

ΤΗ ΠΑΤΡΑΣ
ΤΕΧΝΗ ΛΟΓΙΣΤΙΚΗΣ



« Μέθοδοι και αρχές δειγματοληψίας »



Εισηγητή: Μητροπούλος
Παναγιώτης

Μαθητές: Θάνου Μαριέλα
Μαλέα Ανδρέας

ΑΡΙΘΜΟΣ ΕΙΣΑΓΩΓΗΣ	6093
----------------------	------

Τ.Ε.Ι. ΠΑΤΡΑΣ
ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ
ΕΤΟΣ 2002

**ΘΕΜΑ: «ΜΕΘΟΔΟΙ
ΚΑΙ ΑΡΧΕΣ
ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ»**

ΕΙΣΗΓΗΤΗΣ : ΜΗΤΡΟΠΟΥΛΟΣ ΠΑΝΑΓΙΩΤΗΣ
ΜΑΘΗΤΕΣ : ΘΑΝΟΥ ΜΑΡΙΕΛΑ
ΜΑΛΕΑΣ ΑΝΔΡΕΑΣ

Περιεχόμενα

ΠΡΟΛΟΓΟΣ

ΜΕΡΟΣ Ι

Κεφαλαίο 1^ο Εισαγωγή

1. 1 Βασικές στατιστικές έννοιες.....	7
1. 2 Απογραφή και δειγματοληπτική έρευνα.....	9

Κεφαλαίο 2^ο Δειγματοληψία και σχέδια δειγματοληψίας

2. 1 Εισαγωγή.....	11
2. 2 Βασικές έννοιες της δειγματοληψίας.....	11
2. 2. 1 Δειγματοληπτική μονάδα.....	11
2. 2. 2 Δειγματοληπτικό πλαίσιο.....	12
ή ο δειγματοληπτούμενος πληθυσμός.....	12
2. 2. 3 Δείγματα.....	15
2. 2. 4 Τυχαία και μη τυχαία δειγματοληψία.....	17
2. 2. 5 δειγματοληπτικά σφάλματα.....	18
2. 3 Λόγοι χρησιμοποίησης της δειγματοληψίας.....	20
2. 4 Πλεονεκτήματα και μειονεκτήματα της δειγματοληψίας.....	21
2. 5 Σταδία δειγματοληψίας.....	23
2. 5. 1 Το στάδιο του σχεδιασμού της δειγματοληψίας.....	24
2. 5. 2 Στόχοι.....	25
2. 6. 1 Η ακρίβεια και η μεροληψία.....	26

των δειγματικών εκτιμήσεων.....	26
2.6.2 Το μέγεθος του δείγματος.....	27
2.7 Μέθοδοι διενέργειας της δειγματοληψίας.....	28

ΜΕΡΟΣ ΙΙ

Κεφάλαιο 1^{ος} Απλή τυχαία δειγματοληψία

1.1 Εισαγωγή.....	30
1.2 Επιλογή απλού τυχαίου δείγματος.....	31
1.3 Τρόποι εκλογής ενός δείγματος.....	33
1.3.1 Απλή τυχαία δειγματοληψία χωρίς επαναθέσει.....	34
1.3.2 Απλή τυχαία δειγματοληψία με επαναθέσει.....	34
1.4 Συμβολισμοί.....	35
1.5 Εκτιμήσεις και σφάλματα αυτών.....	36
1.6 Εκτίμηση του τυπικού σφάλματος.....	38
από ένα δείγμα.....	38
1.7 Εκτίμηση ποσοστού ή αναλογίας πληθυσμού.....	40
1.7.1 Ποιοτικά χαρακτηριστικά.....	40
1.7.2 Διακυμάνσεις των δειγματικών εκτιμήσεων.....	41
1.8 Διαστήματα εμπιστοσύνης.....	43
1.9 Εκτίμηση μέσων και συνολικών.....	47
μεγεθών υποπληθυσμών.....	47
1.10 Εκτίμηση του μεγέθους του δείγματος.....	48
1.10.1 Προσδιορισμός μεγέθους δείγματος για την.....	48
εκτίμηση μέσου και συνολικού πληθυσμού.....	48
1.10.2 Προσδιορισμός μεγέθους δείγματος λαμβανομένου.....	51
υπόψη και του κόστους.....	51
1.10.3 Το μέγεθος του δείγματος στην.....	52
δειγματοληψία για ποσοστά.....	52
1.10.4 Εκτιμήσεις των πληθυσμιακών διακυμάνσεων.....	53
για τον καθορισμό του μεγέθους του δείγματος.....	53
1.10.5 Μέγεθος δείγματος όταν οι εκτιμήσεις αφορούν.....	56
σε υποδιαίρεσεις του πληθυσμού.....	56
1.11 Δειγματοληψία με πιθανότητα προς.....	57
το μέγεθος των μονάδων.....	57
1.12 Συμπέρασμα.....	58

Κεφαλαίο 2^ο Συστηματική δειγματοληψία

2. 1 Εισαγωγή.....	60
2. 2 Περιγραφή της μεθόδου της.....	60
συστηματικής δειγματοληψίας.....	60
2. 3 Διακύμανση της εκτιμήσεως του μέσου.....	62
2. 4 Εκτίμηση του μέσου και συνολικού πληθυσμού.....	64
2. 5 Επαναλαμβανόμενη συστηματική δειγματοληψία.....	65
2. 6 Συμπέρασμα.....	67

Κεφαλαίο 3^ο Στρωματοποιημένη τυχαία δειγματοληψία

3. 1 Εισαγωγή.....	71
3. 2 Ορισμοί – Συμβολισμοί.....	73
3. 3 Κατανομή δείγματος στα στρώματα.....	75
3. 3 . 1 Αναλογική κατανομή δείγματος.....	77
3. 3 . 2 Αρίστη κατανομή του δείγματος.....	78
3. 3 . 3 Μη αναλογική κατανομή.....	80
3. 3 . 4 Σύγκριση μεταξύ ανάλογου.....	81
και δυσανάλογου κατανομής.....	81
3. 4 Εκτίμηση μέσου και συνολικού πληθυσμού.....	81
3. 4 . 1 Αναλογική κατανομή (proportional allocation).....	83
3. 4 . 2 Άριστη κατανομή (optimum allocation).....	84
3. 5 Σύγκριση απλής τυχαίας ,.....	85
αναλογικής και αρίστη κατανομή.....	85
3. 6 Εκτίμηση του μεγέθους του δείγματος.....	86
3. 7 Στρωματοποίηση μετά την συλλογή του δείγματος.....	88
3. 8 Συμπέρασμα.....	88

Κεφαλαίο 4^ο Δειγματοληψία κατά ομάδες

4. 1 Εισαγωγή.....	92
4. 2 Συμβολισμοί.....	94
4. 3 Εκτίμηση του μέσου και του συνολικού πληθυσμού.....	95
4. 4 Δειγματοληψία κατά ομάδες συνδυαζόμενη.....	96
με στρωματοποίηση.....	96
4. 5 Δειγματοληψία κατά ομάδες με.....	99
πιθανότητα ανάλογη προς το μέγεθος των ομάδων.....	99
4. 6 Συμπέρασμα.....	100

Κεφαλαίο 5^ο Εκτιμήσεις με τη μέθοδο του λόγου και της παλινδρόμησης

5. 1 Εισαγωγή.....	103
5. 2 Εκτίμηση λόγου.....	104
5. 2. 1 Εκτίμηση λόγου χρησιμοποιώντας απλή τυχαία δειγματοληψία	105
5. 2. 2 Εκτίμηση λόγου στη στρωματοποιημένη δειγματοληψία	107
5. 3 Εκτίμηση παλινδρόμησης.....	109
5. 3. 1 Εκτίμηση παλινδρόμησης χρησιμοποιώντας απλή τυχαία δειγματοληψία	110
5. 3. 2 Εκτίμηση παλινδρόμησης στη στρωματοποιημένη δειγματοληψία	111
5. 4 Συμπέρασμα	112

Κεφαλαίο 6^ο Δισταδιακή δειγματοληψία

6. 1 Γενικά	114
6. 1. 1 Εισαγωγικές έννοιες.....	114
6. 1. 2 Το πρόβλημα της κατασκευής των δεσμίδων	115
6. 2 Εισαγωγή στην δισταδιακή δειγματοληψία	117
6. 3 Συμβολισμοί.....	119
6. 4 Εκτίμηση του μέσου και του συνολικού πληθυσμού	120
6. 5 Προσδιορισμός μεγέθους δείγματος	121
6. 6 Η κατανομή των πόρων στην δειγματοληψία δύο σταδίων.....	123
6. 7 Δισταδιακή δειγματοληψία με πιθανότητα ανάλογη προς το μέγεθος των μονάδων	124
6. 8 Συμπέρασμα	126

Κεφαλαίο 7^ο Άλλη μέθοδοι διενέργειας της δειγματοληψίας

7. 1 Τρισταδιακή δειγματοληψία	127
7. 2 Πολυφασική δειγματοληψία	127
7. 3 Δειγματοληψία λογική (κρίσεως).....	128
7. 4 Δειγματοληψία ποσοστών Quota	130
7. 5 Δειγματοληψία με άνεση (χάριν ευκολίας).....	131
7. 6 Δειγματοληψία με σταθερά δείγματα.....	132

BIBΛΙΟΓΡΑΦΙΑ.....	134
-------------------	-----

ΠΡΟΛΟΓΟΣ

Η δειγματοληψία αποτελεί, σήμερα, χρήσιμη αλλά και αναγκαία μέθοδο αναλύσεως προβλημάτων, τόσο για καθαρά ερευνητικούς σκοπούς, όσο και για την διεκπεραίωση καθημερινής εργασίας σχεδόν σε όλους τους τομείς της κοινωνικής και οικονομικής ζωής του ανθρώπου. Έτσι, ένας μεγάλος αριθμός ερευνητών και υπαλλήλων του δημοσίου και του ιδιωτικού φορέα έχει ως αντικείμενο ενασχολήσεως του τη διενέργεια δειγματοληψιών, πάρα το διαφορετικό επίπεδο εργασίας που εισάγει, σε πολλές περιπτώσεις, η σοβαρότητα διαφοροποίηση του προβληματισμού και της εφαρμοζόμενης μεθοδολογίας.

Η εκπαίδευση σε πανεπιστημιακό επίπεδο και η βιβλιογραφική ενημέρωση στο ελληνικό χώρο, σε θέματα δειγματοληψίας, υπήρξε στο παρελθόν ελλιπείς. Μια αξιοπρόσεκτη και σοβαρή προσπάθεια έγινε στην κατεύθυνση της βιβλιογραφίας, χωρίς όμως να δίνονται όλα εκείνα τα στοιχεία της θεωρίας και της σύγχρονης τεχνικής για τη διεκπεραίωση των δειγματοληπτικών ερευνών. Για αυτούς τους λόγους κρίθηκε αναγκαία η συγγραφή του παρόντος, ώστε αφ' ενός μεν να συμβάλλει στον εμπλουτισμό της βιβλιογραφίας, αφ' ετέρου δε να ενημερώσει πλήρως τους ενδιαφερόμενους και ιδίως τους ασχολούμενους με δημοσκοπήσεις, τόσο ως προς το θεωρητικό υπόβαθρο, όσο και στην πρακτική διεκπεραίωση τους. Έτσι η παράθεση της θεωρίας, μέσα στο πλαίσιο της τεχνικής της δειγματοληψίας, θα δώσει στους ερευνητές όλα εκείνα τα αναγκαία στοιχεία για την σωστή λήψη και την αξιολόγηση των δειγματοληπτικών αποτελεσμάτων.

Σκοπός της πτυχιακής μας εργασίας είναι η παροχή βοήθειας στους μαθητές και στους φοιτητές, σε θέματα δειγματοληψίας.

Σας ευχαριστούμε

Θάνου Μαρίελα

Μαλας Ανδρέας

ΜΕΡΟΣ Ι

ΚΕΦΑΛΑΙΟ	Εισαγωγή
1.	

1. 1 Βασικές στατιστικές έννοιες

Στη Στατιστική και ειδικότερα στην θεωρία της Δειγματοληψίας οι έννοιες και οι όροι, που χρησιμοποιούνται, κυρίως, από την Δημογραφία, η οποία αποτέλεσε, ως γνωστόν, τον πρώτο επιστημονικό κλάδο, όπου εφαρμόστηκε και απ' όπου ξεκίνησε η Στατιστική Ανάλυση. Για το λόγο αυτό είναι αναγκαίο να διευκρινισθούν και να ορισθούν μερικές βασικές έννοιες, οι οποίες μπορεί να είναι άγνωστες ή να χρησιμοποιούνται με διαφορετική σημασία σε άλλους επιστημονικούς κλάδους.

Δειγματοληπτική μονάδα είναι το στοιχείο ενός συνόλου το οποίο ερευνάται ως προς μια ή περισσότερες ιδιότητες (χαρακτηριστικά). Π.χ. η οικογένεια μιας πόλης ή μιας χώρας η οποία ερευνάται ως προς τον αριθμό παιδιών ή ως προς το ύψος του εισοδήματος, ο υπάλληλος μιας επιχείρησης ο οποίος ερευνάται ως προς το μέγεθος του μισθού ή ως προς τον αριθμό των ετών υπηρεσίας κ.ο.κ. αποτελούν στατιστικές μονάδες.

Μεταβλητή είναι η ιδιότητα ή το χαρακτηριστικό ως προς το οποίο ερευνάται ένας πληθυσμός. Π.χ. το προσωπικό μιας επιχείρησης μπορεί να εξετασθεί ως προς το φύλο, την ηλικία, το μηνιαίο μισθό, την ειδικευση κ.λ.π. Τα χαρακτηριστικά του πληθυσμού, μπορούν να καταταχθούν, να διαβαθμιστούν ή να μετρηθούν συγκρίνοντας το μέγεθος τους με ένα άλλο πρότυπο που λαμβάνεται ως μονάδα. Σε οποιοδήποτε κατάταξη ή διαβάθμιση μπορεί να αντιστοιχισθεί, αυθαίρετα, μια ακολουθία σταθερών τιμών.

Κάθε μεταβλητή παίρνει ορισμένες τιμές, οι τιμές αυτές μπορεί να είναι είτε αριθμητικές είτε να εκφράζονται με λέξεις. Με τη μέτρηση ή

την παρατήρηση παίρνονται ορισμένες τιμές των μεταβλητών οι οποίες συμβολίζονται στα επόμενα με τα μικρά γράμματα, π.χ. x, y, z, ενώ αυτές καθ' αυτές οι μεταβλητές με τα αντίστοιχα κεφαλαία X, Y, Z. Μια μεταβλητή επονομάζεται σταθερή, εάν για κάθε παρατήρηση ή μέτρηση καθορίζεται μια μόνο σταθερή τιμή.

Για την κατανόηση των εννοιών της μεταβλητής και των τιμών αυτής αναφέρουμε τα παρακάτω παραδείγματα :

ΜΕΤΑΒΛΗΤΕΣ	ΤΙΜΕΣ
Αριθμός παιδιών σε μια οικογένεια	0, 1, 2, 3 ..
Ύψος μαθητών ενός σχολείου	135, 138, 140
Ρίψη ενός νομίσματος	Πρόσωπο, γράμματα
Φύλο ατόμων	Άρρεν, θήλυ
Οικογενειακή κατάσταση ατόμων	Άγαμος, Έγγαμος ...

Οι μεταβλητές διακρίνονται σε δυο βασικές κατηγορίες :

1. Ποσοτικές μεταβλητές: λέγονται αυτές που οι τιμές τους εκφράζονται με αριθμούς
2. Ποιοτικές μεταβλητές: λέγονται αυτές οι οποίες δεν μπορεί να μετρηθούν, δηλαδή δεν μπορεί να εκφραστούν με αριθμούς. Οι τιμές των ποιοτικών μεταβλητών εκφράζονται με λέξεις οι οποίες επιτρέπουν την ταξινόμηση του ερευνώμενου πληθυσμού σε διάφορες κατηγορίες.

Με την έννοια του πληθυσμού ή της ολότητας ορίζεται ένα βασικό σύνολο στοιχείων που πρόκειται να μελετηθεί, ως προς μια ή περισσότερες χαρακτηριστικές ιδιότητες. Θα πρέπει να διευκρινισθεί, ιδιαίτερα, ότι τα στοιχεία του πληθυσμού μπορεί να είναι συγκεκριμένα άτομα, αλλά μπορεί να είναι και οποιοδήποτε άλλο σύνολο αντικειμένων ή τιμών που προέκυψαν από πειράματα ή παρατηρήσεις.

Το πείραμα ή οι παρατηρήσεις μπορούν να επαναληφθούν, θεωρητικά τουλάχιστον, άπειρες φορές. Έτσι, το μη πεπερασμένο σύνολο τιμών που προκύπτει μετά από άπειρες εκτελέσεις ενός πειράματος, που έγινε πάντα με τις ίδιες προϋποθέσεις, ή γενικότερα το μη πεπερασμένο σύνολο στοιχείων που πρέπει να διερευνηθεί, ως προς μια χαρακτηριστική

ιδιότητα, ονομάζεται άπειρος πληθυσμός. Σε πολλές περιπτώσεις, παρ' ό τι είναι πεπερασμένο το πλήθος των στοιχείων του πληθυσμού που αναφέρεται συνήθως ως μέγεθος του πληθυσμού, εν τούτοις θεωρείται ο πληθυσμός άπειρος, εφ' όσον το μέγεθος είναι αρκετά μεγάλο σε σχέση με την κλίμακα μετρήσεων που χρησιμοποιείται.

Ο πληθυσμός που διερευνάται θα πρέπει να οριστεί με απόλυτη ακρίβεια ως προς τα στοιχεία που τον απαρτίζουν. Π.χ. όταν αναφέρονται οι κάτοικοι μιας πόλης, θα πρέπει να καθορισθεί επακριβώς ποιοι και πόσοι κάτοικοι περιλαμβάνονται. Είναι κατανοητό ότι ανάλογα με τον τρόπο, που ορίζονται οι κάτοικοι της πόλεως, είναι δυνατόν να καθορισθούν εντελώς διαφορετικά σύνολα πληθυσμών.

Τα αποτελέσματα της επιλογής μιας πεπερασμένης σειράς πειραμάτων, παρατηρήσεων ή στοιχείων επονομάζεται, ιδιαίτερα, δείγμα. Γενικότερα, ορίζεται ως δείγμα από ένα πληθυσμό με N στοιχεία, κάθε πεπερασμένο υποσύνολο με n στοιχεία ($n \leq N$), οπού n ονομάζεται το μέγεθος του δείγματος.

Θα πρέπει να διευκρινισθεί ότι, σε αντίθεση με τους πληθυσμούς που μπορούν να περιγραφούν με κάποια απαρίθμηση ή με κάποια ιδιότητα, τα δείγματα ορίζονται, σχεδόν πάντα, με απαρίθμηση των στοιχείων που τα απαρτίζουν.

1.2 Απογραφή και δειγματοληπτική έρευνα

Η μελέτη ενός φαινομένου (οικονομικό, κοινωνικό κ.λ.π.) προϋποθέτει την ύπαρξη στατιστικών στοιχείων. Έτσι στατιστικά δεδομένα ή στατιστικά στοιχεία ή παρατηρήσεις καλείται το σύνολο των πληροφοριών ή μετρήσεων από την έρευνα ενός φαινομένου.

Το πρώτο λοιπόν στάδιο για οποιοδήποτε στατιστική μελέτη είναι η συλλογή στατιστικών στοιχείων. Η συλλογή στατιστικών στοιχείων γίνεται κυρίως με απογραφές ή με δειγματοληπτικές έρευνες.

Οι απογραφές (censuses) αποτελούν καθολικές έρευνες, δηλαδή συγκέντρωση πληροφοριών από όλες τις στατιστικές μονάδες του ερευνώμενου πληθυσμού.

Οι δειγματοληπτικές έρευνες (sample surveys) καλύπτουν μέρος του ερευνώμενου πληθυσμού, δηλαδή ερευνάται ένα μέρος του πληθυσμού του οποίου ονομάζεται δείγμα. Η επιλογή των στατιστικών μονάδων αποτελούν το δείγμα, γίνεται κατά τρόπο, ώστε οι πληροφορίες των χαρακτηριστικών που θα συγκεντρωθούν να ισχύουν για το σύνολο

του ερευνώμενου πληθυσμού. Με άλλα λόγια το δείγμα πρέπει να επιλέγεται κατά τέτοιο τρόπο ώστε να είναι αντιπροσωπευτικό για τον σύνολο του πληθυσμού στον οποίο ανήκει. Ένα δείγμα θεωρείται αντιπροσωπευτικό όταν τα συμπεράσματα μπορούν να γενικευθούν με αξιοπιστία στο σύνολο του πληθυσμού από τον οποίο προέρχεται.

Βασική προϋπόθεση για την επιτυχία μιας δειγματοληπτικής έρευνας είναι ο προσδιορισμός του μεγέθους του δείγματος και η επιλογή της κατάλληλης μεθόδου διενέργειας δειγματοληψίας η οποία πρέπει να βασίζεται στο βαθμό ομοιογενείας των στατιστικών μονάδων. Φυσικά πάντα υπάρχει κίνδυνος να έχουμε κάποιο βαθμό λάθους.

Εκτός των ανωτέρω, μια συγκριτική αξιολόγηση των δύο μεθόδων συλλογής πληροφοριών προδίδει σοβαρά μειονεκτήματα στην μέθοδο της απογραφής και πολλά πλεονεκτήματα στην μέθοδο της δειγματοληψίας.

Τα βασικά μειονεκτήματα της απογραφής είναι :

1. η καθολική απογραφή ενός μεγάλου πληθυσμού συνεπάγεται σημαντική δαπάνη χρήματος και χρόνου.
2. οι απογραφείς θα επιφέρουν τα προσωπικά τους σφάλματα, τα οποία συσσωρευμένα δυνατόν να προκαλέσουν απατηλή εικόνα του συνόλου.
3. υπάρχει διάσταση μεταξύ ημερομηνίας διεξαγωγής της απογραφής και δημοσιεύσεως των αποτελεσμάτων της.

Σχετικά με τα πλεονεκτήματα και μειονεκτήματα της δειγματοληψίας θα αναφερθούν στον επόμενο κεφαλαίο.

ΚΕΦΑΛΑΙΟ 2 .	Δειγματοληψία και σχέδια δειγματοληψίας

2 . 1 Εισαγωγή

Σ' αυτό το κεφάλαιο θα αναφερθούμε στα βασικά στοιχεία της Δειγματοληψίας και τα εναλλακτικά σχέδια δειγματοληψίας. Τα θέματα αυτά αποτελούν μια γενική ανασκόπηση, με την έννοια ότι θα πρέπει να τεθούν και να καθορισθούν ευθύς ώστε να προσδιορισθεί το πρόβλημα αποφάσεως σε στατιστικούς όρους και προς της συγκεντρώσεως των στοιχείων.

Δειγματοληψία είναι η μέθοδος μελέτης ενός πληθυσμού με βάση ένα δείγμα που παίρνουμε από τον πληθυσμό αυτό.

2 . 2 Βασικές έννοιες της δειγματοληψίας

Μερικοί τεχνικοί όροι σχετικοί με την θεωρία της δειγματοληψίας πρέπει να γίνουν κατανοητοί.

2 . 2 . 1 Δειγματοληπτική μονάδα

Κατά τον σχεδιασμό της δειγματοληψίας είναι απαραίτητα, πριν προβούμε στην επιλογή του δείγματος, να ορίσουμε το σύνολο των μονάδων που αποτελούν τον ερευνώμενο πληθυσμό, οι οποίες

ονομάζονται δειγματοληπτικές μονάδες (sampling units). Π.χ. αν θέλουμε να μελετήσουμε τις αντιδράσεις των υπαλλήλων μιας επιχείρησης, σε μια απόφαση της διοίκησης για αναπροσαρμογή των μισθών, δειγματοληπτικές μονάδες θα είναι όλοι η υπάλληλοι της επιχείρησης. Αν θέλουμε να μελετήσουμε το ύψος του εισοδήματος των νοικοκυριών μίας πόλης δειγματοληπτικές μονάδες θα είναι όλα τα νοικοκυριά της πόλης κ.ο.κ.

Η δειγματοληπτική μονάδα πρέπει να ορίζεται με σαφήνεια, ώστε να μπορούμε να γενικεύουμε τα συμπεράσματα που προκύπτουν από το δείγμα στο σύνολο του ερευνημένου πληθυσμού. Π.χ. όταν οι δειγματοληπτικές μονάδες αναφέρονται σε οικογένειες, νοικοκυριά, καταστήματα, επιχείρησης κ.λ.π. πρέπει να δίνεται στους ερευνητές, ο σχετικός ορισμός των μονάδων αυτών, γιατί κάθε μια από αυτές τις μονάδες έχει και διαφορετική έννοια. Για παράδειγμα το «νοικοκυριό» μπορεί να αποτελείται από δύο ή περισσότερα άτομα, συγγενικά ή μη που διαμένουν στην ίδια κατοικία, προμηθεύονται συνήθως από κοινού τα απαραίτητα για την συντήρησή τους και τρώνε κατά κανόνα μαζί (πολυπρόσωπο νοικοκυριό). Επίσης νοικοκυριό αποτελεί κάθε άτομο το οποίο διαμένει μόνο του σε μια κατοικία, αλλά δεν προμηθεύεται από κοινού με αυτά τα απαραίτητα για την συντήρησή του, ούτε τρώει μαζί τους (μονοπρόσωπο νοικοκυριό). Αντίθετα ο όρος «οικογένεια» προϋποθέτει την ύπαρξη συγγένειας μεταξύ των ατόμων.

Είναι φανερό ότι σε μια έρευνα μερικοί ερευνητές συγκεντρώνουν πληροφορίες σε επίπεδο νοικοκυριού και μερικοί σε επίπεδο οικογένειας και ειδικότερα όταν κάποια από τα μέλη της οικογένειας διαμένουν σε άλλη πόλη (π.χ. φοιτητές, στρατιώτες κ.λ.π.) τότε τα στοιχεία που θα συγκεντρωθούν δεν μπορούν να γενικευθούν επαγωγικά στο σύνολο του ερευνώμενου πληθυσμού.

2.2.2 Δειγματοληπτικό πλαίσιο **ή ο δειγματοληπτούμενος πληθυσμός**

Ο καθορισμός του πληθυσμού σε σχέση με την φύση του ερευνώμενου προβλήματος, είναι με κάποια έννοια η ταυτοποίηση των στοιχειωδών μονάδων του πληθυσμού. Οι στοιχειώδεις μονάδες πρέπει να ορισθούν επακριβώς και μάλιστα πριν αρχίσει η δειγματοληψία. Όταν γίνει αυτό, τότε θα έχουμε κάποια ιδέα περί του εάν ο καθορισμός του πληθυσμού είναι λειτουργικά εφικτός (διαθέσιμος για χρησιμοποίηση).

Σε πολλές περιπτώσεις, η προσέγγιση του πληθυσμού δεν παρουσιάζει προβλήματα. Εάν για παράδειγμα, θέλουμε να εκτιμήσουμε την ποιότητα μιας παραγωγικής διαδικασίας, στοιχειώδεις μονάδες θα είναι οι εκροές που δημιουργούνται σε συνεχή ροή, από την παραγωγή. Επίσης, εάν θα θέλαμε να καθορίσουμε τις αντιδράσεις των εργαζομένων σε μια επιχείρηση, για μια πιθανή πολιτική της διεύθυνσής της επιχειρήσεως, που θα επιδράσει στις συνθήκες εργασίας τους, στοιχειώδεις μονάδες θα είναι όλοι οι εργαζόμενοι στην επιχείρηση. Μπορεί όμως να παρουσιασθούν περιπτώσεις που ο υπό έρευνα πληθυσμός εύκολα καθορίζεται αλλά οι στοιχειώδεις μονάδες αυτού είναι δύσκολο να εντοπισθούν. Για παράδειγμα, ένας πληθυσμός που μπορεί εύκολα να ορισθεί είναι «το σύνολο των διαφόρων τύπων συσκευών τηλεόρασης που είναι σε χρήση», αλλά στην περίπτωση αυτή δεν μπορούμε να γνωρίζουμε ακριβώς που είναι τοποθετημένες όλες οι συσκευές τηλεόρασης που αποτελούν τον ανωτέρω πληθυσμό.

Οποτεδήποτε η προσέγγιση του πληθυσμού παρουσιάζει δυσκολίες, το πρώτο πράγμα που πρέπει να φροντίσουμε είναι να ετοιμάσουμε ένα είδος καταλόγου, από τον οποίο να διαπιστώνονται οι στοιχειώδεις μονάδες αυτού. Ένας τέτοιος κατάλογος ονομάζεται πλαίσιο (frame or working population) ή δειγματοληπτούμενος πληθυσμός (sampled population). Με άλλα λόγια δειγματοληπτικό πλαίσιο είναι το σύνολο των δειγματοληπτικών μονάδων του ερευνώμενου πληθυσμού, το οποίο είναι καταχωρημένο σε έναν κατάλογο ή παρουσιάζονται υπό μορφή χαρτογραφικών διαγραμμάτων (οικοδομικά τετράγωνα μιας πόλης κ.λ.π.). Οι ονομασίες αυτές δηλώνουν ότι πλαίσιο συνιστά τον δειγματοληπτικά εφικτό πληθυσμό. Αντιθέτως, ο πληθυσμός όπως αρχικά ορίζεται σαν σύνολο, που σκοπεύουμε να αναφερθούμε δειγματοληπτικά, αποτελεί τον πληθυσμό στόχο (target population).

Ο πληθυσμός στόχος μπορεί (ή και όχι) να συμπίπτει με τον δειγματοληπτούμενο πληθυσμό. Η επιτυχία (ή αποτυχία) μιας στατιστικής έρευνας εξαρτάται από τον δειγματοληπτούμενο πληθυσμό, δηλαδή αυτόν που περιέχεται στο πλαίσιο (frame), και συνεπώς είναι διαθέσιμος. Εάν δεν μπορεί να βρεθεί ένα λογικά επαρκές πλαίσιο, η σχεδιαζόμενη μελέτη θα πρέπει να εγκαταλειφθεί. Εάν το πλαίσιο είναι διάφορο από τον πληθυσμό στόχο, αλλά το πρώτο μπορεί να θεωρηθεί ότι περιέχει επαρκείς πληροφορίες, η μελέτη μπορεί να συνεχισθεί. Δεν πρέπει όμως να μας διαφεύγει ότι τα στατιστικά συμπεράσματα, που θα γίνουν επαγωγικά, θα αφορούν στον δειγματοληπτούμενο πληθυσμό. Δηλαδή, οι στατιστικές μέθοδοι μπορούν να εφαρμοσθούν για να διαμορφώσουμε συμπεράσματα, που αφορούν στο δειγματοληπτούμενο πληθυσμό και ότι αυτά τα συμπεράσματα δεν μπορούν να εφαρμοσθούν εξ' ίσου καλά στον πληθυσμό στόχο, εκτός εάν οι δύο πληθυσμοί συμβαίνουν να συμπίπτουν.

Το δειγματοληπτικό πλαίσιο αποτελεί την βάση προϋπόθεση για την επιτυχία μιας δειγματοληπτικής έρευνας, δεδομένου ότι πρέπει να περιέχει όλο το δειγματοληπτούμενο πληθυσμό, ώστε να ικανοποιεί την υπόθεση, «ότι κάθε δειγματοληπτικά μονάδα έχει την ίδια ευκαιρία επιλογής κατά την δειγματοληψία», δηλαδή δεν υπάρχουν παραλείψεις δειγματοληπτικών μονάδων και διπλές καταχωρήσεις στο δειγματοληπτικό πλαίσιο. Σε περίπτωση που υπάρχουν παραλείψεις ή διπλές καταχωρήσεις στο πλαίσιο, το δείγμα που θα επιλεγεί, δεν θα είναι αντιπροσωπευτικό και κατά συνέπεια τα στατιστικά συμπεράσματα που θα γενικευθούν επαγωγικά στο σύνολο του ερευνώμενου πληθυσμού δεν θα είναι αξιόπιστα.

Οι κατάλογοι που χρησιμοποιούνται ως δειγματοληπτικά πλαίσια και περιέχουν τους παρακάτω τύπους λαθών, οδηγούν σε αναξιόπιστα αποτελέσματα :

ο Λανθασμένη πληροφόρηση

Όταν ο κατάλογος – πλαίσιο (δειγματοληπτικό πλαίσιο) δεν είναι πλήρως ενημερωμένος. Π.χ. υπάρχουν στον κατάλογο λανθασμένα ονοματεπώνυμα λόγω μετακίνησης ή θανάτου ορισμένων ατόμων, λανθασμένες διευθύνσεις κατοικιών κ.ο.κ.

ο Ελλιπής πληροφόρηση

Όταν ο κατάλογος – πλαίσιο δεν περιέχει όλες τις δειγματοληπτικές μονάδες. Π.χ. οι εκλογικοί κατάλογοι δεν περιλαμβάνουν άτομα όλων των ηλικιών παρά μόνο αυτούς που ψηφίζουν, επομένως δεν πρέπει να χρησιμοποιούνται ως πλαίσια για έρευνες που αφορούν πληθυσμούς ατόμων όλων των ηλικιών. Επίσης, οι τηλεφωνικοί κατάλογοι αποτελούν πλαίσια με ελλιπή πληροφόρηση, διότι δεν περιλαμβάνονται σ'αυτούς όλα τα άτομα ή νοικοκυριά και ειδικότερα όταν έχουν πολύ χαμηλό εισόδημα .

ο Περισσότερες πληροφορίες από ότι χρειάζεται

Όταν ο κατάλογος – πλαίσιο περιέχει ορισμένες δειγματοληπτικές μονάδες περισσότερο από μια φορά (π.χ. ονόματα ατόμων να παρουσιάζονται δύο φορές σε κάποιο κατάλογο) ή περιέχει δειγματοληπτικές μονάδες που δεν ανήκουν στο δειγματοληπτούμενο πληθυσμό.

Για τον περιορισμό των παραπάνω λαθών είναι απαραίτητο να ενημερώνονται οι κατάλογοι – πλαίσια, πριν χρησιμοποιηθούν για την επιλογή των τελικών μονάδων (άτομα, νοικοκυριά, καταστήματα κ.λ.π.) του δείγματος, ώστε να προκύπτουν αξιόπιστα αποτελέσματα. Έτσι εάν ένα πλαίσιο που ενδέχεται να περιέχει διπλές εγγραφές δεδομένων στοιχειωδών μονάδων μπορεί συνήθως να βελτιωθεί με λεπτομερείς ελέγχους, αν και μια τέτοια διαδικασία μπορεί να αποδειχθεί ανιαρή και δαπανηρή.

Το πλαίσιο εκείνο που δεν περιέχει όλες τις δειγματοληπτικές μονάδες, παρουσιάζει πολλά και σοβαρά προβλήματα. Στην πρώτη θέση προβλημάτων είναι, ότι η ατέλεια του πλαισίου μπορεί δύσκολα να διαπιστωθεί κατά την εξέλιξη της μελέτης. Ένα άλλο ακόμη περισσότερο σοβαρό πρόβλημα είναι, ότι η ατέλεια μπορεί να αφορά την απουσία από τον πλαίσιο μιας ομάδας ή ομάδων στοιχειωδών μονάδων, που χαρακτηρίζονται ενδεχομένως από κάποιες ειδικές ιδιότητες τις οποίες δεν έχουν οι μονάδες που περιέχονται στο πλαίσιο. Τελικά, είναι συνήθως αδύνατο να διορθώσουμε ένα ατελές πλαίσιο, προσθέτοντας τις χαμένες στοιχειώδεις μονάδες, οι οποίες μπορεί να είναι απρόσιτες. Συνεπώς, γνωρίζοντας ότι τα πλαίσια είναι ατελές, δεν παραέχει μια βάση από την οποία θα κρίνουμε εάν η ατέλεια θα αποκλείσει την μελέτη βάση του πλαισίου, όποτε και θα ενεργήσουμε αντίστοιχα.

Η άγνοια ή η σκόπιμη μη αναγνώριση της ατέλειας του πλαισίου, συχνά οδηγεί σε άχρηστες δειγματοληπτικές μελέτες, διότι για τους λόγους που αναφέραμε ανώτερα, πάντοτε θα έχει επιζήμια αποτελέσματα.

2.2.3 Δείγματα .

Ένα δείγμα μπορεί να το ορίσουμε με περισσότερη ακρίβεια στα πλαίσια της διαδικασίας της δειγματοληψίας, διότι μέχρι τώρα ορίζεται σαν ένα μέρος του πληθυσμού.

Οι μονάδες που περιέχονται σε ένα πληθυσμό μπορούν να ταξινομηθούν σε δύο κατηγορίες: - στοιχειώδεις μονάδες δειγματοληψίας (elementary sampling units) και - πρωταρχικές μονάδες δειγματοληψίας (primary sampling units). Οι στοιχειώδεις μονάδες δειγματοληψίας, όπως ήδη γνωρίζουμε, είναι όλες οι μονάδες που περιέχονται στον πληθυσμό των οποίων οι χαρακτηριστικές ιδιότητες μετρούνται ή απαριθμούνται. Πρωταρχικές μονάδες δειγματοληψίας μπορεί να είναι οι στοιχειώδεις μονάδες δειγματοληψίας αυτές καθαυτές ή ομάδες από στοιχειώδεις μονάδες δειγματοληψίας, όπως π.χ. οι διάφοροι κλάδοι της χαλυβουργίας όπου οι στοιχειώδεις μονάδες δειγματοληψίας, που είναι οι «εργαζόμενοι» ομαδοποιούνται κατά κλάδο.

Ένα δείγμα είναι μια συλλογή πρωταρχικών μονάδων δειγματοληψίας, που επιλέγονται έτσι ώστε να αποτελούν ένα αντιπροσωπευτικό μικρόκοσμο από τον οποίο μπορούν να εξαχθούν (επαγωγικά) συμπεράσματα για τον πληθυσμό.

Στην θεωρία της δειγματοληψίας, είναι σημαντικό να διακρίνουμε τα τυχαία δείγματα από τα λογικά δείγματα (ή δείγματα κρίσεως). Ένα τυχαίο δείγμα επιλέγεται δια της μεθόδου πιθανότητας σύμφωνα με την οποία ούτε ο ερευνητής ούτε οι μονάδες δειγματοληψίας μπορούν να αποφασίσουν, ποιες μονάδες του πληθυσμού θα περιληφθούν στο δείγμα. Η επιλογή επιτυγχάνεται με την λειτουργία της τύχης και μόνο. Έτσι σε ένα τυχαίο δείγμα οι πιθανότητες επιλογής είναι γνώστες και για αυτό το δείγμα κατάλληλα ονομάζεται «δείγμα πιθανότητας». Επί πλέον, για ένα δείγμα πιθανότητας, το σφάλμα δειγματοληψίας (που για αυτό θα αναφερθούμε παρακάτω) μπορεί να μετρηθεί και ελεγχθεί δια της θεωρίας των πιθανοτήτων, και έτσι οι μεροληψίες επιλογής, μη ανταποκρίσεως και εκτιμήσεως μπορούν να εξαλειφθούν ή τουλάχιστον να περιέχονται μέσα σε γνωστά όρια. Αντίθετα, ένα λογικά δείγμα, όπως προκύπτει από την ονομασία του, λαμβάνεται σύμφωνα με την προσωπική κρίση του ερευνητή ή οποιουδήποτε άλλου αρμοδίου για την επιλογή του δείγματος. Οι μονάδες που θα περιληφθούν σε ένα δείγμα κρίσεως είναι αποτέλεσμα εμπειρικής κρίσεως του ερευνητή ως προς την αντιπροσωπευτικότητά τους. Συνεπώς, η πιθανότητα με την οποία κάθε μονάδα του πληθυσμού, μπορεί να περιληφθεί στο δείγμα είναι άγνωστη και η αξιοπιστία των αποτελεσμάτων δεν μπορεί να ελεγχθεί με την θεωρία πιθανοτήτων, αλλά εξαρτάται μόνο από την προσωπική κρίση.

Εκείνο που πρέπει να αναφέρουμε εδώ είναι ότι ένα δείγμα, που περιέχει μόνο μια δειγματοληπτική μονάδα μπορεί συχνά να παράγει χρήσιμη πληροφορία αλλά μετρήσεις για την εξαγωγή στατιστικών συμπερασμάτων δεν μπορούν να εφαρμοσθούν στα αποτελέσματά του. Ένα δείγμα το οποίο περιέχει όλες τις μονάδες του πληθυσμού ονομάζεται

«εξαντλητική δειγματοληψία» (exhaustive sampling), που είναι γνωστή με την ονομασία απογραφή.

2.2.4 Τυχαία και μη τυχαία δειγματοληψία

Το πόσο τυχαίο είναι το δείγμα του ερευνώμενου πληθυσμού, χαρακτηρίζει τη δειγματοληψία, σε τυχαία και μη τυχαία.

Στην τυχαία δειγματοληψία (random sampling) η επιλογή των δειγματοληπτικών μονάδων γίνεται κατά τυχαίο τρόπο, δηλαδή όλες οι μονάδες του ερευνώμενου πληθυσμού έχουν την ίδια ευκαιρία να συμπεριληφθούν στο δείγμα.

Η τυχαία δειγματοληψία συνήθως αναφέρεται και σαν δειγματοληψία πιθανότητας, διότι εάν η δειγματοληπτική διαδικασία είναι τυχαία, οι νομοί πιθανότητας μπορούν να εφαρμοσθούν συνεπώς το υπόδειγμα της κατανομής δειγματοληψίας, που είναι απαραίτητο για να ερμηνεύσουμε τα αποτελέσματα του δείγματος και να προβούμε σε εκτιμήσεις, μας είναι διαθέσιμο. Ο όρος τυχαίο δείγμα δεν χρησιμοποιείται για να περιγράψει τα δεδομένα του δείγματος αλλά την διαδικασία που εφαρμόστηκε για την επιλογή του δείγματος. Συνεπώς, η «τυχειότητα» είναι μια ιδιότητα που χαρακτηρίζει την δειγματοληπτική διαδικασία και όχι ένα συγκεκριμένο δείγμα. Όπως μπορεί να δειχθεί, η «τυχειότητα είναι δυνατόν να εξασφαλισθεί με πολλούς τρόπους σε μια δειγματοληπτική διαδικασία και αυτό γιατί τυχαία δείγματα μπορεί να έχουμε πολλών ειδών.

Στην μη τυχαία δειγματοληψία (nonrandom sampling) η επιλογή των δειγματοληπτικών μονάδων γίνεται κατά τρόπο μη τυχαίο, δηλαδή όλες οι μονάδες του ερευνώμενου πληθυσμού δεν έχουν την ίδια ευκαιρία να συμπεριληφθούν στο δείγμα. Με άλλα λόγια η επιλογή του δείγματος γίνεται με υποκειμενικά κριτήρια.

Τα αποτελέσματα που προκύπτουν από την μη τυχαία δειγματοληψία δεν είναι αντιπροσωπευτικά και επομένως τα στατιστικά συμπεράσματα δεν μπορεί να γενικευθούν στο σύνολο του ερευνώμενου πληθυσμού με αξιοπιστία.

2.2.5 δειγματοληπτικά σφάλματα

Ο όρος «σφάλμα» (error) μιας δειγματοληπτικής συναρτήσεως σημαίνει την διαφορά μεταξύ της τιμής της δειγματικής συναρτήσεως και της τιμής της αντίστοιχης παραμέτρου στον πληθυσμό. Ο συνδυασμός διαφόρων αιτιών, δημιουργεί αποκλίσεις των δειγματικών συναρτήσεων από τις αντίστοιχες παραμέτρους και τα σφάλματα (σύμφωνα με τις διάφορες αιτίες) ταξινομούνται σε δειγματοληπτικά και μη δειγματοληπτικά σφάλματα.

Ο όρος «μη δειγματοληπτικό σφάλμα» (non sampling error) προέρχεται από το γεγονός ότι αυτός ο τύπος σφάλματος μπορεί να συμβεί σε οποιαδήποτε έρευνα, είτε αυτή γίνεται με την μέθοδο της απογραφής είτε με την μέθοδο της δειγματοληψίας. Τα μη δειγματοληπτικά σφάλματα περιέχουν μεροληψίες και λάθη. Μια μεροληψία, μπορούμε να πούμε, ότι υπάρχει όταν η τιμή μιας δειγματικής συναρτήσεως παρουσιάζει μια σταθερή τάση αποκλίσεως προς μια κατεύθυνση, από την πραγματική τιμή της παραμέτρου. Συνεπώς, εάν αυτός ο τύπος αυτός σφάλματος παρουσιασθεί, τα δειγματικά αποτελέσματα θα βρίσκονται πάντοτε προς τα πάνω ή πάντοτε προς τα κάτω της τιμής της παραμέτρου και για αυτό δεν μπορούν να αλληλοεξουδετερωθούν τα σφάλματα δια του μέσου όρου.

Οι σπουδαιότεροι παράγοντες οι οποίοι προκαλούν μη δειγματοληπτικά σφάλματα είναι οι εξής :

□ Ακαταλληλότητα του ερωτηματολογίου

Εάν το ερωτηματολόγιο είναι ακατάλληλο, δηλαδή περιέχει ερωτήματα πολύπλοκα ή ερωτήματα που θίγουν το ερευνώμενο πρόσωπο, οι απαντήσεις μπορεί να είναι εσφαλμένες ή μεροληπτικές. Επειδή το ερωτηματολόγιο αποτελεί τον πιο βασικό παράγοντα για την επιτυχία μιας έρευνας, θα περιγράψουμε σε επόμενο παράγραφο.

□ Σφάλματα ανταπόκρισης

Ο ανταποκρινόμενος μπορεί να δώσει εσφαλμένη απάντηση, είτε επειδή δεν κατάλαβε την ερώτηση του ερευνητή, είτε επειδή δεν θυμάται κάποιο γεγονός, είτε διότι δεν επιθυμεί να δώσει τη σωστή απάντηση. Π.χ. μπορεί να δηλώσει μικρότερη ηλικία από την πραγματική, να δηλώσει υψηλότερο εισόδημα από το

πραγματικό για να φανεί στον ερευνητή αξιοσέβαστος κ.ο.κ.

Επίσης η απαντήσεις του ερευνώμενου προσώπου επηρεάζονται από τα προσωπικά χαρακτηριστικά του ερευνητή, π.χ. φύλο, ηλικία, μόρφωση κ.λ.π.

□ Σφάλματα ερευνητή

Ο ερευνητής μπορεί να καταχωρήσει στο ερωτηματολόγιο εσφαλμένα μια απάντηση που δόθηκε σωστά από τον ανταποκρινόμενο. Π.χ. ο ανταποκρινόμενος λέει ότι είναι 45 ετών και ο ερευνητής γράφει 48.

□ Σφάλματα επεξεργασίας

Κατά τον στάδιο της επεξεργασίας (κωδικογράφηση ερωτημάτων και εισαγωγή στοιχείων στα μαγνητικά μέσα) γίνονται λάθη από τους υπαλλήλους που ασχολούνται με την εργασία αυτή.

«Σφάλμα δειγματοληψίας» (sampling error), προκύπτει από την τυχαία επιλογή των μονάδων δειγματοληψίας. Αυτός ο τύπος σφάλματος συμβαίνει, τότε, διότι μόνο ένα μέρος του πληθυσμού περιέχεται στο δείγμα. Εάν λάβει χώρα μια απογραφή, το σφάλμα δειγματοληψίας θα πρέπει να αναμένεται να εξαφανισθεί. Με τον όρο «σφάλμα δειγματοληψίας» εννοούμε ακριβώς την διαφορά μεταξύ του δειγματικού αποτελέσματος και εκείνου της απογραφής, όταν βέβαια και τα δύο αποτελέσματα λαμβάνονται δια χρησιμοποίησης των ίδιων διαδικασιών (μεθόδων). Η σπουδαιότερη ιδιότητα των σφαλμάτων δειγματοληψίας είναι ότι αυτά ακολουθούν τυχαίες διακυμάνσεις που τείνουν να αλληλοεξουδετερωθούν όταν γίνονται μέσοι όροι. Εξ' αιτίας αυτού του χαρακτηριστικού γνωρίσματος, ακόμη και αν μια δεδομένη δειγματική εκτίμηση μπορεί να διαφέρει από την πληθυσμιακή παράμετρο, ή ακόμη και αν τα αποτελέσματα διαφόρων δειγμάτων που λαμβάνονται από τον ίδιο πληθυσμό (με την ίδια διαδικασία) μπορούν να διαφέρουν το ένα του άλλου, ο μέσος της κατανομής δειγματοληψίας μιας δειγματικής συναρτήσεως, δηλαδή η αναμενόμενη δειγματοληψίας μιας δειγματικής συναρτήσεως, μπορεί να αναμένεται να ισούται με την αντίστοιχη πληθυσμιακή παράμετρο.

Το άθροισμα των μη δειγματοληπτικών και δειγματοληπτικών σφαλμάτων είναι το συνολικό σφάλμα μιας στατιστικής έρευνας. Ένα βασικό πρόβλημα της δειγματοληψίας είναι να κάνουμε τον συνολικό

σφάλμα τόσο μικρό όσο είναι δυνατόν. Έτσι το μη δειγματοληπτικό σφάλμα θα μπορούσε να μειωθεί σημαντικά με την βοήθεια των εξής :

- με τον ακριβή καθορισμό του ερευνητού πληθυσμού και των χαρακτηριστικών αυτού που πρόκειται να μετρηθούν
- με προσεκτική προετοιμασία του πλαισίου
- με την προεξέταση του ερωτηματολογίου
- με λεπτομερή έλεγχο σε όλα τα στάδια επεξεργασίας των δεδομένων
- και το σπουδαιότερο από όλα, με πλήρη εκπαίδευση των ερευνητών κ.λ.π.

2.3 Λόγοι χρησιμοποίησης της δειγματοληψίας

Μετά από την σημαντική πρόοδο που σημειώθηκε στην θεωρία της δειγματοληψίας κατά την διάρκεια των τελευταίων δεκαετιών, είναι πλέον δυνατή η μέτρηση χαρακτηριστικών ιδιοτήτων μαζικών δεδομένων, με δεδομένη ακρίβεια, με βάση τα δείγματα. Κατά συνέπεια, σήμερα σχεδόν όλες οι στατιστικές έρευνες, (είτε αυτές γίνονται προκειμένου να ληφθούν αποφάσεις, είτε για την ανάπτυξη των κοινωνικών ή οικονομικών θεωριών) είναι δειγματοληπτικές. Εκτός του ότι, μέσω των μεθόδων της δειγματοληψίας, λαμβάνονται αξιόπιστα αποτελέσματα, υπάρχουν και άλλοι σπουδαίοι λόγοι για την ευρεία παραδοχή της δειγματοληψίας.

Κατά πρώτο, οι υπό έρευνα πληθυσμοί μπορεί να είναι άπειροι, οπότε, σ' αυτές τις περιπτώσεις η δειγματοληψία είναι η μόνη εφικτή μέθοδος. Επί πλέον, ακόμη και στην περίπτωση ενός πεπερασμένου πληθυσμού, πολύ συχνά διαπιστώνεται ότι η δειγματοληψία είναι η μόνη πρακτικά εφαρμόσιμη μέθοδος. Και αυτό συμβαίνει, διότι ενός πεπερασμένου πληθυσμός μπορεί να αποτελείται από δεκάδες και χιλιάδες ή ακόμη και εκατομμύρια μονάδων και έτσι η πλήρης μελέτη αυτού γίνεται πρακτικά αδύνατη.

Επίσης, η μέτρηση των μονάδων ενός πληθυσμού ως προς το εξεταζόμενο χαρακτηριστικό, συχνά προκαλεί την καταστροφή αυτών. Για παράδειγμα, εάν ο παραγωγός επιθυμεί να διαπιστώσει, εάν η δύναμη αντοχής μιας ποσότητας χαλύβδινων συρμάτων ανταποκρίνεται στις ποιοτικές προδιαγραφές, θα πρέπει για τον έλεγχο αυτό να ασκηθεί πίεση επί των χαλύβδινων συρμάτων, η οποία θα προκαλέσει την καταστροφή αυτών. Στην περίπτωση αυτή, μια απογραφή συνεπάγεται την καταστροφή όλων των συρμάτων, με το τέλος του έλεγχου. Το ίδιο

αποτέλεσμα θα έχουμε και στην περίπτωση δοκιμής ηλεκτρικών ασφαλειών ή κατευθυνόμενων βλημάτων.

Αυτό που δεν πρέπει να ξεχνάμε επίσης είναι ότι για πολλούς τύπους δεδομένων ο πληθυσμός δεν είναι προσιτός. Πραγματικά, οφείλουμε να ασχοληθούμε με οποιοδήποτε τμήμα των δεδομένων είναι διαθέσιμο. Για παράδειγμα, στην ανάλυση χρονολογικών σειρών, οι μελέτες γίνονται αναπόφευκτα από δείγματα, διότι μόνο το πρόσφατο παρελθόν παρέχει διαθέσιμα αξιόπιστα στοιχεία.

Τέλος, ακόμη και όταν είναι οικονομικά και πρακτικά δυνατή η μελέτη ολόκληρου του πληθυσμού, η δειγματοληψία είναι περισσότερο αποτελεσματική διαδικασία. Τα αποτελέσματα που θα λάβουμε από την μελέτη των δειγμάτων μπορεί να είναι ακριβή όπως τα αποτελέσματα της απογραφής ή ακόμη και περισσότερο ακριβή από αυτά.

2.4 Πλεονεκτήματα και μειονεκτήματα της δειγματοληψίας

Οι δειγματοληπτικές έρευνες έχουν τα παρακάτω πλεονεκτήματα σε σύγκριση με την απογραφή :

1. *Εξοικονόμηση χρόνου στην συλλογή και επεξεργασία των στατιστικών στοιχείων.*

Η συλλογή και επεξεργασία στατιστικών στοιχείων γίνεται πιο γρήγορα με την δειγματοληψία παρά με την καθολική έρευνα στατιστικών μονάδων του ερευνώμενου πληθυσμού. Π.χ η συμπλήρωση και επεξεργασία των 500 ερωτηματολογίων γίνεται γρηγορότερα από ότι θα γινόταν σε 5.000 ερωτηματολόγια. Αυτό έχει ιδιαίτερη σημασία όταν πρόκειται για συγκέντρωση πληροφοριών που επείγον.

2. *Χαμηλότερο κόστος*

Κύριος και αντικειμενικός σκοπός κάθε δειγματοληπτικής έρευνας είναι η συλλογή πληροφοριών με την μεγαλύτερη δυνατή ακρίβεια και με το ελάχιστο κόστος. Είναι φανερό ότι κοστίζει λιγότερο η συγκέντρωση και επεξεργασία πληροφοριών από μερικές μόνο μονάδες του ερευνώμενου πληθυσμού παρά από τον σύνολο αυτού.

3. Μεγαλύτερη ακρίβεια

Στην δειγματοληπτική έρευνα, είναι δυνατόν, όταν ο αριθμός των στατιστικών μονάδων του πληθυσμού είναι μικρός, να αφιερωθεί περισσότερος χρόνος, μεγαλύτερη προσοχή στις συνεντεύξεις που παίρνουμε για να συμπληρώσουμε το ερωτηματολόγιο, να γίνει καλύτερη εκπαίδευση και επιβλέψει στους ερευνητές, ώστε να έχουμε μεγαλύτερη ακρίβεια στις πληροφορίες.

Το γεγονός αυτό έχει μεγάλη σημασία, γιατί μια δειγματοληπτική έρευνα που γίνεται με σωστό τρόπο μπορεί να δώσει αποτελέσματα, παρά τα δειγματοληπτικά σφάλματα, που να έχουν μεγαλύτερη ακρίβεια και από εκείνα της καθολικής έρευνας.

4. Μεγαλύτερη ευχέρεια εφαρμογής.

Η δειγματοληπτική έρευνα εφαρμόζεται σε εκείνες τις περιπτώσεις, που η απογραφή είναι δυνατή αλλά παράλογη. Π.χ. εάν ένας γιατρός θέλει να προσδιορίσει τα ερυθρά αιμοσφαίρια ενός ασθενούς. Για να πετύχει αυτό φυσικά δεν θα πάρει όλη την ποσότητα αίματος του ασθενούς αλλά λίγα μόνο γραμμάρια.

5. Ολοκληρωτική αδυναμία εφαρμογής της απογραφής.

Η δειγματοληπτική έρευνα μπορεί να εφαρμοστεί σε ορισμένες περιπτώσεις όπου η καθολική έρευνα είναι αδύνατη, επειδή ο πληθυσμός είναι άγνωστος. Π.χ. για να ερευνήσουμε την σύνθεση μιας θαλάσσιας περιοχής ως προς την υπέρξη ψαριών, όπου ο πληθυσμός είναι άγνωστος, πρέπει να περιοριστούμε σε δειγματοληπτική έρευνα.

Τα μειονεκτήματα της δειγματοληψίας είναι τα ακόλουθα :

- a. Αν οι μονάδες του πληθυσμού που ερευνούμε εμφανίζονται πολύ σπάνια, τότε πρέπει να πάρουμε ένα αρκετά μεγάλο δείγμα, ώστε σ' αυτό να περιλαμβάνεται ικανοποιητικός αριθμός από την κατηγορία των μονάδων που θέλουμε να μελετήσουμε, για να πετύχουμε αξιόπιστες εκτιμήσεις. Σε αντίθετη περίπτωση δεν πρέπει να διενεργηθεί δειγματοληπτική έρευνα. Π.χ. αν θέλουμε

να μελετήσουμε το ποσοστό των καπνιστών από άτομα μιας πόλης που έχουν ηλικία πάνω των 80 ετών, πρέπει να πάρουμε ένα αρκετά μεγάλο δείγμα νοικοκυριών, ώστε να περιλαμβάνεται σ' αυτό ικανοποιητικός αριθμός ατόμων ηλικίας πάνω των 80 ετών, γιατί τα άτομα της ηλικίας αυτής αποτελούν μικρό αριθμό ποσοστό στο συνολικό πληθυσμό της πόλης. Αν πάρουμε μικρό αριθμό νοικοκυριών ηλικιωμένων ατόμων.

- b. Αν ο ερευνώμενος πληθυσμός παρουσιάζει μεγάλη ανομοιογένεια, τότε πρέπει να χωρισθεί σε ομοιογενείς ομάδες και να πάρουμε δείγμα μέσα από κάθε ομάδα. Σε αντίθετη περίπτωση οι εκτιμήσεις δεν θα είναι αξιόπιστες. Π.χ. αν θέλουμε να εκτιμήσουμε τον αριθμό των ελαιόδένδρων που υπάρχουν σε μια περιοχή, με τη βοήθεια της δειγματοληπτικής έρευνας, πρέπει να χωρίσουμε τους κατόχους ελαιόδένδρων κατά τάξεις μεγέθους αριθμού δένδρων και να πάρουμε δείγμα μέσα από κάθε τάξη, ώστε να συμπεριληφθούν σ' αυτό κάτοχοι ελαιόδένδρων από όλες τις τάξεις.
- c. Αν δεν ακολουθείται αυστηρά η θεωρητική διαδικασία για τον προσδιορισμό του μεγέθους του δείγματος και την εφαρμογή της κατάλληλης μεθόδου διενέργειας της δειγματοληψίας, τα αποτελέσματα που θα προκύψουν από την δειγματοληπτική έρευνα θα είναι αναξιόπιστα, επομένως θα οδηγούν σε λανθασμένα συμπεράσματα για τον ερευνώμενο πληθυσμό.

2.5 Σταδία δειγματοληψίας .

Η θεωρία δειγματοληψίας μπορεί να χωρισθεί (χονδρικά) σε δύο αλληλοεπιδρώντα στάδια (διαδικασίες) :

- Στην διαδικασία σχεδιασμού ή της επιλογής του δείγματος και τον σχεδιασμό των εκτιμητών που μας δίδει αυτή η δειγματοληψία.
- Στην διαδικασία ανάλυσης και εξαγωγής συμπερασμάτων (δηλαδή με την επέκταση της εκτίμησης στον υπολογισμό διαστημάτων

εμπιστοσύνης με την μελέτη των κατανομών των εκτιμητών [δειγματικών κατανομών])

2.5.1 Το στάδιο του σχεδιασμού της δειγματοληψίας

Κατά τον σχεδιασμό μιας έρευνας (καθολικής ή δειγματοληπτικής) η κατάρτιση του ερωτηματολογίου αποτελεί το πρώτο βήμα στην διαδικασία συλλογής στατιστικών στοιχείων.

Το ερωτηματολόγιο αποτελεί μέσο συλλογής των στατιστικών στοιχείων κατά την διεξαγωγή μιας έρευνας. Το πρώτο βήμα κατά τον σχεδιασμό του ερωτηματολογίου είναι να αποφασισθεί το είδος και ο αριθμός των ερωτήσεων τα οποία πρέπει να καλύπτουν τα χαρακτηριστικά που παρουσιάζουν ενδιαφέρον. Ειδικότερα κατά τον σχεδιασμό του ερωτηματολογίου πρέπει να λαμβάνονται υπόψη κάποιες αρχές. Ορισμένα από τα οποία είναι :

- τα ερωτηματολόγια δεν πρέπει να είναι μακροσκελή ερωτηματολόγια γιατί γίνονται κουραστικά
- Σωστή διατύπωση και σειρά των ερωτημάτων
- Η σύνταξη των ερωτημάτων θα πρέπει να γίνει ανάλογα με το εκπαιδευτικό επίπεδο των ερευνητών. κ.λ.π.

Η συμπλήρωση του ερωτηματολογίου γίνεται :

- είτε με προσωπική συνέντευξη (personal interview), η οποία είναι ο καλύτερος τρόπος συλλογής στατιστικών στοιχείων που χρησιμοποιείται στις δειγματοληπτικές έρευνες . Διότι :
 - συγκεντρώνονται στοιχεία, ποιοτικά καλύτερα από ότι με την μέθοδο της ταχυδρομικής αποστολής, λόγω στο ότι δίνονται διευκρινήσεις στους ερευνώμενους, έτσι ώστε να δίνουν σωστές απαντήσεις
 - έχουμε μεγαλύτερο ποσοστό ανταπόκρισης των ερευνώμενων
 - και τα στοιχεία μπορούν να συγκεντρωθούν απ' ευθείας με κομπιούτερ. Με τον τρόπο αυτό κερδίζουμε χρόνο και επιτυγχάνεται καλύτερη ποιότητα στοιχείων, διότι περιορίζονται τα σφάλματα επεξεργασίας.

- είτε με ταχυδρομική αποστολή του ερωτηματολογίου. Η μέθοδος αυτός, αποστολής του ερωτηματολογίου στον ερυνόμενο, η συμπλήρωση από αυτόν και η επιστροφή του με το ταχυδρομείο, θεωρείται ο πιο εύκολος τρόπος συλλογής στατιστικών στοιχείων. Η εφαρμογή της μεθόδου αυτού είναι αρκετά περιορισμένη λόγω των μειονεκτημάτων (όπως μόνο ένα μικρό ποσοστό απαντά και τα μη δειγματοληπτικά σφάλματα κατά κανόνα είναι σημαντικά) που παρουσιάζει.

2.6 Στόχοι

Και στα δύο στάδια στο οποίο χωρίζεται η θεωρία της δειγματοληψίας (όπως αναφέραμε και παραπάνω) υπάρχουν ορισμένοι στόχοι. Μεταξύ των κυριότερων στόχων και στα δύο στάδια είναι τα εξής :

- Ελάττωση της διασποράς των εκτιμητών. Με τον όρο διασπορά ή διακύμανση εννοούμε εκείνο τον παράμετρο που μας πληροφορεί αν οι τιμές των παρατηρήσεων είναι συγκεντρωμένες ή διασκορπισμένες σε σχέση με τον μέσα αριθμητικό. Υπάρχουν τρεις παράγοντες που συμβάλουν στην ελάττωση της διασποράς τα οποία είναι τα εξής :
 - η μέθοδος δειγματοληψίας που θα ακολουθηθεί
 - η εκλογή του εκτιμητού
 - και το μέγεθος του δείγματος
- Ελάττωση ή εξάλειψη της μεροληψίας στους εκτίμητες είτε τεχνική είτε οφειλομένη στις μετρήσεις ή στην διατάραξη των πιθανοτήτων επιλογής. Για την μεροληψία τα πράγματα είναι δυσκολότερα εκτός εάν είναι μαθηματική (δηλαδή ζήτημα του σταθερού πολλαπλασιαστή) ή οφείλεται σε συστηματικά λάθη μετρήσεως. Πολλές φορές μεροληπτικοί εκτιμητές είναι πολύ χρήσιμοι όταν έχουν την ιδιότητα της «ευσταθείας» στην παρουσία ακραίων τιμών ή μη – κανονικότητας.

2.6.1 Η ακρίβεια και η μεροληψία των δειγματικών εκτιμήσεων

Τα σφάλματα δειγματοληψίας ή οι διακυμάνσεις της δειγματοληψίας μετρούνται με αυτό που ονομάζεται **ακρίβεια** (precision). Η ακρίβεια, η οποία επίσης δηλώνεται με τον όρο αξιοπιστία, των δειγματικών αποτελεσμάτων είναι ο βαθμός με την οποία διαδοχικές δειγματικές συναρτήσεις που προκύπτουν από διαδοχικά τυχαία δείγματα του ίδιου μεγέθους διαφέρουν μεταξύ τους. Συνεπώς, η ακρίβεια δεν είναι παρά το μέγεθος του τυπικού σφάλματος της δειγματοληπτικής συναρτήσεως. Όσο μικρότερο είναι το τυπικό σφάλμα, τόσο μεγαλύτερη είναι η ακρίβεια (ή αξιοπιστία) της εκτιμήσεως.

Ακρίβεια μιας δειγματικής εκτιμήσεως, είναι η διαφορά μεταξύ της δειγματικής συναρτήσεως του δείγματος και της πραγματικής τιμής της παραμέτρου. Η ακρίβεια μιας εκτιμήσεως όμως, εξαρτάται, και από τον βαθμό αξιοπιστίας και από την απουσία της μεροληψίας. Οι μεροληψίες, μπορεί να σημειωθεί, δεν επηρεάζουν απαραίτητα την αξιοπιστία εάν οι πηγές των μεροληψιών είναι σταθερές και ως εκ τούτου τείνουν να δημιουργήσουν αναπαραγωγικότητα των εκτιμήσεων που γίνονται από διαδοχικά δείγματα. Δηλαδή, μια μεροληπτική διαδικασία μπορεί πράγματι να έχει ένα μικρότερο τυπικό σφάλμα από μια αμερόληπτη διαδικασία. Συνεπώς, οι ευκαιρίες του να πάρουμε μια περισσότερη ακριβή εκτίμηση μπορεί να είναι περισσότερες με μια μεροληπτική εκτίμηση απ' ό,τι με μια αμερόληπτη εκτίμηση. Αυτό που πρέπει να γίνει σαφές είναι ότι, όταν μια μεροληπτική διαδικασία εφαρμόζεται, πρέπει να γνωρίζουμε σαφώς την φύση της μεροληψίας και την ερμηνεία των αποτελεσμάτων αντίστοιχα στις διαδικασίες λήψεως αποφάσεως. Το ότι ένα μεροληπτικό δείγμα μπορεί πράγματι να προκαλέσει μια περισσότερο ακριβή εκτίμηση είναι ένα από τους λόγους, που επιβάλλουν την χρησιμοποίηση άλλων σχεδίων δειγματοληψίας, εκτός από τυχαία δειγματοληψία.

Ο μέσος των σφαλμάτων της εκτιμήσεως, λαμβάνοντας υπόψη τα πρόσημά τους ονομάζεται «μεροληψία» της εκτιμήσεως [ή γενικά του σχεδίου δειγματοληψίας]. Μία θετική μεροληψία υπονοεί ότι το σχέδιο δειγματοληψίας δίνει εκτιμήσεις οι οποίες είναι γενικά πολύ ψηλές. Μια αρνητική μεροληψία, αντιθέτως πολύ χαμηλές. Οι εκτιμήσεις που είναι αμερόληπτες αποτελούν ένα επιθυμητό χαρακτηριστικό γνώρισμα ενός σχεδίου δειγματοληψίας. Εξ' άλλου, ένα σχέδιο δειγματοληψίας που δίνει μια μικρή μεροληψία θα πρέπει να αποκλεισθεί από περαιτέρω εξέταση, διότι μπορεί να έχει άλλα ελκυστικά χαρακτηριστικά. Σαν μέτρο της ακρίβειας του σχεδίου δειγματοληψίας χρησιμοποιούμε το μέσο

τετραγωνικό σφάλμα των εκτιμήσεων, που λαμβάνεται περί την πραγματική τιμή.

Στην θεωρία της δειγματοληψίας είναι απαραίτητο να μελετήσουμε μεροληπτικές εκτιμήσεις για δύο λόγους :

- Σε μερικά από τα πιο γνωστά προβλήματα, ειδικά στην εκτίμηση των λόγων, εκτιμήσεις που κατά τα άλλα είναι εύκολες βρίσκονται να είναι μεροληπτικές .
- Ακόμη και στις εκτιμήσεις οι οποίες είναι αμερόληπτες στην δειγματοληψία πιθανότητας, τα σφάλματα μετρήσεως και μη ανταποκρίσεως μπορεί να παράγουν μεροληψίες στα αριθμητικά αποτελέσματα τα οποία μπορούν να προκύψουν από τα δεδομένα.

2.6.2 Το μέγεθος του δείγματος

Κατά τον σχεδιασμό μιας δειγματοληπτικής έρευνας, μια πρώτη απόφαση που πρέπει να ληφθεί είναι για το μέγεθος του δείγματος, που θα πάρουμε από τον δειγματοληπτούμενο πληθυσμό. Η απόφαση αυτή είναι σπουδαία. Καίτοι μιας ακριβής απάντηση δεν είναι εύκολο να βρεθεί, υπάρχει μια λογική μέθοδος αντιμετώπισεως του προβλήματος.

Προφανώς, επιθυμούμε να αποφύγουμε να πάρουμε δείγμα τόσο μικρό που η εκτίμηση θα είναι πολύ ανακριβής για να χρησιμοποιηθεί. Ισοδύναμα, επιθυμούμε να αποφύγουμε την λήψη ενός δείγματος τόσο μεγάλο, ώστε η εκτίμηση να είναι περισσότερη ακριβής απ' όσο απαιτείται. Συνεπώς, το πρώτο βήμα είναι να αποφασίσουμε πόσο μεγάλο σφάλμα μπορούμε να ανεχθούμε στην εκτίμηση. Αυτό απαιτεί προσεκτική σκέψη περί την χρήση της εκτιμήσεως και περί της συνέπειες ενός αρκετά μεγάλου σφάλματος.

Η απόφαση για το μέγεθος του δείγματος, δεν μπορεί να είναι πάντοτε ικανοποιητική, επειδή συνήθως δεν κατέχουμε αρκετή πληροφόρηση, που να μας καθιστά βέβαιους ότι η επιλογή μας όσον αφορά το μέγεθος του δείγματος είναι η καλύτερη δυνατή. Η θεωρία της δειγματοληψίας παρέχει ένα πλαίσιο μέσα στο οποίο μπορούμε να αντιμετωπίσουμε ικανοποιητικά το πρόβλημα .

2. 7 Μέθοδοι διενέργειας της δειγματοληψίας

Για την διενέργεια των δειγματοληπτικών ερευνών χρησιμοποιούνται οι εξής μέθοδοι :

- Απλή τυχαία δειγματοληψία
- Συστηματική δειγματοληψία
- Στρωματοποιημένη δειγματοληψία
- Δειγματοληψία κατά ομάδες
- Δισταδιακή δειγματοληψία
- Τρισταδιακή δειγματοληψία
- Άλλοι μέθοδοι δειγματοληψίας

Η εφαρμογή της μεθόδου προϋποθέτει την υπάρξει τουλάχιστον ένα από τα παρακάτω κριτήρια :

1. από τον ζητούμενο βαθμό ακρίβειας των αποτελεσμάτων
2. από τα χρονικά και χρηματικά περιθώρια της έρευνας
3. Από την μεταβλητικότητα των μονάδων του ερευνώμενου πληθυσμού
4. από την υπάρξει δειγματοληπτικών πλαισίων των μονάδων του πληθυσμού .
5. από την δυνατότητα που υπάρχει για να διαιρεθεί ο ερευνώμενος πληθυσμός σε υποπληθυσμούς με μεγάλη ομοιογένεια .

ΜΕΡΟΣ ΙΙ

ΚΕΦΑΛΑΙΟ	<u>Απλή τυχαία δειγματοληψία .</u>
1 .	

1.1. Εισαγωγή .

Απλή τυχαία δειγματοληψία (simple random sampling) ονομάζεται η μέθοδος επιλογής n μονάδων από έναν πληθυσμό N μονάδων, κατά τρόπο ώστε κάθε δυνατό δείγμα μεγέθους n έχει την ίδια πιθανότητα να επιλεγεί, δηλαδή όλες οι μονάδες του πληθυσμού να έχουν την ίδια ευκαιρία να συμπεριληφθούν στο δείγμα.

Ένα απλό τυχαίο δείγμα λαμβάνεται με τυχαίο επιλογή, η οποία πρέπει να διαφοροποιείται από την επιλογή «στην τύχη» (στα κουτουρού) ή «χωρίς σχέδιο». Αν δεν προσέξουμε αυτή τη διαφορά, μπορεί να γίνει κακή χρήση τυχαίων μεθόδων και έτσι ένα δείγμα ενώ υποτίθεται ότι είναι τυχαίο, στην πραγματικότητα δεν είναι καθόλου τυχαίο. Συνεπώς, σοβαρά σφάλματα μπορεί να γίνουν κατά την δειγματοληψία. Η τυχαία επιλογή που εξασφαλίζει «τυχειότητα» στην δειγματοληπτική διαδικασία μπορεί να εξασφαλισθεί μόνο με προσεκτική μελέτη και μεθοδική διαδικασία. Για αυτό λέμε ότι για να εξασφαλίσουμε τυχαία επιλογή, η διαδικασία της δειγματοληψίας πρέπει να είναι προσεκτικά ελεγχόμενη.

Θεωρητικά, στην επιλογή ενός απλού τυχαίου δείγματος κάθε μονάδα που επιλέγεται πρέπει να μετρείται, καταγράφεται και επιστρέφεται στον πληθυσμό προτού μια άλλη μονάδα επιλεγεί. Έτσι κάθε μονάδα του πληθυσμού μπορεί να επιλεγεί περισσότερες από μια φορές στο ίδιο δείγμα. Πάντως, στην πράξη, αυτή η διαδικασία σπανία παρατηρείται. Υπάρχουν δυο βασικές αιτίες που επιβάλλουν αυτό. Κατά την πρώτη, εάν ο πληθυσμός είναι μεγάλο σε σύγκριση με το μέγεθος του δείγματος, το σφάλμα που προκύπτει από την μη επιστροφή της μονάδας

είναι χωρίς πρακτική σημασία. Κατά την δεύτερη, σε πολλές περιπτώσεις δειγματοληπτικών ερευνών, οι επί μέρους μονάδες που επιλέγονται καταστρέφονται τελείως κατά τον έλεγχο και ως εκ τούτου είναι αδύνατον να επιστρέψουμε αυτές στον πληθυσμό. Πάντως, εάν ο πληθυσμός είναι πολύ μικρός, η διαδικασία της επιστροφής των ελεγχόμενων μονάδων πρέπει να εφαρμοσθεί.

1.2 . Επιλογή απλού τυχαίου δείγματος .

Προκειμένου να επιλέξουμε ένα δείγμα n μονάδων από έναν πληθυσμό ο οποίος αποτελείται από N μονάδες κατά τρόπο ώστε κάθε δυνατό δείγμα μεγέθους n να έχει την ίδια πιθανότητα επιλογής, πρέπει να επιλέξουμε τυχαίο δείγμα από τον ερυνώμενο πληθυσμό. Η τυχαία επιλογή εξασφαλίζεται με μια από τους παρακάτω μεθόδους :

Μέθοδος της κλήρωσης

Η επιλογή των στατιστικών μονάδων (άτομα, νοικοκυριά, καταστήματα κ.λ.π.) κατά την μέθοδο αυτή γίνεται με κλήρωση. Σύμφωνα με αυτή την διαδικασία, όλες οι μονάδες στον δειγματοληπτούμενο πληθυσμό απαριθμούνται και οι αριθμοί αναγράφονται σε σφαιρίδια, καρτέλες ή άλλα στοιχεία τα οποία φυσικά είναι ομοιόμορφα και ομοιογενή. Τα στοιχεία αυτά τοποθετούνται σε κάλη και αναμιγνύονται καλά. Στη συνέχεια οι συγκεκριμένοι αριθμοί των μονάδων που θα συμπεριληφθούν στο δείγμα επιλέγονται με ή χωρίς επαναθέσει (για αυτό θα αναφερθούμε στο επόμενο παράγραφο). Με επαναθέσει, η πιθανότητα επιλογής κάθε μονάδα είναι ίση με $1/N$. Χωρίς επαναθέσει, η πιθανότητα επιλογής των επόμενων μονάδων αυξάνει καθώς το N μειώνεται με τον αριθμό των μονάδων που ήδη έχουν επιλεγεί. Χωρίς επαναθέσει, π.χ. η πιθανότητα με την οποία κάθε μονάδα μπορεί να επιλεγεί

κατά την πρώτη δοκιμή είναι $1/N$

κατά την δεύτερη δοκιμή είναι $1/(N-1)$

κατά την τρίτη δοκιμή είναι $1/(N-2)$ κ.ο.κ.

Παράδειγμα

Εάν επιθυμούμε να επιλέξουμε ένα τυχαίο δείγμα $\langle n = 4 \rangle$ μονάδων από μια δέσμη 52 παιγνιοχάρτων, θα επιλέξουμε ένα χάρτη, θα καταγράψουμε τα στοιχεία του, θα επιστρέψουμε αυτό στην θέσμια, θα ανακατέψουμε την θέσμια καλά, θα επιλέξουμε πάλι ένα χάρτη, θα καταγράψουμε τα στοιχεία του, θα επιστρέψουμε αυτό στην θέσμια και θα

επαναλάβουμε την διαδικασία με τον ίδιο τρόπο, έως ότου επιλεγούν τα τέσσερα χαρτιά.

Μέθοδος τυχαίων αριθμών

Η διαδικασία της χρησιμοποίησης της κάλπης μπορεί να αντικατασταθεί, και αυτό συμφέρει από απόψεως κόστους και χρόνου, με την διαδικασία επιλογής ενός δείγματος με την χρησιμοποίηση των πινάκων των τυχαίων αριθμών.

Στην πράξη η επιλογή των στατιστικών μονάδων του ερευνώμενου πληθυσμού γίνεται με την χρησιμοποίηση πινάκων τυχαίων αριθμών διότι {εκτός από τους παραπάνω λόγους} ο πληθυσμός N συνήθως είναι μεγάλος αριθμός, οπότε πρακτικά η εφαρμογή της μεθόδου της κλήρωσης είναι πρακτικά δύσκολη και δεν εξασφαλίζεται η τυχαία επιλογή αφού δεν είναι εύκολη η ανάμιξη των καρτών ή των χαρτιών.

Οι πίνακες τυχαίων αριθμών αποτελούνται από σειρές και στήλες των αριθμών 0, 1, 2, ..., 9, οι οποίες έχουν τοποθετηθεί κατά τυχαίο τρόπο. Οι πίνακες αυτοί ελέγχονται ως προς τη συχνότητα των αριθμών 0, 1, 2, ..., 9 με διάφορες στατιστικές μεθόδους, έτσι ώστε να εξασφαλίζεται η επιλογή τυχαίων αριθμών.

Οι πιο γνωστοί πίνακες τυχαίων αριθμών που χρησιμοποιούνται συνήθως είναι των Fisher-Yates, Kendall-Badinton-Smith κ.λ.π. Επίσης χρησιμοποιούνται πίνακες τυχαίων αριθμών που δημιουργούνται με την χρήση ηλεκτρονικών υπολογιστών και ειδικών προγραμμάτων.

Οι πίνακες τυχαίων αριθμών χρησιμοποιούνται ως εξής :

Έστω ότι επιθυμούμε να επιλέξουμε ένα δείγμα n από έναν πληθυσμό με N μονάδες. Αρχικά αριθμούμε τον πληθυσμό από 1 μέχρι N , στη συνέχεια θα επιλέξουμε n αριθμούς οι οποίοι θα είναι μικρότεροι ή ίσοι από το N . Η επιλογή αρχίζει από κάποιον αριθμό στήλης ή σειράς μέχρι να επιλέξουμε είτε οριζόντια είτε κατακόρυφα μέχρι να επιλέξουμε n αριθμούς. Αριθμοί μεγαλύτερη από το N ή αριθμοί που έχουν επιλεγεί μια φορά παραλείπονται.

Παράδειγμα

Έστω ότι επιθυμούμε να επιλέξουμε ένα τυχαίο δείγμα 8% από έναν πληθυσμό με $N = 200$ μονάδες.

Αριθμούμε τις μονάδες από 1 μέχρι 200 και επιλέγουμε 16 μονάδες χρησιμοποιώντας τον πίνακα τυχαίων αριθμών .

Η επιλογή αρχίζει κατά τυχαίο τρόπο από κάποιον αριθμό μιας στήλης ή σειράς. Έστω ότι επιλέγομαι τυχαία την δεύτερη στήλη του πίνακα, που έχει πρώτο αριθμό το 358, επειδή $N = 200$ θα πάρουμε τους αριθμούς οι οποίοι είναι μικρότεροι ή ίσοι από το 200. Αν προχωρήσουμε κατακόρυφα στη δεύτερη στήλη θα επιλεγούν οι αριθμοί :

165, 92, 129, 186, 84, 85, 115, 93, 157, 188, 108

Επειδή δεν υπάρχουν άλλοι αριθμοί μικρότερο από 200, συνεχίζουμε την επιλογή κατά τον ίδιο τρόπο, από την τρίτη στήλη, όποτε θα επιλεγούν οι αριθμοί :

~~89, 17,~~ 181, 48, 117

Οι αριθμοί 358, 681, 515 κ.ο.κ. οι οποίοι είναι μεγαλύτεροι από το 200 παραλείπονται.

Οι 16 αριθμοί που έχουν επιλεγεί με την παραπάνω διαδικασία αποτελούν τις μονάδες του δείγματος που θα ερευνηθούν από τον πληθυσμό N . Οι αριθμοί αυτοί αντιστοιχίζονται με τους A/A του καταλόγου / πλαισίου, οπου είναι καταχωρημένος ο ερευνώμενος πληθυσμός [π.χ. κατάλογος νοικοκυριών, κατάλογος ατόμων, κατάλογος επιχειρήσεων κ.λ.π.] οπότε ο ερευνητής θα ερευνησει, για παράδειγμα, τα νοικοκυριά ή της επιχειρήσεις που έχουν A/A στον κατάλογο 165, 92 κ.ο.κ.

1.3 Τρόποι εκλογής ενός δείγματος .

Η εκλογή του δείγματος γίνεται (συνήθως) παίρνοντας τα στοιχεία του πληθυσμού το ένα κατόπιν του άλλου. Όπως έχουμε αναφέρει τα στοιχεία μπορούν να εκλεγούν χωρίς επαναθέσει (ξανατοποθέτηση) ή με επαναθέσει (ξανατοποθέτηση). Επίσης μπορούν να εκλεγούν κατά τέτοιο τρόπο ώστε είτε όλα τα δυνατά στοιχεία σε δοθέν τράβηγμα να έχουν ίσες (δεσμευμένες) πιθανότητες εκλογής είτε άνισες πιθανότητες εκλογής, εκ των προτέρων καθοριζόμενες από την μέθοδο δειγματοληψίας. Η όλη εκλογή γίνεται με την βοήθεια κατάλογων ή δειγματοληπτικών πλαισίων (frames) του πληθυσμού. Έτσι έχουμε τέσσερους βασικούς τρόπους επιλογής του δείγματος :

**Πιθανότητες εκλογής των δυνατών
στοιχείων σε ένα οποιοδήποτε τράβηγμα**

Ε π α ν α θ έ σ ε ι	Μ ε	Ίσες	Άνισες
		Απλή τυχαία δειγματοληψία με επαναθέσει	Δειγματοληψία με άνισες πιθανότητες (που δεν αλλάζουν από τράβηγμα σε τράβηγμα)
	Χ ω ρ ί ς	Ίσες	Άνισες
		Απλή τυχαία δειγματοληψία χωρίς επαναθέσει [Συστηματική δειγματοληψία]	Δειγματοληψία με άνισες πιθανότητες (που αλλάζουν από τράβηγμα σε τράβηγμα)

1.3.1 Απλή τυχαία δειγματοληψία χωρίς επαναθέσει

Αυτήν εννοούμε συνήθως όταν λέγαμε απλή τυχαία δειγματοληψία. Είναι μια ίσης πιθανότητας επιλογής μέθοδος. Σε αυτήν την μέθοδο κάθε δυνατό υποσύνολο n στοιχείων (μονάδων) από έναν πληθυσμό N μονάδων έχει ίση πιθανότητα να είναι δείγμα μεγέθους n που επιλέγεται. Ίση πιθανότητα εδώ αναφέρεται στο γεγονός ότι η [ολική] πιθανότητα για την μονάδα y_i , για κάθε i να συμπεριληφθεί στο δείγμα, είναι :

$$\frac{(N-1)}{n-1} / \frac{(N)}{n} = n/N$$

= το δειγματοληπτικό κλάσμα

Επίσης η πιθανότητα επιλογής της μονάδος y_i , σε ένα συγκεκριμένο τράβηγμα είναι $1/N$ αλλά η υπό συνθήκη πιθανότητα ότι το i -στο τράβηγμα παράγει y_i γνωρίζοντας ότι ένα προγενέστερο τράβηγμα έδωσε y_j ($i \neq j$), είναι $1 / (N-1)$. Συνεπώς τα τράβηγματα δεν είναι ανεξάρτητα .

1.3.2 Απλή τυχαία δειγματοληψία με επαναθέσει

Επίσης είναι μια ίσης πιθανότητας επιλογής μέθοδος. Η προηγούμενη μέθοδος μπορεί να γίνει με την μέθοδο αυτή αρκεί να αποβάλουμε τις

επαναλήψεις μέχρι ότου η διαφορετικές μονάδες έχουν επιλεγεί. Εδώ οι τυχαίες μεταβλητές του δείγματος y_1, y_2, \dots, y_n δεν είναι μόνο ισόνομες όπως και στην προηγούμενη μέθοδο [πιθανότητα να λάβουμε μια τιμή από τις δείγματος Y_1, Y_2, \dots, Y_N είναι $1/N$] αλλά επί πλέον και ανεξάρτητες. Θα δείξουμε στην συνέχεια ότι η διασπορά του δειγμάτικου μέσου στην μέθοδο αυτή είναι μεγαλύτερη από εκείνη της προηγούμενης μεθόδου. Για αυτό απλή τυχαία δειγματοληψία χωρίς επαναθέσει είναι προτιμότερη. Όμως δεν υπάρχει αισθητή διαφορά όταν το δειγματοληπτικό κλάσμα n/N είναι αμελητέα μικρό εν σύγκριση με την μονάδα δηλαδή $1 - n/N \approx 1$

1.4 Συμβολισμοί.

Σε μια δειγματοληπτική έρευνα οι ιδιότητες των μονάδων δειγματοληψίας, που επιθυμούμε να μετρήσουμε και να καταγράψουμε ονομάζονται «χαρακτηριστικές ιδιότητες».

Οι τιμές που λαμβάνονται για οποιαδήποτε χαρακτηριστική ιδιότητα των N μονάδων που συνιστούν τον πληθυσμό συμβολίζονται με : y_1, y_2, \dots, y_n , ή με : y_i ($i = 1, 2, \dots, n$).

Τα κεφαλαία γράμματα αναφέρονται σε χαρακτηριστικές ιδιότητες του πληθυσμού και τα μικρά γράμματα στις αντίστοιχες του δείγματος. Έτσι έχουμε :

ΠΙΝΑΚΑ

Ονομασία	Πληθυσμός	Δείγμα
Μονάδες	y_1, y_2, \dots, y_N	y_1, y_2, \dots, y_n
Μέσος	$\bar{Y} = \frac{1}{N} \sum_{i=1}^N y_i$	$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$
Ολικό	$\dot{Y} = \sum_{i=1}^N y_i = N\bar{Y}$	$\dot{y} = \sum_{i=1}^n y_i = n\bar{y}$

Διασπορά	$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \dot{Y})^2$ $S^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \dot{Y})^2$	$S^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \dot{y})^2$
Λόγος για δύο χαρακτηριστικά	$R = Y / X = \dot{Y} / X$	$R = y / x = \dot{y} / x$
Ποσοστό	# ατόμων στον πληθυσμό που έχουν κάποια <u>χαρακτηριστική ιδιότητα</u> $P = \frac{\quad}{N}$	# ατόμων στον πληθυσμό που έχουν κάποια <u>χαρακτηριστική ιδιότητα</u> $P = \frac{\quad}{n}$

Το σύμβολο $\dot{\quad}$ δείχνει εκτίμηση ενός χαρακτηριστικού του ερευνώμενου πληθυσμού που προκύπτει από ένα δείγμα. Οι πιο απλές εκτιμήσεις είναι :

ΕΚΤΙΜΗΣΕΙΣ

Πληθυσμιακός μέσος \dot{Y} : $\hat{Y} = \dot{y} = \text{δειγματικός μέσος}$
 Πληθυσμιακό σύνολο Y : $\dot{Y} = N\dot{y} = \sum_{i=1}^n y_i = \dot{y} / n$

τα σύμβολα $V(\dots)$ και $SE(\dots)$ χρησιμοποιούνται αντίστοιχα για την διακύμανση και το τυπικό σφάλμα (standard error) των εκτιμήσεων.

Τέλος ο λόγος n / N ονομάζεται «κλάσμα δειγματοληψίας» (sampling fraction) και συμβολίζεται με το γράμμα f .

1.5. Εκτιμήσεις και σφάλματα αυτών.

Η μέση τιμή του δείγματος \dot{y} είναι μια τυχαία μεταβλητή της κατανομής δειγματοληψίας των μέσων \dot{y} , η οποία έχει μέσο όρο εκτιμήσεων ίσο με τον μέσο όρο του πληθυσμού.

$$E(\hat{y}) = \bar{Y}$$

Δηλαδή ο μέσος όρος \hat{y} του δείγματος αποτελεί αμερόληπτη εκτίμηση του μέσου \bar{Y} του πληθυσμού.

Η διακύμανση της y_i σε ένα πεπερασμένο πληθυσμό ορίζεται ως :

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{Y})^2$$

Μια άλλη ισοδύναμο έκφραση της διακύμανσης αυτής για το N σχετικά μεγάλο είναι :

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})^2$$

ΘΕΩΡΗΜΑ

Η διακύμανση του μέσου \hat{y} από ένα απλό τυχαίο δείγμα είναι :

$$V(\hat{y}) = E(\hat{y} - \bar{Y})^2 = \frac{S^2}{n} \frac{(N-n)}{n} = \frac{S^2}{n} (1-f)$$

Όπου $f = n/N$ είναι το κλάσμα δειγματοληψίας.

ΠΟΡΙΣΜΑ 1

Η τυπική απόκλιση ή τυπικό σφάλμα (standard error) είναι :

$$SE(\hat{y}) = \sqrt{V(\hat{y})} = \frac{S}{\sqrt{n}} \sqrt{\frac{(N-n)}{n}} = \frac{S}{\sqrt{n}} \sqrt{(1-f)}$$

Η τυπική απόκλιση εκφράζει την διασπορά του πλήθους όλων των δυνατών εκτιμήσεων \hat{y} από το μέσο του πληθυσμού \bar{Y} . Όσο μικρότερη είναι η τυπική απόκλιση τόσο μικρότερο θα είναι και το σφάλμα $\hat{y} - \bar{Y}$ στην εκτίμηση του μέσου.

ΠΟΡΙΣΜΑ 2

Η διακύμανση της εκτίμησης του συνόλου του πληθυσμού ($\hat{Y} = N\hat{y}$) είναι:

$$V(\hat{Y}) = E(\hat{Y} - Y)^2 = \frac{N^2 S^2}{n} \frac{(N - n)}{N} = \frac{N^2 S^2}{n} (1 - f)$$

Το τυπικό σφάλμα είναι :

$$SE(\hat{Y}) = \sqrt{V(\hat{Y})} = \frac{\sqrt{N^2 S^2} \sqrt{(N - n)}}{\sqrt{n}} = \frac{NS}{\sqrt{n}} \sqrt{(1 - f)}$$

Σημείωση: τον παραπάνω τύπο μπορεί το βρούμε και αλλιώς δηλαδή αντί S να έχουμε το σ . (όμως είτε το S έχουμε, είτε το σ έχουμε το ίδιο πράγμα είναι).

Ο συντελεστής $(N - n) / N$ ονομάζεται **διόρθωση πεπερασμένου πληθυσμού** (finite population correction). Ο συντελεστής αυτός μπορεί να παραληφθεί όταν το κλάσμα δειγματοληψίας είναι πολύ μικρό, όποτε ο συντελεστής $(N - n) / N = 1 - n / N$ πλησιάζει την μονάδα. Αυτό συμβαίνει όταν ο πληθυσμός είναι αρκετά μεγάλος ή όταν πρόκειται για άπειρους πληθυσμούς δηλαδή σε περίπτωση δειγματοληψίας με επανατοποθέτηση.

Στην πράξη η διόρθωση πεπερασμένου πληθυσμού παραλείπεται όταν το κλάσμα δειγματοληψίας δεν ξεπερνά το 5% .

1.6 Εκτίμηση του τυπικού σφάλματος από ένα δείγμα .

Στην πράξη η διακύμανση του πληθυσμού σ^2 είναι άγνωστη, οπότε για την εκτίμηση των τυπικών σφαλμάτων χρησιμοποιείται η αμερόληπτη εκτίμηση s^2 της S^2 η οποία προκύπτει από τα στοιχεία του δείγματος δηλαδή η διακύμανση :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

Η οποία είναι μια αμερόληπτη εκτίμηση [$E(s^2) = S^2$]

Επομένως στην περίπτωση της εκτίμησης του τυπικού σφάλματος του δείγματος θα έχουμε :

Εκτίμηση διακύμανσης

$$V(\bar{y}) = \frac{s^2}{n} (N - n) = \frac{s^2}{n} (1 - f)$$

Και εκτίμηση τυπικού σφάλματος

$$SE(\bar{y}) = \sqrt{V(\bar{y})} = \frac{s}{\sqrt{n}} \sqrt{N - n} = \frac{s}{\sqrt{n}} \sqrt{1 - f}$$

Στην περίπτωση εκτίμησης του συνόλου του πληθυσμού ($\hat{Y} = N\bar{y}$) θα έχουμε :

Εκτίμηση διακύμανσης

$$V(N\bar{y}) = \frac{N^2 s^2}{n} (N - n) = \frac{N^2 s^2}{n} (1 - f)$$

Και εκτίμηση τυπικού σφάλματος

$$SE(N\bar{y}) = \sqrt{V(N\bar{y})} = \frac{Ns}{\sqrt{n}} \sqrt{N - n}$$

Τα τυπικά σφάλματα εκτιμήσεων που περιγράψαμε παραπάνω εκφράζουν απόλυτα σφάλματα. Αυτά μπορούν να εκφραστούν με σχετικούς αριθμούς, οπότε έχουμε σχετικά τυπικά σφάλματα. Το σχετικό τυπικό σφάλμα μιας εκτίμησης ονομάζεται **συντελεστής**

μεταβλητικότητα (coefficient of variation) και είναι το πηλίκο του τυπικού σφάλματος μιας εκτίμησης προς την τιμή της υπό εκτίμηση παραμέτρου. Συμβολίζοντας το συντελεστή μεταβλητικότητας με CV, για την εκτίμηση του \hat{y} έχουμε :

$$CV(\hat{y}) = \frac{SE(\hat{y})}{\hat{y}}$$

1.7 Εκτίμηση ποσοστού ή αναλογίας πληθυσμού

1.7.1 Ποιοτικά χαρακτηριστικά .

Μερικές φορές επιθυμούμε να εκτιμήσουμε τον συνολικό αριθμό, το ποσοστό, ή την αναλογία των μονάδων του πληθυσμού που έχουν κάποιο χαρακτηριστικό ή ανήκουν σε κάποιο συγκεκριμένη τάξη. Π.χ. θέλουμε να εκτιμήσουμε το ποσοστό των γυναικών από τον συνολικό πληθυσμό μιας πόλης, (δηλαδή έχουμε δύο τάξεις, άνδρες, γυναίκες), το ποσοστό ανέργων σε μια πόλη κ.λ.π.

Έστω πλήθος N μονάδων τις οποίες διακρίνουμε σε δυο τάξεις :

- Την τάξη C την οποία αποτελούν οι μονάδες που έχουν μια συγκεκριμένη ιδιότητα και
- Την τάξη C' των υπόλοιπων μονάδων .

Για παράδειγμα, από τον πληθυσμό μιας περιοχής θεωρούμε τα άτομα που ανήκουν σε μια συγκεκριμένη εθνότητα, ότι απαρτίζουν μια τάξη C . Επίσης από το σύνολο των παραχθέντων τεμαχίων κάποιου προϊόντος σε ένα εργοστάσιο, θεωρούμε, σαν τάξη C το σύνολο των ελαττωματικών τεμαχίων.

Υποθέτουμε ότι κάθε μονάδα του πληθυσμού ανήκει σε μια από τις δύο τάξεις C και C' , τότε :

$$\begin{array}{ll} \text{Αριθμός μονάδων της τάξεως } C : & \\ \underline{\text{Στον πληθυσμό}} & \underline{\text{στο δείγμα}} \\ A & a \end{array}$$

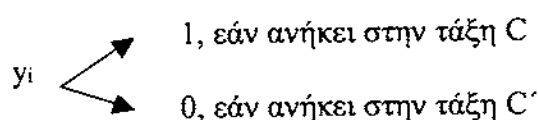
$$\begin{array}{ll} \text{Ποσοστό μονάδων της τάξεως } C : & \\ \underline{\text{Στον πληθυσμό}} & \underline{\text{στο δείγμα}} \\ P = A/N & p = a/n \end{array}$$

Η δειγματική εκτίμηση του P είναι ρ , και η δειγματική εκτίμηση του A είναι $N\rho$ ή Na/n . Στην στατιστική ανάλυση η διωνυμική κατανομή συνήθως εφαρμόζεται για τις εκτιμήσεις a και ρ . Για πεπερασμένους πληθυσμούς ενδείκνυται η υπεργεωμετρική κατανομή, καίτοι η διωνυμική είναι συνήθως μια ικανοποιητική προσέγγιση.

1.7.2 Διακυμάνσεις των δειγματικών εκτιμήσεων.

Τα συμπεράσματα της απλής τυχαίας δειγματοληψίας μπορούμε να χρησιμοποιήσουμε και δίτιμες ποιοτικές μεταβλητές ή όπως αλλιώς λέγονται συνήθως σαν μεταβλητές μηδέν-ένα .

Έτσι για οποιαδήποτε μονάδα στο δείγμα ή τον πληθυσμό, εάν ανήκει στην τάξη C της δίνουμε την τιμή 1 και την τιμή 0 εάν ανήκει στην τάξη C' .



Για τον πληθυσμό των τιμών y_i , θα έχουμε :

$Y = \sum_i^N y_i = A$
$\dot{Y} = \sum_i^N y_i / N = A/N = P$

Επίσης για το δείγμα :

$\dot{y} = \sum_i^n y_i / n = a/n = p$
--

Συνήθως το πρόβλημα της εκτιμήσεως των A και P μπορεί να αντιμετωπισθεί όπως εκείνο της εκτιμήσεως του συνολικού μεγέθους και του μέσου ενός πληθυσμού στον οποίο η y_i παίρνει τις τιμές 1 ή 0 .

$$\sum_{i=1}^N y_i^2 = A = NP \quad , \quad \sum_{i=1}^n y_i^2 = \alpha = np$$

Συνεπώς,

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{Y})^2 = \frac{1}{N} [\sum y_i^2 - N\bar{Y}^2]$$

Επειδή η y_i παίρνει τιμές 1 ή 0

$$= \frac{1}{N} [\sum_{i=1}^N y_i - N\bar{Y}^2] = p - p^2 = P(1-P) = PQ$$

Όπου $Q = 1-P$, το ποσοστό των μονάδων που δεν ανήκουν στην τάξη C

Στην περίπτωση της ισοδύναμης έκφρασης της διακύμανσης του πληθυσμού S^2 έχουμε :

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})^2 = \frac{N}{N-1} PQ$$

Παρομοίως,

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{n}{n-1} pq$$

ΘΕΩΡΗΜΑ 1

Το δειγματικό ποσοστό $p = a/n$ είναι μια αμερόληπτη εκτίμηση του πληθυσμιακού ποσοστού $P = A/N$

ΘΕΩΡΗΜΑ 2

Η διακύμανση του ποσοστού p είναι :

$$V(p) = E(p-P)^2 = \frac{S^2}{n} \frac{(N-n)}{N} = \frac{PQ}{n} \frac{(N-n)}{(N-1)}$$

ΠΟΡΙΣΜΑ

Η διακύμανση της εκτιμήσεως $\hat{A} = Nr$ (εκτίμηση του συνολικού αριθμού των μονάδων της τάξεως C), είναι :

$$V(\hat{A}) = \frac{N^2 PQ}{n} \frac{(N-n)}{(N-1)}$$

ΘΕΩΡΗΜΑ 3

Επειδή στην πράξη η διακύμανση του πληθυσμού είναι συνήθως άγνωστη χρησιμοποιούμε την αμερόληπτη εκτίμηση της διακύμανσης του p που προκύπτει από το δείγμα, δηλαδή

$$V(p) = \frac{s^2}{n} \frac{(N-n)}{N} = \frac{pq}{n-1} \frac{(N-n)}{(N)}$$

Διότι

$$s^2 = \frac{n}{n-1} pq$$

1.8 διαστήματα εμπιστοσύνης

Από τον παραπάνω τύπο για την διακύμανση του δειγματικού ποσοστού p , μιας μορφή της κανονικής προσεγγίσεως για το διάστημα εμπιστοσύνης του πληθυσμού ποσοστό P είναι :

$$p \pm [Z_{\alpha/2} \sqrt{1-f} \sqrt{pq} / (n-1) + 1/2n]$$

Όπου $f = n/N$ και $Z_{\alpha/2}$ η τιμή της τυπικής μεταβλητής $Z \sim N(0,1)$ που αντιστοιχεί στην πιθανότητα εμπιστοσύνης. Συνήθως, χρησιμοποιούμε τον περισσότερο γνωστό όρο \sqrt{pq} / n αντί του $\sqrt{pq} / (n-1)$ που σπάνια δίνει μια αισθητή διαφορά. Ο τελευταίος όρος $1/2n$ είναι η διόρθωση συνεχείας, που δημιουργεί μια ελαφρά βελτίωση στην προσέγγιση. Πάντως, χωρίς την διόρθωση συνεχείας η κανονική προσέγγιση συνήθως δίνει πολύ μικρού πλάτους διάστημα εμπιστοσύνης.

Το σφάλμα που μπορεί να προκύψει κατά την κανονική προσέγγιση εξαρτάται από τις ποσότητες :

- n = μέγεθος δείγματος
- N = μέγεθος πληθυσμού
- p = δειγματικό ποσοστό
- α_u = πιθανότητα με την οποία το ανώτερο όριο του διαστήματος εμπιστοσύνης είναι κατώτερο του πραγματικού ποσοστού P
- α_l = πιθανότητα με την οποία το κατώτερο όριο του διαστήματος εμπιστοσύνης υπερβαίνει το πραγματικό ποσοστό P

Η ποσότητα που επηρεάζει περισσότερο το σφάλμα είναι το np ή περισσότερο συγκεκριμένα ο αριθμός των παρατηρήσεων στην μικρότερη τάξη. Ο πίνακας μας καθοδηγεί να αποφασίσουμε για την χρησιμοποίηση της κανονικής προσέγγισης.

ΠΙΝΑΚΑΣ

Ελάχιστες τιμές του np για χρησιμοποίηση της κανονικής προσεγγίσεως

p	np = αριθμός παρατηρήσεων στην μικρότερη τάξη	n = μέγεθος δείγματος
0,5	15	30
0,4	20	50
0,3	24	80
0,2	40	200
0,1	60	600
0,05	70	1400
*	80	...

* : σημαίνει ότι το ποσοστό p είναι σημαντικά μικρό, ούτως ώστε να ακολουθεί την κατανομή Poisson.

Τα αποτελέσματα του πίνακα έχουν προκύψει έτσι ώστε η συχνότητα, με την οποία το πραγματικό ποσοστό P δεν περιέχεται μεταξύ των ορίων εμπιστοσύνης, να μην υπερβαίνει το 5,5%. Επί πλέον, η πιθανότητα με την οποία το ανώτερο όριο του διαστήματος εμπιστοσύνης είναι κατώτερο του πραγματικού ποσοστού P κυμαίνεται από 2,5% έως 3,5%, και η πιθανότητα με την οποία το κατώτερο όριο του διαστήματος εμπιστοσύνης υπερβαίνει το πραγματικό ποσοστό P κυμαίνεται από 2,5% έως 1,5% .

Διάστημα εμπιστοσύνης όταν υπάρχουν περισσότερες από δύο Τάξεις

Πολύ συχνά, κατά την παρουσίαση των αποτελεσμάτων, οι μονάδες δειγματοληψίας ταξινομούνται σε περισσότερες από δύο τάξεις. Έτσι ένα δείγμα από ένα ανθρώπινο πληθυσμό μπορεί να ταξινομηθεί σε 15 ομάδες ηλικιών πλάτους 5 ετών. Ακόμη όταν μια ερώτηση πρέπει να απαντηθεί με ένα απλό «ναι» ή «όχι» τα αποτελέσματα που θα προκύψουν μπορούν να ταξινομηθούν σε τέσσερις τάξεις: «ναι», «όχι», «δεν γνωρίζω» και «όχι απάντηση». Η επέκταση της προηγούμενης θεωρίας σε τέτοιες περιπτώσεις θα παρουσιασθεί με εφαρμογή στην περίπτωση που υπάρχουν τρεις τάξεις .

Υποθέτουμε ότι ο αριθμός των μονάδων που ανήκουν στην τάξη i είναι A_i στον πληθυσμό και a_i στο δείγμα, οπότε :

$$N = \sum A_i, \quad n = \sum a_i, \quad P_i = \frac{A_i}{N}, \quad p_i = \frac{a_i}{n}$$

Η πιθανότητα επιλογής του παρατηρηθέντος δείγματος είναι :

$$P(a_i / A_i) = \left\{ \frac{A_1}{n} \right\} \left\{ \frac{A_2}{n} \right\} \left\{ \frac{A_3}{n} \right\} \dots \left\{ \frac{A_k}{n} \right\}$$

ή κατά προσέγγιση, όταν το κλάσμα δειγματοληψίας είναι μικρό.

$$P(a_i) = \frac{n!}{a_1! a_2! a_3! \dots a_k!} p_1^{a_1} p_2^{a_2} p_3^{a_3} \dots p_k^{a_k}$$

Όταν υπάρχουν περισσότερες από δύο τάξεις για τον καθορισμό των ορίων εμπιστοσύνης, διακρίνουμε δύο περιπτώσεις .

Περίπτωση 1 : Υπολογίζουμε το ποσοστό

$$\rho = \frac{\text{Αριθμός σε οποιαδήποτε τάξη του δείγματος}}{n} = \frac{a_1}{n}$$

ή

$$\rho = \frac{\text{Συνολικός αριθμός σε μια ομάδα τάξεων}}{n} = \frac{a_1 + a_2 + a_3}{n}$$

σε οποιοδήποτε από αυτές τις καταστάσεις, αν και η αρχική ταξινόμηση περιέχει περισσότερες από δύο τάξεις, το ποσοστό ρ λαμβάνεται από μια υποδιαίρεση του δείγματος $\langle n \rangle$ μόνο σε δύο τάξεις. Η θεωρία που ήδη αναπτύχθηκε στην αρχή της παραγράφου εφαρμόζεται σε αυτή την περίπτωση .

Περίπτωση 2 : Μερικές κάποιες τάξεις παραλείπονται, το ποσοστό ρ υπολογίζεται από την κατάτμηση των υπόλοιπων τάξεων σε δύο μέρη. Για παράδειγμα, θα μπορούσαμε να παραλείψουμε τα άτομα που δεν απάντησαν στο ερωτηματολόγιο μιας έρευνας και να θεωρήσουμε τον λόγο του αριθμού αυτών που απάντησαν «ναι» δια του αριθμού αυτών που απάντησαν «ναι»+ «όχι». Οι λόγοι αυτού του τύπου, συγκεντρώνουν συνήθως και το ενδιαφέρον αυτών που ασχολούνται με δειγματοληπτικές έρευνες. Ο παρονομαστής ενός λόγου όπως περιγράψαμε ανωτέρω δεν είναι $\langle n \rangle$ αλλά κάποιος μικρότερος αριθμός n' ($n' < n$).

Καίτοι ο αριθμός $\langle n' \rangle$ ποικίλει από δείγμα σε δείγμα, τα προηγούμενα αποτελέσματα μπορούν να χρησιμοποιηθούν, εάν θεωρήσουμε την υπό συνθήκη κατανομή του δειγματικού ποσοστού ρ σε δείγματα στα οποία οι αριθμοί $\langle n \rangle$ και $\langle n' \rangle$ είναι σταθεροί .

Έστω ότι,

$$\rho = \frac{a_1}{a_1 + a_2}, \quad n' = a_1 + a_2 \quad n = a_1 + a_2 + a_3$$

όπου a_3 είναι ο αριθμός των μονάδων του δείγματος $\langle n \rangle$, που ανήκουν στις τάξεις για τις οποίες δεν ενδιαφερόμεθα. Τότε, η υπό συνθήκη κατανομή των a_1 και a_2 , είναι η υπεργεωμετρική κατανομή που

προκύπτει όταν το δείγμα είναι μεγέθους $< n' >$ και ο πληθυσμός μεγέθους $N' = A_1 + A_2$.

Συνεπώς, εάν λάβουμε υπ' όψη μας την κανονική προσέγγιση για το διάστημα εμπιστοσύνης, προκύπτει ότι η κανονική προσέγγιση για τα υπό συνθήκη όρια εμπιστοσύνης του $p = A_1 / A_1 + A_2$ είναι :

$$p \pm \left[Z_{\alpha/2} \frac{\sqrt{(1-p') \frac{pq}{n'-1}} + \frac{1}{2n'}}{N'} \right]$$

Εάν η τιμή N' είναι άγνωστη, μπορούμε αντί του λόγου n'/N' να θέσουμε τον λόγο n/N .

1.9 Εκτίμηση μέσων και συνολικών μεγεθών υποπληθυσμών

Σε πολλές έρευνες οι εκτιμήσεις γίνονται κατά κατηγορία ενός πληθυσμού (υποπληθυσμοί). Π.χ. σε μια έρευνα νοικοκυριών μπορεί να εκτιμηθούν τα νοικοκυριά κατά μέγεθος, κατά εισόδημα, κατά επάγγελμα του υπευθύνου του νοικοκυριού κ.λ.π. Επίσης πολλές φορές στα διαθέσιμα δειγματοληπτικά πλαίσια περιέχονται μονάδες οι οποίες δεν ανήκουν στις υπό μελέτη μονάδες και δεν είναι εύκολο να γίνει διαχωρισμός, παρά μόνο μετά την επιλογή του δείγματος. Για παράδειγμα, έστω ότι έχουμε στην διάθεση μας ένα δειγματοληπτικό πλαίσιο (κατάλογο) των νοικοκυριών μιας πόλης και θέλουμε να συγκεντρώσουμε πληροφορίες από νοικοκυριά στα οποία υπάρχουν άτομα ηλικίας πάνω των 65 ετών. Στην περίπτωση αυτή δεν μπορούμε εκ των προτέρων να διακρίνουμε τα νοικοκυριά που έχουν άτομα ηλικίας πάνω των 65 ετών. Για το λόγο αυτό παίρνουμε ένα τυχαίο δείγμα από το σύνολο του πληθυσμού και εξετάζουμε τον αριθμό των νοικοκυριών τα οποία έχουν το χαρακτηριστικό που μας ενδιαφέρει.

Έστω ότι ο αριθμός των μονάδων του πληθυσμού είναι N και N_j είναι ο αριθμός των μονάδων της κατηγορίας (υποπληθυσμού) που μας ενδιαφέρει. Αν ληφθεί δείγμα μεγέθους n από τον πληθυσμό N και ο αριθμός των μονάδων του δείγματος που εμπίπτει στον υποπληθυσμο N_j είναι n_j , τότε συμβολίζοντας με y_k τις μετρήσεις των μονάδων αυτών (όπου $k = 1, 2, \dots, n_j$), ο μέσος του πληθυσμού \bar{Y}_j για την κατηγορία j εκτιμάται από την σχέση:

$$\hat{y}_j = \sum_{k=1}^{n_j} y_{jk} / n_j$$

Η διακόμανση της εκτίμησης \hat{y}_j δίνεται από την σχέση :

$$V(\hat{y}_j) = \frac{(N_j - n_j)}{N_j} \sum_{k=1}^{n_j} (y_{jk} - \hat{y}_j)^2 / n_j (n_j - 1)$$

Το τυπικό σφάλμα είναι :

$$SE(\hat{y}_j) = \sqrt{V(\hat{y}_j)}$$

Εάν η τιμή του N_j είναι άγνωστη μπορεί να χρησιμοποιηθεί το n/N αντί του n_j/N_j για τον υπολογισμό της διόρθωσης πεπερασμένου πληθυσμού.

1. 10 Εκτίμηση του μεγέθους του δείγματος

1.10.1 Προσδιορισμός μεγέθους δείγματος για την εκτίμηση μέσου και συνολικού πληθυσμού .

Κατά τον σχεδιασμό μιας δειγματοληπτικής έρευνας, ο προσδιορισμός του μεγέθους του δείγματος αποτελεί μια από τις σπουδαιότερες αποφάσεις. Πολύ μεγάλο δείγμα συνεπάγεται σπατάλη χρόνου και χρήματος, ενώ πολύ μικρό δείγμα συνεπάγεται αναξιόπιστα αποτελέσματα .

Η επάρκεια δείγματος εξαρτάται, κυρίως, από τον επιθυμητό βαθμό ακρίβειας, με άλλα λόγια τι περιθώρια σφάλματος πρόκειται να δεχθούμε σε μια δειγματοληπτική εκτίμηση. Ο βαθμός ακρίβειας της εκτίμησης δίνεται από τον τυπικό σφάλμα μιας παραμέτρου, π.χ. στην περίπτωση εκτίμησης του μέσου αριθμητικού, το τυπικό σφάλμα εκφράζει τη διασπορά του πλήθους όλων των δυνατών εκτιμήσεων \hat{y} από τον μέσο του πληθυσμού \bar{Y} . Όσο μικρότερο είναι το τυπικό σφάλμα, τόσο μικρότερο θα είναι το δειγματοληπτικό σφάλμα $|\hat{y} - \bar{Y}|$.

Το τυπικό σφάλμα εξαρτάται από το μέγεθος του δείγματος και από την διακύμανση του ερευνώμενου πληθυσμού. Όσο μικρότερη είναι η διακύμανση ενός πληθυσμού, τόσο μικρότερο δείγμα απαιτείται. Η διακύμανση του ερευνώμενου πληθυσμού είναι συνήθως άγνωστη. Στην περίπτωση αυτή μπορεί να βοηθήσουν τα αποτελέσματα προηγούμενης έρευνας ή μια δοκιμαστική έρευνα (pilot survey).

Μια άλλη δυσκολία που αντιμετωπίζεται κατά τον προσδιορισμό του μεγέθους του δείγματος, είναι ότι στην πράξη συγκεντρώνονται πληροφορίες για περισσότερες από μια μεταβλητές (χαρακτηριστικά). Έτσι ένα δείγμα το οποίο μπορεί να είναι ικανοποιητικό για μια μεταβλητή, ενδέχεται να μην είναι αρκετό για τις υπόλοιπες. Ένας τρόπος προσδιορισμού του μεγέθους του δείγματος είναι να καθοριστούν περιθώρια σφάλματος για τις μεταβλητές που θεωρούνται ως πιο σπουδαίες για την έρευνα. Στη συνέχεια, προσδιορίζεται το μέγεθος του δείγματος που χρειάζεται, ξεχωριστά για κάθε μια από τις σπουδαιότερες μεταβλητές. Αν τα μεγέθη (n) δεν διαφέρουν σημαντικά μεταξύ τους, το δε μεγαλύτερο από τα μεγέθη (n) τύχει να περιέχεται εντός του καθορισμένου ορίου, επιλέγεται ως μέγεθος δείγματος. Συνήθως τα μεγέθη (n) των μεταβλητών διαφέρουν μεταξύ τους, οπότε επιλέγεται το μεγαλύτερο (n) γιατί δίνει τη μεγαλύτερη ακρίβεια.

Μερικές φορές τα μεγέθη (n) που απαιτούνται για διάφορες μεταβλητές είναι τόσο ασύμφωνα μεταξύ τους, ώστε μερικά από αυτά πρέπει να παραλειφθούν. Στην πράξη, το επιθυμητό μέγεθος είναι απραγματοποίητο, λόγω περιορισμών στα οικονομικά μέσα και στον απαιτούμενο χρόνο. Σε τέτοιες περιπτώσεις είναι καλύτερα να πάρει κανείς το μεγαλύτερο δυνατό δείγμα, σε σχέση με τα διαθέσιμα οικονομικά, και να απορρίψει ερωτήματα για τα οποία χρειάζεται μεγαλύτερο δείγμα, ώστε να επιτευχθούν χρήσιμα αποτελέσματα από την δειγματοληπτική έρευνα.

Γενικά, μπορούμε να πούμε, ότι ο προσδιορισμός του μεγέθους του δείγματος καθορίζεται από τους εξής παράγοντες :

- τον επιθυμητό βαθμό ακριβείας των αποτελεσμάτων μιας έρευνας
- τη διακύμανση του ερευνώμενου πληθυσμού
- τον αριθμό των μεταβλητών που πρόκειται να ερευνηθούν .
- τα διαθέσιμα οικονομικά και

• το διαθέσιμο χρόνο

Η επιθυμητή ακρίβεια μιας εκτίμηση καθορίζεται από το επιτρεπτό δειγματοληπτικού σφάλματος, έτσι ώστε να εξασφαλίζεται η αξιοπιστία των αποτελεσμάτων μιας δειγματοληπτικής έρευνας. Συνήθως ο προσδιορισμός του επιτρεπτού ορίου δειγματοληπτικού σφάλματος γίνεται σε κάποια όρια διαστήματος εμπιστοσύνης και σε ορισμένο επίπεδο πιθανότητας (π.χ. 95%, 90% κ.ο.κ.). Για παράδειγμα, αν θέλουμε να εκτιμήσουμε τον μέσο του πληθυσμού με δειγματοληπτικό σφάλμα $|\hat{y} - \bar{Y}|$ το οποίο δεν πρέπει να υπερβεί ένα περιθώριο d και με συντελεστή εμπιστοσύνης $\{1 - \alpha\}$ δηλαδή αν θέλουμε

$$Pr\{|\hat{y} - \bar{Y}| \leq d\} = 1 - \alpha$$

Τότε το απαιτούμενο μέγεθος δείγματος προκύπτει ως εξής :

Το διάστημα εμπιστοσύνης για το μέσο \bar{Y} ενός πληθυσμού που ακολουθεί την κανονική κατανομή, με συντελεστή εμπιστοσύνης $1 - \alpha$ είναι :



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

Από το διάστημα εμπιστοσύνης προκύπτει ότι η ακρίβεια της εκτίμησης του μέσου \bar{Y} του πληθυσμού εξαρτάται από το τυπικό σφάλμα. Δηλαδή, όσο πιο μεγάλο είναι το τυπικό σφάλμα, τόσο πιο μεγάλο θα είναι και το διάστημα εμπιστοσύνης και κατά συνέπεια, θα έχουμε πιο μικρή ακρίβεια της εκτίμησης του μέσου \bar{Y} . Η ακρίβεια της εκτίμησης αυτής μπορεί να αυξηθεί, αν αυξηθεί το μέγεθος του δείγματος.

Για να ισχύει ο περιορισμός, ότι το δειγματοληπτικό σφάλμα δεν πρέπει να υπερβεί το περιθώριο d θα πρέπει να ισχύει η παρακάτω εξίσωση:

$$n = \frac{Z_{\alpha/2}^2 \cdot \sigma^2}{d^2}$$

Σε περίπτωση που η πληθυσμιακή διακύμανση σ^2 είναι άγνωστη, μπορούμε να την εκτιμήσουμε από στοιχεία δοκιμαστικής έρευνας, λαμβάνοντας ένα μικρό δείγμα από τον ερευνώμενο πληθυσμό, είτε να χρησιμοποιήσουμε τη διακύμανση από προηγούμενη παρόμοια έρευνα .

Στην περίπτωση που θέλουμε να εκτιμήσουμε το συνολικό πληθυσμό Y με δειγματοληπτικό σφάλμα $|\hat{Y} - Y|$ το οποίο δεν πρέπει να υπερβεί ένα περιθώριο d και με συντελεστή εμπιστοσύνης $(1-\alpha)$ δηλαδή αν θέλουμε

$$P \{ |\hat{Y} - Y| \leq d \} = 1 - \alpha$$

Τότε το απαιτούμενο μέγεθος δείγματος προκύπτει ως εξής :

Το διάστημα εμπιστοσύνης για το συνολικό πληθυσμό Y που ακολουθεί την κανονική κατανομή, με συντελεστή εμπιστοσύνης $(1-\alpha)$ είναι :

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

Οπότε το απαιτούμενο μέγεθος δείγματος προσδιορίζεται από την εξίσωση :

$$n = \frac{Z_{\alpha/2}^2 \cdot \sigma^2 \cdot N^2}{d^2}$$

1.10.2 Προσδιορισμός μεγέθους δείγματος λαμβανομένου υπόψη και του κόστους .

Όταν το κόστος για μια δειγματοληπτική έρευνα είναι δεδομένο, τότε το μέγεθος του δείγματος πρέπει να προσδιορίζεται σε συνάρτηση με το κόστος. Στη συνέχεια, να υπολογίζεται η μέγιστη δυνατή ακρίβεια, που μπορεί να επιτευχθεί με το διαθέσιμο χρηματικό πόσο, και αν είναι εφικτό να προκύψουν αξιόπιστα αποτελέσματα, να διεξάγεται η δειγματοληπτική έρευνα, σε διαφορετική περίπτωση να απορρίπτεται.

Η συνάρτηση κόστους σε μια δειγματοληπτική έρευνα είναι :

$$E = \gamma + n \cdot \varepsilon$$

Όπου

- E είναι το συνολικό κόστος της έρευνας
 γ είναι τα γενικά έξοδα της έρευνας (σχεδιασμός, οργάνωση κ.λ.π.)
 ε είναι η δαπάνη κατά δειγματοληπτική μονάδα (συνεντεύξεις για τη συμπλήρωση των ερωτηματολογίων, επεξεργασία ερωτηματολογίων κ.λ.π.)
 n είναι το μέγεθος του δείγματος .

1. 10.3 Το μέγεθος του δείγματος στην δειγματοληψία για ποσοστά

Οι μονάδες δειγματοληψίας ταξινομούνται σε δυο τάξεις C και C'. Κάποιο όριο του σφάλματος d στον εκτιμώμενο ποσοστό P των μονάδων στην τάξη C έχει συμφωνηθεί και υπάρχει ένας μικρός κίνδυνος α με τον οποίο είμαστε πρόθυμοι να δεχθούμε ότι το πραγματικό σφάλμα είναι μεγαλύτερο από d .

Δηλαδή, συνοπτικά επιθυμούμε :

$$P \{ |p-P| \geq d \} = \alpha$$

Υποθέτουμε, ότι ενεργούμε απλή τυχαία δειγματοληψία και το δειγματικό ποσοστό p κατανέμεται κανονικά. Γνωρίζουμε ότι το διάστημα εμπιστοσύνης για το ποσοστό P ενός πληθυσμού με συντελεστή εμπιστοσύνης 1-α δίνεται από τον παρακάτω τύπο :

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

Όπου SE(p) είναι το τυπικό σφάλμα και δίνεται από την σχέση :

$$SE(p) = \frac{\sqrt{PQ}}{\sqrt{n}} \cdot \frac{\sqrt{N-n}}{\sqrt{N-1}}$$

Κατά συνέπεια, ο τύπος που συνδέει το $\langle n \rangle$ με τον επιθυμητό βαθμό ακρίβειας είναι

$$d = Z_{\alpha/2} \frac{\sqrt{PQ}}{\sqrt{n}} \cdot \frac{\sqrt{N-n}}{\sqrt{N-1}}$$

Αν ο πληθυσμός N είναι αρκετά μεγάλος, τότε παραλείπουμε τον συντελεστή διόρθωσης πεπερασμένου πληθυσμού και λύνοντας ως προς το n έχουμε :

$$n = \frac{Z_{\alpha/2}^2 PQ}{d^2} = \frac{PQ}{V}$$

Όπου $V = d^2 / Z^2 =$ επιθυμητή διακύμανση στο δειγματικό ποσοστό

1.10.4 Εκτιμήσεις των πληθυσμιακών διακυμάνσεων για τον καθορισμό του μεγέθους του δείγματος .

Στις πρακτικές εφαρμογές, υπάρχουν τέσσερις τρόποι εκτίμησης των πληθυσμιακών διακυμάνσεων προκειμένου να καθορίσουμε το μέγεθος του δείγματος.

1. Παίρνοντας το δείγμα σε δύο στάδια, όπου το πρώτο είναι ένα απλό τυχαίο δείγμα μεγέθους $\langle n \rangle$ από το οποίο η τιμή της διακυμάνσεως S^2 ή του ποσοστού P και το απαιτούμενο μέγεθος δείγματος $\langle n \rangle$ θα ληφθούν .
2. Από τα αποτελέσματα ενός δείγματος πιλότου .
3. Από προγενέστερη δειγματοληψία επί του ίδιου ή παρόμοιου πληθυσμού και
4. Από υποθέσεις για την δομή του πληθυσμού .

Μέθοδος 1

Η πρώτη μέθοδος δίνει τις περισσότερες αξιόπιστες εκτιμήσεις της S^2 ή του P , αλλά δεν χρησιμοποιείται συχνά διότι επιβραδύνει την ολοκλήρωση της έρευνας. Όταν η μέθοδος αυτή είναι εφικτή, ο Cox (συγγραφέας) δείχνει πως να υπολογίσουμε το $\langle n \rangle$ από A^2_1 ή το ρ_1 ούτως ώστε η τελική εκτίμηση \hat{y} ή \hat{p} θα έχει μια εκ των προτέρων προσδιορισθείσα διακύμανση V , ένα εκ των προτέρων προσδιορισθέν όριο σφάλματος d , ή ένα εκ των προτέρων προσδιορισθέντα συντελεστή μεταβλητικότητας. Το πρώτο δείγμα υποτίθεται αρκετά μεγάλο ώστε να καθίστανται αμελητέοι όροι της τάξεως $1/n^2_1$. Μερικά αποτελέσματα παρατίθενται κατωτέρα.

Εκτίμηση του \hat{Y} με διακύμανση V

Εάν s^2_1 είναι η διακύμανση από το πρώτο δείγμα, παίρνουμε επιπρόσθετες μονάδες για να φτιάξουμε το συνολικό μέγεθος του τελικού δείγματος

$$n = \frac{s^2_1}{V} (1 + 2/n^2_1)$$

Η κατανομή του y υποτίθεται ότι προσεγγίζει την κανονική. Εάν η τυπική απόκλιση S είναι γνωστή, το απαιτούμενο μέγεθος δείγματος θα ήταν S^2/V . Συνέπεια της μη γνώσεως της τυπικής αποκλίσεως S είναι η αύξηση του μέσου μεγέθους δια του παράγοντα $(1 + 2/n^2_1)$

Εκτίμηση του ποσοστού P με διακύμανση V

Έστω ότι ρ_1 είναι η εκτίμηση του ποσοστού P από το πρώτο δείγμα. Το τελικό δείγμα που προκύπτει από τον συνδυασμό των πρώτων δύο δειγμάτων θα είναι:

$$n = \frac{\rho_1 q_1}{V} + \frac{3 - 8\rho_1 q_1}{\rho_1 q_1} + \frac{1 - 3\rho_1 q_1}{V n_1}$$

Ο πρώτος όρος της σχέσεως είναι το μέγεθος του δείγματος που απαιτείται εάν είναι γνωστό ότι το πληθυσμιακό ποσοστό P είναι ίσο

προς p_1 . Με αυτή την μέθοδο η συνήθης διωνυμική εκτίμηση p , που προκύπτει από το τελικό δείγμα μεγέθους $\langle n \rangle$ είναι ελαφρά μεροληπτική. Για διόρθωση της μεροληψίας, παίρνουμε :

$$p = p + \frac{V(1-2p)}{pq}$$

Εκτίμηση του ποσοστού P με δεδομένο $cv = \sqrt{c}$

Στην περίπτωση αυτή, παίρνουμε :

$$n = \frac{q_1}{cp_1} + \frac{3}{p_1q_1} + \frac{1}{cp_1n_1}$$

Και στα τρία αποτελέσματα που δίνονται ανώτερο ο όρος $(1-f)$ αγνοείται .

Μέθοδος 2

Η δεύτερη μέθοδος, της έρευνας με ένα μικρό δείγμα πλότος, εξυπηρετεί πολλούς σκοπούς, ειδικά εάν το εφικτό της κύριας έρευνας είναι αμφίβολο. Εάν η έρευνα «πλότος» στηρίζεται σε ένα απλό τυχαίο δείγμα, οι προηγούμενες μέθοδοι εφαρμόζονται. Αλλά, πολλές φορές η εργασία δια του δείγματος «πλότος» περιορίζεται σε ένα τμήμα του πληθυσμού το οποίο είναι προσιτό για ερευνητική επεξεργασία ή το οποίο θα αποκαλύψει το μέγεθος ορισμένων προβλημάτων. Η επιλογή του δείγματος «πλότος» πρέπει ιδιαίτερα να προσεχθεί όταν τα αποτελέσματα του χρησιμοποιηθούν για να εκτιμήσουμε την S^2 ή το P. Για παράδειγμα, μια συνηθισμένη πρακτική είναι να περιορίζεται το δείγμα «πλότος» σε λίγες συσσωρεύσεις (ομάδες) μονάδων δειγματοληψίας. Συνεπώς, η υπολογιζόμενη s^2 μετρά ως επί το πλείστον την μεταβλητικότητα μέσα σε μια συσώρευση και μπορεί να είναι μια υποεκτίμηση της σχετικής διακύμανσης S^2 . Το ίδιο πρόβλημα προκύπτει στην συσσωρευτική δειγματοληψία για ποσοστά, στην οποία ο τύπος pq/n μπορεί να υποεκτιμήσει το αποτέλεσμα της μεταβλητικότητας μεταξύ των συσσωρεύσεων.

Μέθοδος 3

Η χρησιμοποίηση των αποτελεσμάτων από τις προηγούμενες έρευνες καταδεικνύει την αξία διαθέσεως, ή τουλάχιστον της τηρήσεως προσιτών, οποιονδήποτε επί των τυπικών αποκλίσεων που ελήφθησαν σε προηγούμενες έρευνες. Δυστυχώς το κόστος υπολογισμού τυπικών αποκλίσεων σε περίπλοκες έρευνες είναι πολύ υψηλό, ακόμη και με ηλεκτρονικές μηχανές, και συχνά μόνο εκείνες οι τυπικές αποκλίσεις που χρειάζονται για να δώσουν μια πρόχειρη ιδέα της ακρίβειας των βασικών εκτιμήσεων, υπολογίζονται και καταγράφονται. Εάν βρεθούν κατάλληλα παρελθόντα δεδομένα, η τιμή της διακυμάνσεως μπορεί να απαιτήσει κατάλληλη προσαρμογή στις διαχρονικές μεταβολές. Στην περίπτωση των ασυμμετρικών δεδομένων όπου ο μέσος \bar{Y} μεταβάλλεται διαχρονικά. Η διακύμανση S^2 βρίσκεται συχνά να μεταβάλλεται με ρυθμό που κυμαίνεται κάπου μεταξύ των ορίων $\kappa\bar{Y}$ και $\kappa\bar{Y}^2$, όπου κ είναι σταθερό. Συνεπώς, εάν ο μέσος \bar{Y} θεωρηθεί ότι έχει αυξηθεί κατά 10% στο χρονικό διάστημα που μεσολαβεί από την προηγούμενη έρευνα, θα μπορούσαμε να αυξήσουμε την αρχική εκτίμηση της διακυμάνσεως από 10 έως 20% .

Τέλος, μια χρήσιμη εκτίμηση της διακυμάνσεως μπορεί μερικές φορές να γίνει από σχετικά μικρή πληροφόρηση γύρω από την φύση του πληθυσμού .

Ελλείπει των προηγούμενων εκτιμήσεων, ο Deming (συγγραφέας) έχει δείξει πως μερικές απλές κατανομές να χρησιμοποιηθούν για να εκτιμήσουμε την διακύμανση από μια πληροφορία του εύρους και μια γενική ιδέα του σχήματος της κατανομής. Εάν η κατανομή είναι περίπου διωνυμική, με ένα ποσοστό p των παρατηρήσεων στο ένα άκρο του εύρους και ένα ποσοστό q στο άλλο άκρο, $S^2 = pqh^2$, όπου h είναι το εύρος. Όταν $p=q=1/2$, η τιμή της διακυμάνσεως S^2 είναι $S^2 = 0,25 h^2$ δηλαδή η μέγιστη δυνατή για δεδομένο εύρος h . Άλλες χρήσιμες σχέσεις είναι ότι $S^2 = 0,83 h^2$ για μια ορθογώνια κατανομή, $S^2 = 0,056 h^2$ για μια κατανομή με σχήμα ορθογώνιο τρίγωνο και $S^2 = 0,045 h^2$ για ένα ισοσκελές τρίγωνο.

Αυτές οι σχέσεις δεν βοηθούν πολύ εάν το εύρος είναι μεγάλο ή λίγο γνωστό. Όμως εάν το εύρος είναι μεγάλο, καλή δειγματοληπτική πρακτική είναι να στρωματοποιήσουμε τον πληθυσμό ούτως ώστε μέσα σε οποιοδήποτε στρώμα το εύρος να είναι αρκετά ελαττωμένο. Συνήθως το σχήμα γίνεται απλούστερο μέσα σε ένα στρώμα. Συνήθως αυτές οι σχέσεις είναι αποτελεσματικές για την πρόβλεψη της διακύμανσης μέσα στα επί μέρους στρώματα .

1. 10 . 5 Μέγεθος δείγματος όταν οι εκτιμήσεις αφορούν σε υποδιαιρέσεις του πληθυσμού .

Συχνά επιδιώκεται να παρουσιασθούν εκτιμήσεις όχι μόνο για τον πληθυσμό σαν σύνολο, αλλά και για ορισμένες υποδιαίρεσεις αυτού. Εάν αυτές μπορούν να καθορισθούν εκ των προτέρων, όπως οι διάφορες γεωγραφικές περιοχές, τότε, θα λάβει χώρα χωριστός υπολογισμός του $\langle n \rangle$ για κάθε περιοχή. Υποθέτουμε ότι ο μέσος κάθε υποδιαίρεσεως πρόκειται να εκτιμηθεί με μια συγκεκριμένη διακύμανση V . Για την i υποδιαίρεση, έχουμε $n_i = S_i^2 / V$, έτσι ώστε το συνολικό δείγμα μεγέθους $\langle n \rangle$ θα είναι $n = \sum S_i^2 / V$

Η ατομική διακύμανση S_i^2 κατά μέσο όρο, θα είναι μικρότερη της πληθυσμιακής διακυμάνσεως S^2 . Έτσι, εάν υπάρχουν k υποδιαίρεσεις $n = kS^2 / V$, ενώ εάν εξητείτο μόνο η εκτίμηση για τον συνολικό πληθυσμό, το μέγεθος του δείγματος θα ήταν :

$$n = S^2 / V$$

Συνεπώς, εάν οι εκτιμήσεις με διακύμανση V ζητούνται για κάθε μια από τις k υποδιαίρεσεις το μέγεθος του δείγματος πρέπει να είναι, περίπου, k φορές μεγαλύτερο από αυτό που θα χρειαζόταν για μια συνολική εκτίμηση της ίδιας ακρίβειας .

Εάν οι υποδιαίρεσεις αντιπροσωπεύουν ταξινομήσεις με μεταβλητές όπως ηλικία, εισόδημα, έτη εκπαίδευσεως, η υποδιαίρεση, στην οποία ένα άτομο ανήκει δεν είναι γνωστή μέχρις ότου το δείγμα επιλεγεί. Εκτιμήσεις του μεγέθους του δείγματος μπορούν ακόμη να γίνουν εάν το ποσοστό p_i των μονάδων δειγματοληψίας οι οποίες ανήκουν στις επί μέρους υποδιαίρεσεις είναι γνωστά. Εάν ένα απλό τυχαίο δείγμα μεγέθους $\langle n \rangle$ πρόκειται να επιλεγεί, το αναμενόμενο μέγεθος του δείγματος από την i υποδιαίρεση του πληθυσμού είναι $n_i = n p_i$.

1. 11 Δειγματοληψία με πιθανότητα προς το μέγεθος των μονάδων

Στις προηγούμενες παραγράφους του κεφαλαίου αυτού, έγινε περιγραφή του μεθόδου της απλής τυχαίας δειγματοληψίας, όπου η επιλογή του δείγματος γίνεται χωρίς επανατοποθέτηση, με πιθανότητα επιλογής ίση για όλες τις μονάδες του ερευνώμενου πληθυσμού και με την υπόθεση ότι οι δειγματοληπτικές μονάδες είναι περίπου του αυτού μεγέθους .

Πολλές φορές οι δειγματοληπτικές μονάδες διαφέρουν σημαντικά σε μέγεθος και υπάρχει έντονη συσχέτιση μεταξύ του μεγέθους της μονάδας και των υπό εξέταση χαρακτηριστικών. Π.χ. η μεταβλητή «αριθμός υπαλλήλων που απουσιάζουν από τη θέση τους λόγω ασθένειας» επηρεάζεται έντονα από το μέγεθος της επιχείρησης [δηλαδή από τον αριθμό υπαλλήλων της επιχείρησης].

Στην περίπτωση αυτή, η εφαρμογή της απλής τυχαιάς δειγματοληψίας με πιθανότητα επιλογής ίση για όλες τις μονάδες, μπορεί να οδηγήσει σε εσφαλμένες εκτιμήσεις, αφού δεν λαμβάνεται υπόψη η επίδραση των μεγαλύτερων δειγματοληπτικών μονάδων στα υπό μελέτη χαρακτηριστικά του ερευνώμενου πληθυσμού.

Το πρόβλημα αυτό μπορεί να βελτιωθεί σε κάποιο βαθμό με την εφαρμογή της στρωματοποιημένης δειγματοληψίας. Ένας άλλος τρόπος βελτίωσης του προβλήματος είναι η επιλογή των δειγματοληπτικών μονάδων με πιθανότητα ανάλογη προς το μέγεθος των μονάδων και με επανατοποθέτηση [δηλαδή μια δειγματοληπτική μονάδα μπορεί να επιλεγεί περισσότερες από μια φορά] .Με τον τρόπο αυτό δίνεται μεγαλύτερη ευκαιρία στις μεγαλύτερες μονάδες , να επιλεγούν στο δείγμα.

1. 12 Συμπέρασμα .

Με λίγα λόγια θα μπορούσαμε να πούμε ότι τυχαιά δειγματοληψία είναι εκείνη κατά την οποία το δείγμα λαμβάνεται από τον πλαίσιο με κλήρωση και κάθε μονάδα του πληθυσμού έχει ίση πιθανότητα με τις άλλες να περιληφθεί στο δείγμα. Αυτό το δείγμα καλείται τυχαιό δείγμα πιθανότητας.

Η απλή τυχαιά δειγματοληψία είναι βέβαια μια αποτελεσματική διαδικασία όταν ο πληθυσμός δεν είναι μεγάλος και όταν είναι σχετικά εύκολο και όχι δαπανηρό να βρούμε τις μονάδες δειγματοληψίας. Επίσης θα ήταν μια πρακτικά εφαρμόσιμη διαδικασία για μεγάλους πληθυσμούς των οποίων οι μονάδες είναι συγκεντρωμένες σε μια μικρή περιοχή.

Ένα μειονέκτημα εφαρμογής της απλής τυχαιάς δειγματοληψίας σε μεγάλους πληθυσμούς είναι ότι οι μονάδες του πληθυσμού πρέπει να έχουν απαριθμηθεί. Θεωρητικά μπορούμε να λάβουμε απλά τυχαιά δείγματα με απαρίθμηση όλων εκείνων που έχουν δικαίωμα ψήφου στις βουλευτικές εκλογές, όλων των ελληνικών νοικοκυριών, όλων των καταστημάτων λιανικών πωλήσεων σε μια περιοχή κ.ο.κ. Μια τέτοια διαδικασία, όμως, θα ήταν πρακτικά δαπανηρή και θα μπορούσε ακόμη και να είναι πρακτικά μη εφαρμόσιμη (αδύνατη). Στην πραγματικότητα, πολλά οικονομικά δεδομένα δεν είναι δυνατόν να χρησιμοποιηθούν αποτελεσματικά, δια της χρησιμοποίησεως της απλής τυχαιάς δειγματοληψίας. Συνήθως, αντί της χρησιμοποίησεως απλής τυχαιάς δειγματοληψίας, εφαρμόζουμε σχέδια περιορισμένου τυχαιού δείγματος. Η μεγαλύτερη αποτελεσματικότητα και η επιθυμία να έχουμε ένα δείγμα πιθανότητας, όταν είναι αδύνατο να χρησιμοποιήσουμε απλή τυχαιά επιλογή, είναι οι βασικοί λόγοι για την προτίμηση των σχεδίων περιορισμένης τυχαιάς δειγματοληψίας.

Τέλος, η μέθοδος της απλής τυχαία δειγματοληψίας παρουσιάζει τα εξής πλεονεκτήματα και μειονεκτήματα :

Πλεονεκτήματα

1. τα αποτελέσματα της απλής τυχαίας δειγματοληψίας είναι αντικειμενικά. Και αυτό οφείλεται στο γεγονός ότι το δείγμα λαμβάνεται με κλήρωση.
2. Επειδή το δείγμα σε αυτή την μέθοδο είναι τυχαίο, επομένως μπορεί να μετρηθεί το μέγεθος του δειγματοληπτικού σφάλματος.
3. Δίνει συνήθως καλές εκτιμήσεις για τα διάφορα χαρακτηριστικά του πληθυσμού με χαμηλό κόστος.
4. Οι εκτιμήσεις των χαρακτηριστικών είναι αμερόληπτες εκτιμήτριες των παραμέτρων με αμερόληπτες εκτιμήτριες διασπορών.

Μειονεκτήματα

1. παρουσιάζει μεγάλη διασπορά σε σχέση με τις άλλες μεθόδους .

ΚΕΦΑΛΑΙΟ	Συστηματική δειγματοληψία
2 .	

2 . 1 Εισαγωγή

Η συστηματική δειγματοληψία (systematic sampling) είναι μια μέθοδος επιλογής δείγματος η οποία χρησιμοποιείται συχνά στην πράξη. Για την εφαρμογή της συστηματικής δειγματοληψίας απαιτείται κατάλληλο δειγματοληπτικό πλαίσιο, δηλαδή κατάλογος στον οποίο να είναι καταχωρημένες όλες οι δειγματοληπτικές μονάδες, ώστε να δίνεται η ίδια ευκαιρία επιλογής, σε όλες τις μονάδες του ερευνώμενου πληθυσμού.

Η συστηματική δειγματοληψία μπορεί να εφαρμοσθεί και σε άγνωστους πληθυσμούς, όποτε δεν απαιτείται δειγματοληπτικό πλαίσιο. Π.χ. αν θέλουμε να εκτιμήσουμε τον αριθμό των ατόμων που εισέρχονται σε ένα νοσοκομείο ημερησίως κ.ο.κ. μπορούμε να εφαρμόσουμε την συστηματική δειγματοληψία.

2 . 2 Περιγραφή της μεθόδου της συστηματικής δειγματοληψίας

Η εφαρμογή της συστηματικής δειγματοληψίας γίνεται ως εξής :

1 . Έστω ότι N είναι οι μονάδες ενός πληθυσμού, οι οποίες είναι καταχωρημένες σε έναν κατάλογο και αριθμημένες από 1 έως N και ότι θέλουμε να επιλέξουμε n μονάδες.

2 . Εάν $n =$ μέγεθος του δείγματος που θέλουμε να επιλέξουμε από τον πληθυσμό N . Έτσι η συστηματική δειγματοληψία είναι μια διαδικασία

Από τον πίνακα φαίνεται ότι ο πληθυσμός έχει διαιρεθεί σε k μεγάλες μονάδες δειγματοληψίας, κάθε μια από τις οποίες περιέχει $\langle n \rangle$ από τις αρχικές μονάδες. Η διαδικασία επιλογής συνίσταται στην επιλογή μιας από αυτές τις μονάδες δειγματοληψίας κατά τυχαίο τρόπο. Έτσι η συστηματική δειγματοληψία συνίσταται στην επιλογή μιας εκ των k μονάδων δειγματοληψίας, η οποία και καθορίζει το συγκεκριμένο δείγμα.

Ένα συστηματικό δείγμα, είναι ένα απλό τυχαίο δείγμα μιας μονάδας δειγματοληψίας (συσσωρευτικής μονάδας) από ένα πληθυσμό k μονάδων δειγματοληψίας (συσσωρευτικών μονάδων).

Για να γίνει κατανοητό η συστηματική δειγματοληψία παίρνουμε ένα παράδειγμα .

Εάν 100 μονάδες πρέπει να επιλεγούν στο δείγμα από έναν πληθυσμό $N=1000$ μονάδων, ο λόγος δειγματοληψίας, $\langle k \rangle$, είναι $10 = N/n = 1000/100$. Για να έχουμε μια τυχαία εκκίνηση όταν ο λόγος δειγματοληψίας είναι 10, επιλέγουμε ένα τυχαίο αριθμό από 1 έως 10 δια της χρησιμοποίησης του πίνακα των τυχαίων αριθμών ή άλλων τυχαίων μεθόδων. Εάν ο τυχαίος αριθμός 3 επιλεγεί, αρχίζουμε με αυτόν και περιλαμβανόμενου στο δείγμα κάθε δέκατη μονάδα από αυτόν, δηλαδή, την δέκατη-τρίτη, την εικοστή-τρίτη, την τριακοστή-τρίτη κ.ο.κ. Η απλότητα στην σχεδίαση είναι ένα βασικό πλεονέκτημα της συστηματικής δειγματοληψίας. Επίσης είναι δυνατόν να διαπιστωθεί ότι μεγαλύτερη αποτελεσματικότητα στην συστηματική δειγματοληψία μπορεί να επιτευχθεί εάν [1] οι μονάδες που είναι κοντά η μια στην άλλη, έχουν μεγαλύτερη ομοιομορφία από ότι [2] οι μονάδες οι οποίες είναι μακριά η μια της άλλης . Αυτές οι προϋποθέσεις είναι ακριβώς οι ίδιες με εκείνες της στρωματοποιημένης δειγματοληψίας .

2. 3 Διακύμανση της εκτιμήσεως του μέσου

Στην ανάλυση που θα ακολουθήσει το σύμβολο y_{ij} δηλώνει την j μονάδα του i συστηματικού δείγματος, έτσι $j=1, 2, \dots, n$ και $i=1, 2, \dots, k$. Ο μέσος του i δείγματος συμβολίζεται με \bar{y}_i .

Δείγμα :	1	2	...	i	...	k
	y_{11}	y_{21}	...	y_{i1}	...	y_{k1}
	y_{12}	y_{22}	...	y_{i2}	...	y_{k2}

	y_{1j}	y_{2j}	...	y_{ij}	...	y_{kj}

	y_{1n}	y_{2n}	...	y_{in}	...	y_{kn}
Μέσοι :	\hat{y}_1	\hat{y}_2	...	\hat{y}_i	...	\hat{y}_k

Ο δειγματικός μέσος \hat{y}_i . Κατά την συστηματική δειγματοληψία αποτελεί αμερόληπτη εκτίμηση του μέσου του πληθυσμού \bar{Y} .

$$E(\hat{y}_i) = \frac{1}{k} \sum_{l=1}^k \hat{y}_i = \frac{\sum_{l=1}^k n \hat{y}_i / nk = Y/N = \bar{Y}}$$

Η διακύμανση του δειγματικού μέσου θα είναι :

$$V(\hat{y}_i) = \frac{1}{k} \sum_{l=1}^k (y_i - \bar{Y})^2$$

Η διακύμανσή αυτή δεν είναι δυνατόν να εκτιμηθεί εκ των στοιχείων ενός μόνο δείγματος, ενώ στην περίπτωση της απλής τυχαίας δειγματοληψίας είναι δυνατόν.

ΘΕΩΡΗΜΑ

Η διακύμανση του μέσου ενός συστηματικού δείγματος είναι :

$$V(\hat{y}_i) = \frac{N-1}{N} S^2 - \frac{k(n-1)}{N} S_w^2$$

Όπου

$$S^2_w = \frac{1}{K(n-1)} \sum_{j=1}^k \sum_{i=1}^n (y_{ij} - \bar{y}_i)^2$$

Είναι η διακύμανση μεταξύ των μονάδων που βρίσκονται στο ίδιο συστηματικό δείγμα. Ο παρονομαστής αυτής της διακυμάνσεως, $k(n-1)$, προκύπτει από τους συνήθεις κανόνες της αναλύσεως διακυμάνσεως : Κάθε ένα από τα k δείγματα συνεισφέρει $(n-1)$ βαθμούς ελευθερίας στο άθροισμα τετραγώνων του αριθμητή .

ΠΟΡΙΣΜΑ

Ο μέσος ενός συστηματικού δείγματος είναι περισσότερο ακριβής από τον μέσο του απλού τυχαίου δείγματος εάν και μόνο εάν

$$S^2_w > S^2$$

Το συμπέρασμα αυτό, δείχνει ότι όσο μεγαλύτερη είναι η τιμή της διακυμάνσεως S^2_w τόσο μικρότερη γίνεται η διακύμανση $V(\bar{y}_i)$, δηλαδή τόσο ακριβέστερη είναι η εκτίμηση του μέσου του πληθυσμού. Αυτό σημαίνει ότι τα διάφορα συστηματικά δείγματα τα οποία μπορούν να καταρτισθούν με διάστημα δειγματοληψίας k , πρέπει να περικλείουν (κάθε ένα) όσο το δυνατόν μη ομοιογενείς μονάδες ως προς το ερευνώμενο χαρακτηριστικό. Εάν η διακύμανση εντός των συστηματικών δειγμάτων είναι μικρή ή μηδέν, τότε η συστηματική δειγματοληψία πρέπει να αποφεύγεται, διότι συνεπάγεται ανακριβείς εκτιμήσεις .

Μια περίπτωση που η διακύμανση των συστηματικών δειγμάτων ισούται με μηδέν, είναι όταν οι μονάδες του πληθυσμού παρουσιάζουν περιοδικότητα, ως προς το μέγεθος του ερευνώμενου χαρακτηριστικού, με συνέπεια κάθε παρατήρηση να είναι ίδια εντός του συστηματικού δείγματος, έτσι το σύνολο των παρατηρήσεων δίνει την ίδια πληροφορία που θα έδινε και μια μόνο παρατήρηση.

2.4 Εκτίμηση του μέσου και συνολικού πληθυσμού

Η εκτίμηση του μέσου πληθυσμού \bar{Y} από ένα συστηματικό δείγμα δίνεται από τον τύπο

$$\hat{y}_{sy} = \sum_{j=1}^n y_{ij}/n$$

Η διακύμανση του \hat{y}_{sy} δεν μπορεί να εκτιμηθεί από τα στοιχεία ενός μόνο συστηματικού δείγματος, εκτός εάν ο πληθυσμός είναι τυχαίος. Ένας πληθυσμός είναι τυχαίος όταν οι μονάδες του είναι διατεταγμένες τυχαία. Οι μονάδες ενός συστηματικού δείγματος που επιλέγονται από έναν τυχαίο πληθυσμό, αναμένονται να είναι ετερογενείς. Όταν το N είναι μεγάλο η διακύμανση του \hat{y}_{sy} είναι περίπου ίση με τη διακύμανση του \hat{y} που βασίζεται στην απλή τυχαία δειγματοληψία, οπότε μπορούμε να χρησιμοποιήσουμε τον τύπο :

$$V(\hat{y}_{sy}) = \frac{s^2}{n} \frac{(N-n)}{N}$$

Η εκτίμηση του συνολικού πληθυσμού Y είναι

$$\hat{Y} = \sum_{j=1}^n y_{ij} \frac{N}{n}$$

Όπου N/n ο συντελεστής αναγωγής, δηλαδή το αντίστροφο του κλάσματος δειγματοληψίας.

Η διακύμανση του \hat{Y}

$$V(\hat{Y}) = N^2 \frac{s^2}{n} \frac{(N-n)}{N}$$

Τέλος η εκτίμηση ποσοστού ενός πληθυσμού (population proportion) από ένα συστηματικό δείγμα, όταν ο πληθυσμός είναι τυχαία, γίνεται όπως και στην περίπτωση της απλής τυχαίας δειγματοληψίας.

2.5 Επαναλαμβανόμενη συστηματική δειγματοληψία

Σε προηγούμενο παράγραφο έχουμε αναφέρει ότι δεν μπορεί να βρεθεί αμερόληπτη εκτίμηση της διακύμανσης του δειγματικού μέσου \hat{y}_{sy}

μόνο από ένα συστηματικό δείγμα, εκτός αν ο πληθυσμός είναι τυχαίος. Μια εναλλακτική μέθοδος, η οποία μπορεί να χρησιμοποιηθεί ώστε εξασφαλισθεί μια ικανοποιητική εκτίμηση της διακυμάνσεως $V(\bar{y}_s)$ είναι η επαναλαμβανόμενη συστηματική δειγματοληψία .

Η επαναλαμβανόμενη συστηματική δειγματοληψία προϋποθέτει τη συλλογή περισσότερων του ενός συστηματικών δειγμάτων. Δηλαδή, αντί να πάρουμε ένα δείγμα μεγέθους n , με διάστημα δειγματοληψίας $k=N/n$, παίρνουμε n συστηματικά δείγματα μεγέθους n/n , με διάστημα δειγματοληψίας $k'=kn$. Το τυχαίο ξεκίνημα για κάθε ένα από τα n συστηματικά δείγματα, επιλέγεται με τυχαίο τρόπο μεταξύ των πρώτων k' αριθμών (1 έως k'). Π.χ. εάν έχουμε έναν πληθυσμό μεγέθους $N=800$ και πάρουμε ένα συστηματικό δείγμα μεγέθους $n=80$, με διάστημα δειγματοληψίας $k=800/80=10$ και με τυχαίο τρόπο ξεκίνημα έναν αριθμό μεταξύ του 1 και 10, θα επιλέγουν 80 μονάδες. Αντί να πάρουμε ένα συστηματικό δείγμα, για την επιλογή των 80 μονάδων, παίρνουμε δέκα ($n=10$) με διάστημα δειγματοληψίας $k'=k'n=10*10=100$. Στη συνέχεια παίρνουμε 10 τυχαίους αριθμούς μεταξύ του 1 και 100. Έστω ότι οι αριθμοί αυτοί είναι :

5, 16, 22, 34, 45, 57, 64, 72, 85, 89 .

οι οποίοι αποτελούν το τυχαίο ξεκίνημα για τα 10 συστηματικά δείγματα, όπως φαίνεται στο παρακάτω πίνακα. Ο δεύτερος αριθμός σε κάθε δείγμα βρίσκεται αν προσθέσουμε στο τυχαίο ξεκίνημα τον αριθμό 100, ο τρίτος αν προσθέσουμε στο δεύτερο τον αριθμό 100, κ.ο.κ.

ΠΙΝΑΚΑΣ
Επιλογή επαναλαμβανόμενων
συστηματικών δειγμάτων

Τυχαίο Ξεκίνημα	Δεύτερος Αριθμός Δείγματος	Τρίτος Αριθμός Δείγματος	...	Όγδοος Αριθμός δείγματος
5	105	205	...	705
16	116	216	...	716
22	122	222	...	722
34	134	234	...	734
45	145	245	...	745
57	157	257	...	757
64	164	264	...	764
72	172	272	...	772
85	185	285	...	785
89	189	289	...	789

Η επιλογή συνήθως δέκα συστηματικά δειγμάτων επιτρέπει τον υπολογισμό αρκετών δειγματικών μέσων, ώστε να εξασφαλισθεί μια ικανοποιητική εκτίμηση της διακύμανσης .

2.6 Συμπέρασμα

Εάν οι δειγματοληπτούμενοι πληθυσμοί είναι επαρκώς μεγάλοι, η συστηματική δειγματοληψία μπορεί τις περισσότερες φορές να αναμένεται να αποδώσει αποτελέσματα τα οποία είναι όμοια με εκείνα της αναλογικής στρωματοποιημένης δειγματοληψίας. Ένεκα της απλότητας της στον σχεδιασμό, η συστηματική δειγματοληψία μπορεί να προτιμηθεί της στρωματοποιημένης δειγματοληψίας εάν ο πληθυσμός μπορεί εύκολα να τοποθετηθεί σε κάποια τακτική διαρρύθμιση (απαρίθμηση). Όταν π.χ. μας δίνεται η λίστα των κερδών επιχειρήσεων, μπορούμε να τα ανακατατάξουμε (τα κέρδη) σύμφωνα με το μέγεθος τους. Η εφαρμογή της συστηματικής δειγματοληψίας σε αυτή την περίπτωση, πράγματι εξασφαλίζεται αναλογική αντιπροσώπευση όλων των τάξεων (μέγεθος κέρδους) χωρίς στρωματοποίηση.

Ένα συστηματικό δείγμα τείνει να είναι περισσότερο αντιπροσωπευτικό από ότι ένα απλό τυχαίο δείγμα του ίδιου μεγέθους εάν όπως, έχει ήδη αναφερθεί ανωτέρω, τα στοιχεία στον πληθυσμό τείνουν να, (ή μπορούν να αναδιαταχθούν για να) μοιάζουν με άλλα στοιχεία που βρίσκονται κοντά περισσότερο από ότι αυτά μοιάζουν με άλλα που βρίσκονται μακριά. Για παράδειγμα, εάν πρόκειται να εκτιμήσουμε το μέσο οικογενειακό εισόδημα στην Αθήνα, μπορούμε να χρησιμοποιήσουμε την συστηματική δειγματοληψία και να επιλέξουμε κάθε εκατοστό νοικοκυριό στο δείγμα μας. Το δείγμα που θα προκύψει μπορεί να είναι περισσότερο αντιπροσωπευτικό από ότι ένα απλό τυχαίο δείγμα. Οι άνθρωποι της ίδιας οικονομικής καταστάσεως, τείνουν να ζουν στην ίδια περιοχή, συνεπώς η συστηματική δειγματοληψία θα δώσει περισσότερο κατάλληλη αντιπροσώπευση σε κάθε εισοδηματική τάξη, ακόμη και αν η απλή τυχαία δειγματοληψία, μπορεί να επιλέξει ένα σχετικά μεγάλο αριθμό νοικοκυριών με τα ίδια εισοδήματα από μια ορισμένη περιοχή .

Η συστηματική δειγματοληψία γίνεται λιγότερο αντιπροσωπευτική από ότι η απλή τυχαία δειγματοληψία εάν εφαρμοσθεί σε πληθυσμούς που παρουσιάζουν αφανείς περιοδικότητες. Για παράδειγμα, εάν οι πωλήσεις κάθε έβδομης ημέρας του έτους αποτελέσουν το δείγμα, τότε το δείγμα θα περιέχει, ας πούμε, όλες τις Δευτέρες ή όλες τις Παρασκευές. Εάν υπάρχει

ένα ορισμένο επαναλαμβανόμενο εβδομαδιαία υπόδειγμα στις πωλήσεις (πράγμα πολύ συνηθισμένο), το δείγμα μας δεν είναι αντιπροσωπευτικό όλων των πωλήσεων για ολόκληρο το έτος και συνεπώς τα δειγματικά αποτελέσματα μπορεί να είναι σοβαρά μεροληπτικά. Παρομοίως, μεροληψίες θα προκύψουν εάν πρόκειται να ελέγξουμε την ποιότητα ενός προϊόντος, παραγωγικής διαδικασίας εν σειρά, παίρνοντας κάθε 50^η μονάδα για παρατήρηση, όταν η μηχανή συμβαίνει να έχει ένα ελάττωμα το οποίο παράγει ατέλειες σε κάθε πενηκοστή μονάδα παραγωγής .

Μέχρι τώρα έχουμε αναφέρει διάφορες τυχαίες διαδικασίες, σαν ανεξάρτητα μεταξύ τους σχέδια. Στην πράξη, συχνά συνδυάζουμε δύο η περισσότερες από αυτές τις μεθόδους, σε ένα απλό σχέδιο. Για παράδειγμα, στην δειγματοληψία για μια εκτίμηση του εργατικού δυναμικού, το γραφείο απογραφών χρησιμοποιεί ένα σχέδιο το οποίο περιλαμβάνει όλους τους τύπους της τυχαίας δειγματοληψίας και άλλες επινοήσεις. Σε μια τέτοια έρευνα στην Ελλάδα επελέγησαν κατά αρχήν 68 γεωγραφικές περιοχές. Κατόπιν, περίπου 2000 πρωταρχικές μονάδες (συσσωρεύσεις που είναι ολόκληροι νομοί ή γειτονικοί νομοί) ομαδοποιήθηκαν σε 55 στρώματα. Στη συνέχεια, μια πρωταρχικές μονάδα επελέγη από κάθε στρώμα. Τέλος, αυτές οι πρωταρχικές μονάδες διαιρέθηκαν σε στοιχειώδεις μονάδες (νοικοκυριά) που αποτέλεσαν το τελικό δείγμα. Σε κάθε στάδιο, μέθοδοι πιθανότητας χρησιμοποιήθηκαν για την επιλογή ούτως ώστε οι κατάλληλοι τύποι να χρησιμοποιηθούν για την μέτρηση της αξιοπιστίας των δειγματικών αποτελεσμάτων.

Η συστηματική δειγματοληψία έχει τα εξής πλεονεκτήματα έναντι της απλής τυχαίας δειγματοληψίας :

1. Η επιλογή του δείγματος και η εφαρμογή της συστηματικής δειγματοληψία είναι ευκολότερη για τους ερευνητές και επομένως υπόκειται λιγότερο σε λάθη που οφείλονται σε αυτόν που παίρνει την συνέντευξη. Δηλαδή παρουσιάζει μικρότερο αριθμό σφαλμάτων συλλογής.
2. Με την εφαρμογή της συστηματικής δειγματοληψίας, περιορίζεται η μεροληπτική επιλογή δειγματοληπτικών μονάδων. Ειδικότερα όταν η επιλογή γίνεται απ'ευθείας από τον ερευνητή κατά την διενέργεια της έρευνας.
3. Η συστηματική δειγματοληψία μπορεί να δώσει περισσότερες πληροφορίες, σε

δεδομένο κατά μονάδα κόστος, διότι ένα συστηματικό δείγμα κατά κανόνα είναι διεσπαρμένο με μεγαλύτερη ομοιογένεια στον ερευνώμενο πληθυσμό.

4. Η συστηματική δειγματοληψία είναι παρά πολύ χρήσιμη όταν ο ερευνώμενος πληθυσμός είναι μεγάλο.
5. τέλος, η επαναλαμβανόμενη δειγματοληψία δίνει σε σχέση με τους μεθόδους δειγματοληψίας που έχουμε αναφέρει μέχρι τώρα μεγαλύτερη ακρίβεια.

Η συστηματική δειγματοληψία έχει ένα σημαντικό μειονέκτημα έναντι της απλής τυχαίας δειγματοληψίας, στις περιπτώσεις που οι μονάδες του ερευνώμενου πληθυσμού παρουσιάζουν περιοδικότητα. Π.χ. αν οι μονάδες που έχουν επιλεγεί σε μια συστηματική δειγματοληψία .

$$y_2, y_{12}, y_{22}, y_{32} \dots \dots$$

παρουσιάζουν περιοδικότητα, τότε θα παρέχουν την ίδια πληροφορία και θα έχουμε

$$y_2 = y_{12} = y_{22} = y_{32} = \dots \dots = \bar{y}$$

δηλαδή η εκτίμηση του μέσου \bar{y} του πληθυσμού από δείγμα μεγέθους n συνεπάγεται την ίδια ακρίβεια με μια εκτίμηση που θα βασίζεται σε δείγμα μεγέθους $n=1$. Επομένως θα έχουμε υπερεκτίμηση ή υποεκτίμηση των υπό μελέτη χαρακτηριστικών. Στην περίπτωση αυτή δεν πρέπει να εφαρμόζεται η συστηματική δειγματοληψία για την επιλογή του δείγματος .

Ένα άλλο μειονέκτημα της συστηματική δειγματοληψία είναι ότι όταν ο πληθυσμός είναι άγνωστος δεν μπορεί να γίνει ακριβός υπολογισμός του βήματος k . Αν το k υπερεκτιμηθεί (αυτό θα συμβεί αν το N υπερεκτιμηθεί) θα ληφθεί μικρότερο δείγμα από αυτό που θέλουμε. Σε πολλές περιπτώσεις υπάρχει η δυνατότητα να επιστρέψουμε και να λάβουμε - το υπόλοιπο συστηματικό δείγμα. - Υπάρχουν όμως και περιπτώσεις, που κάτι τέτοιο είναι αδύνατο. Όταν π.χ. ρωτάμε τα άτομα (1 κάθε k) που περνάνε από κάποιο σημείο τις ώρες κυκλοφοριακής αιχμής .

Το τρίτο μειονέκτημα είναι ότι από τα αποτελέσματα ενός συστηματικού δείγματος δεν υπάρχει αξιόπιστη μέθοδος εκτιμήσεως του

τυπικού σφάλματος του δειγματικού μέσου. Τα διάφορα εγχειρίδια επί της δειγματοληψίας δίνουν διάφορους τύπους, οι οποίοι μπορούν να δοκιμασθούν : Κάθε τύπος ισχύει για ένα ορισμένο τύπο πληθυσμό, άλλα ένας τύπος μπορεί να χρησιμοποιηθεί με εμπιστοσύνη μόνο εάν έχουμε μαρτυρία ότι ο πληθυσμός είναι της μορφής για την οποία ο σχετικός τύπος εφαρμόζεται. Πάντως, η συστηματική δειγματοληψία συχνά αποτελεί μέρος ενός περισσότερο σύνθετου σχεδίου δειγματοληψίας στο οποίο είναι δυνατόν να λάβουμε αμερόληπτες εκτιμήσεις των σφαλμάτων δειγματοληψίας .

ΚΕΦΑΛΑΙΟ	Στρωματοποιημένη τυχαία δειγματοληψία
3.	

3.1 Εισαγωγή.

Η στρωματοποιημένη τυχαία δειγματοληψία (stratified random sampling), είναι μια από τις περιορισμένες τυχαίες μεθόδους, κατά την οποία με τη χρησιμοποίηση διαθέσιμων πληροφοριών που αφορούν στα δεδομένα, προσπαθούμε να σχεδιάσουμε ένα περισσότερο αποτελεσματικό δείγμα από εκείνο, που θα παίρναμε δια της απλής τυχαίας διαδικασίας. Η διαδικασία της στρωματοποίησης απαιτεί όπως ο πληθυσμός διαιρεθεί σε ομοιογενείς ομάδες ή τάξεις που ονομάζονται στρώματα (strata). Στη συνέχεια, ένα δείγμα μπορεί να ληφθεί από κάθε στρώμα δια των απλών τυχαίων μεθόδων και το δείγμα που προκύπτει τελικά, ονομάζεται στρωματοποιημένο δείγμα .

Υπάρχουν τρία στάδια στην μέθοδο της στρωματοποιημένης δειγματοληψίας, τα οποία είναι :

1. ο πληθυσμός διαιρείται σε ένα αριθμό υποπληθυσμών, που ονομάζονται στρώματα.
2. ένα δείγμα λαμβάνεται ανεξάρτητα σε κάθε στρώμα .
3. Σαν μια εκτίμηση του πληθυσμιακού μέσου, χρησιμοποιούμε

$$\hat{y}_{st} = \frac{\sum N_h \hat{y}_h}{N}$$

Όπου

- N_h είναι ο συνολικός αριθμός των μονάδων δειγματοληψίας στο h στρώμα ,
- \hat{y}_h είναι ο δειγμάτικος μέσος στο h στρώμα
- $N = \sum N_h$ είναι το μέγεθος του πληθυσμού .

Γενικά

Στην στρωματοποιημένη τυχαία δειγματοληψία ο ερευνώμενος πληθυσμός χωρίζεται σε k υποπληθυσμούς (στρώματα)

$$N_1, N_2, N_3, \dots, N_k$$

Ούτως ώστε αφ'ένος μεν να μην υπάρχει επικάλυψη μεταξύ των υποπληθυσμών (στρωμάτων) αφ'ετέρου δε το σύνολο των μονάδων των υποπληθυσμών (στρωμάτων) να συνιστούν τον αρχικό πληθυσμό, δηλαδή :

$$N_1 + N_2 + N_3 + \dots + N_k = N$$

Στη συνέχεια επιλέγεται μέσα από κάθε στρώμα με τυχαίο τρόπο ένα δείγμα , το άθροισμα των δειγμάτων αυτών δίνει το συνολικό δείγμα .

Με την στρωματοποιημένη δειγματοληψία επιδιώκεται να χωρισθεί ένας ανομοιογενής πληθυσμός σε ομοιογενείς υποπληθυσμούς. Με τον τρόπο αυτό επιτυγχάνεται μεγαλύτερη ακρίβεια στις εκτιμήσεις των χαρακτηριστικών του ερευνώμενου πληθυσμού, διότι η μετρήσεις που προκύπτουν από ομοιογενή στρώματα διαφέρουν ελάχιστα μεταξύ τους. Επίσης εξασφαλίζεται η αντιπροσωπευτικότητα του ερευνώμενου πληθυσμού, επειδή στο συνολικό δείγμα συμπεριλαμβάνονται μονάδες από όλα τα στρώματα. Ενώ αντίθετα στην απλή τυχαία δειγματοληψία υπάρχει περίπτωση να παραληφθούν μονάδες που αποτελούν ένα ομοιογενές στρώμα, με αποτέλεσμα να μην έχουμε καθόλου πληροφόρηση για την κατηγορία αυτή. Π.χ. αν θέλουμε να διενεργήσουμε μια δειγματοληπτική σωματομετρική έρευνα του πληθυσμού, πρέπει να χωρίσουμε τον πληθυσμό σε ομοιογενή στρώματα με βάση τις σωματικές διαστάσεις των ατόμων κατά φύλο και ηλικία δηλαδή άνδρες ηλικίας μέχρι 65 ετών, άνδρες ηλικίας άνω των 65 ετών, γυναίκες ηλικίας μέχρι

65 ετών, γυναίκες ηλικίας άνω των 65 ετών, αγόρια ηλικίας μέχρι 5 ετών κ.ο.κ. Στη συνέχεια να πάρουμε ένα δείγμα από κάθε στρώμα, έτσι όλες οι ομάδες των ατόμων θα αντιπροσωπεύονται στην έρευνα. Εάν δεν γίνει στρωματοποίηση του πληθυσμού κατά φύλο και ηλικία υπάρχει περίπτωση να μην συμπεριληφθεί για παράδειγμα, στο δείγμα κανένας άνδρας ηλικίας πάνω των 65 ετών, με αποτέλεσμα να μην συγκεντρωθούν πληροφορίες για την ομάδα αυτή των ατόμων .

Οι βασικές αιτίες για την χρησιμοποίηση της στρωματοποιημένης τυχασίας δειγματοληψίας είναι οι εξής:

- Η στρωματοποίηση μπορεί να δώσει μικρότερο σφάλμα εκτιμήσεως από ότι η απλή τυχασία δειγματοληψία του ίδιου μεγέθους δείγματος .
- Το κόστος δειγματοληψίας μπορεί να μειωθεί σημαντικά με τη στρωματοποίηση του πληθυσμού σε ομοιογενείς υποπληθυσμούς, διότι είναι δυνατόν να περιορισθεί το μέγεθος του δείγματος στους επιμέρους ομοιογενείς υποπληθυσμούς .
- Το ενδιαφέρον για εκτιμήσεις των παραμέτρων του πληθυσμού κατά υποπληθυσμό χωριστά .
- Η στρωματοποίηση δίνει τη δυνατότητα αντιμετώπισης των διαφορών που υπάρχουν μεταξύ πληθυσμιακών ομάδων. Π.χ. η συμπεριφορά των μεγάλων επιχειρήσεων , στο σύνολο των οικονομικών τους δραστηριοτήτων, είναι διαφορετική από αυτή των μικρών επιχειρήσεων. Οι απαιτήσεις των ατόμων που ζουν σε συλλογικές κατοικίες (νοσοκομεία, οικοτροφεία κ.λ.π.) είναι διαφορετικές από αυτές των ατόμων που ζουν σε κανονικές κατοικίες κ.ο.κ.

3 . 2 Ορισμοί – Συμβολισμοί

Έστω ότι ένας πληθυσμός μεγέθους N διαιρείται σε k στρώματα. Το γράμμα h δηλώνει την τάξη του στρώματος ($h = 1, 2, \dots, k$) και με το i την τάξη της μονάδας δειγματοληψίας εντός του στρώματος. Περαιτέρω για ένα συγκεκριμένο στρώμα θα έχουμε :

- N_h μέγεθος του στρώματος h
- n_h μέγεθος του δείγματος για το στρώμα h
- n μέγεθος του δείγματος για όλο τον πληθυσμό
- y το εξεταζόμενο χαρακτηριστικό
- y_{hi} τιμή χαρακτηριστικού της i μονάδας
- \bar{Y}_h ο μέσος του πληθυσμού του στρώματος h

$$\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} y_{hi}$$

- \bar{y}_h δειγματικός μέσος του στρώματος h

$$\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$$

- Y_h το ολικό του στρώματος h

$$Y_h = N_h \bar{Y}_h$$

- y_h δειγματικό ολικό του στρώματος h

$$y_h = n_h \bar{y}_h$$

- S_h^2 πληθυσμιακή διασπορά του στρώματος h

$$S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2$$

s^2_h δειγματική διασπορά του στρώματος h

$$s^2_h = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2$$

W_h βάρος του στρώματος h στον πληθυσμό.

$$W_h = \frac{N_h}{N}$$

w_h βάρος του στρώματος h στο δείγμα.

$$w_h = \frac{n_h}{n}$$

f_h διόρθωση για πεπερασμένους πληθυσμούς στο στρώμα h

$$f_h = \frac{n_h}{N_h}$$

Διευκρινίζεται ότι τα δείγματα στα επί μέρους στρώματα επιλέγονται ανεξάρτητα μεταξύ τους. Ο δείκτης f_h δηλώνει την στρωματοποιημένη δειγματοληψία.

3.3 Κατανομή δείγματος στα στρώματα.

Μετά τον διαχωρισμό του ερευνώμενου πληθυσμού σε ομοιογενή στρώματα, γίνεται κατανομή του δείγματος (allocation of the sample) μεταξύ των στρωμάτων. Στη συνέχεια επιλέγεται το δείγμα με τυχαίο τρόπο μέσα από κάθε στρώμα, είτε με τη μέθοδο της συστηματικής δειγματοληψίας, είτε με τη μέθοδο της απλής τυχαίας δειγματοληψίας.

Η επιλογή του απαραίτητου μεγέθους δείγματος μέσα σε κάθε στρώμα, αποσκοπεί στο να πετύχουμε εκτιμήτριες με μικρές διακυμάνσεις στο ελάχιστο δυνατό κόστος. Δηλαδή η κατανομή του συνολικού δείγματος μεταξύ των στρωμάτων πρέπει να γίνεται κατά τρόπο ώστε να επιτυγχάνεται καθορισμένο μέγεθος πληροφοριών στο ελάχιστο κόστος .

Με βάση το σκοπό αυτό για την κατανομή δείγματος μεταξύ των στρωμάτων, πρέπει να λαμβάνονται υπόψη οι παρακάτω παράγοντες :

- ο συνολικός αριθμός μονάδων σε κάθε στρώμα
- η μεταβλητικότητα των μονάδων σε κάθε στρώμα και
- το κόστος κατά δειγματοληπτική μονάδα σε κάθε στρώμα

Ο αριθμός των δειγματοληπτικών μονάδων που επιλέγονται σε κάθε στρώμα επιδρά στο μέγεθος της πληροφόρησης. Π.χ. ένα δείγμα 30 μονάδων από πληθυσμό 300 μονάδων δίνει περισσότερες πληροφορίες, από ότι ένα δείγμα 30 μονάδων από πληθυσμό 30.000 μονάδων. Επομένως, πρέπει να λαμβάνεται μεγαλύτερο δείγμα από στρώματα με μεγαλύτερους πληθυσμούς .

Η μεταβλητικότητα επιδρά στην αξιοπιστία των εκτιμήσεων και επομένως πρέπει να λαμβάνεται μεγαλύτερο δείγμα από στρώματα που παρουσιάζουν μεγαλύτερη ανομοιογένεια .

Εάν το κατά μονάδα κόστος μιας δειγματοληπτικής έρευνας διαφέρει από στρώμα σε στρώμα, πρέπει να πάρουμε μικρότερο δείγμα από το στρώμα με υψηλό κόστος, για να κρατήσουμε το κόστος της δειγματοληπτικής έρευνας σε χαμηλά επίπεδα, υπό την προϋπόθεση ότι δεν επηρεάζεται η αξιοπιστία των αποτελεσμάτων .

Η κατανομή του συνολικού δείγματος στα στρώματα πρέπει να γίνεται με μια από τις παρακάτω μεθόδους, προκειμένου να πετύχουμε εκτιμήτριες με μικρές διακυμάνσεις στο ελάχιστο δυνατό κόστος :

- αναλογική κατανομή
- μη αναλογική κατανομή
- αρίστη κατανομή δείγματος

Πολλές φορές η κατανομή του δείγματος, γίνεται αυθαίρετα στα διάφορα στρώματα, δηλαδή το δείγμα που επιλέγεται σε ένα στρώμα είναι ανεξάρτητο από οποιοδήποτε δείγμα άλλου στρώματος. Ο τρόπος αυτός κατανομής του δείγματος είναι αποδεκτός, με την προϋπόθεση ότι εξασφαλίζεται η ελαχιστοποίηση της διακύμανσης των εκτιμήσεων, ώστε να προκύπτουν αξιόπιστα αποτελέσματα .

Το πεδίο εφαρμογής των παραπάνω είναι το εξής :

- η ανάλογος κατανομή εφαρμόζεται όταν οι διακυμάνσεις των διαφόρων στρωμάτων δεν παρουσιάζουν μεγάλες διαφορές μεταξύ τους οπότε αγνοούνται .
- η δυσανάλογος (ή μη αναλογική) κατανομή εφαρμόζεται όταν έστω και μια διακύμανση στρώματος παρουσιάζει σημαντικά μεγαλύτερη διαφορά από τις άλλες. Από το στρώμα που έχει τη μεγαλύτερη διακύμανση λαμβάνονται περισσότερες μονάδες δείγματος από ότι στην ανάλογη .
- η άριστη κατανομή εφαρμόζεται όταν, πλέον της ανομοιογένειας των διακυμάνσεων, ποικίλει το κόστος μονάδος επιλογής από στρώμα σε στρώμα. Από το στρώμα που έχει το μεγαλύτερο κόστος, λαμβάνονται λιγότερες μονάδες δείγματος σε σχέση με τη δυσανάλογο κατανομή.

Εφ' όσον γίνει ο επιμερισμός, οι μονάδες του δείγματος λαμβάνονται από κάθε στρώμα με τρόπο απεριόριστο. Δηλαδή και η απλή τυχαία (απεριόριστος) δειγματοληψία είναι μια δειγματοληψία κατά στρώμα με ένα μόνο, όμως, στρώμα .

3 . 3 . 1 Αναλογική κατανομή δείγματος

Η αναλογική κατανομή δείγματος (proportional allocation) εφαρμόζεται όταν το κόστος κατά μονάδα δείγματος είναι ίδιο και η διακύμανση είναι περίπου ίδια σε όλα τα στρώματα. Με την αναλογική κατανομή, το δείγμα λαμβάνεται αναλογικά προς το μέγεθος των

στρωμάτων του ερευνώμενου πληθυσμού, δηλαδή με σταθερό κλάσμα δειγματοληψίας, το οποίο δίνει αυτοσταθμιζόμενο δείγμα (self-weighting sample) :

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \dots = \frac{n_k}{N_k} = \frac{n}{N}$$

Οπότε ο τύπος προσδιορισμού του δείγματος στο στρώμα i είναι :

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

Έτσι σε ένα σχέδιο αναλογικής στρωματοποιημένης δειγματοληψίας, ο αριθμός των μονάδων που επιλέγονται από κάθε στρώμα είναι αναλογικός του μεγέθους του στρώματος. Για παράδειγμα, εάν ο δειγματοληπτούμενος πληθυσμός διαιρεθεί σε τέσσερα στρώματα, των οποίων τα μεγέθη είναι αντιστοίχως : 10, 20, 30 και 40% του πληθυσμού και ένα δείγμα των 500 μονάδων πρόκειται να επιλεγεί, το επιθυμητό αναλογικό δείγμα μπορεί να ληφθεί κατά τον ακόλουθο τρόπο :

Από	το	πρώτο	στρώμα	:	500(0,10)= 50	μονάδες
-/-	-/-	δεύτερο	-/-	:	500(0,20) =100	-/-
-/-	-/-	τρίτο	-/-	:	500(0,30) =150	-/-
-/-	-/-	τέταρτο	-/-	:	500(0,40) =200	-/-
Μέγεθος δείγματος < n >				:	=500	-/-

Είναι φανερό από το ανωτέρω παράδειγμα ότι η αναλογική στρωματοποίηση αποδίδει ένα δείγμα το οποίο αντιπροσωπεύει τον δειγματοληπτούμενο πληθυσμό με την σχέση της ποσοστιαίας συμμετοχής κάθε στρώματος στον πληθυσμό. Αυτή η διαδικασία είναι ικανοποιητική εάν δεν υπάρχει μεγάλη διαφορά στην διασπορά από στρώμα σε στρώμα. Αλλά δεν είναι βέβαια η περισσότερο αποτελεσματική διαδικασία, ειδικά όταν οι διακυμάνσεις διαφέρουν ουσιαστικά στα διάφορα στρώματα, επειδή όπως έχει κατ' επανάληψη τονισθεί, ο περισσότερος σπουδαίος παράγοντας για το μέγεθος του δείγματος δεν είναι το μέγεθος αλλά η διακύμανση του πληθυσμού

3.3.2 Αρίστη κατανομή του δείγματος

Όταν ο ερευνώμενος πληθυσμός παρουσιάζει σε μερικά στρώματα μεγαλύτερη μεταβλητικότητα, είναι απαραίτητο στα στρώματα αυτά να ληφθεί μεγαλύτερο δείγμα, προκειμένου να αυξηθεί η ακρίβεια των εκτιμήσεων. Επίσης αν το κόστος συλλογής πληροφοριών είναι μεγαλύτερο σε μερικά στρώματα, μπορεί να ληφθεί μικρότερο δείγμα, προκειμένου να μειωθεί το κόστος της έρευνας. Δηλαδή όταν η διακύμανση της υπό μελέτη μεταβλητής και το κόστος συλλογής πληροφοριών διαφέρουν σημαντικά από στρώμα σε στρώμα, οι μονάδες δείγματος πρέπει να κατανέμονται στα στρώματα κατά τρόπο ώστε η διακύμανση της εκτίμησης να γίνει ελάχιστη για ένα καθορισμένο κόστος ή να γίνει ελαχιστοποίηση του κόστους για μια καθορισμένη τιμή της διακύμανσης της εκτίμησης .

Η κατανομή αυτή του δείγματος η οποία ονομάζεται αρίστη κατανομή (optimum allocation) επιτυγχάνεται αν το κλάσμα δειγματοληψίας στα διάφορα στρώματα λαμβάνεται αναλογικά προς την τυπική απόκλιση και αντιστρόφως ανάλογα προς την τετραγωνική ρίζα του κόστους κατά μονάδα . Εάν το κόστος κατά μονάδα είναι σταθερό στα στρώματα του ερευνώμενου πληθυσμού, τότε η διακύμανση της εκτίμησης ελαχιστοποιείται, αν το κλάσμα δειγματοληψίας στα στρώματα, λαμβάνεται αναλογικά προς την τυπική απόκλιση .

Οπότε ο τύπος προσδιορισμού του δείγματος στο στρώμα i είναι :

$$n_i = n \frac{N_h S_h}{\sum N_h S_h} = n \frac{W_h S_h}{\sum W_h S_h}$$

Η ειδική αυτή περίπτωση της αρίστης κατανομής ονομάζεται **κατανομή Neyman (Neyman allocation)** .

Στην πράξη υπάρχουν οι παρακάτω δυσκολίες κατά την εφαρμογή της αρίστης κατανομής δείγματος :

- Συνήθως η διακύμανση του πληθυσμού κατά στρώμα είναι άγνωστη. Ένας τρόπος να ξεπεραστεί η δυσκολία αυτή είναι να ληφθεί ένα δείγμα μεγέθους m_i από κάθε στρώμα (δηλαδή διεξαγωγή δοκιμαστικής δειγματοληπτικής έρευνας), για την εκτίμηση της S_i η οποία στη συνέχεια θα χρησιμοποιηθεί για την κατανομή του δείγματος n στα στρώματα. Επίσης μπορεί να χρησιμοποιηθούν στοιχεία από

προηγούμενες έρευνες ή απογραφές με ίδιους ή παρόμοιους πληθυσμούς, ώστε να πετύχουμε κάποιες εκτιμήσεις της διακύμανσης

- Όταν σε μια δειγματοληπτική έρευνα υπάρχουν περισσότερα από ένα χαρακτηριστικά, τότε η κατανομή του δείγματος από διάφορα στρώματα, μπορεί να οδηγήσει στην απώλεια ακρίβειας, δεδομένου ότι το δειγματοληπτικό κλάσμα που είναι άριστο για ένα χαρακτηριστικό μπορεί να μην είναι για ένα άλλο. Το πρόβλημα αυτό μπορεί να ξεπεραστεί με κάποια συμβιβαστική λύση. Το πρώτο βήμα είναι να περιορισθεί ο αριθμός των χαρακτηριστικών που θα εξετασθούν σε έναν σχετικά μικρό αριθμό που πιστεύουμε ότι είναι το πιο σπουδαίο. Στη συνέχεια αν υπάρχουν διαθέσιμα στοιχεία από προηγούμενη έρευνα, μπορούμε να προσδιορίσουμε το δείγμα με την αριστή κατανομή για κάθε χαρακτηριστικό ξεχωριστά. Οπότε από τα μεγέθη των δειγμάτων που έχουν καταταξιωθεί κατά χαρακτηριστικό, λαμβάνουμε τον μέσο όρο αυτών, ο οποίος αποτελεί μια ικανοποιητική συμβιβαστική λύση κατανομής του συνολικού δείγματος

3.3.3 Μη αναλογική κατανομή

Για να πετύχουμε την μέγιστη αποτελεσματικότητα στην στρωματοποίηση, θα δώσουμε μεγαλύτερη αντιπροσώπευση σε ένα στρώμα με μεγάλη διασπορά και μικρότερη αντιπροσώπευση σε ένα στρώμα με μικρή διασπορά. Για παράδειγμα, κατά την διεξαγωγή ενός γκάλοπ για τις βουλευτικές εκλογές, μπορούμε να δημιουργήσουμε τόσα στρώματα όσοι είναι οι νομοί της χώρας. Εάν αυτοί που έχουν δικαίωμα ψήφου σε ένα νομό αποτελούν το 5% όλων των εχόντων δικαίωμα ψήφου στην χώρα, αλλά σύμφωνα με την εμπειρία μας το αποτέλεσμα σ' αυτόν τον νομό είναι εκ των προτέρων γνωστό, μπορούμε να πάρουμε από αυτόν

τον νομό μόνο το 2% ή ακόμη και το 1% του δείγματος μας. Εάν, βέβαια, το αποτέλεσμα σε έναν άλλο νομό χαρακτηρίζεται από μεγάλη αβεβαιότητα, μπορούμε να αποφασίσουμε να δώσουμε σ' αυτόν (τον νομό) πολύ μεγαλύτερη αντιπροσώπευση στο δείγμα απ' ότι το σχετικό μέγεθος του νομού καθορίζει. Ένα δείγμα που λαμβάνεται με αυτές τις προϋποθέσεις είναι ένα μη αναλογικό στρωματοποιημένο δείγμα. Η μη αναλογική στρωματοποιημένη δειγματοληψία περιέχει επίσης διαδικασίες λήψεως ίσου αριθμού μονάδων από κάθε στρώμα ανεξαρτήτως του μεγέθους του στρώματος και χορήγηση μόνο μιας μικρής αντιπροσωπεύσεως σε ένα ή περισσότερα στρώματα των οποίων οι μονάδες είναι πολύ δαπανηρό να ερευνηθούν αλλά παρ' όλα αυτά κάποια αντιπροσώπευση αυτών είναι πολύτιμη.

3.3.4 Σύγκριση μεταξύ ανάλογου και δυσανάλογου κατανομής

Η ανάλογος κατανομή αγνοεί τη διακύμανση των στρωμάτων και επιμερίζει το δείγμα σύμφωνα με τον αριθμό των μονάδων του στρώματος.

Η δυσανάλογος κατανομή, εκτός από το μέγεθος των μονάδων του στρώματος N_i , στον επιμερισμό του δείγματος, λαμβάνει υπ' όψιν και τη μέση απόκλιση τετραγώνου του στρώματος.

Μπορούμε λοιπόν να πούμε, ότι ανάλογος δειγματοληψία πλεονεκτεί έναντι της δυσανάλογου, στο ότι δε χρειάζεται τις διακυμάνσεις των στρωμάτων και μειονεκτεί στο ότι αγνοεί έναν παράγοντα, τη μέση απόκλιση, που θα έπρεπε να το λαμβάνει υπ' όψιν.

Γενικά σε περιπτώσεις πληθυσμών που τα στρώματα τους δεν παρουσιάζουν μεγάλες διαφορές διακυμάνσεων καλύτερη κατανομή είναι η ανάλογος. Εάν όμως οι διακυμάνσεις των στρωμάτων παρουσιάζουν μεταξύ τους σημαντικές διαφορές καλύτερη κατανομή είναι η δυσανάλογος.

3.4 Εκτίμηση μέσου και συνολικού πληθυσμού.

Για τον πληθυσμιακό μέσο ανά μονάδα δειγματοληψίας, η εκτίμηση που χρησιμοποιείται στην στρωματοποιημένη δειγματοληψία είναι \hat{y}_{st} , όπου :

$$\hat{y}_{st} = \frac{\sum N_h \hat{y}_h}{N} = \sum W_h \hat{y}_h$$

Όπου

$$N_1 + N_2 + N_3 + \dots + N_k = N$$

Η εκτίμηση \hat{y}_{st} δεν είναι γενικά η ίδια με τον δείγματικο μέσο. Η διαφορά είναι ότι στον \hat{y}_{st} , οι εκτιμήσεις από τα επί μέρους στρώματα χρησιμοποιούν τα βάρη N_h/N . Προφανώς, ο \hat{y}_{st} συμπίπτει με τον \bar{y} , εάν :

$$\frac{n_h}{n} = \frac{N_h}{N}$$

που αυτό σημαίνει ότι το κλάσμα δειγματοληψίας είναι το ίδιο σε όλα τα στρώματα .

ΘΕΩΡΗΜΑ 1

Έστω ότι, κατόπιν επιλογής ενός δείγματος κατά στρώμα, έχουμε υπολογίσει τις αμερόληπτες εκτιμήσεις των μέσων των στρωμάτων, τότε η συνάρτηση του \hat{y}_{st} είναι αμερόληπτη εκτιμήτρια συνάρτηση του μέσου \bar{Y} του πληθυσμού. Δηλαδή :

$$E(\hat{y}_{st}) = \bar{Y}$$

ΘΕΩΡΗΜΑ 2

Στην στρωματοποιημένη δειγματοληψία, η διακύμανση του \hat{y}_{st} είναι :

$$V(\hat{y}_{st}) = \frac{1}{N^2} \sum_{h=1}^k N_h^2 V(\hat{y}_h) = \sum_{h=1}^k W_h^2 V(\hat{y}_h)$$

Όπου, $V(\hat{y}_h) = E(\hat{y}_h - \bar{Y}_h)^2$

Υπάρχουν δύο περιορισμοί σ' αυτό το θεώρημα :

- y_h πρέπει να είναι μια αμερόληπτη εκτίμηση του Y_h
- τα δείγματα πρέπει να επιλέγονται ανεξάρτητα

ΠΟΡΙΣΜΑ

Η διακύμανση της εκτιμήτριας είναι : $\hat{Y}_{st} = N\hat{y}_{st}$
 Του συνόλου του πληθυσμού είναι :

$$V(\hat{y}_{st}) = \sum_{h=1}^K N^2 h \frac{S^2_h}{n_h} (1-f_h)$$

Συνήθως το ολικό μέγεθος του δείγματος $\langle n \rangle$ κατανέμεται στα επί μέρους στρώματα κατά δύο τρόπους που παρουσιάζονται στους παρακάτω παραγράφους .

3 . 4 . 1 Αναλογική κατανομή (proportional allocation)

Σαν δηλωτικό του τρόπου κατανομής του δείγματος στα επί μέρους στρώματα, συμβολίζουμε την εκτίμηση του πραγματικού μέσου \hat{Y} δια \hat{y}_{prop} στην περίπτωση της αναλογικής στρωματοποιημένης δειγματοληψία, αντί \hat{y}_{st} που δηλώνει οποιαδήποτε κατανομή του δείγματος

ΠΟΡΙΣΜΑ 1

Για αναλογική στρωματοποιημένη δειγματοληψία έχουμε

$$V(\hat{y}_{prop}) = \sum_{h=1}^K W_h \frac{S^2_h}{n} (1-f)$$

Εάν επί πλέον $S^2_h = S^2_w$, δηλαδή οι διακυμάνσεις σε όλα τα στρώματα έχουν την ίδια τιμή, τότε έχουμε

$$V(\hat{y}_{prop}) = \frac{S^2_w (N-n)}{n N}$$

3.4.2 Άριστη κατανομή (optimum allocation).

Η εκλογή των δειγματικών μεγεθών των στρωμάτων γίνεται από τον δειγματολήπτη με αντικειμενικό σκοπό την ελαχιστοποίηση του ολικού κόστους της δειγματοληψίας υπό κάποιον περιορισμό (ανώτατο όριο) στη διασπορά ή την ελαχιστοποίηση της διασποράς υπό κάποιον περιορισμό επάνω στο κόστος. Ο βέλτιστος καταμερισμός ή προσδιορισμός των n_h γίνεται όταν εκλέγονται αυτές οι τιμές των n_h που ελαχιστοποιούν την διασπορά του \hat{y}_{st} και ικανοποιούν επίσης την δοσμένη συναρτησιακή σχέση (εξίσωση) του ολικού κόστους.

Θα εξετάσουμε την απλούστατη περίπτωση που το κόστος είναι γραμμική συνάρτηση των n_h και δίνεται από την σχέση

$$C = c_0 + \sum_{h=1}^k c_h n_h$$

όπου

c_0 τα γενικά πάγια έξοδα

c_h το κόστος κατά μονάδα εκλεγμένη από h

Στην περίπτωση της άριστης κατανομής του δείγματος, συμβολίζουμε διακύμανση της εκτίμησης του \hat{y}_{st} με $V(\hat{y}_{st})$ n_e όταν το κόστος κατά μονάδα είναι ίδιο σε όλα τα στρώματα (κατανομή Neyman) και με $V(\hat{y}_{st})_{opt}$ όταν το κόστος κατά μονάδα διαφέρει μεταξύ των στρωμάτων.

Η διακύμανση της εκτίμησης του \hat{y}_{st} προκύπτει από την σχέση αντικαθιστώντας το:



$$n_h = n \frac{N_h S_h}{\sum N_h S_h} = n \frac{W_h S_h}{\sum W_h S_h}$$

Οπότε έχουμε



$$V(\hat{y}_{st}) = \frac{(\sum_{h=1}^k W_h S_h)^2}{n} - \frac{\sum_{h=1}^k W_h S_h^2}{N}$$

Το

$$n_h = n \frac{W_h S_h \sqrt{c_h}}{\sum W_h S_h \sqrt{c_h}}$$

3.5. Σύγκριση απλής τυχαίας, αναλογικής και άριστη κατανομή

Η στρωματοποιημένη δειγματοληψία, εάν χρησιμοποιείται έξυπνα, σχεδόν πάντοτε καταλήγει σε μικρότερη διακύμανση για τον εκτιμώμενο μέσο ή για το συνολικό μέγεθος, από εκείνη που προκύπτει συγκριτικά από την απλή τυχαία δειγματοληψία. Δεν είναι αλήθεια, πάντως, ότι οποιοδήποτε στρωματοποιημένο τυχαίο δείγμα δίνει μια μικρότερη διακύμανση απ' ό,τι ένα απλό τυχαίο δείγμα. Εάν οι τιμές των n_h είναι μακριά από το άριστο μέγεθος, η στρωματοποιημένη δειγματοληψία μπορεί να έχει μια μεγαλύτερη διακύμανση. Πράγματι, ακόμη και η στρωματοποίηση με άριστη κατανομή για σταθερό συνολικό μέγεθος δείγματος μπορεί να δώσει μια μεγαλύτερη διακύμανση, αν και αυτό το αποτέλεσμα είναι μάλλον θεωρητικό παρά κάτι που μπορεί να συμβεί στην πράξη.

Σε αυτή την παράγραφο θα γίνει μια σύγκριση μεταξύ της απλής τυχαίας δειγματοληψίας και της στρωματοποιημένης τυχαίας δειγματοληψίας με αναλογική και άριστη κατανομή. Αυτή η σύγκριση θα δείξει πως επιτυγχάνεται το κέρδος που οφείλεται στην στρωματοποίηση. Οι διακυμάνσεις των εκτιμώμενων μέσων, συμβολίζονται με V_{ran} , V_{prop} , V_{opt} .

ΘΕΩΡΗΜΑ

Εάν οι όροι n_h / N_h αγνοηθούν

$$V_{ran} \leq V_{prop} \leq V_{opt}$$

Όπου

$$V_{ran} = \frac{S^2}{n}$$

$$V_{\text{prop}} = \frac{\sum N_h S_h^2}{n \cdot N}$$

$$V_{\text{opt}} = \frac{(\sum N_h S_h)^2}{n \cdot N^2}$$

Για να συνοψίσουμε, υπάρχουν δύο παράγοντες που συμβάλουν στη μείωση της διακυμάνσεως καθώς μετακινούμεθα από την απλή τυχαία δειγματοληψία στην άριστη κατανομή. Ο πρώτος προέρχεται από την εξάλειψη των διαφορών μεταξύ των μέσων των στρωμάτων. Ο δεύτερος από την εξάλειψη του αποτελέσματος των διαφορών μεταξύ των τυπικών αποκλίσεων των στρωμάτων. Ο δεύτερος παράγοντα αντιπροσωπεύει την διάφορα που προκύπτει στην διακύμανση μεταξύ άριστης και αναλογικής κατανομής.

Εάν η διόρθωση πεπερασμένου πληθυσμού δεν είναι δυνατόν να αγνοηθεί, ο ίδιος τύπος ανάλυσεως οδηγεί στο αποτέλεσμα :

$$V_{\text{ran}} = V_{\text{prop}} + \frac{N-n}{nN(N-1)} \left[\sum N_h (\bar{Y}_h - \bar{Y})^2 - \frac{1}{N} \sum (N-N_h) S_h^2 \right]$$

Αυτό συνεπάγεται ότι η αναλογική στρωματοποίηση δίνει μια υψηλότερη διακύμανση από ότι η απλή τυχαία δειγματοληψία. Εάν

$$\sum N_h (\bar{Y}_h - \bar{Y})^2 < \frac{1}{N} \sum (N-N_h) S_h^2$$

3.6. Εκτίμηση του μεγέθους του δείγματος

Σ' αυτή την παράγραφο θα παρουσιάσουμε τύπους για οποιοδήποτε κατανομή διακύμανση V . Εάν, αντιθέτως, το όριο του σφάλματος d έχει προκαθορισθεί, $V = d^2/Z^2$, όπου Z είναι η τιμή της τυπικής κανονικής μεταβλητής που αντιστοιχεί στην εκχωρούμενη πιθανότητα ότι το σφάλμα θα υπερβαίνει το επιθυμητό όριο.

Εκτίμηση πληθυσμιακού μέσου \bar{Y}

Παίρνοντας τον τύπο του $V(\bar{y}_{st})$ και λύνοντας προς το n τότε θα έχουμε για την

- γενικός

$$n = \frac{(\sum W_h S_h)^2 / W_h}{V + 1/N \sum W_h S_h^2}$$

- άριστη κατανομή

$$n = \frac{(\sum W_h S_h)^2}{V + 1/N \sum W_h S_h^2}$$

- αναλογική κατανομή

$$n_o = \frac{\sum W_h S_h^2}{V}, n = \frac{n_o}{1 + n_o/N}$$

Εκτίμηση για το σύνολο του πληθυσμού

Εάν V είναι η επιθυμητή $V(\bar{Y}_{st})$, οι βασικοί τύποι είναι οι ακόλουθοι :

- γενικός

$$n = \frac{(\sum N_h S_h)^2 / W_h}{V + \sum N_h S_h^2}$$

- την άριστη κατανομή

$$n = \frac{(\sum N_h S_h)^2}{V + \sum N_h S_h^2}$$

- την αναλογική κατανομή

$$n_0 = \frac{N \sum W_h S_h^2}{V}, \quad n = \frac{n_0}{1 + n_0/N}$$

3.7 Στρωματοποίηση μετά την συλλογή του δείγματος

Πολλές φορές το δείγμα των n μονάδων ταξινομείται σε κατηγορίες (στρώματα) όπως κατά φύλο, επάγγελμα κ.λ.π. γιατί μας ενδιαφέρουν χωριστά οι κατηγορίες και μπορεί να μην είναι εύκολη η εκ των προτέρων υποδιαίρεση του πληθυσμού στις κατηγορίες αυτές. Έτσι υπάρχουν περιπτώσεις όπου η στρωματοποίηση του ερευνώμενου πληθυσμού δεν μπορεί να γίνει πριν η συλλογή του δείγματος διότι δεν είναι δυνατόν να γνωρίζουμε εκ των προτέρων σε ποίο στρώμα ανήκει μια δειγματοληπτική μονάδα. Για παράδειγμα, αν επιθυμούμε την στρωματοποίηση των ατόμων μιας πόλης κατά φύλο, ηλικία, επάγγελμα, εκπαίδευση κ.λ.π. δεν μπορούμε να κάνουμε την στρωματοποίηση πριν τη συλλογή του δείγματος.

Στις περιπτώσεις αυτές πρέπει να συλλέγεται απλό τυχαίο δείγμα, το οποίο διαιρείται σε στρώματα μετά την συλλογή των μονάδων (post-stratification) και μεταχειρίζεται σαν να ήταν στρωματοποιημένο δείγμα. Οπότε ο μέσος του πληθυσμού \bar{Y} προσδιορίζεται από την εκτίμηση του \bar{y} αντί του \bar{y} με την προϋπόθεση ότι είναι γνωστό το κλάσμα N_h/N . Εάν το κλάσμα αυτό είναι γνωστό και το δείγμα μεγάλο, $n_h \geq 20$ σε κάθε στρώμα, τότε η μέθοδος στρωματοποίησης μετά τη συλλογή δείγματος έχει περίπου την ακρίβεια της αναλογικής στρωματοποιημένης δειγματοληψίας.

Τέλος, η μέθοδος στρωματοποίησης μετά τη συλλογή του δείγματος, μπορεί να εφαρμοστεί σε ένα δείγμα που ήδη έχει γίνει στρωματοποίησης ως προς κάποια γνωστή μεταβλητή.

3.8 Συμπέρασμα

Η αποτελεσματικότητα στην στρωματοποίηση μπορεί περαιτέρω να αυξηθεί, όταν η φύση των δεδομένων επιτρέπει περαιτέρω υποδιαίρεση του στρώματος σε υπό-στρώματα, τα οποία με την σειρά τους μπορούν να υποδιαιρεθούν σε μικρότερες ομάδες. Για παράδειγμα, γνωρίζοντας, ότι η προτιμήσεις των ψηφοφόρων συχνά επηρεάζονται από διαφορές :

θρησκευτικές, μορφωτικού επίπεδου, οικονομικής καταστάσεως κ.λ.π. στην διεξαγωγή ενός γκάλοπ για τις βουλευτικές εκλογές, πρέπει να υποδιαιρέσουμε τους διάφορους νομούς σύμφωνα με αυτές τις απόψεις. Αυτή η λεπτομερής διαδικασία, κατά την στρωματοποίηση των ομάδων του πληθυσμού, σε σχέση με διαφορετικά σχετικά χαρακτηριστικά, αναφέρεται σαν σταυροειδής- στρωματοποίηση (cross-stratification).

Με τα ανωτέρω, έχει καταδειχθεί ότι η στρωματοποίηση είναι πολύ αποτελεσματική όταν αναφέρεται σε ανομοιογενείς ή με μεγάλη ασυμμετρία πληθυσμούς, τέτοιους όπως : δεδομένα εισοδήματος ή λιανικών πωλήσεων. Σε τέτοιους περιπτώσεις, μπορούμε να στρωματοποιήσουμε τον πληθυσμό με τέτοιο τρόπο ώστε

[1] μέσα σε κάθε στρώμα να υπάρχει τόση ομοιομορφία όση είναι δυνατή

[2] μεταξύ των διαφόρων στρωμάτων οι διαφορές να είναι όσο τον δυνατόν μεγαλύτερες .

Συνεπώς, μπορούμε να πάρουμε ένα δείγμα με ένα μικρότερο δειγματοληπτικό σφάλμα ή ένα μικρότερο δείγμα με την ίδια ακρίβεια που επιτυγχάνεται στην απλή τυχαία δειγματοληψία. Με άλλα λόγια επεξεργαζόμενοι πληθυσμούς με ακραίες τιμές, η στρωματοποίηση είναι περισσότερο αποτελεσματική μέθοδος δειγματοληψίας απ' ότι η απλή τυχαία δειγματοληψία . Αυτό είναι αλήθεια, διότι στην απλή τυχαία δειγματοληψία οι ακραίες τιμές π.χ. υψηλά εισοδήματα, μπορούν να αποκλεισθούν, ενώ αυτές οι τιμές ασφαλώς θα αντιπροσωπευθούν σε ένα στρωματοποιημένο δείγμα. Δεν είναι βέβαια απαραίτητο να σημειώσουμε ότι η στρωματοποιημένη δειγματοληψία δεν μπορεί να έχει κάποιο πλεονέκτημα συγκρινόμενη με την απλή τυχαία δειγματοληψία, κατά την μελέτη συμμετρικών κατανομών ή ακόμη και κατανομών με μέτρια ασυμμετρία. Συνεπώς, πριν αποφασίσουμε την στρωματοποίηση, πρέπει να έχουμε κάποια γνώση για τα χαρακτηριστικά του πληθυσμού. Τέτοια γνώση μπορεί να διαμορφωθεί με βάση δεδομένα του παρελθόντος, ή από εμπειρική κρίση κ.λ.π.

Υποκειμενική κρίση χρησιμοποιούμενη για την διαίρεση του δειγματοληπτούμενου πληθυσμού σε στρώματα δεν σημαίνει ότι τα στρωματοποιημένα δείγματα δεν είναι δείγματα πιθανότητας. Εφαρμόζοντας τυχαία επιλογή από κάθε στρώμα, η πιθανότητα με την οποία κάθε μονάδα μπορεί να επιλεγεί εξασφαλίζεται και είναι γνωστή, ακόμα και στην περίπτωση της μη ανάλογικης δειγματοληψίας όπου οι πιθανότητες με τις οποίες οι επί μέρους μονάδες μπορούν να επιλεγούν δεν είναι ίσες .

Σαν δείγματα πιθανότητας που είναι λοιπόν, ο σταθμικός αριθμητικός μέσος των δειγμάτων μέσων κατά στρώμα είναι μια αμερόληπτη εκτίμηση του μέσου του πληθυσμού. Στην περίπτωση της αναλογικής δειγματοληψίας, όπου κάθε μονάδα έχει ίση πιθανότητα επιλογής, η εκτίμηση του μέσου του πληθυσμού είναι απλώς αριθμητικός μέσος ολόκληρου του δείγματος, επειδή είναι σταθμισμένος με αυτές τις αναλογίες. Πάντως, ο μέσος του μη αναλογικού δείγματος ~~πρέπει να~~ προκύψει με κατάλληλες σταθμίσεις, άλλως σοβαρές μεροληψίες μπορεί να προκύψουν οι κατάλληλες σταθμίσεις, είναι τα σχετικά ποσοστά του πληθυσμού στα διάφορα στρώματα.

Συνοψίζοντας, μπορούμε να πούμε ότι η ακρίβεια των εκτιμήσεων στην στρωματοποιημένη δειγματοληψία μπορεί να μετρηθεί αρκεί το δείγμα να είναι επαρκώς μεγάλο.

Τέλος, η στρωματοποιημένη δειγματοληψία παρουσιάζει τα εξής πλεονεκτήματα και μειονεκτήματα:

Πλεονεκτήματα

1. Παρουσιάζει καλύτερη εκτίμηση ανά μονάδα κόστος σε σχέση με την απλή τυχαία δειγματοληψία και αυτό οφείλεται στο ότι η μεταβλητότητα στα στρώματα είναι μικρότερη από ότι σε όλο τον πληθυσμό.
2. Το δειγματικό κόστος είναι μικρότερο στην στρωματοποιημένη δειγματοληψία, διότι απαιτείται μικρότερο μέγεθος δείγματος για να επιτυγχάνει μικρότερη διασπορά.
3. Δίνει χωριστές εκτιμήσεις για τα διάφορα χαρακτηριστικά κάθε στρώματος χωρίς να χρειάζεται επιπρόσθετη δειγματοληψία.
4. Η εκλογή του δείγματος και η συγκέντρωση των πληροφοριών γίνεται ευκολότερα και αυτό επειδή κάθε στρώμα αποτελείται από λίγες μονάδες και καλύπτει μικρότερη γεωγραφική περιοχή σε σχέση με όλο τον πληθυσμό. Για αυτό και η έρευνα ολοκληρώνεται γρηγορότερο.
5. Οδηγεί σε αμερόληπτες εκτιμήτριες.

Μειονεκτήματα :

1. Υπάρχουν περιπτώσεις που κάποια άτομα να είναι δυνατό να ανήκουν σε περισσότερα από ένα στρώμα, όταν τα όρια δεν είναι σαφή για αυτό παρουσιάζονται δυσκολίες στον ακριβή ορισμό των στρωμάτων και επομένως στον χωρισμό του πληθυσμού σε στρώματα .
2. Αν τα στρώματα δεν παρουσιάζουν ανομοιογένεια προς τον εξεταζόμενο χαρακτηριστικό, η διασπορά θα είναι ίδια με αυτήν της απλής τυχαία δειγματοληψία. Άρα χρειάζεται μεγάλη προσοχή στην εκλογή κριτηρίου, βάσει του οποίου χωρίζεται ο πληθυσμός σε στρώματα ανομοιογενή, διαφορετικά θα υποβληθούμε σε μάταιο κόπο και κόστος .

ΚΕΦΑΛΑΙΟ	Δειγματοληψία κατά ομάδες
4 .	

4 . 1 Εισαγωγή

Η δειγματοληψία κατά ομάδες ή συσσωρευτική ή κατά συστάδες δειγματοληψία (όπως αλλιώς είναι γνωστό) (cluster sampling), συμφώνα με την οποία χωρίζουμε τον πληθυσμό σε διάφορες συστάδες, και στη συνέχεια η επιλογή ενός τυχαίου δείγματος από συστάδες. Συνεπώς, η δειγματοληψία κατά ομάδες είναι μια απλή τυχαία δειγματοληψία στην οποία κάθε δειγματοληπτική μονάδα είναι μια ομάδα στοιχείων του ερευνώμενου πληθυσμού. Π.χ. εάν επιθυμούμε να εκτιμήσουμε το μέσο εισόδημα των νοικοκυριών μιας πόλης μπορούμε να θεωρήσουμε ότι όλα τα νοικοκυριά ομαδοποιούνται κατά οικοδομικό τετράγωνο στο χαρτογραφικό διάγραμμα της πόλης. Έτσι το σύνολο των νοικοκυριών της πόλης ισοδυναμεί με το σύνολο των οικοδομικών τετράγωνων της πόλης, επομένως αντί να πάρουμε τυχαία έναν αριθμό νοικοκυριών της πόλης παίρνουμε τυχαία έναν αριθμό οικοδομικών τετραγώνων και ερευνούμε τον σύνολο των νοικοκυριών που βρίσκονται στα οικοδομικά τετράγωνα του δείγματος .

Με τον τρόπο αυτό συντομεύεται ο χρόνος διεξαγωγής της έρευνας και μειώνεται το κόστος, ειδικότερα όταν τα νοικοκυριά του ερευνώμενου πληθυσμού απέχουν μεταξύ τους μεγάλες αποστάσεις, απαιτείται περισσότερος χρόνος για μετακίνηση των ερευνητών από νοικοκυριό σε νοικοκυριό και κατά συνέπεια μεγαλύτερο κόστος .

Επίσης η δειγματοληψία κατά ομάδες μπορεί να εφαρμοστεί όταν δεν υπάρχουν τα απαιτούμενα δειγματοληπτικά πλαίσια (π.χ. κατάλογοι νοικοκυριών μιας πόλης κ.λ.π.), ενώ αντίθετα υπάρχουν πλαίσια για

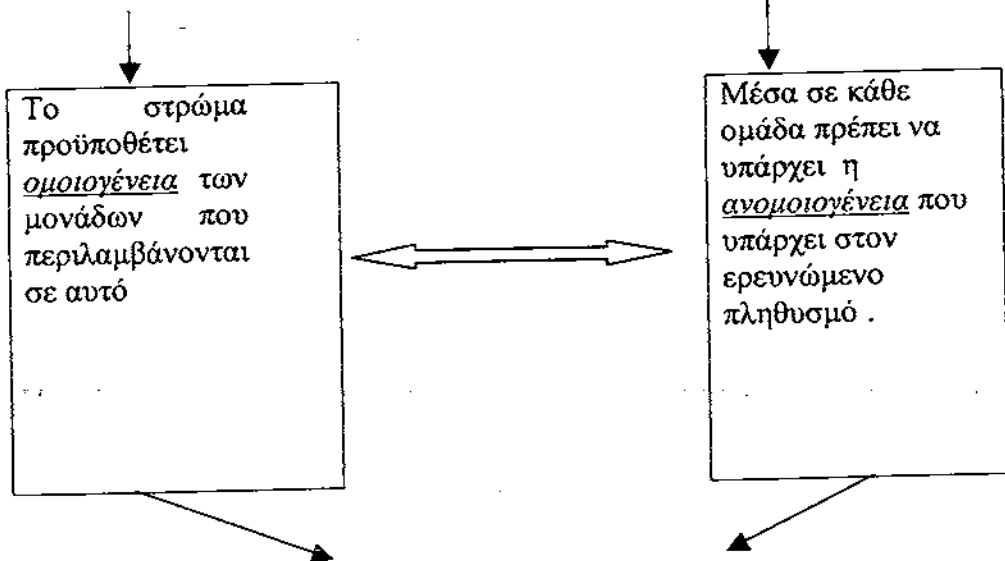
ομάδες στοιχείων (π.χ. οικοδομικά τετράγωνα σε χαρτογραφικά διαγράμματα μιας πόλης).

Ομάδες στοιχείων, στην δειγματοληψία κατά ομάδες, μπορεί να αποτελέσουν τα οικοδομικά τετράγωνα, όταν ως μονάδα του ερευνημένου πληθυσμού είναι το νοικοκυριό, τα σχολεία, όταν ως μονάδα του ερευνούμενου πληθυσμού είναι ο μαθητής, όταν ως μονάδα του ερευνώμενου πληθυσμού είναι ασθενής, ή κατοικία όταν ως μονάδα του ερευνώμενου πληθυσμού είναι το άτομο, οι επιχειρήσεις ενός κλάδου οικονομικής δραστηριότητας όταν ως μονάδα του ερευνώμενου πληθυσμού είναι ο απασχολούμενος στον κλάδο αυτό κ.ο.κ.

Ο χωρισμός του πληθυσμού σε ομάδες - συστάδες φαίνεται να μοιάζει με τη μέθοδο της στρωματοποιημένης δειγματοληψίας όπου πάλι ο πληθυσμός χωρίζεται σε ομάδες - στρώματα. Για αυτό επισημαίνεται ότι δεν πρέπει να γίνεται σύγχυση μεταξύ στρωματοποιημένης δειγματοληψίας και δειγματοληψίας κατά ομάδες διότι υπάρχει μια σημαντική διαφορά. Έτσι στην στρωματοποιημένη δειγματοληψία έχουμε δει ότι για την βέλτιστη κατασκευή (άριστη κατανομή) των στρώματων πρέπει τα στρώματα να είναι όσο τον δυνατόν ομοιογενή ως προς τα στοιχεία (μονάδες) που περιέχουν και να είναι όσο τον δυνατόν περισσότερο ανομοιογενή μεταξύ τους, δηλαδή να διαφέρουν μεταξύ τους, ως προς τα χαρακτηριστικό που μας ενδιαφέρει. Αντίθετα, στην δειγματοληψία κατά ομάδες, ο χωρισμός γίνεται σε όσο τον δυνατόν ομοιογενείς συστάδες με ανομοιογενή στοιχεία η κάθε μια. Για να γίνει κατανοητό αναφέρουμε τον παρακάτω σχήμα :

Στρωματοποιημένη
Δειγματοληψία

Δειγματοληψία
κατά ομάδες



Ωστε να προκύψουν
αξιόπιστα
αποτελέσματα

4.2 Συμβολισμοί

Στη δειγματοληψία κατά ομάδες χρησιμοποιούνται οι παρακάτω συμβολισμοί :

N = ο αριθμός των ομάδων στον πληθυσμό

n = ο αριθμός των μονάδων που έχουν επιλεγεί στο δείγμα

m_i = ο αριθμός των στοιχείων στην ομάδα *i* (*i*=1, 2, ..., N)

\bar{n} = $1/n \sum_{i=1}^n m_i$, μέσος μέγεθος ομάδας στο δείγμα

M = $\sum_{i=1}^N m_i$, ο αριθμός των στοιχείων στον πληθυσμό .

\bar{N} = M/m , μέσο μέγεθος ομάδων στον πληθυσμό .

y_i = συνολική τιμή της μεταβλητής *y* στην *i* ομάδα .

\bar{Y} = $1/M \sum_{i=1}^N y_i$, μέσος πληθυσμού

\bar{y} = $\sum_{i=1}^n y_i / \sum_{i=1}^n m_i$, μέσος δείγματος .

α_i = ο αριθμός των στοιχείων στην ομάδα *i* που έχουν το συγκεκριμένο χαρακτηριστικό, που μας ενδιαφέρει .

ρ = $\sum_{i=1}^n \alpha_i / \sum_{i=1}^n m_i$, ποσοστό στοιχείων δείγματος που ανήκουν στο συγκεκριμένο χαρακτηριστικό .

4.3 Εκτίμηση του μέσου και του συνολικού πληθυσμού

Η εκτίμηση του μέσου πληθυσμού και του συνολικού πληθυσμού γίνεται όπως στην απλή τυχαία δειγματοληψία, αφού η δειγματοληψία κατά ομάδες είναι μια απλή τυχαία δειγματοληψία στην οποία κάθε δειγματοληπτική μονάδα αποτελείται από μια ομάδα στοιχείων του πληθυσμού. Επομένως μια αμερόληπτη εκτίμηση του μέσου \bar{Y} του πληθυσμού είναι ο μέσος του δείγματος και συμβολίζεται με \hat{y}_c (ο δείκτης c σημαίνει δειγματοληψία κατά ομάδες - cluster sampling) δηλαδή έχουμε

$$\hat{y}_c = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n m_i}$$

Η διακύμανση της εκτίμησης του \hat{y}_c είναι

$$V(\hat{y}_c) = \frac{N-n}{N \cdot n} S^2$$

Όπου

$$S^2 = \frac{\sum_{i=1}^N (y_i - \bar{Y})^2}{N-1}$$

Εάν το S^2 είναι άγνωστο χρησιμοποιείται η εκτιμήτρια s^2 :

$$s^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_c m_i)^2}{n-1}$$

Επίσης όταν το M είναι άγνωστο τότε το \bar{N} εκτιμάται από το \hat{n}
Το τυπικό σφάλμα είναι :

$$SE(\hat{y}_c) = \sqrt{V(\hat{y}_c)}$$

Η εκτίμηση του συνολικού πληθυσμού Y δίνεται από την σχέση

$$\hat{Y} = M\hat{y}_c$$

Η διακύμανση της εκτίμησης του συνολικού πληθυσμού Y υπολογίζεται από την σχέση

$$V(\hat{Y}) = V(M\hat{y}_c) = M^2V(\hat{y}_c)$$

Το τυπικό σφάλμα είναι

$$SE(\hat{Y}) = SE(M\hat{y}_c) = \sqrt{V(M\hat{y}_c)}$$

- Εάν ο αριθμός των στοιχείων M στον πληθυσμό είναι άγνωστος τότε η εκτίμηση του συνολικού πληθυσμού δίνεται από την σχέση

$$\hat{Y} = N\hat{y}_t$$

Η διακύμανση της εκτίμησης του συνολικού πληθυσμού δίνεται από την σχέση

$$V(\hat{Y}) = V(N\hat{y}_t) = N^2V(\hat{y}_t)$$

Το τυπικό σφάλμα είναι

$$SE(\hat{Y}) = SE(N\hat{y}_t) = \sqrt{V(N\hat{y}_t)}$$

Σημείωση: Η εκτίμηση ποσοστού ενός πληθυσμού, και ο προσδιορισμός μεγέθους δείγματος τόσο για την εκτίμηση του μέσου και συνολικού πληθυσμού όσο για την εκτίμηση ποσοστών με την μέθοδο της δειγματοληψίας κατά ομάδες, γίνεται όπως και στην απλή τυχαία δειγματοληψία. Όποτε δεν είναι απαραίτητο να αναφέρουμε εδώ τους τύπους αυτούς.

4.4 Δειγματοληψία κατά ομάδες συνδυαζόμενη με στρωματοποίηση.

Η μέθοδος της στρωματοποιημένης δειγματοληψίας κατά ομάδες είναι ένας συνδυασμός της στρωματοποιημένης δειγματοληψίας και της δειγματοληψίας κατά ομάδες. Έτσι η δειγματοληψία κατά ομάδες μπορεί

να εφαρμοσθεί μετά την στρωματοποίηση του ερευνωμένου πληθυσμού σε ομοιογενείς υποπληθυσμούς. Δηλαδή χωρίζεται ο πληθυσμός σε ομοιογενή στρώματα και στη συνέχεια κάθε στρώμα σε ομάδες, όποτε από κάθε στρώμα λαμβάνεται ένα δείγμα ομάδων (π.χ. οικοδομικά τετράγωνα που διαμένουν νοικοκυριά) με τυχαίο τρόπο .

Όπως θα φανεί στην συνέχεια η μέθοδος επιδεικνύεται για μεγάλους πληθυσμούς και μάλιστα εκεί όπου ο χωρισμός σε στρώματα δεν μπορεί να εξασφαλίσει την ομοιογένεια των στρωμάτων, πράγμα που σημαίνει ότι η εφαρμογή μόνο της στρωματοποιημένης θα οδηγούσε σε κακές εκτιμήτριες

Η εκτίμηση του μέσου του πληθυσμού \bar{Y} δίνεται από την σχέση (ο δείκτης cs σημαίνει δειγματοληψία κατά στρώματα και στρωματοποιημένη)

$$\hat{y}_{cs} = \frac{\sum_{j=1}^k N_j \hat{y}_j}{\sum_{j=1}^k N_j \bar{n}_j}$$

Όπου N_j ο συνολικός αριθμός ομάδων μέσα σε κάθε στρώμα j ($j=1, 2, \dots, k$),

$$\hat{y}_j = \frac{\sum_{i=1}^{n_j} y_{ji}}{n_j}$$

Είναι ο μέσος ολικός των ομάδων του δείγματος στο στρώμα j είναι

$$\bar{n}_j = \frac{\sum_{i=1}^{n_j} m_{ji}}{n_j}$$

Το μέσο μέγεθος ομάδων του δείγματος στο στρώμα j

Τέλος, αυτή η μέθοδος παρουσιάζει ορισμένα πλεονεκτήματα και μειονεκτήματα, τα οποία είναι τα εξής :

Πλεονεκτήματα

- I. οι εκτιμήσεις στην μέθοδο αυτό θεωρούνται καλύτερες από τις αντίστοιχες της απλής τυχαίας δειγματοληψίας, της

- στρωματοποιημένης και της δειγματοληψίας κατά ομάδες .
2. για την στρωματοποιημένη δειγματοληψία κατά ομάδες απαιτείται μικρότερο μέγεθος δείγματος, το οποίο θα είναι αντιπροσωπευτικός τους πληθυσμού .
 3. για κάθε στρώμα, γίνονται ξεχωριστές εκτιμήσεις και με τα αποτελέσματα που προκύπτουν δίνεται η δυνατότητα να κάνουμε σύγκριση μεταξύ τους .
 4. στην μέθοδο αυτό παρατηρείτε η δυνατότητα καταμερισμού των εργασιών τόσο στην συλλογή των πληροφοριών όσο και στην επεξεργασία τους. Αποτέλεσμα αυτής της δυνατότητας είναι ότι η εκλογή και η εξέταση του δείγματος είναι ευκολότερη .
 5. η μέθοδος αυτό παρουσιάζεται, όταν δεν υπάρχει ονομαστική λίστα ή όταν είναι δύσκολο να δημιουργηθεί λόγω κόστους ή άλλων παραγόντων.
 6. μειώνει το κόστος της δειγματοληψίας και ενδείκνυται, όταν αύξηση της απόστασης μεταξύ των παρατηρήσεων σημαίνει αύξηση του κόστους.

Μειονεκτήματα

1. ο χωρισμός σε στρώματα και συστάδες θα πρέπει να γίνεται με επιμέλεια, ώστε να τηρούνται οι καθορισμένες προδιαγραφές και να αντιμετωπίζονται με προσοχή περιπτώσεις, όπως εκείνες που δύο ή περισσότερα στρώματα μπορούν να διεκδικήσουν ίδια άτομα από το πληθυσμό. Σε αντίθετη περίπτωση κινδυνεύουμε να

μην πάρουμε αντιπροσωπευτικό δείγμα, να οδηγηθούμε σε μεγάλες αποκλείσεις χάνοντας έτσι τα πλεονεκτήματα της μεθόδου.

2. και τέλος οι εκτιμήτριες της είναι μεροληπτικές.

4.5 Δειγματοληψία κατά ομάδες με πιθανότητα ανάλογη προς το μέγεθος των ομάδων

Μερικές φορές μπορούμε να μειώσουμε τη διακύμανση μιας εκτίμησης, αν οι δειγματοληπτικές μονάδες λαμβάνονται με πιθανότητα ανάλογη προς ένα μέτρο του μεγέθους των μονάδων. Η δειγματοληψία κατά ομάδες αποτελεί μια ιδανική περίπτωση, στην οποία μπορεί να ληφθούν οι δειγματοληπτικές μονάδες με πιθανότητα ανάλογη προς το μέγεθος, διότι ο αριθμός των στοιχείων (m_i) σε μια ομάδα i είναι να φυσικό μέτρο του μεγέθους της ομάδας. Έτσι μπορούμε να περιορίσουμε το σφάλμα μιας εκτίμησης στην περίπτωση που ο αριθμός των στοιχείων διαφέρει σημαντικά μεταξύ των ομάδων του ερευνώμενου πληθυσμού, και υπάρχει έντονη συσχέτιση μεταξύ του μεγέθους της μονάδας και των υπό εξέταση χαρακτηριστικών. Αυτό οφείλεται στο γεγονός, ότι δίνεται μεγαλύτερη πιθανότητα επιλογής στις μεγαλύτερες ομάδες (π.χ. οικοδομικά τετράγωνα με μεγαλύτερο αριθμό νοικοκυριών).

Συμβολίζοντας με π_i την πιθανότητα ότι η i δειγματοληπτική μονάδα εμφανίζεται στο δείγμα, έχουμε :

$$\pi_i = \frac{m_i}{M}$$

Επομένως η εκτίμηση του συνολικού πληθυσμού προκύπτει είναι ως εξής :

$$\hat{Y} = \frac{M}{n} \sum_{i=1}^n \hat{y}_i$$

4.6 Συμπέρασμα

Όταν οι συσσωρευτικές, που αποτελούν και τις πρωταρχικές μονάδες δειγματοληψίας, επιλέγουν, μπορούμε είτε να συμπεριλάβουμε στο δείγμα όλες τις στοιχειώδεις μονάδες των επιλεγείσων ομάδων (συσσωρεύσεων) είτε να πάρουμε ένα δείγμα με λιγότερες πρωταρχικές μονάδες (ή από στοιχειώδεις μονάδες τους) από εκείνες που επιλέξαμε σαν ομάδες δειγματοληψίας.

Όταν όλες οι στοιχειώδεις μονάδες των επιλεγείσων ομάδων είναι διαθέσιμες για παρατήρηση, έχουμε αυτό που είναι γνωστό σαν δειγματοληψία ενός σταδίου (single-stage sampling). Όταν ένα δείγμα στοιχειωδών μονάδων λαμβάνεται από τις ομάδες, έχουμε ένα τύπο σχεδίου που ονομάζεται δειγματοληψία δύο σταδίων (two-stage sampling or sub sampling). Και στα δυο στάδια ένα τυχαίο δείγμα επιλέγεται. Όταν η δειγματοληψία κατά ομάδες περιλαμβάνει περισσότερα από δυο στάδια για την επιλογή του τελικού δείγματος, ονομάζεται πολυσταδιακή δειγματοληψία (multistage sampling). Για παράδειγμα, μπορούμε να πάρουμε τα κολέγια (πρωταρχικές μονάδες) στο πρώτο στάδιο, κατόπιν επιλέγουμε τα τμήματα σαν δεύτερο στάδιο και τέλος επιλέγουμε τους σπουδαστές (στοιχειώδεις μονάδες) σαν τρίτο και τελευταίο στάδιο. Όταν οι συσσωρεύσεις (πρωταρχικές μονάδες) είναι γεωγραφικές περιοχές, η δειγματοληψία κατά ομάδες γίνεται η πολύ γνωστή «δειγματοληψία επιφανείας» (area sampling). Σε κάθε περίπτωση, πάντως, σκοπός μας είναι να μελετήσουμε τις στοιχειώδεις μονάδες.

Οι αρχές που εξασφαλίζουν την μέγιστη αποτελεσματικότητα στην δειγματοληψία κατά ομάδες είναι οι αντίθετες από εκείνες που χρησιμοποιούνται στην στρωματοποιημένη δειγματοληψία. Στην δειγματοληψία κατά ομάδες, είναι αποτελεσματικό να έχουμε :

[1] Διαφορές μεταξύ των στοιχειωδών μονάδων στο ίδιο γκρουπ (ή τάξη) όσο τον δυνατόν μεγαλύτερες και

[2] Διαφορές μεταξύ των πρωταρχικών μονάδων (συσσωρεύσεις) όσο τον δυνατόν μικρότερες.

Προφανώς, αυτές οι προϋποθέσεις είναι πολύ δύσκολο να διαπιστωθούν, διότι στους περισσότερους πληθυσμούς παρόμοιες στοιχειώδεις μονάδες τείνουν να συσσωρεύονται μαζί. Για παράδειγμα, οι πλούσιοι άνθρωποι τείνουν να ζουν στην ίδια περιοχή μιας πόλεως, ενώ οι φτωχές οικογένειες συγκεντρώνονται σε άλλη περιοχή.

Αυτή η δυσκολία προκαλεί στην δειγματοληψία κατά ομάδες ένα μεγαλύτερο τυπικό σφάλμα από εκείνο της στρωματοποιημένης δειγματοληψίας, ακόμη και αν η πρώτη επίσης παράγει αμερόληπτες εκτιμήσεις. Το γεγονός ότι η δειγματοληψία κατά ομάδες παρέχει λιγότερη ακρίβεια για ένα δείγμα ίσου μεγέθους δεν σημαίνει ότι δεν μπορεί να είναι περισσότερο αποτελεσματική από άλλα τυχαία σχέδια δειγματοληψίας. Εάν το κόστος κατά στοιχειώδη μονάδα είναι πολύ χαμηλότερο στην δειγματοληψία κατά ομάδες πολύ περισσότερες στοιχειώδεις μονάδες μπορούν να περιληφθούν στο δείγμα [και είναι συνήθης η περίπτωση που το κόστος κατά συνέντευξη στην δειγματοληψία κατά ομάδες είναι χαμηλότερο από το αντίστοιχο κόστος για άλλους τύπος τυχαίων σχεδίων. Χαμηλότερο κόστος διαπιστώνεται περισσότερες φορές στην δειγματοληψία επιφανείας, όταν ο χρόνος μεταξύ δύο συνεντεύξεων είναι αρκετά περιορισμένος]. Κατά συνέπεια, είναι δυνατόν να έχουμε ένα μεγαλύτερο συσσωρευτικό δείγμα το οποίο μπορεί να παρέχει την ίδια ακρίβεια με χαμηλότερο κόστος απ' ό,τι ένα απλό ή στρωματοποιημένο τυχαίο δείγμα. Εναλλακτικά, είναι δυνατόν, με το ίδιο κόστος να έχουμε ένα συσσωρευτικό δείγμα αρκετά μεγάλο ώστε να έχουμε μεγαλύτερη αξιοπιστία από εκείνη που θα είχαμε αντιστοίχως είτε στην απλή είτε στην στρωματοποιημένη δειγματοληψία. Με άλλα λόγια, πάρα το μειονέκτημα στην επίτευξη ετερογενείς μέσα σε κάθε ομάδα (συσσώρευση) και ομοιομορφίας μεταξύ των διαφόρων ομάδων (συσσωρεύσεων), η δειγματοληψία κατά ομάδες μπορεί ακόμη να είναι ένα περισσότερο αποτελεσματικό σχέδιο δειγματοληψίας .

Έτσι συγκεντρωτικά τα πλεονεκτήματα και τα μειονεκτήματα στην δειγματοληψία κατά ομάδες είναι τα εξής :

Πλεονεκτήματα

1. η δειγματοληψία κατά ομάδες είναι οικονομικότερη στις περιπτώσεις όταν :

α . δεν υπάρχει η λίστα με τις μονάδες του πληθυσμού
ή
β. το κόστος για την δημιουργία τις λίστας αυτής είναι υψηλή .

2. η δειγματοληψία κατά ομάδες απαιτεί λιγότερο κόστος σε σχέση με τις άλλες μεθόδους, στην περίπτωση που το κόστος λήψεως των πληροφοριών αυξάνει όσο η απόσταση μεταξύ των

ατόμων του πληθυσμού μεγαλώνει, για καθορισμένη ακρίβεια εκτιμήσεων .

Μειονεκτήματα

1. στην δειγματοληψία κατά ομάδες, οι συστάδες πρέπει να είναι ομοιογενείς μεταξύ τους και οι μονάδες τους ανομοιογενείς. Στην περίπτωση που δεν ισχύει αυτό, τότε:
 - ή το δείγμα δεν είναι αντιπροσωπευτικό, άρα μικρότερη ακρίβεια εκτιμήσεων
 - ή απαιτείται να παρθούν πληροφορίες από μεγάλο αριθμό συστάδων, άρα θα είναι αντιοικονομική .
2. ακόμα, αυτή η μέθοδος οδηγεί σε μη αμερόληπτες εκτιμήτριες .

Εκείνο που πρέπει να προκύψει σαν συμπέρασμα είναι ότι σκοπός της δειγματοληψίας κατά ομάδες δεν είναι να πετύχουμε το περισσότερο αξιόπιστο δείγμα σχέση με το μέγεθος του, όπως στην περίπτωση της στρωματοποιημένης δειγματοληψίας, αλλά να έχουμε την μέγιστη ακρίβεια στις εκτιμήσεις σε σχέση με το κόστος κατά στοιχειώδη μονάδα

<p style="text-align: center;">ΚΕΦΑΛΑΙΟ</p> <p style="text-align: center;">5 .</p>	<p style="text-align: center;">Εκτιμήσεις με τη μέθοδο του λόγου και της παλινδρόμησης</p>

5 . 1 . Εισαγωγή

Στα προηγούμενα κεφάλαια έγινε περιγραφή της εκτίμησης του μέσου και συνολικού πληθυσμού, χρησιμοποιώντας τις παρατηρήσεις του υπό μελέτη χαρακτηριστικού Y , δηλαδή εξετάσαμε τις μονάδες ενός πληθυσμού ως προς μια μόνο μεταβλητή ιδιότητα τους. Σε πολλές περιπτώσεις όμως, υπάρχει συσχέτιση μεταξύ του υπό μελέτη χαρακτηριστικού Y και ενός άλλου βοηθητικού χαρακτηριστικού X , δηλαδή ασχολούμαστε συγχρόνως με τη μελέτη δυο μεταβλητών, με σκοπό να εξακριβώσουμε αν υπάρχει αλληλεξάρτηση μεταξύ τους .

Υποθέτουμε ότι έχουμε ένα πληθυσμό μεγέθους N για το οποίο ενδιαφερόμαστε να εκτιμήσουμε το ολικό Y μιας μεταβλητής ή τη μέση τιμή \bar{Y} , βάσει ενός τυχαίου δείγματος μεγέθους n που έχουμε πάρει . Έστω ότι για κάθε μονάδα του πληθυσμού εκτός του χαρακτηριστικού που μας ενδιαφέρει (y) είναι γνωστή η τιμή ενός άλλου χαρακτηριστικού (x) που είναι σε υψηλή συσχέτιση με το χαρακτηριστικό x .

Για παράδειγμα, Y μπορεί να είναι η συνολική δαπάνη νοικοκυριού και X ο αριθμός των ατόμων του νοικοκυριού .

Ένα άλλο παράδειγμα, Y να είναι ο αριθμός ημερών παραμονής ασθενών στο νοσοκομείο κατά τη διάρκεια ενός μηνός και X να είναι ο αριθμός κλινών του νοσοκομείου. Για τον χαρακτηριστικό X μπορεί να υπάρχει διαθέσιμα στοιχεία ή να συγκεντρωθούν από τον ερευνώμενο πληθυσμό κατά την διενέργεια της δειγματοληπτικής έρευνας .

Στην περίπτωση αυτή εξασφαλίζεται μεγαλύτερη ακρίβεια των εκτιμήσεων χρησιμοποιώντας τις παρατηρήσεις του υπό μελέτη

χαρακτηριστικού Y και του βοηθητικού χαρακτηριστικού X και εφαρμόζοντας μια από τις παρακάτω μεθόδους :

- Εκτίμηση λόγου (ratio estimate)
- Εκτίμηση παλινδρόμησης (regression estimate)

5.2. Εκτίμηση λόγου

Η εκτίμηση λόγου είναι ένας διαφορετικός τρόπος εκτιμήσεως πληθυσμιακών συνόλων (ή μέσων) που είναι χρήσιμα σε πολλά προβλήματα δειγματοληψίας. Υποθέτουμε ότι έχουμε λάβει ένα δείγμα για να εκτιμήσουμε το πληθυσμιακό σύνολο μιας μεταβλητής, Y, και ότι μια πλήρης μέτρηση του πληθυσμού έγινε σε κάποια προηγούμενη περίοδο (π.χ. σε μια πρόσφατη απογραφή). Έστω X δηλώνει την τιμή της μεταβλητής στην προηγούμενη περίοδο .

Τα σύμβολα που χρησιμοποιούνται για την εκτίμηση του λόγου είναι :

Y_i = τιμή του υπό μελέτη χαρακτηριστικού Y για την i μονάδα του πληθυσμού ($i = 1, 2, \dots, N$)

X_i = τιμή του βοηθητικού χαρακτηριστικού X για την i μονάδα του πληθυσμού .

y_i = τιμή του υπό μελέτη χαρακτηριστικού Y για την i μονάδα του δείγματος ($i = 1, 2, \dots, n$)

x_i = τιμή του βοηθητικού χαρακτηριστικού X για την i μονάδα του δείγματος .

$R = \frac{\bar{Y}}{\bar{X}}$, λόγος του μέσου πληθυσμού .

$r = \frac{\bar{y}}{\bar{x}}$, λόγος του δειγματικού μέσου .

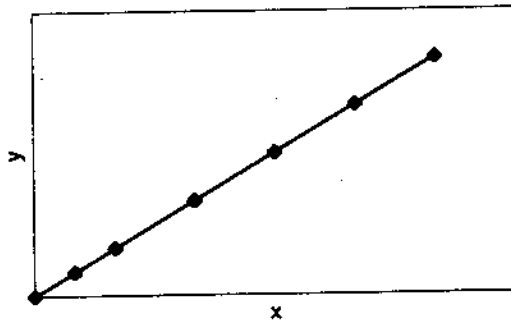
Η εκτίμηση του λόγου χρησιμοποιείται στην ανάλυση στοιχείων ερευνών. Για παράδειγμα, ερευνητές ιατρικής μπορούν με την χρησιμοποίηση της εκτίμησης του λόγου να μετρήσουν την

αποτελεσματικότητα ενός φαρμάκου σε σχέση με άλλο φάρμακο δοκιμασμένο. Σε δειγματοληπτικές έρευνες απασχόλησης μπορεί να χρησιμοποιηθεί η εκτίμηση λόγου ώστε να προσαρμοσθούν αριθμητικά μεγέθη ανέργων κατά ηλικία, φύλο κ.λ.π. Για παράδειγμα, ο λόγος των ανέργων γυναικών που προκύπτει από δείγμα μια περιοχής, μπορεί να χρησιμοποιηθεί σαν μέτρο ανεργίας γυναικών σε μια ευρύτερη περιοχή, αν πολλαπλασιασθεί με τον αριθμό των εργαζομένων γυναικών της ευρύτερης περιοχής .

Στις προβλέψεις συχνά εφαρμόζεται η εκτίμηση του λόγου. Π.χ. πρόβλεψη πωλήσεων, πρόβλεψη αύξησης πληθυσμού κ.λ.π.

Επισημαίνεται ότι η δειγματική κατανομή του λόγου r σε μικρά δείγματα δεν είναι συμμετρική και δίνει συνήθως μεροληπτικές εκτιμήσεις. Όταν το δείγμα είναι μεγάλο η κατανομή του r τείνει προς την κανονικότητα και δίνει αξιόπιστα αποτελέσματα. Δηλαδή όσο το δείγμα αυξάνει τόσο η μεροληψία καθίσταται άνευ σημασία. Επίσης η χρησιμοποίηση της εκτίμησης του λόγου είναι πιο αποτελεσματική όταν :

- η συσχέτιση των μεταβλητών y_i και x_i είναι γραμμική και διέρχεται από την αρχή των αξόνων



- η διακύμανση του y_i είναι ανάλογη προς το x_i

5.2.1 Εκτίμηση λόγου χρησιμοποιώντας απλή τυχαία δειγματοληψία

Έστω ότι θέλουμε να πάρουμε ένα δείγμα μεγέθους n από έναν πληθυσμό N με την μέθοδο της απλής τυχαίας δειγματοληψίας, για να εκτιμήσουμε το μέσο και συνολικό πληθυσμό ή το λόγο R δύο χαρακτηριστικών y_i και x_i των μονάδων του ερευνώμενου πληθυσμού .

Για παράδειγμα, σε έρευνα νοικοκυριών ενδιαφέρει ο μέσος αριθμός κουστουμιών κατά ηλικιωμένο άρρενα, η μέση δαπάνη για κοσμήματα κατά ηλικιωμένη γυναίκα και ο μέσος αριθμός ωρών που αφιερώνονται για παρακολούθηση τηλεοπτικών προγραμμάτων κατά παιδί ηλικίας 10-15 ετών. Για να εκτιμήσουμε τον μέσο αριθμό κουστουμιών κατά ηλικιωμένο άνδρα, αυτών των μονάδων, θα καταγράψουμε για το i νοικοκυριό ($i=1, 2, 3, \dots, n$) τον αριθμό των ηλικιωμένων ανδρών x_i που ζουν εκεί και τον συνολικό αριθμό κουστουμιών y_i τα οποία αυτοί κατέχουν. Η πληθυσμιακή παράμετρος που εκτιμάται είναι ο λόγος :

$$R = \frac{\text{Συνολικός αριθμός κουστουμιών}}{\text{Συνολικός αριθμός ηλικιωμένων ανδρών}} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i}$$

η αντίστοιχη δειγματική εκτίμηση είναι :

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

Η λόγο-εκτίμηση μπορεί επίσης να χρησιμοποιηθεί όταν η X είναι κάποια άλλη κατά κάποιο τρόπο συμπληρωματική μεταβλητή. Οι προϋποθέσεις για μια επιτυχή εφαρμογή αυτής της εκτιμήσεως είναι ότι ο λόγος y_i / x_i θα είναι σταθερός από μονάδα σε μονάδα του πληθυσμού και ότι το πληθυσμιακό σύνολο $X = \sum x_i$ θα είναι γνωστό. Θεωρήσουμε, μια εκτίμηση της συνολικής ποσότητας μιας σοδειάς, αμέσως μετά την συγκομιδή, που έγινε από ένα δείγμα αγρών σε κάποια περιοχή. Για κάθε αγρό στο δείγμα καταγράφουμε την συνολική απόδοση y_i , και την συνολική έκταση (σε εκτάρια) x_i , αυτής της σοδειάς. Σ' αυτήν την περίπτωση ο λόγος R είναι η δειγματική εκτίμηση της μέσης αποδόσεως κατά εκτάριο. Αυτός ο λόγος πολλαπλασιάζεται με τον συνολικό εμβαδόν (σε εκτάρια) της σοδειάς στην περιοχή, το οποίο θα έχει γίνει γνωστό ακριβώς από κάποια άλλη πηγή. Αυτή η εκτίμηση θα είναι ακριβής εάν η μέση απόδοση ανά εκτάριο ποικίλλει λίγο από αγρό σε αγρό.

Η κατανομή δειγματοληψίας της εκτιμήτριας r είναι περισσότερο περίπλοκη από την αντίστοιχη της εκτιμήσεως \hat{y} , διότι και ο αριθμητής \hat{y} και ο παρανομαστής \hat{x} ποικίλλουν από δείγμα σε δείγμα. Σε μικρά δείγματα η κατανομή της εκτιμήσεως r είναι ασυμμετρική και το r είναι

συνήθως μια ελαφρώς μεροληπτική εκτίμηση του R. Σε μεγάλα δείγματα η κατανομή της εκτιμήσεως r τείνει να γίνει κανονική και η μεροληψία γίνεται αμελητέα.

ΘΕΩΡΗΜΑ

Εάν οι μεταβλητές y_i, x_i μετρούνται για κάθε μονάδα ενός απλού τυχαίου δείγματος μεγέθους $<n>$, έστω μεγάλου, η διακύμανση του r είναι :

$$V(r) = \frac{1-f}{n \cdot X^2} \frac{\sum_{i=1}^N (y_i - R x_i)^2}{N-1}$$

Όπου $f = n/N$

Για την εκτίμηση του τυπικού σφάλματος του r , θα έχουμε :

$$SE(r) = \sqrt{V(r)}$$

5.2.2. Εκτίμηση λόγου στη στρωματοποιημένη δειγματοληψία

Η εκτίμηση του λόγου στην περίπτωση της στρωματοποιημένης δειγματοληψίας γίνεται με δύο διαφορετικούς τρόπους.

Ένας τρόπος είναι να γίνει ξεχωριστή εκτίμηση του λόγου (separate ratio estimate) του \hat{Y} και \hat{X} μέσα σε κάθε στρώμα και στη συνέχεια να σχηματισθεί ένα σταθμισμένο ποσοστό (παίρνοντας σταθμίσεις ανάλογες προς το μέγεθος του στρώματος) από τις ξεχωριστές εκτιμήσεις σαν μια απλή εκτίμηση του λόγου του πληθυσμού.

Έστω ότι ένα πληθυσμός μεγέθους N διαιρείται σε k στρώματα. Συμβολίζοντας με το δείκτη i ένα συγκεκριμένο στρώμα ($i=1, 2, \dots, k$) και με το δείκτη j την συγκεκριμένη μονάδα δειγματοληψίας μέσα στο στρώμα, έχουμε τον ακόλουθο συμβολισμό :

- N_i = συνολικός αριθμός μονάδων στρώματος.
- n_i = αριθμός μονάδων δείγματος στο στρώμα i
- n = συνολικός μέγεθος δείγματος
- \underline{n}_i = κλάσμα δειγματοληψίας στο στρώμα i

- N_i
 Y_{ij} = τιμή του υπό μελέτη χαρακτηριστικού Y της j μονάδος ($j=1, 2, \dots, N_i$) του στρώματος i .
 X_{ij} = τιμή του βοηθητικού χαρακτηριστικού X της j μονάδος ($j = 1, 2, \dots, N_i$) του στρώματος i .
 y_{ij} = τιμή του υπό μελέτη χαρακτηριστικού y της j μονάδος του δείγματος ($j = 1, 2, \dots, N_i$) του στρώματος i .
 x_{ij} = τιμή του βοηθητικού χαρακτηριστικού x της j μονάδος του δείγματος του στρώματος i .
 $R_i = \bar{Y}_i / \bar{X}_i$, λόγος του μέσου πληθυσμού του στρώματος i
 $r_i = \bar{y}_i / \bar{x}_i$, λόγος του δειγματικού μέσου του στρώματος i
 $W_i = N_i/N$, στάθμιση στρώματος i

Μια εκτίμηση του λόγου του μέσου πληθυσμού \bar{Y} , είναι ο μέσος σταθμικός των εκτιμήσεων των i στρωμάτων που συμβολίζεται με \hat{y}_{srs} (ο δείκτης srs σημαίνει ξεχωριστή εκτίμηση λόγου στην στρωματοποιημένη δειγματοληψία) και υπολογίζεται από την σχέση :

$$\hat{y}_{srs} = \sum_{i=1}^k W_i r_i \bar{X}_i$$

Θα πρέπει να τονίσουμε ότι η εκτίμηση \hat{y}_{srs} είναι μια μεροληπτική εκτίμηση. Η διακύμανση της εκτίμησης \hat{y}_{srs} είναι :

$$V(\hat{y}_{srs}) = \sum_{i=1}^k W_i^2 \frac{(N_i - n)}{N_i} \frac{\sum_{j=1}^{n_i} (y_{ij} - r_i x_{ij})^2}{n_i(n_i - 1)}$$

Το τυπικό σφάλμα είναι :

$$SE(\hat{y}_{srs}) = \sqrt{V(\hat{y}_{srs})}$$

Η εκτίμηση του συνολικού πληθυσμού Y είναι :

$$\hat{Y} = N \hat{y}_{srs}$$

Με το δεύτερο τρόπο υπολογίζεται πρώτα το \hat{Y} από την εκτίμηση \hat{y}_{st} και το \bar{X} από την εκτίμηση \hat{x}_{st} . Στη συνέχεια ο λόγος $\hat{y}_{st} / \hat{x}_{st}$ χρησιμοποιείται σαν μια εκτίμηση του λόγου \bar{Y} / \bar{X} . Αυτή η εκτίμηση ονομάζεται συνδυασμένη εκτίμηση λόγου (combined ratio estimate).

Μια εκτίμηση του λόγου του μέσου πληθυσμού \dot{Y} δίνεται από την σχέση (ο δείκτης crs σημαίνει συνδυασμένη εκτίμηση λόγου στην στρωματοποιημένη δειγματοληψία) :

$$\dot{y}_{crs} = [\dot{y}_{st} / \dot{x}_{st}] \cdot X = r_{st} \cdot X$$

Η εκτίμηση αυτή είναι μεροληπτική. Στην περίπτωση της συνδυασμένης λόγου, επιτυγχάνεται μείωση της μεροληψίας, όταν ο αριθμός των στρωμάτων k είναι μεγάλος και το μέγεθος των δειγμάτων είναι μικρό .

Η διακύμανση της εκτίμησης \dot{y}_{crs} είναι :

$$V(\dot{y}_{crs}) = \sum_{i=1}^k W_i^2 \frac{(N_i - n_i)}{N_i} \frac{\sum_{j=1}^{n_i} (y_{ij} - r_{st} x_{ij})^2}{n_i(n_i - 1)}$$

Όπου $r_{st} = \dot{y}_{st} / \dot{x}_{st}$.

Στο σημείο αυτό δεν αναφέρουμε τους τύπους του τυπικού σφάλματος και της εκτίμησης του συνολικού πληθυσμού γιατί είναι τα ίδια όπως στην ξεχωριστή εκτίμηση λόγου .

Συγκρίνοντας την ξεχωριστή εκτίμηση λόγου με την συνδυασμένη εκτίμηση λόγου μπορεί να δεχθεί ότι με την ξεχωριστή εκτίμηση λόγου επιτυγχάνεται μεγαλύτερη ακρίβεια, όταν τα δείγμα μέσα σε κάθε στρώμα είναι αρκετά μεγάλο (πάνω από 20 μονάδες). Ενώ αντίθετα αν το δείγμα είναι μικρό μέσα σε κάθε στρώμα, είναι προτιμότερο να χρησιμοποιείται η συνδυασμένη εκτίμηση λόγου

5 - 3 - Εκτίμηση παλινδρόμησης

Όπως αναφέραμε σε προηγούμενη παράγραφο η εκτίμηση λόγου είναι αποτελεσματική όταν υπάρχει γραμμική συσχέτιση μεταξύ της μεταβλητής Y και μιας βοηθητικής μεταβλητής X και η ευθεία γραμμή να διέρχεται από την αρχή των αξόνων .

Πολλές φορές η συσχέτιση είναι περίπου γραμμική και η ευθεία δεν διέρχεται από την αρχή των αξόνων, τότε ενδείκνυται η εκτίμηση που βασίζεται στην γραμμική παλινδρόμηση του y_i επί του x_i αντί της εκτίμησης του λόγου .

Όπως στην εκτίμηση λόγου έτσι και στις εκτιμήτριες παλινδρόμησης το χαρακτηριστικό x (βοηθητική μεταβλητή) είναι γνωστό.

είτε από προηγούμενες μελέτες , είτε μπορεί να μετρηθεί πιο εύκολα παρά το χαρακτηριστικό y . Έτσι το x θεωρείται σταθερά, ενώ το y είναι τυχαία μεταβλητή.

Έστω ότι η τιμές y_i και x_i λαμβάνονται για κάθε μονάδα στο δείγμα και ο μέσος του πληθυσμού X είναι γνωστό. Η εκτίμηση της γραμμικής παλινδρόμησης του μέσου πληθυσμού Y είναι :

$$\hat{y}_{lr} = \hat{y} + b (X - \bar{x})$$

Όπου ο δείκτης lr σημαίνει γραμμική παλινδρόμηση (linear regression) το b ονομάζεται συντελεστής παλινδρόμησης και παριστάνει τη μεταβολή που υφίσταται η μεταβολή y όταν η μεταβλητή x αυξηθεί κατά μονάδα .

Το b στις περισσότερες εφαρμογές υπολογίζεται από στοιχεία δείγματος . Ορισμένες φορές επιλέγεται εκ των προτέρων μια τιμή σταθερή b_0 η οποία βασίζεται σε υπολογισμούς των τιμών b προηγούμενων ερευνών, όπου είχε βρεθεί ότι το b παραμένει σχεδόν σταθερό .

5.3.1. Εκτίμηση παλινδρόμησης χρησιμοποιώντας απλή τυχαία δειγματοληψία

Η εκτίμηση της γραμμικής παλινδρόμησης του μέσου πληθυσμού Y στην περίπτωση της απλής τυχαίας δειγματοληψίας όπου το b_0 επιλέγεται σαν σταθερή τιμή, είναι :

$$\hat{y}_{lr} = \hat{y} + b_0 (X - \bar{x})$$

Η εκτίμηση αυτή είναι μια αμερόληπτη εκτίμηση με διακύμανση :

$$V(\hat{y}_{lr}) = \frac{(N-n)^{-1}}{N} \frac{\sum_{i=1}^N [(y_i - \hat{Y}) - b_0(x_i - \bar{X})]^2}{n(N-1)}$$

Για τον προσδιορισμό της καλύτερης τιμής του b_0 παίρνουμε την τιμή η οποία ελαχιστοποιεί την διακύμανση και δίνεται από την σχέση :

$$b_0 = \frac{\sum_{i=1}^N (y_i - \bar{Y})(x_i - \bar{X})}{\sum_{i=1}^N (x_i - \bar{X})^2}$$

Η επιλογή μιας αξιόπιστης τιμής του b_0 , όπως αναφέραμε παραπάνω μπορεί να προσδιοριστεί από στοιχεία προηγούμενων ερευνών. Όταν υπάρχει δυσκολία προσδιορισμού του b εκ των προτέρων, τότε υπολογίζεται με στοιχεία δείγματος .

5.3.2 Εκτίμηση παλινδρόμησης στη στρωματοποιημένη δειγματοληψία

Στην περίπτωση της στρωματοποιημένης δειγματοληψίας, η εκτίμηση παλινδρόμησης γίνεται με δύο τρόπους όπως και στην εκτίμηση του λόγου .

Δηλαδή ο ένας τρόπος είναι να γίνει ξεχωριστή εκτίμηση παλινδρόμησης (separate regression estimate) μέσα σε κάθε στρώμα i , δηλαδή η εκτίμηση του μέσου σε κάθε στρώμα είναι :

$$\hat{y}_{i|s} = \hat{y}_i + b_i (\bar{X}_i - \bar{x}_i)$$

Που ο δείκτης $i|s$ σημαίνει γραμμική παλινδρόμηση στο στρώμα i . Η εκτίμηση του μέσου πληθυσμού Y είναι :

$$\hat{y}_{lrs} = \sum_{i=1}^k W_i \hat{y}_{i|s}$$

Όπου ο δείκτης lrs σημαίνει ξεχωριστή εκτίμηση παλινδρόμησης και $W_i = N_i / N$.

Η εκτίμηση αυτή θεωρείται κατάλληλη όταν ο συντελεστής παλινδρόμησης διαφέρει από στρώμα σε στρώμα .

Με το δεύτερο τρόπο . ο οποίος ονομάζεται συνδυασμένη εκτίμηση παλινδρόμησης (combined regression estimate), υπολογίζονται πρώτα οι μέσοι των στρωμάτων \hat{y}_i και \bar{x}_i από τις σχέσεις :

$\dot{y}_{st} = \sum_{i=1}^k W_i \dot{y}_i$
$\dot{x}_{st} = \sum_{i=1}^k W_i \dot{x}_i$

Στην συνέχεια προκύπτει η εκτίμηση του μέσου πληθυσμού \dot{Y} από την σχέση :

$$\dot{y}_{lrc} = \dot{y}_{st} + b(\dot{X}_i - \dot{x}_{st})$$

Όπου ο δείκτης lrc σημαίνει συνδυασμένη εκτίμηση παλινδρόμησης .

Η εκτίμηση αυτή θεωρείται κατάλληλη όταν ο συντελεστής παλινδρόμησης είναι ίδιος σε όλα τα στρώματα .

5.4 Συμπέρασμα

Θα πρέπει να επισημανθεί η διαφοροποίηση της μεθόδου της παλινδρόμησης, σε σχέση με την μέθοδο του λόγου. Με τη νέα μέθοδο προσδιορίζεται κατ' αρχήν, η μέση μεταβολή του χαρακτηριστικού Y ανα μονάδα μεταβολής του χαρακτηριστικού X , που εκφράζεται με τον συντελεστή b , και στην συνέχεια πολλαπλασιάζεται, ως εφαρμογή τη μεθόδου λόγου, με την διασπορά της μέσης τιμής του πληθυσμού από την αντίστοιχη δειγματική μέση τιμή του χαρακτηριστικού X . Για το λόγο αυτό η μέθοδος της παλινδρομήσεως θεωρείται καταλληλότερη, ως προς τη μέθοδο του λόγου, και μπορεί να εφαρμοσθεί ακόμα και για διαχρονική ανάλυση, με τη βασική προϋπόθεση ότι δεν σημειώνονται ουσιώδεις αλλαγές στη δομή του πληθυσμού και το δείγμα .

Θα πρέπει να επισημανθεί επίσης ότι, εάν η συναρτησιακή σχέση μεταξύ των χαρακτηριστικών Y και X είναι μη γραμμική . Τότε η εκτίμηση του μεγέθους υπόκειται σε μεροληψία της τάξεως $1/n$, ενώ ο λόγος της μεροληψίας προς το αντίστοιχο τυπικό σφάλμα είναι μικρός για μεγάλα δείγματα .

Έτσι τα πλεονεκτήματα τα οποία παρουσιάζονται στην εκτίμηση λόγου είναι τα εξής :

Πλεονεκτήματα

1. οι εκτιμήσεις που προκύπτουν από την σχέση μεταξύ του μετρούμενου μεγέθους y και του βοηθητικού χαρακτηριστικού x , είναι καλύτερες, υπό την προϋπόθεση ότι υπάρχει γραμμική συσχέτιση δηλαδή ευθεία γραμμή το οποίο περνάει από την αρχή των αξόνων. Καθώς και η μεταβλητότητα του y είναι ανάλογη του x .
2. ότι δίνει την δυνατότητα να χρησιμοποιηθεί και στην στρωματοποιημένη δειγματοληψία .

Μειονεκτήματα

1. υπάρχει η περίπτωση που το κόστος για την συλλογή στοιχείων και πληροφοριών για τον βοηθητικό χαρακτηριστικό να είναι παρά πολύ υψηλό. Έτσι είναι καλύτερα να μην εφαρμόσουμε τον μέθοδο της εκτίμησης του λόγου αλλά κάποιο άλλο .
2. και είναι η μέθοδος που οδηγεί σε αμερόληπτες εκτιμήτριες .

ΚΕΦΑΛΑΙΟ	Δισταδιακή δειγματοληψία
6.	

6.1 Γενικά

6.1.1 Εισαγωγικές έννοιες

Η μεθοδολογία που αναπτύχθηκε στα προηγούμενα κεφάλαια στηρίχθηκε, κατά κύριο λόγο, στην τυχαία εκλογή των στοιχείων του πληθυσμού. Θα πρέπει να τονισθεί, ιδιαίτερα, ότι στη μεθοδολογία αυτή συνέπιπτε, σχεδόν κατά κανόνα, η έννοια της επλεγόμενης με τη διερευνώμενη μονάδα. Η σχηματική αυτή διεργασία, διατηρήθηκε στην απλή τυχαία δειγματοληψία, αλλά και στην στρωματοποιημένη, όπου οι τυχόν ομαδοποιήσεως των στοιχείων συνιστούσαν μια στρωματοποίηση του πληθυσμού και στην συνέχεια επέτρεπαν την εφαρμογή των αρχών την τυχαίας δειγματοληψίας.

Σε πληθυσμούς μεγάλου μεγέθους, η διαδικασία αυτή αποδεικνύεται χρονοβόρα, απαιτεί συνήθως μεγάλο κόστος και μερικές φορές είναι αδύνατον να διεκπεραιωθεί. Για το λόγο αυτό αποδεικνύεται χρήσιμη η εφαρμογή των αρχών της τυχαίας δειγματοληψίας σε δυο ή περισσότερα στάδια. Συγκεκριμένα, τα στοιχεία ενός πληθυσμού διαμοιράζονται σε ομάδες, με βάση ένα συγκεκριμένο χαρακτηριστικό, ώστε να καλύπτεται ολόκληρος πληθυσμός μονοσήμαντα. Δηλαδή, κάθε στοιχείο του πληθυσμού να ανήκει οπωσδήποτε σε κάποια ομάδα και σε μια μόνο. Είναι φανερό ότι το χαρακτηριστικό, στο οποίο θα στηριχθεί η διαμερίσει του πληθυσμού, θα είναι ευρύτερο και δεν θα συμπίπτει με το διερευνώμενο χαρακτηριστικό του πληθυσμού, με τον όρο ότι κάθε ομάδα θα περιέχει περισσότερα από ένα στοιχεία του πληθυσμού.

Ως παράδειγμα μπορεί να αναφερθεί η διερεύνηση μίας ολότητας, την οποία αποτελεί ο εκλογικά νόμιμος πληθυσμός μιας αρκετά μεγάλης πόλεως. Η πόλη είναι διαιρεμένη σε 1.000 εκλογικά τμήματα των 400

περίπου εκλογέων το καθένα . Για τη δειγματική εκλογή 100 εκλογικών τμημάτων μπορεί να εφαρμοσθεί η τυχαία ή ακόμα η συστηματική δειγματοληψία, καθ' όσον δεν εμφανίζεται ικανό κώλυμα για την εφαρμογή της. Για την διερεύνηση, π.χ. ενός κοινωνικού χαρακτηριστικού, είναι δυνατόν, να ληφθούν όλοι οι εκλογές των 100 τμημάτων ή να εφαρμοσθεί σε δεύτερο στάδιο η τυχαία δειγματοληψία. Στην πρώτη περίπτωση γίνεται λόγος για *δειγματοληψία κατά δεσμίδες* και στη δεύτερη για *δισταδιακή δειγματοληψία*. Και στις δύο περιπτώσεις επιτυγχάνεται μια μείωση του κόστους της δειγματοληψίας με την εκλογή συγκεκριμένων *δεσμίδων ή ομάδων* του πληθυσμού που δίνουν συγκεντρωτικότερη αντιπροσώπευση των στοιχείων του. Είναι προφανές ότι στη δισταδιακή δειγματοληψία το κόστος θα είναι ακόμα μικρότερο, λόγω του απαιτούμενου μικρότερου δειγματικού μεγέθους. Είναι, επίσης, προφανές ότι η υποδιαίρεση του πληθυσμού σε δεσμίδες, δεν μπορεί να στηριχθεί στο ιδιαίτερο διερευνώμενο χαρακτηριστικό των στοιχείων του πληθυσμού, αλλά σε πολύ γενικότερα χαρακτηριστικά που θα διευκολύνουν την υποδιαίρεση όσο το δυνατόν περισσότερο από οικονομικής και τεχνικής απόψεως. Ακόμα, είναι δυνατόν το κριτήριο ή το χαρακτηριστικό της τυχαίας εκλογής μιας δεσμίδας, από το σύνολο των δεσμίδων, να στηριχθεί σε εντελώς νέα χαρακτηριστικά που ικανοποιούν περισσότερο τις αρχές της τυχαίας δειγματοληψίας .

Απλό και συνηθισμένο παράδειγμα, για τη δυνατή ή αναγκαία χρησιμοποίηση νέου χαρακτηριστικού στην κατασκευή των δεσμίδων, αποτελεί η διερεύνηση του γεωγραφικά κατανεμημένου πληθυσμού, ως προς διάφορα κοινωνικά και οικονομικά χαρακτηριστικά. Σε πολλές πόλεις αυτό καθιστά αδύνατη την εφαρμογή τυχαίας ή συστηματικής δειγματοληψίας με βάση το διερευνώμενο χαρακτηριστικό, για κάθε στοιχείο του πληθυσμού, που στην περίπτωση αυτή αποτελεί το νοικοκυριό. Για το λόγο αυτό καθορίζονται, π.χ. συγκροτήματα κατοικιών ή, σε πρώτο στάδιο, οικοδομικά τετράγωνα, με βάση την πληθυσμιακή πυκνότητα και εφαρμόζεται εκεί, σε δεύτερο στάδιο, η τυχαία δειγματοληψία .

6. 1. 2. Το πρόβλημα της κατασκευής των δεσμίδων

Βασική απαιτούμενων προϋπόθεση κατά το σχηματισμό των δεσμίδων αποτελεί η μέγιστη δυνατή ανομοιογένεια των στοιχείων κάθε δεσμίδας. Το γεγονός αυτό οδηγεί σε μια ομοιογενοποίηση του πληθυσμού που αποτελούν οι δεσμίδες, στο οποίο μπορεί να εφαρμοσθεί η τυχαία δειγματοληψία. Σε διαφορετική περίπτωση, όταν δηλαδή οι

δεσμίδες περιέχου, κατά το μάλλον ή ήττον, ομοιογενή στοιχεία, έτσι ώστε να δημιουργούνται ανομοιογενείς δεσμίδες εμφανίζεται το επονομαζόμενο φαινόμενο δεσμίδας .

Το κύριο χαρακτηριστικό κατά την πολυσταδιακή δειγματοληψία είναι ότι το επιλεγόμενο στοιχείο-δεσμίδα (πρωτογενές στοιχεία), σε κάθε στάδιο της δειγματοληψίας, περιέχει, συνήθως, ένα αρκετά αριθμό υπό στοιχείων (δευτερογενή στοιχεία) του επομένου σταδίου. Το πρόβλημα που ανακύπτει, αμέσως αφορά, αφ' ενός μεν τον τρόπο κατασκευής των δεσμίδων, αφ' ετέρου δε τη σχέση που πρέπει να υπάρχει μεταξύ των μεγεθών των δεσμίδων, δηλαδή τον αριθμό των στοιχείων που πρέπει να περιέχει κάθε δεσμίδα, σε σχέση με τις υποδεσμίδες - στοιχεία που περιέχει .

Σε πολύ μεγάλες ολότητες συνιστάται η στρωματοποίηση των πρωτογενών στοιχείων-δεσμίδων και η εφαρμογή, στη συνέχεια, μιας πολυσταδιακής στρωματοποιημένης δειγματοληψίας. Στην προκειμένη περίπτωση ενδείκνυται η εφαρμογή συστηματική δειγματοληψίας, καθ' όσον το μεγάλο μέγεθος του πληθυσμού αποκλείει, συνήθως, τα συστηματικά σφάλματα .

Στην πολυσταδιακή δειγματοληψία θεωρείται σκόπιμο κάθε δεσμίδα στην πρώτη διαμερίσει να περιέχει ίσο και αρκετά μεγάλο αριθμό υποδεσμίδων - στοιχείων της δεύτερης διαμερίσει. Παρόμοιος κανόνας θα πρέπει να ισχύει για τη δεύτερη, σε σχέση με την τρίτη διαμερίσει κ.ο.κ.

Ουσιαστικό, επίσης κριτήριο κατά την πολυσταδιακή δειγματοληψία, όπου είναι δεδομένο το συνολικό κόστος, αποτελεί η ελαχιστοποίηση της διασποράς του χαρακτηριστικού, στο οποίο στηρίζεται η δειγματοληψία. Στη περίπτωση αυτή εμφανίζονται ορισμένα προβλήματα προσδιορισμού των παραμέτρων. Συγκεκριμένα, για τον προσδιορισμό της συναρτήσεως κόστους ποσοτικών μεταβλητών είναι αναγκαίο να προσδιορισθούν οι τιμές της διασποράς σε κάθε δεσμίδα και μεταξύ των δεσμίδων από προηγούμενες δειγματοληψίες και μάλιστα για μια σειρά περιόδων, ώστε να αποκτηθεί ουσιαστική εμπειρία για τις τιμές των παραμέτρων .

Κατά τη δειγματοληψία διερευνώντας συγχρόνως πολλά χαρακτηριστικά, όπου οι διασπορές τους διαφέρουν, συνήθως, σημαντικά. Για το λόγο αυτό, θα πρέπει το κριτήριο επιλογής των δεσμίδων να στηριχθεί στα κυριότερα από αυτά, ώστε να αποφευχθούν, κατά το δυνατόν, ισχυρά αντικρουόμενες διαμερίσει του πληθυσμού .

Τέλος ως κατακλείδα θα πρέπει να αναφερθεί ότι η αύξηση των σταδίων της δειγματοληψίας δεν οδηγεί, κατά κανόνα, σε καλύτερα αποτελέσματα, καθ' όσον υφίσταται στενή σχέση μεταξύ σχεδιασμού της δειγματοληψίας και μεγέθους συστηματικού σφάλματος. Θα πρέπει να σημειωθεί, ιδιαίτερα, ότι όσο πολύπλοκος είναι ο σχεδιασμός, τόσο μεγαλύτερος είναι ο κίνδυνος των επιπρόσθετων σφαλμάτων κατά την πρακτική εφαρμογή της δειγματοληψίας. Για το λόγο αυτό, στα επόμενα του κεφαλαίου αυτού, θα περιορισθεί η ανάπτυξη της μεθοδολογίας στη δισταδιακή δειγματοληψία .

6 . 2 Εισαγωγή στην δισταδιακή δειγματοληψία

Θεωρούμε τις ακόλουθες ομάδες προβλημάτων δειγματοληψίας :

1. Μια μελέτη της περιεχόμενης βιταμίνης Α στο βούτυρο που παράγεται από τα βουτυροποιεία .
2. Μια μελέτη της περιεχόμενης προτεΐνης στο σιτάρι των σιτοκαλλιεργειών μιας περιοχής .
3. Μια μελέτη της απარიθμήσεως των ερυθρών αιμοσφαιρίδιων σε ένα πληθυσμό ανδρών ηλικίας 20- 30 ετών .
4. Μια μελέτη των εντόμων που συγκεντρώνονται στα φύλλα των οπωροφόρων δένδρων σε ένα δενδρόκηπο .
5. Μια μελέτη του αριθμού των ελαττωματικών δοντιών από παιδιά της τρίτης τάξεως, των σχολείων μιας μεγάλης πόλεως .

« Τι κοινό έχουν αυτές οι έρευνες μεταξύ τους ; »

Πρώτον : σε κάθε μελέτη η κατάλληλη μονάδα δειγματοληψίας καθορίζεται εκ το φύσεως της - το βουτυροποιείο, : ο αγρός σίτου, ο άνθρωπος, το δένδρο και το σχολείο.

Δεύτερον : και αυτό είναι το κύριο σημείο, σε κάθε μελέτη οι επιλεγόμενες μονάδες δειγματοληψίας μπορούν να υποδιαιρεθούν δειγματοληπτικά, αντί να μετρηθούν τελείως. Πράγματι, η υποδειγματοληψία (sub-sampling) είναι απαραίτητη στις τρεις πρώτες μελέτες. Κανείς δεν πρόκειται να μας επιτρέψει να λάβουμε όλο το βούτυρο που παράχθηκε από ένα βουτυροποιείο, για να καθορίσουμε την

περιεχόμενη βιταμίνη Α ή όλο το σιτάρι ενός αγρού για τον καθορισμό της περιεχόμενης πρωτεΐνης, ή όλο το αίμα ενός ανθρώπου για να επιτύχουμε μια τέλεια μέτρηση των ερυθρών αιμοσφαιρίων του. Στην μελέτη των συγκεντρωμένων εντόμων, μπορεί να είναι εφικτό, καίτοι κουραστικό, να εξετάσουμε όλα τα φύλλα σε οποιοδήποτε δέντρο που επιλέξαμε. Εάν η κατανομή των εντόμων είναι ανομοιογενής, πάντως, πιθανόν να αποφασίσουμε να πάρουμε μόνο ένα μικρό δείγμα φύλλων από κάθε δέντρο που επιλέγουμε, ούτως ώστε να περιλάβουμε στην έρευνα περισσότερα δέντρα. Στην οδοντιατρική έρευνα μπορούμε να πάρουμε όλα τα παιδιά της τρίτης τάξεως σε κάθε επιλεγόμενο σχολείο ή θα μπορούσαμε να καλύψουμε ένα μεγαλύτερο δείγμα σχολείων ή θα μπορούσαμε να καλύψουμε ένα μεγαλύτερο δείγμα σχολείων εξετάζοντας μόνο ένα δείγμα παιδιών από την τρίτη τάξη κάθε επιλεγόμενου σχολείου

Αυτός ο τρόπος δειγματοληψίας ονομάζεται δειγματοληψία σε δυο στάδια (sampling in two stages) ή μερικές φορές υποδειγματοληψία (sub-sampling). Συγκεκριμένα, κατά το πρώτο στάδιο, επιλέγεται με τυχαίο τρόπο ένα δείγμα ομάδων (π.χ. οικοδομικά τετράγωνα). Οι ομάδες οι οποίες αποτελούν τις μονάδες δείγματος κατά το πρώτο στάδιο ονομάζονται πρωτογενείς μονάδες (first stage units or primary units). Κατά το δεύτερο στάδιο επιλέγεται με τυχαίο τρόπο ένας αριθμός στοιχείων (π.χ. νοικοκυριά) μέσα σε κάθε πρωτογενή μονάδα που έχει επιλεγεί στο δείγμα. Τα στοιχεία αυτά αποτελούν τις μονάδες δείγματος κατά το δεύτερο στάδιο και ονομάζονται δευτερογενείς μονάδες (second-stage units or secondary units).

Με τη μέθοδο αυτή συντομεύεται ο χρόνος διεξαγωγής της έρευνας και μειώνεται το κόστος, όπως και στην περίπτωση της δειγματοληψίας κατά ομάδες. Επίσης με την εφαρμογή της δισταδιακής δειγματοληψίας δεν απαιτούνται δειγματοληπτικά πλαίσια για τον σύνολο των μονάδων του ερευνώμενου πληθυσμού παρά μόνο για τις ομάδες που έχουν επιλεγεί στο δείγμα .

Η εφαρμογή της δισταδιακής δειγματοληψίας προϋποθέτει κατάλληλο διαχωρισμό του ερευνώμενου πληθυσμού σε ομάδες . Κατά τον διαχωρισμό του πληθυσμού σε ομάδες, πρέπει να λαμβάνεται υπόψη η ομοιογένεια των ομάδων, ως προς το ερευνώμενο χαρακτηριστικό. Κατά κανόνα τα ερευνώμενα χαρακτηριστικά παρουσιάζουν μεγαλύτερη ανομοιογένεια, στις ομάδες μεγάλου μεγέθους, οπότε απαιτείται μεγαλύτερο δείγμα. Ενώ αντίθετα ομάδες μικρού μεγέθους είναι περισσότερο ομοιογενείς, ως προς το ερευνώμενο χαρακτηριστικό, οπότε απαιτείται μικρότερο δείγμα .

Για παράδειγμα, αν θέλουμε να μελετήσουμε το εισόδημα των νοικοκυριών που διαμένουν σε μια μεγάλη πόλη και χωρίσουμε τα νοικοκυριά σε μεγάλες ομάδες (π.χ. συνοικίες) ή σε μικρότερες ομάδες (οικοδομικά τετράγωνα). Αναμένεται ότι τα νοικοκυριά που διαμένουν σε μια συνοικία θα παρουσιάζουν μεγαλύτερη ανομοιογένεια, ως προς το εισόδημα, από ότι τα νοικοκυριά που διαμένουν στο ίδιο οικοδομικό τετράγωνο. Για το λόγο αυτό ο διαχωρισμός των νοικοκυριών σε ομάδες κατά συνοικία, απαιτεί μεγαλύτερο δείγμα (αριθμό νοικοκυριών που θα ερευνηθούν), από ότι σε ομάδες κατά οικοδομικό τετράγωνο .

6.3 Συμβολισμοί

Στη δισταδιακή δειγματοληψία όταν οι μονάδες στα στάδια είναι άνισου μεγέθους μεταξύ τους, χρησιμοποιούνται οι παρακάτω συμβολισμοί :

N = ο αριθμός των ομάδων (πρωτογενείς μονάδες) στον πληθυσμό
 n = ο αριθμός των ομάδων που έχουν επιλεγεί στο δείγμα
 M_i = ο αριθμός των στοιχείων (δευτερογενείς μονάδες) στην i ομάδα
 m_i = ο αριθμός των στοιχείων που έχουν επιλεγεί από την i ομάδα

$$M = \sum_{i=1}^N M_i, \text{ αριθμός στοιχείων στο πληθυσμό}$$

$$m = \sum_{i=1}^n m_i, \text{ ο συνολικός αριθμός στοιχείων που έχουν επιλεγεί}$$

$$\bar{W} = M / N, \text{ μέσο μέγεθος ομάδων στο πληθυσμό}$$

$$\bar{w} = \frac{1}{n} \sum_{i=1}^n m_i, \text{ μέσο μέγεθος στοιχείων κατά ομάδα στο δείγμα}$$

Y_{ij} = τιμή του υπό μελέτη χαρακτηριστικού για την j δευτερογενή μονάδα ($j = 1, 2, \dots, M$) από την i ομάδα του πληθυσμού .

y_{ij} = τιμή του υπό μελέτη χαρακτηριστικού της j δευτερογενή μονάδα από την i ομάδα του δείγματος .

$$\bar{Y}_i = \frac{1}{M} \sum_{j=1}^M Y_{ij}, \text{ μέσο κατά δευτερογενή μονάδα της } i \text{ πρωτογενούς μονάδας}$$

$$\bar{Y} = \frac{1}{NM} \sum_{j=1}^M \sum_{i=1}^N Y_{ij}, \text{ μέσος πληθυσμού,}$$

$$\bar{y}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{ij}, \text{ ο μέσος δείγματος για την } i \text{ ομάδα}$$

Επειδή στην πράξη οι μονάδες είναι άνισου μεγέθους μεταξύ τους, στα στάδια δειγματοληψίας, για το λόγο αυτό δεν θα ασχοληθούμε με την περιγραφή της διαδικασίας εκτιμήσεων για μονάδες ίσου μεγέθους.

6.4 Εκτίμηση του μέσου και του συνολικού πληθυσμού

Μια αμερόληπτη εκτίμηση του μέσου \bar{Y} του πληθυσμού, είναι ο μέσος του δείγματος ο οποίος συμβολίζεται με \hat{y}_{ts} (ο δείκτης ts σημαίνει δισταδιακή δειγματοληψία-two stage sampling) και είναι :

$$\hat{y}_{ts} = \frac{N}{Mn} \sum_{i=1}^n M_i \hat{y}_i$$

Μια αμερόληπτη εκτίμηση της διακύμανσης του \hat{y}_{ts} είναι :

$$V(\hat{y}_{ts}) = \frac{N-n}{N} \frac{1}{n\omega^2} s^2_1 + \frac{1}{nN\omega^2} \sum_{i=1}^n \frac{M_i(M_i - m_i)}{m_i} s^2_{2i}$$

Όπου

s^2_1 = είναι η διακύμανση του πληθυσμού και
 s^2_{2i} = είναι η διακύμανση του δείγματος στην ομάδα i ($i = 1, 2, \dots, n$)

Το τυπικό σφάλμα είναι :

$$SE(\hat{y}_{ts}) = \sqrt{V(\hat{y}_{ts})}$$

Η διακύμανση της εκτίμησης του συνολικού πληθυσμού Y είναι :

$$V(\hat{Y}) = \frac{N-n}{N} \frac{N^2}{n} s^2_1 + \frac{N}{n} \sum_{i=1}^n \frac{M_i(M_i - m_i)}{m_i} s^2_{2i}$$

Το τυπικό σφάλμα είναι :

$$SE(\hat{Y}) = SE(M\hat{y}_{ts}) = \sqrt{M^2 V(\hat{y}_t)}$$

6.5 Προσδιορισμός μεγέθους δείγματος.

Ο προσδιορισμός μεγέθους στην περίπτωση της δισταδιακής δειγματοληψίας, είναι πολύ πιο δύσκολος από ότι στις άλλες μεθόδους δειγματοληψίας. Αυτό οφείλεται στο γεγονός ότι πρέπει να επιλεγούν σε συνδυασμό, οι μονάδες δείγματος (αριθμός ομάδων δείγματος) και οι μονάδες (αριθμός στοιχείων μέσα σε κάθε ομάδα), διότι μπορεί να υπάρχουν διαφορές μεταξύ των ομάδων και διαφορές μεταξύ των στοιχείων μέσα σε κάθε ομάδα. Η επιλογή των δειγματοληπτικών μονάδων στα δύο στάδια, πρέπει να βασίζεται στην ομοιογένεια των μονάδων. Δηλαδή, αν τα στοιχεία εντός των ομάδων είναι ομοιογενή, αλλά οι μέσοι των ομάδων διαφέρουν σημαντικά από ομάδα σε ομάδα, τότε παίρνουμε μεγαλύτερο δείγμα ομάδων με μικρό αριθμό στοιχείων. Εάν τα στοιχεία μέσα στις ομάδες διαφέρουν σημαντικά, αλλά οι μέσοι των ομάδων είναι ομοιογενείς, τότε παίρνουμε μικρότερο δείγμα ομάδων με μεγαλύτερο αριθμό στοιχείων.

Για απλούστευση του προβλήματος θεωρούμε ότι όλες οι ομάδες είναι ισομεγέθεις, δηλαδή $M_1 = M_2 = M_3 = \dots = M_N = \omega$ και $m_1 = m_2 = m_3 = \dots = m_n = \omega$, οπότε στην παρούσα αντιμετώπιση της δειγματοληψίας δυο σταδίων, υποθέτουμε κατ' αρχή ότι οι πρωταρχικές μονάδες δειγματοληψίας είναι περίπου του ίδιου μεγέθους. Ένα απλό τυχαίο δείγμα n πρωταρχικών μονάδων δειγματοληψίας λαμβάνεται και ο ίδιος αριθμός m υπομονάδων επιλέγεται από κάθε πρωταρχική μονάδα δειγματοληψίας στο δείγμα.

Το εκτιμώμενο τυπικό σφάλμα του δειγματικού μέσου \bar{y} ανά υπό-μονάδα δίνεται από τον τύπο :

$$SE(\bar{y}_s) = \frac{1}{\sqrt{n}} \sqrt{\sum (\hat{y}_i - \bar{y})^2 / (n-1)}$$

Όπου \hat{y}_i είναι ο μέσος ανά υπό-μονάδα στην i πρωταρχική μονάδα δειγματοληψίας. Αυτός τύπος δεν περιέχει την διόρθωση πεπερασμένου πληθυσμού, αλλά είναι αρκετά αξιόπιστος αρκεί το δείγμα να περιέχει λιγότερο από 10% όλων των πρωταρχικών μονάδων δειγματοληψίας. Σημειώστε ότι ο τύπος δεν κάνει χρήση των ατομικών παρατηρήσεων επί των υπό-μονάδων, αλλά μόνο των μέσων \hat{y}_i των πρωταρχικών μονάδων δειγματοληψίας. Εάν τα υπό-δείγματα λαμβάνονται για μια χημική ανάλυση, μια πρακτική είναι να συνθέσουμε το υπό-δείγμα και να κάνουμε ένα χημικό υπολογισμό για κάθε πρωταρχική μονάδα δειγματοληψίας.

Στην ανάλυση διακυμάνσεως, που γίνεται ανάλυση της διακυμάνσεως στα συστατικά μέρη αυτής, δίνονται και εφαρμογές σε προβλήματα δειγματοληψίας σε δύο στάδια.

Με εφαρμογή της τεχνικής της αναλύσεως διακυμάνσεως, μπορούμε να δούμε πώς η διακύμανση του δειγματικού μέσου επηρεάζεται από την μεταβλητικότητα μεταξύ μετρήσεων εντός της ίδιας μονάδας δειγματοληψίας και από την μεταβλητικότητα μεταξύ των μονάδων δειγματοληψίας. Θα μπορούσαμε επίσης να προβλέψουμε πως η διακύμανση του δειγματικού μέσου θα μετεβάλετε με διαφορετικούς αριθμούς μονάδων δειγματοληψίας και μετρήσεων ανά μονάδα δειγματοληψίας στο πείραμα.

Επειδή η τεχνική αυτή είναι η ευρείας χρησιμότητα στη δειγματοληψία δύο σταδίων, θα επαναλάβουμε μερικά από τα αποτελέσματα. Η παρατήρηση σε οποιαδήποτε υπό-μονάδα θεωρείται ότι είναι το άθροισμα δύο ανεξάρτητων όρων. Ένας όρος, που έχει σχέση με την πρωταρχική μονάδα δειγματοληψίας, έχει την ίδια τιμή για όλες τις μονάδες δεύτερου σταδίου της πρωταρχικής μονάδας δειγματοληψίας και ποικίλλει μεταξύ των πρωταρχικών μονάδων δειγματοληψίας με διακύμανση. Ο δεύτερος όρος, ο οποίος βοηθάει να μετρήσουμε διαφορές μεταξύ των υπό-μονάδων με διακύμανση $V(\bar{y}_s)$. Υποθέτουμε ότι ένα δείγμα αποτελείται από n πρωταρχικές μονάδες δειγματοληψίας, από κάθε μια των οποίων λαμβάνονται m υπό-μονάδες. Τότε το δείγμα σαν σύνολο περιέχει n ανεξάρτητες τιμές του πρώτου όρου, ενώ περιέχει nm

ανεξάρτητες τιμές του δεύτερου όρου. Συνεπώς, η διακύμανση του δειγματικού μέσου \bar{y} ανά υπό-μονάδα είναι :

$$V(\bar{y}_{ts}) = \frac{S^2_1}{n} + \frac{S^2_2}{nm}$$

6.6. Η κατανομή των πόρων στην δειγματοληψία δύο σταδίων

Το ίδιο τυπικό σφάλμα μπορεί να επιτευχθεί για τον δειγματικό μέσο με χρησιμοποίηση διάφορων συνδυασμών των τιμών n και m . Ποια από αυτές τις επιλογές είναι η πιο καλή ; Η απάντηση εξαρτάται, φυσικά, από το κόστος προθέσεως μιας επί πλέον πρωταρχικής μονάδας δειγματοληψίας στο δείγμα, σε σχέση με το κόστος προθέσεως μιας επί πλέον υπό-μονάδας σε κάθε πρωταρχική μονάδα δειγματοληψίας. Προφανώς, εάν είναι φτηνό να προσθέσουμε πρωταρχικές μονάδες δειγματοληψίας στο δείγμα αλλά δαπανηρό να προσθέσουμε υπό-μονάδες, το πιο οικονομικό σχέδιο θα είναι αυτό που έχει πολλές πρωταρχικές μονάδες δειγματοληψίας και λίγες (ίσως μια μόνο) υπό-μονάδες ανά πρωταρχική μονάδα δειγματοληψίας. Για μια γενική λύση σε αυτό, πάντως, χρειαζόμαστε μια περισσότερο ακριβή διατύπωση των κοστών των διαφόρων εναλλακτικών σχεδίων.

Προκειμένου να προσδιορίσουμε έναν βέλτιστο συνδυασμό μονάδων δείγματος n , από τις πρωτογενείς μονάδες και m από τις δευτερογενείς μονάδες, επιλέγουμε το n και m κατά τρόπο ώστε να ελαχιστοποιηθεί η διακύμανση για δεδομένο κόστος ή να ελαχιστοποιηθεί το συνολικό κόστος δειγματοληψίας για δεδομένη διακύμανση. Για να επιτευχθεί αυτή η ελαχιστοποίηση χρησιμοποιούμε τη συνάρτηση κόστος

$$C = nc_1 + nmc_2$$

Όπου το

C = συνολικός κόστος

c_1 = είναι το μέσο κόστος ανά πρωταρχική μονάδα δειγματοληψίας, εκείνων των στοιχείων του κόστους που εξαρτώνται αποκλειστικά από τον αριθμό των πρωταρχικών μονάδων δειγματοληψίας και όχι από την έκταση της υπό-δειγματοληψίας.

c_2 = είναι το μέσο κόστος ανά υπό-μονάδα εκείνων των συστατικών του κόστους που είναι ευθέως αναλογικά του συνολικού αριθμού των υπό-μονάδων.

Εάν εκ των προτέρων γίνουν εκτιμήσεις αυτών των συστατικών του κόστους με μια προκαταρκτική μελέτη, τότε μια αποτελεσματική εργασία επιλογής των πιο καλών συνόλων δειγματοληψίας και υπό-δειγματοληψίας μπορεί να γίνει. Το πρόβλημα μπορεί να τεθεί με δύο διαφορετικούς τρόπους. Σε μερικές μελέτες προκαθορίζουμε την επιθυμητή διακύμανση V για τον δειγματικό μέσο και επιθυμούμε να πετύχουμε αυτή (την διακύμανση) όσο το δυνατόν με μικρότερο κόστος. Σε άλλες εφαρμογές το συνολικό κόστος το οποίο πρέπει να μην υπερβούμε επιβάλλεται σε μας και θέλουμε να δώσουμε τόσο μικρή τιμή της διακυμάνσεως όσο μπορούμε για αυτή την προγραμματισμένη δαπάνη. Αυτά τα δύο προβλήματα έχουν βασικά την ίδια λύση. Σε κάθε περίπτωση θέλουμε να ελαχιστοποιήσουμε τον γινόμενο.

Η τιμή του m ελαχιστοποιεί την διακύμανση για δεδομένο κόστος ή ελαχιστοποιεί το κόστος για δεδομένη διακύμανση δίνεται από την σχέση :

$$m = \frac{\sqrt{c_1 S^2_2}}{\sqrt{c_2 S^2_1}}$$

6.7. Δισταδιακή δειγματοληψία με πιθανότητα ανάλογη προς το μέγεθος των μονάδων

Σε πολλά σπουδαία προβλήματα δειγματοληψίας, οι φυσικές πρωταρχικές μονάδες δειγματοληψίας ποικίλλουν σε μέγεθος. Σε εθνικές έρευνες που διεξάγονται για να λάβουμε πληροφορίες περί τα χαρακτηριστικά του πληθυσμού, η πρωταρχική μονάδα δειγματοληψίας είναι συνήθως μια διοικητική περιφέρεια (π.χ. νομός, δήμος κ.ο.κ.). Μια σχετικά μεγάλη πρωταρχική μονάδα δειγματοληψίας αυτού του τύπου περιορίζει τα κόστος ταξιδιών και καθιστά την επιθεώρηση και τον έλεγχο της επιτόπιας εργασίας περισσότερο εύκολη. Ένα δείγμα κατοικιών σε μια πόλη μπορεί να χρησιμοποιήσει οικοδομικά τετράγωνα σαν μονάδες δειγματοληψίας σε πρώτο στάδιο, ο αριθμός των κατοικιών ανά οικοδομικό τετράγωνο ποικίλλει από μηδέν έως σαράντα. Παρομοίως,

σχολεία νοσοκομεία και εργοστάσια περιέχουν διαφορετικούς αριθμούς ατομικών μονάδων έρευνας .

Πολλές φορές το μέγεθος των πρωτογενών μονάδων, διαφέρει σημαντικά ως προς τον αριθμό των στοιχείων (αριθμό δευτερογενών μονάδων) και υπάρχει έντονη συσχέτιση μεταξύ του μεγέθους των πρωτογενών μονάδων και των υπό εξέταση χαρακτηριστικών. Στη περίπτωση αυτή μπορούμε να περιορίσουμε το σφάλμα μιας εκτίμησης εάν η επιλογή των δειγματικών μονάδων γίνει με πιθανότητα ανάλογη προς το μέγεθος των ομάδων (πρωτογενών μονάδων) και με επανατοποθέτηση (δηλαδή μια δειγματοληπτική μονάδα μπορεί να επιλεγεί περισσότερο από μια φορά). Με τον τρόπο αυτό δίνεται μεγαλύτερη πιθανότητα επιλογής στις μεγαλύτερες ομάδες (π.χ. οικοδομικά τετράγωνα με μεγαλύτερο αριθμό νοικοκυριών)

Ανωτέρω, παρουσιάσαμε την απλούστερη περίπτωση, στην οποία υποθέτει ότι οι n μονάδες που επιλέγονται με επαναθέσει. Μερικές περιπλοκές ανακύπτουν όταν επιλέγουμε τις μονάδες χωρίς επαναθέσει. Συνήθως, τα μεγέθη των μονάδων δεν είναι γνωστά επακριβώς και πρέπει να τα εκτιμήσουμε εκ των προτέρων. Παράγοντες, όπως το κόστος ή η δομή της μεταβλητικότητας στον πληθυσμό μπορεί να οδηγήσουν σε επιλογή των μονάδων με πιθανότητες που είναι άνισες, αλλά αναλογικές ως προς κάποια ποσότητα άλλη από το μέγεθος. Σε εκτεταμένες έρευνες, πολυσταδιακή δειγματοληψία με άνισες πιθανότητες επιλογής των πρωταρχικών μονάδων δειγματοληψίας είναι η πιο κοινή μέθοδος στην σημερινή πρακτική .

Γενικά η δειγματοληψία με πιθανότητα ανάλογη προς το μέγεθος των μονάδων, εφαρμόζεται στο πρώτο στάδιο της δισταδιακής δειγματοληψίας, επειδή οι δευτερογενείς μονάδες μέσα στις ομάδες είναι συνήθως του ίδιου περίπου μεγέθους (π.χ. αριθμός μελών των νοικοκυριών μέσα στο οικοδομικό τετράγωνο).

Συμβολίζοντας με π_i την πιθανότητα ότι η i δειγματοληπτική μονάδα εμφανίζεται στο δείγμα, έχουμε :

$$\pi_i = \frac{m_i}{M}$$

Επομένως η εκτίμηση του συνολικού πληθυσμού όπως και στην περίπτωση της δειγματοληψίας κατά ομάδες δίνεται από την σχέση :

$$\hat{Y} = \frac{M}{n} \sum_{i=1}^n \hat{y}_i$$

6.8. Συμπέρασμα

Συνοψίζοντας, η δισταδιακή δειγματοληψία παρουσιάζει ορισμένα προβλήματα .

Πρώτο πρόβλημα που παρουσιάζεται είναι το πώς θα διαχωριστεί ο πληθυσμός σε συστάδες. Αν πρέπει δηλαδή να διαχωριστεί σε πολλές συστάδες και να ληφθεί μεγάλος αριθμός από αυτές ή αν πρέπει να διαχωριστεί σε λίγες, οπότε και θα ληφθεί μικρός αριθμός. Η απάντηση εξαρτάται κατ' αρχήν από το πόση ομοιογένεια υπάρχει. Δηλαδή αν πάρουμε λίγες ομάδες και αυτές είναι ομοιογενείς μεταξύ τους το πρόβλημα αντιμετωπίζεται ικανοποιητικά. Αν δεν είναι ομοιογενείς πρέπει να διαχωρίσουμε τον πληθυσμό σε περισσότερες ομάδες. Ένας άλλος σημαντικός παράγοντας που επηρεάζει καθοριστικά το διαμερισμό είναι το κόστος της δειγματοληψίας .

Μέσα σε όλα αυτά που αναφέραμε παραπάνω, προσπαθήσαμε να δείξουμε ότι, η μέθοδος των δύο σταδίων είναι μερικές φορές ο μόνος πρακτικά εφαρμόσιμος τρόπος με την οποία η δειγματοληψία μπορεί να εφαρμοσθεί. Ακόμη όταν υπάρχει μια επιλογή μεταξύ υποδειγματοληψίας των μονάδων δειγματοληψίας και μετρήσεως αυτών τελείως, η δειγματοληψία δύο σταδίων παρέχει στον δειγματολήπτη μεγαλύτερο πεδίο δράσεως, επειδή αυτός μπορεί να καθορίσει και το μέγεθος του δείγματος των πρωταρχικών μονάδων δειγματοληψίας και το μέγεθος του δείγματος που λαμβάνεται από κάθε πρωταρχική μονάδα δειγματοληψίας. Σε μερικές εφαρμογές ένα σημαντικό πλεονέκτημα της δειγματοληψίας δύο σταδίων είναι ότι διευκολύνει το πρόβλημα καταλογοποίησης του πληθυσμού. Συχνά είναι σχετικά εύκολο να εξασφαλίσουμε μια λίστα των πρωταρχικών μονάδων δειγματοληψίας, αλλά δύσκολο ή δαπανηρό να καταλογοποιήσουμε όλες τις υπό-μονάδες. Για να καταλογοποιήσουμε τα δέντρα σε ένα δενδρόκηπο και να λάβουμε ένα δείγμα από αυτά είναι συνήθως είναι απλό, αλλά το πρόβλημα επιτεύξεως μιας τυχαίας επιλογής των φύλλων ενός δέντρου μπορεί να είναι πολύ δύσκολο και βασανιστικό. Με την δειγματοληψία δύο σταδίων αυτό το πρόβλημα αντιμετωπίζεται μόνο για εκείνα τα δέντρα τα οποία επιλέγονται στο δείγμα. Δεν απαιτείται πλήρης κατάλογος όλων των φύλλων στον δενδρόκηπο.

ΚΕΦΑΛΑΙΟ 7.	Άλλη μέθοδοι διενέργειας της δειγματοληψίας

7.1. Τρισταδιακή δειγματοληψία

Σε έρευνες ευρείας κλίμακας είναι απαραίτητη η εφαρμογή της δειγματοληψίας σε τρία ή περισσότερα στάδια, οπότε ονομάζεται τρισταδιακή δειγματοληψία (three-stage sampling) ή πολυσταδιακή δειγματοληψία (multi-stage sampling). Ένα παράδειγμα τρισταδιακής δειγματοληψίας αποτελεί η εξής διαδικασία : Κατά το πρώτο στάδιο επιλέγεται με τυχαίο τρόπο ένας αριθμός Δήμων και Κοινοτήτων (πρωτογενείς μονάδες) από το σύνολο του αριθμού αυτών που βρίσκονται στην υπό έρευνα επιφάνεια (επαρχία, νομός, διαμέρισμα, χώρα). Στη συνέχεια κατά το δεύτερο στάδιο, μέσα από κάθε πρωτογενή μονάδα που έχει επιλεγεί στο δείγμα, επιλέγεται με τυχαίο τρόπο ένας αριθμός οικοδομικών τετραγώνων (δευτερογενείς μονάδες). Κατά το τρίτο στάδιο, μέσα από κάθε δευτερογενή μονάδα που έχει επιλεγεί στο δείγμα, επιλέγεται τυχαία ένας αριθμός νοικοκυριών (τριτογενείς μονάδες). Τα νοικοκυριά αποτελούν την τελική μονάδα του ερευνώμενου πληθυσμού .

Στην πράξη οι μονάδες είναι άνισου μεγέθους μεταξύ τους, στα διάφορα στάδια, για το λόγο αυτό δεν θα ασχοληθούμε με την περιγραφή της διαδικασίας εκτιμήσεων για μονάδες ίσου μεγέθους .

7.2. Πολυφασική δειγματοληψία

Η εφαρμογή δειγματοληπτικών μεθόδων απαιτεί, πολλές φορές, εκ των προτέρων, πληροφορίες, για το διευρενόμενο χαρακτηριστικό. Οι

πληροφορίες αυτές μπορούν να ληφθούν με την εκτέλεση βοηθητικής δειγματοληψίας είτε ως προς άλλο βοηθητικό χαρακτηριστικό, είτε με την επαναληπτική δειγματοληψία σε τακτά χρονικά διαστήματα .

Η εκτέλεση προκαταρκτικών δειγματοληπτικών ερευνών για τη λήψη προαπαιτούμενων πληροφοριών για το διερευνώμενο χαρακτηριστικό προσκρούει σε σοβαρά οικονομικά προβλήματα, τα οποία αναγκάζουν την εκτέλεση δειγματοληψιών σε μικρό, σχετικά, δειγματικό μέγεθος. Για όλους αυτούς τους λόγους κρίνεται συμφερότερη η λήψη πληροφοριών για το διερευνώμενο χαρακτηριστικό από άλλες δειγματοληψίες, οι οποίες έγιναν για κάποιο άλλο χαρακτηριστικό και οι οποίες μπορούν να θεωρηθούν ως μια προ- ή μετά-δειγματοληπτική έρευνα . Στην περίπτωση αυτή μπορεί να γίνει ένας συνδυασμός των δύο δειγματοληψιών , ο οποίος αναφέρεται στη βιβλιογραφία ως διφασική ή γενικότερα πολυφασική δειγματοληψία .

Η διαχρονική διερεύνηση των χαρακτηριστικών μεταβλητών του πληθυσμού, παρά το απαιτούμενο υψηλό κόστος, αποτελεί επιτακτική ανάγκη για όλες τις περιπτώσεις, όπου το φαινόμενο στο οποίο υπεισέρχεται το διερευνώμενο χαρακτηριστικό έχει μια δυναμική χρονική εξέλιξη. Στην περίπτωση αυτή, ιδιαίτερα στην πολυφασική δειγματοληψία, απαιτείται η ύπαρξη και η τήρηση ενός πρωτοκόλλου-πίνακα, για όλα τα στοιχεία του πληθυσμού, ενημερωμένο σε διαχρονική βάση, ώστε να είναι εύκολη η εφαρμογή των αρχών της τυχαίας δειγματοληψίας. Το γεγονός αυτό δημιουργεί τις μεγαλύτερες τεχνικές και οικονομικές δυσκολίες στην εκτέλεση της πολυφασικής δειγματοληψίας .

Η εφαρμογή αυτής της μεθόδου είναι συμφέρουσα, όταν το κέρδος σε ακρίβεια από την εφαρμογή της μεθόδου του λόγου, της παλινδρόμησης ή της στρωματοποίησης υπερβαίνει τη μείωση στην ακρίβεια λόγω της ελάττωσης του μεγέθους δείγματος .

Τονίζουμε ότι η διφασική ή διπλή δειγματοληψία είναι τελείως διαφορετική από τη διασταδιακή δειγματοληψία όπου από τις αρχικά εκλεγόμενες πρωτογενείς μονάδες του πληθυσμού επιλέγεται ένας αριθμός στοιχείων (δευτερογενείς μονάδες) από κάθε μονάδα των πρωταρχικών δειγμάτων, ώστε το τελικό δείγμα να είναι ένα υπό-δείγμα του αρχικού .

7.3. Δειγματοληψία λογική (κρίσεως)

Ο τύπος της μη τυχαίας δειγματοληψίας που θα εξετάσουμε με την ονομασία «Λογική δειγματοληψία» (judgment sampling), ονομάζεται επίσης «δειγματοληψία σκοπιμότητας» (purposive sampling).

Είναι δε η μέθοδος συλλογής των μονάδων ενός δείγματος με βάση την κρίση του δειγματολήπτη, σύμφωνα με την οποία αυτός θεωρεί ότι τα αποτελέσματα θα είναι αντιπροσωπευτικά .

Με άλλα λόγια, οι μονάδες του πληθυσμού που θεωρούνται ότι είναι αντιπροσωπευτικές, σύμφωνα με τις γνώσεις του ερευνητή, επιλέγονται σε ένα λογικό δείγμα. Στην λογική δειγματοληψία, λοιπόν, η πιθανότητα με την οποία μια ατομική μονάδα του πληθυσμού θα επιλεγεί είναι άγνωστη. Κατά συνέπεια, εκτιμήσεις δια διαστημάτων εμπιστοσύνης δεν μπορούν να γίνουν και ως εκ τούτου η λογική δειγματοληψία οδηγεί μόνο σε σημειακές εκτιμήσεις. Περαιτέρω, μόνο ο ειδικός δειγματολήπτης μπορεί να εξηγήσει την ακρίβεια των αποτελεσμάτων που λαμβάνονται από αυτό το σχέδιο αυτό τα αποτελέσματα δεν είναι δυνατόν να εκτιμηθούν αντικειμενικά .

Ακόμη και αν οι αρχές της θεωρίας της δειγματοληψίας δεν εφαρμόζονται στην λογική δειγματοληψία, εν τούτοις, αυτές συχνά χρησιμοποιούνται για την λύση πολλών τύπων οικονομικών και επιχειρηματικών προβλημάτων. Η χρησιμοποίηση της λογικής δειγματοληψίας κρίνεται σκόπιμη κάτω από μια ποικιλία περιστάσεων. Όταν μόνο ένας μικρός αριθμός μονάδων δειγματοληψίας αποτελεί τον πληθυσμό, η απλή τυχαία επιλογή μπορεί να παραλείψει τις περισσότερες σπουδαιές μονάδες, ενώ η λογική επιλογή οπωσδήποτε θα τις συμπεριλάβει στο δείγμα. Μια μελέτη της βιομηχανίας χάλυβος στις Η.Π.Α. θα ήταν αμφιβόλου αξίας εάν η united states steel corporation δεν είχε συμπεριληφθεί στο δείγμα. Παρομοίως, χωρίς την General Motors στο δείγμα, μια έρευνα για την βιομηχανία αυτοκινήτων θα ήταν εξίσου μη ικανοποιητική .

Η λογική δειγματοληψία χρησιμοποιείται μερικές φορές διότι πρέπει να κρατήσουμε το ύψος του δείγματος μικρό : Αυτό διότι μπορεί να είναι, δύσκολο να εντοπίσουμε κάποιες από τις μονάδες δειγματοληψίας ή απαραίτητο να κρατήσουμε τα κόστη δειγματοληψίας σε χαμηλά επίπεδα. Σε τέτοιες περιπτώσεις, η δειγματοληψία πιθανότητας είναι πρακτικά μη εφαρμόσιμη διότι μεγάλου πλάτους διαστήματα εμπιστοσύνης αντιστοιχούν σε μικρά δείγματα. Συνεπώς, οι δειγματικές εκτιμήσεις, καίτοι αμερόληπτες, είναι μικρής αξίας στην λήψη αποφάσεων .

Επίσης για την λήψη καθημερινών επιχειρηματικών προβλημάτων και για την λήψη αποφάσεων κυβερνητικής πολιτικής, οι διευθυντές των επιχειρήσεων και οι κυβερνητικοί επίσημοι που συχνά πιέζονται από τον χρόνο δεν μπορούν να περιμένουν για την εφαρμογή πιθανοθεωρητικών σχεδίων δειγματοληψίας. Η λογική δειγματοληψία είναι στις περιπτώσεις αυτές, η μόνη πρακτικά εφαρμόσιμη μέθοδος, επειδή οι εκτιμήσεις

μπορούν να είναι διαθέσιμες σε σύντομο χρονικό διάστημα και έτσι θα βοηθήσουν τους διευθυντές των επιχειρήσεων και τους κυβερνητικούς επίσημους να λάβουν αποφάσεις για επείγοντα προβλήματα, οι οποίες θα είναι οπωσδήποτε καλύτερες από αποφάσεις που θα ελαμβάνοντο χωρίς κάποια στατιστικά στοιχεία .

Και πάλι, όταν επιθυμούμε να μελετήσουμε κάποιες άγνωστες ιδιότητες ενός πληθυσμού, του οποίου είναι γνωστές κάποιες χαρακτηριστικές ιδιότητες, μπορούμε να στρωματοποιήσουμε τον πληθυσμό συμφωνά με αυτές τις γνωστές χαρακτηριστικές ιδιότητες και στην συνέχεια να επιλέξουμε τις μονάδες δειγματοληψίας από κάθε στρώμα με βάση την λογική. Αυτή η μέθοδος χρησιμοποιείται για να πάρουμε ένα περισσότερο αντιπροσωπευτικό δείγμα. Αυτό δικαιολογείται από την εξής υπόθεση : «Επειδή ένα δείγμα είναι περισσότερο αντιπροσωπευτικό σε σχέση με τις γνώστες χαρακτηριστικές ιδιότητες του πληθυσμού, μπορεί επίσης αυτό το δείγμα να είναι περισσότερο αντιπροσωπευτικό σε σχέση με τις άγνωστες χαρακτηριστικές ιδιότητες του πληθυσμού.»

Τέλος, η λογική δειγματοληψία μπορεί να χρησιμοποιηθεί για να κατευθύνουμε μελέτες δια δείγματος «πιλότος». Πρέπει να αναφέρουμε επίσης ότι για να έχουμε αποτελεσματική στρωματοποίηση, η γνώση πληθυσμιακών χαρακτηριστικών ιδιοτήτων είναι απαραίτητη. Σε αυτή την περίπτωση μελέτες «πιλότος», είναι ένας τρόπος απόκτησεως των επιθυμητών πληροφοριών. Η μελέτη «πιλότος», είναι απλώς διερευνητική, και γίνεται πριν άλλα σχέδια δειγματοληψίας υιοθετηθούν .

Σε κάθε περίπτωση, η αξιοπιστία των αποτελεσμάτων του δείγματος στην λογική δειγματοληψία εξαρτάται από την ποιότητα των ειδικών γνώσεων του δειγματολήπτη ή την κρίση (λογική) αυτού . Εάν η λογική δειγματοληψία εφαρμοσθεί προσεκτικά και επιδέξια, τα λογικά δείγματα μπορεί να αναμένεται να είναι αντιπροσωπευτικά και να παράγουν πολύτιμα αποτελέσματα . Εξ' άλλου, όταν ένα δείγμα ληφθεί με φτωχή κρίση (λογική), σοβαρές μεροληψίες θα παρουσιασθούν .

7. 4 Δειγματοληψία ποσοστών Quota

Μια μέθοδος δειγματοληψίας που μοιάζει με τη στρωματοποιημένη δειγματοληψία είναι η μέθοδος quota που συνήθως εφαρμόζεται στην έρευνα αγοράς και σε δημοσκοπήσεις γνώμης όπως πολιτικές θέσεις, αντιδράσεις σε κάποιο προτεινόμενο νομοσχέδιο,

κυβερνητική πολιτική για κάποιο θέμα, προτιμήσεις καταναλωτών για κάποιο προϊόν, ακροαματικότητα κάποιας εκπομπής στην Τ.Υ κ.λ.π.

Με τη μέθοδο αυτή κάθε άτομο παίρνει συνεντεύξεις από ένα ορισμένο αριθμό ατόμων. Λαμβάνεται όμως πρόνοια έτσι ώστε διάφορα χαρακτηριστικά του πληθυσμού που είναι ήδη γνωστά από κάποια προηγούμενη απογραφή να περιλαμβάνονται στο εκλεγόμενο δείγμα με την ίδια αναλογία. Τα χαρακτηριστικά κυρίως που λαμβάνονται υπόψη στην επιλογή του δείγματος είναι το φύλο, ηλικία, μόρφωση και οικονομική τάξη. Έτσι επιτυγχάνεται δείγμα της ίδιας σύνθεσης με αυτή που παρουσιάζει ο υπό μελέτη πληθυσμός .

Ο συλλογισμός βάσει του οποίου δημιουργείται το κατευθυνόμενο δείγμα quota είναι :

Εάν είναι γνωστό εκ των προτέρων ότι η υπό έρευνα πληθυσμός έχει ορισμένα χαρακτηριστικά με γνωστές αναλογίες, τότε, το δείγμα μπορεί να επιλεγεί κατά τέτοιο τρόπο ώστε, να περιλαμβάνει τα υπόψη χαρακτηριστικά στην ίδια αναλογία στην οποία και ο υπό έρευνα πληθυσμός. Βάσει αυτού οι συνεντεύκται παίρνουν οδηγίες να κατανεύμουν τις συνεντεύξεις μεταξύ των ανταποκριτών μέσα στο δείγμα quota, σύμφωνα με κάποια καθορισμένα χαρακτηριστικά, έτσι όπως "πόσοι σε κάθε μια από τις εισοδηματικές ομάδες ", "τόσοι σε κάθε φυλή" κ.λ.π., έτσι ώστε να επιτύχουν δείγμα της ίδιας σύνθεσης με αυτή που παρουσιάζει ο υπό έρευνα πληθυσμός .

Ένας σπουδαίος λόγος για την εφαρμογή της δειγματοληψίας quota είναι να εξασφαλίσουμε μερικά από τα πλεονεκτήματα της στρωματοποίησης σε χαμηλό κόστος. Πάντως, κατά την επιλογή των μονάδων δειγματοληψίας η κρίση του συνεντεύκτη αποτελεί τον καθοδηγητικό παράγοντα. Οι συνεντεύκται συχνά επιδιώκουν να παίρνουν συνεντεύξεις από καλύτερα μορφωμένα ή κομψοντυμένα πρόσωπα, ή να ερωτούν οποιονδήποτε που βρίσκεται στο σπίτι κατά την ώρα της επισκέψεως, ή να παραμελούν τις μονάδες δειγματοληψίας οι οποίες δεν είναι εύκολα προσιτές. Συνεπώς, ένα δείγμα quota πράγματι εξαρτάται από την κρίση του συνεντεύκτη. Επειδή τα αποτελέσματα πάντοτε περιέχουν άγνωστες μεροληψίες, το δείγμα δεν μπορεί να θεωρηθεί σαν μια τυχαία μεταβλητή .

7. 5 : Δειγματοληψία με άνεση (γάριν ευκολίας)

Η δειγματοληψία με άνεση (convenience sampling), συνίσταται στο εξής :

Το δείγμα είναι ένα κλάσμα του πληθυσμού που λαμβάνεται για έρευνα εξ' αιτίας της άνετης διαθέσεως αυτού. Συνεπώς, το δείγμα

(κομμάτι του πληθυσμού) επιλέγεται, ούτε με πιθανότητα ούτε με λογική, αλλά με ευκολία. Ένα δείγμα που λαμβάνεται από διαθέσιμες λίστες, τέτοιες όπως ο τηλεφωνικός κατάλογος ή ο κατάλογος ιδιοκτητών αυτοκινήτων είναι ένα δείγμα ευκολίας και όχι ένα τυχαίο δείγμα, ακόμη και αν το δείγμα λαμβάνεται τυχαίο από τις λίστες. Η μη παραδοχή αυτής της διαφοράς συχνά οδηγεί τους δειγματολήπτες σε λάθος δρόμος .

Ένα δείγμα – που είναι απλώς μια άνετα διαθέσιμη φέτα του πληθυσμού – μπορεί μόλις και μετά βίας να είναι αντιπροσωπευτικό του πληθυσμού. Τα αποτελέσματα του είναι συνήθως με μεγάλο βαθμό μεροληπτικά και μη ικανοποιητικά. Άλλοτε, η μέθοδος αυτή είναι συχνά χρησιμοποιούμενη σε έρευνες κοινής γνώμης, όταν οι απογραφείς σταματούσαν κοντά στην στάση του μετρό ή του λεωφορείου για να πάρουν συνέντευξη από διερχόμενα άτομα.

Τέλος, θα μπορούσαμε να πούμε ότι η δειγματοληψία με άνεση μπορεί να χρησιμοποιηθεί σε μελέτες «πιλότος». Ερωτήσεις μπορούν να ελεγχθούν και πρωταρχικές πληροφορίες μπορούν να ληφθούν με ένα τέτοιο δείγμα προτού αποφασισθεί το τελικό σχέδιο δειγματοληψίας .

7. 6 Δειγματοληψία με σταθερά δείγματα

Η δειγματοληψία με σταθερά δείγματα είναι μια μέθοδος που εφαρμόζεται στις περιπτώσεις που συγκεντρώνονται πληροφορίες από ένα πληθυσμό σε τακτά χρονικά διαστήματα (κάθε μήνα, κάθε χρόνο κ.λ.π.). Δηλαδή παίρνουμε από τον ερευνώμενο πληθυσμό ένα δείγμα το οποίο χρησιμοποιούμε κάθε φορά που θέλουμε να μελετήσουμε την κατάσταση του υπό έρευνα φαινομένου, σε επιθυμητά χρονικά διαστήματα. Η επιλογή του δείγματος γίνεται με τυχαίο τρόπο με μια από τις μεθόδους που έχουν περιγραφεί στα προηγούμενα κεφάλαια (απλή τυχαία δειγματοληψία, στρωματοποιημένη κ.λ.π.)

Η δειγματοληψία με σταθερά δείγματα εφαρμόζεται σε έρευνες εισοδήματος νοικοκυριών όταν θέλουμε να μελετήσουμε την εξέλιξη της οικοδομικής κατάστασης αυτών, σε έρευνες αγοράς, όταν θέλουμε να μελετήσουμε τη διαχρονική εξέλιξη της οικονομικής κατάσταση των επιχειρήσεων. Επίσης χρησιμοποιείται σε μηνιαίες έρευνες από τη Γενική Γραμματεία Ε.Σ.Υ.Ε., π.χ. στην μηνιαία έρευνα παραγωγής βιομηχανικών προϊόντων, για την κατάρτιση του μηνιαίου δείκτη βιομηχανικής παραγωγής κ.λ.π.

Τα πλεονεκτήματα της δειγματοληψίας με σταθερά δείγματα είναι τα εξής :

1. Δίνεται η δυνατότητα να μελετηθεί η τάση συμπεριφοράς ατόμων, νοικοκυριών, επιχειρήσεων κ.λ.π. σε επιθυμητά χρονικά διαστήματα .
2. Απαιτείται λιγότερος χρόνος για τον σχεδιασμό μιας έρευνας και μικρότερο κόστος, αφού χρησιμοποιείται το ίδιο δείγμα .
3. Επιτυγχάνεται περιορισμός των μη δειγματοληπτικών σφαλμάτων, λόγω απόκτησης εμπειρίας των ερευνητών, αφού επισκέπτονται τις ίδιες μονάδες δύο και περισσότερες φορές .

Ένα βασικό μειονέκτημα της δειγματοληψίας με σταθερά δείγματα είναι η μείωση του ποσοστού ανταπόκρισης, κάθε φορά που επαναλαμβάνεται η έρευνα. Αυτό οφείλεται στο γεγονός ότι τα άτομα ή τα νοικοκυριά που ερευνώνται δεν είναι πρόθυμα να συνεργαστούν με τον ερευνητή, όταν τα επισκέπτεται περισσότερο από δύο φορές, διότι αρχίζουν να κουράζονται. Το πρόβλημα αυτό μπορεί ορισμένες φορές να περιορισθεί, με αντικατάσταση κάποιων μονάδων που είναι περίπου όμοιες με αυτές που δεν ανταποκρίνονται.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Α. ΕΛΛΗΝΟΓΛΩΣΣΗ

1. Τζιαφέτας . Γ. (1991) – «Θεωρία και Τεχνική της δειγματοληψίας» . ΕΛΚΕΠΑ ,Αθήνα .
2. Δαμιανού .Χ. (1992) – « Μεθοδολογία δειγματοληψίας : τεχνικές και εφαρμογές .» , 2^η έκδοση , Αθήνα .
3. Μπενός . Β.Κ. (1984) – « Ανάλυση και τεχνική της δειγματοληψίας .» - Καραμπερόπουλος , Πειραιάς .
4. Φαρμάκη . Ν. (1992) – «Εισαγωγή στην δειγματοληψία .» - Κ. Χριστοδουλίδη , Θεσσαλονίκη .
5. Τζωρτζόπουλος . Π. (1982) – « Δειγματοληψία.» - Αθήνα .
6. Πανταζίδη . Ν. (1976) – «Εισαγωγή στην δειγματοληψία .»
7. Ν.Α. Τσερπες / Φ. Αλεβιζος (1996) – « Εισαγωγή στην θεωρία δειγματοληψίας.» - Πανεπιστήμιο Πατρών .

Β. ΞΕΝΟΓΛΩΣΣΑ

1. Cochran. W.G. (1977) –“ Sampling techniques.” – John Wiley. New York.
2. Ranjan .K. Som – “ Practical sampling techniques.” – second edition revised and expanded.

