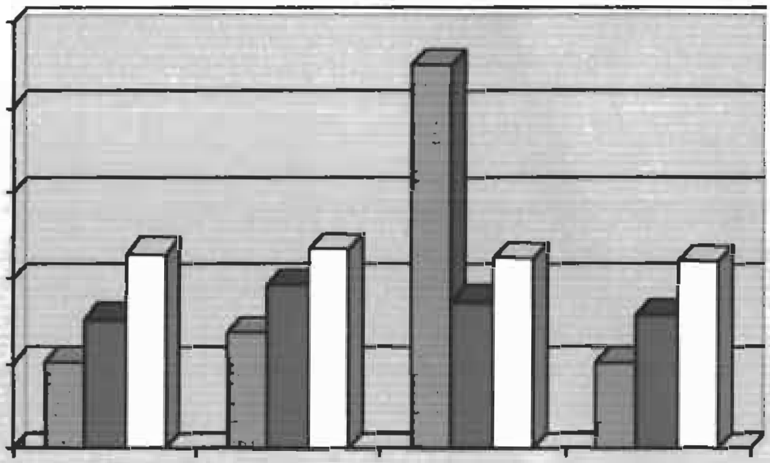
ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΠΑΤΡΑΣ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ
ΠΑΤΡΑ 1999

ΘΕΜΑ

ΕΙΣΑΓΩΓΙΚΕΣ ΤΕΧΝΙΚΕΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΣΤΗΝ ΕΚΤΙΜΗΣΗ ΠΑΡΑΜΕΤΡΩΝ ΤΟΥ ΠΛΗΘΥΣΜΟΥ



ΚΑΘΗΓΗΤΗΣ Κ. ΦΩΤΟΠΟΥΛΟΣ

ΣΠΟΥΔΑΣΤΡΙΕΣ ΜΠΑΚΑΜΗΤΣΟΥ ΑΜΑΛΙΑ
ΛΑΒΑΤΣΗ ΠΑΡΑΣΚΕΥΗ

ΠΕΡΙΕΧΟΜΕΝΑ

Κεφάλαιο 1ο

Εισαγωγή

ΚΕΦΑΛΑΙΟ 1ο

ΕΙΣΑΓΩΓΗ

Η γνώση μας, η συμπεριφορά μας, οι πράξεις μας και οι αποφάσεις μας τόσο στην καθημερινή ζωή όσο και στην επιστημονική έρευνα, βασίζονται πάνω σε δείγματα.

Η συλλογή και επεξεργασία δεδομένων από πεπερασμένους πληθυσμούς είναι μια σημαντική και ενδιαφέρουσα στατιστική περιοχή από πρακτική και ωφελιμιστική πλευρά και αποτελεί ένα σύνθετο τομέα μελέτης και έρευνας ως προς την στατιστική θεωρία και μεθοδολογία.

Στις δειγματοληπτικές έρευνες επιλέγονται δείγματα από πεπερασμένους πληθυσμούς τα οποία χρησιμοποιούνται για να αντικατοπτρίσουν τις ιδιότητες του πληθυσμού που ισχυρίζονται ότι εκπροσωπούν.

Με άλλα λόγια η θεωρία της **δειγματοληψίας** από πεπερασμένους πληθυσμούς χρησιμοποιείται στις δειγματοληπτικές έρευνες (ή επισκοπήσεις) που έχουν ως σκοπό την εκτίμηση ορισμένων χαρακτηριστικών ενός πεπερασμένου πληθυσμού (του πληθυσμού-στόχου όπως λέγεται), με την χρήση ενός καταλλήλως επιλεγμένου (πιθανοτικού) δείγματος από αυτόν τον πληθυσμό.

Τα πλεονεκτήματα της δειγματοληψίας, ως μεθόδου εξαγωγής συμπερασμάτων είναι τα εξής:

- Χαμηλότερο κόστος, διότι οι πληροφορίες και τα δεδομένα προέρχονται από τμήμα μόνο του πληθυσμού
- Μεγαλύτερη ταχύτητα στη συλλογή και επεξεργασία των δεδομένων για τον ίδιο λόγο
- Μεγαλύτερη ακρίβεια, δεδομένου ότι γίνεται πιο προσεκτική εποπτεία της διεξαγωγής της έρευνας και πιο προσεκτική επεξεργασία των αποτελεσμάτων που προκύπτουν από την έρευνα.

Σήμερα έχουν ξεπεραστεί και τα τελευταία ίχνη καχυποψίας που υπήρχαν ως προς την αξιοπιστία των πληροφοριών που περιέχονται σε ένα δείγμα και έτσι η δειγματοληψία είναι πια κοινή πρακτική σε πολλούς τομείς της καθημερινής ζωής, όπως στις βιομηχανίες, στο μάρκετινγκ, στις έρευνες κοινής γνώμης και σε πολλούς άλλους τομείς οικονομικής, κοινωνικής, πολιτικής κλπ., φύσεως. Η σημερινή κοινωνία κατακλύζεται από αριθμούς. Υπάρχει μια τάση να εκφράζονται ποσοτικά όλες οι πλευρές της καθημερινής ζωής, από τις καθημερινές συνήθειες μέχρι τα πολιτικά ιδεώδη. Η παρουσίαση στατιστικών αριθμών έχει βασικό σκοπό να μας κρατήσει ενήμερους για την κατάσταση που επικρατεί γύρω μας και χρησιμοποιείται συχνά για την υποστήριξη κάποιας πρότασης, γνώμης, κριτικής ή ακόμα για την τοποθέτηση ενός θέματος στη σωστή του βάση.

Αναμφισβήτητα τα άτομα σήμερα είναι περισσότερο από ποτέ πληροφορημένα, με την έννοια ότι δέχονται ποσοτικές περιγραφές του κόσμου στον οποίο ζουν. Το γεγονός αυτό είναι βέβαια πολύ καλό, αλλά παράλληλα επιβάλλει σοβαρές απαιτήσεις τόσο στον δέκτη, όσο και στον πομπό των αριθμητικών πληροφοριών. Από την μια πλευρά, ο δέκτης, ο απλός άνθρωπος, χρειάζεται να έχει την ικανότητα να κατανοεί και να ερμηνεύει την πληροφόρηση που του παρέχεται καθώς και να την αξιοποιεί ουσιαστικά. Από την άλλη πλευρά εκείνοι που παρουσιάζουν τα στατιστικά δεδομένα έχουν ευθύνη διπλή. Πρέπει να παρουσιάζουν τα στοιχεία με έντιμο και αντικειμενικό τρόπο και χωρίς την πρόθεση παραπλάνησης του κοινού και επίσης πρέπει να παρέχουν επαρκείς πληροφορίες σχετικά με την πηγή, το στόχο και την μέθοδο συλλογής δεδομένων τα οποία χρησιμοποίησαν, για την κατάλληλη ερμηνεία της οποιασδήποτε παραπέρα ανάλυσης.

Είναι λοιπόν λογικό, ότι στην διάρκεια διεξαγωγής μιας δειγματοληπτικής έρευνας είναι δυνατό να συμβούν λάθη, τα όποια , μπορεί να είναι είτε εγκυρότητας, είτε αξιοπιστίας. Για το λόγο αυτό, η επιλογή της δειγματοληπτικής τεχνικής που θα χρησιμοποιηθεί καθώς και του μεγέθους του δείγματος που απαιτείται, πρέπει να ελεγχθούν σωστά και να ληφθούν υπόψη βασικά πρακτικά προβλήματα που αφορούν τόσο τον σχεδιασμό όσο και την διεξαγωγή της.

Τέτοια προβλήματα μπορεί να είναι:

- Ο ορισμός του πληθυσμού και του χαρακτηριστικού που επιδιώκεται να εκτιμηθεί
- Η μέθοδος λήψης του δείγματος (το οποίο είναι και το αντικείμενό μας)

- Η μεροληπτικότητα των απαντήσεων ή των μη-απαντήσεων
- Οι πιλοτικές έρευνες
- Ο σχεδιασμός του ερωτηματολογίου και η διατύπωση των ερωτήσεων
- Η τεχνική της συνέντευξης
- Ο προϋπολογισμός της έρευνας
- Η χρήση συμπληρωματικής πληροφόρησης κλπ.

Δεδομένου ότι η θεωρία δειγματοληψίας μπορεί χονδρικά να χωριστεί σε δυο βασικές διαδικασίες.

1. Τον σχεδιασμό του δείγματος και τον σχεδιασμό των εκτιμητών που μας δίνει αυτή η δειγματοληψία και

2. Την ανάλυση και εξαγωγή συμπερασμάτων.

Κύριος στόχος των ερευνητών πρέπει να είναι, αφενός η ελάττωση της διασποράς των εκτιμητών (δηλαδή ευρέα διαστήματα εμπιστοσύνης) και αφετέρου η ελάττωση ή εξάλειψη της μεροληψίας στους εκτιμητές.

Ερχόμαστε λοιπόν ξανά σε περιπτώσεις που συνδέονται με πεπερασμένους πληθυσμούς, οι οποίοι περιέχουν ένα περιορισμένο και καλώς ορισμένο σύνολο ατόμων ή μονάδων. Ανεξάρτητα από το μέγεθος του πληθυσμού, ο στόχος είναι η συμπερασματολογία σχετικά με τον πεπερασμένο αυτό πληθυσμό, με βάση τη συλλογή και ανάλυση πληροφοριών που σχετίζονται με ένα μέρος του, αυτό που ονομάζουμε δείγμα από τον πληθυσμό. Οι πληροφορίες αυτές συλλέγονται από έρευνα και επισκόπηση (survey) του πληθυσμού.

Οι δειγματοληπτικές έρευνες μπορούν να καλύψουν οποιοδήποτε θέμα συνδέεται με τα χαρακτηριστικά ενός πεπερασμένου πληθυσμού και να περιγράψουν τους ανθρώπινους πληθυσμούς και το περιβάλλον τους.

Κατηγορίες που αφορούν οι δειγματοληπτικές έρευνες είναι οι εξής:

- Δημογραφικά χαρακτηριστικά του πληθυσμού
- Οικονομική δομή της κοινωνίας
- Τρόπος ζωής
- Κοινωνικό περιβάλλον
- Απόψεις και γνώμες

Η μελέτη του τρόπου διεξαγωγής τέτοιων δειγματοληπτικών ερευνών (sample surveys) αποτελεί το αντικείμενο της εργασίας αυτής.

Προκειμένου δε, να μελετήσουμε τις τεχνικές δειγματοληψίας, είναι απαραίτητο να γίνει πρώτα μια αναφορά σε βασικές έννοιες που θα χρησιμοποιηθούν.

Κεφάλαιο 2º

Ανάλυση Εννοιών

ΚΕΦΑΛΑΙΟ 2ο

ΑΝΑΛΥΣΗ ΕΝΝΟΙΩΝ

2.1 Πεπερασμένος Πληθυσμός

Είναι αυτός που περιέχει περιορισμένο και σαφώς ορισμένο σύνολο γνώσεων, ο οποίος δεν είναι απαραίτητο να είναι ανθρώπινος. Πρέπει να οριστεί με ακρίβεια διότι σ΄ αυτόν θα γενικευτεί το ερευνητικό έργο. Ο ορισμός του συνόλου του πληθυσμού περιλαμβάνει τρία χαρακτηριστικά:

1. Τα χαρακτηριστικά των επιμέρους μονάδων του εξεταζόμενου πληθυσμού

2. Τη διάσταση του χρόνου που σχετίζεται με τα εξεταζόμενα (πρωτογενή) στοιχεία

3. Το γεωγραφικό πεδίο του σχετικού πληθυσμού

2.2 Αντικειμενικός πληθυσμός

Είναι το σύνολο του πεπερασμένου πληθυσμού για τον οποίο ενδιαφερόμαστε να συλλέξουμε πληροφορίες. Π.χ. το σύνολο των εργαζομένων στην Ελλάδα κάτω των 30 ετών.

2.3 Υπό μελέτη πληθυσμός

Είναι το βασικό πεπερασμένο σύνολο ατόμων που θέλουμε να μελετήσουμε . Π.χ. το σύνολο των εργαζομένων στην Ελλάδα ηλικίας κάτω των 30 ετών των οποίων η μόνιμη κατοικία είναι μέσα στην περιοχή των Πατρών. Ο πληθυσμός αυτός ενδέχεται να ταυτίζεται με τον αντικειμενικό πληθυσμό. Συχνά όμως ενδέχεται να είναι ένας περισσότερο περιορισμένος σε μέγεθος αλλά και περισσότερο εύκολο να μελετηθεί, του οποίου οι ιδιότητες ελπίζεται ότι μπορούν να προβληθούν στον ευρύτερο αντικειμενικό πληθυσμό.

2.4 Χαρακτηριστικό του πληθυσμού

Είναι η πλευρά του πληθυσμού που επιθυμούμε να μελετήσουμε (να εκτιμήσουμε). Π.χ. το εισόδημα των εργαζομένων στην περιοχή της Πάτρας ηλικίας κάτω από 30 ετών ή το είδος απορρυπαντικών πλυντηρίου που χρησιμοποιούν οι Αθηναίες νοικοκυρές.

2.5 Δειγματοληπτικές Μονάδες (Sampling Units)

Είναι οι μονάδες που επιλέγουμε με σκοπό να επιτύχουμε πρόσβαση στα άτομα (μονάδες) του υπό μελέτη πληθυσμού. Π.χ. ας υποθέσουμε ότι επιθυμούμε να κάνουμε μια έρευνα οικογενειακών προϋπολογισμών σε κάποια πόλη. Αν και είναι προφανές ότι τα άτομα (μονάδες) του υπό μελέτη πληθυσμού είναι οικογένειες, πρέπει να υιοθετηθεί κάποιος συμβατικός ορισμός της έννοιας της οικογένειας πριν προχωρήσουμε στην έρευνα. Ακόμη και σε αυτή την περίπτωση ενδέχεται να μην υπάρχει εύκολος τρόπος δειγματοληψίας τέτοιων μονάδων (οικογενειών). Θα ήταν ίσως ευκολότερο να ληφθεί ένα δείγμα διευθύνσεων και να αναζητηθούν 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 ουσιώδους σημασίας.

2.6 Δειγματοληπτικό πλαίσιο (Sampling frame)

Είναι το σύνολο των δειγματοληπτικών μονάδων το οποίο αποτελεί και την πηγή του δείγματός μας στα διάφορα στάδια της δειγματοληψίας. Μερικές φορές οι δειγματοληπτικές μονάδες ταυτίζονται με τα μέλη του υπό μελέτη πληθυσμού. Στις περισσότερες περιπτώσεις, όμως, αυτό, δεν συμβαίνει και έτσι το δειγματοληπτικό πλαίσιο είναι μια λιγότερο ακριβής υποδιαίρεση του υπό μελέτη πληθυσμού (coarser cubvision) κάθε μονάδα της οποίας (δειγματοληπτικής μονάδας) περιέχει ένα διακεκριμένο σύνολο μελών του πληθυσμού.

2.7 Κατάλογος ή Λίστα (List)

Για να χρησιμοποιηθεί το δειγματοληπτικό πλαίσιο ως η πρωταρχική πηγή από την οποία θα επιλεγεί το δείγμα θα πρέπει να είναι δυνατός ο προσδιορισμός των δειγματοληπτικών μονάδων. Εξάλλου το δειγματοληπτικό πλαίσιο επιλέγεται με αυτό το σκοπό. Στην καλύτερη περίπτωση ένας πραγματικός κατάλογος όλων των δειγματοληπτικών μονάδων ενδέχεται να υπάρχει, όπως κατάλογος διευθύνσεων, εργαζομένων στην περιοχή Πατρών, λίστα κρατικών απογραφών νοικοκυριών, εκλογικοί κατάλογοι κτλ. Ένας τέτοιος κατάλογος διευκολύνει ιδιαίτερα την επιλογή του δείγματος.

2.8. Απογραφή

Κατά κανόνα η έρευνα θα μπορούσε να βασισθεί στην επισκόπηση όλων των μελών ενός πεπερασμένου πληθυσμού. Μια τέτοια διαδικασία αποτελεί την απογραφή (Census). Η απογραφή είναι δηλαδή μια δειγματοληπτική έρευνα με κάλυψη 100%. Το ενδιαφέρον όμως των ερευνητών εστιάζεται συνήθως σε πολύ χαμηλότερα επίπεδα της τάξεως του 10% ή 5%.

Απογραφή επιχειρείται συνήθως σε τρεις περιπτώσεις:

1. Όταν είναι σπουδαίο να γνωρίζούμε κάτι γύρω από κάθε χαρακτηριστικό (ή μονάδα αν μελετώνται οργανισμοί) στο σύνολο του πληθυσμού.

2. Όταν οι κίνδυνοι μιας σωστής γενίκευσης είναι πολύ μεγάλοι ακόμα και όταν ερευνάται μεγάλο ποσοστό δειγμάτων του συνόλου του πληθυσμού.

3. Όταν το σύνολο του πληθυσμού είναι τόσο μικρό ώστε μπορεί να γίνει απογραφή με τους υπάρχοντες προϋπολογισμούς και χρονικούς περιορισμούς.

Δεδομένου ότι σπάνια είναι απαραίτητο να γνωρίζουμε όλα τα στοιχεία για κάθε μονάδα του συνόλου του πληθυσμού προτιμώνται συνήθως αντιπροσωπευτικά δείγματα τα οποία γενικεύονται στη συνέχεια για το σύνολο του πληθυσμού.

Συμπερασματικά το κόστος διεξαγωγής μιας απογραφής υπερτερεί των κινδύνων μιας λανθασμένης γενίκευσης. Εκτός αυτού όμως, υπάρχουν και άλλοι λόγοι που ένα δείγμα είναι καλύτερο μιας απογραφής, όπως ο κίνδυνος να αποκαλυφθεί μια επιτυχής ιδέα στους ανταγωνιστές, η μεγάλη χρονική διάρκεια που χρειάζεται για τη συλλογή στοιχείων εις όφελος ανταγωνιστών, το μεγάλο κόστος καθώς και η αύξηση του δειγματοληπτικού σφάλματος με την απογραφή λόγω έλλειψης ειδικευμένου προσωπικού. Παρόλα αυτά η απογραφή είναι η σπουδαιότερη πηγή πληροφόρησης, γιατί αποτελεί κύρια πηγή πληροφόρησης σχετικά με οικονομικά, δημογραφικά και κοινωνικά χαρακτηριστικά.

2.9. Εκτίμηση πιθανού σφάλματος πλαισίου της έρευνας

Αναφέρεται στον υπόλοιπο μη δειγματοληπτικό πληθυσμό, δηλαδή στις δειγματοληπτικές μονάδες εκτός αυτών από τις οποίες σχεδιάστηκε το δείγμα. Αυτό συμβαίνει διότι πρακτικά, οι κατάλογοι που χρησιμοποιούνται στην πραγματική δειγματοληψία δεν είναι πάντοτε ταυτόσημοι στον επιθυμητό πληθυσμό. Υπάρχουν αρκετοί λόγοι γι' αυτό. Μπορεί να μην είναι διαθέσιμος κανένας πρόσφατος ακριβής κατάλογος πληθυσμού. Χωρίς αυτόν τον κατάλογο υπάρχει πιθανότητα για σφάλματα πλαισίου ή μια δυσαρμονία του πληθυσμού του δείγματος και του συνολικού πληθυσμού. Η πραγματικότητα είναι ότι σχεδόν κάθε μέθοδος μη τυχαίας δειγματοληψίας έχει κάποιο σφάλμα πλαισίου.

Δυστυχώς υπάρχουν λίγα πράγματα που μπορεί να κάνει ο ερευνητής για τον υπολογισμό του μεγέθους του πιθανού σφάλματος πλαισίου. Θα πρέπει επίσης να λάβει υπόψη του ότι:

1. Το πιθανό σφάλμα πλαισίού είναι αρκετά σοβαρό ώστε να επηρεάσει την αξία του σχεδιασμού.

2. Μπορεί το πιθανό σφάλμα πλαισίου να μειωθεί.

2.10. Δειγματοληπτικά σφάλματα και μη δειγματοληπτικά σφάλματα

Το συνολικό σφάλμα ενός ερευνητικού έργου μπορεί να διαιρεθεί στις εξής δύο κατηγορίες:

- Δειγματοληπτικά σφάλματα (αξιοπιστία): Αυτά λαμβάνουν χώρα όταν τα τμήματα του συνόλου του πληθυσμού που έχουν επιλεγεί, για να ερωτηθούν ή να μελετηθούν, δεν είναι πραγματικά αντιπροσωπευτικά του συνόλου του πληθυσμού. Οφείλονται δηλαδή στην τυχαία επιλογή των μονάδων του δείγματος. Αν υποθέσουμε ότι χρησιμοποιείται η κατάλληλη δειγματοληπτική διαδικασία, είναι προφανές ότι όσο το μέγεθος του δείγματος αυξάνεται, το δειγματοληπτικό σφάλμα μειώνεται, ενώ για μια απογραφή αυτό ισούται με το μηδέν. Όταν συμβαίνει δειγματοληπτικό σφάλμα γίνεται λάθος στην προβολή ευρημάτων της έρευνας στο σύνολο του εξεταζόμενου πληθυσμού.
- Μη δειγματοληπτικά σφάλματα: Αυτά είναι σφάλματα που οφείλονται στην απροσεξία της ομάδας των ατόμων που ασχολούνται με την δειγματοληψία και συγκεκριμένα

  1. Σε λανθασμένες πληροφορίες
  2. Σε λανθασμένες μετρήσεις και καταχωρήσεις
  3. Στο γεγονός ότι το πλαίσιο πού χρησιμοποιούμε για την επιλογή του δείγματος δεν είναι πλήρως ενημερωμένο και σωστό
  4. Στο ότι τα ερωτήματα του ερωτηματολογίου δεν είναι σαφή

5. Στο ότι τα ερωτήματα είναι πολλά, κουραστικά και παραπλανητικά
6. Στη μη ορθή επεξεργασία των ερωτηματολογίων και τη μη σωστή ανάλύση των δεδομένων.

Τα μη δειγματοληπτικά σφάλματα έχουν την τάση να αυξάνονται όσο αυξάνει το μέγεθος του δείγματος. Το σύνολο δειγματοληπτικών και μη δειγματοληπτικών σφαλμάτων πρέπει να επιδιώκουμε να είναι το ελάχιστο δυνατό.

2.11. Συστηματικά και τυχαία σφάλματα

Συστηματικά είναι τα σφάλματα που προέρχονται από τα χρησιμοποιούμενα όργανα μετρήσεως, τα οποία δίνουν μικρότερες ή μεγαλύτερες τιμές στις μετρήσεις π.χ. τα σφάλματα μέτρησης πού γίνονται από το ζυγό, ο οποίος επειδή είναι απορρυθμισμένος προσθέτει π.χ. σε κάθε πρόσωπο που ζυγίζεται δύο κιλά. Συστηματικά είναι ακόμη, τα σφάλματα από την απόκρυψη ηλικίας και εισοδήματος.

Τυχαία είναι τα σφάλματα εκείνα, τα οποία προέρχονται από λανθασμένες μετρήσεις ή καταχωρήσεις και οφείλονται στην απροσεξία και στην υποκειμενική κρίση των ερευνητών. Τα τυχαία σφάλματα, ενώ υπάρχουν στις επιμέρούς μετρήσεις, συνήθως αλληλοεξουδετερώνονται στο σύνολο των μετρήσεων.

Γενικότερα, το τυπικό σφάλμα ή αλλιώς διακυμαντικό σφάλμα ενός εκτιμητή Θ, είναι ένα άλλο όνομα για την τυπική απόκλιση, δηλαδή την τετραγωνική ρίζα της διασποράς του εκτιμητή Var Θ.

2.12. Διαστήματα εμπιστοσύνης

Η έννοια και το περιεχόμενο των διαστημάτων εμπιστοσύνης είναι ουσιώδη, προκειμένου να εξάγουμε στατιστικά συμπεράσματα. Όσο καλή και αν είναι μια δειγματοληψία, η εκτίμηση που προκύπτει από αυτή, από ένα τυχαίο δείγμα, είναι σχεδόν αδύνατο να συμπέσει ακριβώς με την αληθινή παράμετρο του προς εκτίμηση πληθυσμού.

Για το λόγο αυτό προσδιορίζουμε συνήθως ένα διάστημα, μια σειρά από τιμές γύρω από στατιστική δείγματος, μέσα στο οποίο περιμένουμε να βρίσκεται η παράμετρος αυτή του πληθυσμού π.χ. 10000 ± 1750 δηλαδή το διάστημα θα είναι 8248 -11752. Επίσης προσδιορίζουμε και ένα βαθμό εμπιστοσύνης που σχετίζεται με την παραπάνω σειρά των τιμών γύρω από ένα στατιστικό δείγμα π.χ. 90% κτλ. Αν υποθέσουμε ότι το παραπάνω διάστημα 8248-11752 σχετίζεται με ένα επίπεδο εμπιστοσύνης της τάξεως του 95%, αυτό σημαίνει ότι αν ένας μεγάλος αριθμός δειγμάτων ληφθεί από ένα πληθυσμό, το 95% των δειγμάτων αυτών θα περιλαμβάνει την αληθινή τιμή του πληθυσμού μέσα στο διάστημα που έχει τεθεί. Με άλλα λόγια υπάρχει ένα ποσοστό 95% όπου η σωστή τιμή του πληθυσμού θα είναι μέσα στο διάστημα αυτό, ενώ ποσοστό 5% δεν θα περιλαμβάνεται.

Όσο μεγαλύτερο είναι το διάστημα, τόσο μεγαλύτερο θα είναι το διάστημα εμπιστοσύνης που θα περιλαμβάνεται η αληθινή τιμή του πληθυσμού. Όσο μεγαλύτερες είναι οι απαιτήσεις μας για το επίπεδο εμπιστοσύνης τόσο μεγαλύτερο το διάστημα κοκ.

2.13. Μέγεθος Δείγματος

Το πρώτο ερώτημα με το οποίο βρίσκεται αντιμέτωπος ένας ερευνητής, όταν πρόκειται να κάνει μια δειγματοληπτική έρευνα, είναι πόσο μεγάλο πρέπει να είναι το δείγμα που θα χρησιμοποιήσει. Παράγοντες που μπορούν να επηρεάσουν το μέγεθος του δείγματος είναι οι εξής:

- Το μέγεθος ακρίβειας που επιθυμείται
- Το κόστος της δειγματοληψίας
- Ο χρόνος που απαιτείται να διεξαχθεί η δειγματοληψία
- Ο βαθμός ομοιογένειας του πληθυσμού
- Η τεχνική (μέθοδος) δειγματοληψίας

Η τιμή του δείγματος θα εξαρτηθεί - εκτός από τους παραπάνω παράγοντες-άμεσα, από το μέγιστο ανεκτό σφάλμα εκτίμησης και την πιθανότητα με την οποία αυτό είναι επιτρεπτό.

2.14. Παράμετροι

Μετά την συλλογή των στατιστικών στοιχείων, ακολουθεί η ταξινόμηση και η εμφάνιση των πολυάριθμων παρατηρήσεων με μορφή συνοπτικών πινάκων συχνοτήτων, με σκοπό πάντοτε την εύκολη μελέτη και περιγραφή των δεδομένων της δομής του πληθυσμού που ερευνούμε. Οι πίνακες συχνοτήτων όμως αποτελούν μια σύνθετη έκφραση και αυτοί, διότι δεν μπορούν εύκολα να συγκρατηθούν στη μνήμη και δεν διευκολύνουν τις συγκρίσεις.

Για το λόγο αυτό, θεωρείται πολλές φορές αναγκαία μια παραπέρα συμπύκνωση και συγκεφαλαίωση των πινάκων συχνοτήτων. Η συμπύκνωση αυτή γίνεται με την αντικατάσταση των πινάκων συχνοτήτων από ορισμένους αντιπροσωπευτικούς αριθμούς, οι οποίοι ονομάζονται στατιστικές παράμετροι και χαρακτηρίζουν τη θέση, τη διασπορά και μορφολογία του πληθυσμού που διερευνούμε.

Συγχρόνως διευκολύνουν πολύ στις συγκρίσεις ομοειδών ομάδων. π.χ. μέσος αριθμητικός, διάμεσος, επικρατούσα τιμή κτλ.

2.15. Ακρίβεια των εκτιμήσεων

Τα δειγματοληπτικά σφάλματα μετράνε την ακρίβεια των εκτιμήσεων. Η ακρίβεια, είναι η διαφορά μεταξύ της εκτιμήσεως που προκύπτει από ένα δείγμα και της αντίστοιχης παραμέτρου που προκύπτει μέσω απογραφής. Επομένως η ακρίβεια εκτιμήσεων δεν είναι τίποτε άλλο από το μέγεθος του δειγματοληπτικού σφάλματος. Όσο μικρότερο είναι το δειγματοληπτικό σφάλμα, τόσο μεγαλύτερη είναι η ακρίβεια (αξιοπιστία) μιας εκτίμησης. Η ακρίβεια των εκτιμήσεων εξαρτάται από το τυπικό σφάλμα της εκτίμησης και επομένως, από το μέγεθος του δείγματος. Όσο μεγαλύτερο είναι το μέγεθος του δείγματος τόσο μικρότερο θα είναι το τυπικό σφάλμα εκτίμησης και επομένως τόσο μεγαλύτερη η ακρίβεια εκτίμησης.

Υπάρχουν πολλές δειγματοληπτικές τεχνικές για να επιτύχουμε το ζητούμενο αποτέλεσμα - στόχο μιας δειγματοληψίας.

Στην πράξη - συνήθως - χρησιμοποιούνται δύο ή περισσότερες από αυτές, παράγοντας έτσι τους λεγόμενους σύνθετους σχεδιασμούς επισκοπήσεων. Ειδικότερα έχουμε:

1. Απλή τυχαία δειγματοληψία χωρίς επανάθεση
2. Απλή τυχαία δειγματοληψία με επανάθεση
3. Δειγματοληψία κατά στρώματα ή στρωματοποιημένη
4. Συστηματική δειγματοληψία
5. Δειγματοληψία κατά ομάδες
6. Δειγματοληψία ποσοστών ή ποσοστιαία

Κεφάλαιο 3º

Απλή τυχαία δειγματοληψία χωρίς επανάθεση

ΚΕΦΑΛΑΙΟ 3ο

ΑΠΛΗ ΤΥΧΑΙΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΧΩΡΙΣ ΕΠΑΝΑΘΕΣΗ

3.1. Γενικά

Είναι η στοιχειωδέστερη μορφή δειγματοληψίας κατά πιθανότητα. Αυτή συνήθως εννοούμε όταν λέμε απλή τυχαία δειγματοληψία. Είναι μια μέθοδος Ίσης Πιθανότητας Επιλογής Μονάδων (ΙΠΕΜ).

Ορισμός: Σ΄ αυτή τη μέθοδο κάθε δυνατό υποσύνολο n στοιχείων (μονάδων) από ένα πληθυσμό N μονάδων έχει ίση πιθανότητα να είναι το δείγμα μεγέθους n που επιλέγεται (δηλ. πιθανότητα $1/\binom{N}{n}$ αφού υπάρχουν $\frac{N}{n}$ διαφορετικά δείγματα).

Βασική προϋπόθεση για τον σχηματισμό ενός δείγματος από ένα πληθυσμό είναι ο σαφής καθορισμός του πληθυσμού. Για την επιλογή ενός τέτοιου δείγματος μπορούμε ν΄ αριθμήσουμε τις μονάδες του πληθυσμού με διαφορετικούς αριθμούς και μετά κάνοντας χρήση ενός πίνακα τυχαίων αριθμών να τραβήξουμε τυχαίους αριθμούς από τον Πίνακα ώσπου να αντιστοιχήσουμε n διαφορετικά στοιχεία από τον πληθυσμό. Ίση πιθανότητα στη μέθοδο αυτή, αναφέρεται στο γεγονός ότι η ολική πιθανότητα για την μονάδα a_i για κάθε i να συμπεριληφθεί στο δείγμα, είναι $\binom{N-1}{n-1}/\binom{N}{n} = n/N =$ το δειγματοληπτικό κλάσμα.

Επίσης η πιθανότητα επιλογής της μονάδας a_i σ΄ ένα συγκεκριμένο τράβηγμα είναι $1/N$, αλλά η υπό συνθήκη πιθανότητα ότι το i- στο τράβηγμα παράγει το a_i γνωρίζοντας ότι ένα προγενέστερο τράβηγμα έδωσε a_i $(i \neq J)$, είναι $1/N-1$. Συνεπώς τα τραβήγματα (οι μεταβλητές ψi) δεν είναι ανεξάρτητα.

Το σχήμα - τεχνική αυτή χρησιμοποιείται ευρύτατα, κυρίως λόγω της απλότητας του από την άποψη της στατιστικής συμπερασματολογίας. Πέρα απ΄ την αυτοτελή χρήση της, χρησιμεύει και ως βάση για συνθετότερες δειγματοληπτικές Μεθόδους, όπως για παράδειγμα η στρωματατοιημένη απλή τυχαία δειγματοληψία και η δειγματοληψία κατά ομάδες. Υπάρχουν 2 τρόποι λήψης δείγματος στην απλή τυχαία δειγματοληψία.

α) Η μέθοδος των τυχαίων αριθμών

β) Η μέθοδος των λαχνών

Αναλυτικότερα:

α) Η επιλογή ενός απλού τυχαίου δείγματος μπορεί να γίνει με τη βοήθεια πινάκων τυχαίων αριθμών. Αυτοί 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.

Ο πίνακας 1 είναι απόσπασμα 1.000 ψηφίων του πίνακα τυχαίων αριθμών των Snedecar και Cochran (1967).

Για την επιλογή ενός απλού τυχαίου δείγματος μεγέθους n από ένα πληθυσμό μεγέθους N, αντιστοιχούμε καθεμιά από τις μονάδες του πληθυσμού ένα αριθμό από 1 ως N (διαφορετικό για κάθε μονάδα). Διαλέγουμε τυχαία τόσες στήλες όσες τα ψηφία του N και διαβάζουμε προς μία κατεύθυνση π.χ προς τα κάτω την συγκεκριμένη ομάδα στηλών επιλέγοντας τους αριθμούς που είναι ≤ N. Για παράδειγμα έστω N= 198 & n=5. Έστω ότι διαλέγουμε τις στήλες 10-12. Ξεκινώντας απ΄ την γραμμή 00 & διαβάζοντας π.χ προς τα κάτω, οι πρώτοι 5 διακεκριμένοι αριθμοί είναι 188, 112, 106, 108, 72. Το μειονέκτημα της μεθόδου αυτής είναι ότι οι τριψήφιοι αριθμοί 199 & 999 δεν χρησιμοποιούνται (απορρίπτονται).

Μια άλλη μέθοδος που απαιτεί λιγότερες απορρίψεις αριθμών είναι η εξής: αφαιρούμε 200 από τους τριψήφιους αριθμούς μεταξύ 201 και 400, 400 από τους τριψήφιους μεταξύ 401 και 600, 600 από τους τριψήφιους μεταξύ 801 και 999. Φυσικά αφαιρούμε 000 από κάθε τριψήφιο μεταξύ 000 & 200. Στο δείγμα περιλαμβάνουμε τις παραπάνω διαφορές, ενώ απορρίπτουμε τους 000, 199, 200 και όλους τους αριθμούς που μετά την αφαίρεση είναι μεγαλύτεροι του 198. Για π.χ, αν ξεκινήσουμε από την γραμμή 05 του πίνακα και διαβάζουμε προς τα κάτω τους τριψήφιους αριθμούς των στηλών 15-17 επιλέγουμε τους εξής διακεκριμένους αριθμούς : 122, 157, 30, 59, 11.

Πίνακας 1

Τυχαίοι Αριθμοί

	00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
00	54463	22662	65905	70639	79365	67382	29085	69831	47058	08186
01	15389	85205	18850	39226	42249	90669	96325	23248	60933	26927
02	85941	40756	82414	02015	13858	78030	16269	65978	01385	15345
03	61149	69440	11286	88218	58925	03638	52862	62733	33451	77455
04	05219	81619	10651	67079	92511	59888	84502	72095	83463	75577
05	41417	98326	87719	92294	46614	50948	64886	20002	97365	30976
06	28357	94070	20652	35774	16249	75019	21145	05217	47286	76305
07	17783	00015	10806	83091	91530	36466	39981	62481	49177	75779
08	40950	84820	29881	85966	62800	70326	84740	62660	77379	90279
09	82995	64157	66164	41180	10089	41757	78258	96488	88629	37231
10	96754	17676	55659	44105	47361	34833	86679	23930	53249	27083
11	34357	88040	53364	71726	45690	66334	60332	22554	90600	71113
12	06318	37403	49927	57715	50423	67372	63116	48888	21505	80182
13	62111	52820	07243	79931	89292	84767	85693	73947	22278	11551
14	47534	09243	67879	00544	23410	12740	02540	54440	32949	13491
15	98614	75993	84460	62846	59844	14922	48730	73443	48167	34770
16	24856	03648	44898	09351	98795	18644	39765	71058	90368	44104
17	96887	12479	80621	66223	86085	78285	02432	53342	42846	94771
18	90801	21472	42815	77408	37390	76766	52615	32141	30268	18106
19	55165	77312	83666	36028	28420	70219	81369	41943	47366	41067

Η επιλογή των στοιχειωδών παρατηρήσεων μπορεί να επιτευχθεί προφανώς με την εφαρμογή οποιουδήποτε κατάλληλου τυχαίου μηχανισμού επιλογής στις μονάδες του πληθυσμού. Η χρήση ενός πίνακα τυχαίων αριθμών αποτελεί μια τέτοια περίπτωση. Σήμερα βέβαια που η πρόσβαση στους υπολογιστές είναι ευκολότερη και η χρήση στατιστικών πακέτων ευρύτατη, μπορεί να χρησιμοποιηθεί η γεννήτρια τυχαίων αριθμών ενός Η/Υ ή ενός στατιστικού πακέτου, για την παραγωγή των τυχαίων αριθμών που θα αποτελέσουν τους δείκτες των μονάδων του πληθυσμού που θα περιληφθούν στο δείγμα.

β) Η τυχαία επιλογή των μονάδων του δείγματος μπορεί να γίνει και με την γνωστή μέθοδο των λαχνών. Δηλ. κάθε μονάδα του πλαισίου γράφεται σε ένα κλήρο (π.χ σε ένα σφαιρίδιο) που είναι απόλυτα ίδιος για όλες τις μονάδες. Οι κλήροι τοποθετούνται σε μια κληροδόχο (π.χ σε ένα κουτί) και ανακινούνται ώστε ν΄ αναμιχθούν προσεκτικά. Από την κληροδόχο παίρνουμε - με κλειστά μάτια - ένα κλήρο και η μονάδα που γράφεται σε αυτόν αποτελεί μονάδα του δείγματος. Ο κλήρος αυτός παραμένει έξω από την κληροδόχο και συνεχίζουμε όπως και προηγουμένως παίρνοντας 2° κλήρο από την κληροδόχο. Έτσι επιλέγεται η 2η μονάδα δείγματος. Ο 2ος κλήρος, όπως και ο 1ος, παραμένει έξω από την κληροδόχο, και η επιλογή συνεχίζεται, με τον ίδιο τρόπο. Σταματάει Δε, όταν το δείγμα αποτελεστεί από τον αριθμό των μονάδων που έχουμε προκαθορίσει. Ο συγκεκριμένος τρόπος επιλογής δείγματος, χωρίς επανατοποθέτηση εξασφαλίζει, ίση πιθανότητα σε κάθε μονάδα του πλαισίου για να περιληφθεί στο δείγμα. Είναι δε σαφές ότι το δείγμα που σχηματίζεται μέσω αυτής της μεθόδου, δεν περιέχει καμιά μονάδα του πληθυσμού περισσότερες από μία φορές.

Για καλύτερη κατανόηση ας υποθέσουμε ότι έχουμε ένα σαφώς καθορισμένο πληθυσμό μεγέθους 5, έστω {1, 2, 3, 4, 5 }

Τα δυνατά διακεκριμένα δείγματα μεγέθους 2 που μπορούμε να σχηματίσουμε είναι τα εξής 10 :

{ 1, 2 } { 1, 3 } {1, 4 } {1, 5 } {2, 3 } { 2, 4 } { 2, 5 } { 3, 4 } {3, 5 }

{4, 5 }.

Γενικότερα, αν ο πληθυσμός αποτελείται από N μονάδες και επιθυμούμε δείγμα μεγέθους n, το πλήθος των δυνατών διακεκριμένων δειγμάτων είναι :

$$(\frac{N}{n}) = \frac{N!}{n!(N-n)!} = \frac{N(N-1)...(N-n+1)}{n} \quad \text{όπου } n! = 1x2x3x......xn$$

Κάθε ένα από τα δείγματα έχει πιθανότητα ίση με $1/(\frac{N}{n})$ να επιλεγεί.

Στην πράξη δεν είναι εύκολος πάντα ο σχηματισμός όλων των δυνατών διακεκριμένων δειγμάτων. Στο παραπάνω παράδειγμα ήταν δυνατό να χρησιμοποιηθεί η μέθοδος των λαχνών.

Αν όμως το μέγεθος του πληθυσμού N είναι μεγάλο, χρησιμοποιείται συνήθως η μέθοδος των πινάκων τυχαίων αριθμών.

3.2. Εκτιμήσεις στην απλή τυχαία δειγματοληψία

3.2.1 Εκτίμηση του μέσου όρου του πληθυσμού

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

$x_1, x_2, x_3, \ldots, x_n$

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:

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

3.2.2 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 :

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

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.

Παράδειγμα:

Από απλό τυχαίο δείγμα 1000 σπουδαστών της χώρας διαπιστώθηκε ότι 150 είναι εργαζόμενοι. Το ποσοστό εργαζομένων στο δείγμα είναι P = 150/1000 = 0,15. Το τυπικό σφάλμα της P υπολογίζεται ως εξής:

$$Sp^2 = \frac{(0,15)\,(1-0,15)}{1000-1} = \frac{0,1275}{999} = 0,0001276$$

και συνεπώς Sp = 0,0113

3.2.3 Εκτίμηση του τυπικού σφάλματος

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

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

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

α) Δίνει την 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 την δυνατότητα εκτίμησης του μεγέθους του δείγματος που απαιτείται σε μια 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 μια εκτίμηση του σ^2 από τα δεδομένα του δείγματος και ως τέτοια συνήθως θεωρείται η τιμή της εκτιμήτριας.

$$\hat{\sigma}^2 = S^{*2} = \frac{\sum_{i=1}^{n}(Xi-\bar{X}n)^2}{n-1}$$

στην περίπτωση απλού τυχαίου δείγματος.

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

$$S^2 = \sum_{i=1}^{n}\frac{(Xi-\bar{X}n)^2}{n-1}$$

είναι αμερόληπτη εκτιμήτρια της διασποράς σ^2 του πληθυσμού.

Οι στατιστικές συναρτήσεις:

$$\hat{\sigma}^2_{\bar{X}n}\ S^2_{\bar{X}n} = \frac{S^2}{n}\cdot\frac{N-n}{N} = \frac{S^2}{n}\ (1\text{-}f)$$

και

$$\hat{\sigma}^2_{\hat{Y}} = S^2_{\hat{y}} = \frac{N^2S^2}{n}\ \frac{N-n}{N} = \frac{N^2S^2}{n}(1\text{-}f)$$

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)

<u>**Παράδειγμα**</u> :

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 ανά 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.

<u>Λύση</u>

Έχουμε $n=\Sigma vi = 50$

$\Sigma vi\ Xi = 1471$

$\Sigma vi\ Xi^2 = 54497$

Άρα $\overline{X}n = 1471/50 = 29.52$ και επομένως η τιμή της εκτιμήτριας $\hat{Y}$ του συνολικού αριθμού υπογραφών y είναι :

$$\hat{Y} = N \cdot \overline{X}n = 19888$$

Το τυπικό σφάλμα της Y είναι ίσο με

$$S\hat{y} = \frac{N.S}{\sqrt{n}} \sqrt{1-f}$$

Όπου

$$S = \sqrt{\frac{1}{n-1}\Sigma vi\left(Xi - \overline{X}n\right)^2} = \sqrt{\frac{1}{n-1}\left[\Sigma viXi^2 - \frac{(\Sigma viXi)^2}{n}\right]} = \sqrt{229} = 15.13$$

Και $\varphi = \frac{50}{676} = 0.0740$, δηλαδή

$$S\hat{y} = \frac{(676)(15.13)}{\sqrt{50}} \sqrt{1-0.0740} = 1391$$

(Το τυπικό σφάλμα του μέσου του δείγματος είναι

$$S_{\bar{x}_n} = \frac{S}{\sqrt{n}} \sqrt{1-f} = 2.05).$$

3.2.4. Εκτίμηση ενός Λόγου

Έστω πληθυσμός μεγέθους Ν και έστω Xi και Yi οι τιμές δύο διαφορετικών χαρακτηριστικών της i μονάδας του πλυθησμού, i=1,2,....,N.

Τότε, $X = \sum_{i=1}^{N} x_i$ και $y = \sum_{i=1}^{N} y_i$ αντιπροσωπεύουν τις συνολικές τιμές των δυο αυτών χαρακτηριστικών και πολλές φορές στην πράξη μας ενδιαφέρει να εκτιμήσουμε τον λόγο

$$R = \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i} = \frac{\sum_{i=1}^{N} y_i / N}{\sum_{i=1}^{N} x_i / N} = \frac{\mu_y}{\mu_x}$$

Δηλαδή μας ενδιαφέρει να εκτιμήσουμε τον λόγο των συνολικών μεγεθών x και y ή τον λόγο των μέσων τιμών μ_x και μ_y των δύο χαρακτηριστικών.

Η εκτιμήτρια που χρησιμοποιείται για τον σκοπό αυτό είναι η

$$\hat{R} = \frac{\hat{\bar{y}}_n}{\bar{x}_n} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}$$

όπου $\{x_1, x_2, x_n\}$, $\{y_1, y_2, y_n\}$ είναι οι τιμές των μονάδων όπου ενός απλού τυχαίου δείγματος μεγέθους **n** για τα δυο υπό εξέταση χαρακτηριστικά και $\bar{x}_n, \bar{y}_n$ οι μέσοι των τιμών αυτών αντίστοιχα.

Για την αξιολόγηση της ακρίβειας της εκτιμήτριας αυτής και την εξαγωγή συμπερασμάτων, είναι αναγκαίος ο προσδιορισμός της κατανομής της και του τυπικού της σφάλματος. Για μικρές τιμές του n η κατανομή της $\hat{R}$ δεν είναι κανονική. Αντίθετα είναι ασύμμετρη προς τα δεξιά. Επιπλέον, η $\hat{R}$ δεν είναι αμερόληπτη εκτιμήτρια του R. Για μεγάλες όμως τιμές του n η κατανομή της $\hat{R}$ τείνει στην κανονική κατανομή.

Αν $\{x_1, x_2, x_n\}$, $\{y_1, y_2, \ldots, y_n\}$ $\bar{x}_n, \bar{y}_n$ $\{x_1, x_2, \ldots, x_N\}$ $\{y_1, y_2, \ldots y_N\}$ και μ_x μ_y ορίζεται όπως προηγουμένως, τότε για αρκετά μεγάλο n ($n \to N$) η στατιστική συνάρτηση

$$\hat{R} = \frac{\hat{\bar{y}}_n}{\bar{x}}$$

είναι αμερόληπτη εκτιμήτρια του λόγου $R = \mu_y / \mu_x$ και ισχύει ότι

$$V(\hat{R}) = \frac{1-F}{n\mu_x^2} = \frac{\sum_{i=1}^{N} (y_i - Rx_i)^2}{N-1}$$

Παράδειγμα για καλύτερη κατανόηση:

Από απλό τυχαίο δείγμα 33 οικογενειών που επιλέγει από κάποιο πληθυσμό, συνελέγησαν στοιχεία για το μέγεθος ($x^{(1)}$) των οικογενειών, το ετήσιο εισόδημά του ($x^{(2)}$) και την ετήσια δαπάνη τους για φαγητό (Y). Τα αποτελέσματα συνοψίζονται στον παρακάτω πίνακα:

Μέγεθος, ετήσιο εισόδημα και ετήσια δαπάνη για φαγητό 33 οικογενειών (σε εκατοντάδες χιλιάδες δρχ.)

Οικογένεια	$X^{(1)}$	$X^{(2)}$	Y	Οικογένεια	$X^{(1)}$	$X^{(2)}$	Y
1	2	62	14.3	18	4	83	36.0
2	3	62	20.8	19	2	85	20.6
3	3	87	22.7	20	4	73	27.7
4	5	65	30.5	21	2	66	25.9
5	4	58	41.2	22	5	58	23.3
6	7	92	28.2	23	3	77	39.8
7	2	88	24.2	24	4	69	16.8
8	4	79	30.0	25	7	65	37.8
9	2	83	24.2	26	3	77	34.8
10	5	62	44.4	27	3	69	28.7
11	3	63	13.4	28	6	95	63.0
12	6	62	19.8	29	2	77	19.5
13	4	60	29.4	30	2	69	21.6
14	4	75	27.1	31	6	69	18.2
15	2	90	22.2	32	4	67	20.1
16	5	75	37.7	33	2	63	20.7
17	3	69	22.6	ΣΥΝΟΛΟ	123	2394	907.2

Να εκτιμηθεί :

i) η μέση ετήσια δαπάνη για φαγητό ανά οικογένεια

ii) η μέση ετήσια δαπάνη για φαγητό ανά άτομο και

iii) το ποσοστό του εισοδήματος που δαπανάται για φαγητό.

Να υπολογισθεί το τυπικό σφάλμα των εκτιμήσεων αυτών.

Λύση

i) Αγνοώντας τη διόρθωση του πεπερασμένου πληθυσμού έχουμε:

$$\bar{Y}_{33}=\frac{907{,}2}{33}=27{,}49$$

με τυπικό σφάλμα :

$$S_{\bar{Y}_{33}}=\frac{1}{\sqrt{33}}\sqrt{\frac{\sum_{i=1}^{33}Y_i^2-33\bar{Y}_{33}^2}{32}}=\sqrt{\frac{28224-(33)(27{,}49)^2}{(32)(33)}}=1{,}76$$

i) $\hat{R}_1=\frac{\Sigma Y_i}{\Sigma x_i^1}=\frac{907{,}2}{123}=7{,}38$

Με τυπικό σφάλμα

$$S_{\hat{R}_1}=\sqrt{\frac{28224-(14{,}7512)(3595{,}5)+(54{,}3996)(533)}{(33)(32)(3{,}7273)^2}}=0{,}534$$

ii) $\hat{R}_2=100\frac{\Sigma Y_i}{\Sigma X_i^2}=\frac{100(907{,}2)}{2394}\%=37{,}9\%$

Με τυπικό σφάλμα

$$S_{\hat{R}_2}=0{,}0238$$

Παρατήρηση:

Προβλήματα τα οποία απαιτούν την εκτίμηση του λόγου δύο μεταβλητών συναντώνται πολύ συχνά στην πράξη. Η συνηθέστερη περίπτωση είναι όταν η δειγματοληπτική μονάδα αποτελείται από ένα σύνολο στοιχειωδέστερων μονάδων και το ενδιαφέρον του ερευνητή εστιάζεται στην εκτίμηση της μέσης τιμής του πληθυσμού ανά στοιχειώδη μονάδα. Τυπικό παράδειγμα αποτελεί η περίπτωση (ii) του παραπάνω παραδείγματος. Λόγοι δύο μεταβλητών εμφανίζονται επίσης σε προβλήματα στα οποία ενδιαφέρει η μελέτη ενός ποσοτικού χαρακτηριστικού (περίπτωση (iii) παραπάνω). Τέλος, προβλήματα σύγκρισης δύο χαρακτηριστικών ενός πληθυσμού ανάγονται επίσης σε προβλήματα εκτίμησης του λόγου δύο κατάλληλων μεταβλητών.

3.2.5 Εκτίμηση Μέσων Τιμών και Συνολικών Μεγεθών Υποπληθυσμών

Ας υποθέσουμε ότι ένα απλό τυχαίο δείγμα μεγέθους n επιλέγεται από κάποιο πεπερασμένο πληθυσμό μεγέθους N. Έστω No ο αριθμός των μονάδων του πληθυσμού που ανήκει σε κάποιο υποσύνολο του πληθυσμού και no ο αριθμός των μονάδων του δείγματος που ανήκουν σ΄ αυτό το υποσύνολο. Τότε αποδεικνύεται η εξής πρόταση :

- Το υποσύνολο no μονάδων του αρχικού τυχαίου δείγματος αποτελεί ένα απλό τυχαίο δείγμα από τον υποπληθυσμό των No μονάδων και ισχύει ότι :

$$\mathrm{E}\left(\frac{no}{No}\right)=\frac{n}{N}$$

Ουσιαστικά λοιπόν βλέπουμε ότι κάθε απλό τυχαίο δείγμα από ένα υποπληθυσμό, περιέχει ένα απλό τυχαίο δείγμα από οποιονδήποτε υποπληθυσμό.

Η σημασία αυτού του συμπεράσματος όσον αφορά τις εφαρμογές είναι μεγάλη. Παρέχει τη δυνατότητα στον ερευνητή να χρησιμοποιήσει πληροφορίες από το δείγμα για την συναγωγή συμπερασμάτων όχι μόνο για ολόκληρο τον πληθυσμό αλλά και για οποιονδήποτε υποπληθυσμό.

Για παράδειγμα, με βάση ένα απλό τυχαίο δείγμα από κάποιο πληθυσμό εργαζομένων ατόμων είναι δυνατή η μελέτη της κατανομής του εισοδήματος σε ολόκληρο τον πληθυσμό αλλά και παράλληλα σε υποπληθυσμούς, όπως οι άνδρες ηλικίας 30-40 ετών ή οι γυναίκες με δύο παιδιά κλπ.

Οι υποπληθυσμοί στους οποίους μπορεί να διακριθεί ένας πληθυσμός ονομάζονται συνήθως περιοχές μελέτης.

Έστω ο πληθυσμός {y1, y2,.....,yN} και έστω ότι κάθε μονάδα του πληθυσμού αυτού ανήκει σε έναν από k υποπληθυσμούς. Τότε N=N1+N2+......+Nk, όπου N αντιπροσωπεύει το μέγεθός του j υποπληθυσμού. Κατά συνέπεια, x1,x2,.....,xn είναι ένα απολό τυχαίο δείγμα μεγέθους n, ισχύει ότι :

n = n1+n2+.....+nk.

Όπου n_j είναι το μέγεθος του τμήματος του δείγματος του οποίου οι μονάδες ανήκουν στην j περιοχή.

Έστω ότι $y_i^{(j)}$ αντιπροσωπεύει την i μονάδα της j περιοχής του πληθυσμού και $X_i^{(j)}$ την i μονάδα του τυχαίου δείγματος που αντιστοιχεί στην j περιοχή.

Τότε η μέση τιμή και η διασπορά του j υποπληθυσμού δίνονται από τους τύπους

$$\mu_j = \frac{1}{Nj} \sum_{i=1}^{Nj} y_i^{(j)}$$

$$\sigma_j^2 = \frac{1}{Nj-1} \sum_{i=1}^{Nj} (y_i^{(j)} - \mu_j)^2$$

αντίστοιχα, και οι αμερόληπτες εκτιμήτριες τους είναι οι στατιστικές συναρτήσεις

$$\bar{X} n_j = \frac{1}{n_j} = \sum_{i=1}^{nj} X_i^{(j)} \quad \text{και}$$

$$S_j^2 = \frac{1}{nj-1} \sum_{i=1}^{n_j} (X_i^{(j)} - \bar{X} n_j)^2$$

αντίστοιχα. Προφανώς η εκτιμήτρια του τυπικού σφάλματος της εκτιμήτριας $\overline{X} n_j$ είναι ίση με :

$$S\bar{X} nj = \frac{Sj}{\sqrt{nj}} \sqrt{1 - \frac{n_j}{Nj}}$$

Παρατήρηση :

Αν η τιμή του μεγέθους του j υποπληθυσμού, Nj, δεν είναι γνωστή, μπορεί να χρησιμοποιηθεί στη θέση του λόγου nj/Nj ο λόγος n/N για τον υπολογισμό της διόρθωσης πεπερασμένου πληθυσμού.

Αυτό είναι απόρροια του γεγονότος ότι στην απλή τυχαία δειγματοληψία η τιμή nj/Nj είναι αμερόληπτη εκτίμησης της τιμής n/N.

Στην περίπτωση αυτή

$$S\bar{X} nj = Sj \sqrt{\frac{1 - n/N}{nj}}$$

Για την εκτίμηση του συνολικού μεγέθους

$$Y^{(j)} = \sum_{i=1}^{Nj} y_i^{(j)}$$

χρησιμοποιείται η στατιστική συνάρτηση

$$\hat{Y}^{(j)} = Nj\ \bar{X}_{n_j}$$

η οποία είναι αμερόληπτη εκτιμήτρια του yi με τυπικό σφάλμα εκτιμώμενο απ΄ τη στατιστική συνάρτηση

$$S_{\hat{Y}}^{(j)} = Nj\ S\bar{X}_{n_j}$$

Η παραπάνω εκτιμήτρια του $y^{(i)}$ προϋποθέτει τη γνώση της τιμής του Nj. Στην πράξη όμως πολύ σπάνια συμβαίνει να είναι γνωστή αυτή η τιμή.

Στις περιπτώσεις αυτές χρησιμοποιείται ως εκτιμήτρια του μεγέθους του $y^{(j)}$ η συνάρτηση.

$$\hat{Y}^{(j)} = \frac{N}{n} \sum_{i=1}^{nj} X_i^{(j)}$$

Η στατιστική συνάρτηση $\hat{Y}^{(j)}$ είναι αμερόληπτη εκτιμήτρια του συνολικού μεγέθους $y^{(j)}$ με τυπικό σφάλμα εκτιμώμενο από την στατιστική συνάρτηση

$$S_{\hat{Y}}(j) = NS' \sqrt{\frac{1-n/N}{n}}$$

όπου

$$S'^2\ \frac{1}{1-n} \left[\sum_{i=1}^{nj} (x_i^{(j)})^2 - \frac{(\sum_{i=1}^{nj} X_i^{(j)})^2}{n} \right]$$

3.2.6. Άλλες εκτιμήσεις του μέσου $\bar{X}n$ ενός απλού τυχαίου δείγματος

- Ο μέσος $\bar{X}n$ ενός απλού τυχαίου δείγματός μεγέθους n από ένα πληθυσμό N είναι αμερόληπτη εκτιμήτριά της μέσης τιμής μ του πληθυσμού.

- Η διασπορά του μέσου $\bar{X}n$ ενός απλού τυχαίου δείγματός μεγέθους n από ένα πληθυσμό μεγέθους N είναι :

$$V(\bar{X}n) = \frac{\sigma^2}{n} \; \frac{N\text{-}n}{n}$$

- Το τυπικό σφάλμα του μέσου $\bar{X}n$ είναι

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

Παρατήρηση :

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

$$V(\bar{X}n) = \frac{\sigma^2}{n}(1\text{-}f)$$

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

Παρατήρηση :

Μια αμερόληπτη εκτιμήτρια της τιμής $y=\sum_{i=1}^{N} yi$ είναι η στατιστική συνάρτηση $\hat{Y} = N\bar{X}n$ με διασπορά

$V(\hat{Y}) = \frac{N^2\sigma^2}{N}(1\text{-}f)$ 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 την διασπορά της εκτιμήτριας Xn του μ.

Για παράδειγμα, αν δύο πληθυσμοί μεγέθους N_1=200.000 και N_2 =20.000 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≤ 0,05 και πολλές φορές ακόμη και αν f = 0,10. Το αποτέλεσμα είναι να υπερεκτιμάται το τυπικό σφάλμα της εκτιμήτριας $\bar{X}n$.

3.3 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 :

$$\bar{X}n \sim N(\mu, \frac{\sigma^2}{n}) \Leftrightarrow$$

$$\frac{\overline{X}n-\mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

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 25 ($n \geq 25$).

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 ότι $P\left|\frac{X-\mu}{\sigma/\sqrt{n}}\right| \langle Z1-\frac{a}{2} = 1-a$

Όπου $Z_{1-\frac{a}{2}}$ 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

$$\bar{X}n \pm Z1\text{-}\frac{a}{2}\,\frac{\sigma}{\sqrt{n}}$$

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 :

$$\bar{X}n \pm t_{\,n-1,1-\frac{a}{2}}\,\frac{S}{\sqrt{n}}$$

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 ένα 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 ο n›25 G^2_1

Όπου η τιμή G_1 είναι γνωστή ως μέτρο ασυμμετρίας του Fisher και ορίζεται από την 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 μια ιδέα σχετικά με το απαιτούμενο δειγματικό μέγεθος.

3.4 Εκτίμηση Ποσοστών

Η εκτίμηση ποσοστών εμπίπτει στην εκτίμηση των ποιοτικών χαράκτηριστικών μιας δειγματοληψίας.

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

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

Α αν έχουν το χαρακτηριστικό που μας ενδιαφέρει και Α΄ αν όχι.

Ο ακριβής αριθμός των μονάδων που ανήκουν στην κατηγορία Α 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 η :

$\hat{p} = \frac{X}{n}$ 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 $\bar{X}n$ ενός απλού τυχαίου δείγματος με την εισαγωγή των εξής :

Έστω $yi = \begin{cases} 1 & \text{αν η } i \text{ μονάδα πληθυσμού} \in A \\ 0 & \text{διαφορετικά} \end{cases}$

i = 1, 2, 3, ..., N

$$Xi = \begin{cases} 1 & \text{αν η } i \text{ μονάδα δείγματος} \in A \\ 0 & \text{διαφορετικά} \end{cases}$$

i =1, 2,..., n

Τότε έχουμε :

$N_A = \sum_{i=1}^{N} y_i$, $X = \sum_{i=1}^{n} X_i$

$p = \frac{\Sigma y_i}{N} = \mu$, $\hat{p} = \frac{\Sigma X_i}{n} = \bar{X}_n$

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.

Προφανώς :
$$\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} y_i = N_A = N \bullet p$$
$$\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i = X = n \bullet \hat{p}$$

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 μια εκτιμήτριά της παρέχεται από την 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:

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

- Η στατιστική συνάρτηση $\hat{Y}_A = N\hat{p}$ είναι μια 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 $S^2_{\hat{p}} = \frac{\hat{p}(1-\hat{p})}{n-1} \cdot \frac{N-n}{n-1}$ είναι μια 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 μια αμερόληπτη εκτιμήτρια της $V(\hat{Y}_A)$.

Για καλύτερη κατανόηση των παραπάνω, παραθέτουμε το εξής <u>παράδειγμα:</u>

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.

<u>Λύση:</u>

N=3042 n=200 x=38

Άρα $\hat{p} = 0.19$

$\hat{Y}_A = N \cdot \hat{p} = 578$

$S^2_{\hat{Y}_A} = (3042)(2842)(0{,}19)(0{,}81)/199 = 6686$

Άρα $S_{\hat{y}_A} = \sqrt{6686}$ =81,8.

3.5 Πλεονεκτήματα – Μειονεκτήματα Απλής Τυχαίας Δειγματοληψίας

3.5.1 Πλεονεκτήματα

1) Επειδή κάθε μονάδα του πληθυσμού έχει μια γνωστή πιθανότητα να επιλεγεί, το δυνατό δειγματοληπτικό σφάλμα μπορεί να εκτιμηθεί μαθηματικά. Μπορεί δηλαδή να υπολογισθεί ένα αντικειμενικό μέτρο αξιοπιστίας. Αυτό, όχι μόνο επιτρέπει την εκτίμηση των πιθανοτήτων να γίνει ένα σφάλμα στην κρίση, αφού έχει ολοκληρωθεί η μελέτη, αλλά επίσης παρέχει μια μέθοδο για εκτίμηση του κατάλληλου μεγέθους του δείγματος πριν ακόμα αρχίσει η μελέτη.

2) Τα αποτελέσματα της απλής τυχαίας δειγματοληψίας είναι αντικειμενικά, δεδομένου ότι απομακρύνεται η υποκειμενική φύση της επιλογής των αποκρινόμενων και κατά συνέπεια απομακρύνεται η πιθανότητα μεροληψίας.

3) Είναι μια τεχνική εύκολη στην κατανόηση δεδομένου ότι έχει απλό εννοιολογικό σχεδιασμό καθιστώντας την προσιτή ακόμα και σε εκείνους που δεν έχουν ισχυρό στατιστικό υπόβαθρο.

Διαθέτει επίσης απλές τεχνικές ανάλυσης δεδομένων.

3.5.2 Μειονεκτήματα

1) Η μέθοδος αυτή απαιτεί πλαίσια πλήρως ενημερωμένα, είναι δηλαδή αναγκαίο να κατασκευαστεί κατάλογος και να αριθμηθεί κάθε μονάδα του συνόλου του πληθυσμού. Αλλά, ακόμα και αν κατασκευαστεί κατάλογος, υπάρχει η πιθανότητα να μη συμπίπτει

ακριβώς με τον ορισμό του συνόλου του πληθυσμού (όταν π.χ χρησιμοποιείται ένας τηλεφωνικός κατάλογος για την μελέτη μιας έρευνας).

2) Σε πολλές περιπτώσεις και ειδικά όταν το σύνολο του υπό μελέτη πληθυσμού είναι πεπερασμένο, η μέθοδος αυτή είναι ιδιαίτερα χρονοβόρα και έχει μεγάλο κόστος στην συλλογή των δεδομένων.

Κεφάλαιο 4ο

Απλή τυχαία δειγματοληψία με επανάθεση

ΚΕΦΑΛΑΙΟ 4ο

ΑΠΛΗ ΤΥΧΑΙΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΜΕ ΕΠΑΝΑΘΕΣΗ

Αντί της δειγματοληψίας χωρίς επανάθεση, είναι δυνατόν να χρησιμοποιηθεί η δειγματοληψία με επανάθεση (sampling with replacement) .

Στην περίπτωση αυτή, κάθε κλήρος που επιλέγεται από την κληροδόχο, επανατοποθετείται στην κληροδόχο. Αν ο κλήρος αυτός συμβεί να επιλεγεί πάλι σε κάποια επόμενη λήψη, το αποτέλεσμά του δεν αγνοείται, αλλά η αντίστοιχη μονάδα του πληθυσμού παραμένει στο δείγμα το οποίο έτσι την περιέχει δύο ή περισσότερες φορές. Με τον τρόπο αυτό είναι δυνατόν το δείγμα n μονάδων να αποτελεσθεί από την ίδια μονάδα του πληθυσμού, η οποία περιέχεται σ΄ αυτό n φορές.

Αν για παράδειγμα N=5 και n=2, ένα τέτοιο δειγματοληπτικό σχήμα επιτρέπει τον σχηματισμό δειγμάτων μορφής {1,1} {2,2} {3,3} {4,4} {5,5}

Αυτός ο τρόπος δειγματοληψίας είναι γνωστός ως δειγματοληψία χωρίς περιορισμό (unrestricted sampling), σε αντίθεση με την απλή τυχαία δειγματοληψία η οποία διεξάγεται με τον περιορισμό ότι καμιά μονάδα του πληθυσμού δεν μπορεί να βρεθεί στο δείγμα περισσότερες από μία φορές. (restricted sampling).

Στην μέθοδο της απλής τυχαίας δειγματοληψίας η λήψη του δείγματος μπορεί να γίνει με τις ίδιες μεθόδους που χρησιμοποιούνται και στην χωρίς επανατοποθέτηση :

α) Τη μέθοδο των τυχαίων αριθμών

β) Τη μέθοδο των λαχνών

με την διαφορά ότι και στις δυο περιπτώσεις έχουμε επανατοποθέτηση της μονάδας του πληθυσμού που επιλέγεται κάθε φορά.

Ο υπολογισμός της διασποράς και των εκτιμητριών της διασποράς είναι ευχερέστερος όταν έχουμε επανάθεση.

Για τον λόγο αυτό, η δειγματοληψία με επανάθεση χρησιμοποιείται μερικές φορές στα πιο σύνθετα δειγματοληπτικά σχήματα, αν και εκ πρώτης όψεως δεν φαίνεται και τόσο πρακτικό να επιτρέψουμε τη δυνατότητα επιλογής της ίδιας μονάδας του πληθυσμού δύο ή περισσότερες φορές στο δείγμα.

Εντούτοις η διασπορά του δειγματικού μέσου στη μέθοδο αυτή είναι μεγαλύτερη από εκείνη της δειγματοληψίας χωρίς επανάθεση.

Γι΄ αυτό και στην πράξη η απλή τυχαία δειγματοληψία χωρίς επανάθεση είναι προτιμότερη, διότι δίνει εκτιμήσεις με δειγματοληπτικό σφάλμα μικρότερο από το αντίστοιχο της δειγματοληψίας με επανάθεση.

Κεφάλαιο 5ο

Δειγματοληψία κατά στρώματα ή στρωματοποιημένη

ΚΕΦΑΛΑΙΟ 5ο

ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΚΑΤΑ ΣΤΡΩΜΑΤΑ Η ΣΤΡΩΜΑΤΟΠΟΙΗΜΕΝΗ

5.1 Γενικά

Αν ο πληθυσμός που εξετάζεται δεν είναι αρκετά ομοιογενής, η απλή τυχαία δειγματοληψία ενδέχεται να δώσει δείγμα ακραίο, δηλαδή μη αντιπροσωπευτικό. Παραδείγματος χάρη σε απλή τυχαία δειγματοληψία οικογενειών από το λεκανοπέδιο της Αττικής, ίσως προκύψει δείγμα από οικογένειες που όλες τους ζουν στο Παλαιό Ψυχικό. Έτσι, αν ο σκοπός της έρευνας είναι να διαπιστωθεί η κατάσταση απασχόλησης των μελών της οικογένειας, το δείγμα με οικογένειες από το Παλαιό Ψυχικό είναι ακραίο και συνεπώς απρόσφορο για την έρευνα, αφού οι οικογένειες αυτές δεν θίγονται από την ανεργία τόσο πολύ όσο άλλες οικογένειες που ζουν σε περιοχές όπως είναι το Περιστέρι, η Νέα Ιωνία κτλ (εργατικές οικογένειες).

Ωστόσο πρέπει να πούμε ότι μολονότι δεν αποκλείεται δείγμα απλό τυχαίο να εμφανιστεί με ακραίες παρατηρήσεις, η απλή τυχαία δειγματοληψία είναι αμερόληπτο σχέδιο επιλογής των μονάδων του δείγματος. Η παρουσίαση και ακραίων δειγμάτων στην απλή τυχαία δειγματοληψία από ανομοιογενή πληθυσμό σημαίνει ότι οι εκτιμήσεις του δειγματοληπτικού αυτού σχεδίου έχουν μειωμένη ακρίβεια. Αύξηση της ακρίβειας του σχεδίου αυτού μπορούμε να επιτύχουμε με αύξηση του μεγέθους του δείγματος, διότι η αύξηση αυτή μειώνει την τιμή του σφάλματος που οφείλεται στη δειγματοληψία. Αύξηση, όμως, του μεγέθους του δείγματος συνεπάγεται αύξηση και των δαπανών της δειγματοληψίας. Στο σημείο αυτό γίνεται το ερώτημα αν είναι δυνατό να επιτύχουμε μείωση του δειγματοληπτικού σφάλματος (και συνεπώς αύξηση της ακρίβειας των εκτιμήσεων) χωρίς να καταφύγουμε σε αύξηση του μεγέθους του δείγματος και των δαπανών.

Η απάντηση στο ερώτημα αυτό είναι ότι μπορούμε να αυξήσουμε την ακρίβεια των εκτιμήσεων διατηρώντας το μέγεθος του δείγματος σταθερό αρκεί να βελτιώσουμε την αντιπροσωπευτικότητα του δείγματος εφαρμόζοντας για τον σκοπό αυτό δειγματοληψία κατά στρώματα.

Στη δειγματοληψία κατά στρώματα χωρίζουμε το συνολικό πληθυσμό που εξετάζουμε σε ορισμένες κατηγορίες τις οποίες ονομάζουμε στρώματα. Σε κάθε κατηγορία συγκεντρώνουμε όσον το δυνατόν πιο ομοιογενή στοιχεία του πληθυσμού φροντίζοντας ώστε να διαμορφώνεται η πιο μεγάλη δυνατή διαφοροποίηση μεταξύ των κατηγοριών αυτών. Από κάθε στρώμα σχηματίζουμε απλό τυχαίο δείγμα και στη συνέχεια συνενώνουμε τα επιμέρους αυτά δείγματα σε ένα ενιαίο για ολόκληρο τον πληθυσμό. Από την φύση του το ενιαίο αυτό δείγμα είναι περισσότερο αντοπροσωπευτικό σε σύγκριση με εκείνο της απλής τυχαίας δειγματοληψίας, διότι τώρα είναι βέβαιο ότι στο ενιαίο δείγμα θα περιέχονται στοιχεία από όλες τις κατηγορίες (στρώματα) του πληθυσμού ενώ η εξασφάλιση αυτή είναι αδύνατη στην απλή τυχαία δειγματοληψία από ολόκληρο τον πληθυσμό.

Η παραπάνω διαδικασία της στρωματοποιημένης δειγματοληψίας θα μπορούσε να συνοψισθεί στους παρακάτω δύο ορισμούς :

Ορισμός 1: Έστω ότι έχουμε πληθυσμό μεγέθους N και ότι αυτός μπορεί να διαιρεθεί σε k εσωτερικά ομοιογενείς υποπληθυσμούς μεγέθους N1, N2,.....,Nk. Αν αυτοί είναι ξένοι μεταξύ τους ώστε να ισχύει ότι N1+N2+......+Nk=N, οι υποπληθυσμοί αυτοί ονομάζονται στρώματα (strata).

Ορισμός 2: Έστω ότι από κάθε ένα από τα στρώματα ενός πληθυσμού, επιλέγεται ένα απλό τυχαίο δείγμα μεγέθους ni,i=1,2,....,k ανεξάρτητα από τα άλλα. Το δείγμα μεγέθους n=n1+n2+....nk που προκύπτει από την ένωση των k ανεξάρτητων απλών τυχαίων δειγμάτων ονομάζεται στρωματοποιημένο τυχαίο δείγμα (stratified random sample) και η διαδικασία επιλογής του ονομάζεται στρωματοποιημένη τυχαία δειγματοληψία (stratified random sampling).

<u>**Παράδειγμα**</u>:

Ας υποθέσουμε ότι επιδιώκουμε να εκτιμήσουμε το χρόνο τον οποίο διαθέτει ο μέσος σπουδαστής της σχολής μας για την μελέτη των μαθημάτων του. Επειδή έχουμε παρατηρήσει ότι ο χρόνος αυτός διαφέρει σχετικά λίγο, μεταξύ των σπουδαστών του ίδιου έτους σπουδών (ενώ διαφέρει σαφώς περισσότερο από έτος σε έτος σπουδών), χωρίζουμε τους σπουδαστές της σχολής σε όμοια στρώματα, ανάλογα τα έτη σπουδών. Έτσι έχουμε :

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

Σχηματίζουμε δείγμα από κάθε στρώμα (έτος σπουδών) με κλάσμα δειγματοληψίας 10% (το ίδιο ποσοστό για όλα τα στρώματα).

Το συνολικό δείγμα θα αποτελείται από 400 σπουδαστές και θα απαρτίζεται από τα επιμέρους δείγματα των τεσσάρων ετών (στρωμάτων) ως εξής:

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

Είναι σαφές ότι το δείγμα των 400 σπουδαστών αντιπροσωπεύει τους σπουδαστές όλων των ετών σπουδών. Αντίστοιχο δείγμα απλό τυχαίο από 4000 σπουδαστές θα ήταν δυνατό να αποτελεσθεί μόνο από σπουδαστές του Α ή του Β κ.ο.κ.

Με την στρωματοποίηση του πληθυσμού κατορθώνουμε δηλαδή στην ουσία να διασπάσουμε τη συνολική διακύμανση της μεταβλητής που εξετάζουμε στα εξής δύο τμήματα :

α) Διακύμανση μέσα στα στρώματα (within)

β) Διακύμανση μεταξύ των στρωμάτων (between)

Επειδή όμως με τη δειγματοληψία κατά στρώματα στο συνολικό δείγμα αντιπροσωπεύονται όλα τα στρώματα, η διακύμανση της κατανομής δειγματοληψίας δεν επηρεάζεται διόλου από την διακύμανση μεταξύ των στρωμάτων. Έτσι η διακύμανση της κατανομής δειγματοληψίας προσδιορίζεται από τη διακύμανση <u>εντός</u> των στρωμάτων. Συνεπώς, είναι φανερό ότι η στρωματοποίηση πρέπει να διεξάγεται με τέτοιο τρόπο ώστε οι μονάδες του πληθυσμού που εντάσσονται σε κάθε στρώμα να διαφέρουν μεταξύ τους όσον το δυνατόν λιγότερο, μέσα σε κάθε στρώμα.

Στην ακραία περίπτωση που επιδιώκουμε να διαμορφώσουμε απόλυτη ομοιογένεια μέσα σε κάθε στρώμα (ώστε η διακύμανση εντός των στρωμάτων να είναι μηδενική) ίσως χρειαστεί να σχηματίσουμε τόσα στρώματα όσες είναι και οι μονάδες του πληθυσμού.

Η περίπτωση δεν αποτελεί δειγματοληψία, αλλά απογραφική εξέταση του πληθυσμού.

5.2 Εκτιμήσεις στην στρωματοποιημένη δειγματοληψία

5.2.1 Εκτίμηση μέσης Τιμής στρωματοποιημένου πληθυσμού.

Τα δεδομένα της στρωματοποιημένης δειγματοληψίας μπορούμε να τα συνοψίσουμε στον παρακάτω πίνακα:

Πληθυσμός				Δείγμα		
Στρώμα	Μέγεθος Στρώματος	Μέση Τιμή	Διασπορά	Μέγεθος απλού Τυχαίου δειγμ.	Μέσος	Διασπορά
1	N_1	μ_1	σ_1^2	n_1	$\bar{X}_{n_1}$	$\frac{\sigma_1^2}{n_1}\left(1-\frac{n_1}{N_1}\right)$
2	N_2	μ_2	σ_2^2	n_2	$\bar{X}_{n_2}$	$\frac{\sigma_2^2}{n_2}\left(1-\frac{n_2}{N_2}\right)$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	N_i	μ_i	σ_i^2	n_i	$\bar{X}_{n_i}$	$\frac{\sigma_i^2}{n_i}\left(1-\frac{n_i}{N_i}\right)$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	N_k	μ_k	σ_k^2	n_k	$\bar{X}_{n_k}$	$\frac{\sigma_k^2}{n_k}\left(1-\frac{nk}{Nk}\right)$

Βάσει των δεδομένων ισχύει : $\bar{X}_{n_i} = \sum_{j=1}^{n_i} X_j^{(i)}$

Όπου i= 1, 2,.....,k

$X_j^{(i)}$ = n j μονάδα του ι δείγματος.

Είναι προφανές ότι η μέση τιμή μ του πληθυσμού και οι μέσες τιμές $\mu_1, \mu_2, \ldots \mu_k$ των υποπληθυσμών συνδέονται με την σχέση :

$$\mu = \frac{1}{N} \sum_{i=1}^{k} N_i \cdot \mu_i$$

Τα αντίστοιχα δειγματικά μεγέθη συνδέονται επίσης με τη σχέση:

$$\overline{X}n = \frac{1}{n}\sum_{i=1}^{k} n_i \overline{X}n_i$$

όπου $\bar{X}_n$ είναι ο μέσος του συνολικού δείγματος.

Στην πράξη όμως η στατιστική συνάρτηση που χρησιμοποιείται ως εκτιμήτρια του μ δεν είναι η $\bar{X}_n$ αλλά η :

$$\hat{\mu}_n = \frac{1}{N} \sum_{i=1}^{k} N_i \bar{X}_{n_i}$$

στην οποία οι μέσοι των δειγμάτων από τα διάφορα στρώματα σταθμίζονται με τους συντελεστές βαρύτητας

Ni/N, όπου i = 1, 2, ..., k των στρωμάτων.

Επίσης ισχύει : $\hat{\mu}_n = \bar{X}_n$ αν $\frac{ni}{n} = \frac{Ni}{N}$ όπου i = 1, 2, ..., k

$$\Leftrightarrow \frac{ni}{Ni} = \frac{n}{N}$$

$\Leftrightarrow$ fi = f , όπου i = 1, 2, ..., k

Δηλαδή οι δύο στατιστικές συναρτήσεις συμπίπτουν, αν όλα τα δείγματα εκπροσωπούν το ίδιο ποσοστό μονάδων των αντίστοιχων στρωμάτων.

- Η στατιστική συνάρτηση $\hat{\mu}_n$ είναι αμερόληπτη εκτιμήτρια της μέσης τιμής μ του πληθυσμού και ισχύει ότι :

$$V(\hat{\mu}_n) = \sum_{i=1}^{k} \left(\frac{Ni}{N}\right)^2 \frac{\sigma_i^2}{n_i} \left(1 - \frac{n_i}{N_i}\right)$$

- Η στατιστική συνάρτηση $\hat{Y}st = N \cdot \hat{\mu}_n$ είναι αμερόληπτη εκτιμήτρια

του συνολικού μεγέθους y = $\sum_{i=1}^{k}\sum_{j=1}^{k} y_j^{(i)}$ και ισχύει ότι :

$$V(\hat{Y}st) = \sum_{i=1}^{k} N_i^2 \frac{\sigma_i^2}{n_i}\left(1-\frac{n_i}{N_i}\right)$$

- Η στατιστική συνάρτηση $S^2_{\hat{\mu}_n} = \sum_{i=1}^{k}\left(\frac{N_i}{N}\right)^2 . \frac{S_i^2}{n_i}\left(1-\frac{n_i}{N_i}\right)$ είναι μια

αμερόληπτη εκτιμήτρια της $V(\hat{\mu}_n)$

- Η στατιστική συνάρτηση N^2 $S^2_{\hat{\mu}_n}$ είναι μια αμερόληπτη εκτιμήτρια

της $V(\hat{Y}st)$

5.2.2 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 στον πληθυσμό (N).

$$\Sigma\left(\frac{N1}{N}P1+\frac{N2}{N}P2+...+\frac{Nk}{N}Pk\right) = \Sigma(P) = N$$

Η διακύμανση του σταθμικού αθροίσματος P_1, P_2, ..., P_k,

$$S_P^2 = \left(\frac{N_1}{N}\right)^2 \frac{P_1(1-P_1)}{n_1-1}(1-f_1)+...+\left(\frac{Nk}{N}\right)^2 \frac{Pk(1-Pk)}{n_k-1}(1-f_k)$$

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 n_i-1) .

Στην περίπτωση αυτή τίθενται οι ανάλογες προϋποθέσεις

5.3. Επιλογή των n_1, n_2, ..., n_k

Το πρόβλημα που συνδέεται άμεσα με την τεχνική της στρωματοποιημένης δειγματοληψίας είναι ο καταμερισμός του συνολικού δειγματικού μεγέθους n στα k στρώματα δηλαδή ο καθορισμός των τιμών των μεγεθών n_1, n_2, ..., n_k των k απλών τυχαίων δειγμάτων.

Υπάρχουν δύο τρόποι επιλογής του δειγματικού μεγέθους n στα k στρώματα:

α) Δειγματοληψία κατά στρώματα ανάλογη

β) Δειγματοληψία κατά στρώματα δυσανάλογη

Ειδικότερα έχουμε :

α. Δειγματοληψία κατά στρώματα ανάλογη

Το κριτήριο καθορισμού των επιμέρους μεγεθών των δειγμάτων είναι η μεγιστοποίηση της ακρίβειας των εκτιμήσεων βέβαια, αλλά συγχρόνως και ο περιορισμός της συνολικής δαπάνης της δειγματοληπτικής έρευνας.

Αν η δαπάνη αυτή οδηγεί στον καθορισμό του συνολικού μεγέθους του δείγματος σε n μονάδες για όλα τα στρώματα, το δειγματοληπτικό κόστος ανά μονάδα είναι το ίδιο σε όλα τα στρώματα και επίσης οι διασπορές των στρωμάτων δεν διαφέρουν σημαντικά, τότε τα μεγέθη n_1, n_2,......, n_k επιλέγονται ανάλογα με το μέγεθος των στρωμάτων.

Δηλαδή ισχύει η σχέση:

$$\frac{n_1}{N_i} = \frac{n_2}{N_2} = \ldots\ldots = \frac{n_k}{N_k} = \frac{n_1 + n_2 + \ldots + n_k}{N_1 + N_2 + \ldots + N_k} = \frac{n}{N}$$

Από τη σχέση αυτή έχουμε : $n_i = n\frac{Ni}{N}$

Το κλάσμα δειγματοληψίας κατά στρώμα είναι ο αριθμός σταθερός: n/N

Ο σχεδιασμός αυτός είναι γνωστός ως αναλογικός καταμερισμός του n (proportional allocation) και η δειγματοληπτική τεχνική ονομάζεται αναλογική στρωματοποιημένη τυχαία δειγματοληψία (proportional stratified random sampling).

Δηλαδή : $n_i = n\ \frac{Ni}{N}$ όπου i = 1, 2,......, k

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

Εισάγοντας τις ανωτέρω τιμές των ni στους τύπους με τους οποίους υπολογίζονται τα τυπικά σφάλματα του μέσου $\bar{X}$ και του ποσοστού P έχουμε :

$$S^2_{\bar{x}} = \Sigma \left(\frac{N_i}{N}\right)^2 S_i^2 \frac{N-n}{Nn} = \frac{\Sigma niS_i^2}{n}\left(\frac{1}{n} - \frac{1}{N}\right)$$

Αν παραλείψουμε τον διορθωτικό παράγοντα έχουμε :

$S^2_{\bar{X}} = \dfrac{\Sigma n i S_i^2}{n^2}$, όπου S_i^2 = διακύμανση της μεταβλ. Χ μέσα σε κάθε στρώμα.

Επίσης για την διακύμανση της εκτίμησης της αναλογίας, έχουμε :

$$S^2_p = \Sigma \left[\frac{Ni}{N}\right]^2 \frac{Pi(1-Pi)}{n_i - 1}\left[1-\frac{Ni}{N}\right] - \frac{\Sigma n_i Pi(1-Pi)}{n^2}$$

<u>Παράδειγμα</u>:

Από τους σπουδαστές κάθε έτους σπουδών παίρνουμε τυχαίο δείγμα 1% και υπολογίζουμε τη μέση βαθμολογία των μαθημάτων τους $(\bar{X}i)$, τη διακύμανση της βαθμολογίας τους (S_i^2) και την αναλογία των σπουδαστών (Pi)

Στρώμα (έτος)	Πληθυσμός σπουδαστών	Μέγεθος δείγματος n_i	Μέσος όρος $\bar{X}i$	Διακύμανση S_i^2	Αναλογία σπουδαστών εργαζομένων P_i
Α	1500	15	6,5	1,2	0,08
Β	1000	10	7,2	1,1	0,15
Γ	800	8	7,6	1,3	0,20
Δ	700	7	7,0	1,0	0,40
Σύνολα	4000	40			

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

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

Η διακύμανση της κατανομής της δειγματοληψίας της $\bar{X}$ είναι:

$$S^2_{\bar{x}} = \frac{15(1,2)+10(1,1)+8(1,3)+7(1,0)}{40^2} = \frac{46,4}{1600} = 0,029$$

Το αντίστοιχο τυπικό σφάλμα είναι $S_{\bar{X}} = 0,17$

Η αμερόληπτη εκτίμηση της αναλογίας των εργαζομένων σπουδαστών στον πληθυσμό υπολογίζεται ως εξής :

P = 0,08(15/40)+0,15(10/40)+0,20(8/40)+0,40(7/40) = 0,1775

Η διακύμανση της εκτίμησης P υπολογίζεται ως εξής :

$$S^2_p = \frac{15(0,08)(0,92)+10(0,15)(0,85)+8(0,20)(0,80)+7(0,40)(0,60)}{40^2} =$$

$$\frac{5,339}{1600} = 0,003337$$

Το τυπικό σφάλμα της εκτίμησης P είναι Sp = 0,05777

β. Δειγματοληψία κατά στρώματα δυσανάλογη

Αν η διακύμανση της μεταβλητής, η οποία μας ενδιαφέρει, είναι για όλα τα στρώματα ίδια - και επιπλέον - το κόστος της δειγματοληψίας δεν διαφοροποιείται μεταξύ των στρωμάτων, εφαρμόζουμε την ανάλογη δειγματοληψία κατά στρώματα. Στην περίπτωση όμως που το κόστος της δειγματοληψίας (κατά μονάδα του δείγματος) είναι το ίδιο για όλα τα στρώματα, αλλά η διακύμανση διαφέρει μεταξύ των στρωμάτων, η δειγματοληψία είναι δυσανάλογη, δηλαδή από κάθε στρώμα παίρνουμε δείγμα όχι ανάλογο προς το μέγεθος του στρώματος, αλλά ανάλογο προς το γινόμενο του μεγέθους του στρώματος (Ni) και της αντίστοιχης μέσης απόκλισης τετραγώνου (Si) :

$$\frac{n_1}{N_1S_1}=\frac{n_2}{N_2S_2}=......=\frac{nk}{NkSk}=\frac{n_1+n_2+...+n_k}{N_1N_1+...N_kS_k}=\frac{n}{\Sigma NiSi}$$

Από την σχέση αυτή προκύπτει :

$$ni = n\,\frac{NiSi}{\Sigma NiSi}$$

Στη θέση των Si τοποθετούμε τα αντίστοιχα πληθυσμιακά μεγέθη σi αν τα τελευταία είναι γνωστά από προηγούμενες ή ειδικές έρευνες.

Εισάγοντας την τιμή της ni στον τύπο υπολογισμού της διακύμανσης της $\bar{X}$, έχουμε :

$$S\,\bar{X}^2 = \frac{1}{n}(\Sigma WiSi^2)-\frac{1}{N}(\Sigma WiSi^2)$$

όπου Wi = Ni/N.

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

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

Παράδειγμα:

Από 4000 σπουδαστές παίρνουμε δείγμα 40 ατόμων. Η κατανομή του δείγματος τούτου στα επιμέρους στρώματα και τα λοιπά στοιχεία παρουσιάζονται στον ακόλουθο πίνακα :

	Ni	Wi	Si	NiSi	$\bar{X}i$	ni = n ($N_iS_i/\Sigma N_iS_i$)		WiSi	WiSi^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	0,06875	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	1,09375	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 $\bar{X}$ υπολογίζεται με βάση τους προηγούμενους τύπους :

$$S\bar{X}^2 = \frac{1}{40}(1{,}09375)^2 - \frac{1}{4000}(1{,}8275) = 0{,}02945$$

Και το τυπικό σφάλμα της $\bar{X}$ ισούται με S $\bar{X}$ =0,1716

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

Με τα ανωτέρω δεδομένα μπορούμε να υπολογίσουμε την επίδραση του σχεδίου (DEFF) σε σχέση με την απλή τυχαία δειγματοληψία.

5.4. Βέλτιστος καταμερισμός

Υπάρχουν όμως περιπτώσεις όπου οι τιμές του πληθυσμού έχουν μεγαλύτερη διακύμανση σε μερικά στρώματα από ότι σε άλλα. Διαφέρουν δηλαδή σημαντικά οι διασπορές των στρωμάτων. Επομένως για να αντιπροσωπευθούν επαρκώς τα στρώματα αυτά στο δείγμα, θα πρέπει ο λόγος $\frac{ni}{Ni}$ να είναι ανάλογος της τυπικής απόκλισης σ_i του στρώματος. Αυτό σημαίνει ότι τα 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 $V(\hat{\mu}_n)$ γίνεται ελάχιστη εάν τα $n_1, n_2, \ldots, n_k$ 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 (optimum allocation with constant cost per unit) ή καταμερισμός κατά Neyman (Neyman allocation).

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

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

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

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

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

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

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

Η τελευταία σχέση 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.

5.5. Το μέγεθος του Δείγματος

Έστω ότι απαιτείται ένα στρωματοποιημένο τυχαίο δείγμα μεγέθους n του οποίου το $100W_i$% των μονάδων θα προέρχεται από το i στρώμα, i = 1, 2,...., k. Έστω δηλαδή ότι το n καταμερίζεται σύμφωνα με την σχέση

$$n_i = W_i \cdot n, \; \acute{o}\pi o\upsilon \; i = 1,2,\ldots,k$$
$$\Sigma w_i = 1$$

Το πρόβλημα είναι ο προσδιορισμός της τιμής του n που εξασφαλίζει μια δοθείσα τιμή V για την διασπορά της εκτιμήτριας $\hat{\mu}_n$. Πρέπει δηλαδή το n να καθοριστεί έτσι ώστε

$$V(\hat{\mu}_n) = V \Rightarrow \Sigma \left(\frac{N_i}{N}\right)^2 \frac{\sigma_i^2}{n_i}\left(1 - \frac{n_i}{N_i}\right) = V \Rightarrow$$

$$\frac{1}{n}\sum_{i=1}^{k}\frac{(N_i / N)\sigma_i^2}{W_i} - \frac{1}{N}\sum_{i=1}^{k}(N_i / N)\sigma_i^2 = V \Rightarrow$$

$$n = \frac{n_o}{1 + \left[\Sigma(N_i / N)\sigma_i^2\right](NV)}$$

όπου

$$n_o = \frac{1}{V}\sum_{i=1}^{k}\frac{(N_i / N)^2\sigma_i^2}{W_i}$$

Συνήθως το μέγεθος του απαιτούμενου συνολικού δείγματος καθορίζεται με βάση τον βαθμό εμπιστοσύνης με την οποία η εκτίμηση $\hat{\mu}_n$ της μέσης τιμής μ δεν απέχει από την μ περισσότερο από μία δοθείσα τιμή e. Δηλαδή, το πρόβλημα, όπως και στην απλή τυχαία δειγματοληψία, μπορεί να διατυπωθεί ως εξής :

- Ποιο είναι το μέγεθος n του στρωματοποιημένου τυχαίου δείγματος που απαιτείται ώστε

$$P(\left|\hat{\mu}_n - \mu\right| \le e) = 1\text{-}a; \tag{1}$$

Η παραπάνω σχέση είναι ισοδύναμη με την

$$P\left(\left|\frac{\hat{\mu}_n - \mu}{\sqrt{V(\hat{\mu}_n)}}\right| \le \frac{e}{\sqrt{V(\hat{\mu}_n)}}\right) = 1 - a \Leftrightarrow \tag{2}$$

$$\frac{e}{\sqrt{V(\hat{\mu}_n)}} = Z_{1-a/2}$$

Αλλά

$$V(\hat{\mu}_n) = \Sigma \left(\frac{N_i}{N}\right)\frac{\sigma_i^2}{n_i}\left(1 - \frac{n_i}{N_i}\right) = \tag{3}$$

$$= \frac{1}{n}\sum_{i=1}^{k}\frac{(N_i/N)\sigma_i^2}{W_i} - \frac{1}{N}\sum_{i=1}^{k}(N_i/N)\sigma_i^2$$

Αντικαθιστώντας στην (2) την τιμή της $V(\hat{\mu}_n)$ από την (3) και λύνοντας ως προς n έχουμε

$$n = \frac{n_o}{1 + \left[Z_{1-a/2}{}^2\sum_{i=1}^{k}(N_i,N)\sigma_i^2\right]/\left(N_e{}^2\right)} \tag{4}$$

όπου

$$n_o = \frac{Z_{1-a/2}^2}{e^2}\sum_{i=1}^{k}\frac{(N_i/N)^2\sigma_i^2}{W_i} \tag{5}$$

Στην πράξη, η τιμή του n καθορίζεται και στις δυό περιπτώσεις ως εξής :

$$n = \begin{cases} n_o = \dfrac{1}{V}\displaystyle\sum_{i=1}^{k}\dfrac{(N_i/N)^2\sigma_i^2}{W_i} \text{ ή } \quad n_o = \dfrac{Z_{1-a/2}^2}{e^2}\displaystyle\sum_{i=1}^{k}\dfrac{(N_i/N)^2\sigma_i^2}{W_i}, αν \quad n_o/N \le 0.05 \\ n = \dfrac{n_o}{1+\left[\Sigma(N_i/N)\sigma_i^2\right](NV)} \text{ ή } \quad n = \dfrac{n_o}{1+[Z_{1-a/2}^2\sum_{i=1}^{k}(N_i/N)\sigma_i^2](N_e^{\,2})} \end{cases}$$

αν $n_o/N > 0.05$

Παράδειγμα:

Έστω ότι σε κάποιο επάγγελμα, οι μηνιαίοι βασικοί μισθοί ανδρών - γυναικών σε εκατοντάδες χιλιάδες δραχμές είναι οι εξής:

Αποδοχές (Z_i)	10	15	20	25	30	35	40
Άνδρες (a_i)	0	0	100	200	400	200	100
Γυναίκες (γ_i)	500	500	500	0	0	0	0

Ποιο είναι το μέγεθος n του δείγματος που απαιτείται, ώστε με πιθανότητα 95% η εκτίμηση της μέσης τιμής του πληθυσμού να μην διαφέρει από την πραγματική τιμή περισσότερο από 1

α) στην περίπτωση αναλογικού καταμερισμού

β) στην περίπτωση βέλτιστού καταμερισμού

Λύση

e = 1, 1-α = 0,95 Άρα $Z_{0.975} = 1.96$ και επομένως από τις σχέσεις (3) και (4) έχουμε

$$n_o = (1.96)^2\left[\frac{(0.4)^2(30.02)}{W_1}+\frac{(0.6)^2(16.67)}{W_2}\right] = \frac{18.45}{W_1}+\frac{23.05}{W_2}$$

και

$$n = \frac{n_o}{1+\frac{(1.96)^2}{2500}\left[(0.4)(30.02)+(0.6)(16.67)\right]} = \frac{n_o}{1.03382}$$

κατά συνέπεια

α) $W_i = \frac{N_i}{N} = \begin{cases} 0.4, & i=1 \\ 0.6, & i=2 \end{cases} \Rightarrow n_o = 84.54$

οπότε (επειδή $n_o/N = 0.03 < 0.05$) $n = n_o \cong 85$

β) $W_i = \frac{N_i\sigma_i}{N_1\sigma_1 + N_2\sigma_2} = \begin{cases} 0.47, i=1 \\ 0.53, i=2 \end{cases} \Rightarrow n_o = 82.75$

και άρα $n \cong 83$

5.6 Εκτίμηση Ποσοστών

Έστω $N_A^{(i)}$ ο αριθμός των μονάδων του i στρώματος του πληθυσμού που ανήκουν σε μια κατηγορία Α και $X^{(i)}$ ο αριθμός των αντίστοιχων μονάδων του i απλού τυχαίου δείγματος. Τότε, το ποσοστό p των μονάδων του πληθυσμού που ανήκουν στην κατηγορία Α είναι ίσο με :

$$p = \sum_{i=1}^{k} N_A^{(i)}/N = \sum_{i=1}^{k}(N_i/N)p_i$$

όπου

$p_i = \frac{N_A^{(i)}}{N_i} =$ ποσοστό των μονάδων του i στρώματος που ανήκουν στην κατηγορία Α.

Η στατιστική συνάρτηση

$$\hat{p} = \sum_{i=1}^{k} (N_i / N) \hat{p}_i \text{, όπου } \hat{p}_i = \frac{X^{(i)}}{n_i}$$

είναι αμερόληπτη εκτιμήτρια της παραμέτρου p και ισχύει ότι

$$V(p) = \Sigma \left(\frac{N_i}{N}\right)^2 \frac{P_i(1-p_i)}{n_i} \frac{N_i - n_i}{N_i - 1}$$

Αν σε μια δειγματοληπτική έρευνα ενδιαφερόμαστε να κάνουμε συγκρίσεις μεταξύ διαφορετικών στρωμάτων, οι κανόνες για τον καταμερισμό του n είναι διαφορετικοί. Αν, για παράδειγμα ο πληθυσμός αποτελείται από δύο μόνο στρώματα και θέλουμε να συγκρίνουμε τις μέσες τιμές μ_1, μ_2 των στρωμάτων αυτών, τότε μια λογική επιλογή θα ήταν να διαλέξουμε τα n_1 και n_2 έτσι ώστε να ελαχιστοποιείται η $V(\bar{X}_{n_1} - \bar{X}_{n_2})$. Τότε με την υπόθεση μιας γραμμικής συνάρτησης κόστους $c = c_o + c_1 n_1 + c_2 n_2$, οι τιμές των n_1 και n_2 που ελαχιστοποιούν την $V(\bar{X}_{n_1} - \bar{X}_{n_2})$ για σταθερό c είναι

$$n_i = \frac{n\sigma_i / \sqrt{c_i}}{\sigma_1 / \sqrt{\sigma_1} + \sigma_2 / \sqrt{\sigma_2}}, i = 1,2$$

(Για τον καθορισμό των n_i αγνοήθηκε η διόρθωση πεπερασμένου πληθυσμού).

Όταν ο αριθμός k των στρωμάτων είναι μεγαλύτερος του 2, ο βέλτιστος καταμερισμός εξαρτάται από την ακρίβεια που απαιτεί ο ερευνητής για τις διάφορες συκρίσεις.

5.7. Σύγκριση απλής – τυχαίας και στρωματοποιημένης δειγματοληψίας

Αν η στρωματοποίηση γίνει με παράγοντες που δεν συσχετίζονται με το εξεταζόμενο μέγεθος, τα χαρακτηριστικά των επιμέρους στρωμάτων δεν θα διαφέρουν από τα χαρακτηριστικά του πληθυσμού και το όφελος της στρωματοποίησης θα ήταν ανύπαρκτο. Σε σχέση με την απλή τυχαία δειγματοληψία, το όφελος από την ανάλογη δειγματοληψία ισούται με την διαφορά των διακυμάνσεων : Δηλαδή :

Διακύμανση της $\bar{X}$ (απλής τυχαίας δειγμ.) - Διακύμανση της $\bar{X}$ (ανάλογης στρωματοποιημένης δειγμ.) $=\Sigma\ n_i\ (\mu_i-\mu)^2/n^2$

όπου με μ_i συμβολίζεται ο μέσος όρος της $\bar{X}$ στο στρώμα i
και με μ ο αντίστοιχος μέσος όρος του πληθυσμού
Είναι σαφές ότι ο διαχωρισμός των στρωμάτων που γίνεται ανεξάρτητα από το χαρακτηριστικό Χ, θα δώσει διακύμανση μεταξύ των μέσων των στρωμάτων ίση με το μηδέν. Στην περίπτωση αυτή η απλή τυχαία δειγματοληψία δίνει διακύμανση εκτιμήσεων ίση με τη διακύμανση της ανάλογης κατά στρώματα δειγματοληψίας. Συνεπώς, πρέπει να μεγιστοποιηθεί η διακύμανση μεταξύ των στρωμάτων και να ελαχιστοποιηθεί η διακύμανση μέσα στα στρώματα. Αυτό ίσως να σήμαινε ότι κάθε στρώμα πρέπει να αποτελεσθεί από ένα μόνο στοιχείο. Αυτό όμως δεν θα αποτελούσε δειγματοληψία αλλά απογραφή. Για το λόγο αυτό υπάρχει ένα ελάχιστο όριο μεγέθους ίσο με δύο μονάδες κατά δείγμα και συνεπώς κάθε στρώμα πρέπει να περιλαμβάνει τουλάχιστον δύο μονάδες.

Από άποψη ακρίβειας, η στρωματοποιημένη δειγματοληψία έχει τελικά ως αποτέλεσμα μικρότερη διασπορά για την εκτιμήτρια της μέσης τιμής του πληθυσμού. Δεν αληθεύει όμως ότι οποιοδήποτε στρωματοποιημένο δείγμα δίνει μικρότερη διασπορά από ένα απλό τυχαίο δείγμα. Αν οι τιμές n_1, n_2,..., n_k απέχουν πολύ από αυτές του βέλτιστου καταμερισμού, τότε η εκτίμηση της μέσης τιμής μπορεί να έχει μεγαλύτερη διασπορά. Ο διαχωρισμός του πληθυσμού σε στρώματα πρέπει να γίνεται με παράγοντες οι οποίοι σχετίζονται με το εξεταζόμενο χαρακτηριστικό. Αυτό όμως δεν είναι πάντοτε εύκολο να εφαρμοσθεί, δεδομένου ότι πολλά χαρακτηριστικά δεν είναι δυνατόν να ταυτοποιηθούν για κάθε μονάδα του πληθυσμού.

Για παράδειγμα, αν θέλουμε να στρωματοποιήσουμε τους σπουδαστές μιας σχολής βάσει του έτους σπουδών, του φύλου, της κατάστασης απασχόλησης και του συνολικού χρόνου μελέτης, μόνο για τους δύο πρώτους παράγοντες είναι ευχερής η στρωματοποίηση (απ΄ τους ονομαστικούς καταλόγους της σχολής) ενώ για τους άλλους δύο δεν είναι.

5.8. Πλεονεκτήματα – Μειονεκτήματα Στρωματοποιημένης Δειγματοληψίας

5.8.1. Πλεονεκτήματα

1) Όπου υπάρχει μεγάλη απόκλιση στις μεταβλητές και αυτή η απόκλιση μπορεί να συσχετισθεί με πληθυσμιακά χαρακτηριστικά με τρόπο που να επιτρέπει το σχηματισμό πιο ομοιογενών μονάδων η

στρωματοποιημένη δειγματοληψία συνήθως οδηγεί σε καλύτερη εκτίμηση της αληθινής πληθυσμιακής παραμέτρου.

2) Συνήθως έχει ως αποτέλεσμα την σημαντική μείωση της πιθανότητας παραπλανητικών δειγμάτων. Εξασφαλίζει δηλαδή μεγαλύτερη ακρίβεια χωρίς να επιβάλλει αύξηση του μεγέθους του δείγματος.

3) Εξασφαλίζει αντιπροσώπευση συγκεκριμένων μονάδων του πληθυσμού, που διαφορετικά θα μπορούσαν ν' αποκλειστούν διότι αντιπροσωπεύουν μια μικρή αναλογία του συνολικού πληθυσμού.

5.8.2 Μειονεκτήματα

1) Η διαδικασία επιλογής του δείγματος είναι πιο σύνθετη.

2) Η στρωματοποίηση προϋποθέτει περισσότερη προκαταρκτική πληροφόρηση γύρω από το σύνολο του πληθυσμού.

3) Το σύνολο του πληθυσμού πρέπει να ορίζεται και να κατατάσσεται βάσει των χαρακτηριστικών της στρωματοποίησης.

Κεφάλαιο 6ο

Συστηματική δειγματοληψία

ΚΕΦΑΛΑΙΟ 6ο

ΣΥΣΤΗΜΑΤΙΚΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

6.1. Γενικά

Συχνά είναι ταχύτερη και ευκολότερη η επιλογή των μονάδων του πληθυσμού, αν αυτή γίνεται από κάποιο κατάλογο, ξεκινώντας από κάποιο τυχαίο αρχικό σημείο και επιλέγοντας μια μονάδα για κάθε k ($k \rangle 0$) μονάδες, μέχρι να κατασκευασθεί το δείγμα με το δοθέν μέγεθος.

Αν για παράδειγμα θέλουμε να επιλέξουμε 1000 καρτέλες από σύνολο 10.000 καρτελών, είναι σαφώς ταχύτερο να επιλεγεί ένας τυχαίος αριθμός μεταξύ 1 και 10 και να περιληφθεί στο δείγμα η καρτέλα που αντιστοιχεί σ΄ αυτόν τον αριθμό, καθώς και κάθε 10η καρτέλα από εκεί και πέρα, από το να επιλεγούν 1000 τυχαίοι αριθμοί και να περιληφθούν στο δείγμα οι καρτέλες που αντιστοιχούν σ΄ αυτούς.

<u>Ορισμός</u>: Η δειγματοληπτική τεχνική η οποία εισάγει ένα συστηματικό στοιχείο στην διαδικασία επιλογής των μονάδων του πληθυσμού του δείγματος αποτελεί την τεχνική, γνωστή ως συστηματική δειγματοληψία (systematic sampling).

6.2. Περιγραφή Διαδικασίας Λήψεως ενός Συστηματικού Δείγματος

Έστω ότι οι μονάδες ενός πληθυσμού μεγέθους N είναι αριθμημένες από το 1 ως το N. Έστω ότι k ένας θετικός ακέραιος. Για την επιλογή ενός 1 -ανά- k συστηματικού δείγματος μεγέθους n, διαλέγουμε τυχαία μια μονάδα από τις k πρώτες μονάδες και περιλαμβάνουμε στο δείγμα αυτήν και κάθε μονάδα του πληθυσμού που απέχει από αυτή κατά κάποιο πολλαπλάσιο του k.

Η επιλογή της πρώτης μονάδας καθορίζει ολόκληρο το δείγμα. Για παράδειγμα, αν ο πληθυσμός αποτελείται από τις τιμές y_1, y_2,...., y_n και από τις πρώτες k μονάδες του επιλεγεί η k_o, τότε το δείγμα θα αποτελείται από τις εξής μονάδες:

$$y_{k_o},\ y_{k_{o+k}},\ y_{k_{o+2k}},\ldots,\ y_{k_{o+(n-1)k}}$$

Ορισμός: Ένα δείγμα μεγέθους n ονομάζεται 1 - ανά - k συστηματικό δείγμα, αν περιλαμβάνει κάθε k μονάδα του πληθυσμού με αρχικό σημείο επιλεγόμενο τυχαία από τις πρώτες k μονάδες του πληθυσμού.

Αν οι μονάδες του πληθυσμού εμφανίζονται με τυχαία σειρά στη λίστα από την οποία επιλέγουμε το δείγμα, τότε αυτό ονομάζεται ψευτοτυχαίο (pseudorandom).

Παραλλαγές της συστηματικής δειγματοληψίας μπορούν να ορισθούν με βάση διαφορετικές διαδικασίες επιλογής του αρχικού σημείου. Για παράδειγμα είναι δυνατό το αρχικό σημείο να είναι η μονάδα του πληθυσμού με δείκτη

(k+1) / 2, αν ο k είναι περιττός

k / 2 ή (k+2) / 2, αν ο k είναι διαφορετικά

Δηλαδή αντί να ξεκινήσουμε από μία τυχαία μονάδα, ξεκινούμε από το κεντρικό σημείο (ή από ένα σημείο κοντά στο κέντρο) του στρώματος των πρώτων k μονάδων του πληθυσμού.

Επειδή το μέγεθος N του πληθυσμού δεν είναι εν γένει πολλαπλάσιο του k, διαφορετικά συστηματικά δείγματα από τον ίδιο πληθυσμό, ενδέχεται να έχουν διαφορετικό μέγεθος.

Για παράδειγμα, αν N=17 και k=5 τα δυνατά 1 -ανά- 5 συστηματικά δείγματα είναι τα εξής :

1	2	3	4	5
y_1	y_2	y_3	y_4	y_5
y_6	y_7	y_8	y_9	y_{10}
y_{11}	y_{12}	y_{13}	y_{14}	y_{15}
y_{16}	y_{17}			

Για την αποφυγή αυτού του προβλήματος και την επίτευξη σταθερού δειγματικού μεγέθους, χρησιμοποιείται η εξής μέθοδος :

Τα N στοιχεία του πληθυσμού θεωρούνται τοποθετημένα στην περιφέρεια ενός κύκλου. Έστω k ο πλησιέστερος προς το λόγο N/n ακέραιος. Ένα στοιχείο από τα N επιλέγεται τυχαία ως αρχικό σημείο και περιλαμβάνεται στο δείγμα μαζί με όλα τα στοιχεία πάνω στην περιφέρεια του κύκλου, που απέχουν από τα αρχικό κατά πολλαπλάσια του k, καθώς αυτή διατρέχεται κατά τη φορά των δεικτών του ρολογιού.

Η διαδικασία συνεχίζεται ώσπου να επιτευχθεί το επιθυμητό δειγματικό μέγεθος.

Ουσιαστικά, η διαδικασία αυτή ισοδυναμεί με την διαδικασία που επιλέγει στο δείγμα την μονάδα y_{i-N} αν i›N. Επιπλέον είναι εύκολο να διαπιστωθεί ότι η μέθοδος αυτή εξασφαλίζει την ίδια πιθανότητα επιλογής σε κάθε μονάδα του πληθυσμού και συνεπώς οδηγεί σε αμερόληπτες εκτιμήσεις της μέσης τιμής του πληθυσμού.

Για παράδειγμα, αν N=10 και n=4, ακολουθώντας την παραπάνω τεχνική, τα δυνατά 1 - ανά - 3 συστηματικά δείγματα μεγέθους 4 (διότι $k=N/n = 10/4 = 2{,}5 \cong 3$) είναι τα εξής :

1	2	3	4	5	6	7	8	9	10
y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_1	y_2	y_3
y_7	y_8	y_9	y_{10}	y_1	y_2	y_3	y_4	y_5	y_6
y_{10}	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9

6.3. Η Διασπορά της Εκτιμήτριας του Μέσου

Είναι προφανές ότι, αν N = nk, τα k δυνατά 1 -ανά- k συστηματικά δείγματα είναι οι στήλες παρακάτω :

Σύνδεση των k δυνατών συστηματικών δειγμάτων μεγέθους n από ένα 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 την συστηματική δειγματοληψία ο πληθυσμός χωρίζεται σε k σύνθετες μονάδες και η διαδικασία επιλογής ενός συστηματικού δείγματος ισοδυναμεί με την διαδικασία τυχαίας επιλογής μιας συνδέτης μονάδας. Επομένως, ένα 1 -ανά- k συστηματικό τυχαίο δείγμα n μονάδων από ένα πληθυσμό μεγέθους N = nk είναι ένα απλό τυχαίο δείγμα μιας σύνθετης μονάδας από τον πληθυσμό με (σύνθετες) μονάδες τις k στήλες του παραπάνω πίνακα.

Η διαίρεση του αρχικού πληθυσμού των N = nk μονάδων σε k 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 συμβολιστεί με y_{ij}. (Δηλαδή $y_{ij} = y_{(j-1)k+1}$). Τότε, αν μ είναι η μέση τιμή του πληθυσμού και $\bar{X}_n^{(i)}$ ο μέσος του I δείγματος, ισχύει ότι :

$$(N-1)\sigma^2 = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\mu)^2 = (\text{προσθαφαιρώντας το } \bar{X}_n^{(i)}) = \tag{1}$$

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

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

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

είναι η διασπορά μεταξύ μονάδων του πληθυσμού που ανήκουν στο ίδιο συστηματικό δείγμα. Επίσης, αν $\bar{X}_n^*$ συμβολίζει τον μέσο ενός συστηματικού δείγματος, το άθροισμα στον πρώτο προσθετέο της (1) είναι ίσο με k $\sigma^2_{\bar{X}_n}$ αφού

$$\sigma^2_{\bar{X}_n^*} = \frac{1}{k}\sum_{i=1}^{k}[\bar{X}_n^{(i)}-E(\bar{X}_n^{(i)})^2] = \frac{1}{k}\sum_{i=1}^{k}(\bar{X}_n^{(i)}-\mu)^2 \tag{3}$$

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

Τότε:

Η διασπορά του μέσου $\bar{X}_n^*$ ενός 1-ανά-k συστηματικού δείγματος μεγέθους n από ένα πληθυσμό N=nk μονάδων δίνεται από τον τύπο

$$\sigma^2_{\bar{X}^*} = [(N-1)\sigma^2 - k(n-1)\sigma^2_W]/N$$

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

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.

<u>Παρατήρηση:</u>

Είναι προφανές ότι στην περίπτωση εκτίμησης του ποσοστού p των μονάδων του πληθυσμού που ανήκουν σε μια κατηγορία A, εφαρμόζεται η παραπάνω θεωρία, αν η παράμετρος μ αντικατασταθεί με την Παράμετρο p και ο μέσος $\bar{X}_n^{(i)}$ 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 A. Είναι προφανές ότι ο πληθυσμός θα αποτελείται από μονάδες της μορφής

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

Παράδειγμα

Ένας πληθυσμός 360 νοικοκυριών μιας συνοικίας με μικτό πληθυσμό (αριθμημένων από το 1 έως το 360) έχει καταγραφεί σε ένα 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

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

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, από την (3) έχουμε

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

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

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

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 γειτονικών 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.

6.4. Σύγκριση Συστηματικής και Στρωματοποιημένης δειγματοληψίας

Η συστηματική δειγματοληψία εκ πρώτης όψεως φαίνεται να είναι ακριβέστερη από την απλή τυχαία δειγματοληψία. Στην πραγματικότητα στρωματοποιεί τον πληθυσμό σε n στρώματα που αποτελούνται από τις πρώτες k μονάδες, τις δεύτερες k μονάδες κ.ο.κ.

Αναμένεται συνεπώς, ότι ένα συστηματικό τυχαίο δείγμα θα έχει την ίδια περίπου ακρίβεια με ένα στρωματοποιημένο τυχαίο δείγμα με $n_i=1$ όπου $i=1,2,\ldots,n$.

Η διαφορά βρίσκεται στο γεγονός ότι στο συστηματικό δείγμα οι μονάδες έχουν την ίδια σχετική θέση στο στρώμα, ενώ στο στρωματοποιημένο τυχαίο δείγμα η θέση των μονάδων στο στρώμα καθαρίζεται τυχαία.

Το συστηματικό δείγμα δηλαδή είναι πιο ομοιόμορφα κατανεμημένο στον πληθυσμό και αυτό συμβάλλει στο να παρέχει πολύ συχνά ακριβέστερες εκτιμήσεις από ένα στρωματοποιημένο τυχαίο δείγμα.

6.5. Πλεονεκτήματα και Μειονεκτήματα συστηματικής δειγματοληψίας

6.5.1. Πλεονεκτήματα:

1) Είναι σχετικά εύκολη τεχνική ως προς την πραγματοποίησή της και αναλύεται σαν απλό τυχαίο δείγμα.

2) Το δείγμα που λαμβάνεται είναι τυχαίο

3) Ομαδοποιείται ο πληθυσμός, δηλαδή βάσει καταλόγου παίρνει στο δείγμα ένα στοιχείο από κάθε ομάδα. Έτσι τα στοιχεία του πληθυσμού αντιπροσωπεύονται με την ίδια αναλογία στο δείγμα, οπότε το δείγμα είναι περισσότερο αντιπροσωπευτικό.

6.5.3. Μειονεκτήματα

1) Είναι απαραίτητη η ύπαρξη ή ο σχηματισμός καταλόγου με το σύνολο των στοιχείων N του πληθυσμού.

2) Υπάρχει ο κίνδυνος να υπάρχει περιοδικότητα στις τιμές των μονάδων του πληθυσμού όσον αφορά την σειρά εμφάνισής τους στην λίστα, αν το K είναι ίσο με την περίοδο ή ένα πολλαπλάσιό της.

Για παράδειγμα, μια συστηματική επιλογή μονάδων από μια «περιοδική» λίστα οικοδομικών τετραγώνων μιας πόλης, μπορεί να οδηγήσει σ΄ ένα δείγμα που περιέχει τετράγωνα που ανήκουν σε μια γραμμή και επομένως σε αύξηση του σφάλματος των εκτιμητριών.

Εφόσον ο κίνδυνος περιοδικότητας δεν είναι εύκολο να ελεγχθεί, ο κίνδυνος εσφαλμένης χρησιμοποίησης της μεθόδου μπορεί να ελαττωθεί αν το δείγμα είναι αποτέλεσμα ενός αριθμού συστηματικών επιλογών από διαφορετικά στρώματα.

Κεφάλαιο 7ο

Δειγματοληψία κατά ομάδες

ΚΕΦΑΛΑΙΟ 7ο

ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΚΑΤΑ ΟΜΑΔΕΣ

7.1.Γενικά

Από τα προηγούμενα, φαίνεται ότι μια αλλαγή στο σχεδιασμό της δειγματοληψίας και κατά συνέπεια στην μέθοδο εκτίμησης των παραμέτρων ενός πληθυσμού, μπορεί να αυξήσει την ακρίβεια των αποτελεσμάτων, χωρίς απαραίτητα να προϋποθέτει αύξηση του μεγέθους του δείγματος.

Επομένως, μεταξύ δύο δειγματοληπτικών τεχνικών που χρησιμοποιούν ισομεγέθη δείγματα, προτιμητέα είναι αυτή που οδηγεί στο μικρότερο τυπικό σφάλμα.

Σημαντικό επίσης ρόλο στην επιλογή της κατάλληλης τεχνικής παίζει και το κόστος ανά δειγματοληπτική μονάδα όπως και το γεγονός του αν διαθέτουμε δειγματοληπτικό πλαίσιο με τις αρχικές μονάδες του πληθυσμού που συγκεντρώνουν το ενδιαφέρων της έρευνάς μας.

Ουσιαστικά, επιδιώκεται η μέγιστη δυνατή ακρίβεια με το ελάχιστο δυνατό κόστος.

Έτσι οδηγούμαστε στη δειγματοληψία κατά ομάδες. Στην τεχνική αυτή, ο πληθυσμός που πρόκειται να μελετήσουμε είναι δυνατό να θεωρηθεί ως σύνολο που αποτελείται από ορισμένες ξεχωριστές ομάδες δειγματοληπτικών μονάδων. Για καλύτερη κατανόηση, υποθέτουμε πως έχουμε ένα πληθυσμό 2000 νοικοκυριών απ΄ τα οποία θέλουμε να επιλέξουμε δείγμα ίσο με 200 νοικοκυριά.

Για το σκοπό αυτό μπορούμε να ακολουθήσουμε τους εξής δύο τρόπους:

α) Αν υπάρχει ένας κατάλογος των 2000 νοικοκυριών, επιλέγουμε ένα απλό τυχαίο δείγμα μεγέθους 200 απ΄ τον κατάλογο.

β) Διαιρούμε την πόλη σε 400 περιοχές 50 νοικοκυριών και επιλέγουμε ένα απλό τυχαίο δείγμα 4 περιοχών και περιλαμβάνουμε στο δείγμα μας όλα τα νοικοκυριά που ανήκουν σ΄ αυτές (σύνολο 200).

(ή)

Διαιρούμε την πόλη σε 400 περιοχές N_1, N_2, ..., N_{400} νοικοκυριών αντίστοιχα και επιλέγουμε με αναλογική ή βέλτιστη στρωματοποιημένη δειγματοληψία ένα δείγμα 200 νοικοκυριών από τις περιοχές αυτές.

Η (α) μέθοδος οδηγεί σε μικρότερο τυπικό σφάλμα από αυτό της μεθόδου (β). Η (α) όμως συνεπάγεται μεγαλύτερο κόστος γιατί το δείγμα είναι περισσότερο απλωμένο γεωγραφικά, και επομένως το κόστος μετάβασης για τον εντοπισμό των μονάδων που επιλέγονται και το κόστος εποπτείας είναι υψηλότερο.

Στην (β) περίπτωση, το δείγμα είναι λιγότερο απλωμένο μιας και οι δειγματοληπτικές μονάδες είναι σύνθετες.

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

Ακόμη και στον δεύτερο τρόπο της (β) περίπτωσης, το δεύτερο στάδιο της δειγματοληψίας διεξάγεται σε γεωγραφικά μικρότερη έκταση, με αποτέλεσμα τη μείωση του κόστους διεξαγωγής και εποπτείας.

Από τη στιγμή που οι ομάδες (clusters) επιλέγονται, ο ερευνητής μπορεί να περιλάβει στο δείγμα όλες τις στοιχειώδεις μονάδες του αρχικού πληθυσμού που ανήκουν σ΄ αυτές όπως στην (β) περίπτωση τον πρώτο τρόπο, ή μπορεί να επιλέξει ένα δείγμα μικρότερων σύνθετων ομάδων ή και στοιχειωδών μονάδων από τις αρχικές σύνθετες μονάδες όπως στην περίπτωση (β) τον δεύτερο τρόπο.

Οδηγούμαστε στο συμπέρασμα λοιπόν ότι η δειγματοληψία κατά ομάδες μπορεί να γίνει σε ένα ή σε πολλά στάδια. Οι επιμέρους ορισμοί των δύο μεθόδων είναι οι εξής:

Δειγματοληψία κατά ομάδες σε ένα στάδιο. (Single-stage cluster sampling)

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 επιλεγείσες αυτές ομάδες. (όπως ο πρώτος τρόπος της (β) περίπτωσης)

Δειγματοληψία κατά ομάδες σε πολλά στάδια. (Multi-stage cluster sampling)

Ονομάζεται η δειγματοληπτική τεχνική όταν, μετά το πρώτο στάδιο, επιλέγονται δείγματα μικρότερων και μικρότερων ομάδων με τελικό στάδιο την επιλογή δείγματος στοιχειωδών μονάδων (ή την περίληψη όλων των στοιχειωδών μονάδων της τελευταίας κατηγορίας σύνθετων ομάδων). (όπως ο δεύτερος τρόπος της (β) περίπτωσης).

Παρατήρηση:

Αν κατά τα διάφορα στάδια της δειγματοληψίας οι σύνθετες μονάδες επιλέγονται με απλή τυχαία δειγματοληψία, τότε το δειγματοληπτικό σχήμα ονομάζεται απλή δειγματοληψία κατά ομάδες (simple cluster sampling) σε ένα ή περισσότερα στάδια.

Όταν οι σύνθετες μονάδες είναι ομάδες στοιχείων βασισμένες σε γεωγραφικές περιοχές, τότε ο σχεδιασμός ονομάζεται δειγματοληψία κατά περιοχές (area sampling).

Για καλύτερη κατανόηση της διαφοράς των μεθόδων της απλής δειγματοληψίας και της δειγματοληψίας κατά ομάδες, παραθέτουμε το παρακάτω παράδειγμα:

Παράδειγμα:

Έστω ότι πρόκειται να εκτιμηθεί ο μέσος μισθός (σε εκατοντάδες δραχμές) του πληθυσμού του παρακάτω πίνακα με βάση ένα δείγμα μεγέθους 4.

ΑΤΟΜΟ	ΜΙΣΘΟΣ
Α	1300
Β	6300
Γ	3100
Δ	2000
Ε	3600
Ζ	2200
Η	1800
Θ	2700
Ι	1500
Κ	900
Λ	4800
Μ	1900
ΣΥΝΟΛΟ	**32100**
ΜΕΣΗ ΤΙΜΗ	2675

Να υπολογισθεί το τυπικό σφάλμα μου μέσου $\bar{X}_4$ του δείγματος στην περίπτωση απλής τυχαίας δειγματοληψίας και να συγκριθεί με το τυπικό σφάλμα του μέσου $\hat{\mu}_C$ ενός δείγματος που έχει ληφθεί με δειγματοληψία κατά ομάδες σε ένα στάδιο θεωρώντας την εξής διαίρεση του πληθυσμού σε ομάδες (clusters).

ΟΜΑΔΑ (CLUSTER)	ΑΤΟΜΟ	ΜΙΣΘΟΣ (Y)
1	Α	1300
	Δ	2000
	Κ	900
	Μ	1900
2	Ζ	2200
	Η	1800
	Θ	2700
	Ι	1500
3	Β	6300
	Γ	3100
	Ε	3600
	Λ	4800

ΛΥΣΗ:

Α) Περίπτωση δειγματοληψίας κατά ομάδες:

Πρέπει να επιλέξουμε με απλή τυχαία δειγματοληψία μια ομάδα (ένα cluster) και να περιλάβουμε στο δείγμα μας όλα τα άτομα της ομάδας αυτής.

Υπάρχουν 3 δυνατά τέτοια δείγματα (1,2 και 3) και οι μέσοι $\bar{U}_i$, όπου i=1,2,3 των δειγμάτων είναι:

ΔΕΙΓΜΑ	1	2	3
Μέσος $\bar{U}_i$	1525	2050	4450

Αν λοιπόν επιλεγεί μια ομάδα με απλό τυχαίο τρόπο και θεωρήσουμε ως δείγμα μας τα τέσσερα στοιχεία που την απαρτίζουν, ο μέσος του δείγματος αυτού θα είναι η ζητούμενη εκτιμήτρια $\hat{\mu}_C$ του μέσου μισθού μ του πληθυσμού.

Πιο συγκεκριμένα αν επιλεγεί η i_o ομάδα, τότε:

$$\hat{\mu}_C = \bar{U} i_o$$

Επομένως ισχύει: $E(\hat{\mu}_C) = (1525 + 2050 + 4450)/3 = 2675$

Δηλαδή η αναμενόμενη τιμή της θεωρηθείσας εκτιμήτριας με την τιμή του μέσου μισθού που θέλουμε να εκτιμήσουμε, συμπίπτουν.

Άρα, η εκτιμήτρια $\hat{\mu}_C$ είναι αμερόληπτη εκτιμήτρια του μέσου μισθού μ.

Η διασπορά του μέσου $\hat{\mu}_C$ είναι:

$$\sigma^2_{\hat{\mu}_c} = \sum_{i=1}^{3} (\bar{U}_1 - 2675)^2 / 3 = 1621250$$

Το τυπικό σφάλμα είναι:

$$\sigma\mu_C = \sqrt{1621250} = 1273$$

<u>Β) Περίπτωση απλής τυχαίας δειγματοληψίας:</u>

Η διασπορά του μέσου $\bar{X}_4$ ενός απλού τυχαίου δείγματος μεγέθους 4 είναι ίση με

$$\sigma^2_{\bar{x}_4} = \frac{\sigma^2}{4}(1 - \frac{4}{12}) \quad (1)$$

όπου σ^2 είναι η διασπορά ολόκληρου του πληθυσμού:

$$\sigma^2 = \sum_{i=1}^{12} (y_i - 2675)^2 / 11 = 2469318{,}2$$

Οπότε ο $(1) \Rightarrow \sigma^2_{\bar{x}_4} = 411553$

οπότε και : $\sigma_{\bar{x}_4} = 641{,}52$

Δηλαδή, το τυπικό σφάλμα του $\hat{\mu}_C$ είναι 1,98 φορές $(\frac{1273}{641{,}52})$ μεγαλύτερο από το τυπικό σφάλμα $\bar{X}_4$

<u>Παρατήρηση::</u>

Είναι δυνατόν να χωρισθεί ο πληθυσμός σε κατάλληλες ομάδες και να επιλεγεί ένα δείγμα με διασπορά μικρότερη ή ίση της διασποράς ενός απλού τυχαίου δείγματος του ίδιου μεγέθους.

Για παράδειγμα, θα μπορούσαμε να είχαμε ορίσει τις ομάδες ως εξής:

ΟΜΑΔΑ	ΑΤΟΜΑ		ΜΕΣΟΣ ΜΙΣΘΟΣ $\bar{U}_1$
1	ΕΓ	ΗΚ.	2350
2	ΑΘ	ΛΜ	2675
3	ΒΖ	ΔΙ	3000

Και πάλι θα είχαμε: $E(\hat{\mu}_C) = 2675$

$\sigma^2_{\hat{\mu}_c} = 70417$

$\sigma^2_{\hat{\mu}_C} = 265$

Στην περίπτωση αυτή οι ομάδες είναι λιγότερο ομοιογενείς από αυτές της προηγουμένως διαίρεσης.

Γενικά, όσο λιγότερο ομοιογενείς είναι οι ομάδες, τόσο μικρότερο τυπικό σφάλμα επιταχύνεται. Πρακτικά, αυτό είναι αδύνατο να ελεγχθεί.

Πάντως, ο βασικός στόχος της δειγματοληψίας κατά ομάδες δεν είναι η επίτευξη του πιο αξιόπιστου δείγματος στοιχειωδών μονάδων, αλλά η επίτευξη των πιο αξιόπιστων αποτελεσμάτων ανά μονάδα κόστους.

7.2. Εκτίμηση της Μέσης Τιμής

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

Στην περίπτωση της απλής δειγματοληψίας κατά ομάδες, ο πληθυσμός $\{y_1,y_2,...,y_N\}$ διαιρείται σε Μ ομάδες (υποπληθυσμούς, clusters) $U_1,U_2,...,U_M$ μεγέθους $N_1,N_2,...,N_M$, αντίστοιχα, όπου $\sum_{i=1} N_i=N$.

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

$U_i=\{y_{i1},y_{i2},...,y_{iN}\}$ $i=1,2,...,M$

και επομένως, ο αρχικός πληθυσμός είναι η ένωση:

$U_1 U U_2 U \ldots U U_M$

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 ομάδας θα συμβολίζεται με U_{ij}, $j=1,2,\ldots,N_j^1$, $i=1,2,\ldots m$.

Δηλαδή, $U_i = \{U_{i1}, U_{i2} \ldots, U_{iNi}\}$. Εδώ N'_i συμβολίζει το μέγεθος της i επιλεγείσας ομάδας. Αυτό αντιστοιχεί στο πρώτο στάδιο της δειγματοληψίας. Στην περίπτωση που επακολουθεί και δεύτερο στάδιο, το σύνολο $\{U_{i1}, U_{i2}, \ldots, U_{i_{n_i}}\}$, $i=1,2,\ldots,m$ θα συμβολίζει ένα απλό τυχαίο δείγμα n_i μονάδων από την ομάδα U_i που επιλέγει κατά το πρώτο στάδιο ($i=1,2,\ldots m$). Τότε

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

$$\dot{U}_i = \sum_{j=1}^{ni} U_{ij} / n_i = \text{ μέσος του i υποδείγματος}$$

$i=1,2,\ldots,m$

$$S_i^2 = \sum_{j=1}^{n_i} (U_{ij} - \overline{U_i})^2 / (n_i - 1) = \frac{1}{(n_i - 1)} X$$ (άθροισμα τετραγωνικών αποκλίσεων των παρατηρήσεων του i δείγματος από τον μέσο του)=διασπορά του i δείγματος, $i=1,2,\ldots,m$

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

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

$\mu_i = \sum_{j=1}^{N_i} y_{ij} / N$ = μέση τιμή της i ομάδας (του i υποπληθυσμού)

$i=1,2,\ldots,M$

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

Προφανώς

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

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

1) $N_i=N_j, i\neq j$ (<u>Ισομεγέθεις ομάδες,</u> δηλαδή $N_i=\bar{N}$ $i=1,2\ldots,M$)

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

Η στατιστική συνάρτηση

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

είναι μια 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

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

Η στατιστική συνάρτηση

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

αποτελεί μια αμερόληπτη εκτιμήτρια της $V(\hat{\mu}_C)$

Η στατιστική συνάρτηση

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

είναι αμερόληπτη εκτιμήτρια του συνολικού μεγέθους

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

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

2) $N_i \neq N_j$, $i \neq j$ **(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 μια 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 $V(\hat{\mu}_C)$.

7.3. Περίπτωση Ποσοστών

Έστω πληθυσμός μεγέθους N ο οποίος διαιρείται σε M ομάδες μεγέθους N_1, N_2,...,N_M. Έστω p το ποσοστό των μονάδων του πληθυσμού οι οποίες ανήκουν σε μια 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

$$P=\sum_{i=1}^{M} N_i p_i / \sum_{i=1}^{M} N_i \qquad (1)$$

Έστω ότι από το σύνολο των Μ 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 Α, τότε μια αμερόληπτη εκτιμήτρια του p_i είναι η

$$\hat{p}i = x^{(i)} / n_i \quad i=1,2,...,m \qquad (2)$$

Περιοριζόμενοι στην περίπτωση δειγματοληψίας σε ένα στάδιο μόνο, μπορούμε να εφαρμόσουμε 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

$$p_c=\sum_{i=1}^{m} \hat{p}i/m \qquad (4)$$

είναι αμερόληπτη εκτιμήτρια της παραμέτρου p με διασπορά

$$V(\hat{p}_c)=\frac{M-m}{M}\frac{1}{m}\sum_{i=1}^{M}(p_i-p)^2/(M-1) \qquad (5)$$

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

$$S^2_{\hat{p}_c} = \frac{M-m}{M}\frac{1}{m}\sum_{i=1}^{m} (\hat{p}_i - \hat{p}_c)^2/(m-1) \qquad (6)$$

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

Η στατιστική συνάρτηση

$$\hat{P}_c = \sum_{i=1}^{m} x^{(i)} / \sum_{i=1}^{m} N_i' \qquad (7)$$

είναι μια αμερόληπτη εκτιμήτρια της παραμέτρου p με διασπορά

$$V(\hat{p}_c) = \frac{M-m}{M}\frac{1}{m\overline{N}^2}\sum_{i=1}^{M} (N_A^{(i)} - N_i p)^2/(M-1) \qquad (8)$$

Η στατιστική συνάρτηση

$$S^2_{\hat{P}_c} = \frac{M-m}{M}\frac{1}{m\cdot\overline{N}^2}\frac{\sum_{i=1}^{m} X^{(i)2} - 2\hat{P}_c\sum_{i=1}^{m} X^{(i)}N'_i + \hat{P}_c^2\sum_{i=1}^{m} N_i'^2}{m-1} \qquad (9)$$

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

<u>Παράδειγμα</u>:

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 μια 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

$x^{(i)}$	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 (4) έχουμε

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

Επίσης από τη σχέση (6) 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 την εκτίμησης ίσης με

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

και επομένως σε τυπικό σφάλμα ίσο με

$S_{\hat{p}} = 0.0392.$

7.4. Διαφορές της δειγματοληψίας κατά ομάδες με τη στρωματοποιημένη δειγματοληψία

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. Έτσι στη δειγματοληψία κατά ομάδες επιβάλλεται να έχουμε δημιουργήσει τις ομάδες με τέτοιο τρόπο ώστε μέσα σε κάθε μια να υπάρχει τόση ανομοιογένεια όση σε ολόκληρο τον πληθυσμό, γιατί μόνο ορισμένες από τις ομάδες εκπροσωπούνται στο δείγμα.

Στη δειγματοληψία κατά στρώματα, όμως, επιβάλλεται να έχουμε μέσα σε κάθε στρώμα τη μεγαλύτερη δυνατή ομοιογένεια και μεταξύ των στρωμάτων τη μεγαλύτερη δυνατή ανομοιογένεια (ώστε κάθε στρώμα να είναι ομοιογενές και ξεχωριστή ομάδα του πληθυσμού), γιατί στη μέθοδο αυτή εκπροσωπούνται όλα τα στρώματα και πάντα μέσω απλών τυχαίων υποδειγμάτων.

7.5. Πλεονεκτήματα και Μειονεκτήματα της Δειγματοληψίας κατά ομάδες.

7.5.1. Πλεονεκτήματα

1) Είναι οικονομικότερο αντί να παίρνουμε ένα απλό τυχαίο δείγμα απ΄ ευθείας από ένα πολύ μεγάλο πληθυσμό, να παίρνουμε ένα δείγμα αποτελούμενο από n ομάδες από ένα πλήθος N ομάδων.

2) Συντομεύει το χρόνο διεξαγωγής έρευνας.

3) Μειώνεται η πολυπλοκότητα συλλογής δείγματος.

4) Είναι η μόνη εφικτή μέθοδος όταν δεν διαθέτουμε δειγματοληπτικό πλαίσιο με τις αρχικές μονάδες του πληθυσμού που συγκεντρώνουν το ενδιαφέρον της έρευνάς μας.

5) Η πολυσταδιακή δειγματοληψία παρέχει μεγάλη ευελιξία και ποικιλία σύνθετων δειγματοληπτικών σχεδιασμών (συνδυασμών).

7.5.2. Μειονεκτήματα

1) Η αβεβαιότητα ότι τα δείγματα-μονάδες της ομάδας είναι αντιπροσωπευτικά του συνολικού πληθυσμού σαν ολότητας.

Κεφάλαιο 8°

Δειγματοληψία ποσοστών ή Ποσοστιαία Δειγματοληψία

ΚΕΦΑΛΑΙΟ 8ο

ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΠΟΣΟΣΤΩΝ Η ΠΟΣΟΣΤΙΑΙΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

8.1. Γενικά

Μέχρι τώρα, εξετάσθηκαν δειγματοληπτικά σχήματα στα οποία η επιλογή μιας μονάδας του πληθυσμού γινόταν σύμφωνα με μια γνωστή πιθανότητα και επομένως χωρίς διάκριση όσον αφορά το ποιες συγκεκριμένες μονάδες του πληθυσμού θα περιληφθούν στο δείγμα.

Υπάρχει όμως και μια άλλη διαδικασία επιλογής των μονάδων του πληθυσμού που θα περιληφθούν στο δείγμα. Αυτή στηρίζεται στην «κρίση» του ερευνητή ο οποίος επιλέγει μονάδες, έτσι ώστε το δείγμα να είναι «αντιπροσωπευτικό» του πληθυσμού όσον αφορά ορισμένα βασικά χαρακτηριστικά του, όπως ηλικία, φύλο, φυλή, τόπος διαμονής, εισόδημα κτλ. Πρόκειται δηλαδή για μια υποκειμενική δειγματοληψία (subjective sampling) ή δειγματοληψία κατά κρίση (judgemental samplng) ή δειγματοληψία ευκολίας (convenience sampling).

Ένας συνήθης τύπος «κατά κρίση» δειγματοληψίας είναι η δειγματοληψία με προκαθορισμένα ποσοστά.

Ορισμός: Δειγματοληψία ποσοστών ονομάζουμε το δειγματοληπτικό σχέδιο που μοιάζει με τη δειγματοληψία κατά στρώματα, αλλά η επιλογή των μονάδων μέσα σε κάθε στρώμα δεν γίνεται τυχαία. Η επιλογή αυτή γίνεται από τους συνεντεύκτες -ερευνητές με δικά τους προσωπικά κριτήρια.

Αυτή η δειγματοληψία θα μπορούσε να θεωρηθεί μια μορφή στρωματοποιημένης δειγματοληψίας με αναλογικό καταμερισμό μια και τα n_i καθορίζονται έτσι ώστε να εκπροσωπούν το (100 Νι/Ν)% του συνολικού δείγματος. Όμως διαφέρει από την στρωματοποιημένη δειγματοληψία στο ότι η επιλογή των n_i μονάδων από το i στρώμα δεν γίνεται με απλή τυχαία δειγματοληψία. Υπάρχει ένα στοιχείο μη τυχαιότητας που εισάγεται εξ αιτίας του τρόπου επιλογής των μονάδων. Αυτή γίνεται από τον ερευνητή με τρόπο ώστε το (υπο)δειγμα που θα επιλεγεί να εκπροσωπεί ορισμένα προκαθορισμένα ποσοστά όσον αφορά διάφορα άλλα χαρακτηριστικά εκτός από το μέγεθος του i στρώματος. Για παράδειγμα, αν ο πληθυσμός είναι ανθρώπινος, μπορεί να εκπροσωπεί τα ποσοστά 45% και 55% ως προς το χαρακτηριστικό φύλο (άνδρας, γυναίκα) το ποσοστό 85% και 15% όσον αφορά την διάκριση εργαζόμενος-άνεργος και τα ποσοστά 30%, 60%, 10% όσον αφορά την διάκριση στις κλάσεις ηλικίας 15-19, 20-62, 62-άνω. Επομένως ο ερευνητής αποφασίζει αν θα περιλάβει κάποιο άτομο στο δείγμα σύμφωνα με το εάν ανταποκρίνεται στα συγκεκριμένα ποσοστά.

<u>Παράδειγμα</u>

Υποθέτουμε ότι επιθυμούμε να εξετάσουμε το ποσοστό των Ελλήνων που είναι ικανοποιημένοι από την είσοδό μας στην ΕΟΚ. Πρέπει να σχηματίσουμε στρώματα με τα εξής κριτήρια (χαρακτηριστικά ποσόστωσης):

1. Φύλο (άρρενες, θήλυς)

2. Ηλικία (νέοι, ώριμοι, ηλικιωμένοι)

3. Απασχόληση (άνεργοι, απασχολημένοι, μη οικονομικώς ενεργοί)

4. Τομέας (αστικός, αγροτικός)

Πραγματικά, οι άνδρες και οι γυναίκες υπάρχει ένδειξη ότι αξιολογούν διαφορετικά τα αποτελέσματα από την είσοδο της Ελλάδας στην ΕΟΚ. Το ίδιο ισχύει για τις διάφορες ηλικίες και την κατάσταση απασχόλησης. Επίσης, ο γεωργικός πληθυσμός αποκομίζει περισσότερα οφέλη από την ΕΟΚ σε σχέση με τον αστικό και συνεπώς επιβάλλεται και η διάκριση αυτή στη στρωματοποίηση.

Ορισμένα από τα χαρακτηριστικά ποσόστωσης (όπως είναι το φύλο) είναι φανερά και δεν δημιουργούν κανένα σοβαρό πρόβλημα στους συνεντεύκτες κατά το χρόνο διεξαγωγής των συνεντεύξεων. Άλλα χαρακτηριστικά όμως, είναι άδηλα ή αμφίβολα. Παραδείγματος χάρη, ο άνεργος, ο απασχολούμενος και ο οικονομικά μη ενεργός δεν διακρίνεται συνήθως και συνεπώς πρέπει να γίνει από το συνεντεύκτη η ερώτηση αν το συγκεκριμένο άτομο εργάζεται ή ζητάει εργασία ή δεν απασχολείται ούτε ζητάει εργασία. Ανάλογο πρόβλημα έχουμε σε κάποιο βαθμό και στην περίπτωση της ηλικίας. Άλλα προσωπικά χαρακτηριστικά (παντρεμένος, ελεύθερος, χωρισμένος, με παιδιά, χωρίς παιδιά κ.α.) παρουσιάζουν μεγαλύτερες ή μικρότερες δυσκολίες προκειμένου να επισημανθούν από τους συνενετεύκες.

Το ποσοστό των μονάδων που θα επιλεγούν από κάθε στρώμα πρέπει να αντιστοιχεί στην ισχύουσα δομή του πληθυσμού. Τα ποσοστά αυτά υπολογίζονται από τα αποτελέσματα των απογραφών πληθυσμού (φύλο, επάγγελμα, ηλικία κλπ.) μολονότι τα αποτελέσματα αυτά δεν δημοσιεύονται με ικανοποιητική πάντοτε ανάλυση (π.χ. κατανομή του πληθυσμού κατά ηλικία και επάγγελμα δεν γίνεται για μικρούς οικισμούς). Εκτός από τις απογραφές πληθυσμού είναι δυνατό να χρησιμοποιούνται άλλου είδους απογραφές (π.χ. καταστημάτων) ή άλλες ειδικές έρευνες που αποκαλύπτουν τη δομή του εξεταζόμενου πληθυσμού.

Η ποσόστωση από την οποία θα προκύψει το δείγμα μπορεί να γίνει για κάθε χαρακτηριστικό ξεχωριστά και ανεξάρτητα από τα λοιπά χαρακτηριστικά ή σε συσχετισμό με αυτά. Παραδείγματος χάρη, η ποσόστωση ως προς την ηλικία, την απασχόληση και το φύλο δείγματος 200 ατόμων μπορεί να παρουσιάζει την εξής εικόνα (με βάση τα στοιχεία του συνολικού πληθυσμού):

Ηλικία	15-30	31-60	60-	ΣΥΝΟΛΟ
	57	120	23	200

Απασχόληση	εργαζόμενοι	άνεργοι	ΣΥΝΟΛΟ
	180	20	200

Φύλο	άνδρες	γυναίκες	ΣΥΝΟΛΟ
	110	90	200

Την ποσόστωση του ίδιου δείγματος μπορούμε να ενεργήσουμε με συνδυασμό και των τριών χαρακτηριστικών, όπως φαίνεται στον ακόλουθο πίνακα. Αυτό γίνεται για να πετύχουμε μεγαλύτερη αντιπροσωπευτικότητα του δείγματος.

Ηλικία	Εργαζόμενοι		Άνεργοι		ΣΥΝΟΛΟ
	Άνδρες	Γυναίκες	Άνδρες	Γυναίκες	
15-30	30	20	4	3	57
31-60	60	50	4	6	120
61-	10	10	2	1	23
ΣΥΝΟΛΟ	100	80	10	10	200

Ο πιο πάνω πίνακας υποχρεώνει τους συνεντεύκτες να επιλέξουν για το δείγμα τους όχι απλά 110 άνδρες και 90 γυναίκες αλλά τα πρόσωπα αυτά κατανεμημένα και ως προς τα λοιπά χαρακτηριστικά με τον τρόπο που παρουσιάζεται στον αναφερόμενο πίνακα. Αν δεν είναι δυνατό να γίνει η πιο πάνω λεπτομερειακή κατανομή των μονάδων του δείγματος σε όλα τα «φατνία» του πίνακα, η κατανομή αυτή θα περιορισθεί στις περιθωριακές θέσεις του ίδιου πίνακα. Στην περίπτωση αυτή δίνονται οδηγίες στους συνεντεύκτες να εντάσσουν στο δείγμα τους πρόσωπα που συνδυάζουν κατά τρόπο αντιπροσωπευτικό όλα τα χρησιμοποιούμενα χαρακτηριστικά.

8.2. Διαφορά απλής τυχαίας-ποσοστιαίας δειγματοληψίας

Η ποσοστιαία σε σύγκριση με την τυχαία είναι λιγότερο δαπανηρή από άποψη χρήματος και χρόνου, εφόσον ο συνεντεύκτης δεν υπάρχει λόγος να κάνει μεγάλο ταξίδι για να εντοπίσει τον τρόπο διαμονής του προσώπου που επιλέγει στο δείγμα.

Στην ποσοστιαία μπορεί απλώς να αντικαταστήσει το πρόσωπο αυτό με κάποιο άλλο που ανταποκρίνεται στα ποσοστά που έχουν προκαθοριστεί (quotas), χωρίς αυτό να σημαίνει ότι η ταχύτητα και το χαμηλό κόστος μπορούν να θεωρηθούν απόλυτα ως πλεονεκτήματα, εφόσον δεν υπάρχουν συγκριτικά στοιχεία.

Βασικό κριτήριο ακαταλληλότητας της ποσοστιαίας δειγματοληψίας είναι το γεγονός ότι δεν μπορεί να εφαρμοστεί στατιστική μεθοδολογία για την αξιολόγηση συμπερασμάτων που συνάγονται από το δείγμα, δεν μπορεί δηλαδή να γίνει υπολογισμός του σφάλματος που επεισέρχεται κατά την εκτίμηση των παραμέτρων του πληθυσμού.

Βασικό όμως στοιχείο είναι ότι η ποσοστιαία, σε αντίθεση με την τυχαία δειγματοληψία, εξασφαλίζει την ύπαρξη ορισμένων κριτηρίων όσον αφορά το δείγμα και περιλαμβάνει μονάδες οι οποίες θα εξασφαλίσουν αντικειμενικότητα στα αποτελέσματα.

8.3. Πλεονεκτήματα και Μειονεκτήματα Ποσοστιαίας δειγματοληψίας

8.3.1. Πλεονεκτήματα

1) Η δειγματοληψία ποσοστών είναι σχετικά λιγότερο δαπανηρή από την τυχαία δειγματοληψία. Αυτό οφείλεται στο ότι το φαινόμενο της μη ανταπόκρισης και των πρόσθετων επισκέψεων δεν αποτελούν πρόβλημα για το σχέδιο αυτό αφού ο συνεντεύκτης έχει τη δυνατότητα να αντικαταστήσει αμέσως με άλλο πρόσωπο εκείνον τον ερωτώμενο ο οποίος αρνιέται να συνεργασθεί ή δεν έχει χρόνο για μια τέτοια συνεργασία. Επίσης, η μετακίνηση των συνεντευκτών είναι

περιορισμένη, διότι φροντίζουν να εντάσσουν στο δείγμα τους τα πρώτα πρόσωπα που συναντούν με τα επιθυμητά χαρακτηριστικά. Ωστόσο, αξίζει να αναφερθεί εδώ ότι σε περίπτωση λεπτομερειακής και πολυσύνθετης ποσόστωσης του δείγματος ενδέχεται ο συνεντεύκτης να αναζητεί και να μη βρίσκει εύκολα πρόσωπα με σχετικά σπάνιες ιδιότητες (γυναίκα άνεργη, ηλικίας 50-60 ετών, έγγαμη με παιδιά, απόφοιτο τουλάχιστο μέσης εκπαίδευσης, κάτοικο στο Κέντρο της Αθήνας).

2) Η δειγματοληψία ποσοστών απαιτεί σχετικά περιορισμένο χρόνο για να ολοκληρωθεί. Έτσι προσφέρεται ως δειγματοληπτικό σχέδιο για έρευνες που πρέπει να διεξαχθούν μέσα σε μικρό χρονικό διάστημα. Τέτοιες έρευνες είναι όσες αναφέρονται σε ζητήματα που δεν είναι δυνατόν να διακρατηθούν στη μνήμη για μακρό χρονικό διάστημα (ακροαματικότητα, αντίδραση του κοινού σε εξαγγελίες μέτρων από την Κυβέρνηση κλπ.) ή που έχουν επείγοντα χαρακτήρα (σφυγμομέτρηση της κοινής γνώμης προκειμένου να διαμορφωθεί η στάση που θα ακολουθήσει η αντιπολίτευση σε συζήτηση ορισμένου νομοσχεδίου που διεξάγεται στη Βουλή).

3) Η δειγματοληψία ποσοστών είναι δυνατόν να διεξαχθεί σε περιπτώσεις που η τυχαία λήψη των μονάδων του δείγματος δεν μπορεί να γίνει επειδή το αναγκαίο δειγματοληπτικό πλαίσιο είναι ανύπαρκτο. Τέτοιο πλαίσιο, όπως είναι φανερό, δεν χρειάζεται για την πραγματοποίηση της δειγματοληψίας ποσοστών. Ας σημειωθεί εδώ ότι για τα περισσότερα θέματα που συνήθως εξετάζονται με

δειγματοληπτικές έρευνες δεν υπάρχουν έτοιμα δειγματοληπτικά πλαίσια.

4) Τα διοικητικά προβλήματα που παρουσιάζει η δειγματοληψία ποσοστών είναι σχετικά λιγότερα. Αυτό οφείλεται στην έλλειψη των προβλημάτων που προκύπτουν από τις αρνήσεις για συνεργασία και από την επανάληψη των συνεντεύξεων. Άλλωστε, τα πρόσωπα που εντοπίζονται από τους συνεντεύκτες και παρέχουν τις αναγκαίες πληροφορίες κατά τη συνέντευξη, κατά κανόνα παραμένουν άγνωστα και έτσι δεν είναι δυνατόν να γίνει κλασσικός έλεγχος του αν έγινε και πώς έγινε η συνέντευξη (φυσικά, οι συνεντεύκες μπορούν να μαγνητοφωνήσουν τις συνεντεύξεις που παίρνουν).

5) Η δειγματοληψία ποσοστών δίνει τη δυνατότητα αξιολόγησης των πληροφοριών που περιέχονται στο δείγμα με στατιστικές μεθόδους καθώς και υπολογισμού του σφάλματος το οποίο είναι συνυφασμένο με την χρησιμοποιούμενη εκτιμήτρια.

8.3.2. Μειονεκτήματα

1) Η έλλειψη τυχαιότητας κατά την επιλογή των μονάδων του δείγματος δεν επιτρέπει να υπολογισθούν τα τυπικά σφάλματα των εκτιμήσεων της δειγματοληψίας ποσοστών. Δηλαδή συμβαίνει εκείνο που είναι γνωστό στη στατιστική επαγωγή: ότι τα αποτελέσματα μη τυχαίου δείγματος ισχύουν μόνο για το δείγμα από το οποίο υπολογίσθηκαν και δεν είναι δυνατόν να αναχθούν με στατιστικό τρόπο στον αντίστοιχο πληθυσμό. Σε ορισμένες περιπτώσεις όμως, είναι δυνατό να προσεγγισθούν τα δειγματοληπτικά σφάλματα της δειγματοληψίας

ποσοστών. Συγκεκριμένα, το συνολικό δείγμα διαιρείται σε μικρότερα αλλά ανεξάρτητα, δείγματα που αντλούνται από τον ίδιο πληθυσμό. Τα μικρότερα αυτά δείγματα προσεγγίζονται ως δείγματα ποσοστών από τους συνεντεύκτες και παρέχουν τόσες ανεξάρτητες εκτιμήσεις για το επιζητούμενο χαρακτηριστικό όσα είναι τα μερικότερα ανεξάρτητα δείγματα. Η μεταβλητικότητα των εκτιμήσεων αυτών χρησιμοποιείται για να υπολογισθεί η μεταβλητικότητα της δειγματοληψίας ποσοστών.

2) Η δειγματοληψία ποσοστών δεν μπορεί να απαλλαγεί από τη μεροληψία επιλογής των μονάδων του δείγματος που ίσως δημιουργούν οι συνεντεύκτες. Η προσωπική επιλογή ενδέχεται να είναι σοβαρά μεροληπτική.

3) Η δειγματοληψία ποσοστών βρίσκεται «στα χέρια» των συνεντευκτών. Στην πράξη δεν είναι δυνατόν να ελέγξουμε σοβαρά το συνεντεύκτη κατά τον χρόνο που κάνει το έργο εντοπισμού των μονάδων του δείγματος και συλλογής των πληροφοριών.

Κεφάλαιο 9ο

Γενικές Παρατηρήσεις επί των μεθόδων

ΚΕΦΑΛΑΙΟ 9º

ΓΕΝΙΚΕΣ ΠΑΡΑΤΗΡΗΣΕΙΣ ΕΠΙ ΤΩΝ ΜΕΘΟΔΩΝ

Η κατάλληλη χρήση ομάδων (ομαδοποίηση) του πληθυσμού οδηγεί συχνά σε μεγαλύτερη ακρίβεια και ελάττωση της διασποράς των εκτιμητών (σε σύγκριση με την απλή τυχαία δειγματοληψία) είτε είναι στρώματα ή ομάδες στοιχειωδών μονάδων. Γενικά, η δειγματοληψία κατά στρώματα (στρωματοποιημένη) και η δειγματοληψία κατά ομάδες διαφέρουν κυρίως στον τρόπο που το τελικό δείγμα λαμβάνεται, στην μεν στρωματική κάθε στρώμα δειγματολογείται, στην δε κατά ομάδες δειγματοληψία το δείγμα έρχεται σε ομάδες αφού ενώνουμε όλες τις επιλεγμένες ομάδες που λάβαμε στο προηγούμενο στάδιο της δειγματοληψίας (η ένωση των ομάδων δεν θεωρείται ξεχωριστό στάδιο δειγματοληψίας). Τα στρώματα εντός τους τείνουν να είναι ομοιογενή, ενώ οι ομάδες δεν απαιτείται να έχουν ομοιογένεια στο εσωτερικό τους. Επίσης, τα στρώματα (όπως και οι ομάδες) είναι εξαντλητικά του πληθυσμού (ή υποπληθυσμού) και αλληλοαλειόμενα, αλλά οι ομάδες μπορεί να είναι υποδιαιρέσεις στρωμάτων (όπως κοινότητες, οικοκυριά κτλ.).

Η πολυσταδιακή δειγματοληψία εφαρμόζεται συχνά όταν ο πληθυσμός είναι διατεταγμένος σε «ιεραρχικές» ομάδες όπως π.χ. οι ενήλικες σε μια πόλη μπορούν να ομαδοποιηθούν σε τετράγωνα (ή συνοικίες) μέσα σε οικοκυριά μέσα στα τετράγωνα και μετά σε ενήλικες μέσα στα οικοκυριά. Μια τρισταδιακή μέθοδος θα ήταν να επιλέξουμε τυχαίως ένα υποσύνολο τετραγώνων, μετά εντός των επιλεγμένων τετραγώνων να επιλέξουμε ένα υποσύνολο νοικοκυριών, και τελικά ένα δείγμα από ενήλικες εντός των επιλεγμένων νοικοκυριών. Το πρώτο στάδιο, αφού κάθε στρώμα (=τετράγωνο) δεν δειγματολογείται, θεωρείται δειγματοληψία κατά ομάδες. Το δεύτερο στάδιο, μπορεί να γίνει με ομαδοποιημένη ή στρωματική δειγματοληψία αφού ο υποπληθυσμός (στο δείγμα τετραγώνου) είναι φυσικώς στρωματοποιημένος. Ομοίως το δείγμα στο τρίτο στάδιο μπορεί να ληφθεί δειγματολογώντας κάθε νοικοκυριό ή ενώνοντας ένα δείγμα νοικοκυριών (παίρνοντας το δείγμα σε ομάδες). Τα τετράγωνα, δηλαδή τα αντικείμενα της πρώτης δειγματοληψίας (επίπεδο που λαμβάνει χώρα η πρώτη δειγματοληψία), λέγονται πρωτογενείς ή πρωτεύουσες δειγματοληπτικές μονάδες, τα δε νοικοκυριά, δευτερογενείς ή δευτερεύουσες δειγματοληπτικές μονάδες. Η διάκριση σε πρωτογενείς, δευτερογενείς, κτλ., είναι σημαντική, διότι κάθε επίπεδο δειγματοληψίας συμβάλλει μια συνιστώσα στην ολική δειγματική διασπορά.

Στην συστηματική δειγματοληψία η (τυχαία) εκλογή της πρώτης μονάδος a_i, $1\leq i \leq k$ αυτομάτως ορίζει το δείγμα ως α_i, α_{i+k}, α_{2k+i},....$\alpha_{(n-1)k+1}$, όπου $k=N/n=$ το δειγματοληπτικό διάστημα.

Διαισθητικά, αυτή η μέθοδος φαίνεται πιο ακριβής από την απλή τυχαία δειγματοληψία αφού η μέθοδος στρωματοποιεί των πληθυσμό σε n στρώματα αποτελούμενα από τις πρώτες k μονάδες, τις δεύτερες k μονάδες και ούτω καθ' εξής. Ένας άλλος τρόπος θεώρησης της συστηματικής δειγματοληψίας (υποθέτουμε N=nk) είναι να την θεωρήσουμε σαν δειγματοληψία κατά ομάδες, όπου οι k ομάδες είναι:

ΟΜΑΔΑ	ΣΥΝΘΕΣΗ ΟΜΑΔΑΣ
1	1, k+1, 2k+1,..,(n-1)k+1
2	2, k+2, 2k+2 (n-1)k+2
..	....
i	i, k+i, 2k+i,.....,(n-1)k+i
..	....
k	k, 2k, 3k,..........nk

κάθε ομάδα περιέχει n μονάδες και η συστηματική δειγματοληψία είναι ισοδύναμη με την εκλογή (τυχαίως) ενός από τις k ομάδες και θεωρώντας όλες τις μονάδες σ΄ αυτή την ομάδα σαν το δείγμα. Υπάρχει μια περίπτωση που η συστηματική δειγματοληψία είναι ισοδύναμη με την απλή τυχαία δειγματοληψία. Είναι η περίπτωση που η διάταξη των μονάδων του πληθυσμού ως προς τις μετρήσεις μπορεί να θεωρηθεί τυχαία ή μπορεί να γίνει μια τέτοια τυχαία διάταξη. Τότε ένα συστηματικό δείγμα μπορεί να αναλυθεί σαν να ήταν ένα απλό τυχαίο δείγμα και οι εύκολοι τύποι της απλής τυχαίας δειγματοληψίας μπορούν να χρησιμοποιηθούν. Βεβαίως στην απλή τυχαία δειγματοληψία η διάταξη των μονάδων κατά τον χρόνο της λήψης του δείγματος δεν έχει καμιά σημασία.

Βιβλιογραφία

ΒΙΒΛΙΟΓΡΑΦΙΑ

1) Ν.Α. ΤΣΕΡΠΕΣ-Φ.ΑΛΕΒΙΖΟΣ (Πάτρα 1995), «Εισαγωγή στην Θεωρία Δειγματοληψίας»

2) ΠΑΝΑΓΙΩΤΗΣ Θ. ΤΖΟΡΤΖΟΠΟΥΛΟΣ (Αθήνα 1995), «Οργάνωση και Διεξαγωγή δειγματοληπτικών ερευνών

3) Ε. ΞΕΚΑΛΑΚΗ (Αθήνα 1995), «Τεχνικές Δειγματοληψίας»

4) Π. Γ. ΚΥΡΙΑΖΟΠΟΥΛΟΣ-Κ.Κ. ΚΙΟΥΛΑΦΑΣ (Αθήνα 1994), «Εφηρμοσμένο MARKETING (έρευνα αγοράς)»

5) ΠΕΤΡΟΣ Α. ΚΙΟΧΟΣ (Αθήνα 1993), ΣΤΑΤΙΣΤΙΚΗ

6) ΠΕΤΡΟΣ Α. ΚΙΟΧΟΣ (Αθήνα 1993), ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

7) ΒΛΑΣΗΣ ΣΤΑΘΑΚΟΠΟΥΛΟΣ (Αθήνα 1997), Μέθοδοι Έρευνας Αγοράς, ΕΚΔΟΣΕΙΣ Α. ΣΤΑΜΟΥΛΗΣ

Παράρτημα

ΠΑΡΑΡΤΗΜΑ

ΠΙΝΑΚΑΣ ΤΥΧΑΙΩΝ ΑΡΙΘΜΩΝ

13	70	43	69	38	81	87	42	12	20	41	15	26	99	82	78	99
05	22	99	52	32	80	91	72	53	95	81	07	98	14	74	52	58
73	10	22	08	08	68	37	16	36	62	20	02	35	98	21	61	90
53	85	72	86	94	87	18	50	11	47	38	55	66	50	96	78	34
45	52	78	96	68	13	07	31	29	70	09	16	66	81	09	45	92
93	44	87	72	26	75	82	31	72	69	78	85	71	45	32	16	57
91	52	05	93	20	51	99	50	88	62	54	90	51	01	39	18	70
67	62	30	02	88	17	37	25	42	86	00	32	03	08	89	77	12
41	15	25	52	30	93	11	45	10	04	66	94	70	33	74	97	23
40	97	62	48	46	97	04	36	31	27	29	84	85	35	59	59	33
63	53	43	60	30	15	81	67	59	72	63	67	17	24	55	68	32
24	80	13	92	46	28	15	70	28	98	53	36	03	89	83	74	21
03	09	16	31	48	05	10	98	62	14	15	84	82	53	89	92	14
07	84	04	01	66	17	75	68	40	90	39	95	46	10	94	68	39
10	42	77	29	80	73	38	92	11	81	72	50	88	63	55	09	84
66	56	92	13	97	14	87	28	54	29	70	14	85	95	79	72	77
48	57	92	42	97	50	61	19	55	38	55	85	57	85	08	52	30
47	73	26	54	18	05	75	92	95	08	88	44	33	02	47	97	47
04	12	38	93	25	49	91	93	73	14	15	01	47	02	70	30	96
45	42	46	06	93	60	41	09	31	29	52	49	50	69	74	10	51
89	66	51	57	21	54	95	18	56	73	16	02	87	41	05	13	87
13	61	76	96	85	27	81	21	75	39	43	77	80	81	38	51	09
17	41	85	13	20	66	59	22	20	40	91	90	51	74	23	54	88
84	12	16	77	44	53	23	87	91	53	86	97	42	80	83	37	31
25	22	30	16	17	32	34	00	07	25	52	36	35	20	92	81	12
15	28	42	98	67	52	36	12	17	03	83	93	48	64	50	32	57
94	25	51	40	74	85	16	86	09	22	62	06	38	72	38	33	97
36	58	90	91	23	91	19	04	17	20	75	03	85	53	06	41	29
78	51	15	75	57	37	77	67	60	70	44	56	91	03	49	12	47
35	37	15	17	96	24	95	08	39	55	73	67	55	64	16	38	58
74	29	71	49	62	16	02	29	14	16	78	44	49	34	05	46	96
48	98	13	29	19	71	98	71	19	51	86	82	73	65	42	09	39
92	56	68	36	54	55	46	22	96	06	41	55	75	08	62	55	19
15	75	57	26	11	28	98	16	85	39	67	49	02	30	47	76	60
92	22	79	70	66	78	13	97	42	31	80	30	86	08	54	39	88
38	46	74	21	91	55	48	36	26	40	17	70	39	94	05	76	83
70	10	91	20	64	12	33	15	59	43	28	28	35	53	14	30	57

07	34	09	56	26	81	86	91	62	94	83	96	96	17	02	10	89
71	24	86	86	52	67	59	63	22	28	76	43	45	43	73	70	73
19	41	04	60	25	42	09	50	52	69	34	01	65	33	19	62	22
41	29	65	01	15	92	69	53	78	68	58	74	08	05	11	94	46
83	72	49	19	98	09	56	83	25	40	44	42	06	32	95	17	32
67	80	84	09	69	81	58	85	33	16	11	87	12	17	39	12	11
60	25	84	42	22	94	38	96	52	03	38	97	53	12	75	59	76
42	73	48	95	57	51	31	02	68	01	17	09	00	38	12	31	52
22	24	95	68	53	92	82	11	96	03	47	31	35	59	15	77	01
64	69	69	58	40	81	16	60	20	00	84	22	28	26	46	66	36
86	66	17	34	49	85	40	51	40	10	15	33	94	11	65	57	62
15	77	01	64	69	58	40	81	16	60	20	00	84	22	28	26	46
66	36	86	66	17	34	49	85	40	51	40	10	15	33	94	11	65
57	62	94	04	99	05	57	22	71	77	99	68	12	11	14	47	69
35	90	95	16	17	45	86	29	16	70	48	02	00	59	33	93	20
58	34	32	24	34	07	13	26	87	40	20	40	81	46	08	09	74
99	16	92	99	85	19	01	23	11	74	00	74	41	69	10	55	33
20	47	54	16	86	11	16	59	34	71	55	84	03	48	17	60	13
38	71	23	91	83	05	06	67	26	77	14	85	40	52	68	60	41
94	98	18	62	20	94	03	71	60	26	45	17	92	65	50	89	18
75	42	07	50	15	69	86	97	40	25	88	14	17	73	92	07	93
34	61	16	91	59	46	44	45	49	25	36	12	07	25	90	89	55
25	83	47	17	23	93	99	56	14	39	16	63	59	73	21	67	80
00	25	58	25	72	06	12	86	74	54	79	70	85	88	71	58	21
98	48	89	72	47	46	94	78	56	10	65	97	84	79	42	31	49
94	15	31	13	09	45	43	03	82	81	70	51	21	03	18	50	21
99	49	73	06	99	19	24	96	39	43	10	14	12	94	08	55	54
70	14	15	99	60	44	62	72	38	18	36	63	92	61	55	93	77
66	82	10	91	81	51	67	01	47	92	46	90	39	99	64	08	00
07	27	54	96	63	40	54	34	70	27	48	18	68	59	91	83	32
81	23	17	13	01	37	57	92	16	34	15	80	90	25	64	67	77
29	95	84	84	87	22	87	54	42	46	56	28	89	02	06	98	59
90	74	13	38	98	66	23	20	23	90	55	31	83	48	74	73	84
98	13	11	48	25	33	39	27	36	08	99	57	60	42	88	68	25
22	89	67	83	16	94	55	14	00	97	32	51	92	47	03	92	33
73	20	21	29	77	37	06	98	64	63	34	31	43	69	21	94	26
20	73	90	70	92	76	49	14	60	34	43	90	51	72	11	07	75
94	19	49	40	82	36	36	89	29	87	70	08	71	98	49	00	89
89	99	29	08	02	72	32	68	16	29	82	19	25	06	22	30	87
87	44	48	90	91	38	53	10	60	29	40	07	58	97	84	09	04

33	56	72	82	37	97	60	92	76	39	17	84	34	67	65	52	89
90	62	97	04	33	81	91	27	56	46	35	83	71	67	22	15	17
55	56	82	62	88	83	86	38	14	63	89	39	81	90	25	62	58
68	87	73	13	79	15	12	18	34	22	24	75	56	47	45	22	81
30	82	38	34	52	57	48	30	34	17	91	28	00	57	30	92	12
38	95	21	15	70	78	50	88	01	07	90	72	77	99	53	04	34
73	33	47	62	57	08	21	77	31	05	64	04	93	42	20	19	09
71	46	37	32	69	89	56	66	25	32	38	64	70	26	27	67	77
40	04	34	63	98	99	89	31	16	12	90	50	28	96	88	40	52
02	29	82	69	34	50	21	74	00	91	27	52	98	72	03	45	65
30	89	71	45	91	87	63	88	23	62	51	07	69	59	02	89	49
14	98	53	41	92	36	07	76	85	37	84	34	47	32	25	21	15
08	82	34	57	57	35	22	03	33	48	84	37	37	89	76	25	09
31	33	69	45	58	48	00	83	48	94	44	08	67	79	41	41	61
41	15	60	11	88	83	24	82	24	07	78	61	89	42	58	88	22
16	13	24	40	09	00	65	46	38	61	12	90	62	41	11	59	85
18	42	61	29	88	76	04	21	80	78	27	84	05	99	85	75	67
80	05	57	05	71	70	21	31	99	99	06	96	53	99	25	13	63
42	39	30	02	34	99	46	68	45	15	19	74	50	17	44	80	13
86	38	40	45	82	13	44	04	52	43	96	38	13	83	80	72	34
20	84	56	19	49	59	14	85	42	99	71	16	34	33	79	82	85
77	30	16	69	32	46	46	30	84	20	68	72	98	94	62	63	59
44	00	89	06	15	87	38	48	84	88	24	55	46	48	60	06	90
08	83	83	98	40	90	88	25	26	85	74	55	80	85	91	19	05
68	22	58	04	63	21	16	23	38	25	43	32	98	94	65	35	16
91	67	12	43	54	81	87	21	31	40	46	17	62	63	99	71	14
12	64	51	68	50	60	78	22	69	51	98	37	65	43	75	12	91
20	36	25	57	92	33	65	95	48	75	00	06	65	25	90	16	29
34	14	43	49	98	71	31	80	59	57	32	43	07	85	06	64	75
27	29	17	06	11	30	68	70	97	87	21	03	98	68	89	39	71
87	92	14	99	42	10	25	37	30	08	27	75	43	97	54	20	93
50	55	04	21	34	92	89	81	52	15	12	84	11	12	66	87	47
21	06	86	08	35	39	52	28	09	48	09	36	95	36	20	82	53
32	89	92	68.	50	88	17	37	92	02	23	43	63	24	69	80	91
23	97	10	96	57	74	07	95	26	44	93	08	43	30	41	86	45
74	33	78	84	33	38	76	73	43	97	55	98	35	69	45	96	80
46	26	39	96	33	60	20	73	30	79	17	19	03	47	28	40	05
08	50	79	89	58	19	86	48	27	98	99	25	08	94	19	15	81
29	82	14	35	88	03	66	97	10	69	02	25	36	43	71	76	00
67	65	12	69	67	89	55	63	31	50	72	20	33	36	15	62	38

15	62	38	72	92	03	76	09	30	75	77	80	04	24	54	67	60
77	81	15	14	67	55	24	22	20	55	36	93	67	69	37	72	22
18	87	05	09	96	45	14	72	41	46	12	67	46	72	02	59	06
08	58	53	63	66	13	07	04	48	71	39	07	46	96	40	20	86
16	07	79	57	61	42	19	68	15	12	60	21	59	12	07	04	99
54	13	05	46	17	05	51	24	53	57	46	51	14	39	17	21	39
95	27	23	17	39	80	24	44	48	93	75	94	77	09	23	48	75
22	39	44	74	80	25	95	28	63	90	41	19	48	46	72	51	12
69	95	21	30	11	98	81	38	00	53	41	40	04	16	78	67	29
75	75	63	97	12	11	57	05	86	52	82	72	47	72	14	37	72
08	74	79	30	80	70	11	66	79	25	88	01	94	52	31	38	57
04	88	45	98	60	90	92	74	77	87	40	18	65	87	37	08	68
97	35	74	05	75	42	13	49	48	38	74	19	06	42	60	20	79
53	09	93	28	29	80	19	68	30	45	94	49	49	71	21	93	93
26	36	68	48	09	37	69	26	22	80	23	34	10	45	70	83	51
49	16	57	15	79	56	63	22	94	28	11	39	69	55	38	53	06
03	51	79	78	74	75	23	73	75	98	47	85	07	26	02	61	28
21	88	87	28	48	23	44	03	03	80	53	89	07	87	93	30	17
56	41	73	33	41	59	16	59	50	98	24	24	87	06	75	99	52
72	39	19	70	17	01	04	01	22	33	04	84	63	27	65	84	39
97	28	25	81	49	71	69	22	04	51	56	46	56	15	10	69	59
18	87	02	72	08	74	52	16	03	82	20	19	66	23	62	37	51
53	40	11	75	45	13	56	85	31	37	09	17	71	96	79	39	50
60	49	03	41	56	78	33	77	28	92	21	90	10	62	01	97	06
09	16	12	75	04	39	69	95	00	48	26	85	28	73	08	66	92
64	20	19	87	54	88	15	12	54	24	06	99	57	07	28	51	34
31	28	07	58	77	03	98	26	76	09	10	44	57	61	28	60	29
80	04	28	47	76	35	73	67	78	28	09	39	88	63	74	41	26
24	60	22	51	19	34	54	08	24	73	86	72	11	44	69	76	90
59	16	11	26	29	18	97	78	44	43	58	92	78	70	80	09	65
58	54	29	98	27	40	51	92	07	13	58	41	59	56	94	16	32
20	18	34	22	73	57	40	67	17	28	63	57	74	36	18	65	55
53	90	46	56	19	50	58	33	84	53	14	74	17	40	73	86	11
97	16	93	94	65	70	95	95	83	20	91	42	57	95	63	00	86
72	55	71	70	92	04	22	53	19	29	67	29	13	56	70	45	73
99	19	72	58	35	49	09	26	00	74	26	42	94	52	02	83	31
48	21	49	72	97	79	19	64	81	82	78	92	51	96	51	28	79
52	37	68	15	53	22	98	30	16	31	83	24	87	69	29	24	85
97	50	52	53	52	26	98	21	68	69	57	79	42	40	89	55	81
36	05	09	18	11	71	01	63	17	60	11	65	19	43	07	44	86

20	79	79	09	30	81	14	53	80	93	71	94	10	18	14	83	69
13	07	89	72	08	00	37	75	14	94	83	85	06	72	66	07	47
94	26	82	37	43	34	23	00	14	50	96	85	41	17	71	69	20
13	55	88	38	43	75	37	43	83	85	53	74	54	62	99	68	93
02	44	24	97	71	97	93	12	70	89	42	52	33	24	91	05	87
34	90	96	63	54	22	84	36	38	99	85	36	25	03	27	49	24
13	67	06	34	98	04	20	80	12	54	01	18	54	20	76	92	10
18	75	55	82	66	34	77	27	71	79	67	65	85	92	68	16	43
91	25	52	57	15	21	54	40	05	50	67	51	66	45	69	84	72
76	24	00	14	92	14	29	12	17	73	77	46	44	24	30	48	50
10	79	26	21	60	03	48	14	97	58	55	23	12	87	39	84	32
43	46	32	56	15	75	25	12	84	95	87	34	95	31	23	12	64
17	49	12	73	28	23	52	48	11	52	38	09	94	32	47	35	42
79	11	81	74	11	15	23	17	38	69	94	97	10	44	42	85	46
88	22	39	75	16	69	13	84	23	99	36	33	41	99	76	22	29
89	07	35	47	87	44	36	62	09	15	95	74	87	09	63	82	63
91	69	03	55	51	09	74	47	55	75	91	36	57	38	30	89	64
97	39	83	35	33	23	17	29	84	62	29	92	42	03	92	37	46
83	41	18	30	90	44	37	64	79	25	70	07	50	85	32	53	87
69	75	48	72	21	52	51	81	40	10	91	52	27	21	18	64	61
98	71	62	12	56	61	01	54	93	18	86	63	72	22	53	44	23
62	39	52	84	74	90	68	18	63	71	69	30	23	12	85	90	05
90	81	77	18	51	71	27	27	05	29	95	78	06	10	41	62	18
71	30	34	52	65	83	40	13	30	04	29	90	89	64	25	66	36
07	37	44	62	96	74	42	64	75	50	83	42	46	80	76	77	34
97	20	42	09	14	90	43	48	68	82	44	11	33	11	20	42	00
01	22	16	14	12	15	67	22	51	38	78	69	65	25	98	73	40
84	17	74	16	53	31	39	01	98	41	81	63	70	58	43	39	93
09	88	05	86	25	43	50	94	08	44	37	01	53	59	67	11	11
45	55	31	95	88	93	90	37	17	30	92	82	09	42	37	88	43
99	50	29	33	50	16	93	09	74	87	89	10	02	19	45	29	65
04	89	31	32	19	59	85	57	61	81	52	99	80	11	55	21	98
79	27	62	71	14	95	53	03	55	08	43	08	22	50	28	03	18
45	01	19	95	12	24	18	52	44	38	47	15	16	96	03	51	42
10	66	75	62	61	27	82	57	12	45	97	68	57	62	36	61	03
54	98	50	70	88	02	86	48	19	95	23	05	45	01	87	81	18
85	70	79	80	29	19	98	92	71	55	86	72	94	77	08	55	65
92	42	33	06	80	06	33	84	07	32	94	03	20	66	29	98	75
81	17	85	57	47	35	16	84	10	35	58	59	25	89	62	60	77
32	68	26	65	73	90	50	46	62	99	34	08	06	81	46	09	16

51	42	54	77	37	13	85	19	19	44	35	31	20	16	05	25	26
25	50	68	35	90	00	03	38	77	76	94	64	49	45	39	58	07
04	02	04	02	28	49	62	36	97	43	81	59	46	59	26	04	63
29	02	53	02	27	86	70	95	09	77	93	46	95	36	98	08	77
45	05	04	32	43	30	93	41	71	09	43	23	16	33	93	21	87
85	65	66	31	97	67	52	15	25	19	47	70	48	16	91	39	59
13	20	82	34	81	39	46	86	43	23	23	81	42	61	42	37	17
44	25	50	75	62	83	95	41	50	57	81	53	79	98	04	75	77
75	24	52	51	32	79	97	05	81	04	78	50	20	33	21	64	10
19	58	92	23	71	32	96	19	19	62	59	60	23	26	11	30	12
76	53	25	27	36	65	65	05	32	52	48	94	61	60	43	08	29
30	17	11	16	02	63	97	30	79	73	88	64	27	89	92	95	94
15	98	82	79	69	68	50	31	05	12	93	24	38	18	25	64	65
74	43	95	06	26	79	78	87	59	72	45	18	64	49	67	78	83
53	15	77	49	92	83	97	80	22	75	30	52	34	00	43	50	50
72	10	50	95	14	18	26	64	86	21	48	23	45	01	80	49	33
47	04	65	54	45	82	42	90	47	02	27	40	96	41	44	05	54
83	18	74	12	48	68	87	22	36	76	21	72	44	85	55	63	87
74	32	30	17	70	40	90	24	43	84	04	45	20	18	42	25	25
36	30	24	93	08	01	39	37	88	82	00	84	16	82	67	66	77
23	26	91	01	11	26	01	24	06	58	20	33	46	38	86	23	90
75	89	28	38	15	91	81	89	08	86	08	88	20	02	11	67	78
67	39	33	89	97	16	28	94	86	93	86	96	13	43	85	99	51
88	56	56	63	58	22	89	19	26	82	25	94	15	54	65	62	93
19	92	53	92	15	71	47	57	74	69	03	65	57	90	53	17	27
29	84	57	45	80	07	13	57	40	58	34	21	93	90	39	21	79
42	01	84	83	12	79	32	09	56	03	81	90	88	00	71	02	89
19	90	75	68	84	49	53	80	62	19	20	31	14	42	11	17	29
11	33	79	14	20	04	12	40	31	74	39	80	21	37	65	20	97
04	85	55	16	90	71	31	95	15	86	74	87	80	75	71	27	07
89	38	05	46	04	79	67	77	33	21	75	40	51	74	60	53	82
07	67	33	56	52	60	21	50	72	26	28	48	67	31	87	61	59
37	42	91	98	43	33	20	58	62	80	65	19	90	07	84	49	04
41	99	59	15	43	86	34	10	05	99	83	08	02	18	01	22	96
16	04	05	06	28	86	60	70	04	13	28	98	76	78	43	69	31
22	40	03	06	12	45	06	32	34	44	18	01	26	36	78	42	03
31	12	04	99	51	09	49	04	32	68	68	54	64	15	25	68	08
18	54	46	98	33	01	47	85	39	81	11	48	84	07	64	76	37
53	16	98	16	52	52	39	32	22	18	22	04	03	06	77	17	90
35	11	54	89	05	61	10	46	27	43	33	88	92	72	62	01	22

70	77	81	98	78	67	05	62	57	08	79	30	32	62	91	87	05
02	08	26	01	20	16	07	42	88	56	51	31	96	14	85	49	48
00	80	79	60	18	33	92	36	13	50	41	43	59	82	16	65	94
15	35	96	40	87	91	56	91	13	58	85	40	06	36	04	30	15
29	46	60	79	85	99	91	13	99	95	58	75	14	74	88	12	09
92	36	94	07	14	08	90	32	51	29	61	50	60	34	92	25	43
50	33	53	94	81	52	36	31	53	12	74	88	59	99	35	95	31
65	70	30	56	59	08	24	51	75	48	73	11	29	77	08	36	81
71	24	13	38	20	83	02	48	11	67	95	38	97	15	58	18	97
82	95	17	13	46	36	51	36	87	56	10	80	79	40	48	82	37
38	98	94	18	38	88	10	90	29	01	12	48	85	52	97	22	39
88	32	11	43	09	51	32	69	31	63	02	33	47	08	94	85	61
86	87	31	55	50	66	11	37	04	68	14	57	17	08	82	48	38
39	71	44	48	10	19	54	80	24	83	47	06	79	01	78	43	64
89	16	53	05	53	16	98	96	30	89	49	83	32	23	13	32	64
80	66	77	96	02	08	59	58	48	91	81	04	31	64	65	15	13
76	75	40	18	81	33	51	68	04	41	00	72	82	28	68	03	83
30	49	18	17	01	70	06	01	53	04	76	49	93	39	68	00	46
00	49	43	08	86	53	25	50	24	70	63	01	08	52	66	67	54
63	26	60	61	15	83	27	41	02	61	80	72	19	91	56	53	27
67	86	20	90	03	18	48	22	42	82	59	84	31	00	92	15	45
78	40	06	16	28	66	54	93	14	19	00	39	11	13	27	55	35
51	81	15	80	43	36	94	49	89	58	80	80	76	25	65	69	60
66	72	92	63	42	78	21	14	35	00	16	05	92	74	20	31	43
91	10	64	18	60	30	48	99	84	23	37	20	03	50	50	05	07
99	57	92	46	06	55	60	98	81	40	20	72	45	67	83	67	33
76	83	52	32	56	15	09	45	22	54	07	49	70	54	48	84	97
29	62	84	18	48	29	23	75	29	90	68	02	56	04	32	34	68
95	70	15	92	80	82	47	10	21	18	57	83	54	02	09	53	39
89	78	31	98	11	56	27	07	76	59	71	87	56	99	27	28	37
78	82	54	47	20	83	80	10	41	35	22	23	03	98	79	74	41
58	68	87	41	11	08	81	29	89	71	23	10	01	79	25	06	00
42	21	03	88	20	05	35	93	00	68	12	09	55	09	36	54	95
15	07	60	86	67	37	94	24	35	82	44	19	92	96	21	84	29
12	31	66	62	09	54	17	31	23	27	30	37	36	79	75	50	39
44	83	55	47	96	50	93	56	82	58	16	35	18	87	64	08	22
73	43	91	03	57	91	35	40	64	18	61	94	37	16	09	93	96
30	90	00	58	15	99	93	33	67	80	08	59	21	66	13	54	56
33	17	26	25	04	73	18	10	05	34	40	32	65	07	28	68	29
15	44	92	47	28	50	93	03	53	37	70	19	68	59	95	39	87

50	35	50	80	23	67	81	25	02	83	08	12	70	00	25	31	33
21	86	16	30	27	85	16	26	34	50	15	87	22	69	71	36	95
19	60	33	05	29	02	33	74	56	38	84	21	07	35	93	54	70
91	44	09	94	06	89	50	88	83	82	50	11	82	51	30	68	91
71	03	53	38	94	02	52	72	15	44	49	53	42	43	00	36	97
70	22	67	59	98	10	64	68	08	79	06	89	48	41	85	72	10
45	79	46	89	74	73	67	60	15	70	37	61	44	07	67	89	81
80	05	75	64	48	51	68	68	27	71	75	45	32	27	76	35	26
63	56	69	37	19	74	48	63	31	52	36	84	40	66	72	66	03
69	38	02	88	89	71	43	01	87	41	79	42	99	29	41	08	47
79	69	67	64	36	14	82	65	26	40	51	63	42	48	85	48	34
91	53	03	82	64	24	06	31	03	97	44	82	24	89	88	48	66
64	97	27	25	62	23	94	40	54	56	32	97	78	90	58	86	41
85	82	52	08	52	96	26	92	88	93	11	03	23	52	78	23	57
81	37	66	56	99	08	59	19	48	29	69	21	64	95	12	08	15
83	99	02	76	12	16	45	52	66	35	70	93	09	52	75	40	34
98	09	80	62	75	26	64	57	26	46	41	47	90	97	99	46	10
35	42	62	84	37	02	59	78	16	17	96	05	71	39	88	05	34
95	56	39	75	65	47	61	86	33	14	88	55	33	69	70	87	79
63	35	93	23	17	30	14	51	51	17	28	21	74	67	12	11	57
22	96	00	48	52	49	62	09	40	08	30	27	54	70	96	06	52
29	84	34	51	60	19	77	82	16	64	45	02	27	04	65	55	90
84	18	10	29	19	09	66	06	78	37	09	60	50	21	52	72	01
29	48	04	08	55	72	25	25	77	54	26	27	24	39	66	67	06
02	32	99	63	62	42	89	32	20	81	14	08	40	45	22	15	37
83	39	51	30	31	49	94	83	66	02	50	95	18	98	58	84	90
30	90	09	35	41	12	87	93	66	85	96	20	65	34	23	13	05
63	53	97	63	18	86	37	56	20	35	62	66	11	37	30	91	89
43	40	02	41	55	70	52	96	87	02	82	61	21	88	60	65	98
18	65	62	01	97	45	79	51	37	74	47	20	11	48	97	93	73
42	16	13	20	34	51	08	71	52	39	17	71	39	84	97	27	72
92	97	02	34	93	32	95	81	13	92	05	40	70	95	71	66	61
55	35	57	24	52	95	84	90	64	38	39	72	70	17	98	42	85
17	21	09	60	58	86	12	31	11	66	61	43	96	00	93	97	00
85	74	58	28	38	74	68	32	61	87	14	71	83	47	90	11	96
00	29	08	87	42	59	40	24	97	44	99	13	56	87	95	02	47
14	00	42	23	72	03	19	02	41	11	23	36	98	32	19	91	42
58	32	80	82	40	49	71	83	37	93	49	99	60	72	88	14	26
87	38	16	06	82	92	62	32	75	67	64	50	49	39	29	55	53
73	01	84	87	42	88	30	93	75	01	18	34	73	30	28	44	28

35	05	78	73	95	47	83	83	10	03	87	35	31	89	45	64	40
45	80	64	70	95	34	29	36	57	72	32	39	32	78	83	08	53
22	82	48	30	09	56	87	63	83	40	96	33	52	74	75	58	02
04	29	83	32	05	10	48	85	29	24	77	88	45	67	31	74	66
57	12	67	23	22	09	33	91	49	28	72	30	03	03	32	57	95
47	93	86	43	43	30	17	87	60	75	26	58	65	28	70	09	88
25	87	30	23	42	54	31	93	20	96	93	05	95	71	90	83	58
85	25	05	32	03	52	52	94	34	73	87	35	53	33	66	29	50
31	97	89	57	95	55	16	10	42	12	47	60	34	51	84	25	61
90	46	98	64	46	24	71	42	57	65	70	17	49	44	83	03	32
80	06	19	86	14	59	27	14	80	92	91	76	39	57	07	38	03
90	76	90	99	79	63	21	03	00	22	34	59	83	22	53	06	87
18	47	14	62	75	45	02	35	10	31	74	97	55	56	24	84	81
06	28	86	65	17	45	20	75	65	43	98	76	14	42	68	43	55
67	64	12	27	46	00	18	42	44	24	72	86	52	95	61	21	88
87	24	96	04	20	68	00	88	16	39	38	57	70	64	36	88	96
54	26	57	17	63	27	74	13	30	95	40	57	41	19	03	38	11
58	88	67	74	48	90	94	23	36	64	70	84	35	90	33	28	80
41	87	65	29	12	36	64	41	86	43	21	65	27	78	35	17	70
32	19	45	29	59	69	90	23	77	82	23	63	32	70	14	93	29
12	04	33	26	52	26	52	65	67	57	25	37	61	14	90	95	89
54	10	41	27	09	11	61	67	31	48	80	29	29	75	62	66	20
75	19	42	90	85	36	68	94	07	25	78	88	71	92	56	13	38
85	43	53	90	42	22	22	98	40	03	05	12	23	79	44	10	06
24	45	59	25	22	76	96	99	28	14	84	72	37	60	08	57	93
35	62	65	42	27	20	59	50	07	12	34	37	43	89	16	85	44
51	42	73	28	98	89	91	43	43	27	67	69	24	70	00	71	43
05	92	22	43	89	66	89	23	71	33	20	98	75	46	75	11	95
94	46	17	61	72	27	01	33	39	03	80	68	63	76	38	21	28
19	27	38	70	73	82	92	70	60	79	38	93	80	47	18	80	30
12	80	36	12	38	68	05	42	80	22	71	37	41	78	34	80	50
95	04	20	39	29	96	28	70	15	04	92	81	58	71	82	88	34
52	70	29	65	05	37	16	89	87	13	76	96	78	80	29	11	43
40	00	99	35	70	69	58	34	70	21	83	96	26	78	18	33	22
49	38	96	51	19	08	27	01	56	07	44	40	86	57	76	42	27
58	81	00	40	91	12	46	17	09	69	35	42	26	01	07	58	52
41	01	91	48	95	59	45	41	91	77	30	25	56	04	03	64	48
97	51	64	78	05	95	65	36	07	08	01	86	46	94	22	08	18
42	09	03	61	20	83	01	05	84	14	25	59	55	02	52	38	43
86	50	46	61	95	01	24	95	65	59	54	30	45	52	24	91	19

49	42	81	62	32	87	22	02	86	60	83	38	29	99	89	05	41
24	08	77	32	73	66	79	38	42	83	19	25	46	11	48	40	62
96	67	41	11	83	17	78	85	91	80	69	35	56	79	85	78	37
15	20	37	96	73	56	63	12	14	28	93	12	12	71	38	81	60
70	08	67	04	34	46	08	17	80	74	77	75	62	70	70	90	72
97	89	23	51	45	37	83	06	50	91	31	95	98	95	81	72	03
03	58	62	23	74	45	06	26	74	80	36	23	70	43	60	72	48
88	95	48	69	35	40	63	31	60	92	27	82	42	36	64	13	90
92	97	04	48	60	53	90	32	68	19	58	65	03	76	48	17	22
18	01	00	38	26	38	57	90	84	21	49	55	99	37	42	96	82
87	69	15	29	29	11	24	78	09	01	35	27	12	80	54	11	27
66	94	42	87	71	15	33	56	74	49	92	05	98	56	05	89	65
76	46	33	84	55	50	13	03	00	97	36	18	25	57	31	23	27
19	01	67	59	15	00	72	44	40	07	93	42	69	48	89	52	35
76	32	66	08	99	67	99	07	26	67	47	67	42	72	31	67	53
02	47	28	10	24	80	64	66	52	38	34	94	17	54	28	96	96
21	38	74	75	20	54	87	63	74	34	80	48	64	70	79	07	54
15	76	23	67	84	81	43	35	46	20	95	21	48	80	57	84	48
53	60	38	20	38	27	02	61	96	31	36	84	25	44	05	42	61
12	39	71	36	50	82	44	43	32	84	37	77	36	93	64	31	09
63	11	38	68	14	69	63	25	22	53	49	99	06	31	53	17	57
54	14	11	24	60	97	32	52	97	29	86	25	07	61	67	05	06
85	35	29	44	75	95	08	42	39	30	34	12	53	00	91	53	64
53	57	56	76	42	39	29	24	47	87	06	89	29	41	09	87	87
88	80	42	18	56	79	44	06	33	07	20	64	21	48	29	08	44
74	30	45	56	81	14	84	18	09	31	20	98	29	04	09	59	19
31	42	17	42	14	53	97	26	51	79	56	39	88	41	10	79	92
05	54	68	00	81	57	73	00	65	27	37	88	17	32	78	12	96
92	35	32	80	34	12	77	83	97	42	98	41	16	08	33	37	09
63	15	38	22	88	67	54	24	16	70	36	03	43	78	99	13	53
06	95	93	05	22	42	18	68	82	07	35	85	84	11	10	83	89
00	21	35	91	42	57	87	41	13	39	46	60	33	81	89	61	61
62	37	35	74	89	37	32	87	50	25	44	94	20	81	69	94	55
31	46	40	68	65	88	96	64	46	69	09	77	52	93	91	73	15
26	59	20	40	26	77	28	35	71	03	00	56	60	15	16	05	32
41	27	40	43	75	26	53	43	14	09	54	03	37	54	29	51	39

Τυχαίοι Αριθμοί στο διάστημα (0,1)

3407	1440	6960	8675	5649	5793	1514
5044	9859	4658	7779	7986	0520	6697
0045	4999	4930	7408	7551	3124	0527
7536	1448	7843	4801	3147	3071	4749
7653	4231	1233	4409	0609	6448	2900
6157	1144	4779	0951	3757	9562	2354
6593	8668	4871	0946	3155	3941	9662
3187	7434	0315	4418	1569	1101	0043
4780	1071	6814	2733	7968	8541	1003
9414	6170	2581	1398	2429	4763	9192
1948	2360	7244	9682	5418	0596	4971
1843	0914	9705	7861	6861	7865	7293
4944	8903	0460	0188	0530	7790	9118
3882	3195	8287	3298	9532	9066	8225
6596	9009	2055	4081	4842	7852	5915
4793	2503	2906	6807	2028	1075	7175
2112	0232	5334	1443	7306	6418	9639
0743	1083	8071	9779	5973	1141	4393
8856	5352	3384	8891	9189	1680	3192
8027	4975	2346	5786	0693	5615	2047
3134	1688	4071	3766	0570	2142	3492
0633	9002	1305	2256	5956	9256	8979
8771	6069	1598	4275	6017	5946	8189
2672	1304	2186	8279	2430	4896	3698
3136	1916	8886	8617	9312	5070	2720
6490	7491	6562	5355	3794	3555	7510
8628	0501	4618	3364	6709	1289	0543
9270	0504	5018	7013	4423	2147	4089
5723	3807	4997	4699	2231	3193	8130
6228	8874	7271	2621	5746	6333	0345
7645	3379	8376	3030	0351	8290	3640
6842	5836	6203	6171	2698	4086	5469
6126	7792	9337	7773	7286	4236	1788
4956	0215	3468	8038	6144	9753	3131
1327	4736	6229	8965	7215	6458	3937
9188	1516	5279	5433	2254	5768	8718
0271	9627	9442	9217	4656	7603	8826
2127	1847	1331	5122	8332	8195	3322
2102	9201	2911	7318	7670	6079	2676
1706	6011	5280	5552	5180	4630	4747
7501	7635	2301	0889	6955	8113	4364

01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
25	19	17	50	50	46	26	92	62	41	27	66	85	60	70
54	61	41	41	91	88	83	30	32	75	59	03	58	58	83
97	50	71	35	65	67	15	45	73	09	17	60	68	38	05
96	17	27	35	82	80	77	28	97	11	26	72	02	88	96
21	48	84	49	72	93	48	66	75	82	36	33	77	97	35
85	12	09	36	72	81	06	73	04	02	03	10	81	34	44
49	57	40	54	64	88	97	69	03	12	94	45	86	74	66
07	43	79	37	60	96	75	39	46	33	42	41	29	83	73
80	07	51	15	59	55	24	80	49	12	61	68	00	44	58
40	71	81	93	03	03	60	02	42	53	38	35	05	67	73
50	24	44	84	14	02	13	95	71	17	46	16	45	72	36
51	36	08	02	99	65	46	51	84	51	20	85	22	94	38
62	81	28	56	90	81	19	95	58	41	50	80	91	11	62
83	33	85	65	91	68	33	17	85	77	15	53	18	87	75
24	05	75	46	93	05	64	39	09	20	73	52	84	82	81
28	40	31	45	53	96	36	84	57	60	99	82	84	93	66
21	23	47	38	68	53	19	50	06	54	28	00	56	78	63
00	78	78	51	53	72	74	90	79	03	63	27	02	60	44
66	96	71	70	61	05	98	64	67	41	35	00	84	20	51
46	24	17	92	11	04	92	17	17	89	52	52	65	59	36
55	69	47	19	10	36	47	63	23	35	15	03	79	56	48
75	17	81	21	31	84	98	99	77	96	71	72	67	99	24
35	04	66	64	83	34	75	18	40	58	65	35	98	48	02
05	83	68	55	63	72	35	53	51	48	26	41	11	16	45
45	48	17	48	46	21	44	18	99	41	51	94	64	83	03
88	44	33	02	47	97	47	04	12	38	93	25	03	29	72
49	91	93	73	14	15	01	47	02	70	30	96	01	06	30
45	42	46	06	93	60	41	09	31	29	52	49	68	82	39
50	69	74	10	51	89	66	51	57	21	54	95	58	76	46
18	56	73	16	02	87	41	05	13	87	13	61	08	73	29
43	73	70	73	19	41	04	60	25	42	09	50	42	45	01
52	69	34	01	65	33	19	62	22	41	29	65	24	43	22
01	15	92	69	53	78	68	58	74	08	05	11	38	94	28
94	46	83	72	49	19	98	09	56	83	25	40	01	22	61
44	42	06	32	95	17	32	67	80	84	09	69	57	52	92