ΑΝΩΤΑΤΟ ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ

ΠΑΤΡΑΣ (Α.Τ.Ε.Ι.)

ΤΜΗΜΑ ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΥ ΣΧΕΔΙΑΣΜΟΥ &

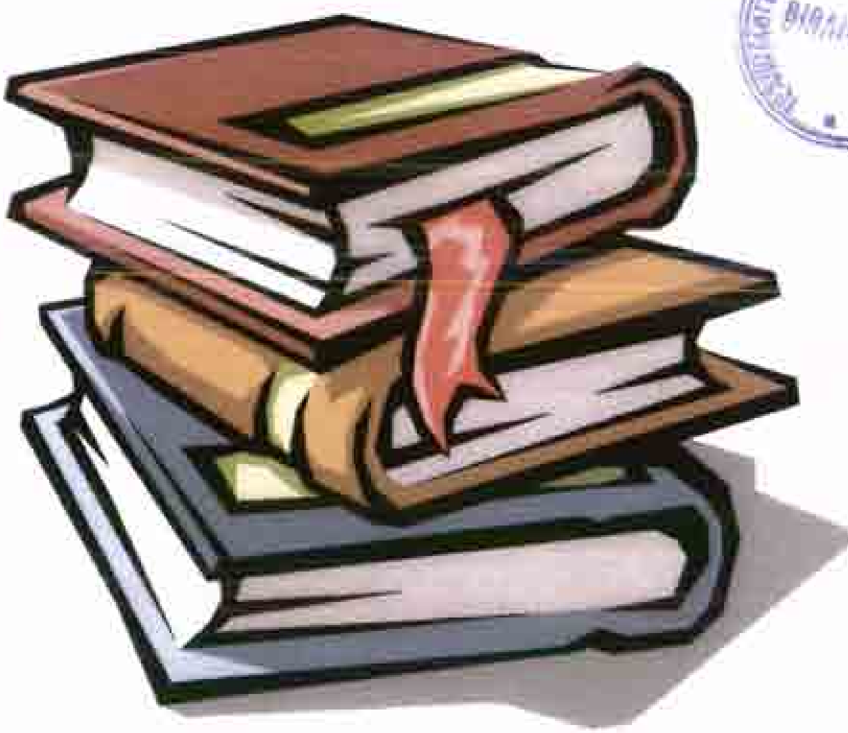
ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΒΑΣΙΚΕΣ ΜΕΘΟΔΟΙ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

ΚΑΙ ΕΦΑΡΜΟΓΕΣ

ΕΚΤΙΜΗΣΕΙΣ ΠΑΡΑΜΕΤΡΩΝ ΤΟΥ ΠΛΗΘΥΣΜΟΥ

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΠΑΤΡΑΣ
ΒΙΒΛΙΟΘΗΚΗ



ΕΙΣΗΓΗΤΡΙΑ : ΜΠΟΥΜΠΟΥΛΗ ΑΘΑΝΑΣΙΑ

ΕΠΙΜΕΛΕΙΑ : ΟΥΚΑ ΧΡΥΣΑΝΘΗ (Θ 341)

ΠΑΤΡΑ , 14 ΝΟΕΜΒΡΙΟΥ 2005

ΑΡΙΘΜΟΣ ΕΙΣΑΓΩΓΗΣ 5869

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ ΤΡΙΤΟ

ΚΕΦΑΛΑΙΟ 1

1.1 ΕΙΣΑΓΩΓΙΚΗ ΕΝΝΟΙΑ

Η δειγματοληψία είναι μέρος της στατιστικής . Η στατιστική προέρχεται από την λέξη status που σημαίνει κράτος-κατάσταση και αναφέρεται στη συλλογή στοιχείων για τις κρατικές ανάγκες . Η στατιστική είναι επίσης υπεύθυνη για την ανάλυση των παρατηρήσεων ή μετρήσεων που αναφέρονται σε ένα συγκεκριμένο αντικείμενο ή γεγονός . Χωρίζεται στην :

1)ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ (descriptive statistics) που είναι η διαδικασία της εξαγωγής σημαντικών πληροφοριών , όπως ο υπολογισμός του μέσου όρου ενός συνόλου αριθμών .

2)ΕΠΑΓΩΓΙΚΗ ΣΤΑΤΙΣΤΙΚΗ (statistical inference) που είναι η διαδικασία κατά την οποία χρησιμοποιούμε τις παρατηρήσεις που έχουμε κάνει σε ένα δείγμα για να εκτιμήσουμε ιδιότητες ενός μεγαλύτερου πληθυσμού .

Απαρτίζεται από τις επιστημονικές μεθόδους συλλογής , οργάνωσης , παρουσίασης , ανάλυσης και ερμηνείας αριθμητικών στατιστικών δεδομένων προκειμένου να βγούν τα απαιτούμενα συμπεράσματα για την λήψη ορθών αποφάσεων . Η στατιστική έχει επαγωγικό χαρακτήρα . Μελετά και αναλύει το μέρος (δείγμα)και βγάζει συμπεράσματα για το όλοι (πληθυσμός) . Χρησιμοποιείται σε όλες τις επιστήμες (Ιατρική , Κοινωνιολογία , Φυσική , Πολιτικές επιστήμες , Μαθηματικά , Οικονομικές και διοικητικές επιστήμες) .

Ως μέθοδος ''ανάλυσης'' η στατιστική εφαρμόζεται στην μελέτη οποιουδήποτε μεγέθους . Τα αποτελέσματα της μεθόδου μπορούν να έχουν ένα χαρακτήρα σταθερότητας και γενικότητας , μόνο εάν το σύνολο είναι αντιπροσωπευτικό . Η στατιστική είναι και μέθοδος περιγραφής ''πολύμορφων συνόλων'' . Αποσκοπεί στην απλοποίηση , σύμπηξη και σύνθεση .

Για την πραγματοποίηση αυτών των μεθόδων και την συγκέντρωση στοιχείων μιας έρευνας η στατιστική εφαρμόζει διάφορες μεθόδους, οι οποίες παρουσιάζουν πλεονεκτήματα και μειονεκτήματα ανάλογα με το είδος της έρευνας και το σκοπό για τον οποίο πραγματοποιείται .

Οι μέθοδοι αυτοί είναι :

ΔΕΙΓΜΑΤΟΛΗΨΙΕΣ ⟷ **ΕΞΑΝΤΛΗΤΙΚΕΣ ΕΡΕΥΝΕΣ**

1.2 ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΚΑΙ ΟΡΙΣΜΟΙ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Όσον αφορά την μέθοδο της δειγματοληψίας υπάρχουν πολλοί ορισμοί που καθορίζουν αυτή την έννοια :

Δειγματοληψία (sampling) ονομάζεται το τμήμα εκείνο της στατιστικής που απασχολείται με τις μεθόδους επιλογής δείγματος , με τα σφάλματα που τυχόν προκύπτουν , τον προσδιορισμό του κατάλληλου μεγέθους του δείγματος κ . α . Άλλωστε , είναι διαπιστωμένο τόσο στην καθημερινή ζωή όσο και στην επιστημονική έρευνα ότι η γνώση μας , η συμπεριφορά μας και οι πράξεις μας βασίζονται πάνω σε δείγματα . Στις καθημερινές μας ενέργειες όπως πριν αγοράσουμε ένα ρούχο , κοιτάμε τις βιτρίνες ορισμένων καταστημάτων . Χρησιμοποιούμε λοιπόν την δειγματοληψία για να κερδίσουμε σε ποιότητα υπηρεσιών , σε χρήμα , χρόνο και γενικότερα για να κερδίσουμε πληροφορίες και γνώσεις σε ένα πληθυσμό . Φυσικά θα μπορούσαμε να έχουμε πλήρη γνώση για ένα πληθυσμό αν κάνουμε απογραφή . Τις περισσότερες όμως φορές αυτό είναι αδύνατο , για αυτό στρεφόμαστε στην δειγματοληψία αρκούμενη στην μερική γνώση ή την γνώση με πιθανότητα . Είναι μεγάλη η πιθανότητα να έχουμε βγάλει σωστά συμπεράσματα αν το δείγμα έχει επιλεγεί με τέτοιο τρόπο που να αντανακλά αρκετά καλά τον πληθυσμό .

Οι δειγματοληψίες συνίστανται στην απογραφή ενός μέρους του πληθυσμού που ονομάζεται δείγμα .

Οι συγκεκριμένες μονάδες που επιλέγονται από το σύνολο των μονάδων του πληθυσμού γίνεται τυχαία ή με τέτοιο τρόπο ώστε να διατηρείται η αντιπροσωπευτικότητα και μας επιτρέπουν να ανάγουμε τα συμπεράσματα που εξάγουμε σε ολόκληρο το πληθυσμό . Ως παράδειγμα δειγματοληψιών μπορούμε να αναφέρουμε τις σφυγμομετρήσεις (gallops) ή την έρευνα της αγοράς για την εισαγωγή ενός νέου προϊόντος κ . λ . π . Έτσι μπορούμε να συμβολίσουμε με n το πλήθος των στοιχείων του δείγματος που παίρνουμε μέσα από το πληθυσμό N . Επειδή όμως στην πράξη το N είναι πολύ μεγάλο , είναι πολύ δύσκολο και οικονομικά ασύμφορο να μελετήσουμε τον πληθυσμό εξετάζοντας ένα προς ένα τα στοιχεία του , για αυτό εξετάζουμε ένα μικρό τμήμα του , το δείγμα n .

Η απογραφή ακόμα και σε πεπερασμένους πληθυσμούς είναι οικονομικά ασύμφορη και πολλές φορές αδύνατη . Δεν είναι δυνατόν π . χ . να κάψουμε όλα τα λάστιχα για αυτοκίνητα που παράγει ένα εργοστάσιο προκειμένου να ελέγξουμε την ποιότητα τους . Για τον λόγο αυτό προτιμούμε τον δεύτερο τρόπο συλλογής στοιχείων , σύμφωνα με τον οποίο εξετάζουμε ένα μέρος , δείγμα του πληθυσμού . Φροντίζουμε δε , να επιλέγουμε το δείγμα , με τέτοιο τρόπο , που να εκπροσωπεί τον πληθυσμό .

1.3 Η ΔΙΑΔΙΚΑΣΙΑ ΤΗΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Αφού ο ερευνητής έχει καθορίσει το πρόβλημα με κάθε λεπτομέρεια, έχει κάνει την επιλογή του τύπου της έρευνας αγοράς που θα ακολουθήσει και έχει σχεδιάσει τα μέσα τα οποία θα χρησιμοποιήσει για να συλλέξει τα στοιχεία , αυτό που πρέπει να επακολουθήσει είναι η επιλογή των ερωτώμενων εκείνων από τους οποίους θα συγκεντρωθούν τα απαραίτητα στοιχεία . Έτσι από την μια πλευρά είναι καλό να ερωτηθούν όλοι οι ερωτώμενοι που κρίνονται ως κατάλληλοι να συμμετάσχουν στην έρευνα , γιατί υποτίθεται ότι το αποτέλεσμα θα είναι πιο ακριβές . Στην έρευνα αγοράς αυτό είναι αδύνατον για διάφορους λόγους που θα αναφερθούν παρακάτω .

Από την άλλη πλευρά όμως ένας ερευνητής μπορεί να μην συγκεντρώνει στοιχεία από το σύνολο του πληθυσμού αλλά να βασιστεί σε ένα δείγμα του πληθυσμού (sample) για να συγκεντρώσει τις απαραίτητες πληροφορίες . Στη συνέχεια , οι πληροφορίες που θα συγκεντρωθούν από το δείγμα χρησιμοποιούνται για να εξαχθούν συμπεράσματα για το σύνολο των ερωτώμενων που αποτελούν το πληθυσμό . Η ικανότητα όμως του ερευνητή να εξάγει συμπεράσματα με βάση το δείγμα για τον πληθυσμό εξαρτάται από τη μέθοδο η οποία χρησιμοποιήθηκε για να επιλεγεί το δείγμα . Οι μέθοδοι της δειγματοληψίας μπορούν να χωριστούν σε δυο μεγάλες κατηγορίες , οι οποίες είναι :

1. **ΔΕΙΓΜΑΤΑ ΠΙΘΑΝΟΤΗΤΩΝ**
2. **ΔΕΙΓΜΑΤΑ ΜΗ ΠΙΘΑΝΟΤΗΤΩΝ**

1 . Στα δείγματα πιθανοτήτων , ο ερευνητής μπορεί να υπολογίσει την πιθανότητα ενός στοιχείου του πληθυσμού να αποτελέσει μέρος του δείγματος . Η πιθανότητα ενός στοιχείου του πληθυσμού να επιλέγει στο δείγμα είναι γνωστή και μη μηδενική . Αυτό σημαίνει ότι στα δείγματα πιθανοτήτων , ο ερευνητής μπορεί να υπολογίσει το μέγεθος του δειγματοληπτικού σφάλματος (sampling error) , κάτι το οποίο δεν είναι δυνατόν με τα δείγματα μη πιθανότητας . Για παράδειγμα , μπορεί να επιλέγει ένα δείγμα πιθανότητας 100 φοιτητών της φιλοσοφικής σχολής του πανεπιστήμιου Πατρών , και οι 100 φοιτητές να είναι όλοι τριτοετής . Αντικειμενικά , αυτό το δείγμα δεν είναι αντιπροσωπευτικό του πληθυσμού της φιλοσοφικής σχολής του πανεπιστήμιου , όσον αφορά τα δημογραφικά στοιχεία και κάποιες άλλες πτυχές . Από το παραπάνω παράδειγμα , μπορούμε να συμπεράνουμε ότι τα δείγματα πιθανότητας έχουν την δυνατότητα να υπολογίσουν το μέγεθος του δειγματοληπτικού σφάλματος . Τρία τέτοια δείγματα πιθανοτήτων είναι : το απλό τυχαίο δείγμα , το στρωματοποιημένο δείγμα και το δείγμα βάσει ομάδων .

2 . Τα δείγματα μη πιθανότητας στηρίζονται στην προσωπική κρίση του ερευνητή , που μερικές φορές μπορεί να αποδειχθεί σωστή και τα αποτελέσματα του δείγματος να αντιπροσωπεύουν σε μεγάλο βαθμό τον πληθυσμό .

Επειδή , η επιλογή των μελών του δείγματος δεν γίνεται με κάποια μηχανιστική διαδικασία , είναι αδύνατο να υπολογιστεί το μέγεθος του στατιστικού σφάλματος .
Ο υπολογισμός του στατιστικού σφάλματος , σε συνδυασμό με την άγνοια του ερευνητή , δεν του επιτρέπει να υπολογίσει με ακρίβεια τα αποτελέσματα που λαμβάνονται από το δείγμα . Τα δείγματα μη πιθανότητας χωρίζονται σε τρία επίπεδα:

1 . Συμβατικό δείγμα: Στο δείγμα αυτό η επιλογή των μελών του δείγματος γίνεται με μόνο κριτήριο την ευκολία . Ο ερευνητής επιλέγει τα πιο εύκολα προσπελάσιμα στοιχεία του πληθυσμού για να συμμετάσχουν στην έρευνα . Παραδείγματα συμβατικών δειγμάτων είναι τα ακόλουθα: η ζήτηση εθελοντών για τη δοκιμή ενός καινούργιου προϊόντος , τα γκάλοπ που διενεργούν συχνά τηλεοπτικοί και ραδιοφωνικοί σταθμοί για θέματα που ενδιαφέρουν τους πολίτες κτλ .
Το μειονέκτημα όμως που παρουσιάζει το συμβατικό δείγμα είναι ο μη υπολογισμός του δειγματοληπτικού σφάλματος , τόσο το συστηματικό όσο και το μεταβλητό . Επικρατεί δε , η αντίληψη ότι όσο πιο μεγάλο είναι το δείγμα της μορφής αυτής τόσο πιο αντιπροσωπευτικό του πληθυσμού θα είναι . Τα συμβατικά δείγματα χρησιμοποιούνται συχνά για εξερευνητικές μελέτες με μικρό κόστος αλλά γρήγορα και εύκολα . Για παράδειγμα , συμβατικό δείγμα μπορεί να χρησιμοποιηθεί για τον προέλεγχο ενός ερωτηματολογίου .

2 . Υποκειμενικό δείγμα: Η επιλογή του δείγματος γίνεται με βάσει την κρίση κάποιου ειδικού ότι συγκεκριμένα στοιχεία θα είναι καλοί πληροφοριοδότες για τους σκοπούς της έρευνας . Η επιλογή των καταστημάτων για την τοποθέτηση των νέων προϊόντων όταν αυτά δοκιμάζονται στην αγορά , η επιλογή των πόλεων για την διενέργεια δοκιμαστικών αγορών (market tests) είναι παραδείγματα υποκειμενικών δειγμάτων . Μια ειδική κατηγορία υποκειμενικού δείγματος αποτελεί το δείγμα χιονοστιβάδας (snowball sample) . Στη μέθοδο αυτή , ο ερευνητής επιλέγει ένα αρχικό δείγμα ερωτώμενων για να συμμετάσχουν στην έρευνα .

Στη συνέχεια , ο ερευνητής ζητάει από κάθε έναν από τους ερωτώμενους αυτούς να υποδείξουν άλλα άτομα που να συγκεντρώνουν τα επιθυμητά χαρακτηριστικά .
Τα υποκειμενικά δείγματα , παρουσιάζουν το πρόβλημα ότι δεν μπορούν να υπολογίσουν το μέγεθος του συστηματικού και τυχαίου σφάλματος . Αν όμως η κρίση του ερευνητή είναι σωστή , τότε τα δείγματα μπορεί να δώσουν πολύ χρήσιμες πληροφορίες , ιδιαίτερα για εξερευνητικές μελέτες .

3 . Δείγμα ποσοστών: Στο δείγμα ποσοστών , σκοπός είναι η επιλογή του δείγματος εκείνου που είναι όμοιο προς τον πληθυσμό με βάση διάφορα προκαθορισμένα χαρακτηριστικά ελέγχου , που είναι συνήθως δημογραφικά . Τα χαρακτηριστικά αυτά αφορούν:

1) Την γνώση των ποσοστών των χαρακτηριστικών ελέγχου στον πληθυσμό .
2) Τα χαρακτηριστικά ελέγχου να είναι περιορισμένα στον αριθμό .
3) Τα χαρακτηριστικά ελέγχου να σχετίζονται άμεσα με τις άλλες μεταβλητές της έρευνας .

Λαμβάνοντας υπόψη τα παραπάνω χαρακτηριστικά για να κάνουμε πιο κατανοητή την μέθοδο του δείγματος ποσοστών θα θέσουμε ένα παράδειγμα .
Ας υποθέσουμε ότι θέλουμε να κάνουμε μια μελέτη για την κατανάλωση γαλακτομικών προϊόντων με μόνα χαρακτηριστικά ελέγχου την ηλικία και το φύλο του ερωτώμενου . Έτσι για κάθε χαρακτηριστικό χρησιμοποιούμε δυο κατηγορίες:

ΗΛΙΚΙΑ	ΦΥΛΟ
Κάτω των 35 ετών	Άρρεν 1
Κάτω των 35 ετών	Θήλυ 2
Άνω των 35 ετών	Θήλυ 3
Άνω των 35 ετών	Άρρεν 4

Αν υποθέσουμε επίσης ότι το μέγεθος του δείγματος είναι 1500 άτομα και ότι η κατανομή των ποσοστών / αναλογιών για κάθε τετράγωνο του παραπάνω πίνακα είναι:

τετράγωνο 1—Άρρεν και Κάτω των 35 ετών :	30%
τετράγωνο 2—Θήλυ και Κάτω των 35 ετών :	20%
τετράγωνο 3—Θήλυ και Άνω των 35 ετών :	30%
τετράγωνο 4—Άρρεν και Άνω των 35 ετών:	20%

Τότε για να υπολογίσουμε τον αριθμό των ατόμων που ανήκουν σε κάθε τετράγωνο θα πρέπει να πολλαπλασιάσουμε το μέγεθος του δείγματος με το ποσοστό του κάθε τετράγωνου .
Έτσι έχουμε:

τετράγωνο 1—1500* 30% =	450 άτομα
τετράγωνο 2—1500* 20% =	300 άτομα
τετράγωνο 3—1500* 30% =	450 άτομα
τετράγωνο 4—1500* 20% =	300 άτομα
Σύνολο =	1500 άτομα

Αφού έχει καθοριστεί ο αριθμός των ατόμων του κάθε τετράγωνου που πρέπει να αποτελέσουν μέρος του δείγματος , στην συνέχεια ο ερευνητής πρέπει να πάρει συνεντεύξεις από όλα τα άτομα του κάθε τετράγωνου χωριστά . Ο τρόπος επιλογής των ατόμων του κάθε τετράγωνου αφήνεται στην προσωπική κρίση του ερευνητή . Το παραπάνω γεγονός έχει ως αποτέλεσμα να επηρεάζεται η αντιπροσωπευτικότητα του δείγματος όσον αφορά τον πληθυσμό . Έτσι , ενώ το δείγμα μπορεί να θεωρηθεί αντιπροσωπευτικό όσον αφορά τα συγκεκριμένα χαρακτηριστικά ελέγχου που χρησιμοποιήθηκαν , για κάποια όμως άλλά , εξίσου σημαντικά χαρακτηριστικά το δείγμα είναι πολύ πιθανόν να μην είναι αντιπροσωπευτικό . Υπάρχει πιθανότητα το πρόβλημα αυτό να λυθεί με τον καθορισμό των ποσοστών όλων των σημαντικών χαρακτηριστικών .

Έτσι όμως , αν χρησιμοποιηθούν πολλά χαρακτηριστικά ελέγχου η μήτρα που θα προκύψει θα είναι πάρα πολύ περίπλοκη . Συμπεραίνοντας , καταλήγουμε ότι το δείγμα ποσοστών είναι κατάλληλο για εξερευνητικές μελέτες .

1.4 ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΚΑΙ ΟΡΙΣΜΟΙ ΤΗΣ ΑΠΟΓΡΑΦΗΣ

Εκτός από την μέθοδο της δειγματοληψίας υπάρχουν και οι εξαντλητικές έρευνες που περιλαμβάνουν την απογραφή .

ΑΠΟΓΡΑΦΗ (census) ονομάζεται η στατιστική έρευνα που γίνεται σε ολόκληρο τον πληθυσμό .
Η απογραφή δηλαδή συνίστανται στην συγκέντρωση πληροφοριών από όλες τις μονάδες ενός πληθυσμού . Σαν παράδειγμα μπορούμε να αναφέρουμε τις απογραφές εμπορίου , βιομηχανίας , γεωργίας , κτηνοτροφίας κ . τ . λ . Ιδιαίτερη σημασία έχει η απογραφή του πληθυσμού μιας χώρας , η οποία αποτελεί την κύρια πηγή πληροφοριών ως προς τα δημογραφικά , οικονομικά και κοινωνικά δεδομένα της χώρας .
Για αυτό η συγκεκριμένη απογραφή αποτελεί σημαντικό γεγονός και γίνεται με μεγάλη κρατική φροντίδα , ώστε τα στοιχεία που θα συγκεντρωθούν να δίνουν την πραγματική εικόνα της χώρας . Στη χώρα μας , απογραφή του πληθυσμού γίνεται κάθε δέκα χρόνια από την Ε . Σ . Υ . Ε . , ενώ απογραφή των βιομηχανικών και εμπορικών επιχειρήσεων κάθε πέντε χρόνια . Απογραφές γεωργίας , κτηνοτροφίας κ . τ . λ δεν έχουν καθορισμένο χρόνο που γίνονται .
Επειδή άλλοτε ο πληθυσμός (με την ευρεία έννοια του όρου) για τον οποίο απαιτούνται πληροφορίες είναι πεπερασμένος (π . χ . οι κάτοικοι μιας πόλης) χρησιμοποιείται ολόκληρος ο πληθυσμός με βάση τα στοιχεία μιας απογραφής .
Η έκταση της άγνοιας γύρω από τον πληθυσμό μπορεί να είναι τόσο μεγάλη ώστε να καθιστά ανεπίτρεπτη οποιαδήποτε συμπερασματολογία στηρίζεται στο δείγμα . Επίσης το κόστος μιας έρευνας (χρηματικά και χρονικά) μπορεί να είναι μεγάλο στην απογραφή .

Η απογραφή είναι δηλαδή μια δειγματοληπτική έρευνα με κάλυψη 100% παρόλο που συνήθως το ενδιαφέρον μας εστιάζεται σε πολύ χαμηλότερα επίπεδα κάλυψης , του ύψους 1% ή 5% .

1.5 ΠΛΕΟΝΕΚΤΗΜΑΤΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ ΕΝΑΝΤΙ ΤΗΣ ΑΠΟΓΡΑΦΗΣ

Οι δειγματοληψίες παρουσιάζουν πλεονεκτήματα έναντι της απογραφής και γενικά προτιμούνται για τους εξής παρακάτω λόγους :

1 . Παρουσιάζουν χαμηλότερο κόστος .

Αυτό συμβαίνει διότι οι πληροφορίες στα δεδομένα προέρχονται από ένα τμήμα του πληθυσμού .

2 . Ο χρόνος δημοσίευσης των πληροφοριών είναι ταχύτερος .

Αντίθετα , η απογραφή είναι μη πρακτική λύση για την εξαγωγή συμπερασμάτων . Αφού η καθυστέρηση τους μπορεί να αποβεί μοιραία . Άρα , ουσιαστικά η εκλογή του ερευνητή είναι μεταξύ του να πάρει έγκαιρα την πληροφορία από ένα κατάλληλο δείγμα ή να μην κάνει τίποτα απολύτως .

3 . Παρουσιάζει μεγαλύτερη ακρίβεια .

Στην περίπτωση του έλεγχου ποιότητας , δεν καταστρέφεται το προϊόν , όπως θα συνέβαινε στην απογραφή . Η διεξαγωγή της έρευνας γίνεται πιο προσεκτικά με αποτέλεσμα η επεξεργασία των αποτελεσμάτων να είναι εφικτή . Έχει διαπιστωθεί από τις δειγματοληπτικές έρευνες ότι όσο μεγαλύτερο είναι το μέγεθος του δείγματος , τόσο πιο αξιόπιστα είναι τα συμπεράσματα που εξάγονται από το δείγμα . Υπάρχουν , βέβαια πιθανότητες , ένα πολύ μεγάλο δείγμα να μην είναι αντιπροσωπευτικό , αν δεν έχει ληφθεί με τον ενδεδειγμένο τρόπο . Συμβάλλει στον περιορισμό της διάρκειας της έρευνας και στην διεύρυνση των δυνατοτήτων της έρευνας .

Παράδειγμα :

Εάν θέλουμε να συζητήσουμε για μερικά επιλεγμένα εκπαιδευτικά θέματα , είναι πολύ δύσκολο να συναντήσουμε όλους τους διευθυντές των σχολικών μονάδων , διότι και ο χρόνος θα είναι πολύς (όπως θα συνέβαινε σε μια απογραφή) .

Για αυτό θα πρέπει να περιοριστούμε σε ένα δείγμα διευθυντών , που θα είναι καταλληλότεροι και πιο σωστά ενημερωμένοι για να απαντήσουν στα συγκεκριμένα εκπαιδευτικά θέματα . Δηλαδή , το δείγμα αυτό των διευθυντών θα έχουν συγκεκριμένες και πλήρεις γνώσεις πάνω στα εκπαιδευτικά θέματα . Με αυτό τον τρόπο μπορεί να διαπιστωθεί ότι τα αποτελέσματα της δειγματοληψίας είναι ακριβέστερα από εκείνα της απογραφής , αφού κατά την απογραφή υπάρχει η πιθανότητα να απαντήσουν άτομα που δεν κατέχουν ακριβώς το θέμα , όπως θα έπρεπε , αλλά έχουν γενικές και μεμονωμένες γνώσεις για τα συγκεκριμένα εκπαιδευτικά θέματα.

Επίσης επιχειρηματίες και βιομήχανοι χρησιμοποιούν δειγματοληπτικές μεθόδους για την αξιολόγηση της επάρκειας των εσωτερικών λειτουργιών των επιχειρήσεων ή βιομηχανιών , για την εκτίμηση των αποθεμάτων , την μελέτη της κατάστασης και της διάρκειας ζωής του εξοπλισμού τους , του ρυθμού παραγωγής εργασίας , για τον έλεγχο ποιότητας των προϊόντων και την μελέτη της κατανομής του χρόνου εργασίας των υπάλληλων τους στα διάφορα καθήκοντα τους .

Τέλος έρευνες κοινής γνώμης πάνω σε τρέχοντα θέματα και έρευνας αγοράς γίνονται πολύ συχνά .

1.6 ΜΕΙΟΝΕΚΤΗΜΑΤΑ ΤΗΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ ΕΝΑΝΤΙ ΤΗΣ ΑΠΟΓΡΑΦΗΣ

Οι δειγματοληψίες παρουσιάζουν μερικές αδυναμίες σε σχέση με τις απογραφές :

1 . Ένα δείγμα παρουσιάζει <u>αδυναμία</u> να συμπεριλάβει με απόλυτη ακρίβεια τα χαρακτηριστικά του πληθυσμού από το οποίο προέρχεται . Αυτό είναι εφικτό , επειδή τα χαρακτηριστικά των ατόμων ενός πληθυσμού διαφέρουν μεταξύ τους .

Η διαφορά αυτή μεταφέρεται και στα δείγματα με αποτέλεσμα τα αποτελέσματα να μην είναι αξιόπιστα .

2 . Ο τρόπος με τον οποίο ορίζεται ένας πληθυσμός μπορεί να είναι λανθασμένος . Παράδειγμα : χρησιμοποιούμε τον τηλεφωνικό κατάλογο για να ερευνήσουμε τα άτομα καθώς και την μόρφωση των ατόμων ενός δήμου .

Ο τρόπος αυτός είναι λανθασμένος .

Το δείγμα που επιλέγουμε , θα περιλάμβανε στοιχεία που δεν ανταποκρίνονται στα χαρακτηριστικά του πληθυσμού που μας ενδιαφέρουν . Διότι είναι δύσκολο να ερευνήσουμε τη μόρφωση των ατόμων του δήμου μέσω τηλεφωνικού καταλόγου .

1.7 ΤΡΟΠΟΙ ΜΕΙΩΣΗΣ ΕΛΑΤΤΩΜΑΤΩΝ ΤΗΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Γενικά για την μείωση των ελαττωμάτων της μεθόδου της δειγματοληψίας θα πρέπει να ακολουθούμε μερικά βήματα :

Α) Ο πληθυσμός πρέπει να οριστεί προηγουμένως με ακρίβεια.

Β) Το αντικείμενο της έρευνας να έχει προσδιοριστεί για την αξιοπιστία των αποτελεσμάτων .

Γ) Η κλίμακα μέτρησης και ο αριθμός των ατόμων που θα χρειαστούν να έχουν προηγουμένως καθοριστεί .

Λαμβάνοντας υπόψη τα πλεονεκτήματα και τα μειονεκτήματα της δειγματοληψίας , το φυσικό ερώτημα που ανακύπτει είναι πως επιλέγεται ένα δείγμα;

Υπάρχουν 4 μέθοδοι που είναι κατάλληλοι για την επιλογή ενός δείγματος και οι οποίοι είναι :

1. ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΕΥΚΟΛΙΑΣ (ΠΡΟΣΙΤΟΤΗΤΑΣ) .

Στην περίπτωση αυτή , επιλέγουμε το δείγμα από ένα τμήμα του πληθυσμού , στο οποίο υπάρχει εύκολη πρόσβαση . Είναι φανερό ότι ένα τέτοιο δείγμα δεν μπορεί να είναι αντιπροσωπευτικό του πληθυσμού από το οποίο προέρχεται . Ως παραδείγματα μπορούμε να αναφέρουμε την σφυγμομέτρηση της κοινής γνώμης αναζητώντας απαντήσεις εθελοντών στο ερωτηματολόγιο , τους αναγνώστες ενός περιοδικού που καλούνται να συμπληρώσουν και να επιστρέψουν ένα ερωτηματολόγιο που επισυνάπτεται στο περιοδικό , το ερωτηματολόγιο που εστάλη από κάποιο Υπουργείο το έτος 1999 κ . τ . λ . Παρατηρούμε ότι ένα τέτοιο δείγμα όσον αφορά την χρησιμοποίηση τους για τη εξαγωγή συμπερασμάτων για ολόκληρο τον πληθυσμό δεν μπορεί να είναι αντιπροσωπευτικό .

2. ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΚΡΙΣΗΣ Ή ΣΚΟΠΙΜΟΤΗΤΑΣ

Στην περίπτωση αυτή η επιλογή του δείγματος από έναν πληθυσμό είναι η εσκεμμένη υποκειμενική επιλογή . Ο ερευνητής δηλαδή επιλέγει από τον ομοιογενή πληθυσμό ένα δείγμα τυπικών μονάδων που κατά την αντίληψη του , πλησιάζουν την μέση τιμή του πληθυσμού.Τα αποτελέσματα αυτής της μορφής της δειγματοληψίας μπορεί να αποβούν σωστά , εάν η κρίση του είναι ακριβής . Το σχήμα αυτό το προτιμούν πολύ .

3 . ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΜΕ ΠΡΟΚΑΘΟΡΙΣΜΕΝΑ ΠΟΣΟΣΤΑ (QUOTA SAMPLING)

Η μέθοδος αυτή συνδυάζει την υποκειμενική κρίση ενός ερευνητή με την προσιτότητα των μονάδων του πληθυσμού .

Παράδειγμα : Ρωτώντας τους περαστικούς ενός δρόμου για την άποψη τους όσον αφορά ένα προϊόν διαφορετικών ηλικιών , φύλου, επαγγέλματος κ . α . παρατηρούμε ότι το δείγμα μπορεί να είναι αντιπροσωπευτικό,όμως δεν επιδέχεται ανάπτυξη δειγματοληπτικής θεωρίας . Αυτό οφείλεται στο γεγονός της έλλειψης της τυχαιότητας κατά την επιλογή των μονάδων που είναι σημαντικό στοιχείο της δειγματοληψίας για να υπάρχει μέτρο σύγκρισης της αντιπροσωπευτικότητας ή της καταλληλότητας των εκτιμητριών ενός δείγματος .

Εμπλέκει επίσης , ένα στοιχείο προσιτότητας που για να το κάνουμε πιο κατανοητό θα θέσουμε:

Επιλέγονται "οι περαστικοί που κρίνονται ότι ανταποκρίνονται περισσότερο από τους υπόλοιπους στα παραπάνω κριτήρια " .

Επομένως , είναι σημαντικό να εισαχθεί ένα στοιχείο τυχαιότητας στην διαδικασία επιλογής δείγματος , ώστε το δείγμα να επιλέγεται με κάποιο μηχανισμό πιθανότητας .

Με αυτό τον τρόπο είναι ευκολότερο να ελεγχθεί η αντιπροσωπευτικότητα και η ακρίβεια μιας δειγματοληπτικής έρευνας και των προκυπτουσών εκτιμητριών . Η εισαγωγή του μηχανισμού πιθανότητας μπορεί να επιτύχει ως εξής:
Έστω ν το μέγεθος του δείγματος που επιθυμούμε να επιλέξουμε . Θεωρούμε στην συνέχεια , όλα τα δυνατά δείγματα μεγέθους ν που θα μπορούσαν να επιλεγούν από τον συγκεκριμένο πληθυσμό . Έστω , λοιπόν S_1 ,S_2 ,..... τα δείγματα που σημαίνει ότι κάθε Si συμβολίζει ένα διακεκριμένο δείγμα μεγέθους ν από ολόκληρο τον πληθυσμό .

4 . ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΚΑΤΑ ΠΙΘΑΝΟΤΗΤΑ (PROBABILITY SAMPLING)

Η μέθοδος αυτή βασίζεται στην αντιστοίχηση μιας πιθανότητας π_1 στο δείγμα s_1 και στην επιλογή ενός δείγματος S_1 από το σύνολο των δειγμάτων S σύμφωνα με τον μηχανισμό πιθανότητας που εξασφαλίζεται στο δείγμα s_1 ,πιθανότητα επιλογής ίση με π_1 (i=1 ,2 ,3 ,...) . Είναι λοιπόν , προφανές ότι υπάρχει μια απειρία διαφορετικών κατά πιθανότητα δειγματοληπτικών σχημάτων αντιστοιχούντων στις διαφορετικές κατανομές πιθανότητας π = { π_1 ,π_2 ,... . } που μπορούν να ορισθούν πάνω στο σύνολο των δυνατών δειγμάτων S_1 ,S_2 ,...... . Τα σχήματα αυτά μπορούν να είναι παρά πολύ απλά έως παρά πολύ επιτηδευμένα και σύνθετα .

ΚΕΦΑΛΑΙΟ 2

ΣΠΟΥΔΑΙΟΤΗΤΑ ΚΑΙ ΚΛΑΔΟΙ ΤΗΣ ΣΤΑΤΙΣΤΙΚΗΣ

2 . 1 ΠΛΗΘΥΣΜΟΙ ΚΑΙ ΔΕΙΓΜΑΤΑ

Ο όρος **πληθυσμός (population)** σε μια στατιστική μελέτη αναφέρεται στο σύνολο των ανθρώπων , πραγμάτων ή μονάδων που ανήκουν στην ομάδα που μας ενδιαφέρει , σε αντίθεση με το δείγμα (sample) που είναι μια ομάδα στοιχείων που έχουν επιλέγει από τον πληθυσμό . Τα στοιχεία του συνόλου αυτού τα λέμε άτομα ή στοιχεία ή στατιστικές μονάδες .
Ο πληθυσμός μπορεί να αναγραφεί και ως στατιστικός πληθυσμός και να συμβολιστεί με Ω . Ο πληθυσμός , στα διάφορα φαινόμενα που μελετάμε μπορεί να είναι πεπερασμένος ή άπειρος . Τα άτομα του πληθυσμού (ο οποίος δεν είναι απαραίτητα ανθρώπινος)έχουν ορισμένα μέτρα των οποίων η εκτίμηση παρουσιάζει ενδιαφέρον .
Συχνά ενδιαφερόμαστε να μελετήσουμε ορισμένα χαρακτηριστικά του πληθυσμού ως προς κάποιο μέτρο , όπως ο μέσος της διάρκειας ζωής ενός τύπου ηλεκτρικού λαμπτήρα , το συνολικό ύψος της παραγωγής λαδιού σε μια περιοχή ή το ποσοστό των οικογενειών με εισοδήματα που υπερβαίνουν τα έξι εκατομμύρια δραχμές κατά την διάρκεια του 1993 .
Κατά την διεξαγωγή μιας δειγματοληπτικής έρευνας, συχνά, δημιουργούνται ασάφειες όσο αφορά την ερμηνεία ορισμένων βασικών εννοιών . Πιο συγκεκριμένα , υπάρχει κάποια ασυνέπεια στον τρόπο με τον οποίο χρησιμοποιούνται οι διαφορετικές έννοιες.
Για τον λόγο αυτό , παραθέτουμε στη συνέχεια μερικές από τις συνηθέστερες βασικές έννοιες:

ΑΝΤΙΚΕΙΜΕΝΙΚΟΣ ΠΛΗΘΥΣΜΟΣ (target population)
Είναι το σύνολο του πεπερασμένου πληθυσμού για τον οποίο ενδιαφερόμαστε να συλλέξουμε πληροφορίες . Για παράδειγμα , ο αντικειμενικός πληθυσμός μιας μελέτης , θα μπορούσε να είναι το σύνολο των εργαζόμενων στην Ελλάδα που είναι κάτω των 30 ετών.

<u>ΥΠΟ ΜΕΛΕΤΗ ΠΛΗΘΥΣΜΟΣ</u> (study population):

Είναι το βασικό πεπερασμένο σύνολο ατόμων που θέλουμε να μελετήσουμε . Για **<u>παράδειγμα</u>** , το σύνολο των εργαζόμενων στην Ελλάδα ηλικίας κάτω των 30 ετών , των οποίων η μόνιμη κατοικία είναι μέσα στην περιοχή του λεκανοπέδιου Αττικής .

Ο πληθυσμός αυτός ενδέχεται να ταυτίζεται με τον αντικειμενικό πληθυσμό . Συχνά όμως ενδέχεται να είναι ένας πληθυσμός περισσότερο περιορισμένος σε μέγεθος , αλλά και περισσότερο εύκολο να μελετηθεί , του οποίου οι ιδιότητες ελπίζεται ότι μπορούν να προβληθούν στο ευρύτερο αντικειμενικό πληθυσμό .

Πληθυσμοί με άπειρο πλήθος στοιχείων δεν εμφανίζονται συχνά στην καθημερινή μας ζωή .

Για παράδειγμα , κατά την εξέταση ποιότητας των προϊόντων που παράγονται σε μια βιομηχανική διαδικασία μπορούμε να θεωρούμε ότι ο πληθυσμός μας αποτελείται από μια άπειρη ακολουθία στοιχείων. Επίσης χρονικά διαστήματα μιας ώρας που συνεχίζονται έπ' αόριστον με στατιστική μονάδα κάθε συγκεκριμένο διάστημα μιας ώρας και μεταβλητή Χ τον αριθμό των τηλεφωνικών κλήσεων που λαμβάνονται σε ένα τηλεφωνικό κέντρο κάθε ώρα . Για να πάρουμε μια απόφαση για έναν πληθυσμό (π.χ οι άνθρωποι που ερωτήθηκαν κατά την διάρκεια μιας δημοσκόπησης για την πρόθεση ψήφου) εξετάζουμε συνήθως ένα δείγμα .

Εάν τύχει το δείγμα που θα επιλέξουμε να μην είναι αντιπροσωπευτικό του πληθυσμού από το οποίο προέρχεται , τότε οι προβλέψεις που θα κάνουμε θα είναι ανακριβείς .

Μελετώντας ολόκληρο το πληθυσμό , αποφεύγουμε μεν αυτό τον κίνδυνο αλλά αντιμετωπίζουμε άλλα προβλήματα , όπως το τεράστιο κόστος και την δυσκολία μελέτης του πληθυσμού .

<u>Παράδειγμα</u> : Είναι μεγάλο κόστος για μια κυβέρνηση να υπολογίσει το μηνιαίο δείκτη ανεργίας μετρώντας πραγματικά το συνολικό αριθμό των άνεργων . Έτσι , αντί για αυτό , ο δείκτης ανεργίας υπολογίζεται κάθε μήνα από την στατιστική υπηρεσία κάθε κράτους με βάση ένα δείγμα του πληθυσμού .

Μερικές φορές λοιπόν , μπορούμε να πετύχουμε μεγαλύτερη ακρίβεια ασχολούμενοι με τα δεδομένα ενός δείγματος παρά με ολόκληρο τον πληθυσμό .

Κύριος στόχος λοιπόν είναι η λήψη ενός δείγματος το οποίο να είναι αντιπροσωπευτικό του πληθυσμού και το οποίο να οδηγεί σε εκτιμήσεις των χαρακτηριστικών του πληθυσμού με όσο το δυνατόν μεγαλύτερη ακρίβεια.

2.2 ΜΕΤΑΒΛΗΤΕΣ ΤΟΥ ΠΛΗΘΥΣΜΟΥ

Ο πληθυσμός αποτελείται από μεταβλητές που συμβολίζονται με (Χ , Ψ , Ζ) και χωρίζονται σε δυο ομάδες .

1) ΣΤΙΣ ΠΟΙΟΤΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ (QUANTITATIVE VARIABLE) που δεν επιδέχονται μέτρηση και οι τιμές τους δεν εκφράζονται με αριθμούς αλλά με λέξεις , όπως το χρώμα ενός αυτοκίνητου , η ένδειξη ενός νομίσματος , το είδος ενός βιβλίου κ . λ . π. Είναι δηλαδή οι τιμές που παίρνει και που είναι κατηγορίες μιας ταξινόμησης και δεν προκύπτουν από αριθμητικές μετρήσεις .

2) ΣΤΙΣ ΠΟΣΟΤΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ (QUANTITIES VARIABLE) που επιδέχονται μέτρηση και οι τιμές τους είναι πραγματικοί αριθμοί , όπως το βάρος ενός μαθητή , η θερμοκρασία , η ηλικία κ.τ.λ. Είναι δηλαδή το σύνολο των τιμών εκείνων που είναι υποσύνολο των πραγματικών αριθμών .

Οι ποσοτικές μεταβλητές χωρίζονται σε δυο άλλες κατηγορίες οι οποίες είναι:

Α) Ασυνεχείς μεταβλητές (discrete variable) που είναι οι μεταβλητές εκείνες που παίρνουν μόνο μεμονωμένες τιμές όπως ο αριθμός παιδιών σε μια οικογένεια .

Β) Συνεχείς μεταβλητές (continuous variable) που είναι οι μεταβλητές εκείνες οι οποίες μπορεί να παίρνουν οποιαδήποτε τιμή μέσα σε ένα διάστημα όπως το βάρος , ύψος , ηλικία κ.τ.λ .

Σε ότι αφορά τον τρόπο συλλογής των δεδομένων υπάρχουν διάφοροι μέθοδοι που αντιστοιχούν στους τρόπους προσέγγισης των στατιστικών μονάδων και της μέτρησης των μεταβλητών που μας ενδιαφέρουν . Μερικές από τις πιο σημαντικές μεθόδους είναι: η προσωπική συνέντευξη (personal interview) ,η τηλεφωνική συνέντευξη (telephone interview) , η δια αλληλογραφίας συλλογή στοιχείων , η παρατήρηση των ατόμων και η καταγραφή των παρατηρήσεων . Φυσικά , είναι σημαντικό και η επιλογή των ατόμων που συλλέγουν τα δεδομένα για την ορθή συμπλήρωση των ερωτηματολογίων διότι επηρεάζουν κατά ένα πολύ σημαντικό βαθμό το αποτέλεσμα μιας έρευνας.

2.3 ΔΕΙΓΜΑΤΟΛΗΠΤΙΚΕΣ ΜΟΝΑΔΕΣ (SAMPLING UNITS)

Δειγματοληπτική μονάδα ονομάζεται κάθε παρατήρηση ή στοιχείο του πληθυσμού . Τις δειγματοληπτικές μονάδες τις επιλέγουμε με σκοπό να επιτύχουμε πρόσβαση στα άτομα (μονάδες) του υπό μελέτη πληθυσμού . Για να είναι πιο σαφές θα θέσουμε ένα παράδειγμα .

Έστω, ότι επιθυμούμε να κάνουμε μια έρευνα οικογενειακών προϋπολογισμών σε μια πόλη . Είναι εύκολο να υποθέσουμε ότι στην περίπτωση αυτή τα άτομα (οι μονάδες) του υπό μελέτη πληθυσμού είναι οι οικογένειες .

Παρόλα αυτά όμως , πριν προχωρήσουμε στην έρευνα πρέπει να ορίσουμε κάποιο συμβατικό ορισμό για τις οικογένειες . Για αυτό , μπορούμε να λάβουμε ένα δείγμα διευθύνσεων και να αναζητήσουμε πληροφορίες για τις οικογένειες που κατοικούν στις επιλεγείσες διευθύνσεις . Έτσι οι διευθύνσεις αποτελούν τις δειγματοληπτικές μονάδες παρά το γεγονός ότι ο πληθυσμός των διευθύνσεων είναι ουσιώδης σημασίας .

Επίσης ας υποθέσουμε ότι θέλουμε να βρούμε όλα τα άτομα / μονάδες που το επώνυμο τους είναι " Παπαδόπουλος " όσον αφορά τον νόμο Αττικής .

Στην περίπτωση αυτή , αυτό που πρέπει να χρησιμοποιήσουμε είναι ένα τηλεφωνικό κατάλογο που περιλαμβάνει όλα τα ονόματα , τηλέφωνα και διευθύνσεις των ατόμων αυτών .

Έτσι , στην περίπτωση αυτή δειγματοληπτική μονάδα αποτελεί ο τηλεφωνικός κατάλογος .

Πολλές φορές , η δειγματοληπτική μονάδα είναι τεχνητή , με την έννοια ότι καθορίζεται αποκλειστικά για τους σκοπούς της δειγματοληψίας. Παραδείγματος χάρη , προκειμένου να σχηματισθεί δείγμα από την έκταση της ορισμένης περιοχής , υποδιαιρείται η επιφάνεια του χάρτη της σε μικρά τετραγωνίδια . Έτσι , η δειγματοληπτική μονάδα είναι το τετραγωνίδιο .

Η δειγματοληπτική μονάδα ενδέχεται να μην είναι σταθερή από άποψη μεγέθους ή σύνθεσής της (από άλλες μονάδες) .

Παράδειγμα:

Ένα αγρόκτημα δεν έχει την ίδια έκταση για όλους τους κτηματίες . Επίσης ενώ ένας άνεργος ως δειγματοληπτική μονάδα αποτελείται πάντα από ένα φυσικό πρόσωπο , η οικογένεια ως δειγματοληπτική μονάδα αποτελείται από κυμαινόμενο αριθμό φυσικών προσώπων (1 . . 2 . . 3 . . 4... . . πρόσωπα) .

Γενικά , κάθε δειγματοληπτική μονάδα πρέπει να προσδιοριστεί με σαφήνεια , γιατί είναι πολύ σημαντική για τον τρόπο επιλογής ενός δείγματος .

2.4 ΔΕΙΓΜΑΤΟΛΗΠΤΙΚΟ ΠΛΑΙΣΙΟ (SAMPLING FRAME)

Δειγματοληπτικό πλαίσιο ονομάζεται το σύνολο των δειγματοληπτικών μονάδων που αντιστοιχούν στο υπό εξεταζόμενο πληθυσμό . Το πλαίσιο αυτό είναι συνήθως μια ονομαστική κατάσταση (ή κατάλογος ή πίνακας) όπου αναγράφονται συστηματικά όλες οι δειγματοληπτικές μονάδες που αντιστοιχούν στον πληθυσμό . Αντί καταλόγου , είναι δυνατό να χρησιμοποιηθεί ως πλαίσιο δειγματοληψίας το σχέδιο ορισμένης πόλης με δειγματοληπτική μονάδα την οδό ή το οικοδομικό τετράγωνο. Επίσης, είναι δυνατόν να χρησιμοποιηθεί σχεδιάγραμμα αγροκτημάτων ορισμένης περιοχής με δειγματοληπτική μονάδα το αγρόκτημα .

Γνωρίζουμε ότι βασικό έργο σε μια έρευνα είναι ο έλεγχος της καταλληλότητας του δειγματοληπτικού πλαισίου .

Παράδειγμα :

Ο κατάλογος των προϊόντων ενός συγκεκριμένου super –market δεν μπορεί να αποτελεί δειγματοληπτικό πλαίσιο του πληθυσμού των προϊόντων που πουλάνε όλα τα super- market της Αθήνας . Επίσης η ενημέρωση ενός συγκεκριμένου τοπικού καναλιού δεν μπορεί να είναι τόσο πλήρης , όσο η ενημέρωση καναλιού γενικού δικτύου .

Το δειγματοληπτικό πλαίσιο λοιπόν, είναι απαραίτητο για την σωστή επιλογή των μονάδων του δείγματος και για τον έλεγχο της καταλληλότητας ή μη του δείγματος , ώστε τυχόν σφάλματα (errors) να μπορούν να αποφευχθούν . Μερικές φορές , οι δειγματοληπτικές μονάδες ταυτίζονται με τα μέλη του υπό μελέτη πληθυσμού .

Στις περισσότερες φορές όμως , αυτό δεν συμβαίνει και έτσι το δειγματοληπτικό πλαίσιο είναι μια λιγότερο ακριβής υποδιαίρεση του υπό μελέτη πληθυσμού , κάθε μονάδα της οποίας περιέχει ένα συγκεκριμένο σύνολο μελών του πληθυσμού.

2.5 ΚΑΤΑΛΟΓΟΣ Ή ΛΙΣΤΑ (LIST)

Για να χρησιμοποιηθεί το δειγματοληπτικό πλαίσιο ως η πρωταρχική πηγή από την οποία θα επιλέγει το δείγμα , θα πρέπει να είναι δυνατός ο προσδιορισμός των δειγματοληπτικών μονάδων. Άλλωστε το δειγματοληπτικό πλαίσιο επιλέγεται με αυτό το σκοπό .

Παράδειγμα: Ένας πραγματικός κατάλογος όλων των δειγματοληπτικών μονάδων ενδέχεται να μην υπάρχει , όπως ο κατάλογος των διευθύνσεων μιας πόλης ή ο κατάλογος που παρέχεται από ιατρικά αρχεία για όλους τους ηλικιωμένους ασθενείς οι οποίοι επισκέπτονται τους γιατρούς τους σε μια συγκεκριμένη περιοχή και για μια ορισμένη χρονική περίοδο .

Ένας τέτοιος κατάλογος διευκολύνει ιδιαίτερα την επιλογή ενός δείγματος .

2.6 ΠΛΑΙΣΙΑ ΑΚΑΤΑΛΛΗΛΑ ΓΙΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

Υπάρχει μια κατηγορία πλαισίων που δεν θεωρούνται κατάλληλα για δειγματοληψία διότι μπορεί να είναι ανακριβή , μη πλήρη , μη ενημερωμένα , με απροσπέλαστες μονάδες ή με διπλές ή πολλαπλές έγγραφες .

1) **Πλαίσια ανακριβή**:

Όταν ορισμένες μονάδες δεν περιέχονται στο πλαίσιο σημαίνει ότι δεν ικανοποιούν το χρησιμοποιούμενο ορισμό της δειγματοληπτικής μονάδας , με αποτέλεσμα να είναι ανακριβή . **Παράδειγμα**: Ανακριβές πλαίσιο είναι η ονομαστική κατάσταση επιδοτούμενων ανέργων που περιέχει ονοματεπώνυμα προσώπων , τα οποία δεν είναι άνεργοι (είναι πρόσωπα που έχουν συνταξιοδοτηθεί ή που εργάζονται) αλλά παραμένουν στον κατάλογο από λάθος ή για να εισπράττουν το επίδομα ανεργίας.

Αν με βάση τα στοιχεία του ανακριβούς πλαισίου γίνεται στρωματοποίηση του πληθυσμού , η ακατάλληλη σύνθεση του οδηγεί σε λανθασμένη στρωματοποίηση και συνεπώς σε μείωση της ακρίβειας των αποτελεσμάτων της δειγματοληψίας.

2) **Πλαίσια μη πλήρη:**

Η έλλειψη της πληρότητας είναι δύσκολο να διαπιστωθεί είτε κατά την εξέταση του περιεχομένου του ίδιου του πλαισίου , είτε κατά την διεξαγωγή της συνέντευξης .

Παράδειγμα:

Ο κατάλογος που αποτελείται από τις επιχειρήσεις των αρχείων της Νομισματικής Επιτροπής δεν αποτελεί πλήρες πλαίσιο επιχειρήσεων , διότι δεν καλύπτει τις αυτοχρηματοδοτούμενες επιχειρήσεις ούτε όσες συνάπτουν τραπεζικά δάνεια κάτω από ορισμένο όριο . Έτσι ένα τέτοιο πλαίσιο δίνει δείγματα μη αντιπροσωπευτικά του πληθυσμού .

3) **<u>Πλαίσια μη ενημερωμένα</u>**:

Πολλά πλαίσια περιέχουν μονάδες που δεν ανήκουν στο πληθυσμό που μας ενδιαφέρει και παραλείπουν άλλες μονάδες που κανονικά ανήκουν στον πληθυσμό αυτό .

Το αρχείο της ιδιωτικής χρήσης επιβατικών αυτοκινήτων , περιέχει και αυτοκίνητα που έχουν αχρηστευθεί ή καταστραφεί , ενώ δεν περιέχει όσα κυκλοφορούν λαθραία ή όσα κυκλοφορούν με ''ελεύθερη χρήση'' .

4) **<u>Πλαίσια με απροσπέλαστες μονάδες</u>**

Ο κατάλογος με τους πτυχιούχους μιας Ανώτατης σχολής που δεν περιέχει τις ταχυδρομικές διευθύνσεις τους (ή το τηλέφωνό τους) , ακόμα και αν είναι ενημερωμένος , δεν είναι δυνατόν να χρησιμοποιηθεί ως πλαίσιο δειγματοληψίας των πτυχιούχων αυτών, διότι ο συνεντευκτής δεν θα κατορθώσει να εντοπίσει τις μονάδες του δείγματος .

5) **<u>Πλαίσια με διπλές και πολλαπλές έγγραφες</u>**

Αν μια δειγματοληπτική μονάδα παρουσιάζεται δυο ή περισσότερες φορές στο πλαίσιο , τότε το πλαίσιο αυτό θα είναι ακατάλληλο .

ΚΕΦΑΛΑΙΟ 3

ΣΦΑΛΜΑΤΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

3.1 ΔΙΕΞΑΓΩΓΗ ΔΕΙΓΜΑΤΟΛΗΠΤΙΚΩΝ ΣΦΑΛΜΑΤΩΝ

Τα αποτελέσματα μιας δειγματοληπτικής έρευνας δεν μπορεί να είναι πάντοτε απολύτως σωστά . Δηλαδή , οι εκτιμήσεις για τις τιμές των παραμέτρων (μέση τιμή , τυπική απόκλιση κ.τ.λ.) ενός πληθυσμού ή οι αποφάσεις σχετικά με κάποιες υποθέσεις που αφορούν τις τιμές αυτών των παραμέτρων δεν ανταποκρίνονται στην πραγματικότητα. Αυτό είναι φυσικό
, γιατί όσο αντιπροσωπευτικό και να είναι ένα δείγμα , δεν μπορεί να συμπεριλάβει με απόλυτη ακρίβεια την μεταβλητότητα των τιμών του . **Έτσι , γενικά , σφάλματα (errors) θα ονομάσουμε τις αποκλίσεις των δειγματοληπτικών αποτελεσμάτων από τις πραγματικές τιμές** .
Στόχος λοιπόν κάθε δειγματοληπτικής έρευνας είναι να μειωθεί όσο το δυνατόν η συμμετοχή αυτών των σφαλμάτων στην διαμόρφωση των τελικών αποτελεσμάτων . Για αυτό , πριν αποφασίσει για την διεξαγωγή μιας δειγματοληπτικής έρευνας ή σφυγμομέτρησης κάποιας δεδομένης μορφής και μεγέθους πάνω σε ένα συγκεκριμένο θέμα , ένας ερευνητής θα πρέπει να μεριμνήσει ώστε ο σχεδιασμός και η διεξαγωγή της έρευνας να ανταποκρίνεται στα παρακάτω σημεία :

- Πως θα τεθούν τα ερωτήματα με τρόπο σαφή και όχι παραπλανητικό .
- Με ποιο τρόπο θα παρακινηθούν οι ερωτώμενοι για να απαντήσουν στις ερωτήσεις .
- Πως θα επιτευχθεί η επαφή με τα άτομα ως προς τα οποία σχεδιάζεται η έρευνα .
- Με ποιο τρόπο θα εξασφαλισθεί ότι ομάδες ατόμων με ειδικά συμφέροντα δεν θα αρνηθούν να απαντήσουν .
- Πως θα αποφασισθεί τι μπορεί να είναι κατάλληλο και σωστό να ερωτηθεί .

- Πως θα αξιολογηθεί το εάν θα επιτύχουμε τους στατιστικούς στόχους της έρευνας όσο αφορά την αμεροληψία και την ακρίβεια , αλλά και το κόστος.

Τα ερωτήματα αυτά στην πραγματικότητα επηρεάζουν τις αποφάσεις μας όσον αφορά την δειγματοληπτική έρευνα , το δειγματοληπτικό σχήμα και το μέγεθος του δείγματος που απαιτείται . Για αυτό είναι σημαντικό , πριν ξεκινήσει η διεξαγωγή της δειγματοληπτικής έρευνας να εξετασθεί εάν είναι δυνατή η συλλογή δεδομένων με τον τρόπο που καθορίζει το δειγματοληπτικό σχήμα και ποιες δυσκολίες ενδέχεται να προκύψουν κατά την πρόοδο της διαδικασίας .

Με αυτή την προκαταρτική εξέταση (pre-survey examination) επιτυγχάνεται η ανατροφοδότηση μας με πληροφορίες από τις οποίες μπορούμε να αναμορφώσουμε τον σχεδιασμό της έρευνας ώστε να είναι περισσότερο πραγματοποιήσιμη . Ο σχεδιασμός των διάφορων δειγματοληπτικών σχημάτων έχει ως στόχο την λήψη ενός δείγματος , το οποίο να είναι ικανοποιητικά αντιπροσωπευτικό του πληθυσμού .

Το δείγμα δηλαδή , πρέπει να αντανακλά τη φυσική μεταβλητότητα στο σύνολο των τιμών , π.χ της μεταβλητής Y . Έτσι η μεταβλητότητα αυτή ονομάζεται **δειγματοληπτική μεταβλητότητα** (sampling variation) ή **δειγματοληπτικό σφάλμα** (sampling error) .

Παράδειγμα:Διαφορετικά άτομα έχουν διαφορετικά εισοδήματα , διαφορετικές επιχειρήσεις έχουν διαφορετικές ετήσιες πωλήσεις κ.τ.λ. Παρατηρούμε λοιπόν , ότι το δειγματοληπτικό σφάλμα επηρεάζει την ακρίβεια και μπορεί να εκτιμηθεί με βάση το δείγμα . Η διαδικασία αυτή απαιτεί την γνώση της τιμής της διασποράς του πληθυσμού , η οποία μπορεί να εκτιμηθεί με την λήψη του δείγματος με διάφορες μεθόδους όπως , πιλοτικές μελέτες (pilot studies) , μέθοδοι jackknife (jackknife methods) και μέθοδος αλληλοεισδυόντων υπό-δειγμάτων (interpenetrating sub samples method) .

1 . ΠΙΛΟΤΙΚΕΣ ΜΕΛΕΤΕΣ:

Το δείγμα που θα χρησιμοποιηθεί εδώ για την διεξαγωγή της έρευνας σχηματίζεται σε δυο φάσεις από δυο υπό-δείγματα . (διπλή δειγματοληψία ή δειγματοληψία σε δυο φάσεις) . Το πρώτο χρησιμοποιείται για την εκτίμηση της διασποράς του πληθυσμού και για τον προσδιορισμό του κατάλληλου δειγματικού μεγέθους . Μια ενδιαφέρουσα επέκταση της διπλής δειγματοληψίας είναι αυτή της επαναλαμβανόμενης δειγματοληψίας (replicated sampling) στην οποία ένα δείγμα μεγέθους ν = μ αποτελείται από κ ανεξάρτητα υπό-δείγματα του ίδιου μεγέθους μ , τα οποία λαμβάνονται σύμφωνα με το δειγματοληπτικό σχήμα της συνολικής έρευνας .

Για παράδειγμα :Έστω ότι θέλουμε να εκτιμήσουμε την μέση τιμή μ με τον μέσο ενός δείγματος . Για αυτό χρειαζόμαστε να προσδιορίσουμε ή να εκτιμήσουμε πρώτα την αναμενόμενη τιμή και την διασπορά της εκτιμήτριας $\overline{\Theta}$ για να αξιολογήσουμε πόσο καλή είναι ως εκτιμήτρια της παραμέτρου θ . Έτσι έχουμε:

$$E(\overline{X}_n) = \mu$$

και

$$V(\overline{X}_n) = \frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)$$

Οπότε αυτό που χρειαζόμαστε μόνο είναι μια εκτίμηση της διασποράς σ^2 του πληθυσμού . Για περισσότερα σύνθετα δειγματοληπτικά σχήματα και περίπλοκες εκτιμήτριες , ενδέχεται να μη γνωρίζουμε τη θεωρητική μορφή της μέσης τιμής της εκτιμήτριας ή της διασποράς της .

Για να προσπεράσουμε όμως την δυσκολία αυτή θα παραθέσουμε ένα παράδειγμα . Έστω ότι κάθε υπόδειγμα οδηγεί σε μια εκτίμηση $\widetilde{\Theta};(i = 1,2,3,...\kappa)$ και έστω $\widetilde{\widetilde{\theta}} = \sum_{i=1}^{\kappa} \widetilde{\theta};/\kappa$

τότε , η τιμή $\widetilde{\widetilde{\Theta}}$ παρέχει μια τετριμμένη εκτίμηση της $E(\overline{\Theta})$,ενώ μια εκτίμηση της $V(\widetilde{\Theta})$ παρέχεται από την σχέση:

$$S_{\tilde{\theta}}^{2} = \frac{\sum_{i=1}^{\kappa}\left(\tilde{\tilde{\Theta}} - \tilde{\Theta}_{i}\right)^{2}}{\kappa(\kappa - 1)}$$

Η επιλογή του αριθμού κ εξισορροπεί την ανάγκη για σχετικά μεγάλο αριθμό υπό-δειγμάτων με σκοπό την εξασφάλιση της μεγάλης ακρίβειας και για αυτό χρησιμοποιείται σε σχέση με οποιαδήποτε μορφή εκτιμήτριας και οποιοδήποτε δειγματοληπτικό σχήμα .

2 . ΜΕΘΟΔΟΙ JACKKNIFE (jackknife methods)

Η τεχνική αυτή βασίζεται στην χρήση υπό-δειγμάτων και είναι γνωστή και ως μέθοδος των αλληλοεισδυόντων υποδειγμάτων (interpenetrating subsamples method) . Είναι κατάλληλη για την μελέτη συσχετισμένων σφαλμάτων και συνίσταται στον χωρισμό ενός τυχαίου δείγματος n μονάδων σε κ υπό-δείγματα με τέτοιο τρόπο ώστε το κάθε υπό-δείγμα να περιέχει m=n/k μονάδες . Η περιοχή έρευνας και η επεξεργασία του συνολικού δείγματος σχεδιάζονται με τέτοιο τρόπο ώστε να μην υπάρχει συσχέτιση μεταξύ των σφαλμάτων μέτρησης οποιωνδήποτε δυο μονάδων διαφορετικών υπό-δειγμάτων . Για αυτό μπορούμε να υποθέσουμε ότι η συσχέτιση αυτή οφείλεται στη μεροληπτικότητα , για παράδειγμα που εισάγουμε από διαφορετικούς συνεντευκτές . Η επίδραση αυτή μπορεί να απομακρυνθεί , αν σε κάθε ένα από τους κ συνεντευκτές ανατεθεί ένα διαφορετικό υπό-δείγμα και δεν υπάρχει συσχέτιση μεταξύ των σφαλμάτων μέτρησης για διαφορετικούς συνεντευκτές .

3.2 ΕΙΔΗ ΣΦΑΛΜΑΤΩΝ

Τα σφάλματα χωρίζονται σε δυο μεγάλες κατηγορίες , οι οποίες είναι:

1) Τα δειγματοληπτικά και μη δειγματοληπτικά σφάλματα.
2) Τα συστηματικά και τυχαία σφάλματα.

1 . Τα δειγματοληπτικά σφάλματα προκύπτουν και οφείλονται στην μέτρηση των χαρακτηριστικών των ατόμων του δείγματος και όχι του συνολικού πληθυσμού . Τα δειγματοληπτικά σφάλματα ονομάζονται και σφάλματα παρατήρησης (observational errors) , τα οποία είναι αποτέλεσμα του ότι , ενώ λαμβάνουμε την πληροφορία από τους επιλεγμένους του δείγματος , η πληροφορία αυτή είναι εσφαλμένη . Μια ερώτηση ενδέχεται να είναι εσφαλμένη ή μη ορθά διατυπωμένη όταν το σφάλμα οφείλεται στον συνεντευκτή (interviewer error) , στην ερώτηση (question error) ,στη καταγραφή (recording error) , στην κωδικοποίηση (coding error) και στην μετάδοση (transmission error) .

Για τα σφάλματα αυτά δεν είναι υπεύθυνο το ερωτώμενο μέλος του δείγματος , όπως συμβαίνει με τα σφάλματα που οφείλονται στο ότι ένα άτομο ενδέχεται να δώσει μια ανακριβή απάντηση σε μια ερώτηση που έχει τεθεί πολύ ορθά , είτε εσκεμμένα για να αποκρύψει πληροφορίες είτε λόγω σύγχυσης που ο συνεντεύκτης δεν μπόρεσε να αποτρέψει . Τα σφάλματα αυτά προκύπτουν κυρίως:

- όταν οι ερωτήσεις αφορούν ευαίσθητα θέματα
- όταν οι ερωτήσεις είναι πολύ λεπτομερείς ως προς την δομή
- όταν οι ερωτήσεις επισύρουν ενοχοποίηση

και τέλος όταν οι ερωτήσεις δημιουργούν ψυχολογική πίεση .

Τα σφάλματα αυτά είναι γνωστά ως **σφάλματα απάντησης** (response error) και συνήθως είναι συνδεδεμένα άμεσα με τον σχεδιασμό του ερωτηματολόγιου και της μεθόδου συλλογής δεδομένων .

Μια ακόμη κατηγορία των δειγματοληπτικών σφαλμάτων είναι τα **σφάλματα μέτρησης** (measurement errors) . Τα σφάλματα αυτά προκύπτουν κυρίως όταν υπάρχει μια συγκεκριμένη τιμή X_i αναφερόμενη στο άτομο i η οποία είναι δύσκολο να παρατηρηθεί χωρίς κάποιο επιπρόσθετο σφάλμα .

Παράδειγμα :Αν συμβολίσουμε με i τον αριθμό των σφυγμών ενός ασθενούς και διαπιστώσουμε ότι οι σφυγμοί του είναι 78 , μπορούμε να θεωρήσουμε ότι :

i = 78 ;

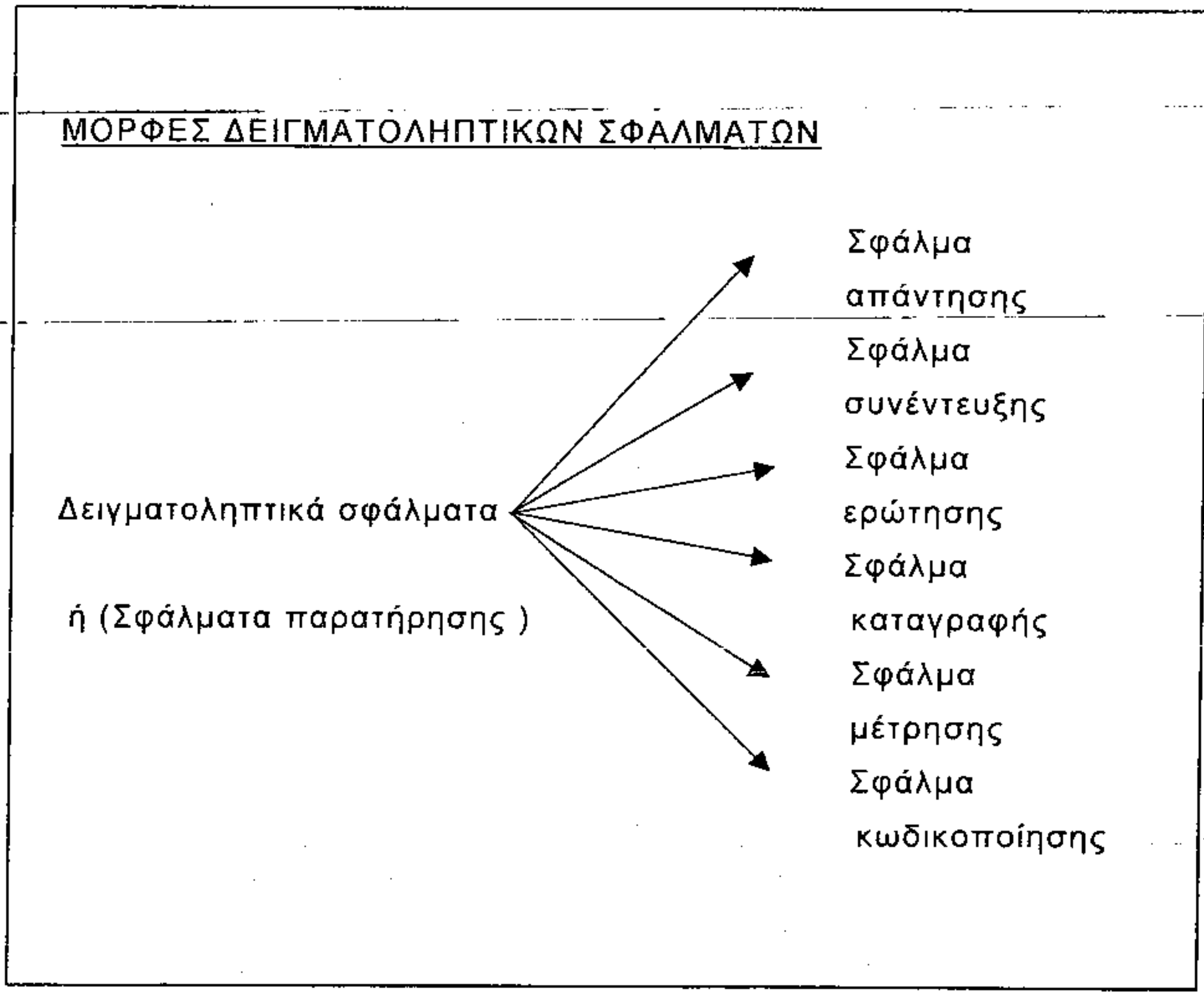
Η απάντηση θα είναι αρνητική και αυτό γιατί υπάρχει σφάλμα που είναι αποτέλεσμα της ανακρίβειας της μέτρησης και οφείλεται στην μεταβλητότητα που προκαλείται από τις φυσικές επιδράσεις .

Έστω η τιμή Y είναι μια τυχαία μεταβλητή . Το αποτέλεσμα που θα παρατηρήσουμε είναι μια παρατήρηση X_i μιας τυχαίας μεταβλητής x_i με μέση τιμή i έτσι ώστε:

$$X_i = Y_i + E_i$$

Όπου E_i συμβολίζει το σφάλμα που οφείλεται στην συνδυασμένη επίδραση της μεταβλητότητας μέτρησης και της φυσικής μεταβλητότητας .

ΠΙΝΑΚΑΣ 1 . 1



Από την άλλη έχουμε τα **μη δειγματοληπτικά σφάλματα (non – sampling error) που προκύπτουν κυρίως κατά την απογραφή και δεν οφείλονται στην επιλογή του δείγματος** .

<u>Παράδειγμα</u> : Πολλές φορές,κάποιες ερωτήσεις σε ερωτηματολόγια δεν είναι εύκολα κατανοητές και σαφής με αποτέλεσμα να παραπλανούν το άτομο που σκοπεύει να απαντήσει .

Τα μη δειγματοληπτικά σφάλματα περιλαμβάνουν κυρίως δυο είδη σφαλμάτων:

- Τα σφάλματα μη περίληψης (non-inclusion errors).
- Τα σφάλματα μη απάντησης (non-response errors).

Τα σφάλματα μη περίληψης εμφανίζονται όταν τα μέλη του αντικειμενικού πληθυσμού (target population) δεν είναι δυνατόν να περιληφθούν στο δείγμα .

Έτσι για παράδειγμα , μια τηλεφωνική έρευνα δεν μπορεί να καλύψει μέλη του πληθυσμού τα οποία δεν έχουν τηλέφωνο . Κατά βάση τα σφάλματα μη περίληψης είναι αποτέλεσμα μιας σοβαρής αναντιστοιχίας μεταξύ του πραγματικού πληθυσμού και του δειγματοληπτικού πλαισίου που σημαίνει ότι είναι δύσκολο να προλαμβάνει κανείς όλες τις επιδράσεις που ενδέχεται να οδηγήσουν σε αδυναμία προσέγγισης κάποιων μελών του πληθυσμού .

Τα σφάλματα μη απάντησης περιλαμβάνουν κατά κανόνα στο δείγμα μέλη του πληθυσμού , τα οποία δεν αποδίδουν μια τιμή για μια μεταβλητή Y που θα μπορούσαμε να ενδιαφερόμαστε να μελετήσουμε . Τα σφάλματα αυτά μπορούν να προκύψουν από διαφορετικούς λόγους , οι οποίοι συνδέονται με την φύση της αναζητούμενης πληροφορίας , (π . χ . γεγονότα ή γνώμες) με τα χαρακτηριστικά των μονάδων του πληθυσμού (πρόσωπα , διοικητικές μονάδες , βιομηχανικά συγκροτήματα) καθώς και με τη μέθοδο με την οποία επιδιώκεται να αποκτηθεί η πληροφορία (π . χ συνέντευξη , ερωτηματολόγιο , τηλεφωνική ή ταχυδρομική συνέντευξη) .

Η πιο δύσκολη μορφή μη απάντησης είναι αυτή που οφείλεται σε άρνηση συνεργασίας από την πλευρά του ερωτώμενου και που μπορεί να έχει την μορφή της έρευνας (π . χ . προσωπική ή ταχυδρομική συνέντευξη) . Μια άλλη μορφή σφάλματος είναι η αδυναμία να εντοπισθούν επιλεγέντα μέλη του δείγματος .

Για **παράδειγμα** , ας υποθέσουμε ότι πρέπει να εκτιμήσουμε το ποσοστό των ατόμων του πληθυσμού μιας περιοχής ηλικίας 18- 45. Στην περίπτωση αυτή θα συμπεριλάβουμε μόνο αυτούς που μένουν στην περιοχή αυτή την στιγμή που διεξάγεται η έρευνα και αυτούς που είναι περαστικοί . Στην περίπτωση αυτή το δείγμα αυτό δεν μπορεί να είναι αντιπροσωπευτικό γιατί υπάρχει επικάλυψη μεταξύ μη απάντησης και μη κάλυψης και έτσι η έρευνα έπρεπε να αναφέρεται στα άτομα που συνήθως ζούν μόνιμα στην περιοχή αυτή .

Ακόμα όμως και στην περίπτωση αυτή τα σφάλματα είναι ενδεχόμενα γιατί υπάρχουν άτομα που ενδέχεται να απουσιάζουν για κάποιο λόγο τη στιγμή της έρευνας (π . χ να εργάζονται σε άλλη πόλη και να διαμένουν στην υπό εξέταση περιοχή μόνο τα Σαββατοκύριακα) . Έτσι λόγω παράλειψης ατόμων που θα έπρεπε να περιληφθούν στο δείγμα ή λόγω περίληψης ατόμων που θα έπρεπε να παραληφθούν ή λόγω ενδεχόμενης διπλής καταχώρησης ατόμων ως κατοίκων δυο διαφορετικών περιοχών (τόπου εργασίας και τόπου κατοικίας) , τα σφάλματα είναι πολύ πιθανά .

Επίσης ακόμα και στην περίπτωση που ο πληθυσμός έχει καθορισθεί με ακρίβεια , τα άτομα που έχουν επιλέγει στο δείγμα και εντοπίζονται ενδέχεται να αρνηθούν να δώσουν την πληροφορία ή να αδυνατούν να δώσουν την πληροφορία λόγω αγνοίας ή λόγω του ότι το χαρακτηριστικό για το οποίο ζητείται η πληροφορία δεν είναι χαρακτηριστικό που τα άτομα αυτά έχουν .

Γενικά τα μη δειγματοληπτικά σφάλματα αποτελούν παραδείγματα μιας γενικότερης κατηγορίας λαθών και οφείλονται σε αδυναμία παρατήρησης .

ΠΙΝΑΚΑΣ 1 . 2

ΜΟΡΦΕΣ ΜΗ ΔΕΙΓΜΑΤΟΛΗΠΤΙΚΩΝ ΣΦΑΛΜΑΤΩΝ

Μη δειγματοληπτικά σφάλματα
ή
σφάλματα μη παρατήρησης
(non - observational errors)

σφάλμα μη απάντησης
(non- response error)

σφάλμα μη περίληψης
(non- inclusion error)

Τα δειγματοληπτικά και μη δειγματοληπτικά σφάλματα κάνουν πιο δύσκολη και περίπλοκη την διεξαγωγή μιας έρευνας και για αυτό είναι χρήσιμο μερικές φορές να γίνει μια προερευνητική ή μια προκαταρκτική δειγματοληψία . Η έκταση της προκαταρκτικής έρευνας εξαρτάται από την περιοχή των προβλημάτων που απαιτούν προκαταρκτική εξέταση καθώς και από τα χρονικά και οικονομικά περιθώρια . Συνήθως τα αποτελέσματα μιας προκαταρκτικής δειγματοληψίας χρησιμοποιούνται για ενδεχόμενες μεταβολές στο σχεδιασμό της δειγματοληψίας (π . χ . επιλογή του μεγέθους του δείγματος) και την μέθοδο λειτουργίας της (π . χ μεταβολές στις ερωτήσεις , εκπαίδευση των συνεντευκτών κ.λ.π).
Τα δεδομένα αυτά δεν θεωρούνται μέρος των δεδομένων της κυρίως έρευνας που πρόκειται να αναλυθούν . Εξαίρεση αποτελούν οι περιπτώσεις της διπλής δειγματοληψίας και επαναλαμβανόμενης δειγματοληψίας .

3.3 ΤΥΧΑΙΑ ΚΑΙ ΣΥΣΤΗΜΑΤΙΚΑ ΣΦΑΛΜΑΤΑ

Τα τυχαία και συστηματικά σφάλματα επηρεάζουν κατά πολύ τα ερωτώμενα πρόσωπα με αποτέλεσμα να προκαλούνται σφάλματα στις απαντήσεις . Τα προσωπικά αυτά χαρακτηριστικά , του οποίου ερευνητή αφορούν σε γενικά πλαίσια το φύλο , ηλικία , εμφάνιση , ένδυση , ύφος , προφορά , προσωπικότητα , μόρφωση κ . τ . λ . και οφείλονται:

1) **Στις απόψεις , γνώμες και ιδέες του ερευνητή .**

Ο ερευνητής οφείλει να τηρεί με πολλή προσοχή απόλυτα τις πολιτικές , κομματικές , κοινωνικές και άλλες ιδεολογικές του θέσεις εκτός πλαισίου της έρευνας που κάνει . Διότι , εάν κάνει φανερές τις προσωπικές του απόψεις για το θέμα που καλύπτει η έρευνα, τότε είναι πιθανό ότι αυτό θα επηρεάσει τον οποιοδήποτε ερωτώμενο με αποτέλεσμα να προκαλέσει αντίστοιχο σφάλμα στην απάντηση που θα δώσει . Επίσης ο ερευνητής μπορεί να επηρεάσει και με άλλους τρόπους , όπως με τον τρόπο που μιλάει και κινείται και με τον τρόπο που ντύνεται και περιποιείται την εμφάνιση του .

2) **Στις προσδοκίες του ερευνητή** .

Οι προσδοκίες συνήθως του ερευνητή , για τις απόψεις και την συμπεριφορά των ερωτώμενων ασκούν επιδράσεις επί των απαντήσεων .

Αυτό οφείλεται σε τρεις κυρίως λόγους:

Α) Στην συνέπεια συμπεριφοράς του ερευνητή :

Που στηρίζεται στην ικανότητα του να αποκομίσει μια ορισμένη εντύπωση για την στάση και συμπεριφορά του ερωτώμενου .

Β) Στην γενική εικόνα :

Ο συνεντευκτής γρήγορα σχηματίζει την γενική εικόνα που παρουσιάζει ο ερωτώμενος με βάση π . χ . την ηλικία , το φύλο , το εισόδημα την οικογενειακή κατάσταση κ . τ . λ .

Έτσι , εάν πάρει κάποιες αόριστες απαντήσεις σε ερωτήσεις για πραγματικά γεγονότα τις ερμηνεύει σύμφωνα με την γενική εικόνα και σύμφωνα με τις απαντήσεις που προσδοκά από ένα πρόσωπο με ανάλογη εικόνα .

Α.Τ.Ε.Ι. ΠΑΤΡΑΣ
ΒΙΒΛΙΟΘΗΚΗ

Παράδειγμα:

Σε ερωτήσεις '' ποσά χρήματα διαθέτετε το μήνα για φροντιστήρια των παιδιών σας'' , όταν ο ερωτώμενος δεν δίνει σαφή απάντηση , τότε ο συνεντευκτής αντιλαμβάνεται ότι ο συγκεκριμένος τύπος οικογενειάρχη πρέπει σίγουρα να στέλνει τα παιδιά του στο φροντιστήριο . Ο συνεντευκτής μπορεί να συμπληρώσει επίσης μόνος του μερικές ερωτήσεις , με βάση την εικόνα που έχει σχηματίσει για τον ερωτώμενο .

Γ) Στην υποκειμενική πιθανότητα :

Στην περίπτωση αυτή , αν συμβεί οι συνεντεύξεις που έχουν προγραμματιστεί να διαμορφώνονται με χαμηλότερο ποσοστό από το αναμενόμενο (από αυτό δηλαδή που είχε προγραμματιστεί αρχικά) , ο συνεντευκτής τείνει να ερμηνεύει τις ασαφές απαντήσεις ως προτιμήσεις υπέρ αυτού που αποσκοπεί .

3) Στον ερωτώμενο :

Τα σφάλματα που δημιουργούνται από τον ερωτώμενο , οφείλονται στην έλλειψη γνώσης , σε περιορισμένη μνήμη , σε αδυναμία να κατανοήσει πλήρως όλες τις απαντήσεις που του υποβάλλονται , σε κούραση που μπορεί να του δημιουργεί ένα μεγάλο ερωτηματολόγιο , στην επιθυμία του να μην δηλώσει την αλήθεια κ . τ . λ . Επίσης , πολλά σφάλματα που κάνει ο ερωτώμενος , οφείλονται στο βαθμό επηρεασμού του από τον συνεντευκτή .

3.4 ΠΗΓΕΣ ΣΦΑΛΜΑΤΩΝ

Οι πήγες αυτών των σφαλμάτων επηρεάζονται από διάφορες συνθήκες . Οι συνθήκες αυτές δημιουργούνται κατά:

1)Το στάδιο υποβολής ερωτήσεων .

Τα σφάλματα του σταδίου αυτού προκαλούνται από τις παραλλαγές που κάνει ο συνεντευκτής στην φραστική διατύπωση και την σειρά των ερωτήσεων .

Αυτού του είδους τα σφάλματα μπορεί να εξουδετερωθούν ή να περιοριστούν σοβαρά στο σύνολο των προσώπων που συνεργάζονται στην έρευνα .

2)Το στάδιο αναζήτησης αλήθειας .

Ορισμένοι ερευνητές δέχονται την απάντηση << δεν γνωρίζω >> . Άλλοι προσπαθούν να διαπιστώσουν αν όντως ο ερωτώμενος δεν γνωρίζει κάνοντας επιπρόσθετες ερωτήσεις . Αυτό μερικές φορές τους οδηγεί στο ότι ο ερωτώμενος γνωρίζει και τελικά δίνει μια πιο ακριβή απάντηση .

3)Το στάδιο της καταγραφής ή κωδικοποίησης απάντησης .

Είναι το στάδιο όπου τα σφάλματα γίνονται τυχαία και όχι συστηματικά .

4)Το στάδιο της παραποίησης και εικονικότητας .

Εδώ τα σφάλματα δημιουργούνται συνειδητά από τον συνεντευκτή και οφείλονται στην μειωμένη ηθική του προσωπικότητα , στις συνθήκες της δουλειάς του , στην αμοιβή και στη στάση του απέναντι στην συγκεκριμένη έρευνα .

Τέτοιου είδους ως παράδειγμα μπορούμε να αναφέρουμε μια μεγάλη εταιρία διεξαγωγής δημοσκοπήσεων που απείλησε να οδηγήσει στα δικαστήρια συνεντευκτές που είχαν συμπληρώσει αυθαίρετα τα ερωτηματολόγιά τους .

Ερωτηματολόγιο και συνθήκες διεξαγωγής συνέντευξης :

Δημιουργούνται συνήθως λάθη από την κόπωση όταν το ερωτηματολόγιο είναι μεγάλο ή όταν οι ερωτήσεις είναι πολύπλοκες και κουραστικές .

Κλασικό είναι το εξής παράδειγμα :

Καπνίζεται;

- ΝΑΙ
- ΟΧΙ
-

Αν ΟΧΙ , παρακαλούμε να απαντήσετε στις εξής ερωτήσεις:

Α)..

Β)..

Γ)..

Δ)..

Έτσι οι ερωτώμενοι , όταν αντιληφθούν ότι η απάντηση **ΟΧΙ** σημαίνει υποβολή περισσότερων ερωτήσεων άρα και απαντήσεων , επιλέγουν την απάντηση **ΝΑΙ** για να τελειώνουν γρηγορότερα .

Σημαντικό τέλος , ρόλο παίζει και αν το θέμα είναι στα ενδιαφέροντα του ερωτώμενου ή αν του κινεί το ενδιαφέρον ώστε να απαντήσει . Αν ισχύουν τα παραπάνω , τότε θα είναι σίγουρα πιο πρόθυμος να απαντήσει σε οποιαδήποτε σχετική ερώτηση .

Γίνεται λοιπόν , σαφές ότι για να επιτύχει μια δειγματοληπτική έρευνα, όσον αφορά την ακρίβεια και την σαφήνεια μιας πληροφόρησης που θα προέλθει από τις απαντήσεις των ερωτημάτων , εξαρτάται από το ερωτηματολόγιο .

Στην επίδραση που έχουν στα αποτελέσματα μιας έρευνας οι παράγοντες που συνδέονται με το ερωτηματολόγιο , προστίθεται και η επίδραση των παραγόντων που συνδέονται με την οργάνωση και την καταγραφή των δεδομένων που θα χρησιμοποιηθούν για τον μετέπειτα υπολογισμό των διάφορων εκτιμήσεων .

Για την αποφυγή όσο το δυνατόν περισσότερων σφαλμάτων , σπουδαίο ρόλο παίζει η **έρευνα αγοράς** , που αν αποδειχθεί επιτυχημένη , τα αποτελέσματα μιας δειγματοληψίας θα είναι περισσότερο ακριβείς και σωστά .

ΚΕΦΑΛΑΙΟ 4

4.1 ΕΡΕΥΝΑ ΑΓΟΡΑΣ

Η έρευνα αγοράς είναι μια ευρύτερη διοικητική δραστηριότητα , όσο αφορά το χώρο των επιχειρήσεων .

Στόχος της είναι η συστηματική , αντικειμενική και λεπτομερής αναζήτηση και μελέτη των μεταβλητών εκείνων που συμμετέχουν σαν ενεργά στοιχεία κάθε προβλήματος που ανάγεται στην αρμοδιότητα του μάρκετινγκ . Δεν περιορίζεται δηλαδή , απλά στο να " ρωτάει " τους τελικούς καταναλωτές για το τί πιστεύουν και σκέφτονται ή αισθάνονται για διάφορα προϊόντα και υπηρεσίες .

Η έρευνα αγοράς είναι ένα από τα πιο χρήσιμα εργαλεία για την διεξαγωγή της δειγματοληψίας πάνω στην οποία βασίζονται τα στελέχη του μάρκετινγκ . Τα στελέχη του μάρκετινγκ για την συλλογή περισσότερων πληροφοριών όσον αφορά τους καταναλωτές και την αγορά γενικότερα , στηρίζονται και σε άλλες μεθόδους και τεχνικές . Οι **μέθοδοι** αυτοί μπορεί να είναι : οι δοκιμαστικές αγορές , η προσωπική και μηχανική παρατήρηση και η χρησιμοποίηση υπαρχόντων δεδομένων . Με την βοήθεια αυτών των μεθόδων , τα στελέχη του μάρκετινγκ έχουν την δυνατότητα να παίρνουν καλύτερες αποφάσεις για τα όποια προβλήματα τυχόν χειρίζονται . Γενικά ως αφορισμό του **MARION HARPER** μπορούμε να υποστηρίξουμε ότι: " το να διοικήσεις μια επιχείρηση καλά , είναι το να μπορέσεις να διοικήσεις το μέλλον και για να το πετύχεις αυτό , θα πρέπει να συλλέξεις και να αξιολογήσεις πληροφορίες . " .

Αυτό λοιπόν , σημαίνει ότι η έννοια της πληροφόρησης είναι πολύ σημαντική και για αυτό πρέπει να περιλαμβάνει στοιχεία τόσο από το εσωτερικό περιβάλλον όσο και από το εξωτερικό περιβάλλον μιας επιχείρησης .

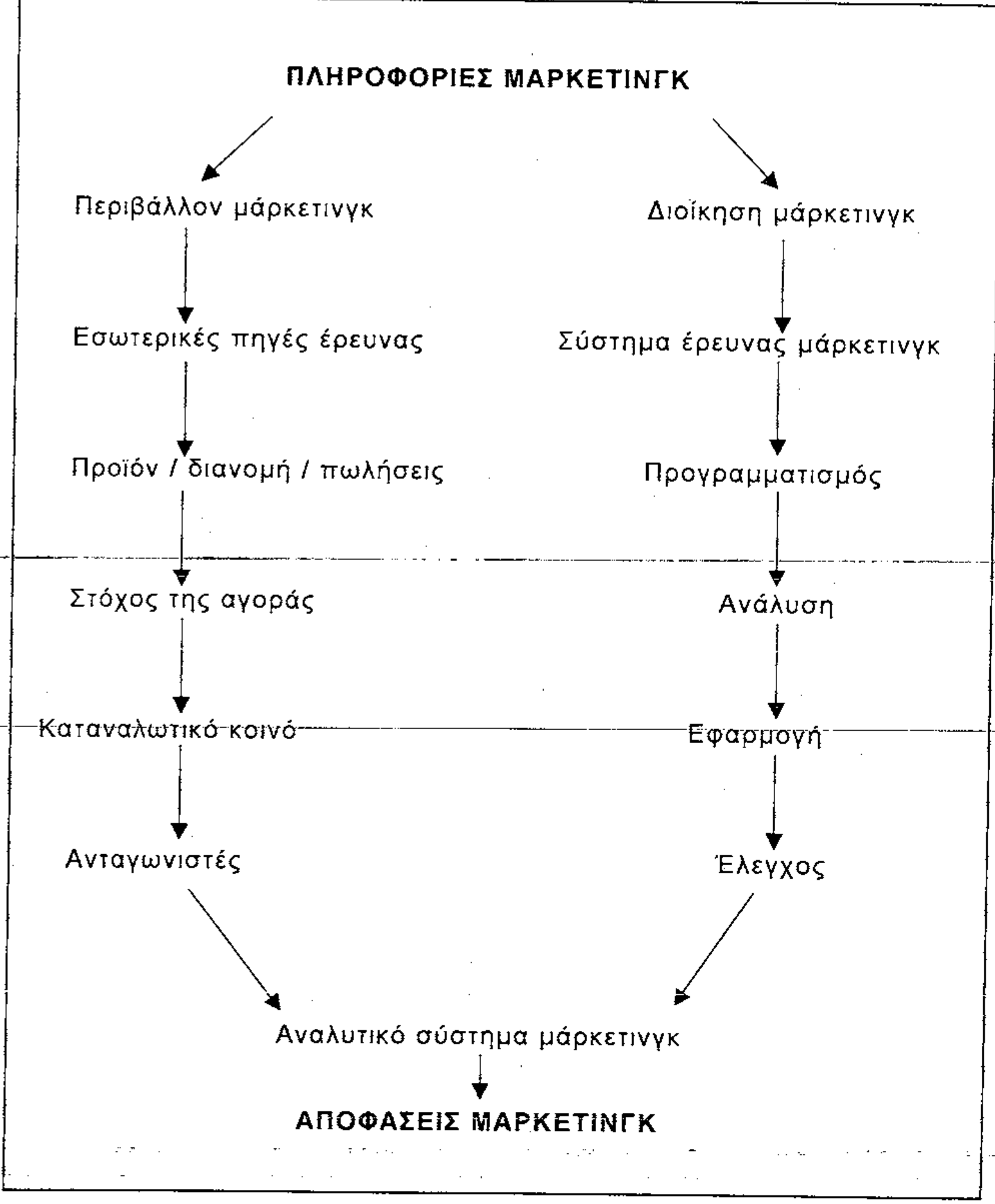
Για να το κάνουμε πιο κατανοητό θα παραθέσουμε ένα σύστημα πληροφόρησης μάρκετινγκ που βασίζεται σε τρεις κύριες μεταβλητές:

1) Το περιβάλλον μάρκετινγκ
2) Το σύστημα πληροφόρησης
3) Τη διοίκηση μάρκετινγκ

Προκειμένου λοιπόν , η διοίκηση μάρκετινγκ αποφασίσει για την ανάλυση , προγραμματισμό και εφαρμογή της έρευνας αγοράς , απαραίτητη προϋπόθεση είναι η πληροφόρηση για τις συνθήκες που επικρατούν στην αγορά που στοχεύει η επιχείρηση , για την διανομή , για τους ανταγωνιστές , για το καταναλωτικό κοινό και για τους συντελεστές του εξωτερικού περιβάλλοντος .

ΠΙΝΑΚΑΣ 4 . 1

ΣΥΣΤΗΜΑ ΠΛΗΡΟΦΟΡΗΣΗΣ ΜΑΡΚΕΤΙΝΓΚ



Το παραπάνω σύστημα πληροφόρησης μάρκετινγκ περιλαμβάνει έρευνες που αφορούν:

1) Τους μακροοικονομικούς συντελεστές του περιβάλλοντος όπως , επιτόκια δανεισμού , ακαθάριστο εθνικό προϊόν , πληθωρισμό και επιτόκια καταθέσεων .

2) Τα προϊόντα μιας επιχείρησης όπως και αυτά των ανταγωνιστών , τα οποία είναι: η αξιολόγηση των προϊόντων νέων ανταγωνιστών και η μελέτη για τους νέους χρήστες των προϊόντων της επιχείρησης .

3) Τους ανταγωνιστές μιας επιχείρησης όσον αφορά την χρηματοοικονομική ανάλυση των ανταγωνιστών και την μελέτη χαρτοφυλακίου προϊόντων ανταγωνιστών .

4) Το καταναλωτικό κοινό όπως: ο καθορισμός των καταναλωτών που αποδέχονται να χρησιμοποιήσουν ένα νέο προϊόν , οι μελέτες που αποβλέπουν στην ''αποβολή'' ενός προϊόντος από τους καταναλωτές , και οι μελέτες που αποσκοπούν στη αποδοχή ή όχι της συσκευασίας των προϊόντων μιας επιχείρησης .

5) Το δίκτυο διανομής και τους τρόπους βελτίωσης της τμηματοποίησης της αγοράς όσο αφορά: τις μελέτες διανομής των προϊόντων , την μέτρηση των μεταβολών σε γεωγραφικές περιοχές παραγωγικών πωλήσεων και την αναθεώρηση τμηματοποίησης της αγοράς με γεωγραφικά κριτήρια .

6) Τον στόχο της αγοράς που είναι και το πιό σημαντικό κομμάτι μιας έρευνας και περιλαμβάνει: την ανάλυση του τμήματος της αγοράς όσο αφορά ένα συγκεκριμένο προϊόν , την ανάλυση των γεωγραφικών περιοχών που θα αναπτυχθεί το ανάλογο προϊόν , την ανάλυση των χαρακτηριστικών γνωρισμάτων της αγοράς , την μελέτη της φύσης της αγοράς , την μελέτη των αλλαγών των προτιμήσεων των καταναλωτών , την μελέτη των τάσεων της αγοράς για συγκεκριμένα προϊόντα , και την μελέτη των οικονομικών παραγόντων που επηρεάζουν συγκεκριμένα προϊόντα .

Για την πραγματοποίηση των παραπάνω ερευνών , η διεύθυνση μάρκετινγκ πρέπει να αναπτύξει ένα σύστημα πληροφόρησης μάρκετινγκ που θα εντάσσεται στο ανάλογο τμήμα της αγοράς . Το τμήμα αυτό περιέχει :

1) Το <u>σύστημα έρευνας μάρκετινγκ</u> που αναφέρεται στην διαδικασία προσέγγισης της έρευνας μάρκετινγκ με την χρήση πρωτογενών ή δευτερογενών στοιχείων από εσωτερικές πηγές ή από εξωτερικούς παράγοντες που αφορούν την έρευνα αγοράς , όπως συνεργάτες – επιχειρήσεις .

2) Τις <u>εσωτερικές πηγές της έρευνας</u> που αναφέρονται στην μελέτη των στοιχείων που αντλούνται από το λογιστήριο και από το τμήμα πωλήσεων των επιχειρήσεων .

3) Το <u>αναλυτικό σύστημα μάρκετινγκ</u> που ασχολείται με την εξαγωγή στοιχείων μέσω της στατιστικής αλλά και της χρήσης διαφόρων μοντέλων διοίκησης .

<u>ΔΙΑΓΡΑΜΜΑ 4 . 2</u>

<u>Αναλυτικό σύστημα μάρκετινγκ</u>

<u>Στατιστικά μοντέλα</u>	<u>Διοικητικά μοντέλα</u>
Παλινδρόμηση	Σχεδιασμός προϊόντος
Συσχέτιση	Τιμολογιακή πολιτική
Παραγοντική ανάλυση	Μίγμα προώθησης

<u>Πληροφορίες μάρκετινγκ</u>

Διακριτική ανάλυση	Προπολογ . Διαφημ .
Ανάλυση ομάδων	Μοντέλο διανομής .

<u>Απόδοση μάρκετινγκ</u>

Λαμβάνοντας υπόψη όλες τις παραπάνω μεταβλητές , για να προσδιορίσουμε ένα πρόβλημα που μπορεί να απασχολεί μια επιχείρηση όσον αφορά την διεξαγωγή ερευνών αγοράς πρέπει να θέσουμε πρώτα μερικούς στόχους .

Κατόπιν απαραίτητη προϋπόθεση είναι η αξιολόγηση της εγκυρότητας των πηγών συλλογής στοιχείων προκειμένου να τεθούν οι σωστές βάσεις για την όλη έρευνα .

Αυτό μπορεί να γίνει είτε μέσω μιας δειγματοληπτικής έρευνας είτε με την μέθοδο του πειραματισμού . Η μέθοδος αυτή ασχολείται με την παρουσίαση ενός προϊόντος σε μια επιλεγμένη περιοχή που περιλαμβάνει όλα τα απαραίτητα χαρακτηριστικά γνωρίσματα της αγοράς που χρειάζονται για την προβλεψιμότητα του μέλλοντος για ένα συγκεκριμένο προϊόν .

Με αυτόν τον τρόπο μπορεί να αντληθούν περισσότερες πληροφορίες για το σχεδιασμό του προϊόντος και την στρατηγική που θα πρέπει να ακολουθηθεί .

Οι βάσεις για το σχεδιασμό ενός προϊόντος σε συνδυασμό με μια καλή στρατηγική που θα πρέπει να ακολουθηθεί , μπορούν να οδηγήσουν στην λήψη σωστών αποφάσεων .

<u>ΥΠΟΔΕΙΓΜΑ 4 . 3</u>

<u>Διαδικασία προσέγγισης στην έρευνα αγοράς</u>

Προσδιορισμός του προβλήματος της έρευνας αγοράς

↓

Προσδιορισμός των στόχων της έρευνας αγοράς

↓

Ανάπτυξη των πηγών συλλογής στοιχείων

↓

Προσδιορισμός των πηγών συλλογής των στοιχείων

↓

Συλλογή των στοιχείων

↓

Ανάλυση των στοιχείων

↓

Παρουσίαση των στοιχείων

4.2 Ο ΡΟΛΟΣ ΤΗΣ ΕΡΕΥΝΑΣ ΑΓΟΡΑΣ

Αφού έχει καθοριστεί και έχει ολοκληρωθεί η διαδικασία προσέγγισης της έρευνας αγοράς , το επόμενο βήμα το οποίο θα πρέπει να εξετασθεί είναι ο σχεδιασμός , η αξιολόγηση και η επιλογή εκείνων των εναλλακτικών λύσεων που θεωρούνται καταλληλότερες για την επίλυση των διοικητικών προβλημάτων μιας επιχείρησης . Βασική προϋπόθεση για την αποτελεσματικότητα αυτών των αποφάσεων είναι ο καθορισμός της ποιότητας των πληροφοριών που είναι διαθέσιμες στην συγκεκριμένη χρονική στιγμή που πρέπει να ληφθεί η απόφαση . Ο ρόλος λοιπόν των στελεχών της διοίκησης είναι να μπορούν να αξιολογούν σωστά την διαδικασία της συλλογής των πληροφοριών πρωτού προχωρήσουν στην λήψη αποφάσεων .

Στην σημερινή εποχή όπου ο χώρος και οι συνθήκες είναι ιδιαίτερα ανταγωνιστικές μεταξύ των επιχειρήσεων , ο ρόλος του Μάρκετινγκ γίνεται όλο και πιο απαραίτητος για την καθορισμό της έρευνας αγοράς .

Μάρκετινγκ σημαίνει : Επίβλεψη της αγοράς και ικανοποίηση των αναγκών των καταναλωτών και εξισορρόπηση των διάφορων παραγόντων ώστε το προϊόν που θα δοκιμαστεί να καταστεί επιτυχημένο στην αγορά . Το κύριο έργο του στελέχους του Μάρκετινγκ είναι να συνδυάσει τις μεταβλητές (εσωτερικές πηγές έρευνας , σύστημα έρευνας Μάρκετινγκ , αναλυτικό σύστημα Μάρκετινγκ) που είναι γνωστές ως μίγμα Μάρκετινγκ σε ένα αποτελεσματικό πρόγραμμα Μάρκετινγκ ώστε κάθε μεταβλητή να συμπληρώνει την άλλη . Οι μεταβλητές αυτές που επηρεάζουν σε μεγάλο βαθμό την συμπεριφορά του καταναλωτή και τελικά τις αποφάσεις τους πρέπει να βρίσκονται υπό τον έλεγχο των στελεχών του μάρκετινγκ . Οι πελάτες που αποτελούν το σημείο εστίασης για όλες τις δραστηριότητες της επιχείρησης αποτελούν το επίκεντρο σημείο του προσφερόμενου μίγματος Μάρκετινγκ .

Η ικανοποίηση των αναγκών και επιθυμιών τους επηρεάζεται και από εξωτερικούς παράγοντες της επιχείρησης .

Το βασικό λοιπόν πρόβλημα που πρέπει να επιλύσει μια επιχείρηση είναι να παρακολουθήσει τις ανάγκες της αγοράς και των καταναλωτών ώστε να μπορεί να προβλέψει τις μελλοντικές αλλαγές όσον αφορά τις συνήθειες τους .

Τα προβλήματα που συνήθως απασχολούν τα στελέχη του Μάρκετινγκ είναι:

1) Σε ποιους καταναλωτές κυρίως απευθυνόμαστε για την αγορά των προϊόντων;

2) Τι εισόδημα κατέχουν οι καταναλωτές αυτοί , που διαμένουν και πόσοι μπορεί να είναι;

3) Υπάρχουν καινούργια τμήματα στην αγορά που θα ήταν ωφέλιμο να εισέλθουμε ;

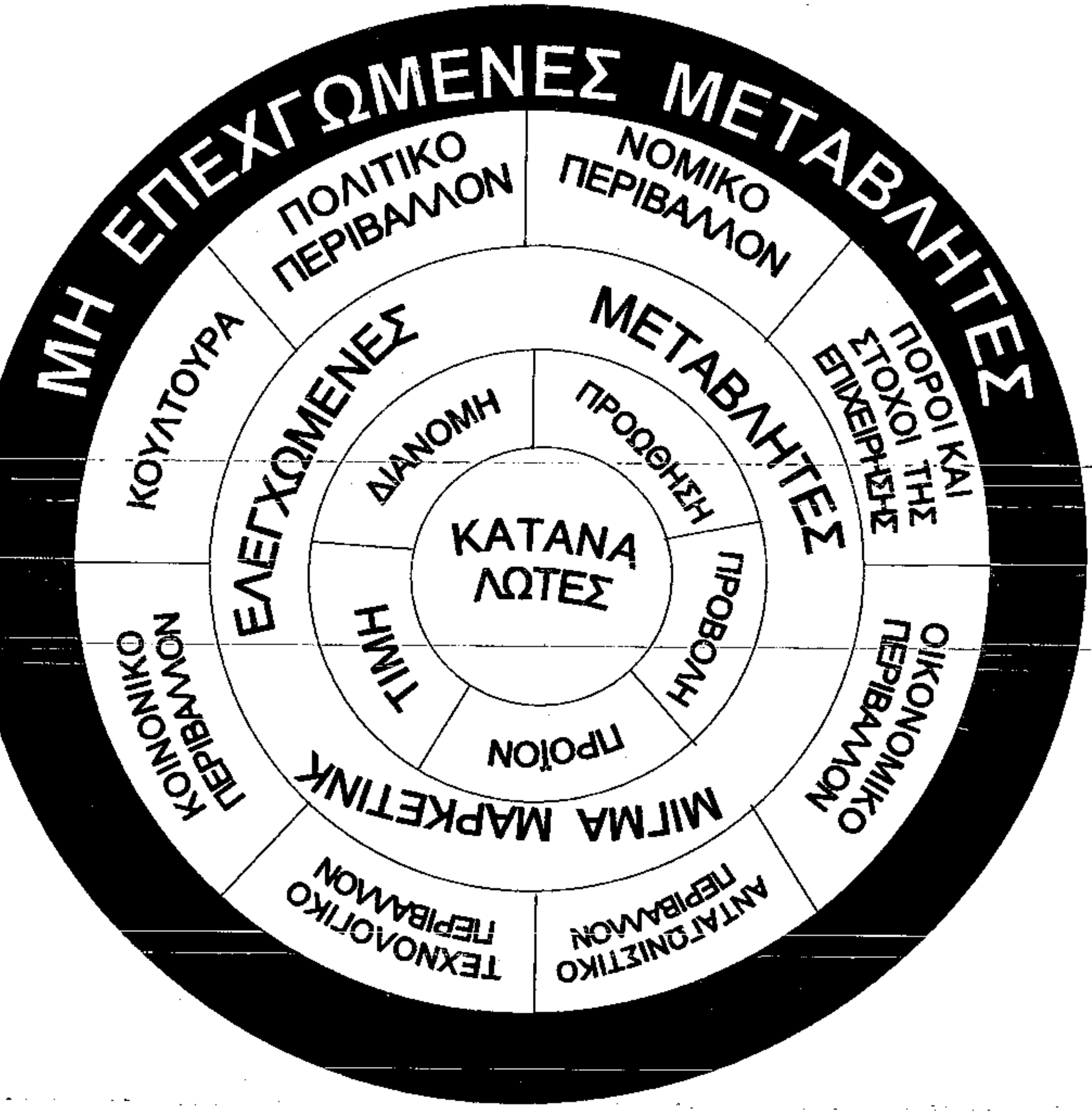
4) Η αγορά για τα προϊόντα που είναι διαθέσιμα αυξάνεται ή μειώνεται;

5) Μπορούμε να εισέλθουμε σε αγορές ξένων χωρών ;

Την λύση των προβλημάτων αυτών μπορεί να επιφέρει η ύπαρξη ενός συστήματος συλλογής πληροφοριών που θα είναι κατάλληλη για την κατανόηση των συνθηκών που επικρατούν στην αγορά.

ΕΙΚΟΝΑ 4 . 4

Σύστημα συλλογής πληροφοριών



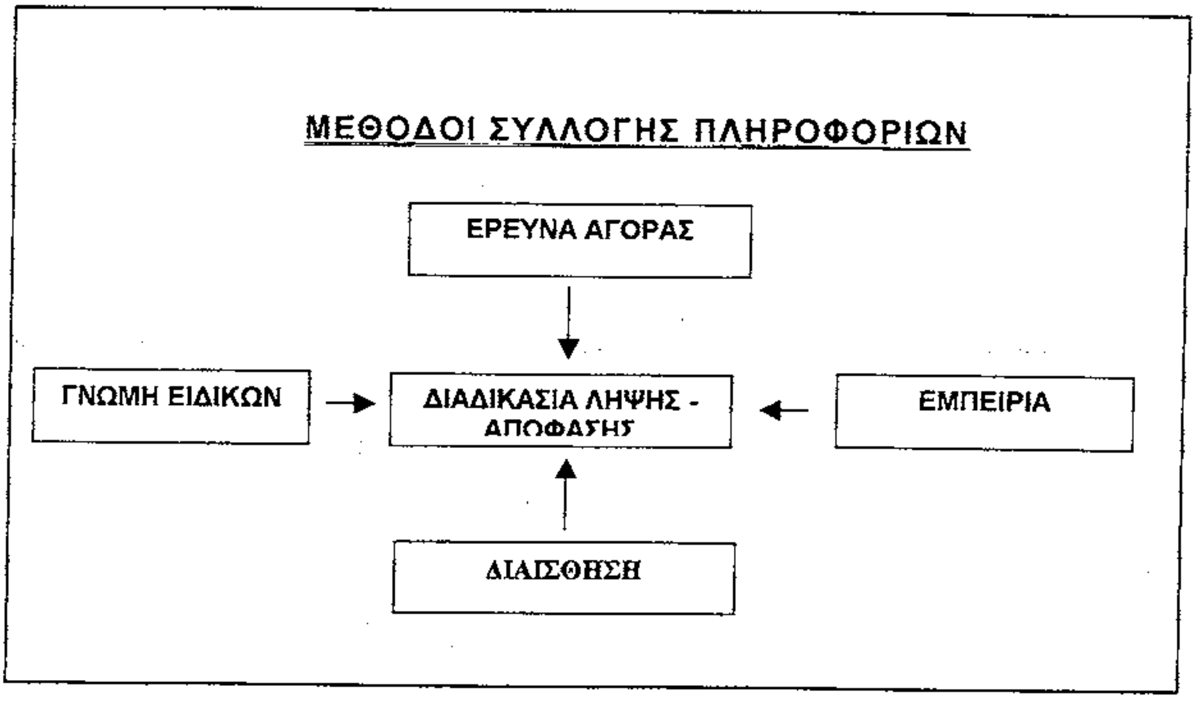
Λαμβάνοντας σοβαρά υπόψη το σύστημα συλλογής πληροφοριών είναι ευκολότερο να επιλύσουμε μερικά από τα κύρια προβλήματα που αφορούν το προϊόν , την τιμή , την διανομή και την προώθηση της έρευνας αγοράς .

Οι παρακάτω ερωτήσεις που μπορεί να μας οδηγήσουν στην επίλυση των προβλημάτων αυτών είναι:

1) Τί συστατικά θα είναι επιθυμητά;
2) Ποιά συσκευασία θα είναι καταλληλότερη και πιο επιτυχημένη;
3) Πόσο πρέπει να κοστίζει το προϊόν ;
4) Που θα πρέπει να οδηγήσει η μείωση τυχόν κόστους παραγωγής ;
5) Μέσω ποιών καναλιών θα πρέπει να διακινηθεί το προϊών;
6) Τί προσφορές πρέπει να γίνουν στους μεσάζοντες για την προώθηση των προϊόντων τους;
7) Ποια στρατηγική προώθησης θα πρέπει να ακολουθηθεί;
8) Τι ρόλο θα πρέπει να παίξει η διαφήμιση;

Οι ερωτήσεις αυτές παίζουν σημαντικό ρόλο για την συγκέντρωση των πληροφοριών και κατόπιν για την λήψη των αποφάσεων όσον αφορά την έρευνα αγοράς . Για την συλλογή στοιχείων και πληροφοριών εκτός από την έρευνα αγοράς , σπουδαίο ρόλο παίζει η <u>διαίσθηση</u> , η <u>εμπειρία</u> και η <u>γνώμη των ειδικών</u> .

<u>ΕΙΚΟΝΑ 4 . 5</u>



1) ΓΝΩΜΗ ΕΙΔΙΚΩΝ:

Η μέθοδος αυτή είναι ευρεία διαδεδομένη και προτιμάται από πολλές επιχειρήσεις γιατί μειώνει δραστικά το χρόνο που απαιτείται για την συλλογή στοιχείων . Τα άτομα αυτά έχουν κάποια συστηματική πείρα στην λήψη των αποφάσεων και πολλές φορές η βοήθεια τους είναι σημαντική .

2) ΔΙΑΙΣΘΗΣΗ Ή ΕΝΣΤΙΚΤΟ:

Η διαίσθηση έχει υποκειμενικό υπόβαθρο και μπορεί να περιλαμβάνει και κάποια εμπειρικά στοιχεία . Συνήθως οι πληροφορίες που απορρέουν από την διαίσθηση δεν συμφωνούν με τα προϋπάρχοντα δεδομένα με αποτέλεσμα να μην αποτελούν έγκυρες πηγές για την λήψη των ορθών αποφάσεων .

3) ΕΜΠΕΙΡΙΑ:

Η εμπειρία βασίζεται σε καταστάσεις και προβλήματα του παρελθόντος , τα οποία είναι παρόμοιας υφής και φύσης με τα τωρινά προβλήματα που αντιμετωπίζει το διοικητικό στέλεχος . Η εμπειρία μπορεί να αποκαλείται και γνώση .

4.3 ΠΛΕΟΝΕΚΤΗΜΑΤΑ ΕΡΕΥΝΑΣ ΑΓΟΡΑΣ ΕΝΑΝΤΙ ΑΛΛΩΝ ΜΕΘΟΔΩΝ

Όλες οι παραπάνω μέθοδοι παίζουν σημαντικό ρόλο στην διαδικασία λήψης των αποφάσεων και αποσκοπούν στην αποτελεσματικότητα των αποφάσεων αυτών . Η έρευνα αγοράς όμως σε σχέση με τις τρεις άλλες μεθόδους είναι πιο αποτελεσματική για τους εξής λόγους:

- Χρησιμοποιεί ένα συστηματικό τρόπο συλλογής και συγκέντρωσης πληροφοριών , δίνοντας έτσι πιο αξιόπιστες πληροφορίες .
- Αποσκοπεί στο να απαντήσει σε διοικητικά προβλήματα στα πλαίσια ενός συστηματικού και ελεγχόμενου εννοιολογικού πλαισίου ώστε να εξασφαλίσει τις καλύτερες δυνατόν πληροφορίες για την λήψη αποφάσεων .

• Παρέχει τα μέσα τα οποία χρειάζεται μια επιχείρηση για να συγκεντρώσει , αναλύσει και να ερμηνεύσει στοιχεία από το περιβάλλον που θα μπορούσαν να επηρεάζουν την επιτυχία των στρατηγικών σχεδίων της .

Η έρευνα αγοράς λοιπόν σύμφωνα με την American Marketing Association είναι μια διαδικασία που συνδέει τον καταναλωτή , τον πελάτη και το περιβάλλον με την επιχείρηση μέσω των πληροφοριών που απαιτούνται για να:

• Προσδιοριστούν ευκαιρίες και προβλήματα του Μάρκετινγκ .

• Σχεδιασθούν , αναθεωρηθούν και να αξιολογηθούν προγράμματα του Μάρκετινγκ .

• Ελεγχθεί η αποτελεσματικότητα των στρατηγικών του Μάρκετινγκ.

• Κατανοηθεί καλύτερα το Μάρκετινγκ ως μια διαδικασία της λήψης των αποφάσεων .

Η έρευνα της αγοράς αποτελεί το συνδετικό κρίκο της επιχείρησης ~~με το περιβάλλον . Αποσκοπεί στον έλεγχο του περιβάλλοντος και~~ της επιχείρησης υποβάλλοντας διαρκώς διάφορες ερωτήσεις όπως:

1) Ποιο θα είναι το μερίδιο της αγοράς;
2) Πόσο ικανοποιημένοι θα μείνουν οι πελάτες;
~~3) Τι επίπεδο εξυπηρέτησης παρέχουμε;~~
4) Τι εικόνα έχουμε στην αγορά;

Γενικά η έρευνα αγοράς είναι πολύ σπουδαίο εργαλείο όσον αφορά την θεωρία της δειγματοληψίας που επιλέγουν πολλές επιχειρήσεις για την επίλυση επιχειρηματικών προβλημάτων .

ΚΕΦΑΛΑΙΟ 5

ΕΚΤΙΜΗΣΗ ΠΑΡΑΜΕΤΡΩΝ ΤΟΥ ΠΛΗΘΥΣΜΟΥ

5.1 ΜΕΓΕΘΟΣ ΔΕΙΓΜΑΤΟΣ

Το μέγεθος ενός δείγματος είναι επίσης πολύ σημαντικό για έναν ερευνητή που σκοπεύει να διενεργήσει μια δειγματοληπτική έρευνα. Για τον καθορισμό του μεγέθους του δείγματος και την διεξαγωγή μιας δειγματοληπτικής έρευνας υπάρχουν μερικοί παράγοντες που πρέπει να ληφθούν υπ' όψιν όπως:

1) Η τιμή του ν δείγματος θα πρέπει να εξαρτάται από το βαθμό ακρίβειας με την οποία ο ερευνητής θα εκτιμήσει την οποιαδήποτε παράμετρο του πληθυσμού . Δηλαδή το μέγεθος του δείγματος θα πρέπει να εξαρτάται από το μέγιστο ανεκτό σφάλμα της εκτίμησης και την πιθανότητα με την οποία αυτό είναι επιτρεπτό . Όσο μεγαλύτερο είναι το δείγμα τόσο μικρότερο είναι το δειγματοληπτικό σφάλμα που σημαίνει ότι η ακρίβεια της έρευνας θα είναι μεγαλύτερη .

2) Η ομοιογένεια του πληθυσμού διαμορφώνει το μέγεθος του δείγματος . Όσο πιο ομοιογενείς είναι ένας πληθυσμός , δηλαδή η διακύμανση του είναι μικρότερη , τόσο λιγότερες μονάδες δείγματος απαιτούνται . Από την άλλη , όσο πιο ανομοιογενείς είναι ένας πληθυσμός , που σημαίνει μεγάλη διακύμανση , τόσο περισσότερες μονάδες δείγματος απαιτούνται .

3) Το κόστος και ο χρόνος της έρευνας επηρεάζουν το μέγεθος του δείγματος . Όσο μικρότερο είναι το δείγμα τόσο γρηγορότερα και με λιγότερα έξοδα διαπράττεται μια έρευνα .

4) Η επιλογή της μεθόδου μιας δειγματοληπτικής έρευνας εξαρτάται και από το μέγεθος ν του δείγματος . Για να είναι ένα δείγμα απολύτως αντιπροσωπευτικό του πληθυσμού δεν αρκεί να αποτελείται από κάποιο μεγάλο αριθμό μονάδων αλλά θα πρέπει να ακολουθείται και από την κατάλληλη μέθοδο δειγματοληπτικής έρευνας .

5.2 ΣΤΑΤΙΣΤΙΚΗ ΕΚΤΙΜΗΣΗ

Η δειγματοληψία μπορεί να μας δίνει εκτιμήσεις που βρίσκονται κοντά στο αντίστοιχο πληθυσμιακό μέγεθος , χωρίς αυτό να σημαίνει ότι βρίσκονται κοντά στο πραγματικό μέγεθος του πληθυσμού που αναζητάμε .

Άρα η **στατιστική εκτίμηση ή εκτιμητική είναι μια διαδικασία με την οποία υπολογίζουμε την πιθανότερη τιμή μιας παραμέτρου ενός πληθυσμού με βάση την τιμή του αντίστοιχου στατιστικού ενός δείγματος που παίρνουμε από τον πληθυσμό .**

Για να γίνει κατανοητό το πόσο σημαντική είναι η διαδικασία για την διεξαγωγή έγκυρων αποτελεσμάτων θα παραθέσουμε ένα παράδειγμα . Αν θέλουμε να εκτιμήσουμε το μέσο ύψος των μαθητών του λυκείου και πάρουμε ένα δείγμα που αποτελείται από μαθητές που παίζουν μόνο μπάσκετ τότε η εκτίμηση και τα συμπεράσματα που θα βγάλουμε θα είναι τελείως λανθασμένα . Για αυτό , εάν θέλουμε να έχουμε μια σωστή εκτίμηση της παραμέτρου του πληθυσμού από το στατιστικό ενός δείγματος , θα πρέπει να εφαρμόσουμε τυχαία δειγματοληψία για την επιλογή των ατόμων του δείγματος . Όσο μεγαλύτερο είναι το μέγεθος του δείγματος τόσο μικρότερο μπορεί να είναι το τυπικό σφάλμα εκτίμησης και επομένως τόσο μεγαλύτερη θα είναι η ακρίβεια εκτιμήσεως .

Υπάρχουν δυο είδη εκτιμήσεων:

- Η μονότονη εκτίμηση ή εκτίμηση σημείου
- Η εκτίμηση διαστήματος εμπιστοσύνης

1) Η σημειακή εκτίμηση (point estimation) περιγράφει τους τρόπους με τους οποίους μπορούμε να βρούμε ένα μοναδικό αριθμό που μπορεί να χρησιμοποιηθεί για την εκτίμηση της τιμής μιας άγνωστης παραμέτρου .

2) Η εκτίμηση διαστήματος εμπιστοσύνης (interval estimate) καθορίζει ένα διάστημα τιμών μέσα στο οποίο αναμένεται με μία ορισμένη πιθανότητα να βρίσκεται η πραγματική τιμή της παραμέτρου του πληθυσμού .

5.3 ΣΗΜΕΙΑΚΗ ΕΚΤΙΜΗΣΗ

Στην καθημερινή μας ζωή , στις επιστήμες και στις επιχειρήσεις παρουσιάζονται συνεχώς προβλήματα εκτίμησης που πρέπει να υπολογιστούν όπως , το ποσοστό των ανθρώπων που προτιμά ένα φάρμακο , την αναμενόμενη διακύμανση της επίδοσης ενός μαθητή , το ποσοστό των λαμπτήρων που είναι ελαττωματικοί κ . τ . λ . Για αυτό είναι απαραίτητο να κάνουμε εκτιμήσεις μέσων τιμών , ποσοστών , μεταβλητότητας κ . τ . λ .

Στην σημειακή εκτίμηση μια παράμετρος του πληθυσμού (μέση τιμή , τυπική απόκλιση) εκτιμάται από τα αντίστοιχα στατιστικά ενός τυχαίου δείγματος .

Από ένα πληθυσμό μπορούμε να πάρουμε πολλά τυχαία δείγματα , οπότε οι εκτιμήσεις για μια παράμετρο του πληθυσμού θα είναι περισσότεροι της μίας . Σκοπός μας είναι η όσο καλύτερη δυνατή εκτίμηση της παραμέτρου του πληθυσμού .

Για αυτό θα παραθέσουμε τους παρακάτω τύπους για την εκτίμηση των αντιστοιχών παραμέτρων:

$$\overline{x} = \frac{1}{\nu} \sum_{i=1}^{\nu} x_i$$

ή

$$\overline{x} = \frac{\sum_{i=1}^{\kappa} x_i \nu_i}{\nu_1 + \nu_2 + \ldots + \nu_\kappa}$$

Ο τύπος αυτός απεικονίζει την εκτιμήτρια της παραμέτρου μ του πληθυσμού . Η αριθμητική τιμή που προκύπτει από τον τύπο αυτό ονομάζεται σημειακή εκτίμηση της παραμέτρου μ του πληθυσμού .

$$s = \sqrt{\frac{\sum_{i=1}^{\nu} \left(x_i - \overline{x}\right)^2}{\nu - 1}} \quad ή \quad s = \sqrt{\frac{\sum_{i=1}^{\kappa} \nu_i \left(x_i - \overline{x}\right)^2}{\nu}}$$

Η αριθμητική τιμή που προκύπτει από αυτόν τον τύπο είναι η εκτίμηση σημείου της τυπικής απόκλισης σ του πληθυσμού . Για την καλύτερη εκτίμηση μιας παραμέτρου του πληθυσμού σημαντικό ρόλο παίζει η εκτιμήτρια . Λαμβάνοντας υπ΄ όψιν τους παραπάνω τύπους θα θέσουμε ένα παράδειγμα:

Προκειμένου να εκτιμήσουμε τη μέση τιμή μιας τυχαίας μεταβλητής Χ που έχει κανονική κατανομή θα σκεφτούμε ως εξής:

Ας πούμε ότι έχουμε στα χέρια μας ν παρατηρήσεις της τιμής της τυχαίας μεταβλητής , τις οποίες ονομάζουμε X_1 ,X_2 ,X_3 ,X_4...... ,X_ν .

Ας πούμε επίσης ότι κάθε τιμή της μεταβλητής Χ είναι ανεξάρτητη από όλες τις άλλες τιμές . Έτσι η εκτίμηση του μέσου θα πρέπει να είναι:

(εκτίμηση του μέσου) $= \bar{x} = \dfrac{x_1 + x_2 + \ldots + x_\nu}{\nu}$

Η ποσότητα αυτή είναι ο μέσος όρος όλων των τιμών της μεταβλητής Χ , δηλαδή ο μέσος του δείγματος και αποτελεί παράδειγμα ενός στατιστικού στοιχείου . **Στατιστικό στοιχείο είναι μια συγκεκριμένη ιδιότητα των στοιχείων ενός τυχαίου δείγματος .** Επίσης εκτός από της εκτίμηση του μέσου $\bar{x}$ του δείγματος μπορούμε να πάρουμε τη διάμεσο , την επικρατούσα τιμή, το μέσο εύρος κ . τ . λ . και έτσι θα έχουμε:

$$\bar{x} = \frac{\sum_{i=1}^{\nu} x_i}{\nu}$$

$$\delta = \alpha_{i-1} + \left(\frac{\nu}{2} - N_{i-1}\right)\frac{c}{\nu_i}$$

$$E.T. = a_{i-1} + \frac{\nu_i - \nu_{i-1}}{2\nu_i - (\nu_i + \nu_{i+1})}c$$

$$\text{Μέσο εύρος} = \frac{\text{μικρότερη τιμή} + \text{μεγαλύτερη τιμή}}{2}$$

Η επόμενη ερώτηση που θα μπορούσε να μας απασχολήσει είναι με ποια κριτήρια θα κάνουμε την επιλογή της εκτίμησης των παραμέτρων αυτών;

Μια εκτιμήτρια για να είναι καλή θα πρέπει να είναι <u>αμερόληπτη</u>, <u>αποτελεσματική</u>, <u>συνεπείς</u> και <u>επαρκής</u>.

1) ΑΜΕΡΟΛΗΠΤΗ ΕΚΤΙΜΗΤΡΙΑ

Μια εκτιμήτρια θεωρείται αμερόληπτη όταν η αναμενόμενη τιμή της ισούται με την πραγματική τιμή της παραμέτρου που θέλουμε να εκτιμήσουμε.

Για παράδειγμα:

$E(\bar{x})=\mu$

Οπότε ο μέσος όρος του δείγματος $\bar{x}$ αποτελεί αμερόληπτη εκτιμήτρια του μέσου πληθυσμού μ.

Η μαθηματική ελπίδα της διακύμανσης του στατιστικού στοιχείου είναι:

$$E (s_1^2) = E (ns_1^2 / n-1) = E (s_1^2) n / n-1 = \sigma^2$$

Κατά συνέπεια αμερόληπτη εκτιμήτρια της διακύμανσης είναι το s_1^2.

Γενικά, ενδέχεται να υπάρχουν πολλές διαφορετικές αμερόληπτες εκτιμήτριες της ίδιας παραμέτρου. Όσο μικρότερη είναι η διακύμανση τόσο καλύτερη θα είναι η τυχαία μεταβλητή της οποίας τη διακύμανση μπορούμε να υπολογίσουμε.

Για παράδειγμα ας υποθέσουμε ότι δυο άνθρωποι προσπαθούν να εκτιμήσουν την τιμή του άγνωστου μέσου μ ενός πληθυσμού. Ο ένας από αυτούς επιλέγει με τυχαίο τρόπο δυο στοιχεία και υπολογίζει το μέσο όρο τους $A_2 =(X_1+X_2)$. Αφού $E(A_2) =\mu$ πρόκειται για μια αμερόληπτη εκτιμήτρια.

Αντίθετα ο άλλος άνθρωπος επιλέγει με τυχαίο τρόπο 100 στοιχεία και υπολογίζει το μέσο όρο τους $A_{100} = (X_1+X_2+X_3+\ldots\ldots\ldots+X_{100})$.

Και οι δυο εκτιμήτριες είναι αμερόληπτες , με την διάφορα ότι το A_{100} είναι πολύ καλύτερη εκτιμήτρια από το A_2 επειδή η διακύμανση του είναι μικρότερη ($\sigma^2/100$ αντί $\sigma^2/2$) .
Βέβαια το γεγονός ότι μια εκτιμήτρια είναι αμερόληπτη δεν σημαίνει ότι είναι και η καλύτερη που θα μπορούσαμε να χρησιμοποιήσουμε.

2) ΑΠΟΤΕΛΕΣΜΑΤΙΚΗ ΕΚΤΙΜΗΤΡΙΑ

Μια άλλη εκτιμήτρια που αναφέρεται στη δειγματική διασπορά είναι η αποτελεσματική εκτιμήτρια . Από όλες τις αμερόληπτες εκτιμήτριες μιας παραμέτρου θ , αποτελεσματικότερη θεωρείται εκείνη που έχει την μικρότερη διασπορά .
Για παράδειγμα αν , για δυο αμερόληπτες εκτιμήτριες ($\overline{\Theta}_1$ και $\overline{\Theta}_2$) της θ είναι $V_{ar}(\overline{\theta}_1) < V_{ar}(\overline{\theta}_2)$ τότε η $\overline{\theta}_1$ είναι αποτελεσματικότερη εκτιμήτρια από τη $\overline{\Theta}_2$. Αποδεικνύεται ότι ο αριθμητικός μέσος δείγματος έχει τη μικρότερη αναμενόμενη διασπορά μεταξύ όλων των άλλων εκτιμητριών (διάμεσο , Ε . Τ μέσο εύρος) της παραμέτρου μ του πληθυσμού και για αυτό θεωρείται η αποτελεσματικότερη εκτιμήτρια της μ .

3) ΣΥΝΕΠΗΣ ΕΚΤΙΜΗΤΡΙΑ

Συνεπής εκτιμήτρια είναι μια εκτιμήτρια της οποίας η διακύμανση πλησιάζει το 0 όσο αυξάνεται το μέγεθος του δείγματος . Δηλαδή μια εκτιμήτρια $\overline{\theta}$ της παραμέτρου θ θα λέμε ότι είναι συνεπής όταν η διακύμανση της $V_{ar}(\overline{\theta})$ τείνει προς το 0 καθώς το μέγεθος ν του δείγματος αυξάνεται απεριόριστα , δηλαδή :

$$\lim_{\nu\to\infty} V_{ar}(\overline{\theta}) = 0$$

Όπως παρατηρούμε , σε μια δειγματικη κατανομή του μέσου $\bar{x}$ είναι $\mu_{\bar{\chi}} = \mu$ και $\sigma_{\bar{\chi}}^2 = \sigma^2 / \nu$. Αφού $\lim_{\nu \to +\infty} \frac{\sigma^2}{\nu} = 0$, συμπεραίνουμε ότι η εκτιμήτρια $\bar{x}$ είναι συνεπής εκτιμήτρια της μ .

Έτσι εξασφαλίζεται ότι η εκτιμήτριά μας θα βρίσκεται πολύ κοντά στην αναμενόμενη τιμή της επιλέγοντας απλώς ένα μεγάλο μέγεθος δείγματος .

Δεν υπάρχει αμφιβολία ότι κανείς δεν μπορεί να εμπιστευθεί ανθρώπους που θεωρούνται ασυνεπείς . Έτσι συμβαίνει και με μια εκτιμήτρια .

Για αυτό υποθέτουμε ότι έχουμε τη δυνατότητα να αυξήσουμε το μέγεθος του δείγματός μας σε μεγάλο βαθμό και συνεπώς να κάνουμε πολύ περισσότερες παρατηρήσεις της τυχαίας μεταβλητής Χ.

Σε αυτή την περίπτωση , η νέα τιμή του x θα είναι πιο κοντά στο μέσο μ ή υπάρχει πιθανότητα να βρίσκεται μακρύτερα .

4) ΕΠΑΡΚΗΣ ΕΚΤΙΜΗΤΡΙΑ

Μια εκτιμήτρια $\bar{\theta}$ της παραμέτρου θ θα λέμε ότι είναι επαρκής , όταν χρησιμοποιεί όλες τις σχετικές με την παράμετρο πληροφορίες που περιέχονται σε ένα δείγμα .

Για παράδειγμα , ο μέσος $\bar{x}$ του δείγματος είναι επαρκής εκτιμήτρια του μέσου μ του πληθυσμού , ενώ η διάμεσος που προκύπτει από το δείγμα δεν είναι επαρκής εκτιμήτρια , αφού δεν χρησιμοποιεί τις τιμές των παρατηρήσεων , αλλά μόνο την θέση τους .

Συμπερασματικά όσον αφορά της εκτίμηση του μέσου μ του πληθυσμού μπορούμε να πούμε ότι σύμφωνα με το κεντρικό οριακό θεώρημα , η δειγματική κατανομή του μέσου $\bar{x}$ για μεγάλα τυχαία δείγματα ακολουθεί (κατά προσέγγιση) την κανονική κατανομή .

Έτσι μπορούμε να ισχυριστούμε ότι με πιθανότητα 1-α (όπου α = 0,05 ή α= 0 ,01 συνήθως) ο δειγματικός μέσος διαφέρει $\bar{x}$ από τον μέσο μ το πολύ κατά $z_\alpha/2$ τυπικά σφάλματα του μέσου και αυτό γιατί είπαμε ότι:

Αν ν≥ 30 , τότε $\bar{x} = N\left(\mu, \frac{\sigma^2}{\nu}\right)$ και $z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{\nu}}} \cong (N/0,1)$ που

σημαίνει ότι το σφάλμα $E = \bar{x} - \mu = z\frac{\sigma}{\sqrt{\nu}} = z_{\alpha/2}\frac{\sigma}{\sqrt{\nu}}$

Έτσι για α = 0 ,05 είναι α/2 = 0 ,025 και από κάποιο πίνακα μιας τυπικής κανονικής κατανομής βρίσκουμε ότι $z_{0.025} = 1.96$.

Έτσι το σφάλμα που κάνουμε με πιθανότητα 95% είναι : $E = 1{,}96\frac{\sigma}{\sqrt{\nu}}$

Όταν όμως δεν γνωρίζουμε το σ και είναι $\nu \geq 30$, μπορούμε σε αυτή την περίπτωση να χρησιμοποιήσουμε την τυπική απόκλιση s του δείγματος .

5.4 ΕΚΤΙΜΗΣΗ ΔΙΑΣΤΗΜΑΤΟΣ - ΔΙΑΣΤΗΜΑ ΕΜΠΙΣΤΟΣΥΝΗΣ

Αναφερθήκαμε προηγουμένως στην σημειακή εκτίμηση που προκύπτει σε ένα τυχαίο δείγμα . Είδαμε ότι μια παράμετρος θ ενός πληθυσμού μπορεί να εκτιμηθεί από μια εκτιμήτρια $\bar{\theta}$, αρκεί να επιλέξουμε την κατάλληλη εκτιμήτρια . Κάποια δείγματα δίνουν σημειακές εκτιμήσεις που βρίσκονται πλησιέστερα στη θ από ότι κάποια άλλα .

Επίσης γνωρίζουμε ότι διάφορα κριτήρια επιλογής εκτιμητριών μας πληροφορούν για τους αναμενόμενους μέσους όρους των σημειακών εκτιμητριών που δίνει η εκτιμήτρια , αλλά δεν μας πληροφορούν για την θέση μίας συγκεκριμένης εκτίμησης της πραγματικής τιμής της παραμέτρου θ , δηλαδή αν η εκτίμηση είναι κοντά ή μακριά από τη θ . Από αυτό διαπιστώνουμε ότι , η εκτίμηση πρέπει να συνοδεύεται από ένα μέτρο πιθανού σφάλματος σε σχέση με την πραγματική τιμή θ .

Για τον λόγο αυτό , καλό είναι να προσδιορίσουμε ένα διάστημα μέσα στο οποίο να βρίσκεται η άγνωστος παράμετρος του πληθυσμού με μια προκαθορισμένη πιθανότητα 0 ,95 ή 0 ,99 . Η πιθανότητα αυτή αντιστοιχεί στην αναλογία των δειγμάτων που δίνουν εκτιμήσεις μέσα στο διάστημα αυτό .

Το υπόλοιπο της πιθανότητας αυτής (0 ,05 ή 0 ,01) δείχνει την αναλογία των δειγμάτων που δίνουν σημειακές εκτιμήσεις έξω από το καθορισμένο διάστημα .

Πιο συγκεκριμένα , ας υποθέσουμε ότι θέλουμε να εκτιμήσουμε την παράμετρο θ ενός πληθυσμού , παίρνοντας ως παράδειγμα ένα τυχαίο δείγμα με εκτιμήτρια της $\bar{\Theta}$. Ας υποθέσουμε επίσης ότι ζητάμε να προσδιορίσουμε δυο αριθμούς $\bar{\theta} - \varepsilon$ και $\bar{\theta} + \varepsilon$ τέτοιους ώστε:

$P(\bar{\theta} - \varepsilon < \Theta < \bar{\theta} + \varepsilon) = 1 - \alpha$,

Όποτε :

$P(\Theta \leq \bar{\theta} - \varepsilon) + P(\Theta > \bar{\theta} + \varepsilon) = \alpha$, όπου το α μπορεί να πάρει οποιαδήποτε τιμή πιθανότητας .

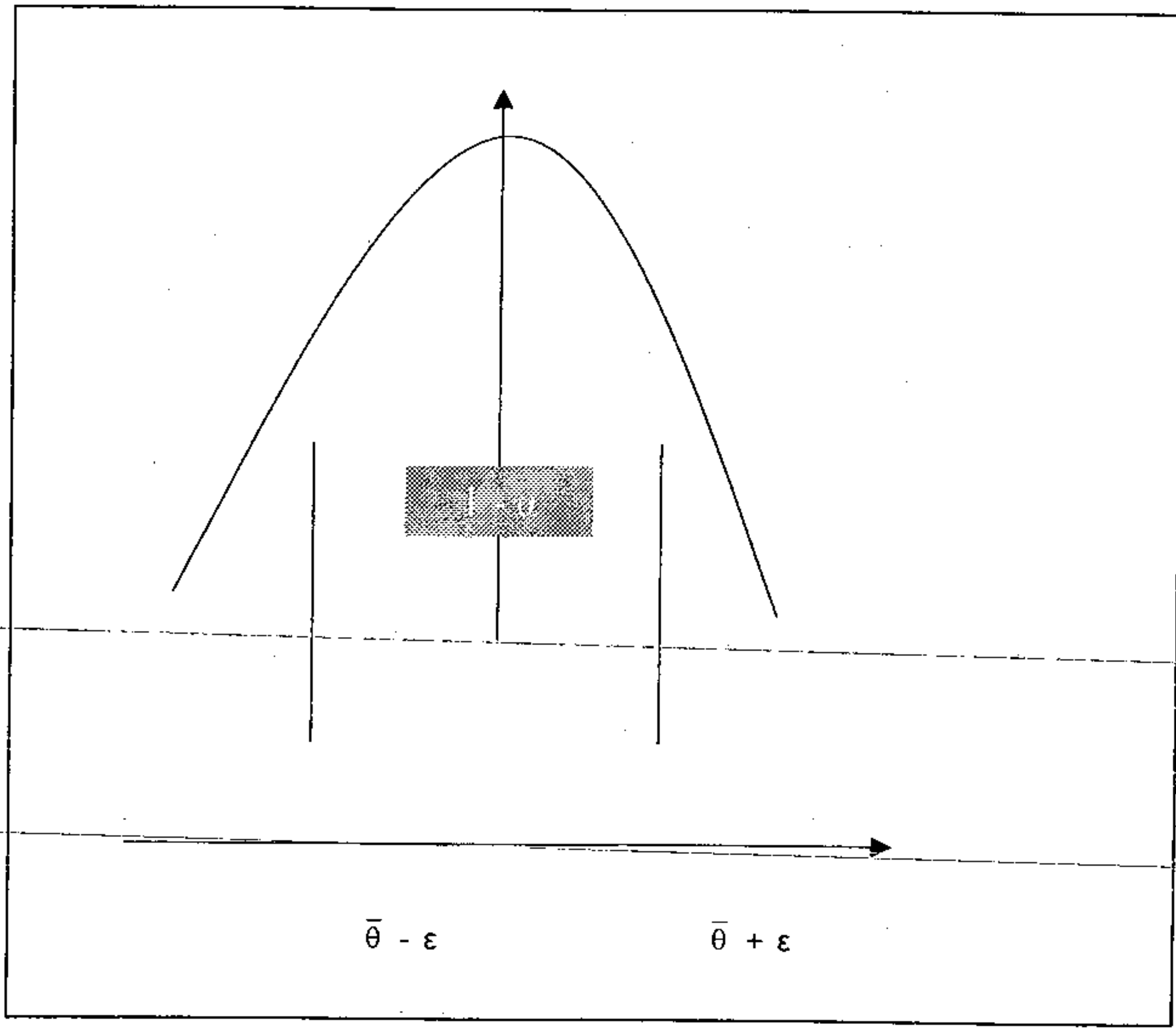
Εδώ λοιπόν , το διάστημα εμπιστοσύνης της παραμέτρου θ ή 100 (1-α)% είναι το διάστημα ($\bar{\theta}$-ε , $\bar{\theta}$+ε) . Τα όρια $\bar{\theta}$-ε και $\bar{\theta}$+ε είναι τα όρια εμπιστοσύνης . Το $\bar{\theta}$-ε ονομάζεται κάτω όριο εμπιστοσύνης και $\bar{\theta}$+ε ονομάζεται άνω όριο εμπιστοσύνης.

Επίσης , η πιθανότητα 1 – α ονομάζεται συντελεστής εμπιστοσύνης ή επίπεδο εμπιστοσύνης και η πιθανότητα α επίπεδο σημαντικότητας . Άρα έχουμε 0 < α < 1 .

Στην πράξη , για να προσδιορίσουμε ένα διάστημα εμπιστοσύνης για την παράμετρο θ ενός πληθυσμού , θα εργαστούμε ως εξής:

1. Παίρνουμε ένα τυχαίο δείγμα μεγέθους ν από το πληθυσμό .
2. Υπολογίζουμε από το δείγμα αυτό μια εκτίμηση $\bar{\theta}$ της θ .
3. Υπολογίζουμε μια εκτίμηση της τυπικής απόκλισης της παραμέτρου του πληθυσμού από τα στοιχεία του δείγματος .
4. Καθορίζουμε τη δειγματική κατανομή που ακολουθεί η παράμετρος θ .
5. Προσδιορίζουμε την πιθανότητα 1 – α με την οποία θα υπολογίσουμε το διάστημα εμπιστοσύνης ($\bar{\theta}$-ε , $\bar{\theta}$+ε) για να έχουμε: $P(\bar{\theta} - \varepsilon \leq \Theta \leq \bar{\theta} + \varepsilon) = 1 - \alpha$

ΔΙΑΓΡΑΜΜΑ 5 . 1



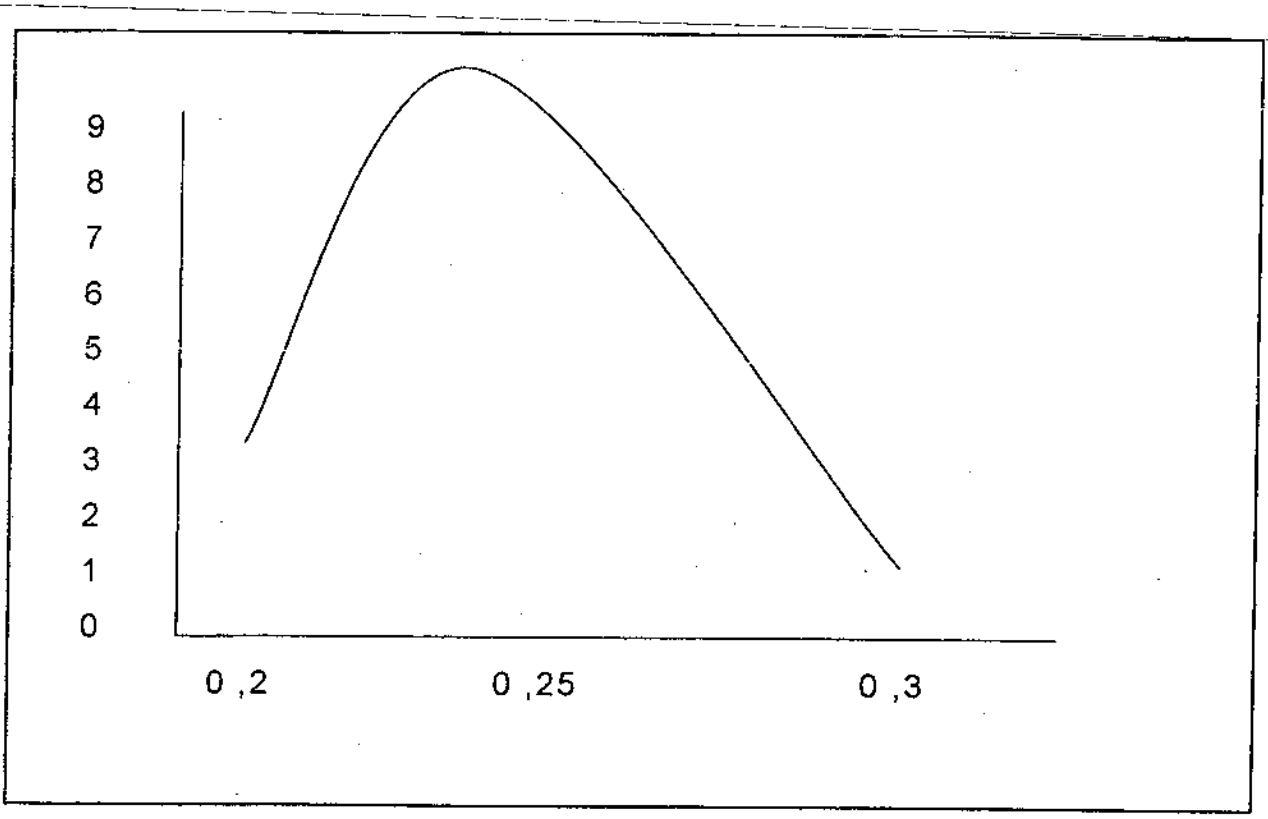
Με βάση το παραπάνω διάγραμμα , μπορούμε να καταλάβουμε ότι βαθμός εμπιστοσύνης είναι το 1-α και αυτό σημαίνει ότι αν πάρουμε διάφορα δείγματα από τον ίδιο πληθυσμό που έχουν το ίδιο μέγεθος ν , τότε κάθε δειγματική εκτίμηση θα δώσει ένα διαφορετικό διάστημα . Κατά μέσο όρο , το 100 (1- α) % από αυτά τα διαστήματα θα περιέχουν την άγνωστη παράμετρο θ του πληθυσμού .

Για **παράδειγμα** :Αν α = 0 ,05 τότε το 1-α = 0 ,95 που σημαίνει ότι με πιθανότητα 95% η τιμή της παραμέτρου θ που εκτιμάμε θα βρίσκεται σε αυτό το διάστημα εμπιστοσύνης και με πιθανότητα 5% θα βρίσκεται έξω από αυτό . Διαπιστώνουμε λοιπόν ότι όσο αυξάνεται ο βαθμός εμπιστοσύνης 1-α για το ίδιο δείγμα , τόσο το πλάτος του διαστήματος εμπιστοσύνης μεγαλώνει , με αποτέλεσμα να μειώνεται η ακρίβεια της εκτίμησης .

5.5 ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΑΝΑΛΟΓΙΩΝ

Τα διαστήματα εμπιστοσύνης αναλογιών μπορούν να υπολογιστούν καλύτερα αν προσεγγιστούν με βάση την κανονική κατανομή .
Έστω λοιπόν , ρ η αναλογία Χ/ν του δείγματος , όπου Χ είναι το πλήθος των ανθρώπων του δείγματος . Επειδή το Χ μπορεί να υπολογιστεί από μια κανονική κατανομή με μέσο όρο $\nu\bar{\rho}$ και διακύμανση νρ (1-ρ) καταλήγουμε στο ότι το ρ έχει κανονική κατανομή με μέσο όρο $\bar{\rho}$ και διακύμανση ρ(1-ρ) / ν .

ΔΙΑΓΡΑΜΜΑ 5 . 2



Το παραπάνω διάγραμμα είναι επιλογή ενός τυχαίου δείγματος από ένα μεγάλο πληθυσμό με αναλογία του πληθυσμού 0,28 και μέγεθος δείγματος 100 . Από το διάγραμμα αυτό διαπιστώνουμε ότι όσο ψηλώνει και στενεύει η συνάρτηση πυκνότητας τόσο αυξάνεται το μέγεθος του δείγματος , που σημαίνει ότι όσο μεγαλύτερο είναι το ν τόσο οι πιθανότητες για το $\bar{p}$ συγκεντρώνονται πιο κοντά στην τιμή του p .

Ένα μικρό ανησυχητικό σημείο σχετικά με την ανάλυση μας όσον αφορά το διάστημα εμπιστοσύνης των αναλογιών είναι το ενδεχόμενο ο πληθυσμός να μην είναι πολύ μεγαλύτερος από το δείγμα , που σημαίνει ότι η προσέγγισή μας είναι πολύ πιθανόν να ανατραπεί πλήρως .

Ευτυχώς σε πολλές πραγματικές περιπτώσεις ο πληθυσμός είναι πολύ μεγαλύτερος από το δείγμα , με αποτέλεσμα η προσέγγιση να δουλεύει μια χαρά . Αν ο πληθυσμός είναι πολύ μεγαλύτερος βέβαια μπορεί να αγνοηθεί .

ΚΕΦΑΛΑΙΟ 6

ΑΠΛΗ ΤΥΧΑΙΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

6.1 Εισαγωγή

Η απλή τυχαία δειγματοληψία είναι μια από τις μεθόδους δειγματοληψίας . Ονομάζετε δε και επιστημονική δειγματοληψία . Είναι η μέθοδος επιλογής n - μονάδων από ένα πληθυσμό N – μονάδων , με τέτοιο τρόπο ώστε κάθε δυνατό δείγμα μεγέθους n να έχει την ίδια πιθανότητα να επιλέγεί (ΧΑΡΙΣΗΣ – ΚΙΟΧΟΣ , 1997) . Με βάση τον παραπάνω ορισμό , αν ο πληθυσμός αποτελείται από N μονάδες κι εμείς επιθυμούμε ένα δείγμα μεγέθους n , το πλήθος των δυνατών διακεκριμένων δειγμάτων θα είναι:

$$\begin{bmatrix} N \\ n \end{bmatrix} = \frac{N!}{n!(N-n)!} = \frac{N(n-1)......(N-n+1)}{n}$$

όπου n ! = 1 x 2 x 3 x .... x n

Στην πράξη όμως , δεν είναι πάντα εύκολος ο σχηματισμός όλων των δυνατών διακεκριμένων δειγμάτων , κυρίως όταν το μέγεθος N του πληθυσμού είναι πολύ μεγάλο .

Γι' αυτόν τον λόγο , πολλές φορές ακολουθούμε μία εναλλακτική διαδικασία επιλογής μονάδας του πληθυσμού , με τέτοιο τρόπο ώστε να εξασφαλίζει την ίδια πιθανότητα επιλογής σε κάθε μία από τις N μονάδες του πληθυσμού .

Η μέθοδος αυτή περιλαμβάνει τα εξής 2 βήματα:

1 . Αντιστοιχούμε σε κάθε μονάδα του πληθυσμού έναν αριθμό από $1 \rightarrow N$ και κατόπιν

2 . Διαλέγουμε μία σειρά n τυχαίων αριθμών από $1 \rightarrow N$ με τη βοήθεια πινάκων τυχαίων αριθμών . Οι πίνακες τυχαίων αριθμών είναι πίνακες ψηφίων 0 , 1 , 2 , 3 , 4, ... , 9 στους οποίους η πιθανότητα επιλογής σε οποιαδήποτε δοκιμή είναι η ίδια (1/10) για κάθε ψηφίο .

Έτσι , με τη βοήθεια αυτών των πινάκων , εύκολα μπορούμε να διαπιστώσουμε ότι κάθε ένα από τα δυνατά $\begin{bmatrix} N \\ n \end{bmatrix}$ διακεκριμένα τυχαία δείγματα έχει πιθανότητα 1 /$\begin{bmatrix} N \\ n \end{bmatrix}$ να επιλεγεί . Έτσι , κατά την πρώτη επιλογή , η πιθανότητα ότι κάποια από τις n συγκεκριμένες μονάδες του δείγματος θα επιλεγεί από τον δοθέντα πληθυσμό είναι $\frac{n}{N}$. Στη δεύτερη δοκιμή , η πιθανότητα ότι κάποια από τις n-1 απομένουσες μονάδες του δείγματος θα επιλεγεί από τις N-1 απομένουσες μονάδες του πληθυσμού είναι $\frac{n-1}{N-1}$ κ . ο . κ . Άρα , η πιθανότητα με την οποία και οι n μονάδες του επιθυμούμενου δείγματος θα επιλεγούν είναι:

$$\frac{n}{N}\times\frac{n-1}{N-1}\times\frac{n-2}{N-2}\times\ldots\times\frac{1}{N-n+1}=\frac{n!(N-n!)}{N!}$$

$$=\frac{1}{\begin{bmatrix} N \\ n \end{bmatrix}}$$

Παρατηρούμε ότι κάθε φορά που απομακρύνεται μία μονάδα δείγματος από τον πληθυσμό δεν επανατοποθετείται . Για τον λόγο αυτό , η δειγματοληψία αυτή ονομάζεται και απλή τυχαία δειγματοληψία χωρίς επανάθεση (sampling without replacement).

Παράδειγμα: Σχηματική παράσταση

Δυνατοί τρόποι επιλογής μίας μονάδα	N-1	N-2	...N-j+1	1	N-j	....	N-n+1
Δοκιμή	1	2	...j-1	j	j+1	....	N
Διαθέσιμες μονάδες του πληθυσμού	N	N-1	...N-j+2	N-j+1	N-j	....	N-n+1

Από την παραπάνω σχηματική παράσταση είναι εύκολο να διαπιστωθεί ότι η πιθανότητα με την οποία η i μονάδα του πληθυσμού επιλέγεται στην j δοκιμή είναι $\frac{1}{N}$

Έτσι P (να επιλεγεί i μονάδα την j δοκιμή) =

$$\frac{(N-1)\,(N-2)\ldots\ldots(N-j+1)\,(N-j)\ldots(N-n+1)}{N(N-1)\,(N-2)\ldots\ldots(N-j+2)\,(N-j+1)\ldots(\mu-n+1)} = \frac{1}{N}$$

Για παράδειγμα , έστω ότι από τους 4 αριστούχους ισοβαθμίσοντες μαθητές μίας τάξης , την Ελένη (Ε) , τον Γιάννη (Γ) , τον Κώστα (Κ) και τη Μαρία (Μ) θέλουμε να πάρουμε ένα δείγμα μεγέθους 2 , δηλαδή 2 μαθητές . Τότε , ο αριθμός των δυνατών δειγμάτων είναι:

$r = \binom{N}{n} = \binom{4}{2} = 6$ και τα δυνατά αυτά δείγματα είναι τα :

δ1 = { Ε , Γ }
δ2 = { Ε , Κ }
δ3 = { Ε , Μ }
δ4 = { Γ , Κ }
δ5 = { Γ , Μ }
και
δ6 = { Κ , Μ }

Αν λοιπόν το κάθε ένα από τα παραπάνω δείγματα επιλεγεί με πιθανότητα $\frac{1}{6}$ δηλαδή $P_1 = P_2 = \ldots = P_6 = \frac{1}{\binom{4}{2}} = \frac{1}{6}$

τότε κάθε δείγμα αποτελεί απλό τυχαίο δείγμα χωρίς επανάθεση και η σειρά με την οποία επιλέγονται , δεν μας ενδιαφέρει .

6 . 2 Εκτιμήσεις στην απλή τυχαία δειγματοληψία

6 . 2 . 1 Εκτιμήσεις του μέσου όρου του πληθυσμού

Από απλό τυχαίο δείγμα τιμών της μεταβλητής χ, μέγεθος n , λαμβάνονται οι εξής τιμές :

x_1 , x_2 , x_3 , ……, x_n

Με βάση τις τιμές αυτές υπολογίζεται ο μέσος αριθμητικός όρος του δείγματος :

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Υπολογίζεται επίσης και η αμερόληπτη εκτίμηση της διακύμανσης του πληθυσμού με τη σχέση :

$$S^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots (x_n - \bar{x})^2}{n-1}$$

6 . 2 . 2 Εκτίμηση αναλογίας πληθυσμού

Η αναλογία η οποία υπολογίζεται από τα δεδομένα του δείγματος , είναι αμερόληπτη εκτίμηση της αντίστοιχης του πληθυσμού . Το τυπικό σφάλμα της αναλογίας υπολογίζεται ως εξής :

$$Sp = \sqrt{\frac{nP(1-P)}{n-1} \frac{1}{n} \frac{N-n}{n}}$$

Στην πιο πάνω σχέση με P συμβολίζεται η αναλογία του πληθυσμού ή (αν αυτή είναι άγνωστη) του δείγματος. Και πάλι η διόρθωση παραλείπεται, εάν ο λόγος n/N είναι αριθμός πολύ μικρός. Επίσης, ο παρανομαστής n-1 μπορεί να αντικατασταθεί με n, εάν η τιμή της n είναι σχετικά μεγάλος αριθμός.

6.2.3 Εκτίμηση του τυπικού σφάλματος

Η διασπορά του μέσου $\overline{X}_n$ ενός απλού τυχαίου δείγματος μεγέθους n από πληθυσμό μεγέθους N είναι:

$$V(\overline{X}_n) = \frac{\sigma^2}{n} \cdot \frac{N-n}{n}$$

Η χρησιμότητα του αποτελέσματος του θεωρήματος αυτού για το τυπικό σφάλμα είναι μεγάλη για τους εξής λόγους:

α) Δίνει τη δυνατότητα μέτρησης του βαθμού ακρίβειας της εκτίμησης της μέσης τιμής του πληθυσμού και σύγκρισής του με τον βαθμό ακρίβειας που παρέχει οποιαδήποτε άλλη μέθοδος δειγματοληψίας.

β) Δίνει τη δυνατότητα εκτίμησης του μεγέθους του δείγματος που απαιτείται σε μία δειγματοληπτική έρευνα ώστε να επιτευχθεί ο επιθυμητός βαθμός ακρίβειας.

Βέβαια η γνώση του σ^2 είναι απαραίτητη. Στην πράξη όμως αυτό συμβαίνει σπάνια.

Για τον λόγο αυτό απαιτείται μία εκτίμηση του σ^2 από τα δεδομένα του δείγματος και ως τέτοια συνήθως θεωρείται η τιμή της εκτιμήτριας.

$$\hat{\sigma}^2 = S^{*2} = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x_n}\right)^2}{n-1}$$

στην περίπτωση απλού τυχαίου δείγματος.

Στην απλή τυχαία δειγματοληψία χωρίς επανάθεση από πεπερασμένο πληθυσμό μεγέθους N, η στατιστική συνάρτηση

$$S^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x_n})^2}{n-1}$$

είναι αμερόληπτη εκτιμήτρια της διασποράς σ^2 του πληθυσμού.

Οι στατιστικές συναρτήσεις:

$$\hat{\sigma}^2_{\hat{X}n} = s^2_{\overline{X}n} = \frac{S^2}{n}\frac{N-n}{N} = \frac{S^2}{n}\left(\frac{N-n}{n}\right)$$

και

$$\hat{\sigma}^2_{\hat{y}} = S^2_{\hat{y}} = \frac{N^2S^2}{n}\frac{N-n}{N} = \frac{N^2S^2}{n}\left(\frac{N-n}{n}\right)$$

είναι αμερόληπτες εκτιμήτριες των διασπορών του μέσου $\overline{x_n}$ και της τυχαίας μεταβλητής $\overline{Y} = N \cdot \overline{X}_n$ αντίστοιχα .

(Για την εκτίμηση των τυπικών σφαλμάτων θεωρούμε τις θετικές τετραγωνικές ρίζες των παραπάνω εκφράσεων)

Παράδειγμα :

Για κάποιο αίτημα, μαζεύτηκαν υπογραφές που κάλυψαν 676 σελίδες . Κάθε σελίδα είχε χώρο για 42 υπογραφές , αλλά σε πολλές από τις σελίδες υπήρχε αριθμός υπογραφών διαφορετικός του 42 . Ένα απλό τυχαίο δείγμα 50 σελίδων επελέγη και ο αριθμός των υπογραφών ανα σελίδα καταγράφηκε . Ο παρακάτω πίνακας συχνότητας συνοψίζει τα αποτελέσματα . (Xi = αριθμός υπογραφών , Vi – αριθμός σελίδων με Xi υπογραφές) .

Xi	42	41	36	32	29	27	23	19	16	15
Vi	23	4	1	1	1	2	1	1	2	2
Xi	14	11	10	9	7	6	5	4	3	Σύνολο
Vi	1	1	1	1	1	3	2	1	1	50

Να εκτιμηθεί ο συνολικός αριθμός των υπογραφών που μαζεύτηκαν στις 676 σελίδες και να υπολογισθεί το τυπικό σφάλμα της εκτίμησης αυτής .

Λύση

Έχουμε $n = \sum \nu_i = 50$

$$\sum \nu_i x_i = 1471$$

$$\sum \nu_i x_i^2 = 54497$$

Άρα $\overline{x}_n$ = 1471/50 = 29.42 και επομένως η τιμή της εκτιμήτριας $\overline{\mathrm{Y}}$ του συνολικού αριθμού υπογραφών y είναι :

$$\overline{Y} = N \cdot \overline{x_n} = 676 \cdot 29.42 = 19888$$

Το τυπικό σφάλμα της Y είναι ίσο με

$$S\overline{y} = \frac{N.S}{\sqrt{n}}\sqrt{\frac{N-n}{N}}$$

Όπου

$$S = \sqrt{\frac{1}{n-1}\sum (x_i - \overline{x}_n)^2} = \sqrt{\frac{1}{n-1}\left[\sum v_i x_i^2 - \frac{\left(\sum v_i x_i\right)^2}{n}\right]} = \sqrt{229} = 15.13$$

$$S_{\hat{y}} = \frac{(676)(15.13)}{\sqrt{50}}\sqrt{\frac{676-50}{676}} = 1391.9$$

(Το τυπικό σφάλμα του μέσου του δείγματος είναι

$$S_{\overline{X}_n} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N}} = \frac{15.13}{\sqrt{50}}\sqrt{\frac{676-50}{676}} = 2.05$$

6.2.4 Άλλες εκτιμήσεις του μέσου $\overline{x}_n$ ενός απλού τυχαίου δείγματος

- Ο μέσος $\overline{x}_n$ ενός απλού τυχαίου δείγματος μεγέθους n από ένα πληθυσμό N είναι αμερόληπτη εκτιμήτρια της μέσης τιμής μ του πληθυσμού .

- Η διασπορά του μέσου $\overline{x}_n$ ενός απλού τυχαίου δείγματος μεγέθους n από έναν πληθυσμό μεγέθους N είναι :

$$V(\overline{x}_n) = \frac{\sigma^2}{n}\frac{N-n}{n}$$

- Το τυπικό σφάλμα του μέσου $\overline{x}_n$ είναι

$$\sigma_{\overline{X}n} = \sigma\sqrt{\frac{\mathrm{N}-n}{n\mathrm{N}}}$$

Παρατήρηση :

Ο λόγος n/N αντιπροσωπεύει την αναλογία δείγματος – πληθυσμού και συνήθως συμβολίζεται με f . Η εισαγωγή του συμβολισμού αυτού στους παραπάνω τύπους , διευκολύνει την απομνημόνευσή τους :

$$V(\overline{x_n}) = \frac{\sigma^2}{n}(1-f)$$

$$\sigma_{\bar{X}n} = \frac{\sigma}{\sqrt{n}}\sqrt{1-f}$$ αντίστοιχα

Παρατήρηση:

Μία αμερόληπτη εκτιμήτρια της τιμής $y = \sum_{i=1}^{N} y_i$ είναι η στατιστική συνάρτηση $\hat{Y} = N\overline{x_n}$ με διασπορά

$V(\hat{Y}) = \frac{N^2\sigma^2}{N}(1-f)$ και τυπικό σφάλμα

$$\sigma_{\hat{y}} = \frac{N\sigma}{\sqrt{n}}\sqrt{1-f}$$

Παρατήρηση :

Είναι γνωστό ότι η διασπορά του μέσου ενός απλού τυχαίου δείγματος μεγέθους n από ένα άπειρο πληθυσμό είναι σ^2 / n . Η διαφορά που υπάρχει στην περίπτωση πεπερασμένου πληθυσμού μεγέθους N είναι ο παράγοντας 1 – f . Ο παράγοντας αυτός ονομάζεται διόρθωση πεπερασμένου πληθυσμού και ο παρανομαστής του είναι **N – 1** , αν τα αποτελέσματα αναφέρονται στην έκφραση $\frac{1}{N}\Sigma(y_i-\mu)^2$ για την διασπορά σ^2 του πληθυσμού . Προφανώς , όταν ο λόγος **f = n/N** είναι μικρός , η διόρθωση 1 – f είναι κοντά στο 1 και το μέγεθος N του πληθυσμού δεν επηρεάζει τη διασπορά της εκτιμήτριας Xn του μ.

Για παράδειγμα , αν δύο πληθυσμοί μεγέθους N_1=200.000 και N_2 = 200.000 αντίστοιχα έχουν την ίδια διασπορά , σ^2 , ένα απλό τυχαίο δείγμα μεγέθους 500 από κάθε ένα από τους δύο πληθυσμούς θα οδηγήσει σε εκτίμηση της μέσης τιμής του κάθε ενός με το ίδιο περίπου τυπικό σφάλμα .

Στην πράξη η διόρθωση 1-f αγνοείται όταν $f \leq 0{,}05$ και πολλές φορές ακόμη και αν f = 0 , 10 . Το αποτέλεσμα είναι να υπερεκτιμάται το τυπικό σφάλμα της εκτιμήτριας $\overline{x}_n$.

6.3 Διαστήματα Εμπιστοσύνης

Η ισχύς της Κανονικής Προσέγγισης

Η κανονική κατανομή είναι η οριακή μορφή της κατανομής του μέσου $\overline{x}_n$ ενός τυχαίου δείγματος μεγέθους n το οποίο προέρχεται από έναν άπειρο πληθυσμό με πεπερασμένη διασπορά , όταν το n τείνει στο ∞ . Αν δηλαδή , μ και σ^2 συμβολίζουν την μέση τιμή και την διασπορά του πληθυσμού αντίστοιχα , τότε όταν το δείγμα είναι αρκετά μεγάλο ($n \to \infty$) ισχύει :

$$\overline{x}_n \cong N(\mu, \frac{\sigma^2}{n}) \Leftrightarrow$$

$$\frac{\overline{X}n - \mu}{\sigma / \sqrt{n}} \cong N(0.1)$$

Το πόσο μεγάλο πρέπει να είναι το μέγεθος του δείγματος n ώστε η προσέγγιση της πραγματικής κατανομής $\overline{X}n$ από την κανονική κατανομή να είναι ικανοποιητική δεν καθορίζεται από κάποιο γενικό κανόνα . Στις περισσότερες εφαρμογές το n δεν συνηθίζεται να είναι μικρότερο του 30 ($n \geq 30$).

Το παραπάνω αποτέλεσμα είναι γνωστό ως κεντρικό οριακό θεώρημα . Η πρακτική αξία του θεωρήματος αυτού είναι μεγάλη εξαιτίας των δυνατοτήτων που δίνει στον ερευνητή όσον αφορά την συναγωγή στατιστικών συμπερασμάτων .

Το γεγονός $P\left|\frac{x-\mu}{\sigma/\sqrt{n}}\right|\langle Z_1-\frac{\alpha}{2}=1-\alpha$

Όπου $Z_1-\frac{\alpha}{2}$ συμβολίζει το (1–α/2) – ποσοστιαίο σημείο της N(0,1) οδηγεί αμέσως στο συμπέρασμα ότι στο 100(1-α)% των περιπτώσεων , η πραγματική τιμή του μ ανήκει στο διάστημα με άκρα

$$\overline{x_n} \pm Z_1-\frac{\alpha}{2}\frac{\sigma}{\sqrt{n}}$$

Στην περίπτωση που η διασπορά σ^2 του πληθυσμού είναι άγνωστη και χρησιμοποιείται η εκτίμησή της S^2 , η κατανομή του λόγου $\frac{\overline{x_n}-\mu}{\sigma/\sqrt{n}}$ είναι η t , με n-1 βαθμούς ελευθερίας και τα άκρα του 100(1-α)% διαστήματος εμπιστοσύνης είναι :

$$\overline{X}n \pm t_{n-1,1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}$$

Εδώ $t_{n-1,\ 1-\alpha/2}$ συμβολίζει το (1-α)/2) – ποσοστιαίο σημείο της κατανομής t με n-1 βαθμούς ελευθερίας .

Μάλιστα , στην περίπτωση δειγματοληψίας χωρίς επανάθεση από έναν πεπερασμένο πληθυσμό μεγέθους N έχει επίσης αποδειχθεί η ισχύς της κανονικής προσέγγισης της κατανομής του μέσου $\overline{X}n$ κάτω από ορισμένες συνθήκες . Και πάλι , το ερώτημα που αντιμετωπίζει ο ερευνητής είναι πόσο μεγάλο πρέπει να είναι το μέγεθος του δείγματος ώστε η κανονική προσέγγιση να είναι ικανοποιητική .

Για τους πληθυσμούς με μακριά δεξιά ουρά , ένας πρόχειρος κανόνας είναι $n>25\ G^2_1$

Όπου η τιμή G_1 είναι γνωστή ως μέτρο ασυμμετρίας του Fisher και ορίζεται από τη σχέση

$$G1=\frac{1}{N\sigma^3}\sum_{i=1}^{N}(yi-\mu)^3 \Leftrightarrow$$

$$G1=\frac{1}{N\sigma^3}\left\{\sum_{i=1}^{N}y_i^3-3\mu\sum_{i=1}^{N}y_i^2+2\mu^3\right\}$$

Ο κανόνας αυτός στηρίζεται στην υπόθεση ότι οποιεσδήποτε ροπές της κατανομής τάξης μεγαλύτερης του 3 παίζουν αμελητέο ρόλο στον καθορισμό του n . Έτσι υπολογίζοντας το G1 ενός συγκεκριμένου πληθυσμού , μπορούμε να έχουμε μία ιδέα σχετικά με το απαιτούμενο δειγματικό μέγεθος .

6.4 Εκτίμηση Ποσοστών

Η εκτίμηση ποσοστών εμπίπτει στην εκτίμηση των ποιοτικών χαρακτηριστικών μίας δειγματοληψίας .

Για παράδειγμα , πολλές δειγματοληπτικές έρευνες έχουν ως σκοπό την εκτίμηση του ποσοστού π .χ . ανεργίας , τροχαίων κ. λ . π .

Στα πλαίσια τέτοιας μορφής προβλημάτων , οι μονάδες ενός πληθυσμού N ταξινομούνται σε δύο κατηγορίες :

A αν έχουν το χαρακτηριστικό που μας ενδιαφέρει και A' αν όχι .

Ο ακριβής αριθμός των μονάδων που ανήκουν στην κατηγορία A δεν είναι γνωστός , έστω ότι είναι N_A

Άγνωστο είναι και το ποσοστό $p = \frac{N_A}{n}$ των μονάδων αυτών .

Έστω πληθυσμός μεγέθους N και έστω p το ποσοστό των μονάδων που ανήκουν σε κάποια κατηγορία A . Ας υποθέσουμε ότι θα θέλαμε να εκτιμήσουμε την παράμετρο p με βάση τις πληροφορίες (εκτιμήσεις) ενός απλού τυχαίου δείγματος μεγέθους n .

Η στατιστική συνάρτηση που χρησιμεύει εδώ είναι n :

$\hat{p} = \frac{X}{n}$ όπου X : ο αριθμός των μονάδων του δείγματος που ανήκουν στην κατηγορία A .

Για τον υπολογισμό του τυπικού σφάλματος της εκτιμήτριας $\hat{p}$ είναι δυνατόν να εφαρμοστεί η θεωρία του μέσου $\overline{x_n}$ ενός απλού τυχαίου δείγματος με την εισαγωγή των εξής :

Έστω $y_i = \begin{cases} \text{αν η i μονάδα πληθυσμού} \in A \\ 0 \text{ διαφορετικά} \end{cases}$, i=1,2,3,...,N

$x_i = \begin{cases} \text{αν η i μονάδα πληθυσμού} \in A \\ 0 \text{ διαφορετικά} \end{cases}$, i=1,2,..., N

Τότε έχουμε :

$$N_A = \sum_{i=1}^{N} y_i \qquad X = \sum_{i=1}^{n} x_i$$

$$\rho = \frac{\sum y_i}{N} = \mu, \qquad \hat{p} = \frac{\sum X_i}{n} = \overline{x_n}$$

Οπότε το πρόβλημα εκτίμησης του p και του N_A είναι ισοδύναμο με το πρόβλημα εκτίμησης της μέσης τιμής και του συνολικού ύψους των τιμών ενός πληθυσμού , του οποίου οι τιμές των μονάδων είναι 0 και 1 .

$$\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} y_i = N_A = Np$$

Προφανώς :

$$\sum_{i=1}^{N} X_i^2 = \sum_{i=1}^{N} X_i = X = np$$

Άρα η διασπορά του πληθυσμού είναι ίση με :

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - p)^2 = \frac{1}{N-1}\left(\sum_{i=1}^{N} y^2 - Np^2\right) = \frac{Np - Np^2}{N-1} = \frac{N}{N-1}p(1-p)$$

και μία εκτιμήτριά της παρέχεται από τη στατιστική συνάρτηση :

$$S^2 = \frac{n}{n-1}\hat{p}(1-\hat{p})$$

- Η στατιστική συνάρτηση $\hat{p}=X/n$ είναι αμερόληπτη εκτιμήτρια της παραμέτρου p και ισχύει :

$$V(\bar{p}) = \frac{p(1-p)}{n} \cdot \frac{N-n}{N-1}$$

- Η στατιστική συνάρτηση $\hat{Y}_A = N\hat{p}$ είναι μία αμερόληπτη εκτιμήτρια του N_A με διασπορά

$$V(\hat{y}_A) = \frac{N^2 p(1-p)}{n} \cdot \frac{N-n}{N-1}$$

- Η στατιστική συνάρτηση $S_{\hat{p}}^2 = \frac{\hat{p}(1-\hat{p})}{n-1} \cdot \frac{N-n}{n-1}$ είναι μία αμερόληπτη εκτιμήτρια της $V(\hat{P})$.

- Η στατιστική συνάρτηση $S_{\hat{Y}_A} = \frac{N(N-n)}{n-1}\hat{p}(1-\hat{p})$ είναι μία αμερόληπτη εκτιμήτρια της $V(\hat{Y}_A)$.

Για καλύτερη κατανόηση των παραπάνω , παραθέτουμε το εξής

Παράδειγμα :

Από τους 3042 φοιτητές ενός πανεπιστημίου επελέγη ένα απλό τυχαίο δείγμα 200 φοιτητών . Αν 38 από τους φοιτητές αυτούς ήταν υπέρ της εξωτερικής πολιτικής της χώρας τους , να εκτιμηθεί ο συνολικός αριθμός των φοιτητών που υποστηρίζουν την εξωτερική πολιτική της χώρας τους . Να υπολογισθεί το τυπικό σφάλμα της εκτίμησης .

Λύση :

N=3042 n=200 x=38

Άρα $\hat{p} = \frac{X}{n} = \frac{38}{200} = 0.19$

$\hat{Y}_A = N \cdot \hat{p} = 3042 \cdot 0.19 = 577.98$

$$S_{\hat{Y}_A}^2 = \frac{N(N-n)}{n-1}\hat{p}(1-\hat{p}) = \frac{(3042)(3042-200)(0{,}19)(1-0.19)}{200-1} =$$

$$= \frac{(3042)(2842)(0{,}19)(0.81)}{199} = 6686.03$$

Άρα $S_{\hat{Y}_A} = \sqrt{6686.03} = 81{,}8$

6.5 Απλή τυχαία δειγματοληψία με επανάθεση

Εκτός από την απλή τυχαία δειγματοληψία χωρίς επανάθεση έχουμε και την απλή τυχαία δειγματοληψία με επανάθεση (Sampling with replacement) . Σ' αυτήν την περίπτωση κάθε μέλος που επιλέγεται επανατοποθετείται στον πληθυσμό και επομένως μπορεί να επιλεγεί και άλλες φορές .

Χρησιμοποιούμε τη δειγματοληψία με επανάθεση όταν έχουμε πολύπλοκες περιπτώσεις , γιατί οι μαθηματικοί τύποι είναι ευκολότεροι .

Ας υποθέσουμε ότι από πλαίσιο 1000 ονομάτων πρέπει να επιλέξουμε δείγμα 50 ονομάτων . Το κλάσμα δειγματοληψίας είναι $n/N=50/1000=1/20$. Δηλαδή , σε κάθε 20 ονόματα ένα πρέπει να επιλέγεται . Σε 20 σφαιρίδια , γράφουμε 1 ,2 ,3... ,20 (έναν αριθμό σε κάθε σφαιρίδιο) και τα τοποθετούμε σε κάποιο κληροδόχο . Παίρνουμε ένα σφαιρίδιο «στην τύχη» και υποθέτουμε ότι σ' αυτό είναι γραμμένος ο αριθμός 11 . Το δείγμα των 50 ονομάτων σχηματίζεται αν πάρουμε τα ονόματα που κατέχουν την εξής σειρά στο πλαίσιο: [11 ,31 ,51 ,71 ,......971 ,991] .

Αυτός ο τρόπος επιλογής εξασφαλίζει ίση πιθανότητα επιλογής μόνο σε 20 δείγματα των 50 ονομάτων .

Συμπερασματικά , μπορούμε να πούμε ότι η απλή τυχαία δειγματοληψία (με επανατοποθέτηση ή μη) είναι μία μέθοδος επιλογής δείγματος , η οποία εξασφαλίζει ίσες ευκαιρίες στα μέλη ενός πεπερασμένου πληθυσμού να επιλεγούν στο δείγμα . Με αποτέλεσμα , το δείγμα που θα επιλεγεί να εκφράζει όλα τα χαρακτηριστικά του πληθυσμού . Αυτός ο τρόπος δειγματοληψίας είναι γνωστός και ως δειγματοληψία χωρίς περιορισμό (unrestricted sampling) , σε αντίθεση με την απλή τυχαία δειγματοληψία , η οποία διεξάγεται με τον περιορισμό ότι καμία μονάδα του πλαισίου δεν μπορεί να βρεθεί στο δείγμα περισσότερες φορές από 1 (restricted sampling) .

Συνοψίζοντας , η δειγματοληψία χωρίς επανατοποθέτηση είναι προτιμότερη , διότι δίνει εκτιμήσεις με δειγματοληπτικό σφάλμα μικρότερο από το αντίστοιχο της δειγματοληψίας με επανάθεση .

6.6 Πλεονεκτήματα-Μειονεκτήματα απλής τυχαίας δειγματοληψίας

6.6.1 Πλεονεκτήματα

1. Είναι μία τεχνική εύκολη στην κατανόηση δεδομένου ότι οι μέθοδοι (τρόποι) επιλογής δεδομένων όπως ο πίνακας των τυχαίων αριθμών , ο κλήρος κ . τ . λ . έχουν απλό εννοιολογικό σχεδιασμό , με αποτέλεσμα να είναι η μέθοδος αυτή προσιτή ακόμα και σε εκείνους που έχουν χαμηλό στατιστικό υπόβαθρο .
2. Τα αποτελέσματα που προκύπτουν κατά την τυχαία δειγματοληψία είναι αντικειμενικά . Το κάθε δείγμα λαμβάνεται με κλήρωση , μειώνοντας έτσι την πιθανότητα της μεροληψίας και καθιστώντας την υποκειμενική φύση της επιλογής των ερωτηθέντων ασήμαντη .
3. Το δείγμα είναι τυχαίο και επομένως μπορεί να υπολογισθεί το μέγεθος του δειγματοληπτικού σφάλματος .
4. Τα αποτελέσματα της απλής τυχαίας δειγματοληψίας είναι σε μεγάλο ποσοστό σωστά στην περίπτωση που ο πληθυσμός είναι σχετικά μικρός .

6.6.2 Μειονεκτήματα

1 . Η μέθοδος αυτή είναι αρκετά χρονοβόρα όταν ο υπό μελέτη πληθυσμός είναι μεγάλος και γεωγραφικά διασκορπισμένος .
2 . Είναι δαπανηρή , επειδή απαιτεί πλήρως ενημερωμένα πλαίσια . Είναι γενικά απαραίτητο να κατασκευασθεί κατάλογος και να αριθμηθεί κάθε μονάδα του συνόλου του πληθυσμού .
3. Το κόστος που απαιτείται για τη συλλογή δεδομένων είναι μεγάλο.
4 . Το δείγμα είναι μη αντιπροσωπευτικό , αν ο πληθυσμός που εξετάζεται δεν είναι αρκετά ομοιογενής . Οι εκτιμήσεις δηλαδή του δειγματοληπτικού σχεδίου έχουν μειωμένη ακρίβεια .

ΚΕΦΑΛΑΙΟ 7

Στρωματοποιημένη τυχαία δειγματοληψία

7.1 Εισαγωγή

Η στρωματοποιημένη τυχαία δειγματοληψία (stratified random sampling) ή κατά στρώματα δειγματοληψία είναι μία μέθοδος τυχαίας δειγματοληψίας κατά την οποία πρώτα διαιρούμε τον πληθυσμό σε ένα συγκεκριμένο αριθμό υποπληθυσμών , στη βάση κάποιου κοινού χαρακτηριστικού – όπως είναι το φύλο , η ηλικία , ο τόπος διαμονής κ.τ.λ. – και στη συνέχεια επιλέγουμε ένα τυχαίο δείγμα από κάθε υποπληθυσμό . Το σύνολο των στοιχείων όλων των επιμέρους δειγμάτων αποτελεί το δείγμα . Οι υποπληθυσμοί , που ονομάζονται στρώματα (strata) είναι ξένοι μεταξύ τους , δηλαδή δεν έχουν κοινά στοιχεία και το σύνολο των στοιχείων τους αποτελούν τον πληθυσμό . Επίσης , κάθε επιμέρους στρώμα (υποσύνολο) πρέπει να συμμετέχει με την ποσοστιαία αναλογία που έχει και στον πληθυσμό . Για παράδειγμα , αν θέλουμε να πάρουμε ένα δείγμα 1000 καθηγητών Δευτεροβάθμιας Εκπαίδευσης για μία έρευνα , τότε έχοντας όλους τους καθηγητές σε ειδικότητες (στρώματα) με ποσοστό 35% φιλολόγους , 20% μαθηματικούς , 20% φυσικούς , 3% θεολόγους κ.λ.π θα πρέπει και στα δείγματά μας οι ειδικότητες να συμμετέχουν με το ίδιο ποσοστό , δηλαδή σ' αυτό θα πρέπει να υπάρχουν 350 φιλόλογοι , 200 μαθηματικοί , 200 φυσικοί κ.λ.π . Για να γίνει πιο κατανοητό θα θέσουμε 2 ορισμούς:

Ορισμός 1

Έστω ότι από κάθε ένα από τα στρώματα ενός πληθυσμού επιλέγεται ένα απλό τυχαίο δείγμα μεγέθους n_i , i=1 ,2 ,... . , κ ανεξάρτητα από τα άλλα . Το δείγμα μεγέθους $n = n_1 + n_2 + \ldots + n_\kappa$ που προκύπτει από την ένωση κ ανεξάρτητων απλών τυχαίων δειγμάτων ονομάζεται στρωματοποιημένο τυχαίο δείγμα και η διαδικασία επιλογής του ονομάζεται στρωματοποιημένη τυχαία δειγματοληψία .

Ορισμός 2

Υποθέτουμε ότι έχουμε έναν πληθυσμό μεγέθους N που μπορεί να διαιρεθεί σε κ εσωτερικά ομοιογενείς υποπληθυσμούς μεγέθους N_1 , N_2 ,...,N_κ . Αν αυτοί είναι ξένοι μεταξύ τους ώστε να ισχύει $N_1+N_2+...+N_\kappa=N$, τότε οι υποπληθυσμοί αυτοί θα ονομάζονται στρώματα (strata) .

<u>**Παράδειγμα**</u>: Υποθέτουμε ότι σκοπεύουμε να εκτιμήσουμε το χρόνο που διαθέτει ένας μέσος σπουδαστής της σχολής Α για τη μελέτη των μαθημάτων του . Παρατηρώντας ότι ο χρόνος αυτός διαφέρει σχετικά λίγο μεταξύ των σπουδαστών του ίδιου έτους σπουδών , χωρίζουμε τους σπουδαστές της σχολής Α σε τόσα στρώματα όσα είναι τα έτη σπουδών . 4000 σπουδαστές χωρίζονται σε 4 έτη σπουδών (στρώματα) ως εξής:

Έτος σπουδών (στρώμα)	Πληθυσμός σπουδαστών / έτος σπουδών (πληθυσμός κατά στρώματα)
Α	1500
Β	1000
Γ	800
Δ	700
Σύνολο	4000

Σχηματίζουμε δείγμα από κάθε στρώμα (έτος σπουδών) με κλάσμα δειγματοληψίας 10% που είναι ίδιο για όλα τα στρώματα . Έτσι , το συνολικό δείγμα θα αποτελείται από 400 σπουδαστές και θα απαρτίζεται από τα επιμέρους δείγματα των 4 στρωμάτων ως εξής:

Έτος σπουδών (στρώμα)	Πληθυσμός σπουδαστών / έτος σπουδών (πληθυσμός κατά στρώματα)
Α	150
Β	100
Γ	80
Δ	70
Σύνολο	400

Με βάση τα παραπάνω , παρατηρούμε ότι το δείγμα αυτό των 400 σπουδαστών αντιπροσωπεύει τους σπουδαστές όλων των ετών σπουδών . Αντίστοιχο δείγμα απλό τυχαίο από τους 400 σπουδαστές θα ήταν δυνατό να αποτελεστεί μόνο από τους σπουδαστές του Α έτους ή μόνο από τους σπουδαστές του Β έτους κ.ο.κ. Η στρωματοποιημένη δειγματοληψία χωρίζεται σε 2 κατηγορίες:

Α) Ανάλογη Δειγματοληψία κατά στρώματα

Β) Δυσανάλογη Δειγματοληψία κατά στρώματα

7.2 Δειγματοληψία κατά στρώματα ανάλογη (Proportionate)

Την ανάλογη δειγματοληψία κατά στρώματα την εφαρμόζουμε πάντα όταν:

1 . Η διακύμανση της μεταβλητής η οποία μας ενδιαφέρει είναι για όλα τα στρώματα ίδια και

2 . Όταν το κόστος της δειγματοληψίας δε διαφοροποιείται μεταξύ των στρωμάτων

Έτσι , ο τρόπος καθορισμού του μεγέθους των δειγμάτων κατά στρώμα (n_i) ποικίλει .

Το κριτήριο καθορισμού των επιμέρους μεγεθών των δειγμάτων τούτων είναι η μεγιστοποίηση της ακρίβειας των εκτιμήσεων , με τον περιορισμό ότι η συνολική δαπάνη της δειγματοληπτικής έρευνας είναι δοσμένη . Αν η δαπάνη αυτή οδηγεί στον καθορισμό του συνολικού μεγέθους του δείγματος σε n μονάδες (για όλα τα στρώματα μαζί) , η κατανομή των μονάδων τούτων στα επιμέρους στρώματα (n_1 , n_2 ,... , n_K) γίνεται ανάλογα με το μέγεθος των στρωμάτων:

$$\frac{n_1}{N_1}=\frac{n_2}{N_2}=\frac{n_3}{N_3}=\dots\dots=\frac{n_K}{N_K}=\frac{n_1+n_2+n_3+\dots+n_K}{N_1+N_2+N_3+\dots+N_K}=\frac{n}{N}$$

Από τις σχέσεις αυτές παίρνουμε τις εξής :

$$n_i = n\cdot\frac{N_i}{N}$$

Το κλάσμα δειγματοληψίας κατά στρώμα είναι αριθμός σταθερός (n/N ή n_i/n) .

Εισάγοντας τις ανωτέρω τιμές των n_i , στους τύπους με τους οποίους υπολογίζονται τα τυπικά σφάλματα του μέσου $\bar{x}$ και του ποσοστού P

εχουμε:

$$S_{\bar{x}^2}=\sum\left[\frac{N_i}{N}\right]^2\cdot S_i^2\cdot\frac{N-n}{Nn}=\frac{\Sigma n_i S_i^2}{n}\left[\frac{1}{n}-\frac{1}{N}\right]$$

Αν παραλείψουμε τον γνωστό διορθωτικό παράγοντα , έχουμε:

$$S\bar{x}^2=\frac{\Sigma n_i S_i^2}{n^2}$$

Στις πιο πάνω σχέσεις με S_i^2 συμβολίζουμε τη διακύμανση της μεταβλητής X μέσα σε κάθε στρώμα .

Για τη διακύμανση της εκτίμησης της αναλογίας έχουμε :

$$S_p^2=\sum\left[\frac{N_i}{N}\right]^2\cdot\frac{P_i(1-P_i)}{n_i-1}\cdot\left[1-\frac{N_i}{N}\right]=\frac{\sum n_iP_i(1-P_i)}{n^2}$$

Το δεύτερο σκέλος της τελευταίας ισότητας αναφέρεται στην περίπτωση παράλειψης του διορθωτικού παράγοντα .

Παράδειγμα

Από τους σπουδαστές κάθε έτους σπουδών του προηγούμενου παραδείγματος παίρνουμε δείγμα τυχαίο 1% και υπολογίζουμε τη μέση βαθμολογία των μαθημάτων τους (X_i) , τη διακύμανση της βαθμολογίας τους (S_i^2) και την αναλογία των εργαζομένων (P_i) :

Στρώματα (έτος)	Μέγεθος δείγματος n_i	Μέσος όρος $\overline{X}_i$	Διακύμανση S^2_1	Αναλογία σπουδ. P_i
Α	15	6,5	1,2	0,08
Β	10	7,2	1,1	0,15
Γ	8	7,6	1,3	0,20
Δ	7	7,0	1,0	0,40
	40			

Η αμερόληπτη εκτίμηση της μέσης βαθμολογίας του πληθυσμού βρίσκεται με στάθμιση των τεσσάρων μέσων βαθμολογιών :

$$\overline{X} = 6{,}5(15/40) + 7{,}2(10/40) + 7{,}6(8/40) + 7{,}0(7/40) = 6{,}9825$$

Η διακύμανση της κατανομής δειγματοληψίας της $\overline{X}$ (χωρίς τη διόρθωση)

$$S_{\overline{\chi}}^2 = \frac{15(1.2) + 10(1.1) + 8(1.3) + 7(1.0)}{40^2} = \frac{46.4}{1600} = 0.029$$

Και το αντίστοιχο τυπικό σφάλμα είναι $S_{\bar{x}} = 0.17$

Η αμερόληπτη εκτίμηση της αναλογίας των εργαζομένων σπουδαστών στον πληθυσμό υπολογίζεται σύμφωνα με όσα αναφέρουμε για το θέμα αυτό πιο πάνω :

$$P = 0.08(15/40) + 0.15(10/40) + 0.20(8/40) + 0.40(7/40) = 0.1775$$

Η διακύμανση της εκτιμήτριας P υπολογίζεται ως εξής (χωρίς τη διόρθωση)

$$S_p^2 \frac{15(0.08)(0.92) + (10(0.15)(0.85) + 8(0.20)(0.80) + 7(0.40)(0.60)}{40^2} = \frac{5.339}{1600} = 0.003337$$

Και το τυπικό σφάλμα της εκτίμησης P είναι S_p=0.05777.

7.3 Δειγματοληψία κατά στρώματα δυσανάλογη (disproportionate)

Σε αντίθεση με την ανάλογη στρωματοποιημένη δειγματοληψία , η δυσανάλογη στρωματοποιημένη δειγματοληψία εφαρμόζεται όταν:

1 . Το κόστος της δειγματοληψίας (κατά μονάδα του δείγματος) είναι το ίδιο για όλα τα στρώματα και

2 . Η διακύμανση διαφέρει μεταξύ των στρωμάτων . Δηλαδή , από κάθε στρώμα παίρνουμε δείγμα όχι ανάλογο προς το μέγεθος του στρώματος , αλλά ανάλογο προς το γινόμενο του μεγέθους του στρώματος (N_i) και της αντίστοιχης μέσης απόκλισης τετραγώνου (S_i):

$$\frac{n_1}{N_1S_1} = \frac{n_2}{N_2S_2} = \frac{n_k}{N_KS_K} = \frac{n_1+n_2+\dots\dots\dots\dots+n_K}{N_1N_1 + N_2N_2 + \dots N_KS_K \sum \mathrm{N}_i S_i}$$

Από την παραπάνω σχέση έχουμε:

$$n_i = n\frac{N_iS_i}{\sum \mathrm{N}_i S_i}$$

Στη θέση των S_i τοποθετούμε τα αντίστοιχα πληθυσμιακά μεγέθη σ_i. Εισάγουμε και την τιμή της n_i στον τύπο υπολογισμού της διακύμανσης της $\overline{X}$ και έχουμε:

$$Sx^2 = \frac{1}{n}(\Sigma W_iS_i^2) - \frac{1}{N}(\Sigma W_i S_i^2)$$

όπου $W_i = N_i / N$

Το τμήμα $(\Sigma W_iS_i^2)/N$ είναι η γνωστή διόρθωση για πληθυσμούς περιορισμένου μεγέθους από τους οποίους η δειγματοληψία γίνεται χωρίς επανατοποθέτηση .

Το τυπικό σφάλμα κατανομής δειγματοληψίας της $\overline{X}$ είναι κατά τα γνωστά $S_{\overline{X}}$ (η αμερόληπτη εκτίμηση του μέσου όρου του πληθυσμού προκύπτει ως σταθμικός μέσος των κατά στρώματα μέσων όρων του δείγματος: $\Sigma W_i \overline{X_i}$) .

Παράδειγμα:

Επανερχόμαστε στο παράδειγμα των σπουδαστών που από τον πληθυσμό τους παίρνουμε δείγμα 40 ατόμων . Η κατανομή τον δείγματος τούτου στα επιμέρους στρώματα και τα λοιπά στοιχεία (διαφορετικά κάπως από τα προηγούμενα) παρουσιάζονται στον ακόλουθο πίνακα:

	N_i	w_i	S_i	N_iS_i	$\overline{X}_i$	$n_i = n(N_iS_i / \Sigma N_iSi)$		W_iS_i	$W_iS_i^2$
Α	1500	0,375	1,0	1500	6,0	11,5	12	0,37500	0,3750
Β	1000	0,250	1,1	1100	7,0	8,4	8	0,2750	0,3025
Γ	800	0,200	1,5	1200	7,5	9,2	9	0,30000	0,4500
Δ	700	0,175	2,0	1400	6,5	10,8	11	0,35000	0,7000
	4000	1.000		5200			40	1,30000	1,8275

Η αμερόληπτη εκτίμηση της μέσης βαθμολογίας είναι:

6 ,0(0 ,375) + 7 ,0(0 ,250) + 7 ,5(0 ,200) + 6 ,5(0 ,175) = 6 ,6375

Το τυπικό σφάλμα της $\overline{X}$ υπολογίζεται με βάση τους προηγούμενους τύπους :

$$S_{\overline{X}}^2 = \frac{1}{40}(1.09375)^2 - \frac{1}{4000}(1.8275) = 0.02945$$

Και το τυπικό σφάλμα της $\overline{X}$ ισούται με $S_{\overline{X}}$= 0 ,176 .

Παρατηρούμε ότι χωρίς το διορθωτικό παράγοντα η διακύμανση της $\overline{X}$ ισούται με 0 ,02991 (ο διορθωτικός παράγοντας είναι 0 ,00046) .

7.4 Εκτίμηση παραμέτρων στρωματοποιημένης δειγματοληψίας

7.4.1 Εκτίμηση μέση τιμής του πληθυσμού

Δειγματοληπτικό σφάλμα εκτιμήσεων δειγματοληψίας κατά στρώματα

1. Εκτίμηση του μέσου όρου του πληθυσμού

Οι μέσοι όροι του δείγματος σε κάθε στρώμα , αφού σταθμισθούν με το σχετικό μέγεθος του αντίστοιχου στρώματος , παρέχουν αμερόληπτη εκτίμηση του μέσου όρου ολόκληρου τον πληθυσμού:

$$E\left[\frac{N_1}{N}\cdot\overline{X}_i+\frac{N_2}{N}\cdot X_2+\ldots\ldots+\frac{N_K}{N}\cdot\overline{X}_K\right]=E\left[\overline{X}\right]=\mu$$

Στην πιο πάνω σχέση με τα σύμβολα N_1 , N_2 ,... . . , N_K εκφράζονται το μέγεθος κάθε στρώματος στον πληθυσμό , με N το μέγεθος ολόκληρου του πληθυσμού και με $\overline{X}_1$, $\overline{X}_2$... . . , $\overline{X}_K$ οι μέσοι όροι της μεταβλητής X από τις τιμές του δείγματος μέσα σε κάθε στρώμα (οι υποδείκτες 1 ,2 , ... , K συμβολίζουν κατά σειρά το πρώτο , το δεύτερο , . . . , το τελευταίο από τα K στρώματα του πληθυσμού) . Η διακύμανση του σταθμικού αθροίσματος των μέσων όρων των δειγμάτων (ένα δείγμα από κάθε στρώμα); δηλαδή της X , δίνεται από τη σχέση:

$$S_{\overline{X}}^{\,2}=\left[\frac{N_1}{N}\right]^2\cdot\sigma^2_{\overline{X}_1}+\left[\frac{N_2}{N}\right]^2\cdot\sigma^2_{\overline{X}_2}+\ldots+\left[\frac{N_K}{N}\right]^2\cdot\sigma^2_{\overline{X}k}$$

$$=\left[\frac{N_1}{N}\right]^2\cdot\frac{\sigma^2_{x1}}{n_1}+\left[\frac{N_2}{N}\right]^2\cdot\frac{\sigma^2_{X_2}}{n_2}+\ldots+\left[\frac{N_K}{N}\right]^2\cdot\frac{\sigma^2_{\overline{X}k}}{n_K}$$

Στις πιο πάνω σχέσεις με σ^2 συμβολίζεται η διακύμανσή της $\overline{X}$ ή της X κάθε στρώματος και με n_1 , n_2... .n_K , το μέγεθος του δείγματος από κάθε στρώμα . Αν η σ^2 για τη X κάθε στρώματος είναι άγνωστη (και τούτο συμβαίνει συνήθως) , στη θέση της τοποθετούμε την S^2 , όπως την υπολογίζουμε από τα στοιχεία των δειγμάτων . Επίσης , παρατηρούμε ότι από τις πιο πάνω σχέσεις έχει παραλειφθεί ο γνωστός διορθωτικός παράγοντας .

Αν , συνεπώς , λάβουμε υπ'όψει τις τιμές S^2 και τους διορθωτικούς παράγοντες , η $S_{\bar{x}}^2$ παίρνει την ακόλουθη διατύπωση:

$$S_{\bar{X}}^2 = \left[\frac{N_1}{N}\right]^2 \cdot \frac{Sx_1^2}{n_1} \cdot \frac{N_i - n_1}{N_1} + \dots + \left[\frac{N_K}{N}\right]^2 \cdot \frac{S_{XK}^{\ 2}}{nK} \cdot \frac{N_K - n_K}{N_K} =$$

$$= \frac{1}{N^2} \sum_{i=1}^{i=K} N_i (N_i - n_i) \frac{S_i^2}{n_i}$$

Το τυπικό σφάλμα της $\overline{X}$ ισούται με $S_{\bar{x}}$.

7.4.2 Εκτίμηση της αναλογίας του πληθυσμού

Από τα στοιχεία των δειγμάτων των στρωμάτων υπολογίζονται οι λόγοι P των οποίων το σταθμικό άθροισμα αποτελεί αμερόληπτη εκτίμηση της αντίστοιχης αναλογίας στο πληθυσμό (N) :

$$\Sigma\left[\frac{N_1}{N} \cdot P_1 + \frac{N_2}{N} \cdot P_2 + \dots + \frac{N_K}{N} \cdot P_K\right] = \Sigma(P) = N$$

Η διακύμανση του σταθμικού αθροίσματος των $P_1, P_2, \dots\dots\dots P_K$ υπολογίζεται ως εξής :

$$Sp^2 = \left[\frac{N_1}{N}\right]^2 \cdot \frac{P_1(1-P_1)}{n_1 - 1} \cdot (1 - f_1) + \dots + \left[\frac{N_K}{N}\right]^2 \cdot \frac{P_K(1-P_K)}{n_K - 1} \cdot (1 - f_K)$$

Από την τελευταία σχέση είναι δυνατό να παραλειφθούν οι διορθωτικοί παράγοντες (1-f) και στους παρανομαστές να τεθεί το μέγεθος του δείγματος n_i (στη θέση του όρου $n_i - 1$) . Τούτο , όμως θα συμβεί κάτω από τις γνωστές προϋποθέσεις . Το τυπικό σφάλμα της P ισούται με S_p .

7.5 Βέλτιστος καταμερισμός στρωματοποιημένης δειγματοληψίας

Ο καταμερισμός του συνολικού δειγματικού μεγέθους n στα κ διαθέσιμα στρώματα , δηλαδή ο καθορισμός των τιμών των μεγεθών n_1 ,n_2 ,... . . n_κ των κ απλών τυχαίων δειγμάτων είναι ένα από τα προβλήματα που συνδέεται άμεσα με την τεχνική της στρωματοποιημένης δειγματοληψίας . Αν το δειγματοληπτικό κόστος ανά μονάδα είναι το ίδιο σε όλα τα στρώματα και οι διασπορές των στρωμάτων δε διαφέρουν σημαντικά , τα μεγέθη n_1 ,n_2 ,... . . n_κ επιλέγονται συνήθως με τον εξής τρόπο:

$$\frac{n_i}{N_i} = \frac{n}{N}, \quad i=1,2, \ldots ,\kappa$$

Ο τρόπος αυτός είναι γνωστός ως αναλογικός καταμερισμός του n και η δειγματοληπτική τεχνική είναι γνωστή ως αναλογική στρωματοποιημένη δειγματοληψία .

Στην περίπτωση αυτή:

$$n_i = n\frac{N_i}{N}, i = 1,2,3,\ldots,\kappa$$ που σημαίνει

ότι το μέγεθος του δείγματος από ένα στρώμα είναι ανάλογο του ποσοστού των μονάδων του πληθυσμού που το στρώμα εκπροσωπεί .

Υπάρχουν όμως περιπτώσεις όπου οι τιμές του πληθυσμού έχουν μεγαλύτερη διακύμανση σε μερικά στρώματα από ό , τι σε άλλα . Διαφέρουν δηλαδή σημαντικά οι διασπορές των στρωμάτων . Επομένως για να αντιπροσωπευθούν επαρκώς τα στρώματα αυτά στο δείγμα , θα πρέπει ο λόγος $\frac{n_i}{N_i}$ να είναι ανάλογος της τυπικής απόκλισης σ_i του στρώματος . Αυτό σημαίνει ότι στρώματα με μεγαλύτερη διακύμανση από άλλα πρέπει να εκπροσωπούνται από μεγαλύτερο τμήμα του δείγματος , για να αυξηθεί η ακρίβεια των εκτιμήσεων .

Υποθέτοντας ότι το δειγματοληπτικό κόστος ανά μονάδα είναι το ίδιο για όλα τα στρώματα , αποδεικνύεται ότι η $v(\bar{\mu}_n)$ γίνεται ελάχιστη αν τα n_1 ,n_2 , . . . , n_K επιλεγούν έτσι ώστε:

$$n_i = n\frac{N_i\sigma_i}{\sum_{j=1}^{k} N_j\sigma_j}, i = 1,2,\ldots\ldots,k$$

Ο σχεδιασμός αυτός είναι γνωστός ως βέλτιστος καταμερισμός του n με σταθερό κόστος ανά δειγματοληπτική μονάδα .

Αν το δειγματοληπτικό κόστος ανά μονάδα διαφέρει από στρώμα σε στρώμα , τότε είναι φυσικό να προσπαθήσει ο ερευνητής να αυξήσει την ακρίβεια των εκτιμήσεων του επιλέγοντας τα n_i , , αντιστρόφως ανάλογα των c_i , i =1 . 2... . k .

Έστω ότι το συνολικό κόστος της c μιας δειγματοληψίας είναι συνάρτηση των c_i , i =1 . 2... . k . δηλαδή έστω ότι

$$c = c_o + \sum_{i=1}^{k} n_i c_i \qquad c_o > 0$$

Τότε αποδεικνύεται ότι αν το κόστος C έχει μια δοθείσα τιμή , οι τιμές των n_i που ελαχιστοποιούν την διασπορα $V(\hat{\mu}_n)$δίνονται από τον τύπο :

$$n_i = n\frac{N_i\sigma_i / \sqrt{c_i}}{\sum_{j=1}^{k} N_j\sigma_j / \sqrt{c_j}}, i = 1,2,\ldots,k \qquad (1)$$

Οι τιμές ελαχιστοποιούν επίσης το κόστος εάν η διασπορά $V(\hat{\mu}_n)$ έχει δοθείσα τιμή.

Ο παραπάνω καταμερισμός του n ονομάζετε βέλτιστος καταμερισμός (optimum allocation).

Η σχέση (1) υπονοεί ότι το μέγεθος του απλού τυχαίου δείγματος , που επιλέγεται από ένα στρώμα , πρέπει να είναι μεγαλύτερο από τα δείγματα άλλων στρωμάτων αν το μέγεθος του στρώματος είναι μεγαλύτερο , η διασπορά του στρώματος είναι μεγαλύτερη και το κόστος ανά μονάδα του στρώματος είναι χαμηλότερο .

7.6 Πλεονεκτήματα-Μειονεκτήματα στρωματοποιημένης δειγματοληψίας

7.6.1 Πλεονεκτήματα

1 . Εξασφαλίζει μεγαλύτερη ακρίβεια , αφού τα δεδομένα σε κάθε στρώμα είναι περισσότερο ομοιογενή (σε σχέση με την απλή τυχαία δειγματοληψία) απ' ότι σε ολόκληρο τον πληθυσμό . Αυτό σημαίνει μικρότερη διασπορά στην εκτίμηση των παραμέτρων .
2 . Κάθε στρώμα έχει μικρότερο μέγεθος σε σχέση με τον πληθυσμό, με αποτέλεσμα να είναι πιο απλό να επιλεγεί αντιπροσωπευτικό δείγμα και να γίνει η συλλογή πληροφοριών .
3 . Εξασφαλίζει την αύξηση της ακρίβειας του σχεδίου , διατηρώντας σταθερό το μέγεθος του δείγματος που οδηγεί στην μείωση της τιμής δειγματοληπτικού σφάλματος . Σκοπός εδώ είναι η βελτίωση της αντιπροσωπευτικότητας του δείγματος , με αποτέλεσμα το κόστος της δειγματοληψίας να μειωθεί σημαντικά .
4 . Η στρωματοποίηση διευκολύνει την αντιμετώπιση κάποιων διαφορών που μπορεί να υπάρχουν μεταξύ των πληθυσμιακών ομάδων π.χ υπάρχει διαφορά στο σύνολο των οικονομικών δραστηριοτήτων μεταξύ των μεγάλων και μικρών επιχειρήσεων .

7.6.2 Μειονεκτήματα

1 . Η διαδικασία που ακολουθείται για την επιλογή ενός δείγματος είναι πιο περίπλοκη .
2 . Απαιτείται περισσότερη προκαταρτική έρευνα για το σύνολο του πληθυσμού που πρόκειται να μελετηθεί .
3 . Το κόστος δειγματοληπτικού σφάλματος είναι υψηλότερο , εάν το δείγμα δεν παραμείνει σταθερό .

4 . Το σύνολο του πληθυσμού πρέπει να πληρεί και την ποσοστιαία αναλογία που θα έχει το όλο δείγμα προς τον πληθυσμό . Πρέπει δηλαδή να κατατάσσεται βάσει των χαρακτηριστικών της στρωματοποίησης .

7.7 Σύγκριση απλής – τυχαίας και στρωματοποιημένης δειγματοληψίας

Αν ο πληθυσμός που εξετάζεται δεν είναι αρκετά ομογενείς , η απλή-τυχαία δειγματοληψία ενδέχεται να δώσει δείγμα ακραίο (μη αντιπροσωπευτικό) . Αντίθετα , η στρωματοποιημένη δειγματοληψία μας δίνει πιο ακριβή αποτελέσματα , μειώνοντας τη διασπορά για την εκτιμήτρια της μέσης τιμής του πληθυσμού . Αυτό βέβαια δε σημαίνει ότι πάντα ένα στρωματοποιημένο δείγμα μας δίνει μικρότερη διασπορά απ' ό ,τι ένα απλό – τυχαίο δείγμα .
Η αύξηση ή η μείωση της διασποράς εξαρτάται από τον σχεδιασμό του βέλτιστου καταμερισμού . Επιπλέον , ο διαχωρισμός του πληθυσμού σε στρώματα (υποπληθυσμούς , υποσύνολα) ξένα μεταξύ τους πρέπει να γίνεται με τρόπους που σχετίζονται με το εξεταζόμενο χαρακτηριστικό , πράγμα το οποίο δεν είναι πάντοτε εύκολο , αφού μερικά χαρακτηριστικά είναι δύσκολο να ταυτοποιηθούν για κάθε μονάδα του πληθυσμού .
Συγκρίνοντας τους δύο τρόπους δειγματοληψίας , παρατηρούμε ότι η στρωματοποιημένη δειγματοληψία , εάν και πιο σύνθετη στη λειτουργία , χρησιμοποιείται περισσότερο από ό ,τι η απλή- τυχαία δειγματοληψία . Αυτό οφείλεται στην αύξηση ακρίβειας που προσφέρει και στην αποφυγή διαχωρισμού του πληθυσμού που σε πολλές περιπτώσεις είναι ήδη χωρισμένος σε στρώματα . Με αυτόν τον τρόπο και σε αυτές τις περιπτώσεις κερδίζουμε χρόνο και χρήμα .

ΚΕΦΑΛΑΙΟ 8

Συστηματική δειγματοληψία (Systematic Sampling)

8.1 Εισαγωγή

Η συστηματική δειγματοληψία αποτελεί έναν πολύ απλό τρόπο επιλογής δείγματος . Η επιλογή δείγματος γίνεται με την εφαρμογή κάποιου κατάλληλου δειγματοληπτικού πλαισίου , δηλαδή καταλόγου , ξεκινώντας από κάποιο τυχαίο αρχικό σημείο και επιλέγοντας μία μονάδα κάθε κ (κ>ο) μονάδες μέχρι να κατασκευαστεί το δείγμα με το αντίστοιχο μέγεθος . Η συστηματική δειγματοληψία είναι δηλαδή μία τεχνική η οποία εισάγει ένα συστηματικό στοιχείο κατά τη διαδικασία επιλογής των μονάδων του πληθυσμού . Έστω ότι έχουμε έναν πληθυσμό με Ν στοιχεία και θέλουμε να πάρουμε ένα δείγμα μεγέθους ν . Χωρίζουμε τον πληθυσμό μεγέθους Ν σε r ισοπληθείς ομάδες για να πάρουμε τυχαία ένα άτομο .

Παράδειγμα : Έστω ότι θέλουμε να μετρήσουμε τους μαθητές ενός σχολείου που θα εξεταστούν στα μαθηματικά . Ο κατάλογος των ερωτήσεων θα περιλαμβάνει 600 ερωτήσεις από τις οποίες θα επιλεγούν οι 20 ερωτήσεις . Όλες οι ερωτήσεις είναι απαραίτητα αριθμημένες από 1-600 . Χωρίζουμε λοιπόν τις 600 ερωτήσεις σε 20 ισοπληθείς ομάδες (600:20=30) . Κατόπιν , με κλήρωση ή με πίνακα τυχαίων αριθμών παίρνουμε μία ερώτηση από την πρώτη ομάδα ερωτήσεων . Υποθέτουμε ότι κληρώθηκε ο αριθμός 17 . Στη συνέχεια , παίρνουμε και τις άλλες ερωτήσεις χωρίς άλλη κλήρωση. Δηλαδή , παίρνουμε τις ερωτήσεις 17 ,47 ,77 ,107...587 . Το δείγμα που επιλέγεται μ' αυτόν τον τρόπο λέγεται <u>συστηματικό δείγμα</u> . Αν όμως οι μονάδες αυτές του πληθυσμού εμφανίζονται με τυχαία σειρά στον κατάλογο ή στη λίστα από όπου επιλέγεται το δείγμα , τότε θα ονομάζεται ψευδοτυχαίο . Επιπλέον , η μέθοδος αυτή εξασφαλίζει την ίδια πιθανότητα επιλογής σε κάθε μονάδα του πληθυσμού με αποτέλεσμα να μας οδηγεί σε αμερόληπτες εκτιμήσεις της μέσης τιμής του πληθυσμού .

8.2 Εκτίμηση του μέσου της συστηματικής δειγματοληψίας

8.2.1 Η διασπορά της εκτιμήτριας του μέσου

Αν υποθέσουμε ότι Ν = nκ , τότε τα κ δυνατά 1 – ανά – κ συστηματικά δείγματα θα είναι οι στήλες του παρακάτω πίνακα :

Πίνακας 8 . 2 . 1

Σύνθεση των k δυνατών συστηματικών δειγμάτων μεγέθους n από έναν πληθυσμό μεγέθους N=nk .

		Δείγμα			
1	2	. . .	i	. . .	k
y_1	y_2	. . .	y_i	. . .	y_k
y_{k+1}	y_{k+2}	. . .	y_{k+i}	. . .	y_{2k}
.	.	. . .	.	. . .	.
.	.	. . .	.	. . .	.
.	.	. . .	.	. . .	.
$y_{(n-1)k+1}$	$y_{(n-1)k+2}$	. . .	$y_{(n-1)k+i}$	. . .	y_{nk}

Άρα με τη συστηματική δειγματοληψία ο πληθυσμός χωρίζεται σε k σύνθετες μονάδες και η διαδικασία επιλογής ενός συστηματικού δείγματος ισοδυναμεί με τη διαδικασία τυχαίας επιλογής μίας σύνθετης μονάδας . Επομένως , ένα 1 - ανά k συστηματικό τυχαίο δείγμα n μονάδων από έναν πληθυσμό μεγέθους N=nk είναι ένα απλό τυχαίο δείγμα μίας σύνθετης μονάδας από τον πληθυσμό με (σύνθετες) μονάδες τις k στήλες του πίνακα 8.2.1

Η διαίρεση του αρχικού πληθυσμού των N=nk μονάδων σε Κ ομάδες έχει ως αποτέλεσμα την δυνατότητα έκφρασης της διασποράς σ^2 του πληθυσμού μέσω της διασποράς « μεταξύ » των k ομάδων και της διασποράς « μέσα » στις ομάδες . Αυτό γίνεται ευκολότερα αντιληπτό αν το j στοιχείο του i συστηματικού δείγματος συμβολισθεί με y_{ij}

(δηλαδή $y_{ij} = y_{(j-1)k+i}$) . Τότε , αν μ είναι η μέση τιμή του πληθυσμού και $\overline{X}_n^{(i)}$ ο μέσος του i δείγματος , ισχύει ότι

$$(N-1)\sigma^2=\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\mu)^2= \quad (\pi\rho o\sigma\theta o\alpha\varphi\alpha\iota\rho\acute{\omega}\nu\tau\alpha\varsigma\ \overline{X}_n^{(i)})^2$$

$$=n\sum_{i=1}^{k}(\overline{X}_n^{(i)}-\mu)^2+\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\overline{X}_n^{(i)})^2 \qquad (8.2.1)$$

Αλλά επειδή κάθε ένα από τα k συστηματικά δείγματα συνεισφέρει n-1 βαθμούς ελευθερίας , ο δεύτερος προσθετέος είναι ίσος με $k(n-1)\sigma_w^2$

όπου

$$\sigma_w^2=\frac{1}{k(n-1)}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\overline{X}_n^{(i)})^2$$

Είναι η διασπορά μεταξύ μονάδων του πληθυσμού που ανήκουν στο ίδιο συστηματικό δείγμα . Επίσης , αν $\overline{X}_n^*$ συμβολίζει τον μέσο ενός συστηματικού δείγματος , το άθροισμα στον πρώτο προσθετέο της (8.2.1) είναι ίσο με $k \quad \sigma^2_{\overline{X}_n}$ αφού

$$\sigma^2_{\overline{X}_n^*}=\frac{1}{k}\sum_{i=1}^{k}\left[\overline{X}_n^{(i)}-E(\overline{X}_n^{(i)})\right]^2=\frac{1}{k}\sum_{i=1}^{k}(\overline{X}_n^{(i)}-\mu)^2$$

Είναι , δηλαδή , ο πρώτος προσθετέος ανάλογος της διασποράς μεταξύ των συστηματικών δειγμάτων .

Η διασπορά του μέσου $\overline{X}_n^*$ ενός 1-ανά-k συστηματικού δείγματος μεγέθους n από έναν πληθυσμό N=nk μονάδων δίνεται και από τον τύπο : $\sigma^2_{\overline{X}^*}=\left[(N-1)\sigma^2-k(n-1)\sigma_w^2\right]/N$

Ο μέσος $\overline{X}_{n^*}$ ενός συστηματικού δείγματος είναι ακριβέστερος από τον μέσο $\overline{X}_{n^*}$ ενός απλού τυχαίου δείγματος του ίδιου μεγέθους τότε και μόνο τότε αν $\sigma_w^2 \rangle \sigma^2$.

Η συστηματική δειγματοληψία οδηγεί σε μικρότερο τυπικό σφάλμα , αν η διασπορά μέσα στο δείγμα είναι μεγαλύτερη από την διασπορά ολόκληρου του πληθυσμού .

Επομένως , μεγαλύτερη ακρίβεια επιτυγχάνεται αν οι μονάδες του δείγματος έχουν μεγαλύτερη ετερογένεια σε σχέση με αυτήν που έχουν οι μονάδες όλου του πληθυσμού .

Παρατήρηση :

Είναι προφανές ότι στην περίπτωση εκτίμησης του ποσοστού p των μονάδων του πληθυσμού που ανήκουν σε μία κατηγορία Α , εφαρμόζεται η παραπάνω θεωρία , αν η παράμετρος μ αντικατασταθεί με την παράμετρο p και ο μέσος $\overline{X}_n^{(i)}$ αντικατασταθεί με τη στατιστική συνάρτηση $\hat{p}^{(i)}$ όπου $\hat{p}^{(i)}$ είναι το ποσοστό των μονάδων του i συστηματικού δείγματος που ανήκουν στην κατηγορία Α . Είναι προφανές ότι ο πληθυσμός θα αποτελείται από μονάδες της μορφής

$$y_{ij}=\begin{cases} 1, \text{ αν η j μονάδα του i δείγματος ανήκει στην Α} \\ 0, \text{ διαφορετικά} \end{cases}$$

Παράδειγμα

Ένας πληθυσμός 360 νοικοκυριών μίας συνοικίας με μικτό πληθυσμό (αριθμημένων από το 1 έως το 360) έχει καταγραφεί σε έναν κατάλογο κατά αλφαβητική σειρά ως προς το επίθετο του αρχηγού του νοικοκυριού . Τα νοικοκυριά των οποίων ο αρχηγός είναι μη λευκός εμφανίζονται με τους εξής αύξοντες αριθμούς .

28 , 31 – 33 , 36 – 41 , 44 , 45 , 47 , 55 , 56 , 58 , 68 , 69 , 82 , 83, 85 , 86 , 89 – 94 , 98 , 99 , 101 , 107 – 110 , 114 , 154 , 156 , 178 , 223 , 224 , 296 , 298 – 300 , 302 – 304 , 306 – 323 , 325 – 331 , 333 , 335 – 339 , 341 , 342 .

Να συγκριθεί η ακρίβεια ενός συστηματικού δείγματος που περιλαμβάνει ένα άτομο ανά 8 άτομα του πληθυσμού με την ακρίβεια ενός απλού τυχαίου δείγματος του ίδιου μεγέθους , αν υποτεθεί ότι θέλουμε να εκτιμήσουμε το ποσοστό των νοικοκυριών με μη λευκούς αρχηγούς .

Λύση

N=360 , k=8 . Άρα , n=N/k=360/8=45 . Επομένως υπάρχουν 8 δυνατά 1-ανά – 8 συστηματικά δείγματα μεγέθους 45 . Έστω

$$y_{ij} = \begin{cases} 1, \text{ αν η j μονάδα του i δείγματος έχει μη λευκό αρχηγό} \\ 0, \text{ διαφορετικά} \end{cases}$$

Οι αριθμοί $\sum_{j=1} y_{ij}, i-1,2,\ldots,8$ των νοικοκυριών με μη λευκό αρχηγό για τα διάφορα δυνατά δείγματα είναι :

ΔΕΙΓΜΑ	1	2	3	4	5	6	7	8	ΣΥΝΟΛΟ
$\sum_{j=1}^{45} y_{ij}$	7	13	10	10	12	9	10	10	81

Έστω $\hat{p}^*$ η εκτιμήτρια του p από ένα συστηματικό τυχαίο δείγμα .

Τότε έχουμε:

$$V(\hat{p}^*) = \tfrac{1}{k} \sum_{i=1}^{k} (\hat{p}^{(i)} - p)^2,$$

όπου $\hat{p}^{(i)} = \sum_{i=1}^{45} y_{ij} / 45, i = 1,2,\ldots k$

και $p = \sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij} / N = \sum_{i=1}^{8} \sum_{j=1}^{45} y_{ij} / 360 = \dfrac{81}{360} = 0.225$

Δηλαδή τελικά

$V(\hat{p}) = 0.001412$

Για την εκτιμήτρια $\hat{p}$ βασισμένη σε ένα απλό τυχαίο δείγμα μεγέθους 45 έχουμε

$$V(\hat{p}) = \frac{N-n}{n} \frac{p(1-p)}{N-1} = \frac{360-45}{45} \frac{(0.225)(0.775)}{359} = 0.00340$$

Είναι δηλαδή , η τιμή της διασποράς της $\hat{p}^*$ ίση με το 41,53% της τιμής της διασποράς της $\hat{p}$.

Παρατήρηση :

Η συστηματική δειγματοληψία χρησιμοποιείται συχνά λόγω της απλότητάς της , σε πληθυσμούς στους οποίους η αρίθμηση των μονάδων είναι τυχαία . Αυτό συμβαίνει , για παράδειγμα , στις περιπτώσεις δειγματοληψίας από ένα αρχείο ονομάτων τα οποία εμφανίζονται κατά αλφαβητική σειρά , με την προϋπόθεση ότι το χαρακτηριστικό που μελετάται δεν σχετίζεται με το επίθετο των συγκεκριμένων ατόμων . Στην περίπτωση αυτή δεν θα υπάρχει τάση ή στρωματοποίηση ως προς τις τιμές y_i των μονάδων του πληθυσμού καθώς θα διατρέχεται το αρχείο και δεν θα υπάρχει συσχέτιση μεταξύ γειτονιών τιμών . Επομένως , αναμένεται ότι η συστηματική δειγματοληψία θα είναι ισοδύναμη με την απλή τυχαία δειγματοληψία και θα οδηγεί στην ίδια διασπορά , οποτεδήποτε ο πληθυσμός έχει τυχαία διάταξη , με την έννοια της τυχαίας αρίθμησης των μονάδων που τον απαρτίζουν . Για οποιονδήποτε συγκεκριμένο πεπερασμένο πληθυσμό , και για οποιεσδήποτε τιμές του n και k , αυτό δεν αληθεύει ακριβώς . Ο λόγος είναι ότι η διασπορά της εκτιμήτριας στην περίπτωση της συστηματικής δειγματοληψίας , η οποία βασίζεται μόνο σε k βαθμούς ελευθερίας συμπεριφέρεται μάλλον ανορθόδοξα όταν η τιμή του k είναι μικρή και , επομένως , ενδέχεται να υπερβαίνει ή να είναι μικρότερη από τη διασπορά της εκτιμήτριας στην περίπτωση της απλής τυχαίας δειγματοληψίας . Μπορεί , όμως , να αποδειχθεί ότι κατά μέσο όρο, οι δύο διασπορές ταυτίζονται .

8.3 Πλεονεκτήματα – Μειονεκτήματα συστηματικής δειγματοληψίας

8.3.1 Πλεονεκτήματα

1. Είναι ευκολότερη στην χρήση της , λόγω της απλότητας της σε πληθυσμούς στους οποίους η αρίθμηση των μονάδων είναι τυχαία . Αυτό έχει ως αποτέλεσμα μικρότερο αριθμό σφαλμάτων .

2. Δίνει περισσότερο πληροφορίες σε δεδομένο με μεγαλύτερη ομοιογένεια στο ευρυνούμενο πληθυσμό .

3. Ένα συστηματικό δείγμα είναι πιο ομοιόμορφα κατανεμημένο στον πληθυσμό και αυτό συμβάλλει στο να παρέχει ακριβέστερες εκτιμήσεις .

4. Περιορίζεται η μεροληπτική επιλογή δειγματοληπτικών μονάδων.

5. Τα στοιχεία του πληθυσμού αντιπροσωπεύονται με την ίδια αναλογία στο δείγμα , οπότε το δείγμα είναι περισσότερο αντιπροσωπευτικό .

6. Είναι περισσότερο ασφαλής η χρησιμοποίηση της συστηματικής δειγματοληψίας , εάν δεν υπάρχουν ενδείξεις περιοδικότητας .

8.3.2 Μειονεκτήματα

1. Εμφανίζεται ο κίνδυνος περιοδικότητας στις τιμές των μονάδων του πληθυσμού όσον αφορά τη σειρά εμφάνισής τους στη λίστα , αν το κ είναι ίσο με την περίοδο ή ένα πολλαπλάσιό της . Για παράδειγμα μία συστηματική επιλογή μονάδων από μία «περιοδική» λίστα οικοδομικών τετραγώνων μίας πόλης μπορεί να οδηγήσει σ' ένα δείγμα που περιέχει τετράγωνα που ανήκουν σε μία γραμμή και επομένως σε αύξηση του σφάλματος των εκτιμητριών . Ο κίνδυνος αυτός μπορεί να ελαττωθεί αν το δείγμα είναι αποτέλεσμα ενός αριθμού συστηματικών επιλογών από διαφορετικά στρώματα .

2. Απαραίτητη προϋπόθεση είναι η ύπαρξη ή ο σχηματισμός καταλόγου με το σύνολο των στοιχείων Ν του πληθυσμού .

8.4 Σύγκριση της Συστηματικής και Στρωματοποιημένης δειγματοληψίας

Η συστηματική και στρωματοποιημένη δειγματοληψία παρέχουν την ίδια σχετικά ακρίβεια εάν επιλεχτεί ένα τυχαίο γενικά δείγμα . Είναι ασφαλής μέθοδος και τα αποτελέσματά τους μας δίνουν συχνά ένα αντιπροσωπευτικό δείγμα . Η μόνη διαφορά που εμφανίζουν οι δύο αυτοί μέθοδοι δειγματοληψίας εμφανίζεται στο γεγονός ότι ενώ οι μονάδες έχουν την ίδια σχετική θέση στο στρώμα στο συστηματικό δείγμα , στο στρωματοποιημένο δείγμα , η θέση των μονάδων στο στρώμα συνήθως καθορίζεται τυχαία . Παρατηρείται επιπλέον , ότι το συστηματικό δείγμα είναι πιο ομοιόμορφα κατανεμημένο στον πληθυσμό , παρέχοντας έτσι ακριβέστερες εκτιμήσεις από ό ,τι ένα στρωματοποιημένο δείγμα .

ΚΕΦΑΛΑΙΟ 9

Δειγματοληψία Κατά Ομάδες

9.1 Εισαγωγή

Η δειγματοληψία κατά ομάδες είναι μία τυχαία δειγματοληψία , όπου ο συνολικός πληθυσμός χωρίζεται σε ομάδες όπου με τυχαίο τρόπο επιλέγονται μόνο ορισμένες ομάδες . Κατόπιν , στο δείγμα περιλαμβάνουμε όλες τις δειγματοληπτικές μονάδες των ομάδων που έχουν επιλεγεί . Δηλαδή από το σύνολο Μ ομάδων του πληθυσμού , παίρνουμε τυχαία Κ ομάδες (K<M) και οι μονάδες που αποτελούν τις ομάδες αυτές συνενώνονται σ' ένα ενιαίο δείγμα . Τη μέθοδο αυτή την χρησιμοποιούμε στην περίπτωση που δε διαθέτουμε δειγματοληπτικό πλαίσιο με τις αρχικές μονάδες του πληθυσμού , υπάρχουν όμως πλαίσια των ομάδων π . χ . σχολεία , άτομα , κατοικίες , νοικοκυριά κ . τ . λ .

Για παράδειγμα , αν επιθυμούμε να πάρουμε δείγμα αγροτικών οικογενειών από τον πληθυσμό της Ελλάδας , δεν μπορούμε να σχηματίσουμε απλό , τυχαίο δείγμα από τέτοιου είδους οικογένειες , διότι δεν υπάρχει κατάλληλος κατάλογος όλων των αγροτικών οικογενειών της Ελλάδας . Ωστόσο , είναι δυνατό να θεωρήσουμε ότι ο πληθυσμός των ελληνικών αγροτικών οικογενειών είναι χωρισμένος σε ομάδες και ότι οι ομάδες αυτές αντιστοιχούν στα χωριά και τις πόλεις που έχουν πληθυσμό μέχρι 1000 κατοίκους . Έτσι , από το σύνολο των Μ οικισμών της Ελλάδας που ο καθένας τους έχει πληθυσμό μέχρι 1000 κατοίκους παίρνουμε τυχαίο δείγμα Κ οικισμών και συνενώνουμε σ' ένα ενιαίο σύνολο τις αγροτικές οικογένειες που τις αποτελούν .

Διαπιστώνουμε λοιπόν ότι η μέθοδος αυτή μας οδηγεί σε μικρότερο τυπικό σφάλμα . Επιπλέον , εάν το δείγμα είναι περισσότερο «απλωμένο» γεωγραφικά το κόστος μετάβασης είναι υψηλό . Εάν όμως το δείγμα είναι λιγότερο «απλωμένο» , που σημαίνει ότι οι δειγματοληπτικές μονάδες είναι σύνθετες , τότε το κόστος είναι μικρότερο .

Ουσιαστικά λοιπόν η δειγματοληψία γίνεται σε πρώτο στάδιο από ένα σύνθετο πληθυσμό του οποίου οι μονάδες είναι ομάδες μονάδων του αρχικού πληθυσμού .

Από τη στιγμή που επιλέγονται οι ομάδες , ο ερευνητής μπορεί να περιλάβει στο δείγμα όλες τις στοιχειώδεις μονάδες του αρχικού πληθυσμού που ανήκουν σ' αυτές ή να επιλέξει ένα δείγμα μικρότερων σύνθετων ομάδων . Η δειγματοληψία κατά ομάδες χωρίζεται σ' ένα στάδιο (single stage cluster sampling) και σε πολλά στάδια (multi-stage cluster sampling) .

1. Δειγματοληψία κατά ομάδες σ' ένα στάδιο ονομάζεται η δειγματοληπτική τεχνική , η οποία διαιρεί τις στοιχειώδεις μονάδες του πληθυσμού σε ομάδες (clusters) , επιλέγει ένα δείγμα των ομάδων αυτών και περιλαμβάνει στο τελικό δείγμα των στοιχειωδών μονάδων όλες τις στοιχειώδεις μονάδες που ανήκουν στις ομάδες αυτές .

2. Δειγματοληψία κατά ομάδες σε πολλά στάδια ονομάζεται η δειγματοληπτική τεχνική , όπου κατόπιν του πρώτου σταδίου επιλέγονται δείγματα μικρότερων ομάδων με τελικό στάδιο την επιλογή του δείγματος στοιχειωδών μονάδων .

Εκτός από τις 2 παραπάνω μεθόδους δειγματοληψίας έχουμε και την δειγματοληψία κατά περιοχές όπου οι σύνθετες μονάδες είναι ομάδες στοιχείων βασισμένες σε γεωγραφικές περιοχές .

Στην περίπτωση που οι σύνθετες μονάδες επιλέγονται με απλή τυχαία δειγματοληψία , κατά τα διάφορα στάδια τηςδειγματοληψίας, τότε το δειγματοληπτικό σχήμα ονομάζεται απλή δειγματοληψία κατά ομάδες σ' ένα ή περισσότερα στάδια .

9.2 Εκτίμηση της μέσης τιμής της δειγματοληψίας κατά ομάδες

Από τα προηγούμενα προκύπτει ότι η δειγματοληψία κατά ομάδες προϋποθέτει ότι ο πληθυσμός διαιρείται σε υποπληθυσμούς . Μερικοί από αυτούς εκπροσωπούνται στο δείγμα είτε εξ ολοκλήρου (ένα στάδιο) είτε εν μέρει μέσω κάποιου δείγματος (δύο ή περισσότερα στάδια) .

Η διαφορά της απλής δειγματοληψίας κατά ομάδες από την απλή στρωματοποιημένη δειγματοληψία έγκειται στο ότι στην πρώτη **μόνο ορισμένα** από τα στρώματα εκπροσωπούνται στο δείγμα , ενώ στην δεύτερη εκπροσωπούνται **όλα** τα στρώματα και πάντα μέσω απλών τυχαίων υποδειγμάτων .

Στην περίπτωση της απλής δειγματοληψίας κατά ομάδες , ο πληθυσμός $\{y_1, y_2, \ldots, y_N\}$ διαιρείται σε M ομάδες (υποπληθυσμός, clusters) $u_1, u_2, \ldots, u_M$ μεγέθους $N_1, N_2, \ldots, N_M$, αντίστοιχα , όπου :

$$\sum_{i=1}^{M} N_i = N.$$

Δηλαδή , αν με y_{ij} συμβολίζουμε την τιμή της j μονάδας της i ομάδας , τότε

$$u_i = \{y_{i1}, y_{i2}, \ldots, y_{iN_i}\} = i = 1,2,\ldots,M$$

και , επομένως , ο αρχικός πληθυσμός είναι η ένωση $\upsilon_1 \cup \upsilon_2 \cup \ldots \cup \upsilon_M$.

Προφανώς

$$\bar{N} = \sum_{i=1}^{M} N_i / M = N/M$$

συμβολίζει το μέσο μέγεθος των ομάδων αυτών .

Έστω $\{U_1, U_2, \ldots, U_m\}$ ένα απλό τυχαίο δείγμα m ομάδων από τον πληθυσμό των M ομάδων . Το j στοιχείο της i επιλεγείσας θα συμβολίζεται με U_{ij}, $j = 1,2,\ldots,N'_i$, $i = 1,2,\ldots m$. Δηλαδή, $U_i = \{U_{i1}, U_{i2}, \ldots, U_{iN'_i}\}$. Εδώ N_i' συμβολίζει το μέγεθος της i επλεγείσας ομάδας . Αυτό αντιστοιχεί στο πρώτο στάδιο της δειγματοληψίας . Στην περίπτωση που επακολουθεί και δεύτερο στάδιο , το σύνολο $\{U_{i1}, U_{i2}, \ldots, U_{in'_i}\}$, i=1,2,...m θα συμβολίζει ένα απλό τυχαίο δείγμα n_i μονάδων από την ομάδα U_i που επελέγη κατά το πρώτο στάδιο (i = 1 , 2 , ... , m) . Τότε

$$\bar{n} = \sum_{i=1}^{m} n_i / m$$ μέσο μέγεθος των m υποδειγμάτων (9.2.1)

$$\bar{U}_i = \sum_{j=1}^{n_i} U_{ij} / n_i$$ μέσος του i υποδείγματος (9.2.2)

i=1,2,...,m

$$S_i^2 = \sum_{j=1}^{n_i} (U_{ij} - \overline{U}_i)^2 / (n_i - 1)$$

$= \frac{1}{(n_i - 1)} x$ (άθροισμα τετραγωνικών αποκλίσεων των παρατηρήσεων του i δείγματος από τον μέσο του)

= διασπορά του i δείγματος , i=1,2,...,m (9.2.3)

Περιοριζόμενοι στην περίπτωση διεξαγωγής της δειγματοληψίας σε ένα στάδιο μόνο ($N'_i \leftrightarrow n_i$) έχουμε

$\mu = \sum_{i=1}^{M} \sum_{j=1}^{N_i} y_{ij} / N =$ μέση τιμή του αρχικού πληθυσμού (μέση τιμή του υπό εξέταση χαρακτηριστικού ανά στοιχειώδη μονάδα) (9.2.4)

$\mu_i = \sum_{j=1}^{N_i} y_{ij} / N_i$ =μέση τιμή της i ομάδας (του i υποπληθυσμού) , i = 1,2,.....M (9.2.5)

$\overline{\mu} = \sum_{i=1}^{M} \sum_{j=1}^{N_i} y_{ij} / M = \sum_{i=1}^{M} N_i \mu_i / M =$ μέση τιμή του χαρακτηριστικού ανά σύνθετη μονάδα (ανά ομάδα)

Προφανώς

$$\mu = \sum_{i=1}^{M} N_i \ \mu_i / \sum_{i=1}^{M} N_i = \overline{\mu} / \overline{N} \qquad 9.2.6$$

Για την εκτίμηση της μέσης τιμής μ του αρχικού πληθυσμού θα περιορισθούμε στην περίπτωση ενός σταδίου $(n_i = N_i \ \kappa\alpha\iota \ \overline{U}_i = \mu_i, i = 1,2,...,M)$ και θα διακρίνουμε τις εξής υποπεριπτώσεις :

(i) $N_i = N_j, \forall i \neq j$ (Ισομεγέθεις ομάδες, δηλαδή $N_i = \overline{N}$ $i = 1,2,....M$).

Στην περίπτωση αυτή , τα μεγέθη N'_i των ομάδων που θα επιλεγούν θα είναι ίσα με $\overline{N}$, δηλαδή $N'_i = \overline{N}$, $i = 1,2,...,m$.

Η στατιστική συνάρτηση :

$$\hat{\mu}_c = \sum_{i=1}^{m}\sum_{j=1}^{N} y_{ij}/(m\bar{N}) \Leftrightarrow \quad \hat{\mu}_c = \sum_{i=1}^{m} \bar{U}_i/m = \sum_{i=1}^{m}\mu_i/m$$

είναι μία αμερόληπτη εκτιμήτρια της μέσης τιμής μ με διασπορά

$$V(\hat{\mu}_c) \cong \frac{M-m}{M}\frac{1}{m}\frac{1}{\bar{N}^2}\sum_{i=1}^{M}(\sum_{j=1}^{\bar{N}} y_{ij}-\bar{\mu})^2/(M-1) = \frac{M-m}{M}\frac{1}{m}\sum_{i=1}^{M}(\mu_i-\mu)^2/(M-1)$$

(Η δεύτερη ισότητα προκύπτει από το γεγονός ότι $\bar{\mu} = \bar{N}\mu$ και

$$\sum_{j=1}^{N_i} y_{ij}\sum_{j=1}^{N} y_{ij} = N_i\mu_i = \bar{N}\mu_i)$$

Η στατιστική συνάρτηση

$$s_{\hat{\mu}} = \frac{M-m}{M}\frac{1}{M}\sum_{i=1}^{m}(\bar{U}_i-\bar{\mu}_c)^2/(m-1)$$

αποτελεί μία αμερόληπτη εκτιμήτρια της $V(\hat{\mu}_c)$

Η στατιστική συνάρτηση

$$\hat{Y} = M\bar{N}\hat{\mu}_c$$

είναι μία αμερόληπτη εκτιμήτρια του συνολικού μεγέθους

$$y = M\bar{\mu} = M\bar{N}\mu$$

Προφανώς $V(\hat{Y}) = (M\bar{N})^2 V(\hat{\mu}_c)$

2) $N \neq N_j, i \neq j$ **(Ανισομεγέθεις ομάδες)**

Τότε ισχύει

Η στατιστική συνάρτηση

$$\hat{\mu}_c = \sum_{i=1}^{m}\sum_{j=1}^{N'_i} U_{ij}/\sum_{i=1} N'_i$$

είναι μία αμερόληπτη εκτιμήτρια της παραμέτρου μ με διασπορά

$$V(\hat{\mu}_c) \cong \frac{M-m}{M}\frac{1}{m}\frac{1}{\bar{N}^2}\sum_{i=1}^{M}(\sum_{j=1}^{N_i} y_{ij}-N_i\mu)^2/(M-1)$$

Η στατιστική συνάρτηση

$$S(\hat{\mu}_c) = \frac{M-m}{M}\frac{1}{m\bar{N}^2}\sum_{i=1}^{m}(\sum_{j=1}^{N'_i} U_{ij}-N'_i\hat{\mu}_c)^2/(m-1)$$

είναι αμερόληπτη εκτιμήτρια της $V(\hat{\mu}_c)$.

9.3 Εκτίμηση ποσοστών της δειγματοληψίας κατά ομάδες

Έστω πληθυσμός μεγέθους N ο οποίος διαιρείται σε M ομάδες (clusters) μεγέθους N_1 , N_2 , ... , N_M . Έστω p το ποσοστό των μονάδων του πληθυσμού οι οποίες ανήκουν σε μία κατηγορία A . Το πρόβλημα που θα αντιμετωπισθεί είναι η εκτίμηση του p με βάση ένα δείγμα που θα επιλεγεί με απλή δειγματοληψία κατά ομάδες . Έστω $N_A^{(i)}$ ο αριθμός των μονάδων της ομάδας i που ανήκουν στην κατηγορία A και $p_i = N_A^{(i)} / N_i$ το ποσοστό των μονάδων της i ομάδας που ανήκουν στην κατηγορία A , i=1,2,...,M . Προφανώς ισχύει ότι

$$P = \sum_{i=1}^{M} N_i P_1 / \sum_{i=1}^{M} N_i \tag{9.3.1}$$

Έστω ότι από το σύνολο των M ομάδων επιλέγεται ένα απλό τυχαίο δείγμα m ομάδων U_1 , U_2 , ... , U_m (πρώτο στάδιο) και έστω ότι από την i επιλεγείσα ομάδα επιλέγεται ένα απλό τυχαίο δείγμα στοιχειωδών μονάδων μεγέθους n_i (δεύτερο στάδιο) . Αν $X^{(i)}$ συμβολίζει τον αριθμό των μονάδων του i δείγματος που ανήκουν στην κατηγορία A , τότε μία αμερόληπτη εκτιμήτρια του p_i είναι η

$$\hat{p}_i = X^{(i)} / n_i, \quad i = 1,2,....,m \tag{9.3.2}$$

Περιοριζόμενοι στην περίπτωση δειγματοληψίας σε ένα στάδιο μόνο, μπορούμε να εφαρμόζουμε , την θεωρία για την εκτίμηση του p , αν θεωρήσουμε τις αντίστοιχες $\mu_i \leftrightarrow p_i$, $\mu \leftrightarrow p$. Συγκεκριμένα οδηγούμαστε στα εξής συμπεράσματα :

1) Περίπτωση ισομεγεθών ομάδων ($N_i = \overline{N}$, $\hat{p} = p_i, i = 1,2,...,M$)

Στην περίπτωση αυτή $N_i = \overline{N}$, i-1 , 2 , ... , m , οπότε και ισχύει ότι Η στατιστική συνάρτηση

$$p_c = \sum_{i=1}^{m} \hat{p}_i / m \tag{9.3.3}$$

είναι αμερόληπτη εκτιμήτρια της παραμέτρου p με διασπορά

$$V(\hat{p}_c) = \frac{M-m}{M} \frac{1}{m} \sum_{i=1}^{M} (p_i - p)^2 / (M-1) \tag{9.3.4}$$

Μία αμερόληπτη εκτιμήτρια της $V(\hat{p}_c)$ είναι η στατιστική συνάρτηση

$$S^2\hat{p}_c = \frac{M-m}{M}\frac{1}{m}\sum_{i=1}^{M}(\hat{p}_i - \hat{p}_c)^2/(m-1) \qquad (9.3.5)$$

2) Περίπτωση ανισομεγεθών ομάδων ($N_i \neq N_j, i \neq j$)

Η στατιστική συνάρτηση

$$P_c = \sum_{i=1}^{m} x^{(1)} / \sum_{i=1}^{m} N_i \qquad (9.3.6)$$

είναι μία αμερόληπτη εκτιμήτρια της παραμέτρου p με διασπορά

$$V(\hat{p}_c) = \frac{M-m}{M}\frac{1}{m\overline{N}^2}\sum_{i=1}^{M}(N_A^{(i)} - N_i p)^2/(M-1) \qquad (9.3.7)$$

Η στατιστική συνάρτηση

$$S^2_{\hat{P}c} = \frac{M-m}{M}\frac{1}{m\cdot\overline{N}^2}\frac{\sum_{i=1}^{m} x^{(i)2} - 2\hat{P}_c\sum_{i=1}^{m} X^{(i)}N'_i + \hat{P}_c^2\sum_{i=1}^{m} N_i'^2}{m-1} \qquad (9.3.8)$$

Είναι μία αμερόληπτη εκτιμήτρια της $V(\hat{P}_c)$.

Παράδειγμα:

Έστω ότι οι 660 φοιτητές του τμήματος Στατιστικής ενός πανεπιστημίου μπορούν να διαιρεθούν σε 110 τάξεις των 6 φοιτητών έτσι ώστε κάθε φοιτητής να ανήκει σε μία μόνο τάξη . Για να εξετασθεί πως αντιμετωπίζεται μία μελετώμενη μεταβολή στο πρόγραμμα σπουδών , επιλέγονται τυχαία 11 τάξεις και όλοι οι φοιτητές των τάξεων αυτών περιλαμβάνονται στο δείγμα . Έστω ότι οι αριθμοί υπέρ της μεταβολής είναι

$\chi^{(i)}$	Σύνολο
3,5,2,3,4,1,4,2,6,1,2	33

Να εκτιμηθεί το πραγματικό ποσοστό p των φοιτητών του τμήματος που είναι υπέρ της μεταβολής .

Λύση

Προφανώς m=11,M=110, Ni=6, i=1,2,.....110 . Τότε από τη σχέση (9.3.3) έχουμε

$p_c = \sum_{i=1}^{11} \hat{P}i/11 = \sum_{i=1}^{11} x^{(i)}/(6x11) = 0.5$

Επίσης από τη σχέση (9.3.5) προκύπτει ότι :

$$S^2_{\hat{P}c} = \frac{110-11}{110}\frac{1}{11}\sum_{i=1}^{11}(\hat{P}_i - 0.5)^2/10 = 0.00591$$

Άρα $S_{\hat{P}c} = 0.0769$

Σημείωση: Ένα απλό τυχαίο δείγμα 66 φοιτητών , που ενδεχομένως θα έδινε την ίδια εκτίμηση για το P , θα οδηγούσε σε διασπορά της εκτίμησης ίσης με

$$S^2_{\hat{P}c} = \frac{P(1-\hat{P})}{n-1}\frac{N-n}{N} = \frac{(0.5)(0.5)}{65}\frac{110-66}{110} = 0.00154$$

και επομένως σε τυπικό σφάλμα ίσο με : $S_{\hat{P}} = 0.0392$

9.4 Πλεονεκτήματα – Μειονεκτήματα της δειγματοληψίας κατά ομάδες

9.4.1 Πλεονεκτήματα

1. Περιορίζει τις δαπάνες μετακίνησης των συνεντευκτών.
2. Συντομεύει το χρόνο διεξαγωγής της έρευνας όταν οι μονάδες που περιέχονται σε κάθε ομάδα βρίσκονται γεωγραφικά συγκεντρωμένες και η προσέγγισή τους γίνεται σχετικά συντομότερα .
3. Το κόστος είναι χαμηλότερο όταν παίρνουμε ένα δείγμα αποτελούμενο από n ομάδες από ένα πλήθος N – ομάδων , αντί να παίρνουμε ένα απλό τυχαίο δείγμα απ' ευθείας από ένα μεγάλο πληθυσμό .
4. Είναι η μόνη μέθοδος που μπορούμε να χρησιμοποιήσουμε όταν δε διαθέτουμε δειγματοληπτικό πλαίσιο με τις αρχικές μονάδες του πληθυσμού για τη συγκέντρωση των στοιχείων μίας έρευνας .
5. Η συλλογή του δείγματος γίνεται με λιγότερο πολύπλοκο τρόπο .
6. Επιδιώκεται η μέγιστη δυνατή ακρίβεια με το ελάχιστο δυνατό κόστος .

9.4.2 Μειονεκτήματα

1. Η αβεβαιότητα που υπάρχει για το εάν είναι μη αντιπροσωπευτικές οι μονάδες της ομάδας του συνολικού πληθυσμού αποτελεί ένα από τα βασικότερα μειονεκτήματα της μεθόδου .

9.5 Σύγκριση δειγματοληψίας κατά ομάδες με την στρωματοποιημένη δειγματοληψία

Η δειγματοληψία κατά ομάδες προϋποθέτει ότι ο πληθυσμός διαιρείται σε υποπληθυσμούς από τους οποίους μερικοί είτε εκπροσωπούνται εξ ολοκλήρου στο δείγμα είτε εν μέρει μέσω κάποιου δείγματος .

Η διαφορά μεταξύ των δύο αυτών μεθόδων έγκειται στο γεγονός ότι στην απλή δειγματοληψία κατά ομάδες , <u>μόνο</u> ορισμένα από τα στρώματα εκπροσωπούνται στο δείγμα , σε αντίθεση με τη στρωματοποιημένη δειγματοληψία , όπου εκπροσωπούνται <u>όλα</u> τα στρώματα μέσω απλών τυχαίων υποδειγμάτων .

Επιπλέον , στη δειγματοληψία κατά στρώματα παίρνουμε τυχαίο δείγμα μονάδων από κάθε στρώμα , ενώ στη δειγματοληψία κατά ομάδες παίρνουμε τυχαία ορισμένες ομάδες και στη συνέχεια όλες τις μονάδες που έχει κάθε ομάδα του δείγματος αυτού . Η δειγματοληψία κατά ομάδες επιβάλλει τη δημιουργία των ομάδων της , με τέτοιο τρόπο , ώστε μέσα σε κάθε μία να υπάρχει τόσο ανομοιογένεια , όσο σε ολόκληρο τον πληθυσμό . Αντίθετα , στη δειγματοληψία κατά στρώματα συμβαίνει ακριβώς το αντίθετο : Είναι απαραίτητο να επικρατεί μέσα σε κάθε στρώμα όσο το δυνατόν μεγαλύτερη ομοιογένεια και μεταξύ των στρωμάτων η μεγαλύτερη δυνατή ανομοιογένεια , ώστε κάθε στρώμα να είναι ομοιογενής και ξεχωριστή ομάδα του πληθυσμού .

ΚΕΦΑΛΑΙΟ 10

Δειγματοληψία ποσοστών

10.1 Εισαγωγή

Δειγματοληψία ποσοστών ονομάζεται το δειγματοληπτικό σχέδιο που μοιάζει με τη στρωματοποιημένη δειγματοληψία , αλλά η επιλογή των μονάδων μέσα σε κάθε στρώμα δε γίνεται τυχαία . Η επιλογή αυτή γίνεται από τους συνεντευκτές-ερευνητές με δικά τους προσωπικά κριτήρια . Η κρίση δηλαδή του ερευνητή παίζει σημαντικό ρόλο στην επιλογή των μονάδων που μας δείχνουν κατά πόσο είναι ή μη αντιπροσωπευτικό ένα δείγμα .

Η μέθοδος αυτή θα μπορούσε να θεωρηθεί μία μορφή στρωματοποημένης δειγματοληψίας με αναλογικό καταμερισμό μία και τα n_i καθορίζονται έτσι ώστε να εκπροσωπούν το ($100\ N_i/N$) του συνολικού δείγματος . Διαφέρει από τη στρωματοποιημένη δειγματοληψία στο ότι η επιλογή των n_i-μονάδων από το i-στρώμα δε γίνεται με απλή τυχαία δειγματοληψία . Υπάρχει ένα στοιχείο μη τυχαιότητας που εισάγεται εξ αιτίας του τρόπου επιλογής των μονάδων . Η επιλογή αυτή γίνεται από τον ερευνητή με τρόπο ώστε το δείγμα που θα επιλεγεί να εκπροσωπεί ορισμένα προκαθορισμένα ποσοστά . Όσον αφορά διάφορα άλλα χαρακτηριστικά εκτός από το μέγεθος του i στρώματος σκοπός των ερευνητών – συνεντευκτών είναι η συνέντευξη που θα κάνουν στα άτομα που θα συναντήσουν να τελειώσει το συντομότερο δυνατό χρόνο . Επίσης επιδίωξή τους είναι να σχηματίσουν στρώματα με τη μεγαλύτερη δυνατή εσωτερική ομοιογένεια και τη μεγαλύτερη δυνατή διαφορά μεταξύ τους (ως προς το χαρακτηριστικό που ερευνούμε) :

<u>Παράδειγμα :</u>

Έστω ότι θέλουμε να εξετάσουμε το ποσοστό των Ελλήνων που είναι ικανοποιημένοι από την είσοδό μας στην ΕΟΚ .

Για το σκοπό αυτό σχηματίζουμε τα εξής στρώματα που είναι απαραίτητο να έχουν τα εξής χαρακτηριστικά για τη μέγιστη δυνατή αντιπροσωπευτικότητα του δείγματος:

- Φύλο (άνδρες , γυναίκες)
- Ηλικία (νέοι , ώριμοι , ηλικιωμένοι)
- Απασχόληση (άνεργοι , απασχολούμενοι , μη οικονομικώς ενεργοί)
- Τομέας (αγροτικός , αστικός)

Παρατηρούμε ότι ορισμένες διακρίσεις (όπως είναι το φύλο) δε δημιουργούν κανένα σοβαρό πρόβλημα στους συνεντευκτές κατά το χρόνο διεξαγωγής των συνεντεύξεων . Άλλες διακρίσεις όμως όπως η ηλικία , απασχόληση , τομέας χρειάζονται περισσότερες διευκρινήσεις και ερωτήσεις ώστε να οδηγήσουν σε αντιπροσωπευτικά αποτελέσματα (για παράδειγμα) όσον αφορά την ηλικία , εάν είναι παντρεμένος , ελεύθερος , χωρισμένος , με παιδιά κ.τ.λ .

Το ποσοστό των μονάδων αυτών που θα επιλέγουν από κάθε στρώμα πρέπει να αντιστοιχεί στην παρούσα δομή του πληθυσμού. Το ποσοστό αυτό υπολογίζεται από τα αποτελέσματα των απογραφών του πληθυσμού (ηλικία , φύλο , επάγγελμα κ.τ.λ.). Εκτός από τις απογραφές του πληθυσμού χρησιμοποιούνται και άλλου είδους απογραφές (π.χ. καταστημάτων) . Η ποσοστιαία μεταβολή από την οποία θα προκύψει το δείγμα είναι δυνατόν να γίνει για κάθε χαρακτηριστικό ξεχωριστά και ανεξάρτητα από τα λοιπά χαρακτηριστικά ή σε συνδυασμό με αυτά .

Ας υποθέσουμε λοιπόν ότι έχουμε 200 άτομα με τα εξής χαρακτηριστικά:

ΠΙΝΑΚΑΣ 10.1

ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ	ΑΤΟΜΑ
1. ΦΥΛΛΟ	200
1.1 Άνδρες	110
1.2 Γυναίκες	90
2. ΗΛΙΚΙΑ	200
2.1 15-30	57
2.2 31-60	120
2.3 61 & άνω	23
3. ΑΠΑΣΧΟΛΗΣΗ	200
2.1 Εργαζόμενοι	180
2.2 Άνεργοι	20

Για να πετύχουμε όμως μεγαλύτερη αντιπροσωπευτικότητα του δείγματος θα αναφέρουμε και τα ποσοστά που αντιστοιχούν στα λοιπά χαρακτηριστικά των μονάδων του δείγματος .

ΠΙΝΑΚΑΣ 10 . 2

Ηλικία	Εργαζόμενοι		Άνεργοι		Σύνολο
	Άνδρες	Γυναίκες	Άνδρες	Γυναίκες	
15-30	**30**	**20**	**4**	**3**	**57**
31-60	**60**	**50**	**4**	**6**	**120**
60 & άνω	**10**	**10**	**2**	**1**	**23**
Σύνολο	**100**	**80**	**10**	**10**	**200**

Ο πίνακας 10.2 υποχρεώνει τους συνεντευκτές να επιλέξουν για το δείγμα τους , όχι απλά 110 άνδρες και 90 γυναίκες αλλά τα πρόσωπα αυτά να είναι κατανεμημένα και ως προς τα λοιπά χαρακτηριστικά με τον τρόπο που παρουσιάζεται στον αναφερόμενο πίνακα . Αν δεν είναι δυνατόν να γίνει η πιο πάνω λεπτομερειακή κατανομή των μονάδων του δείγματος , δίνονται οδηγίες στους συνεντευκτές να εντάσσουν στο δείγμα τους πρόσωπα που συνδυάζουν κατά τρόπο αντιπροσωπευτικό όλα τα χρησιμοποιούμενα χαρακτηριστικά .

10.2 Πλεονεκτήματα – Μειονεκτήματα δειγματοληψίας ποσοστών

10.2.1 Πλεονεκτήματα

1. Είναι σχετικά μη δαπανηρή. Αυτό οφείλεται στο γεγονός ότι σε περίπτωση εμφάνισης τυχόν προβλήματος ο συνεντευκτής έχει τη δυνατότητα σε μικρό χρονικό διάστημα να αντικαταστήσει το Χ ερωτώμενο με κάποιον Υ ερωτώμενο. Επιπλέον, για να είναι το κόστος σε μειωμένο επίπεδο, οι συνεντευκτές φροντίζουν να εντάσσουν το δείγμα τους σ' εκείνα τα πρόσωπα με τα επιθυμητά χαρακτηριστικά, ώστε να μην χρειάζονται μετακινήσεις.
2. Είναι δυνατόν να διεξαχθεί ακόμα και σε περιπτώσεις που η τυχαία λήψη του δείγματος δεν μπορεί να γίνει επειδή το αναγκαίο δειγματοληπτικό πλαίσιο είναι ανύπαρκτο.
3. Δίνει τη δυνατότητα αξιολόγησης των πληροφοριών που περιέχονται στο δείγμα με στατιστικές μεθόδους.
4. Δίνει τη δυνατότητα υπολογισμού του σφάλματος που είναι συνδεδεμένο με τη χρησιμοποιούμενη εκτιμήτρια.
5. Τα διοικητικά προβλήματα σ' αυτήν την περίπτωση είναι λιγότερα. Αυτό οφείλεται στην έλλειψη των προβλημάτων που προκύπτουν από τις αρνήσεις για συνεργασία και από την επανάληψη των συνεντεύξεων. Συνήθως τα πρόσωπα που παρέχουν αναγκαίες πληροφορίες στον συνεντευκτή κατά κανόνα παραμένουν άγνωστα και έτσι δεν είναι δυνατόν να γίνει κλασικός έλεγχος του αν έγινε και πώς έγινε η συνέντευξη (μπορεί βεβαίως να γίνει μαγνητοφώνηση).
6. Η μέθοδος αυτή απαιτεί λιγότερο χρόνο για να ολοκληρωθεί, γι' αυτό είναι κατάλληλη για έρευνες που πρέπει να διεξαχθούν σε μικρό χρονικό διάστημα. Τέτοιες έρευνες είναι όσες αναφέρονται σε ζητήματα που δεν είναι δυνατόν να κρατηθούν στη μνήμη για μακρύ χρονικό διάστημα ή που έχουν επείγοντα χαρακτήρα (π. χ. σφυγμομέτρηση της κοινής γνώμης για διαμόρφωση στάσης που θα ακολουθήσει μία παράταξη για ένα πολιτικό θέμα).

10.2.2 Μειονεκτήματα

1. Ο υπολογισμός των τυπικών σφαλμάτων είναι δύσκολος , λόγω έλλειψης τυχαιότητας κατά την επιλογή των μονάδων του δείγματος. Δηλαδή τα αποτελέσματα ενός μη τυχαίου δείγματος ισχύουν μόνο για το δείγμα από το οποίο υπολογίσθηκαν και δεν είναι δυνατόν να ενταχθούν με στατιστικό τρόπο στον αντίστοιχο πληθυσμό . Είναι δυνατόν όμως να υπολογισθούν , εάν το συνολικό δείγμα διαιρείται σε μικρότερα , αλλά ανεξάρτητα δείγματα που αντλούνται από τον ίδιο πληθυσμό . Κατόπιν , τα μικρότερα αυτά δείγματα προσεγγίζονται ως δείγματα ποσοστών από τους συνεντευκτές και παρέχουν τόσες ανεξάρτητες εκτιμήσεις για το συγκεκριμένο χαρακτηριστικό όσα είναι τα μικρότερα ανεξάρτητα δείγματα . Η μεταβλητικότητα των εκτιμήσεων αυτών χρησιμοποιείται για να υπολογισθεί η μεταβλητικότητα της δειγματοληψίας ποσοστών .
2. Η προσωπική επιλογή μπορεί να είναι σοβαρά μεροληπτική .
3. Η δειγματοληψία ποσοστών δεν μπορεί να απαλλαγεί από τη μεροληψία επιλογής των μονάδων του δείγματος που ίσως δημιουργούν οι συνεντευκτές .
4. Δεν μπορούμε να ελέγξουμε σοβαρά τον συνεντευκτή κατά το χρόνο διεξαγωγής της έρευνας των μονάδων του δείγματος και συλλογής πληροφοριών .

10.3 Σύγκριση Απλής – Τυχαίας και Ποσοστιαίας Δειγματοληψίας

Παρατηρείται ότι η ποσοστιαία δειγματοληψία είναι λιγότερο δαπανηρή και απαιτεί λιγότερο χρονικό διάστημα για τη διεξαγωγή μίας έρευνας σε σχέση με την τυχαία δειγματοληψία . Αυτό οφείλεται στο γεγονός ότι ο συνεντευκτής δεν χρειάζεται να κάνει μεγάλο ταξίδι για να εντοπίσει τον τρόπο διαμονής του προσώπου που επιλέγεται στο δείγμα .

Επιπλέον , εάν το πρόσωπο αρνείται τη συνέντευξη , τότε ο συνεντευκτής μπορεί να αντικαταστήσει το πρόσωπο αυτό με κάποιο άλλο στα ποσοστά που έχουν καθοριστεί .

Από την άλλη όμως , η ποσοστιαία δειγματοληψία δεν μπορεί να εφαρμόσει στατιστική μεθοδολογία για την αξιολόγηση συμπερασμάτων που συνάγονται στο δείγμα , πράγμα το οποίο σημαίνει ότι ο υπολογισμός σφάλματος κατά την εκτίμηση παραμέτρων του πληθυσμού είναι αδύνατος . Η ποσοστιαία δειγματοληψία γενικά εξασφαλίζει την αντικειμενικότητα στα αποτελέσματα .

Κεφάλαιο 11

11.1 Γενικές παρατηρήσεις επί των μεθόδων

Ανακεφαλαιώνοντας , αναφέραμε τις 6 πιο βασικές μεθόδους δειγματοληψίας που μας εξασφαλίζουν την αντικειμενικότητα στα αποτελέσματα και μας διευκολύνουν στη διεξαγωγή οποιασδήποτε έρευνας . Η τυχαία απλή δειγματοληψία είναι μία προσιτή μέθοδος , ακόμα και γι' αυτούς που έχουν χαμηλό στατιστικό υπόβαθρο .

Με τη μέθοδο αυτή μπορεί να υπολογισθεί το μέγεθος του δειγματοληπτικού σφάλματος και τα αποτελέσματά της είναι σχετικά σωστά μόνο όταν ο πληθυσμός είναι μικρός . Η στρωματοποιημένη δειγματοληψία είναι πιο αξιόπιστη μέθοδος και μας δίνει περισσότερο αντικειμενικά και αξιόπιστα αποτελέσματα . Είναι μία μέθοδος τυχαίας δειγματοληψίας κατά την οποία πρώτα διαιρούμε τον πληθυσμό σ' ένα συγκεκριμένο αριθμό υποπληθυσμών και στη συνέχεια επιλέγουμε ένα τυχαίο δείγμα από κάθε υποπληθυσμό .

Η συστηματική δειγματοληψία είναι μία τεχνική η οποία εισάγει ένα συστηματικό στοιχείο κατά τη διαδικασία επιλογής των μονάδων του πληθυσμού . Γίνεται πάντα με την εφαρμογή κατάλληλου δειγματοληπτικού πλαισίου . Είναι εύκολη στη χρήση , δίνει περισσότερες πληροφορίες με λιγότερο δαπάνες και με μεγαλύτερη ταχύτητα λόγω της ομοιογένειάς της . Επιπλέον , είναι αρκετά ασφαλής μέθοδος σε περιπτώσεις μη περιοδικότητας .

Η δειγματοληψία κατά ομάδες είναι μία τυχαία δειγματοληψία , όπου ο συνολικός πληθυσμός χωρίζεται σε ομάδες που επιλέγονται με τυχαίο τρόπο μόνο ορισμένες υποομάδες . Η μέθοδος αυτή είναι η μόνη που μπορούμε να χρησιμοποιήσουμε όταν δε διαθέτουμε δειγματοληπτικό πλαίσιο με τις αρχικές μονάδες του πληθυσμού . Ο τρόπος συλλογής του δείγματος εδώ δεν είναι περίπλοκος και το κόστος και οι δαπάνες μετακίνησης των συνεντευκτών είναι λιγότερες .

Σκοπός γενικά της μεθόδου είναι η μέγιστη δυνατή ακρίβεια με το ελάχιστο δυνατό κόστος . Η δειγματοληψία ποσοστών είναι η μόνη μέθοδος που σε αντίθεση με τις παραπάνω μεθόδους , η επιλογή των μονάδων μέσα σε κάθε στρώμα δε γίνεται τυχαία. Η επιλογή των μονάδων αυτών στηρίζεται κυρίως στην κρίση των συνεντευκτών. Είναι σχετικά μη δαπανηρή και τα διοικητικά προβλήματα είναι λιγότερα αφού τα πρόσωπα που παρέχουν τις αναγκαίες πληροφορίες στον συνεντευκτή κατά κανόνα παραμένουν άγνωστα.

Εκτός από τις παραπάνω μεθόδους δειγματοληψίας υπάρχουν και πολλές άλλες μέθοδοι δειγματοληψίας που θα αναφέρουμε ονομαστικά όπως:

1. Η πολυσταδιακή δειγματοληψία (Multi-stage sampling) όπου μπορεί να εφαρμοστεί είτε ως τυχαία δειγματοληψία είτε ως δειγματοληψία κατά ομάδες.

2. Η δειγματοληψία με μεταβαλλόμενες πιθανότητες, όπου επιδιώκει να εξασφαλίσει κατά την επιλογή των μονάδων ίση πιθανότητα σε κάθε στοιχείο του πληθυσμού που θα περιληφθεί στο δείγμα.

3. Δειγματοληψία σε πολλές φάσεις που είναι δευτερεύουσας σημασίας και συνεπώς η μέτρηση του χαρακτηριστικού που αναφέρεται σ' αυτές είναι δυνατό να διεξαχθεί με σχετικά μικρότερη ακρίβεια, δηλαδή με μεγαλύτερο δειγματοληπτικό σφάλμα.

4. Επιφανειακή δειγματοληψία διεξάγεται συνήθως σε δύο ή περισσότερα στάδια. Όταν πρόκειται να πάρουμε ένα τυχαίο δείγμα από μία γεωγραφική επιφάνεια για την οποία όμως δεν υπάρχουν κατάλογοι των δειγματοληπτικών μονάδων αλλά υπάρχουν χάρτες τοπογραφικοί απολύτως ενημερωμένοι, τότε διενεργείται η επιφανειακή δειγματοληψία.

5. Κατευθυνόμενη δειγματοληψία η οποία εφαρμόζεται όταν η επιλογή των στοιχείων του δείγματος δεν είναι αποτέλεσμα τύχης, αλλά αντίθετα είναι αποτέλεσμα συνειδητής και υποκειμενικής επιλογής τους από τον ερευνητή.

6. Δειγματοληψία με σταθερά δείγματα που εφαρμόζεται σε περιπτώσεις που συγκεντρώνονται πληροφορίες από έναν πληθυσμό σε τακτά χρονικά διαστήματα (κάθε μήνα, χρόνο κ.τ.λ.).

7. Δειγματοληψία από κύρια δείγματα όπου γίνεται η επιλογή με τυχαίο τρόπο από αρκετά μεγάλο , λεπτομερές και αντιπροσωπευτικό δείγμα του ερευνούμενου πληθυσμού και κατόπιν γίνεται επιλογή με τυχαίο τρόπο μικρότερων δειγμάτων όταν θέλουμε να προβούμε στη διενέργεια μίας έρευνας .

8. Πειραματική δειγματοληψία που μας δίνει αποτελέσματα κατόπιν συστηματικών ή συμπτωματικών πειραμάτων .

Κλείνοντας το συμπέρασμα που προκύπτει είναι ότι οι μέθοδοι δειγματοληψίας λαμβάνοντας πάντα υπ' όψιν την εκτίμηση των παραμέτρων του πληθυσμού που κάθε φορά επιλέγουμε , είναι πολύ σημαντικοί για τη διεξαγωγή μίας αξιόπιστης , αξιόλογης και αποτελεσματικής έρευνας .

Κεφάλαιο 12

12.1 Γενικές εφαρμογές επί των μεθόδων της δειγματοληψίας

12.1.1 Παράδειγμα απλής τυχαίας δειγματοληψίας

Παράδειγμα 1:

Έστω ότι θέλουμε να εκτιμήσουμε τη μέση και τη συνολική μηνιαία δαπάνη νοσοκομειακής περίθαλψης των νοικοκυριών που έχουν άτομα ηλικίας 65 ετών και πάνω , μιάς πόλης , όπου διαμένουν 1000 νοικοκυριά . Από διαθέσιμο κατάλογο των νοικοκυριών της πόλης παίρνουμε με τυχαίο τρόπο δείγμα $n = 100$ νοικοκυριά και βρίσκουμε ότι τα νοικοκυριά που έχουν τουλάχιστον ένα άτομο ηλικίας πάνω των 65 ετών ανέρχονται σε $n_j = 60$. Από την έρευνα που έγινε στα 60 νοικοκυριά διαπιστώθηκε ότι οι δαπάνες κατά νοικοκυριό , για νοσοκομειακή περίθαλψη (σε χιλιάδες δραχμές) είναι :

10	9	11	15	11	14	10	13	12	12	14	12
15	12	13	14	12	13	12	14	11	13	15	14
12	14	15	12	9	15	14	11	10	15	13	13
13	15	12	14	14	9	15	9	14	14	14	10
11	10	9	10	13	10	11	10	15	12	12	9

Να εκτιμηθούν η μέση και η συνολική μηνιαία δαπάνη για νοσοκομειακή περίθαλψη των νοικοκυριών της πόλης και να υπολογισθούν το τυπικό σφάλμα, ο συντελεστής μεταβλητικότητας και το διάστημα εμπιστοσύνης των εκτιμήσεων αυτών με πιθανότητα 95% .

Λύση

Η εκτίμηση της μέσης μηνιαίας δαπάνης για νοσοκομειακή περίθαλψη των νοικοκυριών της πόλης , που έχουν άτομα ηλικίας 65 ετών και πάνω είναι :

$$\overline{y}_j = \frac{\sum_{K=1}^{n_j} y_{jK}}{n_j} = \frac{10+9+11...+12+9}{60} = \frac{739}{60} = 12{,}3$$

δηλαδή 12.300 δρχ . τον μήνα .

Εάν ο αριθμός των νοικοκυριών της πόλης που έχουν άτομα ηλικίας πάνω των 65 ετών ανέρχεται σε $N_j = 650$, τότε η διακύμανση του $\overline{y}_j$ είναι :

$$V(\overline{y_j}) = \left(\frac{N_j - n_j}{N_j}\right)\frac{\sum_{K=1}^{n_j}(y_{jK} - \overline{y}_j)^2}{n_j(n_j - 1)} = \frac{650 - 60}{650}\frac{223}{60(60-1)} = 0{,}057$$

Το τυπικό σφάλμα είναι :

$$SE(\overline{y}_j) = \sqrt{V(\overline{y}_j)} = \sqrt{0{,}057} = 0{,}238$$

Ο συντελεστής μεταβλητικότητας είναι :

$$CV(\overline{y}_j) = \frac{SE(\overline{y}_j)}{\overline{y}_j}100\% = \frac{0{,}238}{12{,}3}100\% = 1.93\%$$

Το διάστημα εμπιστοσύνης στο οποίο βρίσκεται η μέση μηνιαία δαπάνη για νοσοκομειακή περίθαλψη των νοικοκυριών $\overline{Y}_j$ της πόλης , με πιθανότητα 95% είναι :

$$\overline{y}_j \pm Z_{a/2} \cdot SE(\overline{y}_j)$$

ή

$$12{,}3 \pm 1{,}96 \cdot 0{,}238$$

ή

$$12{,}3 \pm 0{,}47$$

δηλαδή το $\overline{Y}_j$ βρίσκεται στο διάστημα από 11.830 δρχ. έως 12.770 δρχ. με πιθανότητα 95% .

Η εκτίμηση της συνολικής δαπάνης όλων των νοικοκυριών που έχουν άτομα ηλικίας πάνω των 65 ετών της πόλης Y_j είναι :

$$\hat{Y}_j = N_j\overline{y}_j = 650 \cdot 12{,}3 = 7.995$$

δηλαδή η συνολική δαπάνη των νοικοκυριών ανέρχεται σε 7.995.000 δρχ.

Η διακύμανση της εκτίμησης $\hat{Y}_j$ είναι :

$$V(\hat{Y}_j) = N_j^2 V(\overline{y}_j) = 650^2 \cdot 0{,}057 = 24.082$$

Το τυπικό σφάλμα είναι :

$$SE(\overline{Y}_j) = \sqrt{V(\overline{Y}_j)} = \sqrt{24.082} = 155$$

Ο συντελεστής μεταβλητικότητας είναι :

$$CV(\hat{Y}_j) = \frac{SE(\hat{Y}_j)}{\hat{Y}_j}100\% = \frac{155}{7.995}100\% = 1.94\%$$

Το διάστημα εμπιστοσύνης στο οποίο βρίσκεται η συνολική δαπάνη των νοικοκυριών Y_j , με πιθανότητα 95% είναι :

$\hat{Y}_j \pm Z_{\alpha/2} \cdot SE(\hat{Y}_j)$

ή

$7.995 \pm 1{,}96 \cdot 155$

ή

7.995 ± 304

δηλαδή η συνολική δαπάνη των νοικοκυριών για νοσοκομειακή περίθαλψη βρίσκεται στο διάστημα από 7.691.000 δρχ. έως 8.299.000 δρχ. με πιθανότητα 95% .

Εάν ο αριθμός των νοικοκυριών της πόλης που έχουν άτομα ηλικίας πάνω των 65 ετών είναι άγνωστος, τότε η εκτίμηση της συνολικής δαπάνης των νοικοκυριών Y_j είναι :

$$\hat{Y}_j = \frac{N}{n}\sum_{K=1}^{n_j} y_{jK} = \frac{1.000}{100}739 = 7.390$$

Η διακύμανση της εκτίμησης $\hat{Y}_j$ είναι :

$$V(\hat{Y}_j) = N^2\left(\frac{N-n)}{N}\right)\frac{\sum_{K=1}^{n_j} y_{jK}^2 - \frac{1}{n}(\sum_{K=1}^{n_j} y_{jK})^2}{n(n-1)} =$$

$$1.000^2\left(\frac{1.000-100}{1.000}\right)\frac{8.985 - \frac{1}{100}739^2}{100(100-1)} = 320.363$$

Το τυπικό σφάλμα είναι :

$$SE(\hat{Y}_j) = \sqrt{V(\hat{Y}_j)} = \sqrt{320.363} = 566$$

Ο συντελεστής μεταβλητικότητας είναι :

$$CV(\hat{Y}_j) = \frac{SE(\hat{Y}_j)}{\hat{Y}_j}100\% = \frac{566}{7.390}100\% = 7{,}65\%$$

Το διάστημα εμπιστοσύνης στο οποίο βρίσκεται η συνολική δαπάνη των νοικοκυριών Y_j , με πιθανότητα 95% είναι :

$$\hat{Y}_j \pm Z_{\alpha/2} \cdot SE(\hat{Y}_j) = 7.390 \pm 1{,}96 \cdot 566 = 7.390 \pm 1.109$$

δηλαδή η συνολική δαπάνη των νοικοκυρών για νοσοκομειακή περίθαλψη βρίσκεται στο διάστημα από 6.281.000 δρχ. έως 8.499.000 δρχ.

12.2.2 Παράδειγμα στρωματοποιημένης σε συνδυασμό με κατά ομάδες δειγματοληψία

Παράδειγμα 2:

Έστω ότι θέλουμε να εκτιμήσουμε το μέσο μηνιαίο εισόδημα και το συνολικό εισόδημα των νοικοκυριών σε δύο συνοικίες της Αθήνας . Κάθε συνοικία αποτελεί ένα στρώμα . Εφαρμόζοντας τη δειγματοληψία κατά ομάδες , όπου κάθε οικοδομικό τετράγωνο του χαρτογραφικού διαγράμματος των συνοικιών αποτέλεσε μία ομάδα νοικοκυριών , πήραμε δείγμα 10 ομάδων από τις 150 της συνοικίας Α και 12 ομάδες από τις 200 της συνοικίας Β .
Από την έρευνα προέκυψαν τα παρακάτω αποτελέσματα .

Ομάδα i		Αριθμός νοικοκυριών m_i		Συνολικό εισόδημα κατά Ομάδα (σε χιλ.) y_i	
Στρώμα Α	Στρώμα Β	Στρώμα Α	Στρώμα Β	Στρώμα Α	Στρώμα Β
1	1	10	11	1.200	1.400
2	2	15	14	1.900	2.000
3	3	20	22	2.500	2.800
4	4	12	13	1.300	1.500
5	5	25	23	3.000	3.100
6	6	18	19	2.400	2.600
7	7	16	17	1.980	2.100
8	8	21	20	2.700	2.800
9	9	24	25	2.900	3.200
10	10	17	16	2.350	2.400
	11		19		2.700
	12		21		3.000
Σύνολο		178	220	22.300	29.600

Να εκτιμηθούν :

α) το μέσο μηνιαίο εισόδημα των νοικοκυριών , το τυπικό σφάλμα , ο συντελεστής μεταβλητικότητας και το διάστημα εμπιστοσύνης με πιθανότητα 95% .

β) το συνολικό εισόδημα των νοικοκυριών , το τυπικό σφάλμα , ο συντελεστής μεταβλητικότητας και το διάστημα εμπιστοσύνης με πιθανότητα 95% .

Λύση

α) Από τα δεδομένα του δείγματος υπολογίζουμε τους μέσους ολικούς των ομάδων και τα μέσα μεγέθη ομάδων κατά στρώμα , δηλαδή : μέσοι ολικοί των ομάδων κατά στρώμα

$$\bar{y}_{t1} = \frac{\sum_{i=1}^{n_1} y_{1i}}{n_1} = \frac{22.230}{10} = 2.223$$

$$\bar{y}_{t2} = \frac{\sum_{i=1}^{n_2} y_{2i}}{n_2} = \frac{29.600}{12} = 2.466$$

μέσα μεγέθη ομάδων κατά στρώμα

$$\bar{m}_1 = \frac{\sum_{i=1}^{n_1} m_{1i}}{n_1} = \frac{178}{10} = 17.8$$

$$\bar{m}_2 = \frac{\sum_{i=1}^{n_2} m_{2i}}{n_2} = \frac{220}{12} = 18,3$$

Η εκτίμηση του μέσου μηνιαίου εισοδήματος των νοικοκυριών είναι :

$$\bar{y}_{cs} = \frac{\sum_{j=1}^{K} \bar{y}_{tj}}{\sum_{j=1}^{K} N_j \bar{m}_j} = \frac{N_1 \bar{y}_{t1} + N_2 \bar{y}_{t2}}{N_1 \bar{m}1 + N_2 \bar{m}_2} = \frac{150(2.223) + 200(2.466)}{150 \cdot 17.8 + 200 \cdot 18.3} = 130.59$$

Επειδή το M είναι άγνωστο εκτιμάται από τη σχέση :

$$M = \sum_{j=1}^{K} N_j \bar{m}_j = N_1 \bar{m}_1 + N_2 \bar{m}_2 = 150 \cdot 17.8 + 200 \cdot 18.3 = 6.330$$

Οπότε η διακύμανση του μέσου $\bar{y}_{cs}$ είναι :

$$V(\bar{y}_{cs}) = \frac{1}{M^2} \sum_{j=1}^{K} \frac{N_j(N_j - n_j)}{n_j(n_j - 1)} \sum_{i=1}^{n_j} \left[(y_{ji} - \bar{y}_{tj}) - \bar{y}_{cs}(m_{ji} - \bar{m}_j)\right]^2 =$$

$$= \frac{1}{M^2} \left[\frac{N_1(N_1 - n_1)}{n_1(n_1 - 1)} \left[(y_{1i} - \bar{y}_{t1}) - \bar{y}_{cs}(m_{ji} - \bar{m}_j)\right]^2 + \right.$$

$$\left. \frac{N_2(N_2 - n_2)}{n_2(n_2 - 1)} \left[(y_{2i} - \bar{y}_{t2}) - y_{cs}(m_{2i} - \bar{m}_2)\right]^2 \right] =$$

$$= \frac{1}{6.330} \left[\frac{150(150 - 10)}{10(10 - 1)} \cdot 153.508.6 + \frac{200(200 - 12)}{12(12 - 1)} \cdot 300.455,9 \right] = 3,05$$

Το τυπικό σφάλμα της εκτίμησης $\bar{y}_{cs}$ είναι :

$$SE(\bar{y}_{cs}) = \sqrt{V(\bar{y}_{cs})} = \sqrt{3.05} = 1.75$$

Ο συντελεστής μεταβλητικότητας είναι :

$$CV(\bar{y}_{cs}) = \frac{SE(\bar{y}_{cs})}{\bar{y}cs} = \frac{1.75}{130.59} = 0.013 \quad \text{ή } 1{,}3\%$$

Το διάστημα εμπιστοσύνης στο οποίο βρίσκεται το μέσο μηνιαίο εισόδημα των νοικοκυριών $\bar{Y}$ των δύο συνοικιών της Αθήνας με πιθανότητα 95% είναι :

$\bar{y}_{cs} \pm Z_{\alpha/2} \cdot SE(\bar{y}_{cs})$

ή

$130.59 \pm 1.96 \cdot 1.75$

ή

130.59 ± 3.43

δηλαδή ο μέσος όρος $\bar{Y}$ βρίσκεται στο διάστημα 127.160 δρχ. – 134. 020 δρχ. με πιθανότητα 95% και με την προϋπόθεση ότι η εκτίμηση $\bar{y}_{cs}$ ακολουθεί την κανονική κατανομή .

β) Η εκτίμηση του συνολικού εισοδήματος των νοικοκυριών στις δύο συνοικίες είναι :

$$\hat{Y} = M\bar{y}_{cs} = 6.330 \cdot 130.59 = 826.634{,}7$$

Η διακύμανση της εκτίμησης του συνολικού εισοδήματος των νοικοκυριών είναι:

$$V(\hat{Y}) = V(M\bar{y}_{cs}) = M^2 V(\bar{y}_{cs}) = 121\ .402\ .682$$

Το τυπικό σφάλμα της εκτίμησης $\hat{Y}$ είναι :

$$SE(\hat{Y}) = SE(M\bar{y}_{cs}) = \sqrt{V(M\bar{y}_{cs})} = \sqrt{121.402.682} \simeq 11.020$$

Ο συντελεστής μεταβλητικότητας είναι :

$$CV(\hat{Y}) = \frac{SE(\hat{Y})}{\hat{Y}} = \frac{11.020}{826.634{,}7} = 0.013 \quad \text{ή } 1{,}3\%$$

Το διάστημα εμπιστοσύνης στο οποίο βρίσκεται η εκτίμηση του συνολικού εισοδήματος των νοικοκυριών με πιθανότητα 95% είναι :

$\hat{Y} \pm Z_{\alpha/2} \cdot SE(\hat{Y})$

ή

$826.634{,}7 \pm 1.96 \cdot 11.020$

ή

$826.634{,}7 \pm 21.599{,}2$

δηλαδή το συνολικό εισόδημα των νοικοκυριών στις δύο συνοικίες βρίσκεται στο διάστημα 805.035.500 δρχ. – 848.223.900 δρχ. με πιθανότητα 95% .

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. **Στατιστική Επιχειρήσεων**

   (Γ . Δροσος – Δ . Καραπιστολης) Εκδόσεις Ελλην

2. **Στατιστική Επιχειρήσεων**

   (Βασιλη Χουβαρδα) Μακεδονικές Εκδόσεις

3. **Εισαγωγή Στην Θεωρία Δειγματοληψίας**

   (Ν. Α. Τσέρπες – Φ . Αλεβίζος)

4. **Τεχνικές Δειγματοληψίας**

   (Ε. Ξεκαλάκη)

5. **Μέθοδοι Έρευνας Αγοράς (Βλάσης Σταθόπουλος)**

   Εκδόσεις Α. Σταμούλης

6. **Στατιστική Των Επιχειρήσεων**

   (Douglas Downingphd,Jefreyclark phd) Εκδόσεις Κλειδάριθμος

7. **Στατιστική Περιγραφική Και Επαγωγική**

   (Δημήτρης Καραγεώργος) Εκδόσεις Σαββάλας

8. **Οργάνωση Και Διεξαγωγή Δειγματοληπτικών Ερευνών**

   (Παναγιώτης Θ . Τζορτζόπουλος)

9. **Εφηρμοσμένο Μάρκετινγκ**

   (Έρευνα Αγοράς) (Π. Γ. Κυριαζόπουλος – Κ. Κ. Κιουλάφας)

10. **Θεωρία Δειγματοληψίας Και Εφαρμογές**

    (Κ . Ι . Χαρίσης Πέτρος , Α. Κιόχος) Εκδόσεις Interbox .

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΠΑΤΡΑΣ ΒΙΒΛΙΟΘΗΚΗ