των στατιστικών στοιχείων

Το πέμπτο στάδιο στην διαδικασία διεξαγωγής μιας έρευνας είναι η πρόσληψη και η προετοιμασία του προσωπικού που θα χρησιμοποιηθεί. Η σωστή επιλογή και εκπαίδευση του προσωπικού παίζει σημαντικό ρόλο στην επιτυχία της έρευνας.

α) Πρόσληψη προσωπικού

Ο αριθμός των προσώπων που θα χρησιμοποιηθούν για την διεξαγωγή μιας έρευνας εξαρτάται από το μέγεθος του δείγματος, τα διαθέσιμα χρήματα, το χρονοδιάγραμμα κ.τ.λ.

Τρεις είναι οι κατηγορίες του προσωπικού που θα προσληφθεί για την διεξαγωγή μιας έρευνας.

- ειδικευμένο προσωπικό, που θα ασχοληθεί με το σχεδιασμό της έρευνας, την εποπτεία και την ανάλυση των αποτελεσμάτων.
- Οι υπάλληλοι γραφείου, που θα ασχοληθούν με την επεξεργασία.
- Οι ερευνητές που θα ασχοληθούν με την συμπλήρωση των ερωτηματολογίων.

Οι ερευνητές αποτελούν το βασικό προσωπικό μιας έρευνας, γιατί από τον τρόπο που θα εργαστούν και θα συλλέξουν τα στοιχεία θα εξαρτηθεί η ποιότητα της έρευνας.

β) Εκπαίδευση ερευνητών

Οι εκπαιδευτές πρέπει να είναι εξοικειωμένοι με τους σκοπούς της έρευνας, το ερωτηματολόγιο και τον τρόπο συμπλήρωσής του.

Συγκεκριμένα το εκπαιδευτικό πρόγραμμα των ερευνητών με την μορφή σεμιναρίου πρέπει να περιλαμβάνει:

- Την περιγραφή του σκοπού της έρευνας
- Συνοπτική περιγραφή της διαδικασίας της επεξεργασίας

- Τα της εργασίας του ερευνητή
- Ανάπτυξη της τεχνικής των συνεντεύξεων
- Διευκρινίσεις για τις οδηγίες
- Αναπαραστάσεις συνεντεύξεων
- Δοκιμαστική συμπλήρωση του ερωτηματολογίου
- Ανάλυση λαθών τους και κριτική αυτών
- Κατατοπισμός ερευνητών σε ορισμένα διοικητικά θέματα.

Σε κάθε έρευνα πρέπει να υπάρχουν δύο ομάδες προκειμένου να έχουμε συνεχή καθοδήγηση των ερευνητών, αλλά και βελτίωση της ποιότητας της όλης εργασίας:

- Ομάδα εποπτείας, η οποία θα παρακολουθεί τους ερευνητές κατά την διάρκεια λήψης των συνεντεύξεων για να διαπιστωθεί αν αντιλαμβάνονται σωστά τις έννοιες και αν εφαρμόζουν την επιβαλλόμενη διαδικασία,.
- Ομάδα επανασυνέντευξης, όπου θα επανεξετάζουν ένα μικρό ποσοστό των ατόμων που επισκέπτεται κάθε ερευνητής. Αν υπάρχει διαφορά στις απαντήσεις θα αναζητηθούν οι λόγοι.

1,6 Χρονοδιάγραμμα της έρευνας

Σε αυτό το στάδιο προσδιορίζουμε το χρονικό διάστημα ολοκλήρωσης μιας έρευνας.

Η χρονική διάρκεια μιας έρευνας καθορίζεται με βάση το χρόνο που απαιτείται για το σχεδιασμό και την προετοιμασία της έρευνας, το μέγεθος του δείγματος, τη μέθοδο συλλογής των στοιχείων, το χρόνο για την δοκιμαστική έρευνα, το χρόνο για την συμπλήρωση των ερωτηματολογίων, το χρόνο για την επεξεργασία των στοιχείων και την στατιστική ανάλυση και παρουσίαση των αποτελεσμάτων υπό τη μορφή μελέτης.

Ένα χρονοδιάγραμμα μπορεί να περιλαμβάνει τις εξής φάσεις:

Πρώτη φάση: Διάρκεια 3 μήνες

Σ' αυτή τη φάση γίνονται οι επεξεργασίες

- α) Συσκέψεις των μελών της ερευνητικής ομάδας
- β) Μελέτη της Ελληνικής και διεθνούς βιβλιογραφίας
- γ) Κατάρτιση ερωτηματολογίου και επιλογή δείγματος
- δ) Κατάρτιση του προϋπολογισμού.

Δεύτερη φάση: Διάρκεια 6 μήνες

Στη φάση αυτή γίνεται η συλλογή των στατιστικών στοιχείων από τις πηγές.

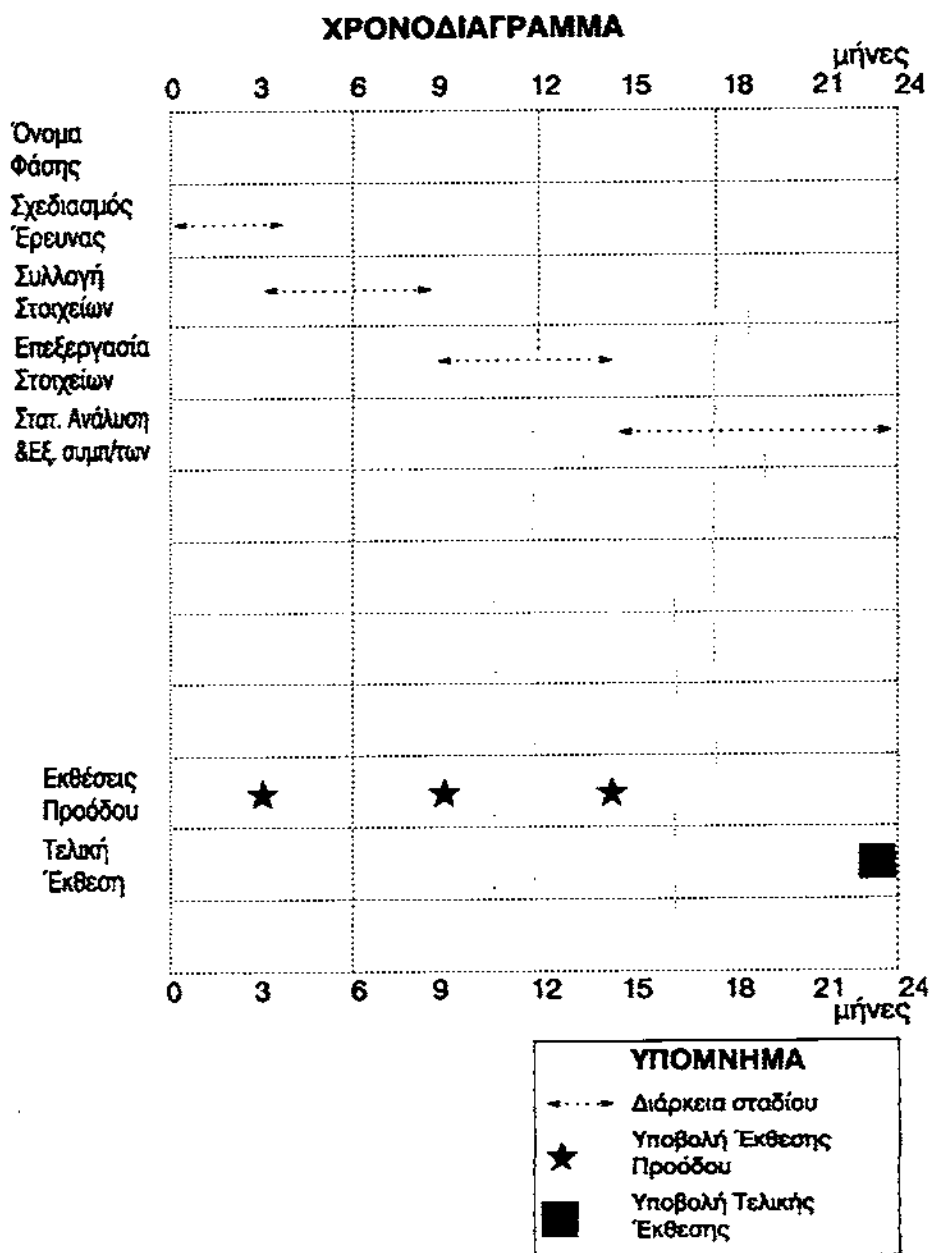
Τρίτη φάση: Διάρκεια 6 μήνες

Εδώ γίνεται η οργάνωση και επεξεργασία των συγκεντρωθέντων στοιχείων, καθώς και η κατάρτιση των απαιτούμενων στατιστικών πινάκων συχνοτήτων.

Τέταρτη φάση: Διάρκεια 9 μήνες

Εδώ γίνεται στατιστική ανάλυση των συγκεντρωθέντων στοιχείων, μελλοντικές προβλέψεις, διατύπωση τελικών συμπερασμάτων, σύντομη και εκτύπωση τελικής μελέτης.

(Η γραφική απεικόνιση του παρακάτω χρονοδιαγράμματος μπορεί να έχει την παρακάτω μορφή:)



1,7 Προϋπολογισμός της έρευνας

Ο προϋπολογισμός μιας έρευνας περιλαμβάνει το σύνολο των δαπανών που απαιτούνται για την διαδικασία διεξαγωγής μιας έρευνας. Αυτό είναι συνάρτηση πολλών παραγόντων, όπως π.χ. το μέγεθος του δείγματος, το επιστημονικό και βοηθητικό προσωπικό, την ακρίβεια της

έρευνας και άλλους παράγοντες. Η ανάλυση του προϋπολογισμού θα έχει την εξής μορφή πχ:

Αμοιβές επιστημονικής ομάδας			
Όνοματεπώνυμο ερευνητών	Διάρκεια απασχόλησης	Μηνιαία αμοιβή	Σύνολο
A.B.	14	75.000	1.050.000
K.N.	14	70.000	980.000
B.Γ.	14	70.000	980.000
N.Π.	8	70.000	560.000
K.Χ.	8	70.000	560.000
	Σύνολο		4.130.000
Γραμματειακή υποστήριξη			100.000
Έξοδα δειγματοληπτικής έρευνας			200.000
Έξοδα προγραμματιστών - αναλυτών Η/Υ			150.000
Δακτυλογραφήσεις			100.000
Υλικά για Η/Υ			50.000
Απρόβλεπτα έξοδα			200.000
Σύνολο αμοιβής Επιστημ. Ομάδας και γενικών εξόδων			4.930.000
Πασσστό ΚΟΔΕ Πανεπιστημίου Πειραιώς 15% x 4.930.000			739.500
Γενικό Σύνολο			5.669.500

1,8 Δοκιμαστική έρευνα

Σ' αυτό το στάδιο πραγματοποιείται μια δοκιμαστική έρευνα. Η διενέργεια μιας δοκιμαστικής έρευνας εξυπηρετεί πολλούς σκοπούς. Ο σημαντικότερος είναι η υποβολή του ερωτηματολογίου σε δοκιμή προκειμένου να γίνουν οι τελικές διορθώσεις. Ακόμα να διαπιστωθεί η επάρκεια των οδηγιών που έχουν δοθεί στους ερευνητές, η δοκιμή της εποπτείας και της οργάνωσης στην πράξη, καθώς και ο υπολογισμός του απαιτούμενου χρόνου για την συμπλήρωση των ερωτηματολογίων και την επεξεργασία τους. Επίσης με την δοκιμαστική έρευνα έχουμε τις πρώτες ενδείξεις σχετικά με το ποσοστό ανταπόκρισης.

Η επιτυχία μιας έρευνας εξασφαλίζεται όταν η δοκιμαστική έρευνα γίνεται με συνθήκες όσο το δυνατόν παρόμοιες με τις αντίστοιχες της κύριας έρευνας.

Μετά τη δοκιμαστική έρευνα, αφού διορθωθούν τυχόν προβλήματα ή λάθη που υπάρχουν γίνεται η σύνταξη του οριστικού ερωτηματολογίου.

1,9 Επεξεργασία στατιστικών στοιχείων

Το επόμενο στάδιο στην διαδικασία διεξαγωγής μιας έρευνας είναι η επεξεργασία και η μηχανογραφική οργάνωση των στοιχείων που έχουν συγκεντρωθεί από το οριστικό ερωτηματολόγιο. Το στάδιο της επεξεργασίας είναι πολύ κουραστικό και δαπανηρό, γιατί περιλαμβάνει τον έλεγχο όλων των ερωτηματολογίων για τυχόν ασυμπλήρωτες, ασαφείς, δυσανάγνωστες ή ασυμβίβαστες απαντήσεις. Ακολουθεί η διαλογή των πληροφοριών των διαφόρων χαρακτηριστικών των στατιστικών μονάδων και η εμφάνιση των χαρακτηριστικών αυτών σε αριθμητικούς πίνακες με βάση διάφορα κριτήρια. Η διαλογή των στατιστικών πληροφοριών μπορεί να γίνει είτε με το χέρι όταν ο αριθμός των ερωτηματολογίων και των χαρακτηριστικών είναι περιορισμένος, είτε με μηχανικά μέσα όταν ο αριθμός των ερωτηματολογίων και των χαρακτηριστικών είναι μεγάλος π.χ. Η/Υ, υπολογιστικές μηχανές, κ.τ.λ.

Η επεξεργασία και ταξινόμηση των στατιστικών στοιχείων με μηχανογραφικά μέσα περιλαμβάνει τέσσερα στάδια:

α) Κωδικογράφηση

Στο στάδιο αυτό τα στοιχεία που περιέχονται στα ερωτηματολόγια μετατρέπονται σε κωδικούς αριθμούς, οι οποίοι σημειώνονται στο δεξιό άκρο του ερωτηματολογίου.

Τα δελτία μετά την κωδικογράφηση παραδίνονται στην χειρίστρια της διατρητικής μηχανής.

β) Διάτρηση

Οι κωδικοί αριθμοί όλων των δελτίων μεταφέρονται σε ειδικές καρτέλες με τη βοήθεια της διάτρησης. Ακολουθεί το στάδιο της

επαλήθευσης με τις επαληθευτικές μηχανές για να εξετάσουμε μήπως οι καρτέλες έχουν τρυπηθεί εσφαλμένα.

γ) Διαλογή

Τα δελτία αφού έχουν διατηρηθεί μεταφέρονται στις διαλογικές μηχανές όπου τα χωρίζουν και τα ταξινομούν σε διάφορες ομάδες. Τα αριθμούν και τα τοποθετούν χωριστά σε 12 υποδοχείς κάθε μηχανής.

δ) Πινακογράφηση

Μετά την διαλογή ακολουθεί η διαδικασία της πινακογράφησης. Με την πινακογραφική μηχανή μπορεί κανείς να μετρήσει τα δελτία κάθε κατηγορίας, να διαβάσει τις διατρήσεις, να κάνει υπολογισμούς πάνω σ' αυτά τα δεδομένα (π.χ. πρόσθεση, πολλαπλασιασμό) αλλά και αν εκτυπώσει ό,τι αναφέρεται στα δελτία.

Παρουσίαση στατιστικών στοιχείων

Σ' αυτό το στάδιο γίνεται η παρουσίαση των στατιστικών στοιχείων υπό μορφή κατανομών συχνοτήτων (πινάκων) και γραφικών παραστάσεων, κατά τρόπο που να διευκολύνει την ανάλυση των στοιχείων αυτών και την εξαγωγή χρήσιμων συμπερασμάτων.

Η παρουσίαση αυτή μπορεί να γίνει με τρεις τρόπους:

- α) με τη μορφή πινάκων
- β) με τη μορφή γραφικών παραστάσεων
- γ) με τη μορφή εκθέσεων ή αναφορών

A) Στατιστικοί πίνακες

Η παρουσίαση των στατιστικών στοιχείων σε πίνακες, γίνεται με την τοποθέτηση των στατιστικών πληροφοριών σε στήλες και γραμμές, με τέτοιο τρόπο που να διευκολύνονται η σύγκριση των στοιχείων και η καλύτερη ενημέρωση του αναγνώστη. Ανάλογα με την χρησιμοποίηση, οι στατιστικοί πίνακες διακρίνονται σε λεπτομερείς και σε συνοπτικούς.

Οι λεπτομερείς πίνακες περιέχουν κάθε διαθέσιμη πληροφορία για μια στατιστική έρευνα και διευκολύνουν τον αναγνώστη στην παροχή κάθε πληροφορίας που τον ενδιαφέρει.

Αντίθετα, οι συνοπτικοί πίνακες παρουσιάζουν τα στοιχεία του πληθυσμού συνοπτικά, τα οποία λαμβάνονται από τους λεπτομερείς πίνακες. Τους συνοπτικούς πίνακες τους χρησιμοποιούμε κυρίως για συγκρίσεις που αποτελούν την βάση της στατιστικής ανάλυσης για εξαγωγή χρήσιμων συμπερασμάτων.

Τους διακρίνουμε δε σε δύο κατηγορίες:

α) Πίνακες απλής εισόδου όπου παρουσιάζουν ένα φαινόμενο από την άποψη μόνο ενός χαρακτηριστικού και χρησιμοποιούνται συνήθως για συγκρίσεις και εξαγωγή συμπερασμάτων,

Π.χ. στον πίνακα δίνεται ο πληθυσμός της Ελλάδας από το 1920 έως το 1991.

Εξέλιξη του ελληνικού πληθυσμού 1920-1991

Χρόνια	Πληθυσμός
1920	5.016.886
1928	6.204.684
1940	7.344.860
1951	7.637.801
1961	8.388.553
1971	8.768.641
1981	9.739.500
1991	10.269.907

Πηγή: Ε.Σ.Υ.Ε.

Επίσης, άλλα παραδείγματα πινάκων απλής εισόδου, με βάση διάφορα γεωγραφικά, ποιοτικά, ποσοτικά κριτήρια είναι και τα παρακάτω

Πληθυσμός της γης στο χρονικό διάστημα 1650-1974

Χρόνια	Πληθυσμός της γης σε εκατομ.
1650	545
1750	728
1800	906
1850	1.171
1900	1.680
1930	2.008
1970	3.635
1974	3.900

Πηγή Ο.Η.Ε
Demographic Yearbook, 1970

Πληθυσμός της γης κατά περιοχές (1970)

Περιοχές	Πληθυσμός το 1970 σε εκατομ.
Ευρώπη	462
Ασία	2.056
Αμερική	511
Αφρική	344
Ωκεανία	19
Ρωσία	243

Πηγή: Ο.Η.Ε. 1970

Τροχαία ατυχήματα στις διάφορες ημέρες της εβδομάδας το 1982 στην Ιταλία.

Ημέρες της εβδομάδας	Ατυχήματα
----------------------	-----------

Δευτέρα	45.154
Τρίτη	50.461
Τετάρτη	43.654
Πέμπτη	42.742
Παρασκευή	43.215
Σάββατο	45.062
Κυριακή	46.195
Πηγή: I.S.T.A.T. Statistica degli, 1972	

**Κατανομή του πληθυσμού της Ελλάδας σε γεωγραφικά
διαμερίσματα σύμφωνα με την απογραφή του 1971**

Γεωγραφικά διαμερίσματα	Αριθμός	Ποσοστό % Κατοίκων
Περιφέρεια πρωτεύουσας Υπόλοιπη Στερεά Ελλάδα & Εύβοια	2.540.241	28,9
Πελοπόννησος	992.077	11,4
Ιόνια νησιά	986.912	11,2
Ήπειρος	184.443	2,2
Θεσσαλία	310.334	3,5
Μακεδονία	659.913	7,5
Θράκη	1.890.684	21,6
Νησιά αιγαίου	329.582	3,7
Κρήτη	417.813	4,7
Κρήτη	456.642	5,3
Σύνολο χώρας	8.768.641	100
Πηγή : Ε.Σ.Υ.Ε.		

Αριθμός τέκνων (X ₁)	Αριθμός οικογενειών
0	200
1	1000
2	2500
3	2000
4	1600
5	700
6	300
7	150
8	80
Σύνολο	8.530

β) Οι πίνακες διπλής εισόδου όπου μας δίνουν πληροφορίες για ένα πληθυσμό από την άποψη δύο ποσοτικών ή ποιοτικών χαρακτηριστικών. Αν πρόκειται για τη μελέτη ενός πληθυσμού από την άποψη δύο ποσοτικών μεταβλητών (π.χ. βάρος, ύψος) θα έχει την εξής μορφή ο πίνακας.

Πίνακας

τάξεις	X ₁	Τάξεις						
		$\beta_0 + \beta_1$	$B_1 - \beta_2$	$\beta_{j-1} - \beta_1$...	$\beta_{\lambda-1} - \beta_\lambda$	
		y _i						
		y ₁	y ₂	y _j	...	y _λ	F ₁
$\alpha_0 - \alpha_1$	X ₁	F ₁₁	F ₁₂	F _{1j}	...	F _{1λ}	F ₁
$\alpha_1 - \alpha_2$	X ₂	F ₂₁	F ₂₂	F _{2j}	...	F _{2λ}	F ₂
.
.
.
$\alpha_{j-1} - \alpha_j$	X _j	F _{j2}	F ₁₂	F _{ij}	...	F _{1λ}	F _j
.
.
.
$\alpha_{κ-1} - \alpha_κ$	X _κ	F _{κ1}	F _{κ2}	F _{κj}	...	F _{κλ}	F _κ
	F _j	F _{.1}	F ₂	F _j		F _{.λ}	N

Αν πρόκειται για ποιοτικά χαρακτηριστικά (π.χ. εθνικότητα, χρώμα μαλλιών) θα έχει την εξής μορφή.

Χαρακτηριστικό Β	Χαρακτηριστικό Α					
	A ₁	A ₂	...	A _j	...	A _κ
B ₁	F ₁₁	F ₁₂	...	F _{1j}	...	F _{1κ}
B ₂	F ₂₁	F ₂₂	...	F _{2j}	...	F _{2κ}
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-

B_i	F_{i1}	F_{i1}	...	F_{ij}	...	F_{ik}
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
B_n	F_{n1}	F_{n2}	...	F_{nj}	...	F_{nk}

Κατά τη σύνταξη των πινάκων όποια μορφή και να έχουν θα πρέπει να τηρούνται ορισμένες προϋποθέσεις:

- 1) να είναι απλοί
- 2) να διευκολύνουν τις συγκρίσεις
- 3) να δίνεται προσοχή σε ορισμένα χαρακτηριστικά στοιχεία
- 4) κάθε πίνακας πρέπει να έχει τίτλο, όπου θα γράφεται στο μέσο του πάνω μέρους του και να είναι περιληπτικός
- 5) να αναφέρεται η πηγή από την οποία προέρχονται τα στοιχεία στο κάτω μέρος του πίνακα
- 6) το κύριο σώμα του πίνακα να περιέχει στατιστικά στοιχεία που είναι δυνατό να αναφέρονται σε γεωγραφικές, χρονολογικές, ποιοτικές και ποσοτικές κατατάξεις.

B) Γραφικές παραστάσεις

Οι γραφικές παραστάσεις αποτελούν το καλύτερο μέσο στατιστικής παρουσίασης, γιατί δίνουν στους αφηρημένους αριθμούς μια συγκεκριμένη μορφή που διευκολύνει να έχουμε μια άμεση αντίληψη της μορφής του φαινομένου το οποίο θέλουμε να μελετήσουμε.

Μια καλοσχεδιασμένη γραφική παράσταση μπορεί να γίνει πιο κατανοητή και να διατηρηθεί στην μνήμη του αναγνώστη ευκολότερα, απ' ότι σε ένα αριθμητικό πίνακα.

Οι γραφικές παραστάσεις δεν πρέπει να περιέχουν πολλές λεπτομέρειες γιατί τότε χάνουν κάθε ενδιαφέρον.

Κάθε γραφική παράσταση, εκτός από το σχέδιο θα πρέπει να περιλαμβάνει:

α) τον τίτλο

β) την κλίμακα των τιμών των μεγεθών που απεικονίζονται

γ) την ένδειξη των πηγών

δ) Υπόμνημα, το οποίο συνήθως γράφεται κάτω δεξιά από το σχήμα και εξηγεί τις διάφορες γραμμές που περιέχει η γραφική παράσταση.

Τις γραφικές παραστάσεις τις διακρίνουμε σε τρεις βασικές κατηγορίες:

i) Διαγράμματα

Για την κατασκευή διαγραμμάτων χρειάζεται εκτός από γνώση των στατιστικών κανόνων και πείρα, δεξιότητες και φαντασία αυτού που τα κατασκευάζει. Υπάρχουν πολλά είδη διαγραμμάτων, τα πιο γνωστά είναι: κυκλικά, τα ακιδωτά, τα χρονολογικά, κ.ά.

ii) Χαρτογράμματα

Τα χαρτογράμματα είναι γραφικές παραστάσεις στατιστικών στοιχείων σε γεωγραφικούς ή τοπογραφικούς χάρτες.

iii) Ειδογράμματα

Πρόκειται για διάφορα σχήματα σε μορφή προσώπων ή πραγμάτων, θεωρούνται ως μέσα πολύ εκφραστικά ενός φαινομένου και μπορεί ο αναγνώστης να συγκρατήσει την εικόνα τους ευκολότερα στη μνήμη του.

Γ) Στατιστικές εκθέσεις ή αναφορές

Άλλη μια μορφή παρουσίασης των στατιστικών στοιχείων είναι οι εκθέσεις ή αναφορές στο κείμενο των οποίων αναφέρονται τα κύρια σημεία των αποτελεσμάτων, σχολιάζεται η σημασία τους και γίνονται παρατηρήσεις από το άτομο που την σύνταξε.

Η μορφή αυτή παρουσίασης στατιστικών στοιχείων έχει πολλά μειονεκτήματα γιατί ο αναγνώστης είναι υποχρεωμένος να την διαβάσει ολόκληρη την έκθεση πολύ προσεκτικά για να μπορεί να συγκρίνει τα διάφορα αριθμητικά στοιχεία και να συγκρατήσει αυτά που τον ενδιαφέρουν.

Στατιστική ανάλυση των εμπειρικών δεδομένων της έρευνας

Στο στάδιο αυτό της στατιστικής ανάλυσης και ερμηνείας των αποτελεσμάτων των εμπειρικών δεδομένων, υπολογίζονται διάφορες παράμετροι, δείκτες, αλληλουχίες και αλληλεξαρτήσεις, ανακαλύπτουμε τους νόμους που διέπουν τα διάφορα φαινόμενα, γίνονται στατιστικοί έλεγχοι, κάνουμε μελλοντικές προβλέψεις, αξιολογούμε τα στατιστικά ευρήματα και διατυπώνονται εισηγήσεις προς τους αρμόδιους φορείς ή γίνονται υποδείξεις για αξιοποίηση των αποτελεσμάτων της έρευνας.

Η οργάνωση, η επεξεργασία των στοιχείων, η κατασκευή των κατανομών συχνοτήτων και η στατιστική ανάλυση είναι τα στάδια εκείνα που αποτελούν την εκτελεστική φάση της έρευνας.

1,10 Παρουσίαση των αποτελεσμάτων της ερευνητικής εργασίας

Στο στάδιο αυτό που είναι και το τελευταίο, γίνεται η σύνταξη μιας μελέτης όπου περιέχει την εισαγωγή όπου αναφέρεται στους σκοπούς της έρευνας, την μεθοδολογία που έχει ακολουθηθεί και εφαρμοστεί, το κύριο μέρος της επιστημονικής έρευνας, τα αποτελέσματα, τα συμπεράσματα, τυχόν αδυναμίες της έρευνας, οι πηγές και τα βοηθήματα που έχουν χρησιμοποιηθεί, η βιβλιογραφία καθώς και παραρτήματα πινάκων και γραφικών παραστάσεων.

ΚΕΦΑΛΑΙΟ 2

ΒΑΣΙΚΕΣ ΑΡΧΕΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Είναι φανερό ότι το πρώτο στάδιο μιας στατιστικής μελέτης είναι η προσεκτική συλλογή των δεδομένων. Σε αντίθεση με την κλασική στατιστική συμπερασματολογία, όπου υποθέτουμε ότι χρησιμοποιούμε δείγματα που προέρχονται από άπειρους πληθυσμούς, σε μια δειγματοσκοπική έρευνα ενδιαφερόμαστε κυρίως για πεπερασμένους πληθυσμούς, οι οποίοι πρέπει να ορίζονται πλήρως, δηλαδή να μπορούν να καταγραφούν και να μετρηθούν.

Οι αρχές και οι μέθοδοι για τη συλλογή και ανάλυση δεδομένων από πεπερασμένους πληθυσμούς είναι ένας κλάδος της Στατιστικής, γνωστός σαν «Μέθοδοι Δειγματοληπτικών Ερευνών» (Sample Survey Methods) ή γενικότερα σαν Δειγματοληψία (Sampling).

Η ανάλυση δειγματοληπτικών ερευνών είναι το κυριότερο εργαλείο έρευνας σε ένα μεγάλο φάσμα εφαρμογών, όπως στη Βιομηχανία, Εκπαίδευση, Ιατρική, Ψυχολογία, Κοινωνιολογία, Γεωπονία, Αθλητισμό κ.λπ.

Οι δειγματοληπτικές έρευνες των ανθρωπίνων πληθυσμών αποτελούν σπουδαίες πηγές βασικής γνώσης των κοινωνικών επιστημών. Οικονομολόγοι, ψυχολόγοι, πολιτικοί επιστήμονες και κοινωνιολόγοι μελετούν ποικίλα θέματα, όπως πρότυπα εσόδων – εξόδων των οικογενειών, την επίδραση της επαγγελματικής απασχόλησης των γυναικών στην οικογενειακή ζωή, τις συγκοινωνιακές και ταξιδιωτικές συνήθειες των κατοίκων μιας πόλης, τις προτιμήσεις των ψηφοφόρων για τους υποψήφιους και τις θέσεις αυτών. Η Εθνική Στατιστική Υπηρεσία κάθε χώρας διενεργεί σε τακτά χρονικά διαστήματα (συνήθως κάθε μήνα) μια δειγματοληπτική έρευνα για να πάρει πληροφορίες για τον

πληθυσμό, την απασχόληση και την ανεργία στη χώρα. Ανάλογα με τα αποτελέσματα αυτά διαμορφώνεται και η κυβερνητική πολιτική στα θέματα αυτά.

Το πρώτο βήμα σε μια δειγματοληπτική έρευνα είναι ο ακριβής προσδιορισμός του πληθυσμού για τον οποίο ενδιαφερόμαστε και από τον οποίο θα πάρουμε το δείγμα. Εάν π.χ. θέλουμε να εξετάσουμε παιδιά μετανάστες πρέπει να ορίσουμε πλήρως τον πληθυσμό που αποτελείται από παιδιά μετανάστες. Δηλαδή να οριστεί η ηλικία που προσδιορίζει τον όρο «παιδί» και ποιες χώρες συνεπάγονται «μετανάστευση». Αν θέλουμε να ορίσουμε τον πληθυσμό των μαθητών που φοιτούν σε γυμνάσιο θα πρέπει να αποφασίσουμε αν θέλουμε να περιοριστούμε μόνο στα δημόσια γυμνάσια ή θα συμπεριλάβουμε και τα ιδιωτικά ή και άλλες ισοδύναμες σχολές, όπως π.χ. σχολές για απροσάρμοστα παιδιά.

Στη θεωρία δειγματοληψίας εκτός των απλών στοιχείων ή μονάδων, χρησιμοποιούμε και την έννοια των **δειγματοληπτικών μονάδων**, οι οποίες είναι μη επικαλυπτόμενες συλλογές απλών στοιχείων του πληθυσμού. Έτσι π.χ. αν ο πληθυσμός μας είναι το σύνολο των κατοίκων μιας περιοχής, θα μπορούσαμε να θεωρήσουμε σαν δειγματοληπτικές μονάδες τα νοικοκυριά ή τα διαμερίσματα ή τις πολυκατοικίες της περιοχής. Είναι φανερό ότι οι δειγματοληπτικές μονάδες δεν συμπίπτουν πάντοτε με τα απλά στοιχεία του πληθυσμού.

Αν φυσικά κάθε δειγματοληπτική μονάδα περιέχει ένα και μόνο ένα στοιχείο από τον πληθυσμό, τότε μια δειγματοληπτική μονάδα και ένα στοιχείο του πληθυσμού είναι τα ίδια.

Το σύνολο των δειγματοληπτικών μονάδων που αντιστοιχούν σε ένα πληθυσμό, λέγεται **δειγματοληπτικό πλαίσιο**. Το δειγματοληπτικό πλαίσιο είναι συνήθως ένας κατάλογος ονομάτων φυσικών προσώπων, αντικειμένων, οικοδομικό σχέδιο πόλεως, μια αεροφωτογραφία ενός νησιού. Έτσι, αν καθορίσουμε σαν δειγματοληπτική μονάδα σε

πληθυσμό με στοιχεία τα άτομα που έχουν δικαίωμα ψήφου τον κάθε ψηφοφόρο χωριστά, τότε ένας κατάλογος με όλους τους εγγεγραμμένους ψηφοφόρους μπορεί να χρησιμοποιηθεί σαν το πλαίσιο για μια δειγματοσκόπηση (γκάλοπ) για τις προσεχείς εκλογές. Είναι φανερό ότι αυτό το πλαίσιο (κατάλογος) δεν περιέχει όλα τα άτομα του πληθυσμού των ψηφοφόρων γιατί είναι αδύνατη η συμπλήρωση του καταλόγου καθημερινά.

Η διαδικασία που θα χρησιμοποιήσουμε για την επιλογή του δείγματος, όπως και το μέγεθος του δείγματος που θα εκλεγεί, είναι σημαντικά στοιχεία για να έχουμε πράγματι ένα αντιπροσωπευτικό δείγμα του υπό εξέταση πληθυσμού. Αυξάνοντας το μέγεθος δείγματος, ασφαλώς θα έχουμε και καλύτερη ακρίβεια στην εκτίμηση των χαρακτηριστικών του πληθυσμού. Από την άλλη πλευρά όμως το κόστος για τη δειγματοληψία, την επεξεργασία και παρουσίαση των αποτελεσμάτων αυξάνει. Πολύ μεγάλο δείγμα συνεπάγεται σπατάλη χρόνου, χρήματος και κόπου. Αντίθετα, πολύ μικρό δείγμα είναι δυνατό να οδηγήσει σε μεροληπτικές εκτιμήσεις, αν και θα πρέπει να σημειωθεί ότι μια «προσεκτική» επιλογή μικρού δείγματος είναι δυνατό να δώσει πολύ καλύτερα αποτελέσματα από ένα μεγαλύτερο δείγμα που δεν έχει εκλεγεί κατάλληλα.

Ενδεικτικό εδώ είναι το παράδειγμα των προεδρικών εκλογών στις ΗΠΑ το 1936, όπου υπήρξε πλήρης αποτυχία της πρόγνωσης των εκλογικών αποτελεσμάτων. Η δειγματοληψία στηρίχθηκε στην εκλογή 10.000.000 ατόμων του πληθυσμού από λίστα – πλαίσιο των τηλεφωνικών καταλόγων των διαφόρων πόλεων. Στο ερωτηματολόγιο που ταχυδρομήθηκε απάντησε περίπου το $\frac{1}{4}$ των επιλεγέντων και από την ανάλυση που έγινε φάνηκε ότι το εκλογικό σώμα θα έδινε πλειοψηφία στους Ρεπουμπλικάνους. Η μέρα των εκλογών όμως επιφύλαξε μια δυσάρεστη έκπληξη τόσο για τους Ρεπουμπλικάνους όσο

και για τους στατιστικούς αναλυτές της δημοσκόπησης: λιγότερο από 40% των ψηφοφόρων προτίμησαν το Ρεπουμπλικανικό κόμμα, και ο Δημοκρατικός υποψήφιος Ρούσβελτ εκλέχθηκε με ιστορική πλειοψηφία. Το λάθος ασφαλώς της πρόγνωσης οφείλετο στο γεγονός ότι οι τηλεφωνικοί κατάλογοι, εκείνη την εποχή τουλάχιστον, δεν περιλάμβαναν τις κατώτερες κοινωνικο-οικονομικές τάξεις.

Το κυριότερο μειονέκτημα μιας δειγματοληπτικής μεθόδου έναντι ολικής απογραφής του πληθυσμού είναι ότι οι εκτιμήσεις των χαρακτηριστικών του πληθυσμού από ένα μέρος αυτού έχουν τα λεγόμενα «σφάλματα δειγματοληψίας». Με κατάλληλο όμως μηχανισμό για την επιλογή των μονάδων του δείγματος είναι ο δυνατός ο περιορισμός των σφαλμάτων αυτών και, το σπουδαιότερο, η πιθανοθεωρητική εκτίμηση του μεγέθους αυτών.

Για την αξιολόγηση της ακρίβειας μιας δειγματοληπτικής έρευνας και για λόγους ευκολίας, τα σφάλματα ταξινομούνται σε δύο κατηγορίες:

1. Δειγματοληπτικά σφάλματα
2. Μη δειγματοληπτικά σφάλματα.

Τα σφάλματα της πρώτης κατηγορίας μπορεί να προκύψουν από τη μη σωστή επιλογή της κατάλληλης μεθόδου δειγματοληψίας, και την επιλογή ακατάλληλου δείγματος, όσον αφορά τη μεταβλητότητα και το μέγεθος του δείγματος.

Μια καλή δειγματοληπτική έρευνα πρέπει να περιλαμβάνει τον υπολογισμό των δειγματοληπτικών σφαλμάτων. Οι πληροφορίες για τα δειγματοληπτικά σφάλματα θα πρέπει να είναι άμεσα διαθέσιμες σε όσους προτίθενται να χρησιμοποιήσουν τα στατιστικά αποτελέσματα. Δηλαδή θα πρέπει να αναφέρεται τόσο η δειγματοληπτική μέθοδος, όσο και το σφάλμα της εκτίμησης μιας παραμέτρου.

Ο όρος «μη δειγματοληπτικό σφάλμα» προέρχεται από το γεγονός ότι αυτός ο τύπος σφάλματος μπορεί να συμβεί σε οποιαδήποτε έρευνα,

είτε αυτή γίνεται με τη μέθοδο της απογραφής, είτε με τη μέθοδο της δειγματοληψίας.

Δυστυχώς, σε αντίθεση με τα δειγματοληπτικά σφάλματα, δεν υπάρχει μια απλή και άμεση μέθοδος για την εκτίμηση του μεγέθους των μη δειγματοληπτικών σφαλμάτων. Τα μη δειγματοληπτικά σφάλματα μπορούν να ταξινομηθούν σε δύο κατηγορίες: α) **Μη συστηματικά** (ή σφάλματα τυχαίου τύπου) των οποίων οι επιδράσεις προσεγγιστικά αλληλοαναιρούνται όταν χρησιμοποιούνται αρκετά μεγάλου μεγέθους δείγματα, και β) **Συστηματικά** (ή μεροληψίες) τα οποία οδηγούν τις εκτιμήσεις μας πάντοτε προς τα πάνω ή πάντοτε προς τα κάτω από την πραγματική τιμή και γι' αυτό δεν μπορούν να αλληλοεξουδετερωθούν, π.χ. μια ζυγαριά ζυγίζει τα πάντα κατά (έστω) 50gr βαρύτερα.

Κλείνουμε την εισαγωγική αυτή παράγραφο, αναφέροντας ορισμένα στοιχεία για τις τεχνικές συλλογής των δεδομένων που απαιτεί μια δειγματοληπτική έρευνα.

Οι πληροφορίες συλλέγονται συνήθως με τη βοήθεια τυποποιημένων ερωτήσεων. Οι ερωτήσεις είναι κοινές για όλα τα άτομα του πληθυσμού και έτσι κάθε εξεταζόμενο άτομο απαντά στις ίδιες ακριβώς ερωτήσεις. Οι πληροφορίες που συγκεντρώνονται είναι εμπιστευτικές και με καμία περίπτωση δεν πρέπει να αποκαλύπτονται. Στην Ελλάδα το απόρρητο των στοιχείων των δειγματοληπτικών ερευνών προστατεύεται από το άρθρο 40 του Ν.Δ. 3627/56 και οι παραβάτες τιμωρούνται με αυστηρές ποινές.

Οι κυριώτεροι τρόποι συλλογής πληροφοριών με ερωτηματολόγιο είναι:

α) **Ταχυδρομικά:** Η μέθοδος αυτή έχει σχετικά μικρό κόστος και είναι αποτελεσματική με τα μέλη ειδικών κατηγοριών ανθρώπων όπως π.χ. τα μέλη ενός επαγγελματικού σωματείου. Το μειονέκτημα της

μεθόδου είναι ότι το ποσοστό αυτών που απαντούν είναι συνήθως μικρό, πολλές φορές μάλιστα κυμαίνεται στο 20-30%.

β) Τηλεφωνικά: Η τηλεφωνική συνέντευξη είναι μια αποτελεσματική μέθοδος συλλογής ορισμένων δεδομένων, έχει και αυτή μικρό κόστος, αλλά έχει το μειονέκτημα ότι πολλά άτομα δεν διαθέτουν τηλέφωνο. Υπάρχει επίσης ο κίνδυνος παρανόησης των ερωτήσεων εκ μέρους του απαντητή.

γ) Προσωπική συνέντευξη: Εδώ έχουμε μεγάλο κόστος, αλλά έχουμε και το πλεονέκτημα της πληρέστερης ενημέρωσης του απαντητή σχετικά με τις ερωτήσεις. Σε περιπτώσεις που πρόκειται να συλλεγούν πολύπλοκες πληροφορίες, η μέθοδος αυτή καθίσταται αναγκαία.

δ) Συνδυασμός των παραπάνω: Μπορούμε π.χ. να τηλεφωνήσουμε στα άτομα που δεν έστειλαν ταχυδρομικά το ερωτηματολόγιό τους, ή επίσης να κλείσουμε ραντεβού για προσωπική συνέντευξη.

ε) Αυτόματα: Η μέθοδος αυτή εφαρμόζεται σε περιπτώσεις που δεν έχουμε ερωτηματολόγιο, αλλά μας ενδιαφέρουν κάποιες μετρήσεις, όπως π.χ. κίνηση οχημάτων σ' ένα δρόμο σε διάφορες χρονικές περιόδους, καταμέτρηση ακροαματικότητας κάποιας τηλεοπτικής εκπομπής, κ.λπ.

ΚΕΦΑΛΑΙΟ 3

Δειγματοληπτικές Μέθοδοι (Τεχνικές)

Υπάρχουν πολλές δειγματοληπτικές μέθοδοι. Παρακάτω θα αναφέρουμε τις κυριότερες πιθανοτικές δειγματοληψίες και μια μη – πιθανοτική μέθοδο.

3,1 Απλή Τυχαία Δειγματοληψία (χωρίς επανάθεση)

Η απλή τυχαία δειγματοληψία είναι μια Ίσης Πιθανότητας Επιλογής Μέθοδος (ΠΙΕΜ). Σ' αυτή τη μέθοδο κάθε δυνατό υποσύνολο n στοιχείων (μονάδων) από έναν πληθυσμό N μονάδων έχει ίση πιθανότητα να γίνει το δείγμα μεγέθους n που επιλέχθηκε. Για να καταφέρουμε να επιλέξουμε αυτό το δείγμα μεγέθους n μπορούμε να ακολουθήσουμε την παρακάτω διαδικασία: αριθμούμε τις μονάδες του πληθυσμού N με διαφορετικούς αριθμούς και μετά με τη βοήθεια ενός πίνακα τυχαίων αριθμών, τραβάμε τυχαίους αριθμούς από τον πίνακα μέχρι να αντιστοιχίσουμε n διαφορετικά στοιχεία από τον πληθυσμό.

Η πιθανότητα επιλογής της μονάδας a_i σε ένα συγκεκριμένο τράβηγμα είναι $1/N$. Με την προϋπόθεση όμως ότι ένα προηγούμενο τράβηγμα έδωσε μια μονάδα a_i ($i \neq j$) τότε η πιθανότητα ότι το i -οστό τράβηγμα θα παράγει το a_i είναι $1/N-1$. Άρα τα τραβήγματα (οι μεταβλητές Y_i) δεν είναι ανεξάρτητα.

3,2 Απλή Τυχαία Δειγματοληψία με επανάθεση

Η απλή τυχαία δειγματοληψία με επανάθεση είναι και αυτή μια Ίσης Πιθανότητας Επιλογής Μέθοδος (ΠΙΕΜ). Η ομοιότητά της με την προηγούμενη μέθοδο είναι ότι και η προηγούμενη μπορεί να γίνει με τη μέθοδο αυτή, αρκεί να αποβάλλουμε τις επαναλήψεις ώσπου να επιλεγούν οι n διαφορετικές μονάδες του πληθυσμού. Η κύρια διαφορά τους είναι ότι στην απλή τυχαία δειγματοληψία με επανάθεση οι

μεταβλητές του δείγματος Y_1, Y_2, \dots, Y_n (τραβήγματα) δεν είναι μόνο ισόνομες όπως και στην προηγούμενη μέθοδο, αλλά και ανεξάρτητες. Δηλαδή η πιθανότητα επιλογής μιας μονάδας αι να λάβει μια τιμή από τις Y_1, Y_2, \dots, Y_n είναι $1/N$ αλλά επίσης το i -οστό τράβηγμα που παράγει το αι δεν αποκλείεται να ταυτίζεται με ένα προηγούμενο τράβηγμα. Στην απλή τυχαία δειγματοληψία με επανάθεση η διασπορά του δειγματικού μέσου είναι μεγαλύτερη από την προηγούμενη μέθοδο. Γι' αυτό η απλή τυχαία δειγματοληψία χωρίς επανάθεση είναι προτιμότερη.

3,3 Συστηματική Δειγματοληψία

Με την μέθοδο αυτή επιλέγουμε κάθε k -οστή μονάδα ενός πληθυσμού N (υποθέτουμε ότι k ακέραιο) αρχίζοντας από κάποια που επιλέγεται τυχαία με έναν τυχαίο αριθμό μεταξύ 1 και k . Η μέθοδος αυτή είναι επίσης μιας Ίσης Πιθανότητας Επιλογής Μέθοδος (ΙΠΕΜ). Η πιθανότητα επιλογής της μονάδας που επιλέχθηκε τυχαία στην αρχή είναι $1/k$, αφού υπάρχουν k συστηματικά δείγματα του ίδιου μεγέθους n . Π.χ. αν τραβήξουμε τυχαία τον αριθμό i , $1 \leq i \leq k$, τότε το i -οστό δείγμα θα είναι $Y_i, Y_{i+k}, Y_{i+2k}, \dots, Y_{i+(n-1)k}$.

Η μέθοδος αυτή είναι εύκολη στην εφαρμογή. Διαταραχές προκύπτουν όταν τα N δεν είναι ακέραια πολλαπλάσια του k οπότε τα συστηματικά δείγματα έχουν μέγεθος άλλα n και άλλα $n+1$. Τότε ο δειγματικός μέσος παρουσιάζει κάποια μεροληψία, όμως για μεγάλη n (π.χ. για $n > 50$) οι διαταραχές αυτές είναι αμελητέες. Όταν οι μετρήσεις παρουσιάζουν συγκεκριμένη περιοδικότητα, το συστηματικό δείγμα k δεν θα πρέπει να συνδέεται με την περίοδο των μετρήσεων, γιατί υπάρχει κίνδυνος να επαναλαμβάνει την ίδια πληροφορία. Όταν η διάταξη των μονάδων του πληθυσμού θεωρείται τυχαία ως προς τις μετρήσεις Y_i , τότε η συστηματική δειγματοληψία είναι ισοδύναμη με τυχαία δειγματοληψία.

3,4 Μέθοδοι Δειγματοληψίας με ομαδοποίηση του πληθυσμού

3,4,1 Στρωματική (Στρωματοποιημένη) δειγματοληψία

Στην στρωματική δειγματοληψία, συμπληρωματικές πληροφορίες για κάποια κοινά χαρακτηριστικά των στοιχείων ενός πληθυσμού N βοηθούν στο να χωριστεί ο πληθυσμός σε ομάδες ή στρώματα και μετά τυχαία δείγματα επιλέγονται από κάθε στρώμα με δειγματικά μεγέθη προσδιοριζόμενα εκ των προτέρων. Συχνά το ίδιο δειγματοληπτικό κλάσμα $n_h/N_h = n/N$, $h=1, 2, \dots, \leq H$, ο αριθμός των στρωμάτων χρησιμοποιείται για όλα τα στρώματα (χωρίς αυτό να είναι απαραίτητο), οπότε μιλάμε για την αναλογική στρωματοποίηση, μια ΠΙΕΜ μέθοδο. Η στρωματική δειγματοληψία είναι μια μέθοδος, η οποία χρησιμοποιείται πολύ συχνά, γιατί αυξάνει την ομοιογένεια των ομάδων στις οποίες χωρίζεται ο πληθυσμός και έτσι βελτιώνει την ακρίβεια των εκτιμητών της επισκόπησης. Η ποσότητα $N_h/N = W_h$ λέγεται το βάρος (στάθμη) για το στρώμα h , όπου N_h το μέγεθος του στρώματος n .

Αξίζει να σημειωθεί ότι η συστηματική δειγματοληψία που εξετάσαμε προηγουμένως στρωματοποιεί τον πληθυσμό σε n στρώματα τα οποία αποτελούνται από τα πρώτα k στοιχεία, τα δεύτερα k στοιχεία κ.τ.λ. και από κάθε στρώμα επιλέγεται ένα δείγμα μεγέθους I με τη διαφορά ότι δεν επιλέγεται τυχαία (εκτός από το πρώτο).

3,4,2 Μονοσταδιακή και Πολυσταδιακή Δειγματοληψία

Δειγματοληψία κατά ομάδες (σμάρια)

Στην δειγματοληψία κατά ομάδες τα μέλη που απαρτίζουν το δείγμα που θα εξετάσουμε είναι ομάδες από στοιχεία που μας ενδιαφέρουν. Οι ομάδες αυτές είναι συνήθως ομαδοποιήσεις του πληθυσμού όπως γεωγραφικές περιοχές, σωματεία, σχολεία, τάξεις, κ.τ.λ. Λαμβάνουμε ένα δείγμα από αυτές τις ομάδες. Αν όλα τα στοιχεία των ομάδων στο δείγμα συμπεριλαμβάνονται στο τελικό δείγμα, τότε έχουμε την μονο-σταδιακή δειγματοληψία κατά ομάδες. Η διαφορά από την μέθοδο που εξετάσαμε προηγουμένως είναι ότι στην στρωματική

δειγματοληψία λαμβάνουμε ένα μέρος του τελικού δείγματος από κάθε ομάδα (στρώμα), ενώ σε αυτήν την μέθοδο το τελικό δείγμα λαμβάνεται κατά ομάδες (σμάρια), ενώνοντας όλα τα στοιχεία των ομάδων που επιλέχθηκαν στο δείγμα στο πρώτο στάδιο. Στην περίπτωση που δείγματα στοιχείων λαμβάνονται από τις ομάδες στο δείγμα στο πρώτο στάδιο, τότε έχουμε δειγματοληψία κατά ομάδες σε δύο στάδια (δι-σταδιακή). Οι μονάδες στο πρώτο στάδιο (οι πρώτες ομάδες) λέγονται πρωτεύουσες δειγματικές μονάδες.

Παράδειγμα: Ομαδοποιούμε τα διαμερίσματα ενός κράτους σε τρεις ομάδες, σύμφωνα με το βαθμό εκβιομηχάνισης (χαμηλό, μεσαίο, υψηλό). Στο πρώτο στάδιο επιλέγουμε τυχαία ένα διαμέρισμα από κάθε ομάδα. Στο δεύτερο στάδιο παίρνουμε ένα δείγμα αυτοδιοικήσεων από το επιλεγμένο δείγμα διαμερισμάτων. Στο τρίτο στάδιο θα μπορούμε να πάρουμε ένα δείγμα νοικοκυριών από το δείγμα των επιλεγμένων αυτοδιοικήσεων κ.τ.λ.

Διαπιστώνουμε λοιπόν ότι η πολύ-σταδιακή δειγματοληψία είναι μια αρκετά σύνθετη μέθοδος, η οποία συνδυάζει προηγούμενες μεθόδους. Ένα πλεονέκτημα αυτής της μεθόδου είναι ότι είναι οικονομικότερη σε χρήμα και χρόνο. Η πολύ-σταδιακή δειγματοληψία είναι πολύ χρήσιμη όταν έχουμε ιεραρχική διάταξη των στοιχείων του πληθυσμού. Π.χ. Έλληνες ενήλικοι ομαδοποιούνται σε τετράγωνα πόλης, μετά σε νοικοκυριά εντός των τετραγώνων και μετά σε ενήλικες μέσα στα νοικοκυριά. Στα πρώτα στάδιο θα μπορούσαμε να επιλέξουμε τυχαία ένα δείγμα τετραγώνων, στο δεύτερο στάδιο ένα δείγμα νοικοκυριών από τα ήδη επιλεγμένα τετράγωνα και στο τρίτο στάδιο ένα δείγμα ενηλίκων από το επιλεγμένο δείγμα νοικοκυριών. Έτσι θα είχαμε μια τρι-σταδιακή μέθοδο, με τρία επίπεδα όπου τα τετράγωνα είναι πρωτεύουσες δειγματοληπτικές μονάδες, τα νοικοκυριά δευτερεύουσες και οι ενήλικες

τριτεύουσες. Κάθε ένα από αυτά τα επίπεδα συμβάλλει μια συνιστώσα στην ολική διασπορά.

3,5 Δειγματοληψία με πιθανότητες ανάλογες του μεγέθους (ΠΑΜ)

(Δειγματοληψία με άνισες πιθανότητες)

Οι ομάδες στην πολυσταδιακή δειγματοληψία είναι συνήθως φυσικές υποδιαιρέσεις του πληθυσμού, όπως νοσοκομεία, σχολεία, τάξεις κ.τ.λ. Επειδή λοιπόν είναι οι ομάδες αυτές πολλές φορές διαφορετικές ως προς το μέγεθος, τόσο τα δειγματικά μεγέθη όσο και οι πιθανότητες επιλογής πρέπει να είναι ανάλογες με το μέγεθος των ομάδων, τουλάχιστον στο πρώτο στάδιο. Υπάρχουν διάφορες μέθοδοι για δειγματοληψία με ΠΑΜ, όπως με επανάθεση, συστηματική με ΠΑΜ κ.τ.λ. Στις πιο πολλές από αυτές η ιδέα είναι να φανταστούμε ότι κάθε ομάδα εμφανίζεται στην λίστα με έναν αριθμό φορών ανάλογο του μεγέθους της. Αυτό γίνεται συνήθως με την προοδευτική συσσώρευση των μεγεθών των μονάδων και τον προσδιορισμό διαστημάτων επιλογής για κάθε ομάδα. Π.χ. αν η πιθανότητα (ΠΑΜ) επιλογής σε ένα τράβηγμα είναι $p_i = M_i/M_0$ ($M_0 = \sum M_i$), τότε στη δειγματοληψία με επανάθεση, η πιθανότητα p_i , η μονάδα a_i να είναι στο δείγμα δεν είναι ακριβώς nM_i/M_0 , αλλά στη συστηματική δειγματοληψία είναι $p_i = np_i$.

Συστηματική δειγματοληψία με ΠΑΜ

Μια μέθοδος που κρατάει τα $p_i = np_i$ και όπου, υπό την προϋπόθεση ότι $nM_i < M_0$ για κάθε i , καμιά μονάδα δεν επιλέγεται παραπάνω από μια φορά είναι η εξής: τραβάμε ένα τυχαίο αριθμό και τότε το δείγμα είναι $k, k+M_0, \dots, k+(n-1)M_0$ υπό την προϋπόθεση ότι κάθε μονάδα a_i έχει πιθανότητα να είναι στο δείγμα $p_i = np_i$.

Θα μπορούσαμε να συσσωρεύσουμε τα M_i και να πάρουμε σαν δειγματοληπτικό διάστημα $k=Mo/n$ οπότε μια τυχαία αρχή ρ μεταξύ 1 και k και τότε το δείγμα θα είναι $\rho, \rho+k, \rho+2k, \dots, \rho+(n-1)k$. Άρα η πιθανότητα n μονάδα i να είναι στο δείγμα είναι $M_i/k=nM_i/Mo=n\rho_i$ ($\rho_i=M_i/Mo$).

Η συστηματική δειγματοληψία με ΠΑΜ μπορεί να χρησιμοποιηθεί και για να δώσει έναν ΙΠΕΜ σχεδιασμό δείγματος π.χ. σ' ένα δι-σταδιακό σχεδιασμό οι πρωτεύουσες μονάδες επιλέγονται με ΠΑΜ και μετά m στοιχεία επιλέγονται με απλή τυχαία δειγματοληψία για κάθε ομάδα. Η πιθανότητα επιλογής για ένα στοιχείο στην i ομάδα είναι $(nM_i/Mo)(m/M_i)=nm/Mo$. Έτσι με τη διαδικασία αυτή το ολικό μέγεθος δείγματος είναι nm με το ίδιο μέγεθος m για κάθε επιλεγμένη πρωτεύουσα ομάδα και με ίση ολική πιθανότητα επιλογής για όλα τα στοιχεία.

3,6 Ποσοστιαία δειγματοληψία (κατά ποσοστά)

Αντίθετα με τις προηγούμενες μεθόδους η ποσοστιαία δειγματοληψία δεν είναι πιθανοτική μέθοδος. Αυτοί που παίρνουν τις συνεντεύξεις των ατόμων είναι ελεύθεροι να επιλέξουν τους συνεντευξιαζόμενους, αρκεί το δείγμα να συμπληρώνει προκαθορισμένους αριθμούς (quotee) φύλου, εισοδήματος, ηλικίας κ.τ.λ. Αυτές οι quotee προκαθορίζονται ώστε να δίνουν αναλογική εκπροσώπηση κάθε στρώματος. Άρα η ποσοστιαία δειγματοληψία είναι μια μορφή στρωματικής δειγματοληψίας, όπου η εκλογή του δείγματος μονάδων από τα στρώματα είναι μη - πιθανοτική. Οι πιο συχνές στρωματικές μεταβλητές που χρησιμοποιούνται είναι γεωγραφική περιοχή, φύλο, ηλικία, κοινωνικό status.

3,7 Σφάλματα, μεροληψία, συνέπειες στα διαστήματα εμπιστοσύνης

Η λέξη «σφάλμα» στην Στατιστική δεν σημαίνει αναγκαστικά λάθος ή γκάφα π.χ. Το τυπικό σφάλμα ενός εκτιμητή θ είναι ένα άλλο όνομα για την τυπική απόκλιση, δηλαδή την τετραγωνική ρίζα της διασποράς του εκτιμητή, $\text{Var}(\hat{\theta})$. Αυτά τα σφάλματα που οφείλονται στη φυσική μεταβλητικότητα των τυχαίων δειγμάτων και στη χρήση ενός μόνο δείγματος λέγονται δειγματοληπτικά σφάλματα. Υπάρχουν και τα μη-δειγματοληπτικά σφάλματα που συμβάλλουν στο μέσο τετραγωνικό σφάλμα του εκτιμητή θ , δηλαδή της ποσότητας $\text{MTE}(\hat{\theta}) = E(\hat{\theta} - \theta)^2$.

$$\text{Αποδεικνύεται ότι } E(\hat{\theta} - \theta)^2 = E[\hat{\theta} - E(\hat{\theta})]^2 + E[E(\hat{\theta}) - \theta]^2.$$

Το πρώτο μέλος είναι εξ ορισμού η $\text{Var}(\hat{\theta})$ και η διαφορά $E(\hat{\theta} - \theta)$ λέγεται μεροληψία του εκτιμητή θ . Επομένως $\text{MTE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + (\text{Μεροληψία } \hat{\theta})^2$.

Τα μη δειγματοληπτικά σφάλματα οφείλονται σε πολλούς παράγοντες, όπως π.χ. η συστηματική ή σταθερή μεροληψία. Αυτή είναι δύσκολο να ανιχνευτεί αφού οι διαφορές των μετρήσεων και η διασπορά του μέσου δεν αλλάζει και μόνο ο δειγματικός μέσος μετατίθεται στην κλίμακα, π.χ. ένα διάστημα της μορφής $\bar{y} - 1,96(s\sqrt{n})$, $\bar{y} + 1,96(s\sqrt{n})$ (1), που είναι ονομαστικώς διάστημα εμπιστοσύνης 95% μπορεί στην ουσία να είναι 50%, δηλαδή άχρηστο αφού στην πραγματικότητα είναι της μορφής $\bar{y} - \text{Μεροληψία}(\bar{y}) \pm 1,96(s\sqrt{n})$.

Εφόσον $E(\bar{y}) = \bar{y} + B$, όπου $B = \text{Μεροληψία} \neq 0$, μπορεί τα διαστήματα της μορφής (1) να περικλείουν τον μέσο του πληθυσμού \bar{y} λιγότερες φορές, αφού έχουν μετατεθεί προς τα δεξιά.

Άλλος ένας παράγοντας μη-δειγματοληπτικού σφάλματος είναι η μεροληψία εισαγόμενη από την απουσία δεδομένων στο δείγμα, που οφείλεται στην μη-ανταπόκριση επιλεγμένων ατόμων να απαντήσουν στα ερωτηματολόγια. Ας υποθέσουμε ότι $N = N_1 + N_2$ όπου N_1 ο αριθμός των

ατόμων που απάντησαν σε μια επισκόπηση και N_2 ο αριθμός αυτών που δεν απάντησαν. Επιλέγεται ένα τυχαίο δείγμα n και οι απαντήσεις παίρνονται από n_1 μέλη του δείγματος, ενώ $n-n_1$ δεν απαντούν. Αν οι πληθυσμιακοί μέσοι των N_1 και N_2 είναι \bar{Y}_1 και \bar{Y}_2 , ο ολικός μέσος θα είναι $\bar{Y} = 1/N(N_1\bar{Y}_1 + N_2\bar{Y}_2)$. Ο δειγματικός μέσος \bar{Y}_1 βασισμένος στους n_1 που απάντησαν είναι αμερόληπτος εκτιμητής του \bar{Y}_1 , δηλαδή $E(n_1) = \bar{Y}_1$. Αν ο \bar{Y}_1 χρησιμοποιηθεί σαν εκτιμητής του y , τότε η μεροληψία του \bar{Y}_1 είναι $E(\bar{Y}_1) - \bar{y} = \bar{Y}_1 - 1/N(N_1\bar{Y}_1 + N_2\bar{Y}_2) = N_2/N(\bar{Y}_1 - \bar{Y}_2)$.

3,8 Γενικές παρατηρήσεις επί των μεθόδων

Η ομαδοποίηση του πληθυσμού είτε οι ομάδες είναι στρώματα είτε σμάρια οδηγεί συχνά σε μεγαλύτερη ακρίβεια και ελάττωση της διασποράς των εκτιμητών. Η δειγματοληψία κατά στρώματα και η δειγματοληψία κατά σμάρια διαφέρουν κυρίως στον τρόπο που το τελικό δείγμα λαμβάνεται. Στην στρωματική κάθε στρώμα δειγματοληπτείται ενώ στην κατά σμάρια το δείγμα έρχεται σε σμάρια, ενώνοντας όλα τα επιλεγμένα σμάρια που λάβαμε στο προηγούμενο στάδιο. Τα στρώματα τείνουν να είναι ομοιογενή, ενώ στα σμάρια δεν απαιτείται να έχουν ομοιογένεια. Επίσης, τα στρώματα είναι εξαντλητικά του πληθυσμού και αλληλοαποκλειόμενα, ενώ τα σμάρια μπορεί να είναι υποδιαίρεσεις στρωμάτων.

Στην συστηματική δειγματοληψία, η τυχαία εκλογή της πρώτης ομάδας $a_i, 1 \leq i \leq k$, αυτομάτως ορίζει το δείγμα ως $a_i, a_i+k, a_i+2k, \dots, a_i+(n-1)k+1$, όπου $k=N/n$ το δειγματοληπτικό κλάσμα. Αυτή η μέθοδος είναι πιο ακριβής από την απλή τυχαία δειγματοληψία, αφού στρωματοποιεί τον πληθυσμό σε n στρώματα, αποτελούμενα από τις πρώτες k μονάδες, τις δεύτερες k μονάδες κ.τ.λ. Ένας άλλος τρόπος θεώρησης της

συστηματικής δειγματοληψίας (αν $N=nk$) είναι να την θεωρήσουμε σαν δειγματοληψία κατά σμάρια όπου τα k σμάρια είναι:

Σμάρι	Σύνθεση σμαριού
1	1, $k+1$, $2k+1$, ..., $(n-1)k+1$
i	i , $k+i$, $2k+i$, ..., $(n-1)k+i$
K	k , $2k$, $3k$, ..., nk

Κάθε σμάρι περιέχει n μονάδες και η συστηματική δειγματοληψία είναι ισοδύναμη με την τυχαία εκλογή ενός από το k σμάριο και θεωρώντας όλες τις μονάδες σ' αυτό το σμάρι σαν το δείγμα.

ΚΕΦΑΛΑΙΟ 4

Συλλογή στατιστικών στοιχείων

4,1 Ορισμοί

Κάθε σύνολο εμπύχων ή μη μπορεί να θεωρηθεί ότι αποτελεί έναν **πληθυσμό**, π.χ. το σύνολο των κατοίκων μιας χώρας, το σύνολο των φοιτητών των Πανεπιστημίων, το σύνολο των εργοστασίων, των ξενοδοχείων κ.τ.λ. Κάθε στοιχείο ενός πληθυσμού ονομάζεται μονάδα ή άτομο.

Ως **στατιστικό πληθυσμό** θα εννοούμε το σύνολο των μετρήσεων ή γενικότερα των παρατηρήσεων, οι οποίες αναφέρονται σ' ένα χαρακτηριστικό ή μια ιδιότητα των μονάδων του πληθυσμού που εξετάζουμε.

Στην στατιστική ως **τυχαία μεταβλητή** ορίζουμε κάθε χαρακτηριστικό ή ιδιότητα των στατιστικών μονάδων.

Το εισόδημα, η ηλικία, το επάγγελμα, η ιπποδύναμη, ο κυβισμός, το χρώμα (ενός μη επιβατικού αυτοκινήτου) είναι μερικές από τις ιδιότητες των μονάδων ενός πληθυσμού.

Οι ιδιότητες αυτές ονομάζονται **μεταβλητές**, γιατί κάθε μια από αυτές παίρνει τουλάχιστον δύο διαφορετικές τιμές.

Αν επιλέξουμε μια μονάδα από κάποιον πληθυσμό, τον οποίο μελετάμε ως προς τη μεταβλητή π.χ. οικογενειακή κατάσταση, τότε η μονάδα αυτή μπορεί να είναι είτε χήρος, είτε διαζευγμένος, είτε άγαμος. Επειδή η επιλογή της μονάδας γίνεται τυχαία, η τιμή που αντιστοιχεί στη συγκεκριμένη μονάδα μπορεί να είναι οποιαδήποτε από τις τιμές της μεταβλητής «οικογενειακή κατάσταση», γεγονός που δικαιολογεί και τον όρο **τυχαία μεταβλητή**.

Επιπλέον ορίζουμε ως **στατιστικό πληθυσμό** το σύνολο των δυνατών μετρήσεων ή παρατηρήσεων που αναφέρονται σε κάποια μεταβλητή του πληθυσμού, δηλαδή το σύνολο των τιμών που μπορεί να πάρει μια μεταβλητή X.

4,2 Είδη μεταβλητών

Οι μεταβλητές διακρίνονται σε **ποσοτικές**, **κατηγορικές** και **ποιοτικές**.

Ποσοτικές λέγονται οι μεταβλητές που παίρνουν πραγματικές τιμές, όπως το ανάστημα, ο αριθμός μελών μιας οικογένειας κ.τ.λ. Το «ανάστημα» λέγεται **συνεχής** μεταβλητή γιατί μπορεί να πάρει οποιαδήποτε τιμή του συνόλου των πραγματικών αριθμών, ενώ ο «αριθμός μελών μιας οικογένειας» λέγεται **ασυνεχής** ή **αναριθμητής**, αφού δέχεται μόνο φυσικές τιμές.

Οι μη ποσοτικές μεταβλητές διακρίνονται σε **κατηγορικές**, όπως το φύλο, το θρήσκευμα κ.τ.λ. και **ποιοτικές**, όπως η κατάσταση υγείας, η διαγωγή ενός μαθητή κ.τ.λ. Η κύρια διαφορά τους είναι ότι στις ποιοτικές μεταβλητές οι τιμές του επιδέχονται κάποια διάταξη ή ιεράρχηση π.χ. την κατάσταση υγείας ενός ανθρώπου μπορούμε να την χαρακτηρίσουμε από κακή έως άριστη, κάτι που δεν συμβαίνει με τις τιμές των κατηγορικών μεταβλητών π.χ. θρήσκευμα.

4,3 Μέθοδοι συλλογής στατιστικών στοιχείων

Το πρώτο στάδιο για τη στατιστική μελέτη ενός προβλήματος είναι η συλλογή των αριθμητικών στοιχείων. Η συλλογή του στατιστικού υλικού αποτελεί το θεμέλιο πάνω στο οποίο στηρίζεται κάθε στατιστική έρευνα. Αν τα πρωτογενή δεδομένα είναι ακριβή, τότε θα προκύψουν ορθά συμπεράσματα. Αν τα δεδομένα είναι λανθασμένα, τότε το ίδιο θα συμβεί και στα συμπεράσματα.

Η συλλογή των στατιστικών στοιχείων γίνεται κυρίως με δύο μεθόδους, την απογραφή και τη δειγματοληψία.

Όταν οι παρατηρήσεις προκύπτουν από όλες τις μονάδες του πληθυσμού, τότε μιλάμε για απογραφή.

Όταν οι παρατηρήσεις προκύπτουν από ένα υποσύνολο του αρχικού πληθυσμού, με σκοπό τη γενίκευση των συμπερασμάτων για όλο τον πληθυσμό, τότε μιλάμε για δειγματοληψία και το εξεταζόμενο υποσύνολο ονομάζεται δείγμα.

Σε αρκετές περιπτώσεις η απογραφή είναι πρακτικά αδύνατη ή περιττή π.χ. για να εξεταστεί η καταλληλότητα ενός τροφίμου, αρκεί να εξεταστεί μια μικρή ποσότητα από αυτό. Σε άλλες περιπτώσεις η απογραφή κοστίζει σημαντικά σε χρόνο και χρήμα, γεγονός που αποθαρρύνει τη μέθοδο της απογραφής π.χ. η ακροαματικότητα ενός σταθμού, η δημοτικότητα ενός πολιτικού κ.τ.λ.

Η δειγματοληψία αποτελεί ξεχωριστό κλάδο της Στατιστικής με μεγάλη ανάπτυξη κατά τα τελευταία χρόνια και εκτός από τη σημαντική οικονομία σε χρόνο και χρήμα, προσφέρει εκτιμήσεις πολύ κοντά στην πραγματικότητα, αρκεί να γίνεται με σύγχρονες μεθόδους και με κατάλληλα εκπαιδευμένο προσωπικό.

Υπάρχουν βέβαια και περιπτώσεις όπου η απογραφή αποτελεί τον μοναδικό τρόπο συλλογής στοιχείων, π.χ. οι βουλευτικές, δημοτικές εκλογές, τα δημοψηφίσματα κ.τ.λ.

4,4 Δειγματοληπτικά σφάλματα

Ένα καλό δείγμα μπορεί να αναπαράγει τα χαρακτηριστικά ενός πληθυσμού με μεγάλη ακρίβεια. Το δείγμα αυτό χαρακτηρίζεται ως τυχαίο και είναι εκείνο όπου όλες οι μονάδες έχουν την ίδια πιθανότητα επιλογής. Π.χ. για να εκτιμηθεί το ποσοστό των ψηφοφόρων κάποιου

κόμματος, πρέπει μέσα στο δείγμα να αντιπροσωπεύονται αναλογικά όλες οι περιοχές της χώρας, αστικές, ημιαστικές και αγροτικές.

Όλες οι μονάδες του δείγματος θα πρέπει να επιλεγούν κατά τέτοιο τρόπο ώστε οι μονάδες του πληθυσμού να έχουν ακριβώς την ίδια πιθανότητα επιλογής. Βέβαια, αυτή η θεμελιώδης απαίτηση δεν είναι πάντα εύκολο να ικανοποιηθεί. Όταν ο πληθυσμός είναι σαφώς καθορισμένος π.χ. οι συνδρομητές του ΟΤΕ, οι υπάλληλοι στο δημόσιο, κ.τ.λ. για τους οποίους υπάρχουν κρατικά στοιχεία, η επιλογή ενός τυχαίου δείγματος είναι εφικτή.

Αντίθετα, οι δημοσκοπήσεις που αναφέρονται σε ποσοστά κομμάτων, ακροαματικότητα προγραμμάτων κ.τ.λ. παρουσιάζουν σοβαρές δυσκολίες στην επιλογή ενός αντιπροσωπευτικού δείγματος, π.χ. το μέσο ανάστημα 500 αρρένων σπουδαστών, που προέκυψε με απογραφή είναι 1.72 cm. Αν με κάποια τυχαία διαδικασία (π.χ. κλήρωση) επιλέξουμε 50 από αυτούς και το μέσο ανάστημα του δείγματος αυτού προκύψει 1.70cm, τότε η διαφορά των δύο εκατοστών είναι ένα δειγματοληπτικό σφάλμα.

Ακόμα υπάρχουν και τα σφάλματα μέτρησης που μπορούν να προκύψουν π.χ. από ένα κακό ερωτηματολόγιο ή από απροθυμία ειλικρινούς συνεργασίας των μονάδων του δείγματος, όταν η δειγματοληψία γίνεται από ανθρώπινους πληθυσμούς.

Ένα άλλο δειγματοληπτικό σφάλμα είναι η αντικατάσταση μονάδων του πληθυσμού από άλλες, όταν μερικές από τις αρχικά επιλεγμένες αρνούνται να συνεργαστούν π.χ. δειγματοληψία με αλληλογραφία.

4,5 Το μέγεθος του δείγματος

Το πλήθος των μονάδων του δείγματος λέγεται μέγεθος. Η αξία μιας εκτίμησης επηρεάζεται πολύ από το μέγεθος του δείγματος. Όσο

μεγαλύτερο είναι το μέγεθος του δείγματος, τόσο η εκτίμηση θα διαφέρει της πραγματικής τιμής. Το σφάλμα μηδενίζεται όταν το μέγεθος του δείγματος φτάσει το μέγεθος του πληθυσμού, δηλαδή όταν η δειγματοληψία μετατραπεί σε απογραφή. Το ζητούμενο όμως δεν είναι να μηδενιστεί το σφάλμα, αλλά να περιοριστεί μέσα σε επιθυμητά όρια, ώστε η μεγαλύτερη δυνατή απόκλιση να μην υπερβαίνει κάποια κρίσιμη τιμή.

Το μέγεθος του δείγματος είναι ανάλογο της διακύμανσης του πληθυσμού. Όσο μεγαλύτερη είναι η διακύμανση του πληθυσμού, τόσο μεγαλύτερο δείγμα χρειαζόμαστε. Όσο περισσότερο απέχουν οι επιμέρους τιμές μιας μεταβλητής από τη μέση τιμή τους, τόσο αυτή είναι μεγαλύτερη και μηδενίζεται μόνο όταν όλες οι τιμές ταυτίζονται. Π.χ. για να γίνει ανάλυση αίματος σ' ένα ασθενή, είναι αρκετό ένα μικρό φιαλίδιο αίματος, αφού όσα φιαλίδια και αν ληφθούν από τον ίδιο ασθενή θα προκύψει το ίδιο αποτέλεσμα, εφόσον η διακύμανση είναι μηδενική από δείγμα σε δείγμα για όλες τις μεταβλητές του αίματος.

Σημαντικό ρόλο στον προσδιορισμό του μεγέθους του δείγματος παίζει και το κόστος της έρευνας. Έτσι, αν η κάθε μονάδα του δείγματος κοστίζει 1.000 δρχ. για αμοιβή των ερευνητών και για τη συνολική έρευνα διατίθεται 1.000.000 δρχ., τότε ο μέγιστος αριθμός των μονάδων του δείγματος δεν μπορεί να είναι μεγαλύτερος από χίλιες μονάδες. Εάν οι χίλιες μονάδες είναι αρκετές για τον περιορισμό της εκτίμησης, τότε δεν υπάρχει πρόβλημα. Σε αντίθετη περίπτωση πρέπει να αυξηθεί το κόστος της έρευνας.

4,6 Το ερωτηματολόγιο

Ένα βασικό στοιχείο για την επιτυχία μιας έρευνας, όταν η δειγματοληψία γίνεται από ανθρώπινους πληθυσμούς, είναι η σωστή κατασκευή του ερωτηματολογίου. Πολύ συχνά μέσα σ' ένα

ερωτηματολόγιο περιέχονται πολλές δευτερεύουσες ερωτήσεις, με το σκεπτικό ότι δεν επιβαρύνουν ιδιαίτερα το κόστος της έρευνας. Αυτό έχει ως αποτέλεσμα ο ερωτώμενος μετά από τρεις – τέσσερις ερωτήσεις να αρχίσει να δυσανασχετεί. Διαπιστώνουμε λοιπόν ότι ένα μεγάλο ερωτηματολόγιο έχει σαν συνέπεια τη συλλογή κακής ποιότητας στοιχείων και συνήθως την αποτυχία της έρευνας.

Σε περιπτώσεις όμως ερευνών ευρείας έκτασης όπως π.χ. η απογραφή του ελληνικού πληθυσμού, ένα περιορισμένο ερωτηματολόγιο δεν αρκεί για να συλλεγούν πλήρη στοιχεία. Γι' αυτό τα ερωτηματολόγια αυτά συμπληρώνονται από ειδικευμένους υπαλλήλους.

Μεγάλη προσοχή πρέπει να δίνεται στη φρασεολογία των ερωτήσεων, ώστε να είναι σαφείς και κατανοητές. Επίσης πρέπει να προηγούνται οι ερωτήσεις που θα κερδίσουν το ενδιαφέρον ή θα τονώσουν την εμπιστοσύνη του ερωτώμενου. Για το ίδιο ερώτημα δεν θα πρέπει να υπάρχει δυνατότητα δύο ή περισσότερων απαντήσεων, γιατί έτσι η ανάλυση των δεδομένων θα είναι πολύπλοκη. Ένα σωστά σχεδιασμένο ερωτηματολόγιο δίνει σωστές απαντήσεις, τα δεδομένα αναλύονται ευκολότερα και τα σφάλματα ελαχιστοποιούνται.

4,7 Δειγματοληπτικές μέθοδοι

Οι κυριότερες μέθοδοι επιλογής ενός δείγματος, είναι:

α) Απλή τυχαία δειγματοληψία

Με τη μέθοδο αυτή επιλέγεται τυχαία μια μονάδα του πληθυσμού και στη συνέχεια πάλι τυχαία μια μονάδα απ' τις υπόλοιπες κ.τ.λ. μέχρι να επιλεγεί ένα επιθυμητό πλήθος μονάδων.

β) Συστηματική δειγματοληψία

Με τη μέθοδο αυτή χωρίζεται ο πληθυσμός σε τμήματα και από κάθε τμήμα επιλέγεται ένας αριθμός μέχρι να εξαντληθεί ο πληθυσμός.

γ) Στρωματική δειγματοληψία

Με τη μέθοδο αυτή ο πληθυσμός διαιρείται σε υποπληθυσμούς που ονομάζονται στρώματα. Στη συνέχεια από κάθε στρώμα επιλέγεται ένα απλό τυχαίο δείγμα που είναι συνήθως ανάλογο του μεγέθους του στρώματος. Η μέθοδος αυτή είναι πολύ διαδεδομένη και γίνεται για πολλούς λόγους. Οι σπουδαιότεροι είναι:

i) Για λόγους διοικητικούς π.χ. αν κάθε στρώμα είναι ένας νομός της χώρας, οι τοπικές αρχές μπορούν να εποπτεύουν την έρευνα.

ii) Για λόγους τοπικού ενδιαφέροντος αφού τα αποτελέσματα της έρευνας θα μπορούν να αναφέρονται χωριστά για κάθε νομό.

iii) Για λόγους στατιστικούς διότι οι μονάδες εντός των στρωμάτων είναι πιο ομοιογενείς γεγονός που συνεπάγεται οικονομία.

δ) Δισταδιακή δειγματοληψία

Η δισταδιακή δειγματοληψία είναι μια πολύ διαδεδομένη μέθοδος. Μ' αυτή τη μέθοδο ένας πληθυσμός διαιρείται σε υποπληθυσμούς. Αρχικά επιλέγεται τυχαία ένας αριθμός υποπληθυσμών και στη συνέχεια από κάθε επιλεγμένο υποπληθυσμό επιλέγεται τυχαία ένας ορισμένος εκ των προτέρων αριθμός μονάδων.

ε) Διφασιακή δειγματοληψία

Ας υποθέσουμε ότι μελετούμε τα χαρακτηριστικά των ψηφοφόρων ενός κόμματος. Στην περίπτωση αυτή η δειγματοληψία γίνεται σε δύο φάσεις. Κατά την πρώτη φάση επιλέγεται μια μονάδα τυχαία και ερωτάται αν ψήφισε το κόμμα αυτό. Αν απαντήσει ΟΧΙ τότε σταματούν οι ερωτήσεις, ενώ αν απαντήσει ΝΑΙ τότε συνεχίζονται οι ερωτήσεις για τα χαρακτηριστικά που ερευνούμε.

Συνεχείς εγγραφές στατιστικών στοιχείων

Σε πολλές περιπτώσεις στοιχεία συλλέγονται σε μόνιμη βάση από δημόσιες κυρίως υπηρεσίες. Τέτοια στοιχεία είναι π.χ. γεννήσεις, θάνατοι που συλλέγονται από τα Ληξιαρχεία της χώρας, στοιχεία που συλλέγονται από τα τελωνεία και αφορούν το εξωτερικό εμπόριο, τα αεροδρόμια, τα λιμάνια, οικονομικά στοιχεία που συλλέγει το υπουργείο οικονομικών, η Τράπεζα της Ελλάδος, στατιστικά στοιχεία που συλλέγουν όλα τα Υπουργεία, οι δημόσιες επιχειρήσεις κ.τ.λ. Η σπουδαιότερη κρατική υπηρεσία συλλογής στατιστικών στοιχείων είναι η Εθνική Στατιστική Υπηρεσία (Ε.ΣΥ.Ε.).

4,8 Η δημοσκόπηση

Τα τελευταία χρόνια έχει αναπτυχθεί πολύ στη χώρα μας η επιστημονική δημοσκόπηση, κυρίως μετά την ίδρυση εξειδικευμένων εταιριών. Οι εταιρίες αυτές εργάζονται συνήθως για λογαριασμό των μέσων μαζικής ενημέρωσης, για διάφορα πολιτικά κόμματα κ.ά. Αντικείμενο των εταιριών αυτών είναι η σφυγμομέτρηση της κοινής γνώμης για διάφορα επίκαιρα θέματα γενικού ενδιαφέροντος, όπως η δημοτικότητα ενός πολιτικού, η τηλεθέαση μιας εκπομπής κ.τ.λ.

Τα αποτελέσματα της δημοσκόπησης υπόκεινται σε δειγματοληπτικά σφάλματα. Εκτός όμως από τα σφάλματα, μερικές φορές γίνεται σκόπιμη αλλοίωση των αποτελεσμάτων της δειγματοληψίας, κυρίως όταν αυτά ανακοινώνονται από τα ΜΜΕ. Οι εταιρίες βέβαια φροντίζουν να τηρούν τα προσχήματα της αληθοφάνειας έτσι ώστε τα αποτελέσματα που δίνονται στη δημοσιότητα να επιδέχονται περισσότερες από μια ερμηνείες ειδικά όταν πρόκειται για ποσοστά κομμάτων. Αυτό το πετυχαίνουν κυρίως αυξάνοντας υπερβολικά τον αριθμό των αναποφάσιστων ψηφοφόρων.

Ο κυριότερος λόγος που σε μια καλοπροαίρετη δημοσκόπηση το ποσοστό των αναποφάσιστων είναι μεγάλο, είναι το γεγονός ότι πολλοί

ερωτώμενοι δηλώνουν αναποφάσιστοι όταν θέλουν να αποφύγουν την απάντηση για ευνόητους λόγους. Επίσης στις εκλογές ένας αριθμός ψηφοφόρων ψηφίζει λευκό ή άκυρο και αυτοί δεν θεωρούνται αναποφάσιστοι.

Ένας τρόπος για να μειωθεί ο φαινομενικός αριθμός των αναποφάσιστων είναι η χρήση της κάλπης με ψηφοδέλτια κατά τη δειγματοληψία, ώστε οι ερωτώμενοι να εκφράζονται χωρίς να υπάρχει κίνδυνος να αποκαλυφθεί τι ψηφίζουν.

Παρά την αναξιοπιστία ορισμένων δημοσκοπήσεων, η στατιστική επιστήμη με τη βοήθεια των ηλεκτρονικών υπολογιστών έχει όλα τα μέσα ώστε να γίνονται ασφαλείς προβλέψεις, αρκεί να χρησιμοποιούνται οι κατάλληλες μέθοδοι για κάθε περίπτωση και να ικανοποιούνται με τον καλύτερο τρόπο οι βασικές αρχές της δειγματοληψίας.

4.9 Βασικές προϋποθέσεις για την δημοσιοποίηση μιας δημοσκόπησης

Τα μέσα μαζικής ενημέρωσης δεν θα πρέπει να δημοσιοποιούν δημοσκοπήσεις αν αυτές δεν πληρούν ορισμένες προϋποθέσεις. Οι βασικότερες είναι:

1. Να αναφέρεται η εταιρία που έκανε τη δημοσκόπηση
2. Να αναφέρεται η ακριβής ημερομηνία της δειγματοληψίας
3. Να αναφέρεται ο αριθμός των μονάδων του δείγματος
4. Να αναφέρεται η δειγματοληπτική μέθοδος
5. Να αναφέρεται η γεωγραφική περιοχή που καλύπτει το δείγμα.
6. Να αναφέρονται τα ακριβή ερωτήματα της δημοσκόπησης.

7. Να αναφέρεται κάθε άλλη πληροφορία σχετικά με το σχεδιασμό, τη διεξαγωγή ή την ανάλυση των στοιχείων της έρευνας που θα μπορούσε να ενδιαφέρει το κοινό, π.χ. αν έγινε χρήση κάλπης, αν έγινε τηλεφωνικά, αν υπήρξαν μονάδες που αρνήθηκαν να απαντήσουν κ.τ.λ.

4,10 Παρουσίαση στατιστικών στοιχείων

Εισαγωγή

Η παρουσίαση του στατιστικού υλικού γίνεται με δύο τρόπους. Ο πρώτος τρόπος συνίσταται στην κατασκευή ενός στατιστικού πίνακα, που ονομάζεται πίνακας συχνοτήτων ή κατανομή συχνοτήτων. Ο δεύτερος τρόπος συνίσταται στην κατασκευή ενός κατάλληλου για την περίπτωση διαγράμματος. Και οι δύο τρόποι παρέχουν σε κάθε ενδιαφερόμενο με τρόπο συνοπτικό μια εικόνα για τη μορφολογία του πληθυσμού και αποτελούν μια πρώτη μορφή στατιστικής ανάλυσης.

Μονομεταβλητοί πίνακες συχνότητας

Σε κάθε πίνακα συχνοτήτων μιας κατηγορικής ή ποιοτικής μεταβλητής, διακρίνουμε δύο κυρίως στήλες. Η πρώτη σχηματίζεται από τις τιμές της μεταβλητής, ενώ η δεύτερη από τις συχνότητες των μεταβλητών.

Απόλυτη συχνότητα μιας τιμής ονομάζεται το πλήθος των μονάδων του πληθυσμού που παίρνουν την τιμή.

Σχετική συχνότητα μιας τιμής ονομάζουμε το ποσοστό των μονάδων του πληθυσμού που παίρνουν την τιμή αυτή.

Στην περίπτωση που η μεταβλητή είναι συνεχής, η πρώτη στήλη συνίσταται από κατάλληλα επιλεγμένα διαστήματα που λέγονται ομάδες ή τάξεις. Τα διαστήματα αυτά είναι κλειστά κατά το ένα άκρο το οποίο

μπορεί να είναι είτε το δεξί είτε το αριστερό. Οι συχνότητες αυτών των διαστημάτων ονομάζονται ταξικές συχνότητες.

Αθροιστικές συχνότητες

Εστω μια ποσοτική μεταβλητή. Απόλυτη αθροιστική συχνότητα μιας τιμής της μεταβλητής ονομάζουμε το πλήθος των μονάδων του πληθυσμού για τις οποίες η μεταβλητή παίρνει τιμές μικρότερες ή ίσες από την τιμή αυτή. Σχετική αθροιστική συχνότητα μιας τιμής της μεταβλητής θα ονομάζουμε το ποσοστό των μονάδων για τις οποίες η μεταβλητή παίρνει τιμές μικρότερες ή ίσες από την τιμή αυτή.

Η αθροιστική συχνότητα μιας τιμής της μεταβλητής προκύπτει σαν άθροισμα όλων των ταξικών συχνοτήτων για τις οποίες η μεταβλητή παίρνει τιμές μικρότερες ή ίσες από την τιμή αυτή με την προϋπόθεση ότι η τιμή αυτή είναι κάποιο από τα δεξιά άκρα των τάξεων (κλειστά δεξιά διαστήματα).

Στην περίπτωση που τα διαστήματα είναι ανοικτά δεξιά, τότε ως αθροιστική συχνότητα μιας τιμής που συμπίπτει με το δεξιό άκρο κάποιου διαστήματος, ορίζεται ο αριθμός των μονάδων του πληθυσμού, για τις οποίες η μεταβλητή παίρνει τιμές μικρότερες από την τιμή αυτή.

4,11 Κατασκευή των τάξεων μιας συνεχούς μεταβλητής

Στην περίπτωση που η μεταβλητή είναι κατηγορική ή ποιοτική, ή και ποσοτική αλλά συνεχής, τότε η πρώτη στήλη κάθε πίνακα συχνοτήτων συνίσταται από τις επιμέρους τιμές της μεταβλητής. Αν η μεταβλητή είναι συνεχής οι τιμές που παίρνει είναι άπειρες. Γι' αυτό επιλέγονται κατάλληλα διαστήματα που λέγονται τάξεις ή ομάδες.

Δύο βασικά κριτήρια για την αναζήτηση των τάξεων είναι το κριτήριο της ομοιογένειας και το κριτήριο της απλότητας.

Το κριτήριο της ομοιογένειας υπαγορεύεται να φτιάξουμε πολλές τάξεις, έτσι ώστε μέσα στην ίδια τάξη οι τιμές της μεταβλητής να μη διαφέρουν ουσιαστικά. Π.χ. δεν είναι δυνατό μέσα σε μια τάξη να περιλαμβάνονται κοντά και ψηλοί ή πλούσιοι και φτωχοί.

Από την άλλη το κριτήριο της απλότητας υπαγορεύει να κατασκευαστούν όσο το δυνατόν λιγότερες τάξεις, έτσι ώστε με ένα μικρό σχετικά πίνακα να συνοψίζεται ικανοποιητικά η κατανομή της μεταβλητής.

Τα δύο κριτήρια αυτά αλληλοσυγκρούονται αλλά στις εφαρμογές ένας αριθμός από πέντε ως δεκαπέντε τάξεις είναι συνήθως αρκετός για την κατασκευή του πίνακα συχνοτήτων, μιας συνεχούς μεταβλητής.

Αν υποθέσουμε ότι μια βιοτεχνία κατασκευής ανδρικών κουστουμιών επιθυμεί να προγραμματίσει την παραγωγή της, έτσι ώστε σε κάθε τάξη του πίνακα να αντιστοιχεί και διαφορετικό μέγεθος κουστουμιών, τότε πρέπει να κατασκευάσει εννέα διαφορετικά μεγέθη για να καλύψει τις απαιτήσεις των πελατών της.

Αν κάποια άλλη ομοειδής βιοτεχνία κρίνει ότι τα εννέα μεγέθη είναι υπερβολικά και ότι πέντε διαφορετικά μεγέθη θα είναι αρκετά, τότε θα ήταν κατάλληλος γι' αυτήν ένας πίνακας με τάξεις πλάτους 10 εκατοστών.

Μια τρίτη βιοτεχνία κατασκευής ενδυμάτων ενδεχομένως να ήταν διατεθειμένη να περιοριστεί σε τρία μεγέθη. Τότε θα αρκούσαν τρεις τάξεις πλάτους 15 εκατοστών η κάθε μια.

Διαπιστώνουμε λοιπόν ότι το πλήθος των τάξεων, καθώς και το πλάτος τους αποφασίζεται με βάση το εκάστοτε πρόβλημα και σε πολλές περιπτώσεις για να ικανοποιηθούν και τα δύο κριτήρια η κατασκευή τάξεων ίσου πλάτους είναι ανέφικτη.

Τεχνικά χαρακτηριστικά των στατιστικών πινάκων

Σε κάθε στατιστικό πίνακα πρέπει να υπάρχουν τα παρακάτω χαρακτηριστικά:

α) **Ο τίτλος.** Δηλαδή μια σαφής περίληψη για το περιεχόμενο του πίνακα.

β) **Η επικεφαλίδα.** Δηλαδή ένας χαρακτηρισμός που τοποθετείται στην κορυφή κάθε στήλης του πίνακα, με τον οποίο γίνεται φανερή η φύση των στατιστικών στοιχείων της κάθε στήλης.

γ) **Το κύριο σώμα** στο οποίο περιλαμβάνονται οι τάξεις και οι συχνότητες.

δ) Τα **σύνολα**, δηλαδή τα αθροίσματα των στηλών των συχνοτήτων.

ε) **Η πηγή.** Δηλαδή η υπηρεσία από την οποία έχουν ληφθεί τα ονομαστικά στοιχεία, καθώς και το έτος ή η ημερομηνία συλλογής των στοιχείων.

Γραφικές απεικονίσεις

Η γραφική απεικόνιση των στατιστικών στοιχείων γίνεται με τη χρήση κατάλληλων διαγραμμάτων, έτσι ώστε να υπάρχει μια πολύ καλή εικόνα της μορφολογίας του πληθυσμού. Υπάρχουν πολλοί τύποι διαγραμμάτων ανάλογα με το είδος της μεταβλητής και τον επιδιωκόμενο σκοπό.

α) Γραφική απεικόνιση πινάκων συχνοτήτων.

Για την γραφική απεικόνιση των πινάκων συχνοτήτων χρησιμοποιούνται τα ραβδωτά και τα κυκλικά διαγράμματα. Το ύψος κάθε ράβδου ή το εμβαδό κάθε κυκλικού τομέα είναι ανάλογο προς τη συχνότητα που απεικονίζει. Τα είδη αυτή των διαγραμμάτων αναφέρονται σε όλα τα είδη των μεταβλητών.

Στην ειδική περίπτωση που ενδιαφερόμαστε να απεικονίσουμε ομαδοποιημένα στατικά στοιχεία που αναφέρονται σε συνεχείς μεταβλητές, χρησιμοποιούμε το ιστόγραμμα συχνοτήτων.

Όταν ένα διάγραμμα αναφέρεται στη γραφική απεικόνιση παρατηρήσεων που λαμβάνονται σε χρονικά σημεία ή διαστήματα, τότε ονομάζεται χρονοδιάγραμμα ή διάγραμμα χρονολογικής σειράς.

Επίσης υπάρχουν και τα εικονοδιαγράμματα, δηλαδή εικόνες αντικειμένων διαφορετικού μεγέθους ώστε να απεικονιστεί κάποιος πίνακας συχνοτήτων ή η διαχρονική εξέλιξη μιας μεταβλητής.

Τέλος, πολύ χρήσιμα είναι και τα χρονοδιαγράμματα με τα οποία επιτυγχάνεται η γραφική απεικόνιση στατιστικών στοιχείων που αναφέρονται σε γεωγραφικά διαμερίσματα, πόλεις, χώρες κ.τ.λ.

Κύριο χαρακτηριστικό όλων των τύπων διαγραμμάτων είναι η γραφική απεικόνιση των πραγματικών μεγεθών με την κατάλληλη κλίμακα, έτσι ώστε να εμφανίζεται καλύτερα η νομοτέλεια των μεγεθών. Τα διαγράμματα πλεονεκτούν των στατιστικών πινάκων γιατί κεντρίζουν το ενδιαφέρον κάθε ενδιαφερόμενου και μπορούν με μια ματιά να του δώσουν σημαντικές πληροφορίες. Με τη χρήση των ηλεκτρονικών υπολογιστών, η σχεδίαση, ο χρωματισμός κ.τ.λ. ενός διαγράμματος έχει τελειοποιηθεί.

4,12 Επιλογή τύπου διαγράμματος

Κυκλικό διάγραμμα ή διάγραμμα πίτας (Pie Chart). Το διάγραμμα πίτας αποτελεί εξαιρετική επιλογή όταν θέλουμε να δείξουμε πως μια σειρά ομάδων στοιχείων σχηματίζει ένα σύνολο. Κάθε κομμάτι τις πίτας τιλοφορείται με τίτλους ή ποσοστά και παριστάνει την συνεισφορά της αντίστοιχης μονάδας. Τα διαγράμματα αυτά είναι πολύ χρήσιμα ότι δίνουμε έμφαση σε ένα σημείο και αυτό μπορεί να χωριστεί από τα άλλα. Χαρακτηριστικά παραδείγματα διαγραμμάτων πίτας είναι η

κατανομή ποσοστών στα κόμματα, η κατανομή κρατικών κονδυλίων κ.τ.λ.

Διάγραμμα Δακτυλίων (Donought Chart). Αποτελεί παραλλαγή του διαγράμματος πίτας με τη διαφορά ότι εδώ μπορεί να απεικονισθεί άνω της μιας σειράς. Κάθε σειρά απεικονίζεται ως ένας διαφορετικός ομόκεντρος κύκλος.

Γραμμικά διαγράμματα (Line Charts). Ένα γραμμικό διάγραμμα δείχνει μια συναρτησιακή σχέση μεταξύ μιας ή περισσότερων μεταβλητών σε σχέση με μια ανεξάρτητη μεταβλητή. Τα γραμμικά διαγράμματα χρησιμοποιούνται για αποκάλυψη τάσεων και την παρακολούθηση μεταβλητών μιας ανεξάρτητης μεταβλητής. Παραδείγματα τέτοιων διαγραμμάτων είναι οι απεικονίσεις μεταβολών πληθυσμού ως προς το χρόνο, το Μέγιστο – Ελάχιστο – Κλείσιμο (High – Low – Close) και Άνοιγμα – Μέγιστο – Ελάχιστο (Open – High – Low – Close). Τα δύο τελευταία είναι κατάλληλα για να δείξουν την περιοχή κύμανσης δεδομένων σε χρονικό διάστημα π.χ. παρακολούθηση τιμών μετοχών.

Διαγράμματα εμβαδού (Area Charts). Και τα διαγράμματα αυτά αποκαλύπτουν τάσεις με έμφαση όμως στο ποσοστό μεταβολής αντί του ρυθμού. Παραδείγματα τέτοιων διαγραμμάτων είναι η απεικόνιση όγκου παραγωγής, πωλήσεων κ.τ.λ.

Ραβδογράμματα (Bar & Column Charts). Τα ραβδογράμματα παριστάνουν τιμές δεδομένων με ύψη ράβδων σε σχέση με μια βασική γραμμή. Είναι εξαιρετική επιλογή για τη σύγκριση μεταβλητών που μετρήθηκαν σε διακριτές χρονικές στιγμές. Αυτό γιατί τα μήκη - ή ύψη - των ράβδων είναι πολύ εύκολα συγκρίσιμα μια και οι ράβδοι είναι η μια δίπλα στην άλλη. Επίσης τα διαγράμματα εμβαδού αποκαλύπτουν και τάσεις. Οι ράβδοι μπορεί να είναι είτε οριζόντιες (bar) είτε κάθετες (column).

Σωρευμένα ραβδοδιαγράμματα (Stecked bar Chart). Τα διαγράμματα αυτά αποτελούν παραλλαγή των απλών ραβδοδιαγραμμάτων. Σ' αυτά οι ράβδοι σωρεύονται η μια πάνω στην άλλη και έτσι γίνεται ορατή και η σχετική συνεισφορά αλλά και το σύνολο. Χρησιμοποιούνται εναλλακτικά προς τα ραβδοδιαγράμματα όταν ο χώρος είναι περιορισμένος. Μια υποπερίπτωση είναι το ραβδόγραμμα 100% που μπορεί να χρησιμοποιηθεί εναλλακτικά προς το διάγραμμα πίτας.

Ακτινωτό διάγραμμα (Radar Chart). Χρησιμοποιείται για να δείξει μεταβολές συχνοτήτων μεταξύ των σειρών σε σχέση με ένα κεντρικό σημείο αλλά και μεταξύ τους. Κάθε κατηγορία αντιστοιχίζεται με άξονα τύπου ακτίνας.

Διάγραμμα X-Y (Διασποράς – Scatter). Απεικονίζει επιμέρους σημεία δεδομένων με ή χωρίς γραμμές σύνδεσης. Χρησιμοποιούνται συνήθως για να δείξουν μια στατιστική σχέση αντί για μια συναρτησιακή. Τα διαγράμματα αυτά μπορούν και να χρησιμοποιηθούν και για γενική σχεδίαση δύο ομάδων αριθμών με βάση σύστημα αξόνων XY. Αποτελούν την καλύτερη επιλογή διδιάστατων διαγραμμάτων για επιστημονικές εφαρμογές. Τα διαγράμματα XY χειρίζονται τα δεδομένα διαφορετικά από τα γραμμικά. Τα XY πρέπει να χρησιμοποιούνται για διασπαρμένα δεδομένα ή για δεδομένα σε άνισα διαστήματα.

Διάγραμμα συνδυασμών (Combination). Με αυτό υπερτίθεται ένας τύπος διαγράμματος πάνω σε έναν άλλον. Χρησιμεύει για σύγκριση δεδομένων διαφορετικών τύπων ή δεδομένων που απαιτούν διαφορετικές κλίμακες αξόνων.

Τρισδιάστατα διαγράμματα (3D Charts). Αποτελούν επέκταση των διδιάστατων τύπων. Η διαφορά τους είναι η τρίτη διάσταση η οποία δίνει μια πιο επαγγελματική αίσθηση στις απεικονίσεις, ιδίως όταν προορίζονται για παρουσιάσεις. Π.χ. ένα τρισδιάστατο ραβδόγραμμα

δίνει μεγαλύτερη έμφαση στις τιμές και τις συγκρίσεις, ενώ ένα 3D-διάγραμμα πίτας προσελκύει περισσότερο την προσοχή, από το αντίστοιχο δισδιάστατο. Όμως τα διαγράμματα αυτά πρέπει να χρησιμοποιούνται με προσοχή, γιατί όταν τα δεδομένα είναι πολλά, οι μικρότερες τιμές μπορεί να γίνουν δυσδιάκριτες από τις συγκεκριμένες γωνίες.

β) Ιστογράμματα

Στην περίπτωση που τα διαγράμματα αφορούν συνεχείς μεταβλητές η γραφική απεικόνιση των δεδομένων ονομάζεται ιστόγραμμα. Κάθε ορθογώνιο του διαγράμματος ονομάζεται ιστός.

Αν θεωρήσουμε ότι οι τάξεις έχουν ίδιο πλάτος και ότι το πλάτος αυτό είναι ίσο με τη μονάδα, τότε το εμβαδό κάθε ιστού θα συμπίπτει αριθμητικά με το ύψος του. Αν χαράξουμε τους ιστούς έτσι ώστε το ύψος τους να είναι ίσο με την αντίστοιχη συχνότητα, τότε το άθροισμα των εμβαδών των ιστών θα συμπίπτει αριθμητικά με το μέγεθος του πληθυσμού. Το εμβαδόν που περικλείεται από το πολύγωνο συχνοτήτων και τον οριζόντιο άξονα είναι ίσο προς το άθροισμα των εμβαδών των ιστών και το εμβαδό αυτό έχει μέτρο όσο και το μέγεθος του πληθυσμού.

Η καμπύλη συχνοτήτων είναι μια οριακή κατάσταση του πολυγώνου συχνοτήτων όταν το πλάτος των τάξεων συνεχώς μικραίνει.

γ) Αθροιστικά διαγράμματα

Αναφέρονται στη γραφική απεικόνιση των αθροιστικών συχνοτήτων μιας ποσοτικής μεταβλητής.

Η κατασκευή τους γίνεται ως εξής: στο τέλος κάθε τάξης υψώνουμε κάθετη ίση προς την αντίστοιχη συχνότητα και ενώνουμε τα άκρα των κάθετων αυτών με μια τεθλασμένη γραμμή, η οποία έχει σχήματα τελικού σίγμα. Όταν η μεταβλητή είναι συνεχής έχει τη μορφή μιας γνησίως αύξουσας συναρτήσεως ενώ όταν είναι συνεχής έχει τη μορφή μιας κλιμακωτής αύξουσας συνάρτησης.

4,13 Στατιστικές εκθέσεις ή αναφορές

Κάθε αναφορά που στηρίζεται σε στατιστικά δεδομένα ονομάζεται στατιστική έκθεση. Όταν η έκθεση στηρίζεται στην απλή περιγραφή ενός φαινομένου μπορεί να διεξαχθεί από μη ειδικούς επιστήμονες όπως γιατρούς, ψυχολόγους κ.τ.λ. Όταν όμως περιέχει προβλέψεις για την μελλοντική εκτίμηση κάποιου φαινομένου ή ακόμα πολύπλοκες αναλύσεις, τότε είναι έργο ενός ειδικού στατιστικολόγου.

Υπάρχουν βέβαια και ορισμένες κακές εκθέσεις οι οποίες οφείλονται σε αμφίβολης ποιότητας στατιστικό υλικό και γίνονται κυρίως για πολιτικές ή άλλες σκοπιμότητες.

ΚΕΦΑΛΑΙΟ 5

Διαστήματα, φράγματα και περιοχές εμπιστοσύνης

Ορισμός 1: Εστω X_1, X_2, \dots, X_n ένα τυχαίο δείγμα από την κατανομή $f(x, \theta)$, $\theta \in \Theta$ και $g(\theta)$ μια παραμετρική συνάρτηση και $\delta = \delta(x)$ εκτιμητήρια της $g(\theta)$. Ορίζουμε τυχαία φράγματα της παραμέτρου $g(\theta)$ όπου να την περιέχουν

$$[\delta(x) - d, \delta(x) + d]$$

$$L = L(x) = \delta(x) - d, u = u(x) = \delta(x) + d$$

Το τυχαίο διάστημα

$$I = I(x) = (L, U) = [L(x), U(x)]$$

Με άκρα τις στατιστικές συναρτήσεις

$$L = L(x), U = U(x) \quad [L(x) < U(x)]$$

Αποτελεί ένα διάστημα εμπιστοσύνης για την $g(\theta)$ με επίπεδο σημαντικότητας

$$1 - \alpha, 0 < \alpha < 1$$

αν για όλα τα $\theta \in \Theta$ ισχύει

$$P\{g(\theta) \in I(x)\} \geq 1 - \alpha$$

Αν $P\{g(\theta) \in I(x)\} = 1 - \alpha$, ισχύει δηλ. σαν ισότητα, τότε το διάστημα I έχει συντελεστή εμπιστοσύνης $1 - \alpha$.

Παράδειγμα 1: Εστω X_1, X_2, \dots, X_n ένα τυχαίο δείγμα από την κανονική κατανομή $N(\mu, \sigma^2)$ με $\mu \in \mathbb{R}$ άγνωστη παράμετρο και $\sigma^2 > 0$ γνωστό.

Για να δημιουργήσουμε ένα φράγμα εμπιστοσύνης για το μ με συντελεστή εμπιστοσύνης $1 - \alpha$ παίρνουμε τη σχέση

$$P\{|x - \mu| \leq d\} = P\left\{\left|\frac{x - \mu}{\sigma/\sqrt{n}}\right| \leq \frac{d/\sigma/\sqrt{n}}{1} \right\} = 1 - \alpha$$

Λύνοντας τη σχέση έχουμε

$$P[z > z_p] = 1 - \Phi(z_p) = \alpha/2$$

Όπου z μια τυποποιημένη κανονική τυχαία μεταβλητή $N(0,1)$ και Z_p το άνω ποσοστιαίο σημείο της κατανομής $N(0,1)$.

Οπότε το διάστημα εμπιστοσύνης για το μ με συντελεστή $1-\alpha$ δίνεται από τον τύπο

$$I = \left[x - \frac{\sigma}{\sqrt{v}} Z_{\alpha/2}, x + \frac{\sigma}{\sqrt{v}} Z_{\alpha/2} \right]$$

Το πλάτος του διαστήματος θα είναι ίσο με

$$\mu(I) = \left(x + \frac{\sigma}{\sqrt{v}} Z_{\alpha/2} \right) - \left(x - \frac{\sigma}{\sqrt{v}} Z_{\alpha/2} \right) = 2 \frac{\sigma}{\sqrt{v}} Z_{\alpha/2}$$

Αν το μήκος $\mu(I) = \mu[I(x)]$ ενός διαστήματος εμπιστοσύνης είναι μια τυχαία μεταβλητή και όχι σταθερή ποσότητα, τότε εξετάζουμε το μέσο μήκος $E\theta[\mu(I(x))]$ για τις διάφορες τιμές της παραμέτρου θ .

Μπορούμε αντί να φράσουμε μια παραμετρική συνάρτηση $g(\theta)$ μεταξύ ενός διαστήματος εμπιστοσύνης (σε δύο ποσότητες L και U) αρκεί να ορίσουμε ένα άνω ή κάτω φράγμα εμπιστοσύνης. Έτσι ορίζουμε:

Ορισμός 2: Εστω X_1, X_2, \dots, X_n ένα τυχαίο δείγμα από την κατανομή $f(x, \theta)$ $\Theta \in \Theta$ και $g(\theta)$ μια παραμετρική συνάρτηση, τότε

i) Η συνάρτηση $L=L(x)$ θα ονομάζεται **κάτω φράγμα εμπιστοσύνης** για την $g(\theta)$ σε επίπεδο σημαντικότητας $1-\alpha$ αν για κάθε $\Theta \in \Theta$ ισχύει

$$P[L(x) \leq g(\theta)] \geq 1-\alpha$$

ii) Η συνάρτηση $U=U(x)$ θα ονομάζεται **άνω φράγμα εμπιστοσύνης** για την $g(\theta)$ σε επίπεδο σημαντικότητας $1-\alpha$ αν για κάθε $\theta \in \Theta$ ισχύει:

$$P[U(x) \leq g(\theta)] \geq 1-\alpha$$

Ομοίως σε περίπτωση ισότητας των πιθανοτήτων με $1-\alpha$ τότε θα έχουμε ένα διάστημα εμπιστοσύνης με συντελεστή εμπιστοσύνης $1-\alpha$.

1. Γενική μέθοδος κατασκευής διαστημάτων εμπιστοσύνης

Εστω X_1, X_2, \dots, X_n ένα τυχαίο δείγμα από την κατανομή $f(x, \theta)$, $\theta \in \Theta \subset \mathbb{R}^s$ και $g(\theta)$ μια παραμετρική συνάρτηση. Και θέλουμε να κατασκευάσουμε ένα διάστημα εμπιστοσύνης για την $g(\theta)$ με συντελεστή εμπιστοσύνης $1-\alpha$, το οποίο να χρησιμοποιεί το δείγμα $x=(x_1, x_2, \dots, x_n)$.

Τα επόμενα βήματα μας δείχνουν μια γενική μέθοδο κατασκευής τέτοιων διαστημάτων, η οποία μπορεί να εφαρμοστεί με απόλυτη επιτυχία στις περισσότερες από τις γνωστές κατανομές.

Αρα έχουμε:

B1. Προσδιορίζουμε μια στατιστική συνάρτηση $T=T(x)$ της οποίας η κατανομή να εξαρτάται από την παράμετρο θ .

B2. Κατασκευάζουμε μια συνάρτηση

$$Y=\Psi[T, g(\theta)]$$

Του T και του $g(\theta)$ τέτοια ώστε η κατανομή της Y να μην εξαρτάται από το θ .

B3. Υπολογίζουμε δύο σταθερές $C_1 \leq C_2$ ώστε να ισχύει

$$P[C_1 \leq Y \leq C_2] = 1-\alpha$$

οι οποίες να εξαρτώνται μόνο από την κατανομή της τυχαίας μεταβλητής Y και το $\alpha \in (0, 1)$.

B4. Λύνουμε τη σχέση

$$C1 \leq \Psi[T(x), g(\theta)] \leq C2$$

Ως προς το $g(\theta)$ όπου παίρνουμε μια διπλή ανισότητα

$$L=L(x) \leq g(\theta) \leq U(x)=U$$

Το διάστημα που βρέθηκε

$$I = [L, U]$$

είναι ένα διάστημα εμπιστοσύνης για το $g(\theta)$ με συντελεστή εμπιστοσύνης $1-\alpha$.

Αριθμητική εφαρμογή 1: Οι χρόνοι ζωής ενός δείγματος από $V=5$ λυχνίες ήταν 7, 8, 11, 10, 14. Να βρεθεί ένα διάστημα εμπιστοσύνης με συντελεστή εμπιστοσύνης $1-\alpha=90\%$ για το μέσο χρόνο ζωής των λυχνιών, αν υποθεθεί ότι οι χρόνοι ζωής ακολουθούν εκθετική κατανομή.

Με βάση τους τύπους θα έχουμε

$$\alpha) L=L(x) = \frac{x_{2v}^2 \left(1 - \frac{\alpha}{2}\right)}{2 \sum_{i=1}^v x_i} \quad (\text{κάτω άκρο})$$

$$L(x) = \frac{x_{2 \cdot 5}^2 \left(1 - \frac{1-90\%}{2}\right)}{2 \cdot 50} = \frac{x_{10}^2 (0,95)}{100} = 0,0394$$

$$\beta) U=U(x) = \frac{x_{2v}^2 \left(\frac{\alpha}{2}\right)}{2 \cdot \sum_{i=1}^v x_i} \quad (\text{άνω άκρο})$$

$$U(x) = \frac{x_{2v}^2 (0,05)}{2 \cdot 50} = 0,1831$$

Αρα ένα διάστημα εμπιστοσύνης για την παράμετρο θ θα είναι το $[0,0394, 0,1831]$

Αρα το διάστημα εμπιστοσύνης για το μέσο χρόνο ζωής $1/\theta$ θα είναι

$$\left[\frac{1}{0,1831}, \frac{1}{0,0394} \right] = [5,46, 25,38]$$

Αριθμητική εφαρμογή 2: Ο χρόνος ζωής X κάποιου μικροοργανισμού ακολουθεί ομοιόμορφη κατανομή στο διάστημα $(0, \theta)$. Να δοθεί διάστημα εμπιστοσύνης με συντελεστή εμπιστοσύνης $1-\alpha=0,90$ για

(i) το μέσο χρόνο ζωής

(ii) τη διασπορά του χρόνου ζωής αν ένα δείγμα μικροοργανισμών καταστράφηκε σε 40 ώρες.

Παίρνουμε το διάστημα (L, U) με άκρα

$$L=L(x)=\frac{X(v)}{\sqrt{1-\frac{\alpha}{2}}}=\frac{40}{\sqrt{1-\frac{0,10}{2}}}=40,06$$

$$U=U(x)=\frac{X(v)}{\sqrt{\frac{\alpha}{2}}}=\frac{40}{\sqrt{\frac{0,10}{2}}}=44,20$$

Άρα το διάστημα εμπιστοσύνης για το μέσο χρόνο ζωής $\theta/2$ θα είναι:

$$[40,06/2, 44,20/2] = [20,03, 22,10]$$

και για τη διασπορά του χρόνου ζωής

$$[40,06^2/12, 44,20^2/12]$$

2. Διαστήματα εμπιστοσύνης για τις παραμέτρους ενός κανονικού πληθυσμού

Σ' αυτή την παράγραφο θα προσπαθήσουμε να εξετάσουμε το πρόβλημα εύρεσης διαστήματος εμπιστοσύνης για τις παραμέτρους ενός κανονικού πληθυσμού.

Υποθέτουμε ότι έχουμε ένα τυχαίο δείγμα X_1, X_2, \dots, X_n μεγέθους n από την κανονική κατανομή $N(\mu, \sigma^2)$ και αναζητούμε διαστήματα εμπιστοσύνης για τις παραμέτρους μ και σ^2 με συντελεστή εμπιστοσύνης $1-\alpha$.

Οπότε έχουμε τις εξής περιπτώσεις.

I) Διάστημα εμπιστοσύνης για τον μέσο μ όταν η διαφορά σ^2 είναι γνωστή

τότε η τυχαία μεταβλητή

$$Y = \frac{x - \mu}{\sigma/\sqrt{v}}$$

ακολουθεί την τυποποιημένη κανονική κατανομή $N(0,1)$ οπότε

$$P[-Z_{\alpha/2} \leq Y \leq Z_{\alpha/2}] = 1 - \alpha$$

Αρα το διάστημα εμπιστοσύνης για το μ θα είναι το

$$I = \left[x - \frac{\sigma}{\sqrt{v}} Z_{\alpha/2}, x + \frac{\sigma}{\sqrt{v}} Z_{\alpha/2} \right]$$

Το πλάτος του διαστήματος εμπιστοσύνης θα είναι:

$$\mu(I) = \frac{2\sigma}{\sqrt{v}} Z_{\alpha/2}$$

και δεν αποτελεί τυχαία μεταβλητή.

II. Διάστημα εμπιστοσύνης για το μέσο μ , όταν η διασπορά σ^2 είναι άγνωστη

Και εδώ η τυχαία μεταβλητή

$$Z = \frac{x - \mu}{\sigma/\sqrt{v}}$$

ακολουθεί την τυποποιημένη κανονική κατανομή $N(0,1)$, επειδή όμως περιέχει την άγνωστη παράμετρο σ^2 δεν μπορούμε να την χρησιμοποιήσουμε απ' ευθείας για να δημιουργήσουμε διάστημα εμπιστοσύνης. Έτσι παίρνουμε και την τυχαία μεταβλητή

$$Q = \frac{(v-1)S^2}{\sigma^2} = \sum_{i=1}^v \left(\frac{x_i - \bar{x}}{\sigma} \right)^2$$

η οποία είναι ανεξάρτητη του Z και ακολουθεί την κατανομή χ^2 με $v-1$ βαθμούς ελευθερίας και σχηματίζει το πηλίκο

$$\frac{Z}{\sqrt{\frac{\alpha}{v-1}}} = \frac{\bar{x} - \mu}{S/\sqrt{v}}$$

το οποίο θα ακολουθεί την κατανομή t με $v-1$ βαθμούς ελευθερίας.

Άρα

$$P[-t_{v-1}(\alpha/2) \leq \frac{\bar{x} - \mu}{S/\sqrt{v}} \leq t_{v-1}(\alpha/2)] = 1 - \alpha$$

και το διάστημα εμπιστοσύνης για το μ θα δίνεται από τον τύπο

$$I = \left[\bar{x} - \frac{S}{\sqrt{v}} t_{v-1}(\alpha/2), \bar{x} + \frac{S}{\sqrt{v}} t_{v-1}(\alpha/2) \right]$$

III. Διαστήματα εμπιστοσύνης για την διασπορά σ^2 όταν ο μέσος μ είναι γνωστός

Η τυχαία μεταβλητή:

$$Y = \sum_{i=1}^v \left(\frac{x_i - \mu}{\sigma} \right)^2 = \frac{\sum_{i=1}^v (x_i - \mu)^2}{\sigma^2}$$

ακολουθεί την κατανομή χ^2 με v βαθμούς ελευθερίας. Άρα θα ισχύει η σχέση

$$P[\chi^2_v(1-\alpha/2) \leq Y \leq \chi^2_v(\alpha/2)] = 1 - \alpha$$

και ένα διάστημα εμπιστοσύνης θα δίνεται από τη σχέση

$$I = \left[\frac{\sum_{i=1}^v (x_i - \mu)^2}{\chi^2_v(\alpha/2)}, \frac{\sum_{i=1}^v (x_i - \mu)^2}{\chi^2_v(1-\alpha/2)} \right]$$

IV. Διάστημα εμπιστοσύνης για την διασπορά σ^2 όταν ο μέσος μ είναι άγνωστος

Παίρνουμε την τυχαία μεταβλητή

$$Y = \sum_{i=1}^v \left(\frac{x_i - \bar{x}}{\sigma} \right)^2 = \frac{(v-1)S^2}{\sigma^2}$$

η οποία ακολουθεί την κατανομή χ^2 με $v-1$ βαθμούς ελευθερίας. Άρα θα ισχύει η σχέση

$$P[\chi^2_{v-1}(1-\alpha/2) \leq Y \leq \chi^2_{v-1}(\alpha/2)] = 1-\alpha$$

Και ένα διάστημα εμπιστοσύνης για το σ^2 θα δίνεται από την

$$I = \left[\frac{(v-1)S^2}{\chi^2_{v-1}(\frac{\alpha}{2})}, \frac{(v-1)S^2}{\chi^2_{v-1}(1-\frac{\alpha}{2})} \right]$$

V. Περιοχή εμπιστοσύνης για το ζεύγος $\theta=(\mu, \sigma^2)$ των παραμέτρων της κανονικής κατανομής

Όταν και οι δύο παράμετροι της κανονικής κατανομής $N(\mu, \sigma^2)$ είναι άγνωστοι, θα πρέπει εκτός από το διάστημα εμπιστοσύνης για το μ και για το σ^2 (περίπτωση II και III αντίστοιχα) θα πρέπει να αναζητήσουμε και περιοχές εμπιστοσύνης για το ζεύγος $\theta=(\mu, \sigma^2)$, δηλ. τυχαία υποσύνολα $I=I(x)$ του $R \times R^*$ ώστε να ισχύει

$$P[\theta \in I] = 1-\alpha$$

παίρνουμε το α σαν άθροισμα δύο θετικών προσθετέων

$$\alpha = \alpha_1 + \alpha_2$$

και ονομάζουμε

$$I_1 = I_1(x) = \left[x - \frac{S}{\sqrt{v}} t_{v-1}(\alpha_1/2), x + \frac{S}{\sqrt{v}} t_{v-1}(\alpha_2/2) \right]$$

το διάστημα εμπιστοσύνης για το μ με συντελεστή εμπιστοσύνης $1-\alpha_1$ και

$$I_2 = I_2(x) = \left[\frac{(v-1)S^2}{X^2_{v-1}(\frac{\alpha_2}{2})}, \frac{(v-1)S^2}{X^2_{v-1}(1-\frac{\alpha_2}{2})} \right]$$

το διάστημα εμπιστοσύνης για το σ^2 με συντελεστή εμπιστοσύνης $1-\alpha_2$. Τότε θα έχουμε

$$P[\theta \in I_1 \times I_2] = P[\mu \in I_1 \text{ και } \sigma^2 \in I_2] = 1 - P[\mu \in I_1 \text{ ή } \sigma^2 \in I_2]$$

και αφού ισχύει

$$P[A \cup B] = P[A] + P[B] - P[A \cap B] \leq P[A] + P[B]$$

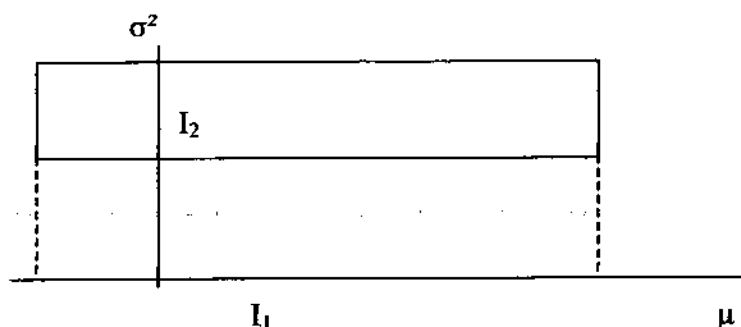
τότε

$$P[\theta \in I_1 \times I_2] \geq 1 - P[\mu \in I_1] - P[\sigma^2 \in I_2] \text{ και}$$

$$P[\theta \in I_1 \times I_2] \geq 1 - (\alpha_1 + \alpha_2) = 1 - \alpha$$

Η τελευταία σχέση μας δείχνει ότι το τυχαίο κλειστό ορθογώνιο $I_1 \times I_2$ (σχ. 1) αποτελεί ένα διάστημα εμπιστοσύνης για το $\theta = (\mu, \sigma^2)$ με συντελεστή εμπιστοσύνης τουλάχιστον $1-\alpha$.

Σχήμα 1



3. Προσεγγιστικά διαστήματα εμπιστοσύνης για μεγάλα δείγματα

Έστω ότι έχουμε ένα τυχαίο δείγμα X_1, X_2, \dots, X_n από άγνωστη κατανομή $f(x; \theta)$, $\theta \in \theta \subset \mathbb{R}^s$ και θέλουμε ένα διάστημα εμπιστοσύνης για την παραμετρική συνάρτηση $g(\theta)$. Αν το μέγεθος n του δείγματος είναι μεγάλο ($n \rightarrow \infty$), μπορούμε, να δημιουργήσουμε απλά διαστήματα εμπιστοσύνης αν θεωρήσουμε ότι η ακολουθία των δειγματικών μέσων

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

συγκλίνει κατά νόμο στην κανονική κατανομή (κεντρικό οριακό θεώρημα) δηλ. ότι η τυχαία μεταβλητή

$$Z_n = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}, \quad \mu = E(x_i), \sigma^2 = \Delta(x_i)$$

Ακολουθεί προσεγγιστικά την τυποποιημένη κανονική κατανομή $N(0,1)$.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. **Θ. Αποστολόπουλος** : *Στατιστική Επιχειρήσεων*, ΤΕΙ Αθήνας
2. **Χ. Δαμιανού – Μ. Κούτρας** : *Εισαγωγή στη Στατιστική*, Εκδόσεις Αίθρα
3. **Α. Κιντής** : *Συγκριτική Στατιστική Ανάλυση*, Εκδόσεις Gutenberg
4. **Κιόχος** : *Μεθοδολογία διεξαγωγής ερευνών*, Εκδόσεις Σταμούλης
5. **Ο. Παπαδήμας – Χ. Κούλιας** : *Εφαρμοσμένη Στατιστική*, Εκδόσεις Νέων Τεχνολογιών
6. **Φράγκος** : *Στατιστική Επιχειρήσεων* , Εκδόσεις Σταμούλης
7. **Ν. Τσερπές – Φ. Αλεβίτος** : *Εισαγωγή στη θεωρία Δειγματοληψίας*, Εκδόσεις Πανεπιστημίου Πατρών.

