

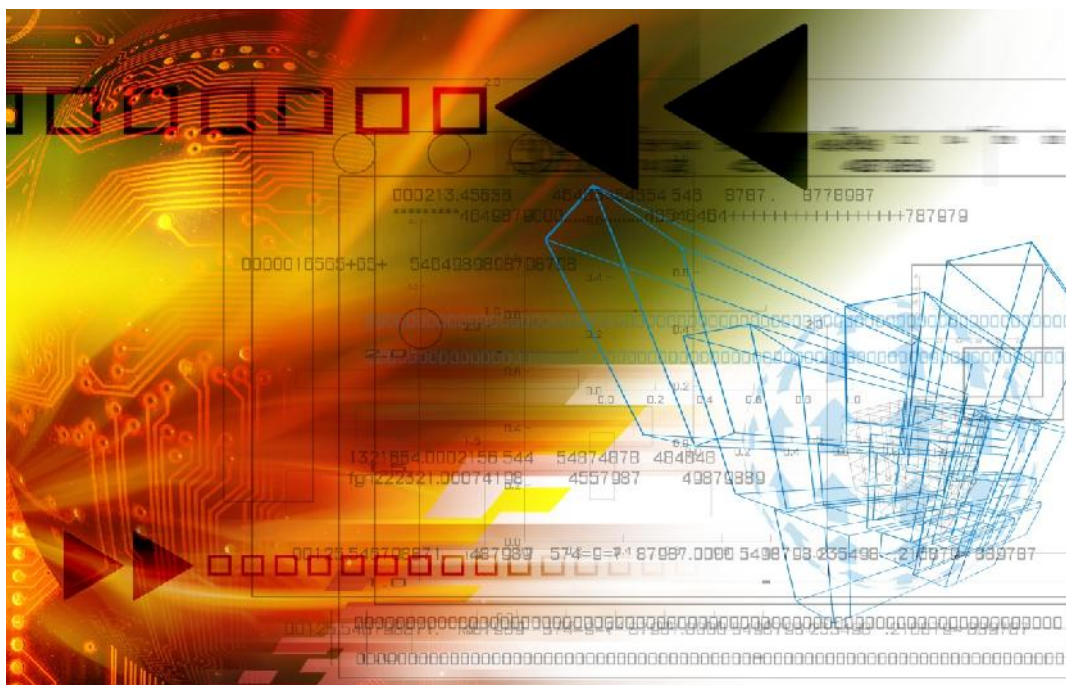
ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΠΑΤΡΩΝ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ

ΤΕΧΝΙΚΕΣ ΔΕΙΓΜΑΤΟΛΗΠΤΙΚΩΝ ΕΡΕΥΝΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ



ΟΝΟΜΑΤΕΠΩΝΥΜΟ ΣΠΟΥΔΑΣΤΩΝ

ΔΗΜΟΠΟΥΛΟΥ ΜΑΡΙΑ

ΚΟΥΡΗ ΜΑΡΙΑ

ΕΠΟΠΤΕΥΩΝ ΚΑΘΗΓΗΤΗΣ

ΚΑΡΥΩΤΗ ΒΑΣΙΛΙΚΗ

ΠΕΡΙΛΗΨΗ

Η στατιστική είναι επιστήμη που επιχειρεί να εξαγάγει γνώσεις με χρήση εμπειρικών δεδομένων. Βασίζεται στη χρήση της στατιστικής θεωρίας, ενός κλάδου των εφαρμοσμένων μαθηματικών. Στη στατιστική, η τυχαιότητα και η απροσδιοριστία ορίζονται στα πλαίσια της θεωρίας πιθανοτήτων. Η πρακτική της στατιστικής περιλαμβάνει τη σχεδίαση, συλλογή και ερμηνεία δεδομένων που προκύπτουν από αβέβαιες παρατηρήσεις. Επειδή η στατιστική αποσκοπεί στην εξαγωγή των «καλύτερων» πληροφοριών από τα διαθέσιμα δεδομένα, κατατάσσεται από μερικούς σαν κλάδος της θεωρίας των αποφάσεων.

Η κυβέρνηση, οι επιχειρηματίες και οι επαγγελματίες σε μία ελεύθερη και προοδευτική κοινωνία, αναζητούν την ευρύτερη δυνατή πραγματική βάση για τη λήψη αποφάσεων, την τυποποίηση της πολιτικής, και την ανάπτυξη της κοινωνικής και οικονομικής θεωρίας. Η από κοινού εμφάνιση αυτών των διαφορετικών ενδιαφερόντων στην αναζήτηση οικονομικών και κοινωνικών στόχων έχει συμβάλει σημαντικά στην μεγάλη σπουδαιότητα της στατιστικής σήμερα. Εκατοντάδες εκατομμύρια δολάρια έχουν δαπανηθεί για την στατιστική τόσο από το Δημόσιο όσο και από τον ιδιωτικό τομέα κατά τη διάρκεια των τελευταίων δεκαετιών. Η ανάπτυξη της δειγματοληψίας κατά την ίδια περίοδο κατέστησε την αξία της αδιαμφισβήτητη.

Ο σχηματισμός της μεθοδολογίας λήψεως αποφάσεων, εξαρτάται από την γνώση μας για τις συνέπειες που μπορεί να προκύψουν από διαφορετικές πράξεις επί μιας δεδομένης καταστάσεως και από την επικρατούσα κατάσταση της φύσεως κατά την στιγμή λήψεως μιας αποφάσεως. Συχνά όμως διαπιστώνουμε, ότι οι ακριβείς ιδιότητες του πληθυσμιακού μοντέλου ή της κατάστασης της φύσεως, δεν είναι γνωστές. Για τη λήψη ορθών αποφάσεων, πρέπει να γνωρίζουμε το μοντέλο τουλάχιστον κατά προσέγγιση. Ένας τρόπος προσεγγίσεως των ιδιοτήτων ενός πληθυσμιακού μοντέλου είναι η από ευθείας δειγματοληψία του πληθυσμού.

Εάν πάρουμε ένα δείγμα, με μία τυχαία διαδικασία, τότε οι δειγματικές συναρτήσεις είναι τυχαίες μεταβλητές οι οποίες μπορούν να χρησιμοποιηθούν για την εκτίμηση των αντίστοιχων πληθυσμιακών παραμέτρων. Οι στατιστικές μέθοδοι που μας επιτρέπουν να συμπεράνουμε από περιορισμένα δεδομένα (δείγματα) τις αναμενόμενες μακροπρόθεσμες συμπεριφορές (πληθυσμοί) αποτελούν την “Στατιστική Συμπερασματολογία” ή “Επαγωγική Στατιστική”. Στην διαδικασία του σχηματισμού των επαγωγικών συμπερασμάτων, μπορεί να κάνουμε σφάλματα εξαιτίας της φύσεως των χαρακτηριστικών του πληθυσμού που εξετάζονται. Πρέπει να μπορέσουμε να εκτιμήσουμε αυτά τα σφάλματα, τα οποία είναι το αποτέλεσμα τυχαίων ή πιθανών διακυμάνσεων και μπορούν να εκτιμηθούν μόνο σε όρους πιθανοτήτων. Μια δειγματική συνάρτηση, σαν τυχαία μεταβλητή, πρέπει να έχει την δική της κατανομή πιθανότητας. Η κατανομή της δειγματικής συνάρτησης, έχει συγκεκριμένες ιδιότητες, όπως κάθε μοντέλο πιθανότητας. Από τις ιδιότητες της κατανομής αυτής μπορούμε να υπολογίσουμε τα σφάλματα που οφείλονται στην τύχη και που ανέρχονται στις γενικεύσεις που κάνουμε για τους πληθυσμούς με βάση τα δείγματα.

Από τα παραπάνω προκύπτει το ερώτημα: “Πως είναι δυνατόν, ένα τυχαίο δείγμα, που πήραμε από ένα άγνωστο πληθυσμό, είναι δυνατόν να παρέχει πληροφορίες για τον πληθυσμό, από τον οποίο προήλθε;”

Απάντηση στο ερώτημα αυτό θα δώσει η θεωρία της δειγματοληψίας (η οποία αποτελεί κατά κάποιον τρόπο τον σύνδεσμο μεταξύ της πραγματικότητας και των πορισμάτων της θεωρίας πιθανοτήτων), που επιτρέπει να υπολογίσουμε χαρακτηριστικά μέτρα του δείγματος και στη συνέχεια να τα χρησιμοποιήσουμε σαν ικανοποιητικές προσεγγίσεις (εκτιμήσεις) των αγνώστων παραμέτρων του δειγματοληπτικού πληθυσμού.

ABSTRACT

Statistics is a science that attempts to export knowledge by using practical data. It is based on the use of statistics theory, a sector of applicative mathematics. In statistics, the accidental and the infinite are defined according to the possibility theory. The practice of statistics includes designing, collecting and interpreting data that arise by doubtful comments. Due to the fact that statistics aims at the export of “better” information from the available data, it is rated by some as a sector of the decision theory.

The government, the businessmen and the professionals in a free and progressive society, seek the widest possible base for decision making, the standardizing of politics and the development of social and economical theory. Appearance of these different interests in seeking economic and social targets has contributed essentially to the great importance of statistics nowadays. Hundreds of million dollars have been spent for statistics not only by the public but also by the private sector during the last decades. The development of sampling during the same period has proved its value undoubtable. Forming a methodological decision making, depends on our knowledge for the consequences that can occur by different actions on a given situation and the existing situation during the moment of reaching a decision. We usually realize that the exact qualities of the population model or the existing situation are not known. So, in order to make sensible decisions, we need to know the model at least roughly. A way of approaching the qualities of a population model, is the direct sampling of the population.

If we get a sample with a random procedure, then the sampling functions are random variables, which can be used for the evaluation of the corresponding population parameters. The Statistics methods which allow us reach a result of limited data (samples) the expected long-term behaviors (population) consist of the “Statistic Deductions” or “Statistics Inductions”. During the process of forming the deductions, we may get some errors that are related to the kind of the examined properties of the population. We have to be able to evaluate these errors, which are the result of random or possible variations and can only be evaluated in possibility terms. One sample function, as a random variable, must have its own possibility distribution. The distribution of the sample function has specific qualities, just like every possibility model. By the qualities of this distribution, we can estimate the errors owing to luck and that arise to the generalizations we made for the populations based on samples.

From the above a question arises: “How is possible, a random sample, which we took from an unknown population to give us information about the population it came from”. A reply to this question will be given by the theory of sampling (which is in a way the connection between reality and the conclusions of the possibility theory), which allows us estimate

typical measures of the sample and later on use them as satisfying approaches of the unknown parameters of the sampling population.

Περιεχόμενα

ΕΙΣΑΓΩΓΗ	10
Κεφάλαιο 1: Βασικές Έννοιες.....	11
1.1 Εισαγωγή	11
1.2 Πληθυσμός	11
1.3 Στατιστική Μονάδα και Μεταβλητές	12
1.4 Δείγμα.....	15
1.4.1 Τυχαίο Δείγμα.....	17
Κεφάλαιο 2: Δειγματοληψία	18
2.1 Εισαγωγή	18
2.2 Ορισμός δειγματοληψίας.....	19
2.3 Θεωρητική βάση της δειγματοληψίας.....	19
2.4 Πλεονεκτήματα της δειγματοληψίας.....	21
2.5 Μειονεκτήματα της δειγματοληψίας.....	22
2.6 Ο υπό Δειγματοληψία Πληθυσμός	22
2.7 Δειγματοληπτική Μονάδα	23
2.7.1 Δειγματικό πλαίσιο ή δειγματοληπτικός σκελετός.....	23
2.7.2 Κατάλογος ή λίστα.....	24
2.8 Μέθοδοι συγκέντρωσης στατιστικών δεδομένων	25
2.9 Απογραφή	26
2.9.1 Πλεονεκτήματα της απογραφής.....	27
2.9.2 Μειονεκτήματα της απογραφής.....	27
2.10 Απογραφές του πληθυσμού στην Ελλάδα	28
2.11 Μερική απογραφή	29
2.12 Επιλογή της απογραφικής ή της δειγματοληπτικής έρευνας.....	30
2.13 Προσδιορισμός του μεγέθους του δείγματος	30

Κεφάλαιο: 3 Δειγματοληπτική έρευνα.....	33
3.1 Εισαγωγή	33
3.2 Ορισμός και ανάλυση	33
3.3 Διαδικασία Δειγματοληπτικής Έρευνας.....	34
3.3.1 Κατάρτιση πλαισίου δειγματοληψίας	36
3.3.2 Επεξεργασία Δειγματοληπτικών στατιστικών στοιχείων	37
3.3.3 Παρουσίαση Δειγματοληπτικών Στοιχείων σε Μορφή Πινάκων	37
3.3.3.1 Τεχνική κατασκευής στατιστικών πινάκων	38
3.3.3.2 Τύποι στατιστικών πινάκων	39
3.3.4 Το ερωτηματολόγιο	41
3.3.4.1 Κατάρτιση ερωτήσεων του ερωτηματολογίου.....	42
3.3.4.2 Δημιουργία ερωτηματολογίου	42
3.3.4.3 Τύποι ερωτήσεων	44
3.3.4.4 Σειρά των ερωτήσεων	48
3.3.4.5 Διατύπωση των ερωτήσεων	49
3.3.4.6 Υλική και τυπογραφική παρουσίαση	49
3.3.4.7 Το μέγεθος του ερωτηματολογίου	50
3.3.4.8 Δοκιμή του ερωτηματολογίου.....	50
3.3.5 Πρόσληψη προσωπικού και εκπαίδευση	51
3.3.5.1 Εκπαίδευση ερευνητών	51
3.3.5.2 Αρχική εκπαίδευση	51
3.3.6 Τελική δοκιμαστική έρευνα.....	52
3.3.7 Μέθοδοι Επιλογής στατιστικών στοιχείων	52
3.3.7.1 Η μέθοδος της παρατήρησης.....	52
3.3.7.2 Η μέθοδος της συνέντευξης	53
3.3.7.3 Η μέθοδος του Ταχυδρομείου	53
Κεφάλαιο: 4 Στατιστικά Δειγματοληπτικά σφάλματα.....	54

4.1 Εισαγωγή	54
4.2 Ορισμός Στατιστικών Σφαλμάτων	55
4.3 Κατηγορίες σφαλμάτων.....	55
4.3.1 Δειγματοληπτικά σφάλματα	55
4.3.2 Τα μη δειγματοληπτικά σφάλματα	56
4.3.3 Τα τυχαία σφάλματα είναι αναπόφευκτα και περιγράφονται με τη στατιστική θεωρία	57
4.4 Ακρίβεια και αξιοπιστία στα σφάλματα	58
4.5 Είδη σφαλμάτων	59
4.5.1 Συστηματικά σφάλματα	59
4.5.2 Τυχαία σφάλματα.....	59
4.6 Μέση τιμή και σφάλμα	59
Κεφάλαιο 5: Απλή τυχαία δειγματοληψία	62
5.1 Εισαγωγή	62
5.2 Ορισμός και ανάλυση απλής τυχαίας δειγματοληψίας.....	62
5.3 Πλεονεκτήματα – Μειονεκτήματα απλής τυχαίας δειγματοληψίας.....	63
5.4 Μέθοδος της κλήρωσης.....	64
5.5 Μέθοδος τυχαίων αριθμών	64
5.5.1 Κατασκευή πίνακα τυχαίων αριθμών	65
5.6 Δειγματοληψία χωρίς επανατοποθέτηση.....	68
5.7 Δειγματοληψία με επανατοποθέτηση	69
5.8 Σύγκριση των μεθόδων.....	69
Κεφάλαιο 6: Συστηματική δειγματοληψία.....	70
6.1 Εισαγωγή	70
6.2 Ορισμός της συστηματικής δειγματοληψίας.....	71
6.2.1 Ορισμός Συστηματικού δείγματος.....	71
6.2.2 Περιγραφή της Διαδικασίας.....	71

6.2.3 Εφαρμογή της συστηματικής δειγματοληψίας	73
6.3 Πλεονεκτήματα της συστηματικής δειγματοληψίας	73
6.4 Μειονεκτήματα της συστηματικής δειγματοληψίας	73
6.5 Το μέγεθος του δείγματος.....	74
6.6 Η Διασπορά της Εκτιμήτριας του Μέσου	75
Κεφάλαιο 7: Δειγματοληψία κατά στρώματα ή στρωματοποιημένη δειγματοληψία.....	78
7.1 Εισαγωγή	78
7.2 Ορισμοί και συμβολισμοί	79
7.3 Χρήσεις της τυχαίας στρωματοποιημένης δειγματοληψίας	81
7.4 Πλεονεκτήματα της στρωματοποιημένης δειγματοληψίας	82
7.5 Μειονεκτήματα της στρωματοποιημένης δειγματοληψίας	82
7.6 Εκτίμηση Μέσης Τιμής Στρωματοποιημένου Πληθυσμού	83
7.7 Το μέγεθος του δείγματος.....	87
7.7.1 Στρωματοποίηση με αναλογικό καταμερισμό	88
7.7.2 Δειγματοληψία κατά στρώματα δυσανάλογη	90
7.8 Βέλτιστος καταμερισμός	90
7.9 Σύγκριση Απλής Τυχαίας Και Στρωματοποιημένης Δειγματοληψίας	93
Κεφάλαιο 8: Δειγματοληψία Κατά Ομάδες – Πολυσταδιακή Δειγματοληψία.....	94
8.1 Εισαγωγή	94
8.2 Ορισμός Δειγματοληψίας κατά ομάδες.....	95
8.2.1 Δειγματοληψία κατά ομάδες σε ένα στάδιο	95
8.2.2 Δειγματοληψία κατά ομάδες σε πολλά στάδια	95
8.3 Χρήσεις της δειγματοληψίας κατά ομάδες.....	96
8.4 Πλεονεκτήματα της δειγματοληψίας κατά ομάδες	97
8.5 Μειονεκτήματα της δειγματοληψίας κατά ομάδες.....	97
8.6 Εκτίμηση της Μέσης Τιμής.....	97
8.7 Περίπτωση Ποσοστών.....	102

1) Περίπτωση ισομεγεθών ομάδων.....	103
2) Περίπτωση ανισομεγεθών ομάδων.....	103
8.8 Σύγκριση της Δειγματοληψίας Κατά Ομάδες με την Στρωματοποιημένη Δειγματοληψία	105
Κεφάλαιο 9: Έρευνα	107
9.1 Εισαγωγή.....	107
9.2 Ταυτότητα της έρευνας.....	107
9.3 Το ερωτηματολόγιο	109
9.4 Ανάλυση αποτελεσμάτων της έρευνας.....	111
9.4.1 Πρώτο μέρος της έρευνας.....	111
9.4.2 Δεύτερο μέρος της έρευνας.....	130
Κεφάλαιο 10: Συμπεράσματα.....	139
Βιβλιογραφία.....	141
Παραρτήματα	142
Παράρτημα Α: Πίνακας Τυχαίων Αριθμών 2500 ψηφίων.....	142
Παράρτημα Β: Ερωτηματολόγιο Α	143
Παράρτημα Γ : Ερωτηματολόγιο Β.....	146

ΕΙΣΑΓΩΓΗ

Στην εργασία αυτή θα αναφερθούν τεχνικές δειγματοληψίας οι οποίες είναι απαραίτητες για τη συγκρότηση οποιουδήποτε είδους έρευνας. Από αυτή τη πλευρά και μέσω αυτής της εργασίας θα αναπτυχθούν μέθοδοι και τεχνικές δειγματοληψίας που θα χρησιμοποιούνται στη δομή μιας σωστής έρευνας.

Προκειμένου να μελετήσουμε τις τεχνικές δειγματοληψίας, είναι απαραίτητο να γίνει πρώτα μια αναφορά σε βασικές έννοιες που θα χρησιμοποιηθούν, οι οποίες θα δείξουν πώς είναι δυνατόν, ένα σχετικά μικρό δείγμα, να εξασφαλίσει αποτελέσματα τα οποία προσεγγίζουν πολύ ικανοποιητικά τα χαρακτηριστικά ενός μεγάλου πληθυσμού και θα γίνει μια εκτενής αναφορά στο τυχαίο δείγμα. Από την εφαρμογή των μεθόδων και των τεχνικών δειγματοληψίας προκύπτουν στατιστικά δειγματοληπτικά στοιχεία και σφάλματα τα οποία θα επεξεργαστούν και θα αναλυθούν έτσι ώστε να προκύψουν σωστές εκτιμήσεις.

Επίσης, θα παρουσιαστεί μια αντιπαράθεση της απλής τυχαίας δειγματοληψίας με επανάθεση και χωρίς επανάθεση καθώς και μια αναλυτική παρουσίαση άλλων μεθόδων και τεχνικών δειγματοληψίας όπως η στρωματοποιημένη δειγματοληψία, η δειγματοληψία κατά ομάδες και η συστηματική δειγματοληψία.

Θα δούμε επίσης τα βήματα που ακολουθούνται για την επιλογή και κατάρτιση ενός ερωτηματολόγιου, τις μεθόδους που χρησιμοποιήσαμε και τα αποτελέσματα της έρευνας. Το θέμα της έρευνας που πραγματοποιήθηκε ήταν οι διατροφικές συνήθειες των παιδιών αλλά και τα ενδιαφέροντα που έχουν οι μαθητές και οι δραστηριότητες τους εκτός σχολείου.

Τα αποτελέσματα της έρευνας θα εξαρτηθούν από:

Την κοινωνική τάξη των παιδιών

Την οικονομική τους κατάσταση

Τέλος, θα γίνει μια επισκόπηση των αποτελεσμάτων και μια παρουσίαση των συμπερασμάτων με μορφή πίνακα.

Κεφάλαιο 1: Βασικές Έννοιες

1.1 Εισαγωγή

Σ' αυτό το κεφάλαιο θα παρουσιαστεί μια συνοπτική θεώρηση των βασικών εννοιών της δειγματοληψίας και των εναλλακτικών σχεδίων δειγματοληψίας. Θα παρουσιαστεί πως είναι δυνατόν, με ένα σχετικά μικρό δείγμα, να εξασφαλίσουμε αποτελέσματα τα οποία προσεγγίζουν ικανοποιητικά τα υπό έρευνα χαρακτηριστικά ενός μεγάλου πληθυσμού, και θα δοθούν σε απλοποιημένες παρουσιάσεις μερικές από τις βασικές αρχές σχεδιασμού ενός δείγματος. Επίσης, θεωρείται ιδιαίτερα σημαντικό να αναλυθεί όσο γίνεται περισσότερο η έννοια του μέτρου της ακρίβειας και η σχέση αυτή με το βαθμό εμπιστοσύνης που δύναται να αποδοθεί στα αποτελέσματα μιας έρευνας δειγματοληψίας.

Πρέπει να επισημανθεί, ότι η ακρίβεια των αποτελεσμάτων που προέρχεται από μια έρευνα δειγματοληψίας, εξαρτάται όχι μόνο από το μέγεθος του δείγματος, αλλά επίσης και από άλλους παράγοντες, που αφορούν τον σχεδιασμό του δείγματος. Για να επιτευχθεί ένας αποτελεσματικός σχεδιασμός του δείγματος θα πρέπει να χρησιμοποιηθούν αποτελεσματικά όλες οι διαθέσιμες πληροφορίες και πηγές.

1.2 Πληθυσμός

Πληθυσμός (population) είναι ένα σύνολο από στατιστικές μονάδες πάνω στις οποίες γίνονται οι μετρήσεις των μεταβλητών, όπως είναι οι άνθρωποι, ζώα ή αντικείμενα του ίδιου είδους και από το οποία θα επιλεγεί ένα υποσύνολο για τη μελέτη των χαρακτηριστικών του που μας ενδιαφέρουν. (W. Cochran, 1997)

Ο **πληθυσμός** μπορεί να περιλαμβάνει πεπερασμένο ή άπειρο πλήθος στοιχείων. Το πλήθος αυτό θα συμβολίζεται με N . Με τον όρο **πληθυσμό** ορίζουμε το σύνολο ατόμων ή αντικειμένων (ή άλλων οντοτήτων) όπου βασικός σκοπός είναι η μελέτη, ανάλυση και διεξαγωγή αποτελεσμάτων, τα οποία θα ερμηνεύουν με τον καλύτερο δυνατό τρόπο το υπό-μελέτη σύνολο.

Μπορούμε να ορίσουμε διαφορετικούς πληθυσμούς ανάλογα με τα χαρακτηριστικά των μελών του (ανθρώπων, φυτών, ζώων, ποδοσφαιρικών ομάδων, εκλογικών τμημάτων).

Για κάθε στατιστική μελέτη, ο πληθυσμός πρέπει να είναι καλά ορισμένος, να περιγράφεται, όσο το δυνατό καλύτερα με βάση τα κοινά χαρακτηριστικά που τον απαρτίζουν.

Για παράδειγμα, όπως το σύνολο των Ελλήνων, οι μαθητές της 3^{ης} Λυκείου, οι φοιτητές του Πανεπιστημίου Αιγαίου, αποτελούν μερικά χαρακτηριστικά παραδείγματα πληθυσμών.

Ένας πληθυσμός μπορεί να χαρακτηριστεί ως υπαρκτός, όταν καθορίζεται από συγκεκριμένα χαρακτηριστικά γνωρίσματα (ύψος φοιτητών ενός Πανεπιστημίου) ή ως ιδεατός όταν καθορίζεται από αφηρημένα (ασαφή) χαρακτηριστικά γνωρίσματα (όπως το δείκτη που οι φοιτητές ενός Πανεπιστημίου είναι ικανοποιημένοι με ένα συγκεκριμένο μάθημα).

Ιδεατοί πληθυσμοί χρησιμοποιούνται τις περισσότερες φορές σε μελέτες έρευνας αγοράς ή σε κλινικές μελέτες.

Πληθυσμός με πεπερασμένο πλήθος στατιστικών χαρακτηριστικών ονομάζεται **πεπερασμένος**. Σε αντίθετη περίπτωση, ονομάζεται **μη πεπερασμένος ή άπειρος**.

Επειδή τις περισσότερες φορές οι πληθυσμοί που μελετώνται είναι μεγάλοι σε μέγεθος, καταλήγουμε στη διαδικασία της δειγματοληψίας. Στην περίπτωση αυτή ο ερευνητής είναι αυτός που ορίζει τον υπό-μελέτη πληθυσμό ανάλογα με τις ανάγκες και τις απαιτήσεις της μελέτης.

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 μία χαρακτηριστική ιδιότητα, καλείται **άπειρος πληθυσμός**. Σε πολλές περιπτώσεις, παρ' ότι είναι πεπερασμένο το πλήθος των στοιχείων του πληθυσμού αναφέρεται συνήθως ως μέγεθος του πληθυσμού. Ο πληθυσμός θεωρείται άπειρος, εφ' όσον το μέγεθος είναι αρκετά μεγάλο σε σχέση με την κλίμακα μετρήσεων που χρησιμοποιείται.

Ο πληθυσμός που διερευνάται θα πρέπει να ορισθεί με απόλυτη ακρίβεια ως προς τα στοιχεία που τον απαρτίζουν. Για παράδειγμα, 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.

1.3 Στατιστική Μονάδα και Μεταβλητές

Στατιστικές μονάδες είναι τα αντικείμενα από τα οποία λαμβάνουμε τις πληροφορίες που επιθυμούμε να επεξεργαστούμε και να αναλύσουμε στατιστικά. Η στατιστική μονάδα μπορεί να είναι ένα αντικείμενο, ένα άτομο, μία επιχείρηση, ένα ίδρυμα, κάποιο γεγονός (εκλογικές αναμετρήσεις, αθλητικοί αγώνες) ή ακόμα και έννοια. Πριν από οποιαδήποτε ενέργεια για τη συλλογή στατιστικών δεδομένων είναι απαραίτητο να προδιαγραφεί με ακρίβεια και σαφήνεια η στατιστική μονάδα

Σκοπός είναι η παροχή λεπτομερέστερων ή και περισσότερο ομοιογενών δεδομένων από αυτά που θα μπορούσαμε να αποκτήσουμε με άλλους τρόπους.

Μεταβλητή (variable) είναι η ιδιότητα ή το χαρακτηριστικό ως προς το οποίο ερευνάται ένας πληθυσμός. Για παράδειγμα, οι φοιτητές του ΤΕΙ Πατρών μπορούν να εξεταστούν ως προς το φύλο, την ηλικία, το βαθμό σε ένα μάθημα κ.λ.π. (ΔΑΜΙΑΝΟΥ Χ.-1992)

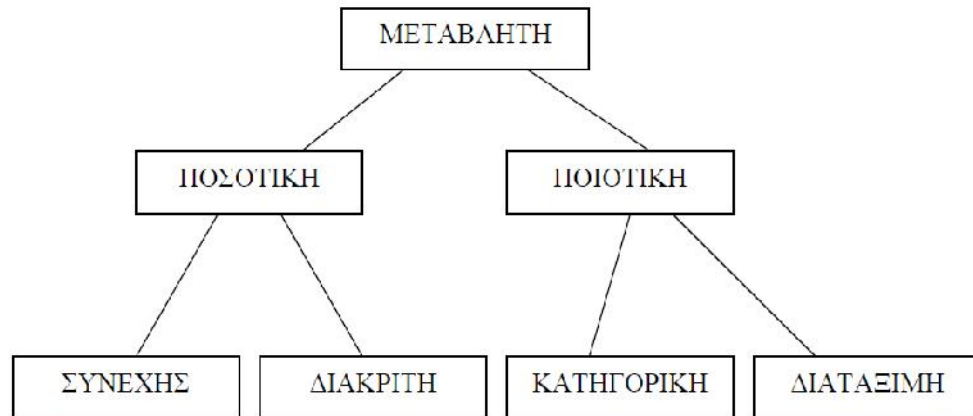
Κάθε μεταβλητή συμβολίζεται με ένα από τα κεφαλαία γράμματα X, Y, Z ενώ οι τιμές του, με τα αντίστοιχα μικρά γράμματα $x_1, x_2, x_3, \dots, x_n$ ή $y_1, y_2, y_3, \dots, y_k$.

Εάν μια μεταβλητή μπορεί να πάρει κάποιον αριθμό διαφορετικών τιμών ούτως ώστε κάποιο συγκεκριμένο αποτέλεσμα να καθορίζεται τυχαία, καλείται **τυχαία μεταβλητή (random variable)**.

Στην πραγματικότητα, η τυχαία μεταβλητή είναι μια πραγματική συνάρτηση που ορίζεται στον δειγματικό χώρο ενός πειράματος. Οι τυχαίες μεταβλητές συμβολίζονται με κεφαλαία γράμματα και οι τιμές τους με τα αντίστοιχα μικρά. Διακρίνονται σε διακριτές και συνεχείς τυχαίες μεταβλητές.

Για την καλύτερη κατανόηση της έννοιας της μεταβλητής, δίνονται μερικά παραδείγματα στον παρακάτω πίνακα:

Μεταβλητές	Τιμές
Αριθμός αγοριών σε μία οικογένεια	0,1,2,3,4,5,...
Ένδειξη ενός ζαριού	1,2,3,4,5,6
Ένδειξη ενός νομίσματος	«κεφαλί», «γράμματα»
Αριθμός δωματίων ενός διαμερίσματος	1,2,3,4
Φυλή ατόμων	λευκός, μαύρος, κίτρινος, ερυθρόδερμος
Ύψος ατόμων	145, 146, 147,...
Υγεία ατόμων	Άριστη, καλή, μέτρια, κακή
Φύλο ατόμων	Αρσενικό, θηλυκό



Οι μεταβλητές χωρίζονται σε δύο κύριες κατηγορίες:

α) Οι ποσοτικές μεταβλητές είναι οι μεταβλητές των οποίων οι τιμές έχουν αριθμητικές ιδιότητες και εκφράζονται με μία μονάδα μέτρησης. Για παράδειγμα, το εισόδημα, το βάρος, το ύψος, ο αριθμός των παιδιών μιας οικογένειας.

Διακρίνονται με την σειρά τους σε **συνεχείς** και **διακριτές** ανάλογα με το αν είναι συνεχείς ή διακριτές.

Συνεχείς μπορούν να καλούνται, οι μεταβλητές όπου μπορούν να πάρουν οποιαδήποτε τιμή σε κάποιο διάστημα (ύψος, βάρος, μήκος, αριθμός ωρών που εργάζεται κάποιος, ύψος μηνιαίου μισθού).

Ενώ **διακριτή** καλείται, μια μεταβλητή που παίρνει ακέραιες τιμές από εάν πεπερασμένο ή αριθμήσιμο σύνολο (η ηλικία, αριθμός παιδιών σε μία οικογένεια, αριθμός φοιτητών σε ένα τμήμα πανεπιστημίου, αριθμός τηλεφωνημάτων που δέχεται κάποιος σε μια μέρα).

β) Οι ποιοτικές μεταβλητές είναι αυτές που δεν επιδέχονται μέτρηση και αναφέρονται σε κάποιο ποιοτικό χαρακτηριστικό, όπως το φύλο, το επίπεδο μόρφωσης, περιοχή καταγωγής.

Για τον προσδιορισμό των μεταβλητών είναι απαραίτητη η ύπαρξη μίας κλίμακας μέτρησης. Ευρέως χρησιμοποιούνται οι εξής τέσσερις κλίμακες, διάταξης, κατηγορίας, διαστήματος και αναλογίας. Οι δύο πρώτες αφορούν τις ποιοτικές μεταβλητές ενώ οι δύο τελευταίες τις ποσοτικές.

Διάταξης (ordinal) είναι οι μεταβλητές που για το σύνολο τιμών τους μπορούμε να ορίσουμε μία σχέση διάταξης, δηλαδή να τοποθετηθούν στη σειρά. Η διάταξη μπορεί να είναι από την μικρότερη τιμή προς την μεγαλύτερη ή και αντίστροφα. Οι ίσες διαφορές μεταξύ των τιμών μίας τέτοιας μεταβλητής δεν συνεπάγονται και ίσες διαφορές για το χαρακτηριστικό που μετράει η μεταβλητή. Δηλαδή δεν υπάρχει αντιστοίχιση σε υποδιαίρεσεις ή πολλαπλάσια

κάποιας μονάδας. Η διάταξη δηλαδή, το μόνο που εξασφαλίζει είναι τον προσδιορισμό της μεγαλύτερης, καλύτερης, προτιμότερης κατηγορίας αλλά όχι πόσο μεγαλύτερη, καλύτερη, προτιμότερη είναι σε σχέση με κάποια από τις υπόλοιπες. Για παράδειγμα, επίπεδα εκπαίδευσης, αγωνίσματα, αξιολόγηση μίας ταινίας.

Κατηγορίας (nominal) είναι οι μεταβλητές των οποίων το σύνολο των τιμών δεν έχει καμία ιδιότητα. Για τη μεταβλητή αυτή, μοναδική σημασία έχουν οι διαφορετικές τιμές (το πλήθος των κατηγοριών της) που μπορεί να πάρει. Η μοναδική σχέση που μπορεί να προσδιοριστεί μεταξύ των κατηγοριών αυτών είναι απλά η ύπαρξη διαφοράς. Για παράδειγμα, χρώμα ματιών, φύλο, οικογενειακή κατάσταση.

Διαστήματος (interval) είναι οι μεταβλητές των οποίων οι ίσες διαφορές μεταξύ των τιμών τους συνεπάγονται και ίσες διαφορές για το χαρακτηριστικό που μετράει η μεταβλητή. Για παράδειγμα, ηλικία, θερμοκρασία. Η κλίμακα αυτή δεν επιτρέπει μόνο την ιεράρχηση των υποκειμένων αλλά προσδιορίζει επίσης και την ακριβή διαφορά τους. Η απόσταση μεταξύ δυο οποιονδήποτε διαδοχικών τιμών της μεταβλητής αυτής είναι ίση με την απόσταση δυο άλλων τυχαίων διαδοχικών τιμών της. Επίσης, δεν έχει νόημα ο υπολογισμός αναλογιών. Για παράδειγμα, βασικό χαρακτηριστικό των μεταβλητών αυτής της διάταξης είναι ο αυθαίρετος ορισμός του μηδενός, που δεν υποδηλώνει παντελή έλλειψη του μετρήσιμου χαρακτηριστικού.

Αναλογίας (rate) είναι οι μεταβλητές των οποίων οι τιμές αντιστοιχούν αναλογικά στην ποσότητα του χαρακτηριστικού που μετρούν. Εδώ το μηδέν ανήκει στο διάστημα τιμών της μεταβλητής και δηλώνει την πλήρη απουσία. Επίσης, για τις τιμές των μεταβλητών αυτών έχει έννοια ο υπολογισμός των αναλογιών.

1.4 Δείγμα

Κάθε υποσύνολο του πληθυσμού αποτελεί ένα δείγμα (sample) από τον πληθυσμό.

Το πλήθος των στοιχείων του δείγματος, που λαμβάνεται μέσα από τον πληθυσμό, θα συμβολίζεται με n .

Το μέγεθος του δείγματος δεν χρειάζεται να είναι κάποιο συγκεκριμένο ποσοστό του μεγέθους του πληθυσμού. Αντίθετα, προσδιορίζεται σε σχέση με το μέγεθος του σφάλματος εκτίμησης που παράγεται εξαιτίας της χρήσης δειγματοληψίας, της μεθόδου δειγματοληψίας, της σχετικής ομοιογένειας του πληθυσμού και του κόστους δειγματοληψίας.

Ο όρος δείγμα εμπεριέχει τρεις βασικές έννοιες:

α) Ένα δείγμα μπορεί να περιέχει πρωταρχικές μονάδες δειγματοληψίας, όπως για παράδειγμα οι επιχειρήσεις ενός βιομηχανικού κλάδου. Ακόμη και αν ο σκοπός του

δείγματος είναι να μελετήσουμε τις ιδιότητες των στοιχειωδών μονάδων, όπως για παράδειγμα είναι οι εργαζόμενοι στον βιομηχανικό κλάδο.

β) Ένα δείγμα δεν είναι απλώς ένα κομμάτι του πληθυσμού που λαμβάνεται για να εξασφαλισθεί η άνετη μελέτη του πληθυσμού, αλλά ένα υποσύνολο (ένας μικρόκοσμος), που αναμένεται να είναι αντιπροσωπευτικό του γεννήτορα (υπό δειγματοληψία) πληθυσμού.

γ) Ο τελικός σκοπός της δειγματοληψίας δεν είναι απλώς η εξασφάλιση δειγματικών συναρτήσεων, αλλά η εξαγωγή επαγωγικών συμπερασμάτων που αφορούν τον υπό δειγματοληψία πληθυσμό.

Παράδειγμα:

Μετρήσαμε τα ύψη (μονάδα μέτρησης αποτελεί το ένα εκατοστό του μέτρου – 1 cm) των μαθητών μιας τάξης η οποία έχει $n = 12$ μαθητές και βρήκαμε τα αποτελέσματα του πίνακα:

156,4	162,3	149,7	152,5	155,3	167,1
148,2	160,4	170,1	157,5	162,7	163,2

Θεωρούμε δείγμα μεγέθους $n = 3$. Το δείγμα είναι τυχαίο. Έστω X_1 το ύψος του 1^{ου} μαθητή του δείγματος, X_2 το ύψος του 2^{ου} μαθητή και X_3 το ύψος του 3^{ου} μαθητή. Αν το δείγμα περιέχει τον 4^ο μαθητή τον 7^ο μαθητή και τον 12^ο μαθητή της τάξης, τότε οι τιμές των τυχαίων μεταβλητών στο δείγμα αυτό είναι:

$x_1 = 152,5$	$x_2 = 148,2$	$x_3 = 163,2$
---------------	---------------	---------------

Ένα άλλο δείγμα ίδιου μεγέθους μπορεί να περιέχει τον 2^ο μαθητή, τον 5^ο μαθητή και τον 7^ο μαθητή. Τώρα, οι τιμές των τυχαίων μεταβλητών X_1, X_2, X_3 στο δείγμα είναι:

$x_1 = 162,3$	$x_2 = 155,3$	$x_3 = 148,2$
---------------	---------------	---------------

Με τον ίδιο τρόπο, μπορούμε να θεωρήσουμε και άλλα δείγματα από το συγκεκριμένο πληθυσμό. Όλα τα δείγματα έχουν το ίδιο μέγεθος, κατόπιν συμφωνίας. Βεβαίως θα μπορούσαμε να είχαμε συμφωνήσει για διαφορετικό μέγεθος δείγματος.

1.4.1 Τυχαίο Δείγμα

Τυχαίο δείγμα (random sample) είναι το δείγμα του πληθυσμού, όπου τα άτομα διαλέγονται το ένα μετά το άλλο, με κύριο χαρακτηριστικό, ότι τα υπόλοιπα άτομα του πληθυσμού κάθε φορά, έχουν τις ίδιες πιθανότητες να περιληφθούν στο τυχαίο δείγμα.

Ένα (απλό) **τυχαίο δείγμα**, είναι ένα δείγμα, το οποίο επιλέγεται κατά τέτοιο τρόπο ώστε κάθε μονάδα του πληθυσμού έχει την ίδια πιθανότητα να επιλεγεί στο δείγμα, ανεξάρτητα από το ποιες μονάδες του πληθυσμού έχουν ήδη επιλεγεί. (B. Μπένου ,1997).

Έστω η τυχαία μεταβλητή X που περιγράφει το χαρακτηριστικό που ερευνούμε.

Επιλέγουμε με τυχαίο τρόπο n στοιχεία του πληθυσμού μεγέθους N και συμβολίζουμε με $x_i, i \in \{1, \dots, n\}$ την τιμή που παίρνει η τυχαία μεταβλητή για το στοιχείο αυτό. Το διάνυσμα $X = (X_1, \dots, X_n) \in R^n$ ονομάζεται τυχαίο δείγμα μεγέθους n . Ο όρος "τυχαίο δείγμα" χρησιμοποιείται και για την αντίστοιχη τυχαία μεταβλητή $X = (X_1, \dots, X_n)$, όπου οι τυχαίες μεταβλητές $x_i, i \in \{1, \dots, n\}$ θεωρούνται ανεξάρτητες και ισόνομες, δηλαδή την ίδια συνάρτηση κατανομής.

Για παράδειγμα, το να επιλέξουμε τρία ονόματα από ένα καπέλο που περιέχει όλα τα ονόματα των φοιτητών της τάξης είναι ένα παράδειγμα απλού τυχαίου δείγματος, όπου κάθε ομάδα τριών ατόμων έχει την ίδια πιθανότητα να επιλεγεί.

Κεφάλαιο 2: Δειγματοληψία

2.1 Εισαγωγή

Η δειγματοληψία ασχολείται με το πώς εμείς μπορούμε να επιλέξουμε το δείγμα μας. Δηλαδή με τις μεθόδους επιλογής δείγματος, με τον καθορισμό του μεγέθους του, καθώς και με τα σφάλματα που υπάρχουν στις δειγματοληπτικές έρευνες.

Η πιθανότητα να επιλεγεί μια οποιαδήποτε μονάδα σε μια τυχαία επιλογή είναι $1/N$. Η πιθανότητα μια μονάδα του πληθυσμού να επιλεγεί σε n δοκιμές είναι n/N . Το τυχαίο δείγμα (random sample) εκλέγεται με τη μέθοδο των πιθανοτήτων κατά τυχαίο τρόπο και αυτό δεν επιτρέπει στον ερευνητή να καθορίζει ποιες μονάδες του πληθυσμού θα περιληφθούν στο δείγμα. Ένα τυχαίο δείγμα, που οι πιθανότητες εκλογής του είναι γνωστές, ονομάζεται **δείγμα πιθανότητας (probability sample)**.

Έχουν αναπτυχθεί πολλές μέθοδοι επιλογής δείγματος που μας εξασφαλίζουν την τυχειότητα (η μέθοδος των λαχνών, η μέθοδος των τυχαίων αριθμών). Αν η επιλογή του δείγματος γίνει τυχαία τότε θεωρείται αμερόληπτο, και οδηγεί σε αμερόληπτη στατιστικά εκτίμηση των παραμέτρων του πληθυσμού. Ο τρόπος με τον οποίο παίρνουμε ένα δείγμα από τον πληθυσμό μας, λέγεται δειγματοληπτικό σχέδιο.

Για παράδειγμα, οι κερδίζοντες αριθμοί των λαχείων θα αποτελούν ένα τυχαίο δείγμα του πληθυσμού των λαχείων, αν ο τρόπος κληρώσεως των λαχείων που κερδίζουν, είναι πράγματι τυχαίος.

Συμπεραίνουμε ότι, το τυχαίο δείγμα έχει την ιδιότητα η επιλογή των μονάδων του να γίνεται αμερόληπτα. Δηλαδή, όλες οι μονάδες του πληθυσμού να έχουν την ίδια πιθανότητα να επιλεγούν στο δείγμα. Για να είναι το δείγμα αντιπροσωπευτικό και αμερόληπτο, θα πρέπει οι περιοχές από τις οποίες παίρνουμε τις δειγματοληπτικές μονάδες να είναι αντιπροσωπευτικές του πληθυσμού. Πολλές φορές όμως το κόστος της έρευνας είναι τόσο υψηλό, ώστε να αναγκάζομαστε να πάρουμε το δείγμα μόνο από μια περιοχή οπότε υπάρχει κίνδυνος το δείγμα να είναι μεροληπτικό και να βγάλουμε λανθασμένα συμπεράσματα. (Lindqvist M., 1993)

Παράδειγμα μεροληπτικού δείγματος είναι η πρόβλεψη των προεδρικών εκλογών των Η.Π.Α το 1936 από ένα αμερικάνικο περιοδικό. Το δείγμα του περιοδικού περιείχε 2.000.000 απαντήσεις και προέβλεψε ότι ο Ρούζβελτ θα έπαιρνε το 41% των ψήφων, ενώ στην πραγματικότητα πήρε το 61% των ψήφων. Αυτό οφειλόταν στο ότι το δείγμα είχε ληφθεί από τηλεφωνικούς καταλόγους και καταλόγους Ι. Χ. αυτοκινήτων και δεν ήταν αντιπροσωπευτικό δείγμα του εκλογικού σώματος των Η. Π. Α. Δηλαδή, το δείγμα ήταν μεροληπτικό και έδωσε λανθασμένα προγνωστικά αποτελέσματα. (Lindqvist M., 1993)

2.2 Ορισμός δειγματοληψίας

Δειγματοληψία (sampling) ονομάζεται η διαδικασία με την οποία επιλέγεται το τυχαίο δείγμα.

2.3 Θεωρητική βάση της δειγματοληψίας

Η δειγματοληψία είναι ένα ισχυρό εργαλείο στη διάθεση κάθε στατιστικού, με το οποίο προβλέπουμε και γενικεύουμε (επαγωγικά) την συμπεριφορά μαζικών φαινομένων. Είναι πράγματι ένα μεγάλο επιστημονικό επίτευγμα “για επιστημονική γνώση του κόσμου γύρω μας” αναφέρει ο Roy Jastram στο “Elements of Statistical Inference 1948”. “Για να γενικεύσουμε λογικά και με ακρίβεια σε πληθυσμό χιλιάδων μονάδων, που δεν έχουμε δει, συμπεράσματα που προκύπτουν από εκτίμηση που αφορούν σε πενήντα ή εκατό από αυτές τις μονάδες”. Η θεωρία της δειγματοληψίας, που καθιστά δυνατή αυτή την επαγωγική διαδικασία, έχει σαν βάση τα κύρια χαρακτηριστικά των μαζικών δεδομένων, τα οποία μπορούν να συνοψίσουν απλά και με ακρίβεια τη φράση “ποικιλία σε ενιαίο σύνολο”.

Εξ’ άλλου, οι στοιχειώδεις μονάδες οποιουδήποτε πληθυσμού επηρεάζονται από μια πολλαπλότητα δυνάμεων. Αυτή η τεράστια σύνθεση δυνάμεων, που κατά κανόνα σχετίζονται μεταξύ τους, επιδρά επί των επί μέρους στοιχειωδών μονάδων με σημαντικό βαθμό ανεξαρτησίας. Αυτό ερμηνεύει τις μεταβολές (αποκλίσεις) από μονάδα σε μονάδα του πληθυσμού.

Για παράδειγμα, στην χρηματιστηριακή αγορά, κατά την ίδια μέρα, διαπιστώνουμε ότι σε κάποιες μετοχές αυξάνουν οι τιμές, σε άλλες μειώνονται και σε άλλες παραμένουν σχετικά αμετάβλητες. Η απαρίθμηση παραδειγμάτων μπορεί να γίνει κατά χιλιάδες.

Αν και η ποικιλία είναι μια καθολική ιδιότητα των μαζικών δεδομένων, δεν υπάρχει στατιστικός πληθυσμός, του οποίου οι μονάδες θα ποικίλλουν μεταξύ τους ως προς το ερευνώμενο χαρακτηριστικό, χωρίς όριο.

Για παράδειγμα, οι έφηβοι σε ένα ανθρώπινο πληθυσμό μπορεί να διαφέρουν μεταξύ τους ως προς το χαρακτηριστικό “βάρος” αλλά κανένα συνηθισμένο άτομο αυτής της κατηγορίας δεν θα είναι τόσο κοντό όπως ένα μέτρο ή τόσο ψηλό όπως δύομιση μέτρα.

Λόγω της στατιστικής, εάν ένα μεγάλο τυχαίο δείγμα επιλεγεί, τα χαρακτηριστικά του θα διαφέρουν ελάχιστα από τα αντίστοιχα του πληθυσμού. Εξ’ αιτίας της αποκλίσεως, εάν λάβουμε έναν αριθμό τυχαίων δειγμάτων, αν και όμοια από πολλές απόψεις, τα δείγματα δεν είναι δυνατόν να έχουν αυστηρή κανονικότητα. Η κανονικότητα ή ομοιομορφία, αναφέρεται στην τάση των υπό μέτρηση χαρακτηριστικών να συσσωρεύονται γύρω από ένα συγκεκριμένο “κέντρο έλξεως”.

Το μέτρο για μια τέτοια κεντρική τάση είναι ουσιαστικά ένας μέσος όρος, από τον οποίο, οι επί μέρους παρατηρήσεις αποκλίνουν με κάποιο συγκεκριμένο τρόπο. Οι μέσοι παρουσιάζουν μεγαλύτερη σταθερότητα απ' ό,τι οι επιμέρους παρατηρήσεις και γίνονται περισσότερο σταθεροί όταν περιλαμβάνονται στο δείγμα περισσότερες παρατηρήσεις.

Σύμφωνα και με το “**Νόμο των μεγάλων αριθμών**”, αν και τα δείγματα ποικίλλουν, οι μέσοι ποικίλλουν λιγότερο σε μεγάλα δείγματα σε σύγκριση με τα μικρά δείγματα, που έχουν ληφθεί από τον ίδιο πληθυσμό.

Αυτός ο νόμος έχει δύο απόψεις βασικής σημασίας για την αξιοπιστία των δειγμάτων και την επάρκεια των μεθόδων δειγματοληψίας:

α. Εάν όλο ένα μεγαλύτερο και μεγαλύτερο τυχαίο δείγμα λαμβάνεται από τον ίδιο πληθυσμό, τα χαρακτηριστικά μέτρα κάθε μεγαλύτερου δείγματος θα τείνουν να διαφέρουν όλο και λιγότερο από τα αντίστοιχα χαρακτηριστικά μέτρα του πληθυσμού. Δηλαδή, όσο μεγαλύτερο είναι το δείγμα τόσο πιο αντιπροσωπευτικό του πληθυσμού είναι.

β. Εάν ένας μεγάλος αριθμός μεγάλων τυχαίων δειγμάτων λαμβάνεται από τον ίδιο πληθυσμό, τότε όχι μόνο κάθε δείγμα θα διαφέρει από τον πληθυσμό πολύ λίγο, αλλά και τα δείγματα μεταξύ τους θα έχουν πολύ μικρές αποκλίσεις.

Συνεπώς, είτε το δείγμα είναι μεγάλο, είτε μικρό, υπάρχει βεβαιότητα ότι τα χαρακτηριστικά του δεν είναι ακριβώς εκείνα του πληθυσμού. Σχεδόν κάθε μέθοδος δειγματοληψίας παρουσιάζει κάποια μεταβλητότητα.

Όμως, η φύση της μελετώμενης κατανομής συχνοτήτων, μπορεί να είναι γνωστή, μόνο όταν η δειγματοληψία είναι τυχαία, καθώς αυτή η γνώση μπορεί να επιτευχθεί με την εφαρμογή νόμων πιθανοτήτων, οι οποίοι με την σειρά τους εφαρμόζονται μόνο σε τυχαία δείγματα.

Από τα παραπάνω μπορούν να προκύψουν πολλά χρήσιμα συμπεράσματα για το θεωρητικό υπόβαθρο της δειγματοληψίας.

Πρώτον, ένα τυχαίο δείγμα μπορεί να αντιπροσωπεύσει αρκετά καλά έναν πληθυσμό καθώς κάθε πληθυσμός παρουσιάζει σχετικά σταθερά χαρακτηριστικά γνωρίσματα και περιορισμένες αποκλίσεις.

Επιπλέον, τυχαία δείγματα που επιλέγονται από τον ίδιο πληθυσμό αναμένεται να έχουν πολλά κοινά χαρακτηριστικά, αλλά ποτέ δεν θα έχουν αυστηρή ομοιομορφία λόγω της καθολικής ιδιότητας, “απόκλιση στην ενότητα των μαζικών δεδομένων”. Ένα μεγάλο δείγμα θα διαφέρει λιγότερο από ένα μικρότερο δείγμα, από τον πληθυσμό και όσο αυξάνει το μέγεθος των δειγμάτων τόσο η μεταβλητότητα μεταξύ τους θα μειώνεται.

Τέλος, η δειγματοληπτική μεταβλητότητα μπορεί να μετρηθεί εάν και μόνο εάν η δειγματοληψία είναι τυχαία, διότι οι νόμοι των πιθανοτήτων εφαρμόζονται σε τυχαία

δείγματα και συνεπώς καθιστούν δυνατή τη γνώση της φύσης της συγκεκριμένης κατανομής συχνοτήτων.

2.4 Πλεονεκτήματα της δειγματοληψίας

Τα κυριότερα πλεονεκτήματα της δειγματοληπτικής μεθόδου είναι:

α. Μεγαλύτερη ταχύτητα πληροφοριών

Η συλλογή του στατιστικού υλικού που χρειάζεται για να μελετηθούν οι χαρακτηριστικές ιδιότητες ενός πληθυσμού μπορεί να γίνει πιο γρήγορα με τη δειγματοληψία παρά με την καθολική απογραφή των μονάδων του φαινομένου που εξετάζουμε. Αυτό έχει ιδιαίτερη σημασία για συγκέντρωση πληροφοριών που επείγουν.

β. Μεγαλύτερη ακρίβεια

Σε δειγματοληπτικές έρευνες είναι δυνατόν, όταν ο αριθμός των μονάδων του πληθυσμού είναι μικρός, να αφιερωθεί περισσότερος χρόνος και μεγαλύτερη προσοχή στις συνεντεύξεις που παίρνουμε για να συμπληρώσουμε το ερωτηματολόγιο και να γίνει καλύτερη εκπαίδευση και επίβλεψη στους απογραφείς. Το γεγονός αυτό έχει μεγάλη σημασία, γιατί μια δειγματοληπτική έρευνα που έγινε με σωστό τρόπο μπορεί να δώσει αποτελέσματα, παρά τα δειγματοληπτικά σφάλματα, με μεγαλύτερη ακρίβεια και από εκείνα ακόμα τις καθολικής απογραφής.

γ. Μεγαλύτερη ευχέρεια απογραφής

Η δειγματοληπτική έρευνα εφαρμόζεται σε εκείνες τις περιπτώσεις που η γενική απογραφή είναι δυνατή αλλά παράλογη. Για παράδειγμα, έστω ότι γιατρός θέλει να ερευνήσει τα ερυθρά αιμοσφαίρια ενός ασθενούς. Για να το πετύχει αυτό δεν θα πάρει όλη την ποσότητα αίματος του ασθενούς, αλλά λίγα μόνο γραμμάρια. Στις περιπτώσεις αυτές δεν υπάρχει πρόβλημα επιλογής για το αν θα ακολουθήσουμε δηλαδή τη δειγματοληπτική μέθοδο ή τη μέθοδο της καθολικής απογραφής. Εκείνο που θα μας προβληματίσει περισσότερο είναι να μείνουμε χωρίς πληροφορίες ή να τις πάρουμε από τη εξέταση ενός δείγματος του πληθυσμού, γιατί η μέθοδος της γενικής απογραφής είναι παράλογο να χρησιμοποιηθεί.

δ. Χαμηλό κόστος

Κύριος και αντικειμενικός σκοπός κάθε δειγματοληπτικής έρευνας είναι η λήψη μιας πληροφορίας με τη μεγαλύτερη δυνατή ακρίβεια και με το ελάχιστο κόστος. Για παράδειγμα, είναι πιο λίγο δαπανηρή η συγκέντρωση και ανάλυση πληροφοριών από μερικές μόνο χιλιάδες ή εκατοντάδες κατοίκων μιας πόλης για την προτίμηση ενός αγαθού, παρά από το σύνολο των κατοίκων της.

ε. Ολοκληρωτική δύναμη εφαρμογής της γενικής απογραφής

Η δειγματοληπτική έρευνα μπορεί να εφαρμοστεί σε εκείνες τις περιπτώσεις που η καθολική έρευνα είναι αδύναμη. Για παράδειγμα, για να ερευνήσουμε αν μια θαλάσσια περιοχή έχει ψάρια, περιοριζόμαστε σε μερική έρευνα, γιατί είναι αδύνατη η καθολική.

2.5 Μειονεκτήματα της δειγματοληψίας

Παράλληλα με τα πλεονεκτήματα της δειγματοληψίας υπάρχουν και μερικά μειονεκτήματα. Αναφέρουμε τα ακόλουθα:

α. Αν οι μονάδες του πληθυσμού που ερευνούμε εμφανίζονται πολύ σπάνια, πρέπει να μελετήσουμε ένα σημαντικό μεγάλο δείγμα, αν θέλουμε να έχουμε αξιόπιστη εκτίμηση των χαρακτηριστικών ιδιοτήτων του πληθυσμού.

Για παράδειγμα, αν θέλουμε με την βοήθεια δείγματος να μελετήσουμε το ποσοστό καπνιστών από άτομα μίας πόλης που έχουν ηλικία πάνω από 8 χρόνων, πρέπει να πάρουμε ένα αρκετά μεγάλο δείγμα ώστε σε αυτό να περιλαμβάνεται ικανοποιητικός αριθμός ατόμων με ηλικία 8 ετών και πάνω, γιατί τα άτομα με τέτοια ηλικία αποτελούν μικρό ποσοστό στο συνολικό πληθυσμό της πόλης.

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

γ. Διάφοροι παράγοντες, όπως η κακή σχεδίαση και εκτέλεση της δειγματοληψίας, η μη αντιπροσωπευτικότητα του δείγματος, η μη κατάλληλη μέθοδος διενέργειας της δειγματοληψίας και τα ανεπαρκή δεδομένα, οδηγούν σε αποτυχία της μερικής έρευνας.

δ. Άλλο βασικό μειονέκτημα της δειγματοληψίας είναι τα δειγματοληπτικά σφάλματα, τα οποία θα αναφέρουμε παρακάτω.

2.6 Ο υπό Δειγματοληψία Πληθυσμός

Ο **υπό δειγματοληψία πληθυσμός (sampled population)** ονομάζεται ο πληθυσμός από τον οποίον επιλέγεται ένα δείγμα για τη μελέτη των χαρακτηριστικών του, που μας ενδιαφέρουν.

Ο υπό δειγματοληψία πληθυσμός πρέπει συνήθως να ταυτίζεται με τον πληθυσμό στόχο.

Ο **πληθυσμός στόχος (target population)** είναι ο σωστός πληθυσμός ως προς το αντικείμενο της έρευνας.

Για να γίνει κατανοητή η διαφορά ανάμεσα στον υπό δειγματοληψία πληθυσμό και στον πληθυσμό στόχο δίνεται το ακόλουθο παράδειγμα.

Εάν το αντικείμενο της έρευνας είναι η μελέτη του ποσοστού των Ελλήνων καπνιστών, ο **πληθυσμός στόχος** είναι οι Έλληνες. Στην πράξη, όμως, πολλές φορές ο υπό δειγματοληψία πληθυσμός δεν ταυτίζεται με τον πληθυσμό στόχο. Υπάρχουν Έλληνες που κατοικούν σε απομακρυσμένες περιοχές όπου για να επισκεφθούμε είναι δύσκολο ή χρονοβόρο, οπότε πρέπει να τους εξαιρέσουμε από τον πληθυσμό και να περιοριστούμε στους Έλληνες που κατοικούν σε προσιτές περιοχές. Ο πληθυσμός στόχος είναι όλοι οι Έλληνες που κατοικούν σε προσιτές περιοχές και ασφαλώς οι δύο πληθυσμοί είναι διαφορετικοί. Αυτό που χρειάζεται να προσέξουμε εδώ πέρα είναι ότι τα αποτελέσματα της στατιστικής μας ανάλυσης θα αναφέρονται στον υπό δειγματοληψία πληθυσμό κι όχι στον πληθυσμό στόχο, δηλαδή θα αναφέρονται στους Έλληνες που κατοικούν σε προσιτές περιοχές κι όχι σε όλους τους Έλληνες. (W. Cochran, 1997)

2.7 Δειγματοληπτική Μονάδα

Δειγματοληπτική μονάδα (sampling unit) ονομάζεται ένα υποσύνολο του πληθυσμού, δηλαδή ένα σύνολο αποτελούμενο από κάποιες στατιστικές μονάδες του πληθυσμού.

Ας υποθέσουμε ότι επιθυμούμε να κάνουμε μια έρευνα οικογενειακών προϋπολογισμών σε κάποια πόλη. Αν και είναι προφανές ότι οι μονάδες του υπό μελέτη πληθυσμού είναι οικογένειες, πρέπει να υιοθετηθεί κάποιος συμβατικός ορισμός της έννοιας της οικογένειας. Ακόμη και σε αυτή την περίπτωση ενδέχεται να μην υπάρχει εύκολος τρόπος δειγματοληψίας τέτοιων μονάδων (δηλαδή οικογενειών). Θα ήταν ευκολότερο να ληφθεί ένα δείγμα διευθύνσεων και να αναζητηθούν πληροφορίες για τις οικογένειες που κατοικούν στις επιλεγείσες διευθύνσεις. Έτσι οι διευθύνσεις αποτελούν τις δειγματοληπτικές μονάδες.

2.7.1 Δειγματικό πλαίσιο ή δειγματοληπτικός σκελετός

Δειγματοληπτικό πλαίσιο (sampling frame) ή **δειγματοληπτικός σκελετός (sampling frame)** ονομάζεται ένα σύνολο από δειγματοληπτικές μονάδες που αποτελούν διαμερισμό του πληθυσμού. Δηλαδή η ένωση των δειγματοληπτικών μονάδων του πληθυσμού μας δίνει τον ίδιο τον πληθυσμό. Οι δειγματοληπτικές αυτές μονάδες ανά δύο είναι ξένες μεταξύ τους και η τομή τους είναι το κενό. (W. Cochran, 1997)

2.7.2 Κατάλογος ή λίστα

Το πλαίσιο δειγματοληψίας είναι ένας κατάλογος που περιέχει όλες τις μονάδες που αναφέρονται στη χαρακτηριστική ιδιότητα που θέλουμε να μελετήσουμε, από τον οποίο με την εφαρμογή διαφόρων μεθόδων θα επιλέξουμε το επιθυμητό δείγμα. Η επιλογή του δείγματος γίνεται ευκολότερη αν υπάρχει ένας πραγματικός κατάλογος των δειγματοληπτικών μονάδων, όπως κατάλογος διευθύνσεων, εργαζομένων στη περιοχή Πατρών, λίστα κρατικών απογραφών νοικοκυριών, εκλογικοί κατάλογοι κτλ. Ένας τέτοιος κατάλογος διευκολύνει ιδιαίτερα την επιλογή του δείγματος.

Υπάρχουν κατάλληλα και ακατάλληλα δειγματοληπτικά πλαίσια για την εξαγωγή μιας δειγματοληπτικής έρευνας. Εάν ένα δειγματοληπτικό πλαίσιο περιέχει ολόκληρο τον υπό δειγματοληψία πληθυσμό είναι **ακατάλληλο**.

Για παράδειγμα, οι τηλεφωνικοί κατάλογοι που περιέχουν λανθασμένες διευθύνσεις κατοικιών λόγω μετακίνησης ή θανάτου ορισμένων ατόμων.

Κατάλληλο, είναι το δειγματοληπτικό πλαίσιο που είναι καλά ενημερωμένο ώστε να μπορεί να χρησιμοποιηθεί σε μια έρευνα.

Για παράδειγμα, οι κατάλογοι φορολογίας φυσικών προσώπων ή τα μητρώα ληξιαρχείων. Η εξασφάλιση ενός κατάλληλου δειγματοληπτικού πλαισίου, είναι το πρώτο βήμα στην διαδικασία επιλογής δείγματος. Αυτό σημαίνει ότι αν δεν υπάρχει κατάλληλο πλαίσιο, θα πρέπει να κατασκευαστεί με τέτοιο τρόπο ώστε να ικανοποιεί τις ανάγκες της έρευνας. (Θεοδώρου Η. Αποστολόπουλου, 1988)

Το πρώτο βήμα στην διαδικασία επιλογής δείγματος είναι η εξασφάλιση ενός **κατάλληλου** καταλόγου – πλαισίου. Αυτό σημαίνει ότι αν δεν υπάρχει κατάλληλο πλαίσιο, θα πρέπει να κατασκευαστεί με τρόπο ώστε να ικανοποιεί τις ανάγκες της έρευνας.

Για να θεωρηθεί κατάλληλος ένας κατάλογος – πλαίσιο θα πρέπει να συγκεντρώνει τις ακόλουθες ιδιότητες:

- Να αποτελείται από τον ίδιο τύπο μονάδων με τον πληθυσμό που θέλουμε να μελετήσουμε και επομένως με το δείγμα που θέλουμε να επιλέξουμε. Στην περίπτωση που δεν ισχύει αυτό, οι μονάδες του καταλόγου θα πρέπει να είναι εύκολα μετατρέψιμες στην επιθυμητή κατεύθυνση.
- Να είναι ενημερωμένος:
 1. Να είναι πλήρης, δηλαδή να περιλαμβάνει όλες τις μονάδες, που μέχρι πρόσφατα ανήκουν στον υπό εξέταση πληθυσμό. Είναι επίσης ανάγκη, οι κατάλογοι να είναι ενημερωμένοι και για τα στοιχεία αναγνώρισης των μονάδων (Διεύθυνση, κλπ.) προκειμένου να είναι δυνατή η εύρεσή τους κατά την έρευνα.

2. Να έχουν αφαιρεθεί, τουλάχιστον, κατά το μεγαλύτερο μέρος τους, οι μονάδες που δεν ανήκουν πλέον στον πληθυσμό της έρευνας. Επίσης, πρέπει να υπάρχει δυνατότητα εντοπισμού και εξάλειψης των διπλοεγγραφών γιατί δημιουργούν μεροληπτικό δείγμα.

Να παρέχουν τη δυνατότητα εφαρμογής της ενδεδειγμένης μεθόδου δειγματοληψίας. Για παράδειγμα, να είναι εύκολη η μετατροπή τους, ώστε να μπορεί να επιλεγεί δείγμα, όχι μόνο από το σύνολο του πληθυσμού με ενιαίο τρόπο, αλλά και κατά στρώματα. Για να εξασφαλιστεί η δυνατότητα αυτή, ο κατάλογος θα πρέπει να βρίσκεται σε ηλεκτρονική μορφή. (Hoinville G. et al, 1983)

2.8 Μέθοδοι συγκέντρωσης στατιστικών δεδομένων

Υπάρχουν δύο ειδών μέθοδοι συγκέντρωσης στατιστικών δεδομένων, οι εξαντλητικές έρευνες και οι δειγματοληπτικές έρευνες:

- η **εξαντλητική έρευνα (exhaustive)** χρησιμοποιείται για την συγκέντρωση στατιστικών δεδομένων από ολόκληρο των πληθυσμό. Ένα είδος εξαντλητικής έρευνας είναι η απογραφή (census).
- η **συνεχής καταγραφή**, η οποία χρησιμοποιείται από διάφορες υπηρεσίες για την καταχώρηση γεγονότων που εμπίπτουν στην αρμοδιότητα τους. Ένα παράδειγμα είναι, η καταγραφή από τις τοπικές αρχές των γεννήσεων, θανάτων, γάμων και άλλων γεγονότων που επηρεάζουν τη φυσική κίνηση του πληθυσμού.

Όπως έχουμε ήδη τονίσει, όμως, στην πράξη συχνά εφαρμόζονται σε δειγματοληπτικές έρευνες (sampling methods), είτε επειδή αυτές κοστίζουν λιγότερο από τις εξαντλητικές είτε επειδή είναι οι μόνες δυνατές. Ένα ερώτημα που προκύπτει σε αυτή την περίπτωση, είναι πόσο αντιπροσωπευτικό είναι το δείγμα και συνεπώς πόσο αξιόπιστα θα είναι τα συμπεράσματά μας για τον πληθυσμό. Είναι ευνόητο ότι όσο μεγαλύτερο είναι το μέγεθος του δείγματος τόσο πιο αξιόπιστα θα είναι τα αποτελέσματα.

Οι δειγματοληπτικές μέθοδοι διακρίνονται:

- Σε τυχαίες (random), όπου η επιλογή του δείγματος γίνεται κατά τρόπο μηχανικό και όπου η πιθανότητα μίας στατιστικής μονάδας είναι γνωστή και μεγαλύτερη από το μηδέν.
- Σε μη πιθανοθεωρητικές (nonprobabilistic) όπου η επιλογή γίνεται με υποκειμενικά κριτήρια, οπότε κάποιες στατιστικές μονάδες μπορεί να αποκλεισθούν και να μην έχουν πιθανότητα να επιλεγούν.

2.9 Απογραφή

Απογραφή γίνεται συνήθως όταν το σύνολο του πληθυσμού είναι τόσο μικρό ώστε να μπορεί να γίνει απογραφή με τους υπάρχοντες προϋπολογισμούς. Σαν εξαίρεση αποτελεί η περίπτωση στην οποία θα πρέπει να γνωρίζουμε με απόλυτη ακρίβεια κάθε χαρακτηριστικό από το σύνολο του πληθυσμού. (W. Cochran, 1997)

Η **απογραφή (census)** συνίσταται στη συγκέντρωση στοιχείων από όλες τις στατιστικές μονάδες του πληθυσμού που επιθυμούμε να μελετήσουμε. Ανάλογα με το είδος του πληθυσμού που απογράφουμε, διακρίνουμε σε απογραφές γεωργίας, κτηνοτροφίας, βιομηχανικών και εμπορικών επιχειρήσεων, απογραφές πληθυσμού κλπ. Από όλες τις μορφές απογραφών, η απογραφή του πληθυσμού είναι η σπουδαιότερη γιατί αποτελεί την κύρια πηγή πληροφοριών από άποψη δημογραφικών, οικονομικών και κοινωνικών χαρακτηριστικών.

Τα σπουδαιότερα χαρακτηριστικά που μελετάμε σε έναν πληθυσμό με την βοήθεια των γενικών απογραφών είναι:

- α. η σύνθεση του πληθυσμού κατά ηλικία
- β. η οικογενειακή κατάσταση (παντρεμένοι, ανύπαντροι, χωρισμένοι, χήροι)
- γ. η σύνθεση κατά φύλο
- δ. η σύνθεση κατά επάγγελμα
- ε. η ανεργία και απασχόληση
- στ. η εκπαίδευση
- ζ. η φυσική κίνηση του πληθυσμού, η μετανάστευση

Την κατάσταση του πληθυσμού με τα διάφορα δημογραφικά χαρακτηριστικά της, σε μία δεδομένη χρονική στιγμή, τη διαπιστώνουμε με περιοδικές δεκαετείς γενικές απογραφές, που μας δίνουν μία φωτογραφία του πληθυσμού σε μία ορισμένη χρονική στιγμή και αποτελούν την κύρια βάση όλων των μετέπειτα δημογραφικών υπολογισμών.

Η απογραφή του πληθυσμού γίνεται από την στατιστική υπηρεσία της κάθε χώρας.

Ως προς την συχνότητα διεξαγωγής των απογραφών, συνήθως διεξάγετε κάθε 10 χρόνια και στα έτη που λήγουν σε 0 ή 1.

Σχετικά με την εποχή διεξαγωγής των γενικών απογραφών, οι απογραφές γίνονται κατά την άνοιξη ή το φθινόπωρο. Η προτίμηση αυτή δεν είναι τυχαία, και αυτό επειδή τις εποχές που αναφέρουμε οι γεωργικές απασχολήσεις του πληθυσμού είναι μεγάλες και έτσι δεν έχουμε πληθυσμιακές μετακινήσεις. Ακόμα τις εποχές αυτές δεν υπάρχει μεγάλη τουριστική κίνηση και αυτό βοηθάει τους υπαλλήλους της στατιστικής υπηρεσίας να είναι άνετοι στις μετακινήσεις τους για να κάνουν την απογραφή. Σχετικά με την ημέρα που προτιμάται να

γίνεται η απογραφή είναι Κυριακή. Η διάρκεια συμπλήρωσης των απογραφικών δελτίων, συνήθως γίνεται σε μία ημέρα, αν όμως δεν επαρκούν οι απογραφείς μπορεί να διαρκέσει και περισσότερες μέρες.

2.9.1 Πλεονεκτήματα της απογραφής

Η μέθοδος της απογραφής έχει σημαντικά πλεονεκτήματα:

- α. Δεν υπόκεινται, όπως θα δούμε, σε δειγματοληπτικό σφάλμα.
- β. Δίνουν την δυνατότητα παραγωγής στατιστικών δεδομένων σε μεγάλη θεματική και γεωγραφική ανάλυση.
- γ. Τέλος, αποτελούν τη βάση για τη διενέργεια άλλων στατιστικών ερευνών, καθώς από την πλήρη καταγραφή των μονάδων ενός πληθυσμού δημιουργούμε δειγματοληπτικό πλαίσιο για τον πληθυσμό αυτό.

2.9.2 Μειονεκτήματα της απογραφής

Η μέθοδος της απογραφής παρουσιάζει τα ακόλουθα μειονεκτήματα:

- α. Απαιτεί μεγάλο χρόνο για την προετοιμασία και διεξαγωγή της.
- β. Απαιτεί μεγάλο κόστος.
- γ. Επειδή ο αριθμός των στατιστικών δεδομένων (νοικοκυριών) και το πλήθος των πληροφοριών είναι μεγάλο, η δημοσίευση των αποτελεσμάτων καθυστερεί και σχεδόν πάντοτε στη χώρα μας τα στοιχεία που δημοσιεύονται περιορίζονται σε ένα δείγμα 5%, ή σπάνια το πολύ 25% επί του συνόλου των στοιχείων που συγκεντρώθηκαν κατά την απογραφή.

2.10 Απογραφές του πληθυσμού στην Ελλάδα

Η πρώτη απογραφή του πληθυσμού έγινε το 1828 και έδειξε ότι ο πληθυσμός της χώρας μας κατά την εποχή εκείνη ήταν 753.400 άτομα. Με βάση τον αριθμό αυτό, υπολογίστηκε ότι ο πληθυσμός της χώρας κατά το 1821 ήταν 938.765 άτομα.

Κατά τα έτη, από το 1836 μέχρι το 1845 έγινε ονομαστική καταγραφή των κατοίκων σε βιβλίο κατά φύλλο και ηλικία.

Κατά τα έτη 1848, 1853, 1856 έγινε καταγραφή των κατοίκων όπως και στα παραπάνω έτη.

Στις 12 Μαρτίου 1861 χρησιμοποιήθηκε δελτίο απογραφής μόνο για τους κατοίκους Αθηνών και Πειραιώς.

Στο χρονικό διάστημα 2-17 Μαρτίου 1870 έγινε απογραφή με δελτίο.

Στο χρονικό διάστημα 15-21 Απριλίου 1878 έγινε απογραφή με οικογενειακό δελτίο.

Στις 16 Απριλίου 1889 επίσης με οικογενειακό δελτίο.

Στις 6 Οκτωβρίου 1896 γενική χρήση δελτίου.

Στις 27 Οκτωβρίου 1907 με ατομικό δελτίο.

Στις 19 Δεκεμβρίου 1920 με ατομικό και οικογενειακό δελτίο.

Στις 15 Μαΐου 1928 με ατομικό και οικογενειακό δελτίο.

Στις 16 Οκτωβρίου 1940 κατά τον ίδιο τρόπο.

Στις 8 Απριλίου 1951 απογραφή με ερωτηματολόγιο με διάφορα δημογραφικά, οικονομικά και κοινωνικά χαρακτηριστικά του πληθυσμού.

Στη συνέχεια, συνεχίζονται οι απογραφές κανονικά και βελτιωμένες ανά δεκαετία. 1961, 1971, 1981, 1991, 2001.

Έτος	Πληθυσμός	Συνολική Μεταβολή
1821	938.765	
1828	753.400	753.400
1840	850.246	96.846
1853	1.035.527	185.281
1861	1.096.810	61.283
1870	1.457.894	361.084
1879	1.679.470	221.576
1889	2.187.208	507.738
1896	2.433.806	246.598
1907	2.631.952	198.146
1920	5.531.474	2.899.522
1928	6.204.684	673.210
1940	7.344.860	1.140.176
1951	7.632.801	287.941
1961	8.388.553	755.752
1971	8.768.641	380.088
1981	9.740.417	971.776
1991	10.259.900	519.483
2001	10.964.200	704.120

Πίνακας 1 - Απογραφές πληθυσμού στην Ελλάδα

2.11 Μερική απογραφή

Όταν μόνο ένα τμήμα του πληθυσμού απαριθμείται και μελετάται, τότε πρόκειται για μερική απογραφή. Για παράδειγμα οι γεννήσεις και οι θάνατοι σε ένα γεωγραφικό διαμέρισμα, η μετανάστευση σε ορισμένες μόνο περιοχές, η ανεργία των νέων σε ορισμένο γεωγραφικό διαμέρισμα. Με την παραπάνω έννοια οι μερικές απογραφές διαφέρουν από τις δειγματοληπτικές έρευνες.

2.12 Επιλογή της απογραφικής ή της δειγματοληπτικής έρευνας

Μια βασική επιλογή του ερευνητή είναι να αποφασίσει για την έκταση των μονάδων του πληθυσμού που θα ερευνηθεί. Ουσιαστικά έχει δύο επιλογές. Την απογραφική ή την δειγματοληπτική έρευνα.

Η απογραφή απαιτεί τη καταγραφή των χαρακτηριστικών ιδιοτήτων όλων των μονάδων του εκάστοτε εξεταζόμενου πληθυσμού. Αντίθετα, οι δειγματοληπτικές έρευνες αποτελούν στατιστικές εργασίες, στις οποίες η συλλογή των δεδομένων περιορίζεται σε ένα αντιπροσωπευτικό δείγμα του εξεταζόμενου πληθυσμού. Η στατιστική πληροφόρηση που αφορά στα χαρακτηριστικά ολόκληρου του πληθυσμού προκύπτουν κατόπιν επεξεργασίας και αξιολόγησης του εν λόγω δείγματος, που γίνεται με ειδικές στατιστικές μεθόδους. Με την διαδικασία αυτή θα ασχοληθούμε στις επόμενες ενότητες του κεφαλαίου αυτού.

Οι απογραφές έχουν σημαντικά πλεονεκτήματα και μειονεκτήματα τα οποία αναφέραμε προηγουμένως. Όπως είδαμε, οι απογραφές, δεν υπόκεινται σε δειγματοληπτικό σφάλμα.

Είναι αυτονόητο, ότι μια τέτοια απογραφή που αναφέρεται σε έναν εθνικό πληθυσμό δεν είναι δυνατόν να διενεργείται συχνά. Μια απογραφή λόγω του κόστους της συλλέγει μια μεγάλη γκάμα στοιχείων για τα χαρακτηριστικά του πληθυσμού που μελετά, ενώ για ειδικότερα θέματα χρησιμοποιείται η δειγματοληπτική έρευνα που ερευνά σε βάθος.

Η δειγματοληπτική έρευνα υπόκειται, βέβαια, σε δειγματοληπτικό σφάλμα, αλλά υπάρχουν όπως θα δούμε μέθοδοι περιορισμού του. Σε οποιαδήποτε δε περίπτωση το σφάλμα αυτό είναι μετρήσιμο όπως θα αναλυθεί παρακάτω.

Επίσης, πρέπει να σημειώσουμε ότι στην απογραφή, λόγω της έκτασής της και της εμπλοκής σε αυτήν πολλών ατόμων, αυξάνεται η συχνότητα μη δειγματοληπτικών λαθών.

Τέλος, η απογραφική μέθοδος είναι αδύνατη όταν ο πληθυσμός είναι άπειρος ή όταν με την επιλογή μιας στατιστικής μονάδας αυτή καταστρέφεται.

2.13 Προσδιορισμός του μεγέθους του δείγματος

Ο προσδιορισμός του μεγέθους του δείγματος είναι το δεύτερο στάδιο μετά την εξασφάλιση του καταλόγου – πλαισίου. Το μέγεθος του δείγματος, γενικά, προσδιορίζεται από τους ακόλουθους παράγοντες:

- Την επιθυμητή ακρίβεια της έρευνας (το επιθυμητό μέγεθος του δειγματοληπτικού σφάλματος).
- Το διαθέσιμο κόστος και χρόνο.

- Την ομοιογένεια (το μέγεθος της διακύμανσης) του πληθυσμού ως προς το διερευνώμενο χαρακτηριστικό.
- Τον αριθμό των μεταβλητών και των υποομάδων αυτών που πρέπει να ερευνηθούν.
- Την επιλεγμένη μέθοδο δειγματοληψίας.

Το μέγεθος του δειγματοληπτικού σφάλματος μειώνεται (η ακρίβεια της έρευνας αυξάνει) με την αύξηση του δείγματος. Η αύξηση όμως, δεν είναι αναλογική. Η αύξηση του δείγματος συνεπάγεται αύξηση του κόστους και του απαιτούμενου χρόνου διενέργειας της έρευνας. Εκτιμάται ότι το κόστος της έρευνας αυξάνει σχεδόν αναλογικά με την αύξηση του δείγματος. Όσο πιο ομοιογενής είναι ένας πληθυσμός, δηλαδή όσο μικρότερη διακύμανση παρουσιάζει ως προς το διερευνώμενο χαρακτηριστικό αυτού, τόσο μικρότερο δείγμα απαιτείται.

Για να γίνει κατανοητή η σχέση της ομοιογένειας του πληθυσμού με το μέγεθος του δείγματος παραθέτουμε το ακόλουθο παράδειγμα:

Αν θέλουμε να εκτιμήσουμε το μέσο ύψος των παιδιών ηλικίας 12–14 ετών μιας γειτονιάς, απαιτούνται λίγες σχετικά μονάδες δείγματος γιατί η διακύμανση του ύψους των ίδιων περίπου ηλικιών και μάλιστα ίδιας κατά τεκμήριο κοινωνικοοικονομικής προέλευσης (που εξασφαλίζεται από την κατοικία στην ίδια περιοχή) είναι σχετικά μικρή, δηλαδή ο προς μελέτη πληθυσμός είναι σχετικά ομοιογενής.

Αν όμως ο πληθυσμός που πρόκειται να διερευνήσουμε περιλαμβάνει τα παιδιά ηλικίας 7–14 ετών μιας πόλης, το δείγμα πρέπει να είναι (ως ποσοστό επί του πληθυσμού) αυξημένο, σε σχέση με την προηγούμενη περίπτωση, γιατί οι διαφορετικές ηλικίες και η διαφορετική προέλευση των παιδιών δημιουργούν μεγαλύτερη απόκλιση (διακύμανση) από το μέσο επίπεδο του ύψους των παιδιών, δηλαδή ο πληθυσμός είναι λιγότερο ομοιογενής.

Μεγάλος αριθμός προς διερεύνηση μεταβλητών και μεγάλη διάκριση αυτών σε υποομάδες (ομάδες ηλικιών, γεωγραφικές κατανομές) που αποτελούν την συνήθη πρακτική στις στατιστικές έρευνες, δημιουργεί δυσχέρειες στον προσδιορισμό του μεγέθους του δείγματος και απαιτούν σε γενικές γραμμές μεγαλύτερο δείγμα.

Ο παράγοντας αυτός αναφέρεται με τον όρο «βαθμός ανάλυσης» εννοείται μ' αυτόν η έκταση στην οποία προβλέπεται να παράγουμε στοιχεία για τις διάφορες υποομάδες των μεταβλητών που εξετάζουμε.

Αν στους σκοπούς μιας έρευνας είναι να διερευνηθεί, όχι μόνο ο συνολικός πληθυσμός της χώρας αλλά και τα χαρακτηριστικά του κατά ηλικία, φύλο, επίπεδο εκπαίδευσης, συνθήκες απασχόλησης, είναι επόμενο να απαιτείται αυξημένο δείγμα, με τρόπο ώστε να επιλεγούν μονάδες σε κάθε μία από τις υποομάδες (για παράδειγμα άνδρες και γυναίκες αναφορικά με το φύλο).

Σε πολύ γενικές γραμμές η μικρότερη υποδιαίρεση μιας μεταβλητής θα πρέπει να περιλαμβάνει τουλάχιστον 50–100 μονάδες για να θεωρηθεί το δείγμα ικανό να δώσει αξιόπιστα αποτελέσματα.

Τέλος, η επιλογή της κατάλληλης μεθόδου δειγματοληψίας είναι δυνατόν να μειώσει το απαιτούμενο μέγεθος του δείγματος, όπως θα δούμε στη παρουσίαση των μεθόδων επιλογής δείγματος.

Πριν από αυτό και στη κατεύθυνση αξιολόγησης των παραπάνω προσδιοριστικών παραγόντων του μεγέθους του δείγματος και λαμβάνοντας υπόψη την πρακτική πλευρά του θέματος, έχουμε να παρατηρήσουμε τα εξής:

Η επιθυμητή ακρίβεια της κάθε έρευνας είναι δύσκολο να προσδιοριστεί από τους ερευνητές (Hoinville G., 1983) ή είναι ανέφικτο να επιτευχθεί λόγω οικονομικών περιορισμών. (Χαρίσης Κ. 1992)

Ακόμη, ο βαθμός ομοιογένειας είναι δύσκολο να αποτελέσει κριτήριο προσδιορισμού του μεγέθους του δείγματος όταν διερευνώνται πολλά χαρακτηριστικά, στα οποία ο πληθυσμός παρουσιάζει διαφορετικό βαθμό διακύμανσης.

Στην πράξη οι βασικοί παράγοντες που πρέπει να λαμβάνονται υπόψη είναι ο απαιτούμενος βαθμός ανάλυσης των δεδομένων και οι οικονομικοί και χρονικοί περιορισμοί.

Κεφάλαιο: 3 Δειγματοληπτική έρευνα

3.1 Εισαγωγή

Η δειγματοληπτική έρευνα είναι μια στατιστική εργασία κατά την οποία η συλλογή των δεδομένων **περιορίζεται** σε ένα **αντιπροσωπευτικό δείγμα** του ερευνώμενου πληθυσμού. Η στατιστική πληροφόρηση που αφορά χαρακτηριστικά ολόκληρου του πληθυσμού προκύπτουν κατόπιν επεξεργασίας και αξιολόγησης του εν λόγω δείγματος, που γίνεται με ειδικές στατιστικές μεθόδους. Με την διαδικασία αυτή θα ασχοληθούμε στη συνέχεια.

3.2 Ορισμός και ανάλυση

Η **δειγματοληπτική έρευνα (sample survey)** είναι η έρευνα που βασίζεται στο δείγμα με σκοπό να προκύψουν συμπεράσματα για τον πληθυσμό. (W. Cochran, 1997).

Η **δειγματοληπτική έρευνα** υπόκειται, βέβαια, σε δειγματοληπτικό σφάλμα, αλλά υπάρχουν όπως θα δούμε μέθοδοι περιορισμού του σφάλματος αυτού. Σε οποιαδήποτε δε περίπτωση το σφάλμα αυτό είναι μετρήσιμο όπως θα αναλυθεί παρακάτω. Η απογραφική μέθοδος είναι αδύνατη όταν ο πληθυσμός είναι άπειρος ή όταν με την επιλογή μιας στατιστικής μονάδας αυτή καταστρέφεται.

Στη συνέχεια, θα αναφερθούν μερικά παραδείγματα δειγματοληπτικών ερευνών, για να γίνει κατανοητή αυτή η διαφορά:

- Το Υπουργείο Γεωργίας των Η.Π.Α πήρε ένα δείγμα φτωχών πολιτών για να δει πως χρησιμοποιούν τα κουπόνια διατροφής τους.
- Το γραφείο Απογραφής των Η.Π.Α ένα δείγμα πολιτών για να μελετήσει την απασχόληση και την ανεργία στη χώρα.
- Το Συμβούλιο Ανώτερης Εκπαίδευσης του Ιλλινόις πήρε ένα δείγμα κατοίκων για να μελετήσει το ενδιαφέρον των κατοίκων για την εκπαίδευση ενηλίκων.
- Το Εθνικό Κέντρο Στατιστικής για την Υγεία της Ελλάδας, μλέτησε ένα δείγμα πολιτών ώστε να εξετάσει πόσα χρήματα ξοδεύουν οι πολίτες για είδη ιατρικής περίθαλψης.

Η δειγματοληπτική έρευνα, λοιπόν, βασίζεται στο δείγμα για να εξάγει συμπεράσματα για τον πληθυσμό. Δειγματοληπτικές έρευνες έχουν πραγματοποιηθεί σε όλες τις χώρες του κόσμου, πάνω σε διάφορα θέματα που αφορούν τον άνθρωπο, όπως:

- i. Οι γεννήσεις και οι θάνατοι
- ii. Η υγεία και η περίθαλψη
- iii. Η εκπαίδευση
- iv. Το περιβάλλον
- v. Ο πλούτος
- vi. Το εμπόριο και η βιομηχανία
- vii Το έγκλημα (W. Cochran, 1997)

3.3 Διαδικασία Δειγματοληπτικής Έρευνας

Όταν θα γίνει μια δειγματοληπτική έρευνα, χρειάζονται δυο αποφάσεις, κυρίως τεχνικής φύσεως και προκαταρκτικές ενέργειες. Για τον λόγο αυτό, είναι αναγκαία η συνεργασία ενός στατιστικού, με μεγάλη εμπειρία στη διεξαγωγή ερευνών, με ειδικούς επιστήμονες που είναι σχετική με το εξεταζόμενο πρόβλημα. Το σύνολο αυτών των επιστημών αποτελεί την λεγόμενη ομάδα εργασίας.

Αυτή είναι υπεύθυνη για το σχεδιασμό, την εκτέλεση και ανάλυση των αποτελεσμάτων της έρευνας. της έρευνας να προσδιορίσουν τη φύση του προς εξέταση του προβλήματος και τη διατύπωση των αντικειμενικών σκοπών της. Η ομάδα εργασίας θα πρέπει να μελετήσει παρόμοιες έρευνες που έχουν γίνει στο παρελθόν, τόσο στο εσωτερικό της χώρας όσο και στο εξωτερικό.

Ακόμα, ανάλογα με το μέγεθος του δείγματος, θα πρέπει να γίνει προϋπολογισμός των δαπανών στα διάφορα στάδια και ένα λεπτομερές ημερολογιακό πρόγραμμα εκτέλεσης των διάφορων σταδίων της έρευνας από τον αρχικό σχεδιασμό μέχρι τη δημοσίευση των αποτελεσμάτων της έρευνας.

Στο σχεδιασμό της δειγματοληπτικής έρευνας βασική θέση έχουν η κατάρτιση του πλαισίου, το ερωτηματολόγιο, το μέγεθος του δείγματος και η μέθοδος δειγματοληψίας. Το παρακάτω διάγραμμα θα μας βοηθήσει να κατανοήσουμε καλύτερα την διαδικασία της δειγματοληπτικής έρευνας .



Διάγραμμα 1 - Η Διαδικασία Δειγματοληπτικής Έρευνας

3.3.1 Κατάρτιση πλαισίου δειγματοληψίας

Για την πραγματοποίηση μιας δειγματοληψίας είναι απαραίτητος ένας κατάλογος που θα περιέχει όλες τις μονάδες που αποτελούν τον προς εξέταση πληθυσμό. Ο κατάλογος αυτός ονομάζεται πλαίσιο, χρησιμεύει για την επιλογή του δείγματος.

Το πλαίσιο για να χρησιμοποιηθεί θα πρέπει να πληρεί ορισμένους όρους. Όπως, να είναι ενημερωμένο, να μην έχει παραλείψεις, διπλές εγγραφές. Οι χρησιμοποιούμενες συνήθως δειγματοληπτικές μονάδες διακρίνονται στις παρακάτω κατηγορίες:

α. Γεωγραφικές και διοικητικές διαιρέσεις

Τέτοιες είναι οι επαρχίες, οι δήμοι, οι κοινότητες, τα χωριά, οι ενορίες, τα οικοδομικά τετράγωνα. Αφού επιλεγεί ένα δείγμα μονάδων για ένα από τα παραπάνω είδη, θα εξαρτηθεί μετά από το σχέδιο του δείγματος που έχει αποφασιστεί, αν ερευνηθούν όλες οι μονάδες (άτομα, οικογένειες) οι οποίες περιλαμβάνονται σε κάθε επιλεγείσα μονάδα δειγματοληψίας, ή αν θα πάρουμε ένα δείγμα από αυτές σε ένα δεύτερο στάδιο δειγματοληψίας.

β. Διάφορες κοινωνικές ομάδες

Κοινωνικές ομάδες είναι οργανισμοί, σύλλογοι, σχολεία κ.λ.π. Οι μονάδες αυτές μπορούν να χρησιμοποιηθούν για να ερευνηθούν ειδικοί πληθυσμοί (μαθητές, εργάτες) και όχι ο γενικός πληθυσμός της χώρας.

γ. Νοικοκυριά, κατοικία

Οι μονάδες αυτές μπορούν να αποτελούν συγχρόνως μονάδες δειγματοληψίας. Και μονάδες ενδιαφέροντος, όπως οι έρευνες των οικογενειακών προϋπολογισμών ή οι έρευνες συνθηκών διαβίωσης.

δ. Το άτομο

Αν και αυτό αποτελεί τη συνηθισμένη μονάδα έρευνας, σπάνια χρησιμοποιείται ως δειγματοληπτική μονάδα, γιατί δεν υπάρχουν ολοκληρωμένοι κατάλογοι (πλαίσια) που να περιέχουν ολόκληρο τον πληθυσμό των ατόμων, εκτός και αν πρόκειται για πληθυσμό περιορισμένης έκτασης. Για το λόγο αυτό, στις περισσότερες περιπτώσεις γίνεται χρήση μιας από τις παραπάνω μονάδες δειγματοληψίας και ερευνώνται άτομα (όλα ή δείγμα σε δεύτερο στάδιο)τα οποία απαρτίζουν το δείγμα. Ένας παράγοντας που είναι αποφασιστικός για την εκλογή της δειγματοληπτικής μονάδας είναι η ύπαρξη κατάλληλου πλαισίου.

3.3.2 Επεξεργασία Δειγματοληπτικών στατιστικών στοιχείων

Η Επεξεργασία Δειγματοληπτικών στατιστικών στοιχείων θα γίνει αφού συγκεντρωθούν τα στατιστικά στοιχεία που θέλουμε τα οποία βρίσκονται σε πρωτογενή μορφή, δηλαδή ερωτηματολόγια στα οποία έχουν δοθεί οι σχετικές απαντήσεις από τους ερευνώμενους, ακολουθεί το στάδιο της επεξεργασίας. Περιλαμβάνει τον έλεγχο όλων των ερωτηματολογίων, για τυχόν ασαφείς, δυσανάγνωστες ή ασυμπλήρωτες απαντήσεις.

Στη συνέχεια θα γίνει η διαλογή των πληροφοριών των διαφόρων χαρακτηριστικών των στατιστικών μονάδων και η εμφάνισή τους, σε κατάλληλους αριθμητικούς πίνακες, με βάση διάφορα κριτήρια.

Η διαλογή μπορεί να γίνει είτε με το χέρι είτε με μηχανικά μέσα ανάλογα με τον αριθμό των ερωτηματολογίων και τον αριθμό των χαρακτηριστικών που πρόκειται να μελετήσουμε. Η διαλογή των στοιχείων γίνεται με το χέρι όταν ο αριθμός των ερωτηματολογίων και των χαρακτηριστικών είναι περιορισμένος.

Οι εργασίες για την επεξεργασία των στοιχείων με μηχανογραφικά μέσα (π.χ δημογραφικά προβλήματα) εκτελούνται σε 4 στάδια:

1. **Η κωδικογράφηση**: τα στοιχεία μετατρέπονται σε κωδικούς αριθμούς.
2. **Η διατήρηση**: οι κωδικοί μεταφέρονται σε ειδικές καρτέλες με τη βοήθεια της διατήρησης και ακολουθεί η επαλήθευση με τις επαληθευτικές μηχανές για τυχόν λάθη.
3. **Η διαλογή**: τα δελτία μεταφέρονται στις διαλογικές μηχανές όπου χωρίζονται και ταξινομούνται σε ομάδες, αριθμούνται και τέλος τοποθετούνται χωριστά σε 12 υποδοχές κάθε μηχανής.
4. **Η πινακογράφηση**: εδώ η πινακογραφική μηχανή μετρά τα δελτία κάθε κατηγορίας, διαβάζει τις διατρήσεις, κάνει ορισμένους υπολογισμούς πάνω σε αυτά τα δεδομένα και εκτυπώνει τα αποτελέσματα των υπολογισμών.

3.3.3 Παρουσίαση Δειγματοληπτικών Στοιχείων σε Μορφή Πινάκων

Μετά από τα στάδια συλλογή και επεξεργασία δεδομένων, επακολουθεί το στάδιο παρουσίασης των δεδομένων. Τα δεδομένα θα παρουσιαστούν με τη μορφή στατιστικού πίνακα, αλλά πρώτα θα αναφέρουμε τις κατηγορίες ενός πίνακα, τη μορφή που μπορεί να έχει και τη τεχνική κατασκευής του.

Η παρουσίαση του στατιστικού υλικού με τη μορφή πίνακα έχει αποδειχθεί ένας πιο σωστός τρόπος να συλλέξουμε τα δεδομένα σε αντίθεση με την ενσωμάτωση στατιστικών δεδομένων

στο κείμενο εκθέσεων, μελετών και άρθρων του ημερήσιου και περιοδικού τύπου ο οποίος είναι πιο εξεζητημένος.

Τα αριθμητικά δεδομένα κατατάσσονται σε στήλες και γραμμές και η παρουσίαση τους πρέπει να γίνεται με τέτοιο τρόπο, ώστε να επιτυγχάνεται η συνοπτική εμφάνιση των δεδομένων, η οποία διευκολύνει τη σύγκριση μεταξύ αυτών και την εύκολη ανάγνωση και κατανόηση τους. Οι στατιστικοί πίνακες διακρίνονται σε δύο κατηγορίες:

1. στους γενικούς
2. στους ειδικούς πίνακες

Οι **γενικοί** πίνακες περιέχουν πρωτογενές στατιστικό υλικό, γι 'αυτό και αποτελούν κύρια πηγή αντλήσεως πληροφοριών. Έτσι τοποθετούνται στα παραρτήματα μελετών ή στατιστικών δελτίων.

Οι **ειδικοί** πίνακες είναι συνοπτικοί και τα δεδομένα τους αντλούνται από τους γενικούς πίνακες, Τα χαρακτηριστικά των στατιστικών μεγεθών που ερευνούμε αποτυπώνονται πιο εύκολα στα κείμενα εκθέσεων και μελετών, αν χρησιμοποιηθούν οι ειδικοί πίνακες.

3.3.3.1 Τεχνική κατασκευής στατιστικών πινάκων

Υπάρχουν κάποιοι κανόνες για την κατασκευή των στατιστικών πινάκων οι οποίοι αποβλέπουν στην γρήγορη και εύκολη ενημέρωση του αναγνώστη.

Ένας σωστά κατασκευασμένος πίνακας θα πρέπει να περιέχει τα εξής:

1) Τίτλο

Ένας σαφής τίτλος κατατοπίζει τον αναγνώστη για το περιεχόμενο του πίνακα. Αν σε μια μελέτη υπάρχουν πολλοί πίνακες, πρέπει να έχουν αύξοντα αριθμό, ο οποίος γράφεται πάνω από τον τίτλο.

2) Επικεφαλίδες στηλών – γραμμών

Στη κορυφή κάθε στήλης ή στην αρχή κάθε γραμμής πρέπει να γράφεται η φύση των στατιστικών δεδομένων που περιέχει η στήλη ή η γραμμή και η μονάδα μέτρησης των δεδομένων.

3) Κύριο σώμα (κορμός)

Το κύριο σώμα του πίνακα περιέχει τα στατιστικά δεδομένα, τα οποία μπορεί να είναι ποσοτικής ή ποιοτικής φύσεως.

4) Σύνολα στηλών – γραμμών

Αν είναι επιτρεπτή η άθροιση των δεδομένων των στηλών και των γραμμών, τα σύνολα των στηλών γράφονται στο κάτω μέρος του πίνακα και τα σύνολα των γραμμών στα δεξιά του πίνακα.

5) Υποσημειώσεις

Στο κάτω μέρος του πίνακα γράφονται τυχόν υποσημειώσεις, οι οποίες περιέχουν επεξηγήσεις για όλα τα δεδομένα του πίνακα, ακόμα και για αριθμούς που μπορεί να μεταβάλλονται με το χρόνο και για ποιο λόγο γίνεται αυτό. Οι υποσημειώσεις είναι καλό να είναι αριθμημένες.

3.3.3.2 Τύποι στατιστικών πινάκων

Οι στατιστικοί πίνακες, όπως είπαμε διακρίνονται σε γενικούς και ειδικούς. Παρακάτω θα αναλύσουμε τους τύπους των ειδικών πινάκων.

Είναι οι **πίνακες απλής εισόδου**, στους οποίους γίνεται παρουσίαση στατιστικών δεδομένων που εξετάζονται με βάση ένα ποσοτικό ή ποιοτικό χαρακτηριστικό του ερευνώμενου δημογραφικού ή οικονομικού μεγέθους.

Έτος	Πληθυσμός(σε εκατομ.)
1950	2.504
1960	3.014
1965	3.324
1970	3.683
1975	4.076
1980	4.453
1983	4.685
1984	4.763

Πίνακας 1 - Εξέλιξη του πληθυσμού της Γής 1950 – 1984

Ήπειροι	Πληθυσμός (σε εκατομ)	Ποσοστόις %
Αφρική	537	11.3
Αμερική	658	13.8
Ασία	2.777	58.3
Ευρώπη	490	10.3
Ωκεανία	25	0.5
ΕΣΣΔ	276	5.8
Σύνολο	4.763	100.0

Πίνακας 2 - Πληθυσμός της Γής κατά Ηπείρους έτους 1984

Πηγή : United Nations, “Demographic Yearbook” 1984.

Οι πίνακες διπλής εισόδου προκύπτουν από τον συνδυασμό δύο ή περισσότερων πινάκων απλής εισόδου και παρέχουν πληροφορίες για έναν πληθυσμό (δείγμα) του οποίου τα στοιχεία εξετάζονται συγχρόνως ως προς δύο ποσοτικά ή ποιοτικά χαρακτηριστικά.

Η διαδικασία για να προκύψει ένας πίνακας διπλής εισόδου, είναι η εξής: η ταξινόμηση των αριθμητικών δεδομένων, με βάση δύο κριτήρια κατατάξεων, γίνεται με αναγραφή των χαρακτηριστικών ιδιοτήτων της μιας μεταβλητής στις στήλες του πίνακα και των χαρακτηριστικών ιδιοτήτων της άλλης μεταβλητής στις γραμμές του πίνακα.

3.3.4 Το ερωτηματολόγιο

Το ερωτηματολόγιο είναι ένα έντυπο στο οποίο καταχωρούνται οι λαμβανόμενες πληροφορίες από τις ερευνώμενες δειγματοληπτικές μονάδες. Αποτελεί το μέσον επικοινωνίας (interface) μεταξύ του ερευνητή και των ερωτώμενων, με άμεσο ή έμμεσο τρόπο, ανάλογα με τη μέθοδο συλλογής των δεδομένων. Η κατάρτιση του ερωτηματολογίου, λόγω των ιδιοτήτων που έχει, αποτελεί την πλέον κρίσιμη και λεπτή εργασία, καθοριστικής σημασίας για την επιτυχία μιας στατιστικής έρευνας. Το ερωτηματολόγιο πρέπει να συνταχτεί με μεγάλη προσοχή από το τεχνικό προσωπικό της έρευνας.

3.3.4.1 Κατάρτιση ερωτήσεων του ερωτηματολογίου

Για την κατάρτιση του κατάλληλου ερωτηματολογίου θα πρέπει να έχουν προηγηθεί οι ακόλουθες ενέργειες:

- **Προσδιορισμός και εξειδίκευση του στόχου της έρευνας**

Όπως αναφέρθηκε και παραπάνω, ο προσδιορισμός και η εξειδίκευση του στόχου της έρευνας είναι οι έννοιες που πρέπει να καθοριστούν πρώτες. Καθώς πάνω σε αυτές θα στηριχθεί η διαμόρφωση των ερωτήσεων. (Ρόντος - Παπάνης, 2007)

- **Επιλογή της μεθόδου συλλογής των δεδομένων**

Η επιλογή της μεθόδου συλλογής είναι επίσης καθοριστικής σημασίας για τον σχεδιασμό του κατάλληλου ερωτηματολογίου. Για παράδειγμα, αν το ερωτηματολόγιο πρόκειται να συμπληρωθεί από τον ίδιο τον ερωτώμενο, χωρίς την παρέμβαση ερευνητή, τότε πρέπει να δοθεί έμφαση στην τεχνική αρτιότητα του ερωτηματολογίου και στη διευκρίνιση, μέσω εγγράφων επεξηγήσεων στο ερωτηματολόγιο, των σημείων που είναι δυνατόν να παρερμηνευτούν. Αν όμως το ερωτηματολόγιο συμπληρωθεί από τον ερευνητή, κατά τη διάρκεια συνεντεύξεως, τότε το ερωτηματολόγιο θα είναι πλήρως κατανοητό και εύχρηστο, αφού θα μπορεί να επεξηγηθεί, σε τυχόν απορίες, από τον ίδιο τον ερευνητή.

- **Κατανόηση των χαρακτηριστικών των ερωτώμενων.**

Τα χαρακτηριστικά του ερωτώμενου πληθυσμού πρέπει να λαμβάνονται υπόψη στο σχεδιασμό των ερωτηματολογίων, στη διατύπωση των ερωτήσεων και στη χρησιμοποίηση των κατάλληλων λέξεων.

3.3.4.2 Δημιουργία ερωτηματολογίου

Για τη δημιουργία ενός ερωτηματολογίου πρέπει να λάβουμε υπ όψιν μας ορισμένα χαρακτηριστικά ώστε να αποτελεί μια επιτυχημένη και ορθή έρευνα:

- 1) **Πληρότητα**

Η **πληρότητα** αναφέρεται ακριβώς στην ανάγκη κάλυψης όλων των πτυχών του ερευνώμενου χαρακτηριστικού, για το οποίο έχει ήδη γίνει αρκετός λόγος.

2) Σαφήνεια

Η **σαφήνεια** δεν αναφέρεται μόνο στο περιεχόμενο των πληροφοριών αλλά και στο άτομο το οποίο πρέπει να δώσει τις απαντήσεις.

3) Συνοχή

Η **συνοχή** αναφέρεται στην ανάγκη οργανικής σύνδεσης των επιμέρους ερωτημάτων μεταξύ τους. Συγγενή ερωτήματα πρέπει να εμφανίζονται στο ερωτηματολόγιο ομαδοποιημένα και να ερωτώνται μαζί, προκειμένου η σκέψη και η μνήμη του ερωτώμενου να κατευθύνεται ευκολότερα στις σωστές απαντήσεις.

4) Κατάλληλη δομή

Η **κατάλληλη δομή** του ερωτηματολογίου, δηλαδή η σειρά με την οποία θα τεθούν οι ομάδες ερωτήσεων, είναι επίσης μεγάλης σημασίας στην αύξηση του βαθμού ανταπόκρισης του κοινού. Είναι αυτονόητο, αλλά δεν εφαρμόζεται πάντοτε, ότι προσωπικές ή γενικότερα ερωτήσεις που δεν απαντά εύκολα το κοινό (εισόδημα, ύπαρξη διαζυγίου) δεν τίθενται στην αρχή ενός ερωτηματολογίου.

5) Να περιλαμβάνει ερωτήματα ελέγχου

Τα **ερωτήματα ελέγχου** υπάρχουν για τον έλεγχο της ορθότητας των απαντήσεων σε βασικές ερωτήσεις.

6) Να είναι κατά το δυνατόν σύντομο

Ένα αποτελεσματικό ερωτηματολόγιο πρέπει επίσης να είναι και **σύντομο**. Ερωτηματολόγια τα οποία επεκτείνονται σε μεγάλο αριθμό ερωτημάτων κουράζουν τον ερωτώμενο ή του δημιουργούν την αίσθηση ότι θα χάσει πολύ χρόνο και είναι δυνατόν να μην απαντηθούν. Αυτό ισχύει σε μεγαλύτερο βαθμό όταν το ερωτηματολόγιο πρόκειται να συμπληρωθεί από το ίδιο το κοινό.

7) Να έχει τελειότητα παρουσίασης από τεχνικής πλευράς

Η **αρτιότητα εμφάνισης** του ερωτηματολογίου από **τεχνική άποψη** επηρεάζει, επίσης, σημαντικά το βαθμό ανταποκρίσεως του κοινού για δύο κυρίως λόγους:

α. Η ποιότητα του χαρτιού, της εκτύπωσης, κλπ., δημιουργεί ευνοϊκή προδιάθεση για τη σοβαρότητα της έρευνας, με αποτέλεσμα να δείχνουν μεγαλύτερο ενδιαφέρον και να καταβάλλουν μεγαλύτερη προσπάθεια για πλήρεις και σωστές απαντήσεις.

β. Η χρήση δύο ή περισσότερων χρωμάτων, η χρήση κατευθυντήριων τόξων και άλλων συμβόλων, καθοδηγούν τον ερευνητή ή τον ερευνώμενο και διευκολύνουν το έργο του.

8) Να περιλαμβάνει βασικές οδηγίες συμπλήρωσης και εννοιολογικές επεξηγήσεις

Προς την κατεύθυνση αύξησης του βαθμού ανταπόκρισης και υποβοήθησης των ερευνητών ή των ερωτώμενων για όσο το δυνατόν ορθότερες απαντήσεις, συνηθίζεται στο ερωτηματολόγιο να συμπεριλαμβάνονται βασικές σύντομες **οδηγίες** για τον τρόπο συμπλήρωσης των ερωτημάτων ως και βασικές έννοιες και ορισμοί για θέματα που ερωτώνται.

9) Να επιδέχεται κωδικογραφική και μηχανογραφική επεξεργασία

Τέλος, σε κάθε ερωτηματολόγιο θα πρέπει να έχουν προβλεφθεί ειδικοί χώροι σε κάθε ανοιχτή ερώτηση για τη **κωδικογράφηση** της κάθε απάντησης, με τρόπο ώστε να είναι εφικτή η εισαγωγή της, υπό μορφή αριθμού, στον Ηλεκτρονικό Υπολογιστή για περαιτέρω **επεξεργασία**. Επίσης το ερωτηματολόγιο θα πρέπει να έχει ειδικό σχεδιασμό αν πρόκειται να εφαρμοστούν ειδικές μέθοδοι εισαγωγής των δεδομένων στον Η/Υ (μέθοδοι οπτικής ανάγνωσης, κλπ). (Ρόντος - Παπάνης, 2007)

3.3.4.3 Τύποι ερωτήσεων

Ο ερευνητής έχει την δυνατότητα επιλογής μεταξύ διαφορετικών τύπων ερωτήσεων. Παρόλα αυτά η επιλογή δεν είναι τυχαία. Κάθε τύπος ανταποκρίνεται σε ειδικές ανάγκες της έρευνας. (Javeau, 2000)

Υπάρχουν δύο τύποι ερωτήσεων, οι **ανοιχτές ερωτήσεις** και οι **κλειστές ερωτήσεις** (με περαιτέρω υποκατηγορίες).

1. Ανοιχτές Ερωτήσεις:

Στις ερωτήσεις αυτές ο ερωτώμενος μπορεί να εκφράσει την γνώμη του ελεύθερα και χωρίς περιορισμούς

Για παράδειγμα:

Θα μπορούσατε, παρακαλώ πολύ, να προτείνετε ορισμένες νέες θεματικές κατηγορίες, που θα επιθυμούσατε να συμπεριλάβει η Ιατρική Βιβλιοθήκη στην προσπάθεια του εμπλουτισμού της; (όπως: γενετική, ιατρική πληροφορική)

.....
.....
.....

2. Κλειστές Ερωτήσεις που συμπεριλαμβάνουν:

- **Διχοτομικές ερωτήσεις:**

Οι ερωτήσεις αυτές επιτρέπουν στον ερωτώμενο να επιλέξει μόνο μία από τις δύο απαντήσεις που δίνονται. Είναι πολύ ξεκάθαρες και γι' αυτό ακριβώς η επεξεργασία είναι πολύ πιο εύκολη.

Για παράδειγμα:

Καλύπτει η Ιατρική Βιβλιοθήκη του Π.Γ.Ν.Α. τις πληροφοριακές σας ανάγκες;

Ναι

Όχι

- **Ερωτήσεις Βαθμονόμησης:**

Στις ερωτήσεις αυτές ο ερωτώμενος μπορεί να απαντήσει σε μόνο μία από τις υπάρχουσες κατηγορίες.

Για παράδειγμα:

Σας δυσκολεύει στη χρήση του, ο κατάλογος OPAC της Βιβλιοθήκης;

1. Καθόλου

2. Ελάχιστα

3. Μέτρια

4. Αρκετά

5. Πάρα πολύ

- **Ερωτήσεις Κατάταξης:**

Ο ερωτώμενος καλείται να επιλέξει τις απαντήσεις με σειρά προτεραιότητας ανάλογα με το ποια απάντηση θεωρεί πιο σημαντική.

Για παράδειγμα:

Ποιοι είναι οι κύριοι λόγοι για τους οποίους επισκέπτεστε τη βιβλιοθήκη συνήθως;
(παρακαλούμε επιλέξτε με σειρά προτεραιότητας τη σημαντικότητα των λόγων: 1 για το πρώτο, 2 για το δεύτερο, 3 για το τρίτο, 4...,5...)

1. Έρευνα
2. Ενημέρωση
3. Υλικό για εργασία
4. Δανεισμό
5. Μελέτη
6. Διάφοροι λόγοι (παρακαλώ προσδιορίστε)

• **Ερωτήσεις Διαβαθμισμένης Κλίμακας:**

Ο ερωτώμενος ζητείται να βαθμολογήσει με μια συγκεκριμένη κλίμακα μια κατηγορία ερωτήσεων.

Για παράδειγμα:

Προς ποια κατεύθυνση νομίζετε ότι πρέπει να δραστηριοποιηθεί στο άμεσο μέλλον η Ιατρική Βιβλιοθήκη, ώστε να βελτιωθούν οι παρεχόμενες υπηρεσίες και να εξυπηρετούνται καλύτερα οι επιστημονικές ανάγκες σας;

(Παρακαλούμε βαθμολογήστε τη σημασία που νομίζετε ότι έχουν για τις ανάγκες σας τα παρακάτω, με βάση τη κλίμακα)

1. Καθόλου σημαντική
2. Λιγότερο σημαντική
3. Αδιάφορο
4. Αρκετά σημαντική
5. Πολύ σημαντική

- Εγκατάσταση φωτοτυπικού μηχανήματος εντός του χώρου της Βιβλιοθήκης.

1

2

3

4.....

5.....

- Αύξηση Ωρών Λειτουργίας της Βιβλιοθήκης (π.χ. Σάββατο πρωί, 12ώρη λειτουργία)

1

2

3

4.....

5.....

- Αυτοματοποίηση Υπηρεσιών (π.χ. αυτόματος δανεισμός και επιστροφή βιβλίων, αυτόματη έκδοση κάρτας χρήστη).

1

2

3

4.....

5.....

- Αύξηση των θέσεων των Αναγνωστήριων.

1

2

3

4.....

5.....

- Δημιουργία ειδικών - "κλειστών" χώρων ατομικής ή ομαδικής μελέτης.

1

2

3

4.....

5.....

- **Ερωτήσεις Πολλαπλής Επιλογής:**

Στις ερωτήσεις αυτές ο ερωτώμενος μπορεί να επιλέξει περισσότερες από μία απαντήσεις.

Για παράδειγμα:

Παρακαλούμε προσδιορίστε ποιον τρόπο ενημέρωσης θεωρείτε πιο έγκυρο και έγκαιρο:
(μπορείτε να προσδιορίσετε περισσότερους από έναν τρόπο)

• Τηλεφωνικώς	
• Με FAX	
• Με προσωπικό e-mail	
• Με ανάρτηση σχετικής ανακοίνωσης στους πίνακες ανακοινώσεων κάθε Γραμματείας.	
• Με την περιοδική έκδοση ενημερωτικού φυλλαδίου για τις δραστηριότητες της Ιατρικής Βιβλιοθήκης.	
• Με τη διοργάνωση ενημερωτικών ημερίδων.	

3.3.4.4 Σειρά των ερωτήσεων

Η σειρά που ακολουθείται για την διάταξη των ερωτήσεων μέσα στο ερωτηματολόγιο είναι:

- Ερωτήσεις στοιχείων ταυτότητας (φύλλο, ηλικία) συγκεντρώνονται στην αρχή, έτσι ώστε η πρώτη εντύπωση που σχηματίζει ο ερωτώμενος να είναι θετική.
- Εύκολες ερωτήσεις στην αρχή, κατάλληλες να ευαισθητοποιήσουν και να προκαλέσουν ενδιαφέρον στον ερωτώμενο.
- Οι δύσκολες ερωτήσεις στο τέλος οπότε είναι δύσκολο να αρνηθεί να απαντήσει.
- Οι ερωτήσεις που αναφέρονται στο ίδιο θέμα να είναι συγκεντρωμένες σε ενότητες, και οι ερωτήσεις γενικού τύπου να προηγούνται των ειδικών έτσι ώστε το ερωτηματολόγιο να έχει μία όψη συναφή και λογική για να μην νιώθει ο ερωτώμενος ότι εκτροχιάζεται. (Javeau, 2000)

3.3.4.5 Διατύπωση των ερωτήσεων

Σχετικά με την διατύπωση των ερωτήσεων που αφορούν το λεξιλόγιο και την φρασεολογία του ερωτηματολογίου, αποφασίσθηκε:

- 1) Οι ερωτήσεις προφανώς να ακολουθούν τους κανόνες της γραμματικής και του συντακτικού.
- 2) Να είναι διατυπωμένες με σαφήνεια και με τέτοιο τρόπο ώστε να είναι κατανοητές από όλους τους χρήστες.
- 3) Να μην περιέχουν τεχνικούς όρους, ιδιωματισμούς, λέξεις σπάνιες και δύσκολες. Υπήρξαν όμως και κάποιοι τεχνικοί όροι οι οποίοι ήταν απαραίτητο να χρησιμοποιηθούν για την διατύπωση της ερώτησης. Έτσι, προστέθηκε κάποια επεξήγηση μέσα σε παρένθεση προκειμένου να κατανοήσουν πλήρως όλοι οι χρήστες τις έννοιες αυτές.
- 4) Τέλος από πλευράς περιεχομένου οι ερωτήσεις, αποφασίσθηκε να είναι απλές και περιεκτικές.

Σχετικά με το περιεχόμενο των ερωτήσεων θα πρέπει να έχουμε υπόψη μας τα εξής:

1. Την ικανότητα του ατόμου που θα ερωτηθεί. Αν δηλαδή ο πληθυσμός που θα ερωτηθεί έχει τις κατάλληλες γνώσεις να απαντήσει. Πολλές φορές δηλώνουν άγνοια για ψυχολογικούς λόγους και φυσικά οι απαντήσεις δεν ανταποκρίνονται στην πραγματικότητα.
2. Τη θέληση του ερωτωμένου. Συνήθως αποφεύγονται ερωτήσεις που από το περιεχόμενο τους προξενούν έντονη κοινωνική αντίδραση.

3.3.4.6 Υλική και τυπογραφική παρουσίαση

Για τη διάταξη και εμφάνιση του ερωτηματολογίου ο ερευνητής λαμβάνει υπόψη μερικές από τις βασικές αρχές του ερωτηματολογίου. Σύμφωνα μ' αυτές έχει τα παρακάτω χαρακτηριστικά:

1) Εύκολο στην χρήση

Δηλαδή όσον αφορά το σχήμα, το βάρος, τον τρόπο διασύνδεσης των φύλλων.

2) Εύκολο στην ανάγνωση

Δηλαδή, ευανάγνωστοι τυπογραφικοί χαρακτήρες, αραιά γραμμένο ερωτηματολόγιο, οι ερωτήσεις να διαχωρίζονται καλά μεταξύ τους και να αριθμούνται με απλό τρόπο.

3) Εύκολο στην συμπλήρωση

Για τις κλειστές ερωτήσεις, χρησιμοποιήθηκαν τα τετραγωνάκια (□) μέσα στα οποία οι ερωτώμενοι σημειώνουν τις επιλογές τους. Για τις ανοιχτές ερωτήσεις έχει προβλεφθεί πλαίσιο ικανοποιητικών διαστάσεων, ανάλογο με την υπολογιζόμενη έκταση των πιθανών απαντήσεων (.....).

4) Αισθητικά όμορφο

Να ελκύει και να προκαλεί το ενδιαφέρον του ερωτώμενου με την χρήση εικόνων και σχημάτων τα οποία τραβούν την προσοχή του και συγχρόνως προσθέτουν μια ευχάριστη νότα στο ερωτηματολόγιο (Για παράδειγμα •).

5) Πρόλογος

Ένας μικρός πρόλογος με εννοιολογικές επεξηγήσεις και ευχαριστίες ολοκληρώνει την αρτιότητα της εμφάνισης του ερωτηματολογίου.

3.3.4.7 Το μέγεθος του ερωτηματολογίου

Σύμφωνα με τον Πιέρρο (2000), ένα μεγάλο ερωτηματολόγιο είναι εξαρχής αποθαρρυντικό. Το μέγεθος του ερωτηματολογίου γίνεται εκ πρώτης όψεως αντιληπτό από τον όγκο των σελίδων του και από τον αριθμό των ερωτήσεων. Έτσι για να μειώσουμε το μέγεθος σε σελίδες χρησιμοποιούμε την κατάλληλη γραμματοσειρά, η οποία είχε ένα αρκετά ικανοποιητικό μέγεθος ώστε και να μην καλύπτει πολύ χώρο αλλά και οι χαρακτήρες να είναι ευανάγνωστοι.

Ακόμη, αποφασίσαμε οι ερωτήσεις να είναι συνοπτικές και περιεκτικές ώστε να συμπληρώνονται εύκολα. Έγινε προσπάθεια να αποφευχθούν άσκοπες και χωρίς νόημα ερωτήσεις, ώστε το ικανοποιητικό σύνολο των 20 ερωτήσεων να επαρκεί για χρήσιμα συμπεράσματα.

Τέλος, το ερωτηματολόγιο πρέπει να προσπαθήσουμε να γίνει όσο τον δυνατόν λιγότερο φορτικό και χρονοβόρο.

3.3.4.8 Δοκιμή του ερωτηματολογίου

Ο σχεδιασμός ενός καλού ερωτηματολογίου είναι δύσκολο έργο. Για να τελειοποιηθεί, στο βαθμό που να εξασφαλίζει τη συλλογή των επιθυμητών πληροφοριών, το τεχνικό επιτελείο της έρευνας θα πρέπει να είναι καλά προετοιμασμένο πριν αρχίσει η έρευνα. Έτσι, το ερωτηματολόγιο αναθεωρείται και βελτιώνεται.

3.3.5 Πρόσληψη προσωπικού και εκπαίδευση

Ο αριθμός των απασχολούμενων προσώπων σε μια έρευνα εξαρτάται από το μέγεθος του δείγματος, τα διαθέσιμα χρήματα και το χρονικό διάστημα αποπεράτωσης της έρευνας.

Οι κατηγορίες του προσωπικού που θα προσληφθεί πρέπει να είναι:

α. Ειδικευμένο προσωπικό, που θα ασχοληθεί με το σχεδιασμό, την έρευνα, τη εποπτεία και την ανάλυση των αποτελεσμάτων.

β. Υπάλληλοι γραφείου, που θα ασχοληθούν με την επεξεργασία.

γ. Ερευνητές, που θα ασχοληθούν με τη συμπλήρωση του ερωτηματολογίου.

Η κατηγορία αυτή αποτελεί το βασικό προσωπικό μιας έρευνας, γιατί από τον τρόπο που θα ασχοληθούν με τη συγκέντρωση των στοιχείων θα εξαρτηθεί πάρα πολύ η ποιότητα της έρευνας.

3.3.5.1 Εκπαίδευση ερευνητών

Πρωταρχικός σκοπός της εκπαίδευσης των ερευνητών, είναι η εξοικειώσή τους με τους σκοπούς της έρευνας, το ερωτηματολόγιο, τον τρόπο συμπλήρωσης του και ο καταρτισμός τους σε ορισμένα διοικητικής φύσεως θέματα.

3.3.5.2 Αρχική εκπαίδευση

Αυτή συνιστάται σε ένα ειδικό εντατικό εκπαιδευτικό πρόγραμμα που περιλαμβάνει:

- 1) Την ανάλυση των σκοπών της έρευνας
- 2) Τους τρόπους αξιοποίησης των αποτελεσμάτων
- 3) Συνοπτική περιγραφή της διαδικασίας της επεξεργασίας
- 4) Τις εργασίες του ερευνητή
- 5) Ανάπτυξη της τεχνικής των συνεντεύξεων
- 6) Διευκρινίσεις στις οδηγίες
- 7) Αναπαραστάσεις συνεντεύξεων
- 8) Δοκιμαστική συμπλήρωση του ερωτηματολογίου
- 9) Δοκιμαστική συμπλήρωση σε τυχαίο δείγμα

10) Ανάλυση λαθών τους και κριτικής αυτών

11) Κατατοπισμός ερευνητών σε ορισμένα διοικητικά θέματα

3.3.6 Τελική δοκιμαστική έρευνα

Η διενέργεια μιας τελικής δοκιμαστικής έρευνας εξυπηρετεί πολλούς σκοπούς. Ο κυριότερος σκοπός είναι η υποβολή σε τελική δοκιμή του ερωτηματολογίου πριν από την έναρξη της κύριας έρευνας. Μεταξύ των άλλων σκοπών είναι και η δοκιμή της επάρκειας των οδηγιών που έχουν δοθεί στους ερευνητές, η δοκιμή της εποπτείας και της οργάνωσης στην πράξη.

Από τη δοκιμαστική έρευνα έχουμε τις πρώτες ενδείξεις σχετικά με το ποσοστό των αρνήσεων, μη επαφών. Η δοκιμαστική έρευνα πρέπει να γίνεται με συνθήκες, όσο το δυνατόν παρόμοιες με την κύρια έρευνα. Η επιλογή της μεθόδου για τη συλλογή των στατιστικών στοιχείων στη δειγματοληψία επηρεάζεται από το αντικείμενο της έρευνας, τη μονάδα της έρευνας.

3.3.7 Μέθοδοι Επιλογής στατιστικών στοιχείων

Οι μέθοδοι συλλογής των στοιχείων στις δειγματοληπτικές έρευνες είναι η παρατήρηση, οι προσωπικές συνεντεύξεις και το ταχυδρομείο.

3.3.7.1 Η μέθοδος της παρατήρησης

Είναι η διαδικασία όπου κάποιο φαινόμενο ή συμπεριφορά παρατηρούνται κατά τρόπο προγραμματισμένο, οργανωμένο, συστηματοποιημένο, από άτομα εξειδικευμένα ή εκπαιδευμένα για τον ρόλο αυτό. Χρησιμοποιεί την καταγραφή γεγονότων και υφίσταται επαλήθευση. Τα είδη της παρατήρησης είναι: η άμεση, η συμμετοχική και η έμμεση ή προκαλούμενη ή πειραματική παρατήρηση. (Δημητρόπουλος, 1994)

3.3.7.2 Η μέθοδος της συνέντευξης

Είναι μία από τις πιο γνωστές μεθόδους συλλογής υλικού όπου ο ερευνητής υποβάλλει στον ερωτώμενο μια σειρά από ερωτήσεις στις οποίες καλείται να απαντήσει. Αυτό που ενδιαφέρει τον ερευνητή είναι να ανακαλύψει τι σκέφτεται ο ερωτώμενος σε σχέση με κάποιο θέμα και να συγκρίνει τις γνώμες και τις απόψεις των ερωτώμενων. Στην συνέχεια ο ερευνητής ενδιαφέρεται να συγκρίνει και να ομαδοποιήσει τις απόψεις των ερωτώμενων.

Το έργο των ερευνητών συνιστάται:

1. Στον εντοπισμό των μονάδων του δείγματος
2. Στην πραγματοποίηση των συνεντεύξεων
3. Στον τρόπο υποβολής των ερωτήσεων
4. Στην καταγραφή των απαιτήσεων

Η καταγραφή φαίνεται απλή εργασία αλλά παρουσιάζονται λάθη που οφείλονται, κυρίως στο ότι το ερωτηματολόγιο είναι κουραστικό και προκαλεί διάσπαση προσοχής, ενώ ο τρόπος συμπλήρωσης του πολλές φορές είναι πολύπλοκος.

3.3.7.3 Η μέθοδος του Ταχυδρομείου

Παλιά είχε μεγάλη εφαρμογή η αποστολή, η συμπλήρωση του ερωτηματολογίου και η επιστροφή του με το ταχυδρομείο. Σήμερα η μέθοδος αυτή έχει περιοριστεί αρκετά. Σε ορισμένες όμως περιπτώσεις, είναι ο μόνος τρόπος συλλογής των στοιχείων. Όπως, όταν ο πληθυσμός είναι διασκορπισμένος σε μεγάλες εκτάσεις, καθώς και όταν ο διαθέσιμος χρόνος και οι χρηματικοί πόροι είναι περιορισμένοι.

Τα πλεονεκτήματα της μεθόδου αυτής είναι το μικρό κόστος, η ταχεία διεξαγωγή της έρευνας, η αποφυγή σφαλμάτων του ερευνητή, ή καλύτερη συγκέντρωση εμπιστευτικών πληροφοριών συγκεντρώνονται καλύτερα με τον τρόπο αυτό.

Το μειονέκτημα της μεθόδου αυτής είναι ότι μόνο ένα μικρό ποσοστό απαντά και ότι η πληροφορίες δεν είναι όλες ορθές. (Cochran W,1977)

Κεφάλαιο: 4 Στατιστικά Δειγματοληπτικά σφάλματα

4.1 Εισαγωγή

Στην σημερινή εποχή είναι πλέον δυνατή η μέτρηση χαρακτηριστικών ιδιοτήτων μαζικών δεδομένων με συγκεκριμένη (υπολογιστικά) ακρίβεια με βάση τα δείγματα. Αυτό έχει σαν συνέπεια, σχεδόν όλες οι στατιστικές έρευνες σε όλους τους τομείς (επιχειρηματικό, κυβερνητικής πολιτικής, κοινωνικό και οικονομικό) να είναι δειγματοληπτικές. Πέραν του ότι χρησιμοποιώντας δειγματοληπτικές έρευνες, παράγονται αξιόπιστα αποτελέσματα υπάρχουν και άλλοι λόγοι που τις θεωρούμε σημαντικές.

Ένας κύριος λόγος είναι ότι οι υπό έρευνα πληθυσμοί μπορεί να είναι άπειροι, με συνέπεια η δειγματοληψία να είναι η μόνη εφαρμόσιμη μέθοδος. Ακόμα όμως και αν ο πληθυσμός είναι πεπερασμένος μπορεί να αποτελείται από δεκάδες χιλιάδες ή και εκατομμύρια μονάδες και άρα η πλήρης μελέτη του να είναι πρακτικά αδύνατη. Άλλος ένας λόγος είναι ότι η εξέταση των μονάδων ενός πληθυσμού ως προς το επιθυμητό χαρακτηριστικό μπορεί να προκαλεί την καταστροφή τους (έρευνα δύναμης αντοχής υλικών).

Επίσης, πολλές φορές ο τύπος δεδομένων του πληθυσμού προς εξέταση μπορεί να μην είναι διαθέσιμος (ανάλυση χρονολογικών σειρών). Ακόμα, και αν ήταν πρακτικά δυνατή η μελέτη ολόκληρου του πληθυσμού, τα αποτελέσματα μέσω δειγματοληψίας μπορεί να είναι εξίσου ακριβή.

Ο σχεδιασμός και η εκτέλεση μιας στατιστικής έρευνας, είναι επιτυχής, όταν περιορίζει το κόστος και το χρόνο αυτής. Ο περιορισμός όμως αυτός, δεν θα πρέπει να έχει αρνητική επίδραση στην ποιότητα της έρευνας με την αύξηση των σφαλμάτων, που πιθανόν θα προέλθουν αν σχεδιαστεί πλημμελώς η έρευνα ή ενημερωθούν ανεπαρκώς οι ερευνητές ή περιοριστούν οι έλεγχοι ποιότητας, αν περιοριστεί η επιτήρηση και ο έλεγχος των εργασιών της έρευνας.

Στη δειγματοληπτική έρευνα υπάρχουν σφάλματα. Τα αποτελέσματα μιας δειγματοληπτικής έρευνας δε μπορεί να είναι πάντοτε ορθό. Δηλαδή, οι εκτιμήσεις για τις τιμές των παραμέτρων (μέση τιμή, τυπική απόκλιση) ενός πληθυσμού ή οι αποφάσεις σχετικά με κάποιες υποθέσεις που αφορούν τις τιμές αυτών των παραμέτρων, δεν ανταποκρίνονται στην πραγματικότητα. Αυτό είναι φυσικό, γιατί όσο αντιπροσωπευτικό και αν είναι το δείγμα, δε μπορεί να συμπεριλάβει με απόλυτη ακρίβεια, τη μεταβλητότητα των τιμών του.

4.2 Ορισμός Στατιστικών Σφαλμάτων

Τα **στατιστικά σφάλματα (errors)**, είναι οι αποκλίσεις των αποτελεσμάτων της δειγματοληπτικής έρευνας από τις πραγματικές τους τιμές. Δηλαδή, η διαφορά που υπάρχει μεταξύ μιας εκτίμησης που προκύπτει από ένα δείγμα από την αντίστοιχη παράμετρο του πληθυσμού που προκύπτει από την απογραφή.

Με τον όρο **Σφάλμα (error)** μιας δειγματικής συναρτήσεως σημαίνει τη 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.

4.3 Κατηγορίες σφαλμάτων

Τα σφάλματα διακρίνονται σε δυο κατηγορίες:

1. **Τα δειγματοληπτικά σφάλματα**
2. **Τα μη δειγματοληπτικά σφάλματα**

4.3.1 Δειγματοληπτικά σφάλματα

Τα **Σφάλματα Δειγματοληψίας (sampling error)** προκύπτουν από την τυχαία επιλογή των μονάδων δειγματοληψίας. Αυτός ο τύπος σφάλματος προκύπτει, γιατί ένα μέρος του πληθυσμού περιέχεται στο δείγμα. Εάν λάβει χώρα μία 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 ιδίων διαδικασιών (μεθόδων).

Το δειγματοληπτικό σφάλμα αναφέρεται σε διαφορές μεταξύ του δείγματος και του πληθυσμού οι οποίες υπάρχουν επειδή αυτές οι συγκεκριμένες παρατηρήσεις έτυχε να επιλεγθούν. Ένας άλλος τρόπος για να δούμε αυτό είναι: οι διαφορές αποτελεσμάτων για διαφορετικά δείγματα (ιδίου μεγέθους) οφειλόμενα καθαρά δειγματοληπτικό σφάλμα.

Για παράδειγμα, δύο δείγματα μεγέθους 10 από 1.000 νοικοκυριά. Εάν συνέβη να πάρουμε τα δεδομένα με το υψηλότερα εισοδήματα στο πρώτο μας δείγμα και όλα τα χαμηλότερα στο δεύτερο, αυτή η διαφορά οφείλεται καθαρά σε δειγματοληπτικό σφάλμα. Αύξηση του δείγματος, μειώνει το δειγματοληπτικό σφάλμα.

Τα **δειγματοληπτικά σφάλματα ή σφάλματα παρατήρησης** οφείλονται στην μέτρηση των χαρακτηριστικών των ατόμων του δείγματος και όχι του συνολικού πληθυσμού. Λαμβάνουμε δηλαδή την πληροφορία από τους επιλεγμένους του δείγματος και η πληροφορία αυτή είναι εσφαλμένη. Μια ερώτηση ενδέχεται να είναι εσφαλμένη όταν το σφάλμα οφείλεται στο συνεντευκτή, στην ερώτηση, στη καταγραφή, στην κωδικοποίηση και στη μετάδοση. Για τα σφάλματα αυτά δεν είναι υπεύθυνο το ερωτώμενο μέλος του δείγματος.

Η άλλη κατηγορία των δειγματοληπτικών σφαλμάτων είναι τα **σφάλματα μέτρησης**. Τα σφάλματα αυτά προκύπτουν όταν υπάρχει μια συγκεκριμένη τιμή X_i αναφερόμενη στο άτομο i η οποία είναι δύσκολο να παρατηρηθεί χωρίς κάποιο επιπρόσθετο σφάλμα.

4.3.2 Τα μη δειγματοληπτικά σφάλματα

Ο όρος **μη δειγματοληπτικά σφάλματα (nonsampling error)** προέρχεται από το γεγονός ότι, αυτός ο τύπος σφάλματος μπορεί να συμβεί σε οποιαδήποτε έρευνα, είτε αυτή γίνεται με τη μέθοδο της απογραφής είτε με τη μέθοδο της δειγματοληψίας.

Τα **μη δειγματοληπτικά σφάλματα** περιέχουν μεροληψίες και λάθη.

Μία μεροληψία εμφανίζεται συνήθως 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 σφάλματος παρουσιασθεί, τα δειγματικά αποτελέσματα θα βρίσκονται πάντοτε προς τα πάνω ή πάντοτε προς τα κάτω της τιμής της παραμέτρου και γι' αυτό δεν μπορούν να αλληλοεξουδετερωθούν τα σφάλματα δια του μέσου όρου.

Παράγοντες οι οποίοι προκαλούν μεροληψίες:

- 1) Ελλιπής προσδιορισμός του προς έρευνα πληθυσμού
- 2) Ατέλεια του πλαισίου (ανακρίβεια, έλλειψη πληρότητας, διπλοεγγραφές)
- 3) Αοριστία του ερωτηματολογίου
- 4) Όχι σαφής καθορισμός των επιθυμητών πληροφοριών
- 5) Υπερβολικές ή άσχετες απαντήσεις σε ερωτήσεις
- 6) Ανακριβείς μεθόδους συνεντεύξεως

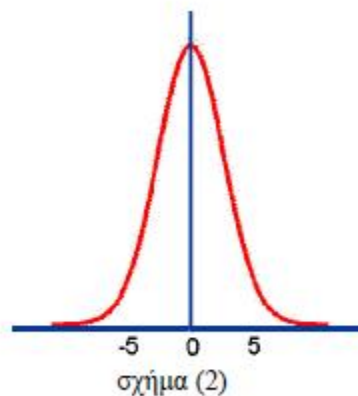
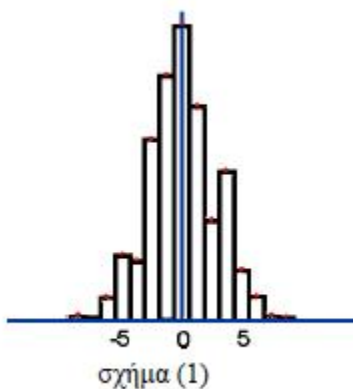
Τα **μη δειγματοληπτικά σφάλματα** οφείλονται συνήθως σε:

- 1) Ατέλειες του σχεδιασμού και της οργάνωσης της έρευνας για λόγους ταχύτητας και περιορισμού του κόστους

- 2) Λάθη των ερευνητών, κωδικογράφων, κλπ
- 3) Άλλες αντικειμενικές δυσκολίες που ανακύπτουν κατά την εκτέλεση μιας στατιστικής έρευνας. Συμβαίνουν δε ανεξάρτητα της επιλογής δείγματος ή της πλήρους καταγραφής (απογραφής).

4.3.3 Τα τυχαία σφάλματα είναι αναπόφευκτα και περιγράφονται με τη στατιστική θεωρία

Σύμφωνα με τη στατιστική θεωρία εάν ένα φαινόμενο είναι πράγματι τυχαίο τότε η οριακή κατανομή που θα προκύψει (μετά από άπειρες προσπάθειες) θα είναι μια κανονική κατανομή ή κατανομή Gauss. Η κατανομή Gauss, είναι ίσως η κοινή κατανομή στη θεωρία των πιθανοτήτων. Δηλαδή εάν επαναλάβουμε ένα πείραμα. Για παράδειγμα, ρίχνοντας ένα βέλος να πετύχουμε το 0. Το αποτέλεσμα που παίρνουμε για παράδειγμα 100 προσπάθειες φαίνεται στο Σχήμα 1, ενώ για άπειρες στο Σχήμα 2 και περιγράφεται μαθηματικά από την καμπύλη.



ή από τη σχέση

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-x_0)^2}{2\sigma^2}}$$

4.4 Ακρίβεια και αξιοπιστία στα σφάλματα

Τα σφάλματα δειγματοληψίας ή οι διακυμάνσεις της δειγματοληψίας μετρούνται με αυτό που ονομάζεται **ακρίβεια (precision)**. Η ακρίβεια, η οποία δηλώνεται με τον όρο **αξιοπιστία (reliability)**, είναι ο βαθμός με τον οποίο 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 δειγματικής συναρτήσεως. Όσο μικρότερο είναι το τυπικό σφάλμα, τόσο μεγαλύτερη είναι η ακρίβεια (ή αξιοπιστία) της εκτιμήσεως.

Ακρίβεια (accuracy) μιας δειγματικής εκτιμήσεως, είναι η διαφορά μεταξύ της δειγματικής συναρτήσεως του δείγματος και της πραγματικής τιμής της παραμέτρου. Η ακρίβεια μια εκτιμήσεως δεν μπορεί άμεσα να μετρηθεί, διότι η πραγματική τιμή της παραμέτρου είναι μια άγνωστη σταθερά. Η ακρίβεια μιας εκτιμήσεως όμως, εξαρτάται, και από το βαθμό αξιοπιστίας και από την απουσία μεροληψίας. Η ακρίβεια είναι ο σκοπός της έρευνας.

Αξιοπιστία (reliability) είναι τι μπορεί να μετρηθεί. Ο έλεγχος και υπολογισμός της αξιοπιστίας των αποτελεσμάτων του δείγματος δια των νόμων της θεωρίας των πιθανοτήτων είναι η βασική αποστολή της δειγματοληψίας, διότι αυτό βοηθάει να επιτευχθεί η ακρίβεια της δειγματικής εκτιμήσεως.

Οι μεροληψίες, πρέπει να σημειωθεί ότι 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, μία μεροληπτική διαδικασία μπορεί πράγματι να έχει ένα μικρότερο τυπικό σφάλμα από ότι μία αμερόληπτη διαδικασία. Συνεπώς, οι ευκαιρίες του να πάρουμε περισσότερο ακριβή εκτίμηση μπορεί να είναι περισσότερες με μια μεροληπτική εκτίμηση από ότι από μια αμερόληπτη εκτίμηση.

Αυτό που πρέπει να γίνει σαφές είναι ότι, όταν μία μεροληπτική διαδικασία εφαρμόζεται, πρέπει να γνωρίζουμε σαφώς τη φύση της μεροληψίας και την ερμηνεία των αποτελεσμάτων αντίστοιχα στις διαδικασίες λήψεων αποφάσεως. Το ότι ένα μεροληπτικό δείγμα μπορεί πράγματι να προκαλέσει μια περισσότερο ακριβή εκτίμηση είναι ένας από τους λόγους, που επιβάλλουν την χρησιμοποίηση άλλων σχεδίων δειγματοληψίας εκτός από την τυχαία δειγματοληψία. (B. MIPENY, 1991)

4.5 Είδη σφαλμάτων

Υπάρχουν δύο ειδών σφάλματα, τα συστηματικά και τα τυχαία.

4.5.1 Συστηματικά σφάλματα

Συστηματικά σφάλματα είναι τα σφάλματα που σχετίζονται με την αξιοπιστία μιας μέτρησης και μπορεί να οφείλονται στην κακή βαθμονόμηση των οργάνων, στη λανθασμένη χρήση των οργάνων ή στην παράβλεψη ορισμένων φαινομένων ή σε εξωτερικά αίτια που μπορεί να αλλάξουν τα αποτελέσματα του πειράματος (υγρασία, πίεση, θερμοκρασία κ.λ.π). Τα συστηματικά σφάλματα τείνουν να μετατοπίσουν όλες τις μετρήσεις με συστηματικό τρόπο έτσι ώστε η μέση τιμή να είναι μετατοπισμένη προς μία διεύθυνση.

Για παράδειγμα, εάν ένας χάρακας είναι φθαρμένος, η μέτρηση κάθε μήκους θα έχει σταθερό συστηματικό σφάλμα που είναι ανεπηρέαστο από την επαναληψιμότητα της μέτρησης.

Ο μόνος τρόπος να ποσοτικοποιήσουμε το σφάλμα, δηλαδή να εκτιμήσουμε σωστά την τάξη μεγέθους του είναι να συγκρίνουμε το όργανο που χρησιμοποιούμε με άλλο που θεωρείται πρότυπο. Σε ένα σωστό πείραμα τα μεγάλα συστηματικά σφάλματα περιορίζονται με σύγκριση των τιμών με διαφορετικές μεθόδους.

4.5.2 Τυχαία σφάλματα

Τυχαία σφάλματα είναι τα σφάλματα που σχετίζονται με την ακρίβεια μιας μέτρησης και δείχνουν τις διακυμάνσεις που έχουν οι μετρήσεις ενός επαναλαμβανόμενου πειράματος που γίνεται κάτω από τις ίδιες φαινομενικά συνθήκες και τα οποία οδηγούν στην κατανομή των αποτελεσμάτων γύρω από μία μέση τιμή. Μπορεί να οφείλονται στην έλλειψη ευαίσθητης απόκρισης του οργάνου ή στον παρατηρητή (σφάλματα ανάγνωσης), στον εξωτερικό «θόρυβο», ή σε στατιστικές διαδικασίες. Όπως είναι η ρίψη ενός ζαριού.

4.6 Μέση τιμή και σφάλμα

Επειδή σε πολλές περιπτώσεις, όταν μετρούμε πολλές φορές, στις ίδιες συνθήκες την ίδια ποσότητα βρίσκουμε διαφορετικά αποτελέσματα, βρίσκουμε τη μέση τιμή και το **απόλυτο σφάλμα της μέσης τιμής** (ή τυπική απόκλιση της μέσης τιμής). Εάν σε ένα πείραμα η μέτρηση του μεγέθους x επαναληφθεί N φορές, και οι μετρούμενες τιμές είναι $x_1, x_2, x_3 \dots x_N$ τότε ως πραγματική θεωρούμε την μέση τιμή.

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_k + \dots + x_N}{N} = \frac{\sum_{k=1}^N x_k}{N}$$

Εάν τα σφάλματα των παραπάνω μετρήσεων είναι τυχαία θα διαφέρουν ως προς το πρόσημο και ως προς το μέγεθος. Έτσι στον υπολογισμό της μέσης τιμής κάποια από τα τυχαία σφάλματα αλληλοαναιρούνται στο άθροισμα. Η επανάληψη πολλών μετρήσεων είναι και ο καλύτερος τρόπος περιορισμού των τυχαίων σφαλμάτων. Επιπλέον μπορεί να υπολογιστεί η απόκλιση των μετρήσεων από τη μέση τιμή. Το απόλυτο σφάλμα της μέσης τιμής μπορεί να χρησιμοποιηθεί για να εκφράσει τη βεβαιότητα των N μετρήσεων μας για τη μέση τιμή του x.

$$\delta_{\bar{x}} = \sqrt{\frac{\sum_k (x_k - \bar{x})^2}{N(N-1)}} \equiv \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}}$$

Συμβολίζουμε το απόλυτο σφάλμα της μέσης τιμής με το ελληνικό γράμμα «δ» για να συμφωνούμε με τα διεθνώς κείμενα και ισχύοντα. Παρατηρούμε ότι το απόλυτο σφάλμα της μέσης τιμής δεν είναι τίποτε άλλο από αυτό που είχαμε ορίσει πριν «τυπική απόκλιση της μέσης τιμής».

Παράδειγμα: Σε ένα πείραμα καταμέτρησης κοσμικών ακτίνων που προσπίπτουν σε έναν ανιχνευτή ανά ώρα, έγιναν εννέα μετρήσεις και τα αποτελέσματα καταγράφονται στον παρακάτω πίνακα.

i	x_i	$(x_i - \bar{x})^2$
1	80	400
2	95	25
3	100	0
4	110	100
5	90	100
6	115	225
7	85	225
8	120	400
9	105	25
Σύνολο	900	1500

$$\bar{x} = \frac{1}{9} \sum_{i=1}^9 x_i = \frac{900}{9} = 100$$

Και

$$\delta_x = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{N(N-1)}} = \sqrt{\frac{1500}{8}} = \sqrt{188} = 14$$

Άρα τελικά ο αριθμός των κοσμικών ακτίνων που μετρήθηκαν είναι $x = 100 \pm 14$ με σφάλμα $14/100=14\%$.

Κεφάλαιο 5: Απλή τυχαία δειγματοληψία

5.1 Εισαγωγή

Σε αυτή την ενότητα θα παρουσιαστεί η απλή τυχαία δειγματοληψία, κύριο χαρακτηριστικό της οποίας είναι ότι κάθε στοιχείο του πληθυσμού έχει την ίδια πιθανότητα να συμπεριληφθεί στο δείγμα.

Το δείγμα αποτελεί ένα υποσύνολο του πληθυσμού που επιθυμούμε να εξετάσουμε κατά τη διεξαγωγή ενός πειράματος. Τα δεδομένα που λαμβάνονται από ένα πείραμα δέχονται μεταβολές, έτσι ώστε μια εκτίμηση που προέρχεται από αυτά να μεταβάλλεται και αυτή, και συνεπώς αυτή η εκτίμηση είναι αβέβαιη. Εάν όμως, ήταν δυνατόν να επαναληφθεί το πείραμα πολλές φορές τότε εξετάζοντας τα αποτελέσματα συνολικά, η εκτίμηση θα κατέληγε σε κάποια αμετάβλητη τιμή η οποία είναι το οριστικό και πραγματικό αποτέλεσμα του πειράματος.

Ο σκοπός της ανάλυσεως ενός πειράματος, είναι να αποκαλύψει τι μπορούν να δώσουν τα δεδομένα του γύρω από αυτό το πραγματικό αποτέλεσμα. Κατά τη διάρκεια των τελευταίων ετών η δειγματοληψία έχει προχωρήσει τόσο, έτσι ώστε να αποτελεί τη βάση για τη λήψη αποφάσεων σε πολλές περιπτώσεις και σε πολλούς τομείς όπως κυβερνητικές υπηρεσίες, γραφεία έρευνα αγοράς και δημοσκοπήσεις κοινής γνώμης. (*B. Μπένου, 1991*)

5.2 Ορισμός και ανάλυση απλής τυχαίας δειγματοληψίας

Ορισμός: Η απλή τυχαία δειγματοληψία (**simple random sampling**) είναι η μέθοδος επιλογής n – μονάδων από έναν πληθυσμό N – μονάδων, κατά τέτοιο τρόπο ώστε κάθε δυνατό δείγμα μεγέθους n να έχει την ίδια πιθανότητα να επιλεγεί. (*Χαρίσης – Κιόχος, 1997*)

Αυτή η σχέση περιγράφεται από τον τύπο που ακολουθεί:

$$C_n^N = \binom{N}{n} = \frac{N!}{n!(N-n)!}$$

Περιγραφή: Στην απλή τυχαία δειγματοληψία κάθε στοιχείο του δείγματος επιλέγεται τυχαία έτσι ώστε κάθε τέτοιο στοιχείο να έχει την ίδια πιθανότητα να επιλεγεί από τον πληθυσμό σε οποιοδήποτε στάδιο κατά τη διάρκεια της διαδικασίας της δειγματοληψίας.

(*Yates, Daniel S. David S. Moore, Daren S. Starnes 2008*)

Η απλή τυχαία δειγματοληψία δεν χρησιμοποιείται συχνά στην πράξη, όμως αυτό δεν αναιρεί την σπουδαιότητά της στη θεμελίωση της θεωρίας της δειγματοληψίας. Πολύ σημαντικά χαρακτηριστικά της απλής τυχαίας δειγματοληψίας είναι η απλότητά της και η ικανοποιητική προσέγγιση σε ορισμένους τύπους δειγματοληψίας, που χρησιμοποιούνται στην πράξη ευρέως.

Παράδειγμα: Έστω ότι έχουμε έναν πληθυσμό που αποτελείται από 10 μονάδες και θέλουμε να πάρουμε δείγμα 4 μονάδων, τότε οι δυνατές τετράδες θα είναι όσοι είναι οι συνδυασμοί 10 ανά 4, δηλαδή:

$$\binom{10}{4} = \frac{10!}{4!(10-4)!} = 210$$

Συνεπώς μπορούν να επιλεγούν 210 δείγματα των 4 μονάδων. Κάθε ένα από τα 210 δείγματα έχει την ίδια πιθανότητα να επιλεγεί:

$$P = 1 : \binom{10}{4}$$

5.3 Πλεονεκτήματα – Μειονεκτήματα απλής τυχαίας δειγματοληψίας

Πλεονεκτήματα

- Το δείγμα είναι τυχαίο και επομένως μπορεί να υπολογιστεί το μέγεθος του δειγματοληπτικού σφάλματος. Τα αποτελέσματα της δειγματοληψίας είναι αντικειμενικά, δεδομένου ότι απομακρύνεται η υποκειμενική φύση της επιλογής των ερωτηθέντων και κατά συνέπεια απομακρύνεται η πιθανότητα μεροληψίας (αφού το δείγμα λαμβάνεται με κλήρωση).
- Είναι μια τεχνική εύκολη στην κατανόηση, δεδομένου ότι έχει απλό εννοιολογικό σχεδιασμό καθιστώντας την προσιτή ακόμα και σε εκείνους που δεν έχουν στατιστικό υπόβαθρο.

Μειονεκτήματα

- Είναι δαπανηρή, επειδή η μέθοδος αυτή απαιτεί πλαίσια πλήρως ενημερωμένα. Είναι δηλαδή αναγκαίο, να κατασκευαστεί κατάλογος και να αριθμηθεί κάθε μονάδα του συνόλου του πληθυσμού.
- Σε πολλές περιπτώσεις και ειδικά όταν το σύνολο του υπό μελέτη πληθυσμού είναι πεπερασμένο, η μέθοδος αυτή είναι ιδιαίτερα χρονοβόρα και έχει μεγάλο κόστος στην συλλογή των δεδομένων.

5.4 Μέθοδος της κλήρωσης

Υποθέτουμε ότι έχουμε στη διάθεσή μας ένα φάκελο με καρτέλες στις οποίες περιέχονται γραμμένα τα ονόματα π.χ. 10.000 νεοσύλλεκτων. Για να μελετήσουμε το ύψος τους, παίρνουμε δείγμα από 500 ονόματα νεοσυλλέκτων ως ακολούθως:

Τα ατομικά στοιχεία των 10.000 νεοσύλλεκτων είναι δυνατόν να αντιγραφούν σε ισομεγέθη τεμάχια από χαρτόνι ή σε ομοιόμορφα σφαιρίδια και ύστερα να ανασυρθούν από ένα άτομο τυχαία 500 καρτέλες ή 500 σφαιρίδια από μια κάλπη στην οποία θα έχουμε τοποθετήσει το σύνολο των 10.000 καρτελών ή σφαιριδίων. Για να μην αλλοιώνεται σε κάθε κλήρωση η σύνθεση της κάλπης, θα πρέπει οι αριθμοί οι οποίοι έχουν κληρωθεί, αφού σημειωθούν, να ξαναριχτούν στην κάλπη πριν από κάθε νέα κλήρωση, γιατί τότε και μόνο τότε κάθε καρτέλα ή κάθε σφαιρίδιο έχει την πιθανότητα να συμπεριληφθεί στο δείγμα. Το δείγμα των 500 στρατιωτικών προέρχεται από ολόκληρο τον πληθυσμό των 10.000 μονάδων και κάθε νεοσύλλεκτος από τους 10.000 έχει την ίδια πιθανότητα να επιλεγεί. Το δείγμα το οποίο το παίρνουμε με αυτό τον τρόπο λέγεται δείγμα με επανατοποθέτηση.

Το δείγμα των 500 μονάδων το επιλέξουμε με τον εξής τρόπο:

Εξάγουμε μια σφαίρα ή μία καρτέλα από την κάλπη και σύμφωνα με την ένδειξή της εντοπίζουμε την αντίστοιχη μονάδα του πληθυσμού. Ύστερα χωρίς επανατοποθέτηση της σφαίρας ή της καρτέλας που βγάλαμε από την κάλπη, εξάγουμε μια νέα σφαίρα ή καρτέλα για τον εντοπισμό της δεύτερης μονάδας του πληθυσμού που επιλέξαμε. Εκτελώντας $n=500$ τέτοιες εξαγωγές, έχουμε τους ακόλουθους αριθμούς που έχουν εκλεγεί από τον πληθυσμό: X_1, \dots, X_{500} . Αυτοί θα αποτελούν το τυχαίο δείγμα χωρίς επανατοποθέτηση.

Οι αριθμοί που έχουν επιλέγει με τον προηγούμενο τρόπο δεν μπορούν να θεωρηθούν τιμές ανεξάρτητων μεταβλητών που ακολουθούν τον ίδιο νόμο, γιατί η μη επανατοποθέτηση κάθε σφαίρας ή καρτέλας μέσα στην κάλπη μεταβάλλει σε κάθε εξαγωγή τη σύνθεση της κάλπης, άρα και τον αριθμό που αντιπροσωπεύει η κάλπη. Όταν όμως το πλήθος N των μονάδων του πληθυσμού είναι πολύ μεγάλος αριθμός σε σχέση με το μέγεθος του δείγματος που παίρνουμε, τότε η μέθοδος χωρίς επανατοποθέτηση συμπίπτει με τη μέθοδο της επανατοποθέτησης.

5.5 Μέθοδος τυχαίων αριθμών

Στην πράξη, όταν ο πληθυσμός N είναι αριθμός μεγάλος και, επιπλέον, η ελάχιστη εκτροπή από τις αυστηρές απαιτήσεις της απλής τυχαίας δειγματοληψίας δεν παραμένει χωρίς συνέπειες, η μέθοδος της κλήρωσης σπάνια εφαρμόζεται.

Πραγματικά, η ανάμειξη των καρτελών για να έχουμε τυχαίο δείγμα στην κάλπη δεν είναι τόσο απλή όσο φαίνεται αρχικά, γιατί ενδέχεται οι καρτέλες ή οι σφαίρες να βρεθούν στις πλευρές της κάλπης, οπότε η πιθανότητα της λήψης μιας καρτέλας ή μιας σφαίρας δεν είναι η ίδια για όλες τις καρτέλες ή σφαίρες της κάλπης.

Αν ο πληθυσμός που ερευνάμε περιλαμβάνει N συνολικά μονάδες αριθμημένες από το 1 ως το N διαδοχικά, μπορούμε εύκολα να πάρουμε τυχαίο δείγμα κάνοντας χρήση των πινάκων των τυχαίων αριθμών. Οι πίνακες αυτοί περιέχουν σειρές και στήλες αριθμών κατά τυχαία τάξη τοποθετημένων και είναι μία ακολουθία ψηφίων από το 0 μέχρι το 9.

5.5.1 Κατασκευή πίνακα τυχαίων αριθμών

Παίρνουμε 10 σφαιρίδια και πάνω σε κάθε σφαιρίδιο αναγράφουμε έναν από τους αριθμούς από το 0 μέχρι το 9. Στη συνέχεια, τοποθετούμε τα σφαιρίδια σε μια κάλπη και εξάγουμε τυχαία ένα, σημειώνουμε τον αριθμό και επανατοποθετούμε το σφαιρίδιο μέσα στην κάλπη πριν κάνουμε νέα εξαγωγή. Με τον τρόπο αυτό εξασφαλίζουμε το σχηματισμό τυχαίων αριθμών με όσα ψηφία επιθυμούμε. Στους πίνακες τυχαίων αριθμών, κάθε σελίδα, κάθε γραμμή και κάθε στήλη είναι διαφορετική από τις άλλες. Η χρησιμοποίηση του πίνακα των τυχαίων είναι πολύ εύκολη. Δίνουμε παρακάτω ένα απόσπασμα από πίνακα τυχαίων αριθμών για να εξηγήσουμε τη χρήση τους.

1 5 2 8 4 2 9 8 6 7 5 2 3 8 3 0 1 2 8 5 9 8 6 8
 4 8 6 4 5 0 3 2 5 7 9 4 6 4 8 7 5 5 6 8 7 2 0 6
 8 6 0 9 2 2 6 2 0 6 3 8 1 6 7 4 7 1 2 7 6 9 8 3
 9 0 9 1 2 3 9 1 1 9 0 4 1 6 3 1 2 5 9 6 6 5 3 2
 0 6 4 1 2 9 7 8 5 1 1 5 4 9 0 1 2 6 8 8 4 5 7 6

7 0 4 4 5 6 3 1 0 3 1 9 6 6 1 9 6 9 6 5 2 7 2 8
 9 6 2 4 9 5 0 8 3 9 5 5 1 5 3 3 1 9 5 0 9 8 2 6
 5 8 7 4 2 9 7 1 4 9 6 2 1 3 2 9 9 0 8 0 9 3 6 6
 4 4 4 9 3 4 0 5 4 6 9 6 8 3 7 4 5 1 0 3 3 9 6 4
 9 8 7 1 1 9 5 1 8 6 8 2 9 5 9 3 8 4 1 3 0 2 6 2

5 6 6 8 3 6 5 4 5 5 4 6 1 3 5 8 8 3 6 1 6 6 6 7
 0 8 6 2 5 5 1 9 1 5 7 5 7 7 7 4 6 5 0 3 4 2 7 8
 8 5 3 9 6 7 4 9 0 2 3 0 4 4 5 5 6 7 1 0 5 9 3 4
 7 0 6 6 7 8 1 2 9 7 4 2 8 1 5 4 1 0 5 7 4 2 1 7
 3 9 8 8 3 8 4 6 7 4 2 1 1 3 7 4 3 6 8 5 5 2 1 9

1 7 7 0 3 9 9 4 0 5 7 6 1 2 9 8 6 5 9 7 7 4 1 8
 1 2 3 3 1 5 5 9 4 3 2 8 5 7 8 8 6 0 6 4 8 0 3 5
 0 7 3 4 0 9 5 6 2 6 8 1 4 1 1 4 9 9 9 6 7 2 0 5
 9 6 1 7 0 0 1 0 8 9 7 1 7 6 5 3 3 7 8 0 0 3 5 8
 6 3 2 2 2 8 7 6 4 2 4 5 9 7 8 7 1 1 6 8 5 7 7 4
 0 0 0 4 0 0 2 5 1 0 1 5 0 0 0 9 2 3 1 5 4 3 1 5

Έστω ότι επιθυμούμε να καθορίσουμε ένα δείγμα 5% από ένα πληθυσμό $N = 5000$ μονάδες, οι οποίες έχουν απαριθμηθεί από το 1 μέχρι και το 5.000. Από τις 5.000 μονάδες εκλέγουμε, με τη βοήθεια του πίνακα των τυχαίων αριθμών, 250 μονάδες.

Ξεκινάμε από κάποιο σημείο του πίνακα και προχωράμε, είτε κατακόρυφα, είτε οριζόντια, είτε διαγώνια, είτε κατά άλλη διαφορετική έννοια, μέχρι να πάρουμε 250 τετραψήφιους αριθμούς.

Στην απλή τυχαία δειγματοληψία, όλες οι στατιστικές μονάδες και όλα τα δείγματα μεγέθους n έχουν την ίδια πιθανότητα να επιλεγούν. Η μέθοδος αυτή μπορεί να εφαρμοσθεί κάνοντας χρήση πινάκων τυχαίων αριθμών, οι οποίοι κατασκευάζονται με την βοήθεια ηλεκτρονικών υπολογιστών, κατά τρόπο ώστε κάθε ένας από τους αριθμούς 0, 1, 2, ..., 9 να έχει πιθανότητα $1/10$ να αποτελέσει ψηφίο του πίνακα.

Παράδειγμα: Οι παρακάτω αριθμοί αποτελούν τμήμα ενός πίνακα τυχαίων αριθμών. Η επιλογή του τμήματος έγινε κατά τρόπο αυθαίρετο. Οι αριθμοί παρουσιάζονται σε στήλες για οπτικούς μόνο λόγους, δηλαδή τα διαστήματα που φαίνονται ουσιαστικά δεν υπάρχουν. Έτσι, διαβάζοντας τους αριθμούς οριζόντια, παίρνουμε 1,7,6,2,3,4,7,4,4,1,2,7 κλπ.

17623	47441	27821	91845	01654	50375	23941	44848
45054	58410	92081	97624	73750	68343	40727	81203
73700	58730	06111	64486	64163	22132	22896	14305

Έστω ότι ο αριθμός που μας ενδιαφέρει αποτελείται από $N=500$ καταστήματα και ότι θέλουμε να πάρουμε ένα δείγμα μεγέθους $n=10$ καταστημάτων. Χρησιμοποιώντας τους παραπάνω τυχαίους αριθμούς, ποια καταστήματα θα συμπεριληφθούν στο δείγμα;

Λύση: Κατ' αρχή, αριθμούμε τα καταστήματα του πληθυσμού με τους αριθμούς 000, 001, 002, ..., 499. Επειδή οι αριθμοί αυτοί είναι τριψήφιοι, θα αρχίσουμε να επιλέγουμε από τον παραπάνω πίνακα τριάδες διαδοχικών ψηφίων, απορρίπτοντας τις τριάδες που σχηματίζουν αριθμούς μεγαλύτερους του 499, μέχρι να συμπληρώσουμε τους 10 αριθμούς του δείγματος.

Διαβάζοντας τους 10 αριθμούς κάθετα, παίρνουμε τους εξής 10 αποδεκτούς αριθμούς: 176, 450, 234, 005, 127, 092, 006, 081, 111, 247.

Το τυχαίο δείγμα θα αποτελείται από τα 10 καταστήματα που έχουν αυτούς σαν αύξοντες αριθμούς. Ας σημειωθεί ότι οι αριθμοί 737, 545, 744, 841, 873, 821, 918, 976 και 644 απορρίφθηκαν επειδή είναι μεγαλύτερη του 499, ενώ ο αριθμός 450 απορρίφθηκε τη δεύτερη φορά που σχηματίστηκε.

Ανάλυση: Στη συνέχεια για να γίνει μια εκτενής ανάλυση της απλής τυχαίας δειγματοληψίας θα γίνει η χρήση ενός υποθετικού παραδείγματος. Έστω, ότι έχουμε έναν πληθυσμό αποτελούμενο από 12 μονάδες όπως φαίνεται στον επόμενο πίνακα.

A/A	Μονάδες	Εισόδημα(€)
1	A	1.300
2	B	6.300
3	Γ	3.100
4	Δ	2.000
5	E	3.600
6	Z	2.200
7	H	1.800
8	Θ	2.700
9	I	1.500
10	K	900
11	Λ	4.800
12	M	1.900
ΣΥΝΟΛΙΚΟ ΕΙΣΟΔΗΜΑ 32.100		
ΜΕΣΟ ΕΙΣΟΔΗΜΑ 2.675		

Πίνακας 1 - Εισοδήματα ενός υποθετικού πληθυσμού 12 μονάδων

Σκοπός μας, είναι να γίνει μια εκτίμηση για το μέσο εισόδημα του πληθυσμού από ένα δείγμα. Για την επιλογή του δείγματος υπάρχουν διάφοροι διαθέσιμοι τρόποι, όπως η κληρωτίδα και η χρήση πινάκων τυχαίων αριθμών. Ο πιο κλασσικός τρόπος επιλογής αποτελεί η κληρωτίδα, τρόπος που θα χρησιμοποιήσουμε στο υποθετικό πείραμα.

Έστω λοιπόν, ότι χρησιμοποιούνται δώδεκα χαρτάκια ίσου μεγέθους στα οποία αναγράφονται τα ονόματα των μονάδων (A,B,Γ,...,M) και τοποθετούνται σε μία γυάλα. Στη συνέχεια ανακατεύουμε καλά και επιλέγουμε δύο χαρτάκια με τέτοιο τρόπο που να μην επηρεάζονται οι επιλεγόμενες μονάδες. (Πέτρος Α. Κιόχος,1993)

5.6 Δειγματοληψία χωρίς επανατοποθέτηση

Η απλή τυχαία δειγματοληψία χωρίς επανάθεση (simple random sampling without replacement) είναι αποτελεσματική σε απαριθμημένους πληθυσμούς.

Στη δειγματοληψία χωρίς επανατοποθέτηση έχουμε ένα πληθυσμό n στοιχείων απ' όπου επιλέγουμε διαδοχικά r στοιχεία παίρνοντάς τα ένα-ένα χωρίς να τα επαναθέτουμε στον πληθυσμό. Το δείγμα επομένως που προκύπτει είναι μία διάταξη των n ανά r .

$$\Delta_n^r = (n)_r = (n)(n-1)\dots(n-r+1)$$

Στην πράξη δεν είναι πάντα εύκολος ο σχηματισμός όλων των δυνατών διακεκριμένων δειγμάτων, ιδίως όταν το μέγεθος N του πληθυσμού είναι πολύ μεγάλο. Στην θέση της διαδικασίας που περιγράφηκε παραπάνω ακολουθείται η εξής εναλλακτική διαδικασία: Μια μονάδα του πληθυσμού επιλέγεται τυχαία, δηλαδή με τρόπο που εξασφαλίζει την ίδια πιθανότητα επιλογής σε κάθε μια από τις N μονάδες του πληθυσμού. Η μέθοδος που χρησιμοποιείται για τον σκοπό αυτό αποτελείται από τα εξής βήματα: (α) Αντιστοιχίζουμε σε κάθε μονάδα του πληθυσμού έναν αριθμό από το 1 μέχρι το N και (β) διαλέγουμε μια σειρά n τυχαίων αριθμών από το 1 μέχρι το N με την βοήθεια πινάκων τυχαίων αριθμών. Εύκολα μπορεί να διαπιστωθεί ότι κάθε ένα από τα δυνατά $\binom{N}{n}$ διακεκριμένα τυχαία δείγματα έχει πιθανότητα $1/\binom{N}{n}$ να επιλεγεί. Πράγματι, ας θεωρήσουμε ένα τέτοιο δείγμα, δηλαδή ένα τέτοιο σύνολο n διακεκριμένων μονάδων. Κατά την πρώτη δοκιμή, η πιθανότητα ότι κάποια από τις n συγκεκριμένες μονάδες του δείγματος θα επιλεγεί από τον δοθέντα πληθυσμό είναι $\frac{n}{N}$.

Στην δεύτερη δοκιμή η πιθανότητα ότι κάποια από τις $n-1$ απομένουσες μονάδες του δείγματος θα επιλεγεί από τις $N-1$ απομένουσες μονάδες του πληθυσμού είναι $\frac{n-1}{N-1}$

Επομένως, η πιθανότητα με την οποία και οι n μονάδες του επιθυμούμενου δείγματος θα επιλεγούν είναι

$$\frac{n}{N} \cdot \frac{n-1}{N-1} \cdot \frac{n-2}{N-2} \dots \frac{1}{N-n+1} = \frac{n!(N-n)!}{N!} = \frac{1}{\binom{N}{n}}$$

5.7 Δειγματοληψία με επανατοποθέτηση

Στην περίπτωση αυτή η επανατοποθέτηση επιτρέπει το ίδιο στοιχείο να επαναλαμβάνεται στο δείγμα. Έτσι το δείγμα που προκύπτει είναι μία διάταξη με επανάληψη των n ανά r . Θα υπάρχουν επομένως

$$A_n^r = n^r$$

τυχαία δείγματα στη δειγματοληψία με επανατοποθέτηση.

5.8 Σύγκριση των μεθόδων

Η δειγματοληψία χωρίς επανατοποθέτηση δίνει περισσότερες πληροφορίες για τον πληθυσμό, όμως η δειγματοληψία με επανατοποθέτηση αναλύεται ευκολότερα.

Όταν το δείγμα είναι μικρό σε σχέση με τον πληθυσμό οι δύο τύποι δειγματοληψίας είναι προσεγγιστικά ισοδύναμοι.

Από τις δύο αυτές μεθόδους η απλή δειγματοληψία χωρίς επανατοποθέτηση είναι προτιμότερη, διότι δίνει εκτιμήσεις με μικρότερο δειγματοληπτικό σφάλμα από το αντίστοιχο της δειγματοληψίας με επανατοποθέτηση.

Κεφάλαιο 6: Συστηματική δειγματοληψία

6.1 Εισαγωγή

Σε αρκετές περιπτώσεις ο πληθυσμός είναι τόσο μεγάλος που η απαρίθμηση όλων των μελών του, η καταχώρησή τους σε πίνακες και η επιλογή δείγματος καθίσταται εξαιρετικά επίπονη και πρακτικά αδύνατη. Όταν όμως, υπάρχει μια ενημερωμένη καταχώρηση των μελών ενός πληθυσμού και ο πληθυσμός παρουσιάζει ομοιογένεια στη σύστασή του, τότε συχνά χρησιμοποιείται η απλή μέθοδος της συστηματικής δειγματοληψίας. Συγκεκριμένα, η μέθοδος αυτή απαιτεί πρώτα να επιλέξουμε τυχαία το πρώτο μέλος του δείγματος και τα επόμενα μέλη να επιλέγονται από τον ίδιο κατάλογο μετά από κάθε σταθερό διάστημα διαδοχικών μονάδων.

Για παράδειγμα, αν ένας ερευνητής θέλει να επιλέξει από ένα αλφαβητικό κατάλογο 500 περίπου μαθητών ένα δείγμα 50 μαθητών, επιλέγει τυχαία ένα αριθμό μεταξύ 1 και 10 (π.χ. το 7) και στη συνέχεια επιλέγει κάθε δέκατο μαθητή (17° , 27° , ..., 497°). Η συστηματική δειγματοληψία είναι μια πολύ απλή και εύχρηστη μέθοδος που επιτρέπει την επιλογή ενός δείγματος ακόμα και όταν τα στοιχεία του πληθυσμού που είναι καταγεγραμμένα σε ένα κατάλογο δεν έχουν αριθμηθεί. Ωστόσο, για να μπορεί να χρησιμοποιηθεί η συστηματική δειγματοληψία στη θέση της τυχαίας δειγματοληψίας, το σύνολο των μελών του πληθυσμού πρέπει να είναι κατανομημένα με εντελώς τυχαίο τρόπο. Αλλιώς, υπάρχει ο κίνδυνος το συστηματικό δείγμα να είναι μεροληπτικό.

Για να έχουμε μια συστηματική καταγραφή των μονάδων του πληθυσμού θα πρέπει:

- Να έχουμε διαθέσιμες όλες τις μονάδες του πληθυσμού που θέλουμε να μελετήσουμε σε μορφή κατάστασης ή σε αρχείο του υπολογιστή. Για παράδειγμα, excel.
- Να αριθμηθούν όλες οι μονάδες του πληθυσμού ώστε επιλέγοντας τυχαία τόσους αριθμούς όσο το επιθυμητό μέγεθος του δείγματος, να γνωρίζουμε σε ποιες μονάδες αντιστοιχούν αυτοί οι αριθμοί.

Με αυτόν τον τρόπο έχουμε μια συστηματική καταγραφή των μονάδων του πληθυσμού. Το πλαίσιο διευκολύνει σημαντικά την οργάνωση της δειγματοληψίας.

Στην περίπτωση που ένα δειγματοληπτικό πλαίσιο είναι διαθέσιμο σε μορφή λίστας, μπορούμε να εφαρμόσουμε συστηματική δειγματοληψία. Η συστηματική δειγματοληψία διαφέρει από την απλή τυχαία δειγματοληψία, στο ότι κάθε στοιχείο έχει την ίδια πιθανότητα να επιλεγεί, ενώ κάθε δείγμα δεν έχει την ίδια πιθανότητα να επιλεγεί.

6.2 Ορισμός της συστηματικής δειγματοληψίας

Ορισμός: Η συστηματική δειγματοληψία (**systematic sampling**) είναι η δειγματοληπτική τεχνική, η οποία εισάγει ένα συστηματικό στοιχείο στην διαδικασία επιλογής των μονάδων του πληθυσμού. Δηλαδή, επιλέγουμε με τυχαίο τρόπο έναν αριθμό k και οι ομάδες του δείγματος προκύπτουν με βάση αυτόν και το διάστημα της δειγματοληψίας, που είναι ο λόγος του μεγέθους του πληθυσμού N προς το μέγεθος του πληθυσμού n .

$$k = \frac{n}{N}$$

6.2.1 Ορισμός Συστηματικού δείγματος

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

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

6.2.2 Περιγραφή της Διαδικασίας

Έστω ο ότι ο υπό μελέτη πληθυσμός έχει μέγεθος N που είναι αριθμημένος από το 1 έως το N και διαιρούμε το σύνολο των στοιχείων του δειγματοληπτικού πλαισίου, δηλαδή το μέγεθος του πληθυσμού, με το μέγεθος του δείγματος n .

Επιλέγουμε με τυχαίο τρόπο έναν αριθμό x ανάμεσα στο ένα και το N/n . Το άτομο που αντιστοιχεί στο συγκεκριμένο αριθμό είναι το πρώτο στοιχείο του δείγματος. Στη συνέχεια επιλέγεται το άτομο με αύξοντα αριθμό $x + N/n$, μετά το άτομο με αριθμό $x + 2N/n$.

(Κ. Κουτσικόπουλος, 2002)

Για να είναι αντιπροσωπευτικός ο πληθυσμός, θα προηγηθεί η ταξινόμηση της λίστας ως προς εύστοχα χαρακτηριστικά για την στρωματοποίηση του πληθυσμού.

Η συστηματική δειγματοληψία είναι μια στρατηγική που διευκολύνει την επιλογή των δειγματοληπτικών μονάδων, γι' αυτό επιλέγεται πιο εύκολα από τους ερευνητές.

Παράδειγμα:

Παρακάτω θα δείξουμε πως η μέθοδος αυτή, απαιτεί πρώτα να επιλέξουμε τυχαία το πρώτο μέλος του δείγματος και τα επόμενα μέλη να επιλέγονται από τον ίδιο κατάλογο μετά από κάθε σταθερό διάστημα διαδοχικών μονάδων.

Αν $N = 100$ (ολόκληρος ο πληθυσμός)

Και θέλουμε δείγμα $n = 20$

$N/n = 5$ άρα χρειαζόμαστε έναν ανά 5.

Επιλέγουμε τυχαία έναν αριθμό από το 1 έως το 5 (π.χ. το 4)

Ξεκινάμε από τον 4^ο και επιλέγουμε κάθε 5^ο (9^ο,14^ο,19^ο,24^ο,...,99^ο)

1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100

6.2.3 Εφαρμογή της συστηματικής δειγματοληψίας

Σε όλες τις μονάδες του ερευνώμενου πληθυσμού αν υπάρχει το κατάλληλο δειγματοληπτικό πλαίσιο, δηλαδή κατάλογος στον οποίο είναι να είναι καταχωρημένες όλες οι δειγματοληπτικές μονάδες ώστε να δίνεται η ίδια ευκαιρία επιλογής, σε όλες τις μονάδες του ερευνώμενου πληθυσμού.

Σε άγνωστους πληθυσμούς, χωρίς δειγματοληπτικό πλαίσιο. Για παράδειγμα αν θέλουμε να εκτιμήσουμε τον αριθμό των ατόμων που εισέρχονται σε μια Δημόσια Υπηρεσία ημερησίως.

6.3 Πλεονεκτήματα της συστηματικής δειγματοληψίας

- Η κατανομή των επιλεγμένων μονάδων, η οποία είναι ελεγχόμενη και κανονική, και όλα τα μέρη του πληθυσμού τόσο φαίνονται όσο και αντιπροσωπεύονται στο δείγμα.
- Το σχέδιο κατανομής των μονάδων δειγματοληψίας είναι απλό και μπορεί συνήθως να αναγνωριστεί και να εντοπισθεί εύκολα, διευκολύνοντας έτσι τις διεργασίες στο πεδίο έρευνας.
- Η κανονική κατανομή των μονάδων δειγματοληψίας διευκολύνει τη χαρτογράφηση των πληροφοριών.
- Η συστηματική δειγματοληψία πιθανό να είναι πιο ακριβής από την απλή τυχαία δειγματοληψία και στην πράξη θα πρέπει να περιμένουμε να είναι τόσο ακριβής όσο το αντίστοιχο στρωματοποιημένο τυχαίο δείγμα με μια μονάδα ανά στρώμα (και με κάθε μονάδα δειγματοληψίας μέσα σε ένα στρώμα διαστάσεων όσο το πλέγμα δειγματοληψίας).

6.4 Μειονεκτήματα της συστηματικής δειγματοληψίας

Η συστηματική δειγματοληψία παρουσιάζει μεροληψία:

- Όταν τα στοιχεία του πληθυσμού είναι διατεταγμένα σύμφωνα με κάποιο κριτήριο οπότε το σημείο εκκίνησης επιδρά συστηματικά στο αποτέλεσμα. Για παράδειγμα, αν τα άτομα είναι διατεταγμένα σύμφωνα με το εισόδημα τους το δείγμα που θα ξεκινήσει από το 1ο άτομο και θα περιλαμβάνει το 51^ο, 101^ο κλπ, θα δώσει χαμηλότερη εκτίμηση του μέσου εισοδήματος από το δείγμα που θα ξεκινήσει από το 50^ο.

- Όταν το πλαίσιο έχει περιοδικότητα υποπολλαπλάσια του μεγέθους του δείγματος. Έτσι π.χ. αν θέλουμε να πάρουμε τυχαίο δείγμα 30 ημερών με συστηματική δειγματοληψία τότε θα καταλήξουμε να παρατηρούμε την ίδια μέρα κάθε μήνα.
- Το συστηματικό δείγμα δεν είναι ένα τυχαίο δείγμα του πληθυσμού, με συνέπεια να δυσκολεύει στην εκτίμηση της διασποράς και βασικών παραμέτρων.

6.5 Το μέγεθος του δείγματος

Το μέγεθος του δείγματος που επιθυμούμε να μελετήσουμε καθορίζει και τον τρόπο επιλογής των ατόμων. Έτσι αν θέλουμε να σχηματίσουμε ένα δείγμα που να αντιπροσωπεύει το $1/k$ του πληθυσμού, αρκεί από την λίστα των ατόμων του πληθυσμού να επιλέγουμε ένα κάθε k άτομα.

$$k = N / n$$

N = τα άτομα του πληθυσμού και

n =το μέγεθος του δείγματος.

Για την επιλογή του δείγματος με **συστηματικό τρόπο** από πληθυσμούς για τους οποίους δεν υπάρχει κατάλογος των ατόμων (σχεδόν όλοι οι φυσικοί πληθυσμοί) η διαδικασία διαφέρει:

Χωρίζουμε το πληθυσμό μας σε ισομεγέθεις πρωτογενείς μονάδες.

Κάθε πρωτογενής μονάδα χωρίζεται με τον ίδιο τρόπο σε k δευτερογενείς.

Από τις k μονάδες μίας πρωτογενούς μονάδος επιλέγεται τυχαία μία, π.χ. πίνακας τυχαίων αριθμών, και στη συνέχεια σε κάθε πρωτογενή μονάδα επιλέγεται η δευτερογενής με τις ίδιες συντεταγμένες. Το σύνολο των δευτερογενών αυτών μονάδων αποτελεί ένα συστηματικό δείγμα ίσο με το $1/k$ του πληθυσμού στόχου.

Συμπεραίνουμε ότι μικρές και πολυπληθείς πρωτογενείς μονάδες οδηγούν σε μια καλύτερη "κάλυψη" του πληθυσμού, αυτό σημαίνει ότι διατρέχουμε όλον το πληθυσμό με λεπτομερή τρόπο και δεν αφήνουμε ακάλυπτες μεγάλες περιοχές.

6.6 Η Διασπορά της Εκτιμήτριας του Μέσου

Αν $N = nk$, τα k δυνατά 1-ανά- k συστηματικά δείγματα είναι οι στήλες του παρακάτω πίνακα

Δείγμα					
1	2	k
Y_1	Y_2	...	Y_i	...	Y_k
Y_{k+1}	Y_{k+2}	...	Y_{k+i}	...	Y_{2k}
.
.
.
$Y_{(n-1)k+1}$	$Y_{(n-1)k+2}$...	$Y_{(n-1)k+i}$...	Y_{nk}

Με την συστηματική δειγματοληψία, ο πληθυσμός χωρίζεται σε k σύνθετες μονάδες, και η διαδικασία επιλογής ενός συστηματικού δείγματος ισοδυναμεί με την διαδικασία τυχαίας επιλογής μιας σύνθετης μονάδας.

Επομένως, ένα 1-ανά- k συστηματικό τυχαίο δείγμα n μονάδων από ένα πληθυσμό μεγέθους $N = nk$, είναι ένα απλό τυχαίο δείγμα μιας σύνθετης μονάδας από τον πληθυσμό με (σύνθετες) μονάδες τις k στήλες του παραπάνω πίνακα.

Είναι σημαντικό να πούμε, ότι η διαίρεση του αρχικού πληθυσμού των $N = n \cdot k$ μονάδων σε k ομάδες έχει ως αποτέλεσμα την δυνατότητα έκφρασης της διασποράς σ^2 του πληθυσμού μέσω της διασποράς "μεταξύ" των k ομάδων και της διασποράς μέσα στις ομάδες.

Ίσως να γίνει πιο αντιληπτό αν πούμε ότι το j στοιχείο του i συστηματικού δείγματος συμβολισθεί με y_{ij} , δηλαδή,

$$y_{ij} = Y_{(j-1)k+i}$$

Τότε, αν μ είναι η μέση τιμή του πληθυσμού και $\bar{X}_n^{(i)}$ ο μέσος του i δείγματος, ισχύει ότι

$$\begin{aligned} (N-1)\sigma^2 &= \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \mu)^2 = \\ &= n \sum_{i=1}^k (\bar{X}_n^{(i)} - \mu)^2 + \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{X}_n^{(i)})^2 \end{aligned} \quad (1)$$

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

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

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

$$\sigma_{\bar{X}_n^*}^2 = \frac{1}{k} \sum_{i=1}^k \left[\bar{X}_n^{(i)} - E(\bar{X}_n^{(i)}) \right]^2 = \frac{1}{k} \sum_{i=1}^k (\bar{X}_n^{(i)} - \mu)^2 \quad (3)$$

Είναι δηλαδή, ο πρώτος προσθετός ανάλογος της διασποράς μεταξύ των συστηματικών δειγμάτων. Τότε, μπορεί να αποδειχθεί το εξής θεώρημα.

ΘΕΩΡΗΜΑ

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

$$\sigma_{\bar{X}_n^*}^2 = \left[(N-1)\sigma^2 - k(n-1)\sigma_w^2 \right] / N \quad (4)$$

ΑΠΟΔΕΙΞΗ

Προφανώς η σχέση (1) είναι ισοδυναμεί με την σχέση

$$(N-1)\sigma^2 = nk\sigma_{\bar{X}_n^*}^2 + k(n-1)\sigma_w^2,$$

Η οποία οδηγεί στην σχέση (4)

ΠΟΡΙΣΜΑ

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

$$\sigma_w^2 > \sigma^2$$

ΑΠΟΔΕΙΞΗ

Ισχύει ότι $V(\bar{X}_n) = \frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)$. Τότε, από την σχέση (4), ισχύει ότι

$$\begin{aligned} V(\bar{X}_n^*) &< V(\bar{X}_n) \\ \Leftrightarrow \frac{N-1}{N} \sigma^2 - \frac{k(n-1)}{N} \sigma_w^2 &< \frac{N-n}{N} \frac{\sigma^2}{n} \\ \Leftrightarrow k(n-1) \sigma_w^2 &> \left(N-1 - \frac{N-n}{N}\right) \sigma^2 \equiv k(n-1) \sigma^2 \end{aligned}$$

Δηλαδή, η συστηματική δειγματοληψία οδηγεί σε μικρότερο τυπικό σφάλμα, αν η διασπορά μέσα στο δείγμα είναι μεγαλύτερη από τη διασπορά ολόκληρου του πληθυσμού. Επομένως, μεγαλύτερη ακρίβεια επιτυγχάνεται, αν οι μονάδες του δείγματος έχουν μεγαλύτερη ετερογένεια σε σχέση με αυτήν που έχουν οι μονάδες όλου του πληθυσμού.

Παρατήρηση

Είναι προφανές ότι στην περίπτωση εκτίμησης του ποσοστού p των μονάδων του πληθυσμού που ανήκουν σε μια κατηγορία A , εφαρμόζεται η παραπάνω θεωρία, αν η παράμετρος μ αντικατασταθεί από την παράμετρο p και ο μέσος $\bar{X}_n^{(i)}$ αντικατασταθεί από την στατιστική συνάρτηση $\hat{p}^{(i)}$, όπου $\hat{p}^{(i)}$ είναι το ποσοστό των μονάδων του i συστηματικού δείγματος που ανήκουν στην κατηγορία A . Είναι προφανές ότι ο πληθυσμός θα αποτελείται από μονάδες της μορφής $y_{ij} = \begin{cases} 1, & \text{αν η } j \text{ μονάδα του } i \text{ δείγματος ανήκει στην } A \\ 0, & \text{διαφορετικά} \end{cases}$

Κεφάλαιο 7: Δειγματοληψία κατά στρώματα ή στρωματοποιημένη δειγματοληψία

7.1 Εισαγωγή

Στην ενότητα αυτή, θα δούμε ότι η στρωματοποιημένη δειγματοληψία σε σύγκριση με την απλή τυχαία δειγματοληψία μας δίνει ένα περισσότερο αντιπροσωπευτικό δείγμα. Η απλή τυχαία δειγματοληψία είναι ένα αμερόληπτο σχέδιο επιλογής των μονάδων του δείγματος άρα και με μειωμένη ακρίβεια. Η αύξηση του μεγέθους του δείγματος θα μειώσει τα δειγματοληπτικά σφάλματα αλλά θα αυξήσει τις δαπάνες της δειγματοληψίας. Ο τρόπος λοιπόν για να έχουμε ένα δείγμα σταθερού μεγέθους, είναι να χρησιμοποιήσουμε τη μέθοδο της δειγματοληψίας κατά στρώματα.

Στη στρωματοποιημένη δειγματοληψία χωρίζουμε το συνολικό πληθυσμό N που εξετάζουμε, σε k ορισμένες κατηγορίες, τα στρώματα, έτσι ώστε $N_1 + N_2 + \dots + N_k = N$. Σε κάθε τέτοια κατηγορία επιλέγονται όσον το δυνατόν πιο ομοιογενή στοιχεία του πληθυσμού έτσι ώστε να πετύχουμε τη μεγαλύτερη δυνατή διαφοροποίηση μεταξύ των κατηγοριών αυτών.

Από κάθε στρώμα επιλέγουμε ένα απλό τυχαίο δείγμα, με δειγματικά μεγέθη προσδιορισμένα εκ των προτέρων, και συνενώνουμε τα επιμέρους αυτά δείγματα σε ένα, για ολόκληρο τον πληθυσμό.

Η στρωματοποιημένη δειγματοληψία είναι πολύ διαδεδομένη μέθοδος δειγματοληψίας και υπαγορεύεται από πολλούς λόγους, οι σημαντικότεροι των οποίων είναι:

1. Αυξάνει την ομοιογένεια των επί μέρους ομάδων (υποπληθυσμών)
2. Η δυνατότητα που παρέχει, για εκτίμηση παραμέτρων εντός των στρωμάτων
3. Λόγοι Διοικητικοί, παραδείγματος χάριν, εάν ένα δείγμα αναφέρεται σε μια ευρεία γεωργική περιοχή, οι τοπικές αρχές μπορούν να εποπτεύσουν την έρευνα που εκτείνεται στην διοικητική τους αρμοδιότητα
4. Λόγοι καθαρά στατιστικοί, παραδείγματος χάριν, μπορεί να δειχθεί ότι διαφορά μεταξύ των μέσων των στρωμάτων στον πληθυσμό δεν συμβάλουν στο σφάλμα δειγματοληψίας της εκτιμήσεως \bar{x}_{st} δηλαδή το σφάλμα δειγματοληψίας της εκτιμήσεως \bar{x}_{st} απορρέει αποκλειστικά από διακυμάνσεις μεταξύ των μονάδων δειγματοληψίας που ανήκουν στο ίδιο στρώμα.

7.2 Ορισμοί και συμβολισμοί

Ο αριθμός των στρώματων στα οποία διαιρείται ο υπό μελέτη πληθυσμός θα συμβολίζεται με k . Το N αντιπροσωπεύει τον συνολικό αριθμό των στοιχειωδών μονάδων (ή μονάδων δειγματοληψίας) στον δειγματοληπτούμενο πληθυσμό και το N_h θα αντιπροσωπεύει τον αντίστοιχο αριθμό στο h στρώμα έτσι ώστε:

$$N = \sum_{h=1}^k N_h = N_1 + N_2 + \dots + N_k$$

Αντίστοιχα το μέγεθος του δείγματος που λαμβάνεται από το h στρώμα θα συμβολίζεται με n_h , και είναι το συνολικό μέγεθος του δείγματος που λαμβάνεται από όλα τα στρώματα.

$$n = \sum_{h=1}^k n_h$$

Από αυτή τη διαδικασία προκύπτει ότι σε σύγκριση με την απλή τυχαία δειγματοληψία ο δείκτης h που προστίθεται δείχνει ένα συγκεκριμένο στρώμα και ότι ο δείκτης i θα καθορίζει την αντίστοιχη μονάδα δειγματοληψίας, αλλά οι μονάδες θεωρούνται και αριθμούνται χωριστά μέσα σε κάθε στρώμα και όχι ενιαία όπως στην απλή τυχαία δειγματοληψία. Έτσι η τιμή του χαρακτηριστικού X της i μονάδας δειγματοληψίας στο h στρώμα θα συμβολίζεται με X_{hi} .

Συνεπώς, εάν ο δείκτης h συμβολίζει το h στρώμα τότε το συνολικό μέγεθος του χαρακτηριστικού για όλες τις μονάδες του στρώματος θα είναι:

$$X_h = \sum_{i=1}^{N_h} X_{hi}$$

Όταν το άθροισμα λαμβάνεται για όλες τις μονάδες ενός πληθυσμού τότε θα συμβολίζεται με κεφαλαίο γράμμα ενώ το άθροισμα επί όλων των μονάδων του δείγματος θα συμβολίζεται με ένα μικρό γράμμα. Παραδείγματος χάριν, X_{hi} είναι η τιμή κάποιου χαρακτηριστικού για την i μονάδα στον πληθυσμό μέσα στο h στρώμα, ενώ x_{hi} είναι η τιμή του χαρακτηριστικού για την i μονάδα του δείγματος από το h στρώμα, ώστε:

$$X_h = \sum_{i=1}^{N_h} X_{hi} \text{ και } x_h = \sum_{i=1}^{n_h} x_{hi}$$

Ομοίως:

$$X = \sum_{h=1}^k X_h = \sum_{h=1}^k \sum_{i=1}^{N_h} X_{hi}$$

αντιπροσωπεύει το άθροισμα επί όλων των μονάδων του δειγματοληπτούμενου πληθυσμού, και:

$$x = \sum_{h=1}^k x_h = \sum_{h=1}^k \sum_{i=1}^{n_h} x_{hi}$$

Είναι η συνολική τιμή του χαρακτηριστικού για όλες τις μονάδες του δείγματος, μεγέθους:

$$n = \sum_{h=1}^k n_h$$

Το \bar{X} αντιπροσωπεύει τον μέσο του δειγματοληπτούμενου πληθυσμού, και:

$$\bar{X} = \frac{X}{N}$$

Ο μέσος μέσα στο h στρώμα συμβολίζεται με:

$$\bar{X}_h = \frac{X_h}{N_h}$$

Και ο μέσος ενός δείγματος με n_h μονάδες από το h θα συμβολίζεται με:

$$\bar{x}_h = \frac{x_h}{n_h}$$

Η διακύμανση του χαρακτηριστικού μεταξύ των στοιχειωδών μονάδων μέσα στο h στρώμα θα συμβολίζεται με σ_h^2 , έτσι ώστε:

$$\sigma_h^2 = \frac{\sum_{i=1}^{N_h} (X_{hi} - \bar{X}_h)^2}{N_h}$$

και

$$s_h^2 = \frac{N_h}{N_h - 1} \sigma_h^2 = \frac{\sum_{i=1}^{N_h} (X_{hi} - \bar{X}_h)^2}{N_h - 1}$$

η αντίστοιχη σχετική διακύμανση είναι:

$$v_h^2 = \frac{S_h^2}{\bar{X}_h^2}$$

Επίσης, η διακύμανση του δειγματοληπτούμενου πληθυσμού χωρίς να λαμβάνουμε υπόψη τα επί μέρους στρώματα είναι:

$$\sigma^2 = \frac{\sum_{h=1}^k \sum_{i=1}^{N_h} (X_{hi} - \bar{X})^2}{N}, \text{ ή } s^2 = \frac{\sum_{h=1}^k \sum_{i=1}^{N_h} (X_{hi} - \bar{X})^2}{N-1}$$

Και η σχετική διακύμανση είναι :

$$v^2 = \frac{s^2}{\bar{X}^2},$$

όπου είναι ακριβώς η ίδια διακύμανση και σχετική διακύμανση όπως στην απλή τυχαία δειγματοληψία.

7.3 Χρήσεις της τυχαίας στρωματοποιημένης δειγματοληψίας

Στρωματοποιημένη τυχαία δειγματοληψία χρησιμοποιείται όταν ο ερευνητής επιθυμεί να τονίσει μια συγκεκριμένη υποομάδα του πληθυσμού. Η τεχνική αυτή είναι χρήσιμη σε αυτές τις έρευνες, δεδομένου ότι εξασφαλίζει την παρουσία των βασικών υποομάδων εντός του δείγματος.

Οι ερευνητές χρησιμοποιούν επίσης στρωματοποιημένη τυχαία δειγματοληψία, όταν θέλουν να τηρούν τις υπάρχουσες σχέσεις ανάμεσα σε δύο ή περισσότερα υποσύνολα. Με μια απλή τυχαία τεχνική δειγματοληψίας, ο ερευνητής δεν είναι σίγουρος για το αν είναι οι υποομάδες που θέλει να παρατηρήσει εκπροσωπούνται εξίσου ή αναλογικά μέσα στο δείγμα.

Με στρωματοποιημένη δειγματοληψία, ο ερευνητής μπορεί να έχει αντιπροσωπευτικά δείγματα ακόμη και τα μικρότερα και πιο απρόσιτες υποομάδες του πληθυσμού.

Αυτό επιτρέπει στον ερευνητή να δοκιμάσει τα σπάνια άκρα του συγκεκριμένου πληθυσμού.

Επειδή αυτή η τεχνική έχει υψηλή στατιστική ακρίβεια, αυτό σημαίνει επίσης ότι απαιτείται ένα μικρό μέγεθος του δείγματος που μπορεί να σώσει πολύ χρόνο, χρήμα και προσπάθεια από τους ερευνητές.

7.4 Πλεονεκτήματα της στρωματοποιημένης δειγματοληψίας

- 1) Μειώνει σημαντικά την εμφάνιση παραπλανητικών δειγμάτων. Δηλαδή, με τη στρωματοποιημένη δειγματοληψία έχουμε μεγαλύτερη ακρίβεια χωρίς να επιβάλλεται αύξηση του μεγέθους του δείγματος.
- 2) Βελτιώνει την ακρίβεια της εκτίμησης όπου υπάρχει μεγάλη απόκλιση στις μεταβλητές όταν έχει σχέση με πληθυσμιακά χαρακτηριστικά, χωρίς να αποτρέπει τον σχηματισμό των ομοιογενών μονάδων.
- 3) Εξασφαλίζει την αντιπροσώπευση συγκεκριμένων μονάδων του πληθυσμού, που διαφορετικά θα μπορούσαν να αποκλειστούν διότι αντιπροσωπεύουν μια μικρή αναλογία του συνολικού πληθυσμού.

7.5 Μειονεκτήματα της στρωματοποιημένης δειγματοληψίας

- 1) Η διαδικασία επιλογής του δείγματος είναι πιο σύνθετη.
- 2) Η στρωματοποίηση προϋποθέτει περισσότερη προκαταρκτική πληροφόρηση γύρω από το σύνολο του πληθυσμού.
- 3) Το σύνολο του πληθυσμού πρέπει να ορίζεται και να κατατάσσεται βάσει των χαρακτηριστικών της στρωματοποίησης.

7.6 Εκτίμηση Μέσης Τιμής Στρωματοποιημένου Πληθυσμού

Σύμφωνα με τους ορισμούς τα δεδομένα του προβλήματος συνοψίζονται στον πίνακα

Στρώμα	Πληθυσμός				Δείγμα		
	Μέγεθος στρώματος	Μέση τιμή	Διασπορά		Μέγεθος απλού τυχ. δείγματος	Μέσος	Διασπορά
1	N_1	μ_1	σ_1^2		n_1	\bar{X}_{n_1}	$\frac{\sigma_1^2}{n_1} \left(1 - \frac{n_1}{N_1}\right)$
2	N_2	μ_2	σ_2^2		n_2	\bar{X}_{n_2}	$\frac{\sigma_2^2}{n_2} \left(1 - \frac{n_2}{N_2}\right)$
.
.
.
I	N_i	μ_i	σ_i^2		n_i	\bar{X}_{n_i}	$\frac{\sigma_i^2}{n_i} \left(1 - \frac{n_i}{N_i}\right)$
.
.
.
K	N_k	μ_k	σ_k^2		n_k	\bar{X}_{n_k}	$\frac{\sigma_k^2}{n_k} \left(1 - \frac{n_k}{N_k}\right)$

Εδώ $\bar{X}_{n_i} = \sum_{j=1}^{n_i} X_j^{(i)}$, $i=1,2,\dots,k$, όπου $X_j^{(i)}$ είναι η j μονάδα του i δείγματος.

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

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

Τα αντίστοιχα δειγματικά μεγέθη συνδέονται με μια παρόμοια σχέση:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^k n_i \bar{X}_{ni},$$

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

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

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

στην οποία οι μέσοι των δειγμάτων από τα διάφορα στρώματα σταθμίζονται με τους συντελεστές βαρύτητας $\frac{N_i}{N}, i=1,2,\dots,k$ των στρωμάτων.

Προφανώς,

$$\hat{\mu}_n = \bar{X}_n, \text{ αν } \frac{n_i}{n} = \frac{N_i}{N}, i=1,2,\dots,k$$

ή ισοδύναμα

$$\text{αν } \frac{n_i}{N_i} = \frac{n}{N} \Leftrightarrow f_i = f, i=1,2,\dots,k$$

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

Θεώρημα:

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

$$V(\hat{\mu}_n) = j$$

Απόδειξη:

Αμεροληψία:

Ισχύει εδώ ότι :

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

Αλλά, οι στατιστικές συναρτήσεις $\bar{X}_{n_1}, \bar{X}_{n_2}, \dots, \bar{X}_{n_k}$ ως μέσοι απλών

τυχαίων δειγμάτων είναι αμερόληπτες εκτιμήτριες των μέσων τιμών $\mu_1, \mu_2, \dots, \mu_k$ αντίστοιχα των (υπο)πληθυσμών, από τους οποίους τα δείγματα αυτά επελέγησαν. Άρα,

$$E(\bar{X}_{n_i}) = \mu_i$$

και, επομένως,

$$E(\hat{\mu}_n) = \frac{1}{N} \sum_{i=1}^k N_i \mu_i = \mu.$$

Η τελευταία σχέση αποδεικνύει ότι η $\hat{\mu}_n$ είναι αμερόληπτη εκτιμήτρια του μ .

$$\begin{aligned} \text{Ισχύει επίσης ότι } V(\hat{\mu}_n) &= V\left(\frac{1}{N} \sum_{i=1}^k N_i \bar{X}_{n_i}\right) \\ &= \frac{1}{N^2} \left\{ \sum_{i=1}^k N_i^2 V(\bar{X}_{n_i}) + 2 \sum_{1 \leq i < j \leq k} N_i N_j \text{Cov}(\bar{X}_{n_i}, \bar{X}_{n_j}) \right\}. \end{aligned}$$

Αλλά, $\text{Cov}(\bar{X}_{n_i}, \bar{X}_{n_j}) = 0$ τα k απλά τυχαία δείγματα είναι αμοιβαία ανεξάρτητα. Άρα, ισχύει η σχέση

$$V(\hat{\mu}_n) = \frac{1}{N^2} \sum_{i=1}^k N_i^2 V(\bar{X}_{n_i}),$$

Η οποία οδηγεί στην απόδειξη του θεωρήματος.

Έστω $y_j^{(i)}$ η j μονάδα του πληθυσμού που ανήκει στο i στρώμα.

Τότε ισχύει το εξής πόρισμα.

Πόρισμα: Η στατιστική συνάρτηση

$$\hat{Y}_{st} = N \hat{\mu}_n$$

Είναι αμερόληπτη εκτιμήτρια του συνολικού μεγέθους

$$y = \sum_{i=1}^k \sum_{j=1}^{n_i} y_j^{(i)}$$

Και ισχύει ότι

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

Απόδειξη : Ισχύει προφανώς ότι

$$y = \sum_{i=1}^k \sum_{j=1}^{n_i} y_j^{(i)} = \sum_{i=1}^k N_i \mu_i = N \mu.$$

Επομένως, $E(\hat{Y}_{st}) = NE(\hat{\mu}_n) = N\mu = y$ και $V(\hat{Y}_{st}) = N^2 V(\hat{\mu}_n)$.

Επειδή από κάθε στρώμα επιλέγεται ένα απλό τυχαίο δείγμα, έπεται ότι η στατιστική συνάρτηση

$$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (X_j^{(i)} - \bar{X}_{n_i})^2$$

Είναι η αμερόληπτη εκτιμήτρια της διασποράς σ_i^2 του i στρώματος.

(Εδώ, $X_j^{(i)}$ είναι η j παρατήρηση του i δείγματος).

Κατά συνέπεια, ισχύει το εξής

$$S_{\mu_n}^2 = \sum_{i=1}^k \left(\frac{N_i}{N}\right)^2 \frac{S_i^2}{n_i} \left(1 - \frac{n_i}{N_i}\right)$$

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

7.7 Το μέγεθος του δείγματος

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

Αν το δειγματοληπτικό κόστος ανά μονάδα είναι το ίδιο σε όλα τα στρώματα και οι διασπορές των στρωμάτων δεν διαφέρουν σημαντικά, τα μεγέθη n_1, n_2, \dots, n_k συνηθίζεται να επιλέγονται έτσι ώστε

$$\frac{n_i}{N_i} = \frac{n}{N}, i = 1, 2, \dots, k.$$

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

$$n_i = n \frac{N_i}{N}, i = 1, 2, \dots, k,$$

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

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

σημαίνει ότι στρώματα με μεγαλύτερη διακύμανση από άλλα πρέπει να εκπροσωπούνται από μεγαλύτερο τμήμα του δείγματος, για να αυξηθεί η ακρίβεια των εκτιμήσεων. Υποθέτοντας ότι το δειγματοληπτικό κόστος ανά μονάδα είναι το ίδιο για όλα τα στρώματα, αποδεικνύεται ότι η $V(\hat{\mu}_n)$ γίνεται ελάχιστη αν τα n_1, n_2, \dots, n_k επιλεγούν έτσι ώστε

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

Ο σχεδιασμός αυτός είναι γνωστός ως βέλτιστος **καταμερισμός του n με σταθερό κόστος ανά δειγματοληπτική μονάδα (optimum allocation with constant cost per unit)** ή **καταμερισμός κατά Neyman (Neyman allocation)**.

7.7.1 Στρωματοποίηση με αναλογικό καταμερισμό

Το μέγεθος του δείγματος της, για κάθε στρώμα, σε αυτήν την τεχνική είναι ανάλογη με το μέγεθος του πληθυσμού του στρώματος όταν αντιμετωπίζεται από το σύνολο του πληθυσμού. Αυτό σημαίνει ότι το κάθε στρώμα έχει το ίδιο κλάσμα δειγματοληψίας. (Castillo, Joseph Juan -2009)

Ο τύπος προσδιορισμού του δείγματος στο στρώμα i είναι :

$$n_i = n N_i / N$$

Στο σημείο αυτό θα δώσουμε ένα παράδειγμα αναλογικού επιμερισμού του δείγματος. Είναι μια μέθοδος που χρησιμοποιείται όταν οι διακυμάνσεις των διαφόρων στρωμάτων (ομάδων) δεν διαφέρουν σημαντικά και το κόστος επιλογής των μονάδων δεν διαφέρει, επίσης, από στρώμα σε στρώμα.

Το σημαντικό πράγμα που πρέπει να θυμόμαστε σε αυτήν την τεχνική είναι να χρησιμοποιηθεί το ίδιο κλάσμα δειγματοληψίας για κάθε στρώμα, ανεξάρτητα από τις διαφορές στο μέγεθος του πληθυσμού των στρωμάτων.

Παράδειγμα:

Επιμερίζουμε αναλογικά δείγμα $n=100$ οικογενειών, που επιλέγεται από ένα σύνολο 10.000 οικογενειών, διακρινόμενο σε τρία στρώματα με κριτήριο το εισόδημά τους:

A/A	Στρώμα	Αριθμός Οικογενειών
1	N_1	3.200
2	N_2	3.800
3	N_3	3.000
Σύνολο $N = 10.000$		

Ο αριθμός των μονάδων δείγματος που θα ληφθεί από κάθε στρώμα είναι ο εξής:

$$n_1 = \frac{N_1}{N} n = \frac{3.200}{10.000} 100 = 32$$

$$n_2 = \frac{N_2}{N} n = \frac{3.800}{10.000} 100 = 38$$

$$n_3 = \frac{N_3}{N} n = \frac{3.000}{10.000} 100 = 30$$

ΣΥΝΟΛΟ: 100

Η κατά στρώματα δειγματοληψία παρέχει τη δυνατότητα λήψης πληροφοριών και για κάθε στρώμα, είναι όμως περισσότερο δαπανηρή από τις προηγούμενες διότι απαιτεί ξεχωριστά πλαίσια για κάθε ένα απ' αυτά.

Παράδειγμα:

Υποθέστε ότι σε μια επιχείρηση υπάρχει το ακόλουθο προσωπικό:

-άντρες, μόνιμη απασχόληση: 90

-άντρες, μερική απασχόληση: 18

-γυναίκες, μόνιμη απασχόληση: 9

-γυναίκες, μερική απασχόληση: 63

Σύνολο: 180

και καλούμαστε να επιλέξουμε ένα δείγμα 40 ατόμων, που στρωματοποιείται σύμφωνα με τις παραπάνω κατηγορίες.

Το πρώτο βήμα είναι να υπολογιστεί το ποσοστό κάθε στρώματος.

Άντρες, μόνιμη απασχόληση $(90/180) \times 100 = 50\%$

Άντρες, μερική απασχόληση $(18/180) \times 100 = 10\%$

Γυναίκες, μόνιμη απασχόληση $(9/180) \times 100 = 5\%$

Γυναίκες, μερική απασχόληση $(63/180) \times 100 = 35\%$

Αυτό μας λέει ότι από τα 40 άτομα του δείγματός μας,

50% πρέπει να είναι άντρες, μόνιμη απασχόληση (20)

10% πρέπει να είναι άντρες, μερική απασχόληση (4)

5% πρέπει να είναι γυναίκες μόνιμη απασχόληση (2)

35% πρέπει να είναι γυναίκες με μερική απασχόληση (14)

Από τα παραπάνω γίνεται φανερό ότι, για να εφαρμοστεί η δειγματοληψία κατά στρώματα, πρέπει να είναι εκ των προτέρων γνωστή η κατανομή των χαρακτηριστικών του πληθυσμού που σχετίζονται με τη μεταβλητή που αποτελεί το θέμα της έρευνας.

7.7.2 Δειγματοληψία κατά στρώματα δυσανάλογη

Με δυσανάλογη διαστρωμάτωση, το κλάσμα της δειγματοληψία μπορεί να διαφέρει από το ένα στρώμα στο άλλο.

Η ακρίβεια του σχεδιασμού μπορεί να είναι πολύ καλή ή πολύ κακή, ανάλογα με το πώς είναι τα σημεία δειγματοληψίας που διατίθενται για στρώματα. Όταν οι διακυμάνσεις διαφέρουν μεταξύ των στρωμάτων, η δυσανάλογη διαστρωμάτωση μπορεί να προσφέρει μεγαλύτερη ακρίβεια από ότι η ανάλογη διαστρωμάτωση.

Με δυσανάλογη διαστρωμάτωση, ο ερευνητής μπορεί να μεγιστοποιήσει την ακρίβεια στην έρευνα που μελετά. Ωστόσο, τα κέρδη ως προς την ακρίβεια δεν μπορούν να καταλογισθούν σε άλλα μέτρα έρευνας.

7.8 Βέλτιστος καταμερισμός

Είναι σαφές, ότι μπορούμε να έχουμε μεγαλύτερη αποτελεσματικότητα στην στρωματοποίηση, αν η φύση των δεδομένων επιτρέπει μεγαλύτερη υποδιαίρεση του στρώματος σε υπο-στρώματα, τα οποία μπορούν να διαιρεθούν σε μικρότερες ομάδες.

Σε περιπτώσεις, που οι διασπορές των στρωμάτων διαφέρουν, λόγω διαφοράς της διακύμανσης, θα πρέπει ο λόγος n_i/N_i να είναι ανάλογος της τυπικής απόκλισης σ_i του στρώματος.

Η ακρίβεια των εκτιμήσεων στα στρώματα με μεγαλύτερη διακύμανση επιτυγχάνετε αν αυξήσουμε το μέγεθος του δείγματος. Αυτό συνεπάγεται κάποιο δειγματοληπτικό κόστος. Αν το κόστος ανά μονάδα είναι το ίδιο για όλα τα στρώματα, αποδεικνύεται ότι η $V(\hat{\mu}_n)$ (διακύμανση) γίνεται ελάχιστη εάν τα n_1, n_2, \dots, n_k επιλεγούν έτσι ώστε :

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

Παράδειγμα:

Έστω ότι ένας πληθυσμός 12 στοιχειωδών μονάδων διαιρείται σε 3 στρώματα, με 5 μονάδες το πρώτο στρώμα, με 3 μονάδες το δεύτερο στρώμα και 4 μονάδες το τρίτο στρώμα όπως φαίνεται στον επόμενο πίνακα:

Στρώμα 1	Στρώμα 2	Στρώμα 3
$X_{11} = 6$	$X_{21} = 9$	$X_{31} = 20$
$X_{12} = 10$	$X_{22} = 18$	$X_{32} = 26$
$X_{13} = 2$	$X_{23} = 12$	$X_{33} = 16$
$X_{14} = 4$		$X_{34} = 26$
$X_{15} = 8$		

Να βρεθούν:

- i. Οι μέσοι των στρωμάτων
- ii. Ο μέσος του πληθυσμού

Λύση

i.

Για να υπολογιστούν οι μέσοι των στρωμάτων θα πρέπει να χρησιμοποιηθεί ο τύπος:

$$\bar{X}_h = \frac{X_h}{N_h}, \text{ όπου } X_h = \sum_{i=1}^{N_h} X_{hi} .$$

Συνεπώς για το Στρώμα 1, έχουμε:

$$\bar{X}_1 = \frac{X_1}{N_1} = \frac{\sum_{i=1}^{N_1} X_{1i}}{N_1} = \frac{X_{11} + X_{12} + X_{13} + X_{14} + X_{15}}{N_1} = \frac{6+10+2+4+8}{5} = 6$$

για το Στρώμα 2, έχουμε:

$$\bar{X}_2 = \frac{X_2}{N_2} = \frac{\sum_{i=1}^{N_2} X_{2i}}{N_2} = \frac{X_{21} + X_{22} + X_{23}}{N_2} = \frac{9+18+12}{3} = 13$$

και για το Στρώμα 3, έχουμε:

$$\bar{X}_3 = \frac{X_3}{N_3} = \frac{\sum_{i=1}^{N_3} X_{3i}}{N_3} = \frac{X_{31} + X_{32} + X_{33} + X_{34}}{N_3} = \frac{20+26+16+26}{4} = 22$$

ii.

Ο μέσος του πληθυσμού υπολογίζεται από:

$$\bar{X} = \frac{X}{N}, \text{ όπου } X = \sum_{h=1}^k X_h = \sum_{h=1}^k \sum_{i=1}^{N_h} X_{hi}$$

Άρα έχουμε:

$$\begin{aligned} \bar{X} &= \frac{X}{N} = \frac{\sum_{h=1}^k X_h}{N} = \frac{\sum_{h=1}^k \sum_{i=1}^{N_h} X_{hi}}{N} = \\ &= \frac{X_{11} + X_{12} + X_{13} + X_{14} + X_{15} + X_{21} + X_{22} + X_{23} + X_{31} + X_{32} + X_{33} + X_{34}}{N} = \\ &= \frac{6+10+2+4+8+9+18+12+20+26+16+26}{12} = 13,08 \end{aligned}$$

Πιθανές πηγές πληροφοριών για τη διάρθρωση του πληθυσμού μπορεί να είναι κάποιες προηγούμενες μελέτες που καλύπτουν παρόμοια θέματα ή τα αποτελέσματα μιας έρευνας πιλότου (pilot study).

Αν όμως, δεν υπάρχουν οι απαιτούμενες πληροφορίες σχετικά με τη διαστρωμάτωση του πληθυσμού, τότε πρέπει να αποφεύγεται η χρησιμοποίηση της μεθόδου αυτής αφού, σε αυτή την περίπτωση, είναι δυνατό να επιτευχθεί ένας ίσος βαθμός ακριβείας χρησιμοποιώντας την τυχαία δειγματοληψία.

7.9 Σύγκριση Απλής Τυχαίας Και Στρωματοποιημένης Δειγματοληψίας

Με την στρωματοποιημένη δειγματοληψία, έχουμε μεγάλη στατιστική ακρίβεια σε σύγκριση με την απλή τυχαία δειγματοληψία. Αυτό οφείλεται στο γεγονός ότι η μεταβλητότητα εντός των στρωμάτων του είναι χαμηλότερη σε σύγκριση με τις παραλλαγές όταν ασχολείται με το σύνολο του πληθυσμού.

Αν η στρωματοποίηση γίνει με παράγοντες που δεν συσχετίζονται με το εξεταζόμενο μέγεθος, τα χαρακτηριστικά των επιμέρους στρωμάτων δεν θα διαφέρουν από τα χαρακτηριστικά του πληθυσμού και το όφελος της στρωματοποίησης θα είναι ανύπαρκτο.

Αντίθετα, στην απλή τυχαία δειγματοληψία, το όφελος από την ανάλογη δειγματοληψία ισούται με τη διαφορά των διακυμάνσεων.

Ο διαχωρισμός των στρωμάτων που γίνεται ανεξάρτητα από το χαρακτηριστικό X , θα δώσει διακύμανση μεταξύ των μέσων στρωμάτων ίση με το μηδέν. Ενώ, η απλή τυχαία δειγματοληψία δίνει διακύμανση των εκτιμήσεων ίση με τη διακύμανση της ανάλογη κατά στρώματα δειγματοληψίας.

Συνεπώς, πρέπει να μεγιστοποιηθεί η διακύμανση μεταξύ των στρωμάτων και να ελαχιστοποιηθεί η διακύμανση μέσα στα στρώματα.

Από άποψη ακρίβειας, η στρωματοποιημένη δειγματοληψία έχει ως αποτέλεσμα μικρότερη διασπορά για την εκτιμήτρια της μέσης τιμής του πληθυσμού. Δεν αληθεύει όμως ότι οποιοδήποτε στρωματοποιημένο δείγμα δίνει μικρότερη διασπορά από ένα απλό τυχαίο δείγμα.

Κεφάλαιο 8: Δειγματοληψία Κατά Ομάδες – Πολυσταδιακή Δειγματοληψία

8.1 Εισαγωγή

Οι προηγούμενες μέθοδοι δειγματοληψίας βασίζονται στην επιλογή των μελών ενός πληθυσμού ένα-προς-ένα. Ωστόσο, συχνά οι συμμετέχοντες ανήκουν ήδη σε ομάδες-συστάδες και ο ερευνητής μπορεί να επιλέξει με τυχαίο τρόπο ομάδες αντί υποκείμενα. Μεταξύ δύο δειγματοληπτικών τεχνικών που χρησιμοποιούν ισομεγέθη δείγματα, προτιμάτε αυτή που οδηγεί στο μικρότερο τυπικό σφάλμα, στη μέγιστη δυνατή ακρίβεια και με το ελάχιστο δυνατό κόστος.

Έτσι οδηγούμαστε στη δειγματοληψία κατά ομάδες. Η τεχνική αυτή μας δίνει τη δυνατότητα να μελετήσουμε τον πληθυσμό που εξετάζουμε ως σύνολο από ορισμένες ξεχωριστές ομάδες δειγματοληπτικών μονάδων.

Εξηγώντας πιο θεωρητικά τη διαδικασία της δειγματοληψίας κατά ομάδες θα υποθέσουμε ότι παίρνουμε ένα δείγμα αποτελούμενο από n ομάδες από ένα πλήθος N ομάδων, όπου i έχει πλήθος υπό-ομάδων (στοιχείων) M_i , $i=1,2,\dots,N$. Μια τέτοια δειγματοληψία απαιτεί κατάλληλη ομαδοποίηση του πληθυσμού σε ομάδες και είναι χρήσιμη όταν ο πληθυσμός είναι ιεραρχικώς διατεταγμένος.

Σε αυτό το κεφάλαιο θα μιλήσουμε για την δειγματοληψία κατά ομάδες σε ένα στάδιο (single-stage cluster sampling). Η μέθοδος αυτή ενδείκνυται όταν δεν υπάρχουν δειγματοληπτικά πλαίσια (κατάλογοι) του πληθυσμού ή είναι σχεδόν αδύνατο να φτιαχτούν. Εάν δεν υπάρχει υπό-δειγματοληψία από τις ήδη n επιλεγείσες στο δείγμα ομάδες, τότε το τελικό μας δείγμα είναι η ένωση όλων των επιλεγμένων μονάδων $\sum_{i=1}^n M_i$ και το άθροισμα της

παρατήρησης y πάνω στις υπό - μονάδες θα είναι $\sum_{i=1}^n y_i$ όπου $i=1,2,\dots,n$. (N. A. Τσερπές – Φ. Αλεβίζος, ΠΑΤΡΑ 1996)

8.2 Ορισμός Δειγματοληψίας κατά ομάδες

8.2.1 Δειγματοληψία κατά ομάδες σε ένα στάδιο

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

8.2.2 Δειγματοληψία κατά ομάδες σε πολλά στάδια

Δειγματοληψία κατά ομάδες σε πολλά στάδια (multi-stage cluster sampling) ονομάζεται η δειγματοληπτική τεχνική, όπου κατόπιν του πρώτου σταδίου επιλέγονται δείγματα μικρότερων ομάδων με τελικό στάδιο την επιλογή του δείγματος στοιχειωδών μονάδων.

Στο παρακάτω παράδειγμα ο πληθυσμός που πρόκειται να μελετήσουμε είναι δυνατό να θεωρηθεί ως σύνολο που αποτελείται από ορισμένες ξεχωριστές ομάδες δειγματοληπτικών μονάδων.

Έστω ότι έχουμε ένα πληθυσμό 2000 νοικοκυριών από τους οποίους θέλουμε να επιλέξουμε δείγμα ίσο με 200 νοικοκυριά.

Έχουμε δυο τρόπους να επιλέξουμε το δείγμα:

1. Αν υπάρχει ένας κατάλογος των 2000 νοικοκυριών, επιλέγουμε ένα απλό τυχαίο δείγμα μεγέθους 200 από τον κατάλογο.
2. Διαιρούμε την πόλη σε 400 περιοχές 50 νοικοκυριών και επιλέγουμε ένα απλό τυχαίο δείγμα τεσσάρων περιοχών και περιλαμβάνουμε στο δείγμα μας όλα τα νοικοκυριά που ανήκουν σε αυτές (σύνολο 200).

ή

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

Η πρώτη περίπτωση (1) οδηγεί σε μικρότερο τυπικό σφάλμα από αυτό της δεύτερης περίπτωσης (2). Η πρώτη περίπτωση όμως συνεπάγεται μεγαλύτερο κόστος γιατί το δείγμα είναι περισσότερο απλωμένο γεωγραφικά, και επομένως το κόστος μετάβασης για τον εντοπισμό των μονάδων που επιλέγονται και το κόστος εποπτείας είναι υψηλότερο.

Στη δεύτερη περίπτωση το δείγμα είναι λιγότερο απλωμένο μιας και οι δειγματοληπτικές μονάδες είναι σύνθετες.

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

Ακόμα και στον δεύτερο τρόπο της δεύτερης (2) περίπτωσης, το δεύτερο στάδιο της δειγματοληψίας διεξάγεται σε γεωγραφικά μικρότερη έκταση, με αποτέλεσμα τη μείωση του κόστους διεξαγωγής και εποπτείας.

Από τη στιγμή που οι ομάδες επιλέγονται, ο ερευνητής μπορεί να περιλάβει στο δείγμα όλες τις στοιχειώδεις μονάδες του αρχικού πληθυσμού που ανήκουν σ'αυτές όπως στον πρώτο τρόπο της περίπτωσης (2), ή μπορεί να επιλέξει ένα δείγμα μικρότερων σύνθετων ομάδων ή και στοιχειωδών μονάδων από τις αρχικές σύνθετες μονάδες όπως στον δεύτερο τρόπο της περίπτωσης (2).

Το συμπέρασμα είναι ότι η δειγματοληψία κατά ομάδες μπορεί να γίνει σε ένα ή σε πολλά στάδια.

8.3 Χρήσεις της δειγματοληψίας κατά ομάδες

Η μέθοδος της κατά ομάδες τυχαίας δειγματοληψίας (cluster random sampling), χρησιμοποιείται στις εξής περιπτώσεις:

α. όταν οι στατιστικές μονάδες του πληθυσμού είναι γεωγραφικά η μια μακριά από την άλλη, οπότε χρήση μίας από τις δύο προηγούμενες μεθόδους θα είναι δαπανηρή, αν οι συνεντεύξεις γίνονται από κοντά.

β. όταν δεν υπάρχει ένας πλήρης κατάλογος των στατιστικών μονάδων του πληθυσμού και η κατάρτισή του θα κόστιζε πολύ. Σε αυτές τις περιπτώσεις, ο ερευνητής μπορεί να κατατάξει τις στατιστικές μονάδες του πληθυσμού σε μικρές ομάδες (clusters). Αρχικά να πάρει ένα τυχαίο δείγμα ομάδων (π.χ με την μέθοδο της απλής τυχαίας δειγματοληψίας) και στην συνέχεια να κάνει απογραφή σε κάθε μία από τις επιλεγείσες ομάδες του δείγματος.

8.4 Πλεονεκτήματα της δειγματοληψίας κατά ομάδες

Η δειγματοληψία κατά ομάδες έχει τα ακόλουθα πλεονεκτήματα:

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

8.5 Μειονεκτήματα της δειγματοληψίας κατά ομάδες

Η δειγματοληψία κατά ομάδες έχει τα ακόλουθα μειονεκτήματα:

- 1) Η αβεβαιότητα που υπάρχει για το εάν είναι μη αντιπροσωπευτικές οι μονάδες του συνολικού πληθυσμού αποτελεί ένα από τα βασικότερα μειονεκτήματα της μεθόδου.
- 2) Η πολύ-σταδιακή δειγματοληψία είναι χρήσιμη μόνο όταν έχουμε μια ιεραρχική διάταξη των μονάδων του πληθυσμού.

8.6 Εκτίμηση της Μέσης Τιμής

Από τα προηγούμενα, προκύπτει ότι η δειγματοληψία κατά ομάδες, προϋποθέτει ότι ο πληθυσμός διαιρείται σε υποπληθυσμούς.

Μερικοί από αυτούς, εκπροσωπούνται στο δείγμα είτε εξ ολοκλήρου (ένα στάδιο) είτε εν μέρει μέσω κάποιου δείγματος (δύο ή περισσότερα στάδια).

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

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

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

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

και, επομένως, ο αρχικός πληθυσμός είναι η ένωση $u_1 \cup u_2 \cup \dots \cup u_M$. Προφανώς,

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

συμβολίζει το μέσο μέγεθος των ομάδων αυτών.

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

$$\bar{n} = \sum_{i=1}^m n_i / m$$

αντιστοιχεί με το μέσο μέγεθος των m υποδειγμάτων

$$\bar{U}_i = \sum_{j=1}^{n_i} U_{ij} / n_i$$

είναι ο μέσος του i υποδείγματος, $i = 1, 2, \dots, m$

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

όπου το $\sum_{j=1}^{n_i} (U_{ij} - \bar{U}_i)^2$ είναι το άθροισμα τετραγωνικών αποκλίσεων των παρατηρήσεων του i δείγματος από τον μέσο του και το

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

είναι η διασπορά του i δείγματος $i = 1, 2, \dots, m$

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

$$\mu = \sum_{i=1}^M \sum_{j=1}^{N_i} y_{ij} / N$$

μέση τιμή του αρχικού πληθυσμού (μέση τιμή του υπό εξέταση χαρακτηριστικού ανά στοιχειώδη μονάδα)

$$\mu_i = \sum_{j=1}^{N_i} y_{ij} / N_i$$

μέση τιμή της i ομάδας (του i υποπληθυσμού),

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

$$\bar{\mu} = \sum_{i=1}^M \sum_{j=1}^{N_i} y_{ij} / M = \sum_{i=1}^M N_i \mu_i / M$$

μέση τιμή του χαρακτηριστικού ανά σύνθετη μονάδα (ανά ομάδα).

Προφανώς,

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

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

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

Στην περίπτωση αυτή, τα μεγέθη N_i' των ομάδων που θα επιλεγούν θα είναι ίσα με \bar{N} , δηλαδή, $N_i' = \bar{N}, i = 1, 2, \dots, M$.

Η στατιστική συνάρτηση

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

ή, ισοδύναμα, η στατιστική συνάρτηση

$$\hat{\mu}_c = \sum_{i=1}^m \bar{U}_i / m = \sum_{i=1}^m \mu_i / m$$

είναι η αμερόληπτη εκτιμήτρια της μέσης τιμής μ με διασπορά

$$\begin{aligned} V(\mu_c) &\cong \frac{M-m}{M} \frac{1}{m} \frac{1}{\bar{N}^2} \sum_{i=1}^M \left(\sum_{j=1}^{\bar{N}} y_{ij} - \bar{\mu} \right)^2 / (M-1) \\ &= \frac{M-m}{M} \frac{1}{m} \sum_{i=1}^M (\mu_i - \mu)^2 / (M-1) \end{aligned}$$

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

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

Η στατιστική συνάρτηση

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

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

Η στατιστική συνάρτηση

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

είναι αμερόληπτη εκτιμήτρια του συνολικού μεγέθους $y = M \bar{\mu} = M \bar{N} \mu$

Προφανώς,

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

2) $N_i \neq N_j, i \neq j$ (Ανισομεγέθεις ομάδες).

Τότε, ισχύει το εξής θεώρημα:

Η στατιστική συνάρτηση

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

είναι μια αμερόληπτη εκτιμήτρια της παραμέτρου μ με διασπορά

$$V(\hat{\mu}_c) \cong \frac{M-m}{M} \frac{1}{m} \frac{1}{\bar{N}^2} \frac{1}{M-1} \sum_{i=1}^m \left(\sum_{j=1}^{N'_i} y_{ij} - N'_i \mu \right)^2.$$

Πόρισμα: Η στατιστική συνάρτηση

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

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

8.7 Περίπτωση Ποσοστών

Έστω πληθυσμός μεγέθους N , ο οποίος διαιρείται σε M ομάδες (clusters) μεγέθους N_1, N_2, \dots, N_M . Έστω p το ποσοστό των μονάδων του πληθυσμού, οι οποίες ανήκουν σε μια κατηγορία A . Το πρόβλημα που θα αντιμετωπισθεί είναι η εκτίμηση του p με βάση ένα δείγμα που θα επιλεγεί με απλή δειγματοληψία κατά ομάδες.

Έστω $N_A^{(i)}$ ο αριθμός των μονάδων της ομάδας i που ανήκουν στην κατηγορία A και $P_i = N_A^{(i)} / N_i$ το ποσοστό των μονάδων της i ομάδας που ανήκουν στην κατηγορία A , $i = 1, 2, \dots, M$.

Προφανώς, ισχύει ότι:

$$p = \sum_{i=1}^M N_i p_i / \sum_{i=1}^M N_i$$

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

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

Περιοριζόμενοι στην περίπτωση δειγματοληψίας σε ένα στάδιο μόνο, μπορούμε να εφαρμόσουμε τη θεωρία της προηγούμενης ενότητας για την εκτίμηση του p , αν θεωρήσουμε τις αντιστοιχίες $\mu_i \leftrightarrow p_i, \mu \leftrightarrow p$. Συγκεκριμένα, οδηγούμεθα στα εξής συμπεράσματα.

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

Στην περίπτωση αυτή,

$$N'_i = \bar{N}, i = 1, 2, \dots, m$$

οπότε,

$$\hat{p}_i = X^{(i)} / \bar{N}, \quad i = 1, 2, \dots, m$$

και ισχύει το εξής θεώρημα.

Θεώρημα: Η στατιστική συνάρτηση:

$$\hat{p}_c = \sum_{i=1}^m \hat{p}_i / m$$

είναι αμερόληπτη εκτιμήτρια της παραμέτρου p με διασπορά

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

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

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

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

Η στατιστική συνάρτηση

$$\hat{p}_c = \sum_{i=1}^m X^{(i)} / \sum_{i=1}^m N'_i$$

είναι μια αμερόληπτη εκτιμήτρια της παραμέτρου P με διασπορά

$$V(\bar{p}_c) = \frac{M-m}{M} \frac{1}{m\bar{N}^2} \frac{1}{M-1} \sum_{i=1}^M (N_A^{(i)} - N_i p)^2$$

Η στατιστική συνάρτηση

$$S_{\hat{p}_c}^2 = \frac{M-m}{M} \frac{1}{m\bar{N}^2} \frac{1}{m-1} \sum_{i=1}^m \left(X^{(i)} - N_i' \hat{p}_c \right)^2$$

ή η υπολογιστικά περισσότερο προσφερόμενη μορφή της

$$S_{\hat{p}_c}^2 = \frac{M-m}{M} \frac{1}{m\bar{N}^2} \frac{1}{m-1} \left(\sum_{i=1}^m X^{(i)2} - 2\hat{p}_c \sum_{i=1}^m X^{(i)} N_i' + \hat{p}_c^2 \sum_{i=1}^m N_i'^2 \right)$$

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

Παράδειγμα :

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

Τάξη(i)	Αριθμός φοιτητών υπέρ ($X^{(i)}$)
1	3
2	5
3	2
4	3
5	4
6	1
7	4
8	2
9	6
10	1
11	2
Σύνολο	33

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

Λύση :

Προφανώς, $m=11$, $M=110$, $N_i=6, i=1,2,\dots,110$.

Τότε έχουμε:

$$\hat{p}_c = \sum_{i=1}^{11} \hat{p}_i / 11 = \sum_{i=1}^{11} X^{(i)} / (6 \cdot 11) = 0.5$$

Επίσης, προκύπτει ότι

$$S_{\hat{p}_c}^2 = \frac{110-11}{110} \frac{1}{11} \frac{1}{10} \sum_{i=1}^{11} (\hat{p}_i - 0.5)^2 = 0.00591$$

Άρα,

$$S_{\hat{p}_c} = 0.0769$$

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

$$S_{\hat{p}}^2 = \frac{\hat{p}(1-\hat{p})}{n-1} \frac{N-n}{N} = \frac{(0.5)(0.5)}{65} \frac{660-66}{660} = 0.00346$$

και, επομένως, σε τυπικό σφάλμα ίσο με

$$S_{\hat{p}} = 0.0588$$

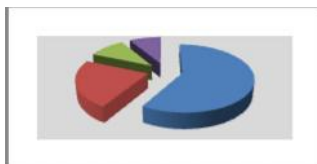
Παρατήρηση: Η 1 ανά-k συστηματική δειγματοληψία είναι μια μορφή δειγματοληψίας κατά ομάδες σε ένα στάδιο. Οι ομάδες είναι τα k δυνατά συστηματικά δείγματα ($M=k$) και το απλό τυχαίο δείγμα ομάδων που επιλέγεται είναι μεγέθους 1 ($m=1$).

8.8 Σύγκριση της Δειγματοληψίας Κατά Ομάδες με την Στρωματοποιημένη Δειγματοληψία

Η διαφορά της απλής δειγματοληψίας κατά ομάδες από την απλή στρωματοποιημένη δειγματοληψία έγκειται στο ότι, στην πρώτη, μόνο ορισμένα από τα στρώματα εκπροσωπούνται στο δείγμα, ενώ, στην δεύτερη, εκπροσωπούνται όλα τα στρώματα και πάντα μέσω απλών τυχαίων υποδειγμάτων. Στη δειγματοληψία κατά στρώματα επιβάλλεται να έχουμε μέσα σε κάθε στρώμα τη μεγαλύτερη δυνατή ομοιογένεια και μεταξύ των στρωμάτων τη μεγαλύτερη δυνατή ανομοιογένεια (ώστε κάθε στρώμα να είναι ομοιογενές και ξεχωριστή ομάδα του πληθυσμού), γιατί στη μέθοδο αυτή εκπροσωπούνται όλα τα στρώματα και πάντα μέσω απλών τυχαίων υποδειγμάτων.

Παρατηρούμε ότι ενώ στη στρωματοποιημένη δειγματοληψία παίρνουμε ένα μέρος του τελικού δείγματος από κάθε ομάδα (στρώμα), εδώ το δείγμα λαμβάνεται κατά ομάδες (σμάρια) ενώνοντας όλα τα στοιχεία των ομάδων που επιλέξαμε στο δείγμα στο πρώτο στάδιο.

Κεφάλαιο 9: Έρευνα



Έρευνα για τις διατροφικές συνήθειες των μαθητών
του 19^{ου} Δημοτικού σχολείου Πατρών

9.1 Εισαγωγή

Η εργασία αυτή είναι το αποτέλεσμα της έρευνας που διεξήχθη για την πτυχιακή εργασία «Τεχνικές δειγματοληπτικών ερευνών» στο τμήμα της λογιστικής του Τ.Ε.Ι Πατρών και πραγματοποιήθηκε στο 19^ο Δημοτικό σχολείο Πατρών στην περιοχή Αγίας Βαρβάρας.

Αντικείμενο μελέτης είναι οι διατροφικές συνήθειες των μαθητών αλλά και οι εξωσχολικές τους δραστηριότητες.

9.2 Ταυτότητα της έρευνας

Η έρευνα πραγματοποιήθηκε στις 6 Απριλίου 2012 στο 19^ο Δημοτικό σχολείο του Δήμου Πατρών. Η συλλογή των πληροφοριών έγινε με την χρήση δύο ερωτηματολογίων. Τα θέματα της έρευνας είναι οι διατροφικές συνήθειες των μαθητών και πόσο χρόνο αφιερώνουν οι μαθητές καθημερινά σε παιχνίδια και ψυχαγωγία.

Για το πρώτο θέμα της έρευνας χρησιμοποιήθηκε η μέθοδος στρωματοποιημένης δειγματοληψία. Ο λόγος ο οποίος χρησιμοποιήσαμε αυτή την τεχνική είναι επειδή επιθυμούσαμε να μελετήσουμε μια συγκεκριμένη υποομάδα του πληθυσμού. Η τεχνική αυτή είναι χρήσιμη σε αυτές τις έρευνες, δεδομένου ότι εξασφαλίζει την παρουσία των βασικών υποομάδων εντός του δείγματος. Ακόμα έχει υψηλή στατιστική ακρίβεια σε σχέση με τις άλλες μεθόδους. Είναι σημαντικό να αναφέρουμε ότι, παρέχει παρόμοια αποτελέσματα με την απλή τυχαία αλλά το μέγεθος του δείγματος είναι πολύ πιο μικρό.

Στο δεύτερο θέμα για την επιλογή του τυχαίου δείγματος χρησιμοποιήθηκε η μέθοδος των τυχαίων αριθμών. Σε σύγκριση με την μέθοδο της κλήρωσης, η μέθοδος των τυχαίων αριθμών είναι πιο απλή και αποτελεί έναν πιο αξιόπιστο τρόπο επιλογής του τυχαίου δείγματος.

Το δείγμα του πρώτου θέματος είναι 68 μαθητές και του δεύτερου θέματος 30 μαθητές.

Για το πρώτο θέμα από κάθε στρώμα επιλέγουμε ένα απλό τυχαίο δείγμα, με δειγματικά μεγέθη προσδιορισμένα εκ των προτέρων και συνενώνουμε τα επιμέρους αυτά δείγματα σε ένα. Δηλαδή από της 4 τάξεις στις οποίες έγινε η έρευνα Γ, Δ, Ε και ΣΤ εμείς πήραμε Γ1, Δ2, Ε1 και ΣΤ2.

Ενώ για το δεύτερο θέμα το οποίο η επιλογή του δείγματος έγινε με την βοήθεια του προγράμματος **minitab** στο οποίο δώσαμε το πληθυσμό των μαθητών της Δ, Ε, ΣΤ. Παρακάτω δίνεται ο πίνακας των αποτελεσμάτων που έδωσε το πρόγραμμα.

	Δ	Ε	ΣΤ
1	2	16	19
2	16	11	11
3	22	5	30
4	12	2	28
5	4	17	13
6	1	1	29
7	4	12	13
8	17	7	24
9	12	16	13
10	14	15	16
11	16		32
12	12		24
13	22		26
14	5		34
15	16		40
16			37
17			5
18			36
19			28
20			40

Η στατιστική ανάλυση έγινε με τη χρήση του προγράμματος διαχείρισης φύλλων εργασίας (Microsoft Office Excel).

Η συλλογή των στοιχείων έγινε ως εξής : κατά τη διάρκεια κάποιων διδακτικών ωρών με την παρουσία των δασκάλων και έχοντας λάβει άδεια από τον διευθυντή του 19^{ου} Δημοτικού σχολείου μοιράσαμε τα ερωτηματολόγια στα παιδιά που είχα επιλεγεί από την εκάστοτε μέθοδο. Επιπλέον, πρέπει να σημειωθεί ότι, οι ερευνητές και οι δάσκαλοι, δεν έβλεπαν τις απαντήσεις που έδιναν οι μαθητές ώστε να διασφαλιστεί το απόρρητο των απαντήσεων.

Η ΤΑΥΤΟΤΗΤΑ ΤΗΣ ΕΡΕΥΝΑΣ	
ΦΟΡΕΑΣ	Τμήμα Λογιστικής, Τεχνολογικού Ιδρύματος Πατρών.
ΤΥΠΟΣ ΚΑΙ ΜΕΘΟΔΟΣ	Έρευνα με προσωπική συμπλήρωση δομημένου ερωτηματολογίου από τους μαθητές στις αίθουσες του 19 ^{ου} Δημοτικού Σχολείου Πατρών.
ΠΛΗΘΥΣΜΟΣ	Σύνολο των μαθητών του 19 ^{ου} Δημοτικού Σχολείου Πατρών.
ΔΕΙΓΜΑ	Το επιδιωκόμενο μέγεθος του δείγματος ανέρχεται σε 98 μαθητές.
ΔΕΙΓΜΑΤΟΛΗΨΙΑ	Ακολουθήθηκε η μέθοδος στρωματοποιημένης δειγματοληψίας και η μέθοδος τυχαίων αριθμών. Στρώματα είναι οι ηλικίες των μαθητών.
ΧΡΟΝΟΣ ΔΙΕΞΑΓΩΓΗΣ	Η έρευνα πραγματοποιήθηκε 6 Απριλίου μεταξύ των ωρών 9-12.

9.3 Το ερωτηματολόγιο

Το ερωτηματολόγιο του πρώτου θέματος περιέχει 21 ερωτήσεις, ο τύπος ερωτήσεων είναι κλειστός (έχει προτεινόμενες απαντήσεις) και καλύπτει 2 σελίδες. Το ερωτηματολόγιο του δεύτερου θέματος περιέχει 10 ερωτήσεις εκ των οποίων οι έξι είναι κλειστού τύπου (έχει προτεινόμενες απαντήσεις) και οι άλλες τέσσερις είναι και ανοιχτού και κλειστού τύπου (έχει προτεινόμενες απαντήσεις αλλά δίνουμε και την επιλογή στον μαθητή με κενό χώρο να εκφράσει την άποψή του).

Ο στόχος του θέματος της πρώτης έρευνας, είναι να μελετήσουμε της διατροφικές συνήθειες των παιδιών ηλικίας 8 έως 12 ετών κατά τη διάρκεια της ημέρας τους. Πιο συγκεκριμένα, θα θέλαμε να μελετήσουμε αρχικά εάν τρώνε πρωινό στο σπίτι ή στο σχολείο, τι τρώνε για πρωινό καθώς και αν παίρνουν τις βασικές τροφές που χρειάζεται να παίρνει ένα παιδί αυτής της ηλικίας σε εβδομαδιαία βάση.

Ο στόχος του δεύτερου ερωτηματολογίου, είναι να μελετήσουμε και να αξιολογήσουμε πόσο χρόνο αφιερώνουν τα παιδιά καθημερινά σε παιχνίδια και ψυχαγωγία. Τι είδους παιχνίδια προτιμούν να παίζουν και αν είναι ικανοποιημένα από τους τρόπους ψυχαγωγίας της περιοχής τους. Ακόμα προσπαθούμε να μελετήσουμε τη σχέση που έχουν τα παιδιά με το

internet και εάν το χρησιμοποιούν στη καθημερινότητά τους, καθώς και ποιοι είναι οι λόγοι που το χρησιμοποιούν.

Κατά την κατάρτιση του ερωτηματολογίου, δόθηκε ιδιαίτερη προσοχή στην διατύπωση των ερωτήσεων και των προτεινόμενων απαντήσεων ώστε να είναι κατανοητές και εύστοχες. Συγκεκριμένα, σχετικά με τις προτεινόμενες επιλογές απαντήσεων εξαντλήθηκαν όλες οι πιθανές απαντήσεις.

9.4 Ανάλυση αποτελεσμάτων της έρευνας

9.4.1 Πρώτο μέρος της έρευνας

Στην έρευνα που ακολουθεί θα γίνει στατιστική ανάλυση του θέματος για τις διατροφικές συνήθειες των μαθητών.



Το γράφημα που ακολουθεί παρουσιάζει το ποσοστό των παιδιών που συμμετείχαν. Οπότε παρατηρούμε ότι το 47% είναι αγόρια και 53% είναι κορίτσια.



Γράφημα 1 - Φύλο των παιδιών (Ποσοστό παιδιών ανά φύλο)



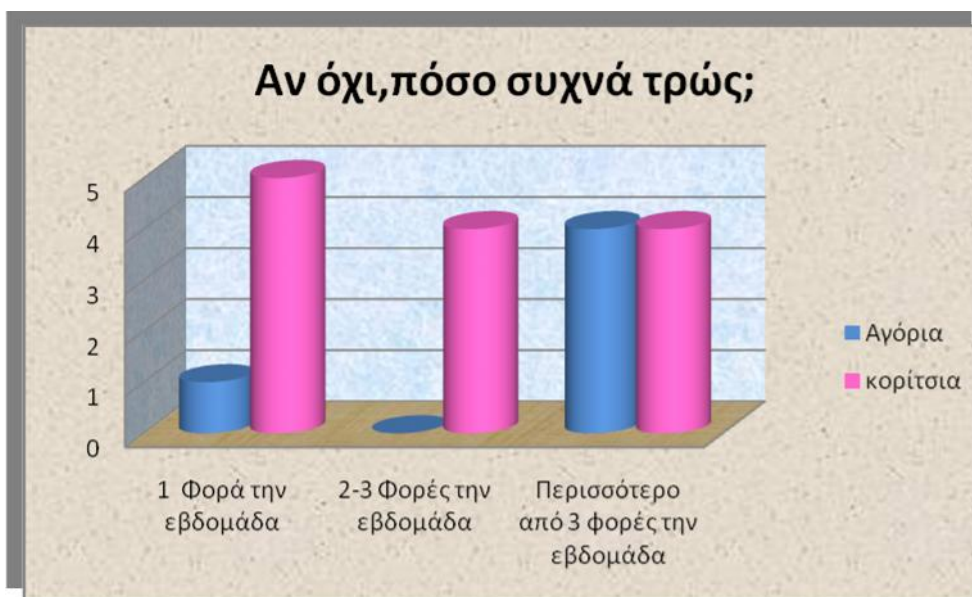
Γράφημα 2 - Φύλο των παιδιών (Αριθμός παιδιών ανά φύλο)

Τα γραφήματα που ακολουθούν παρουσιάζουν τα αποτελέσματα της έρευνας που αφορούν τις επιλογές των μαθητών ανά φύλλο.



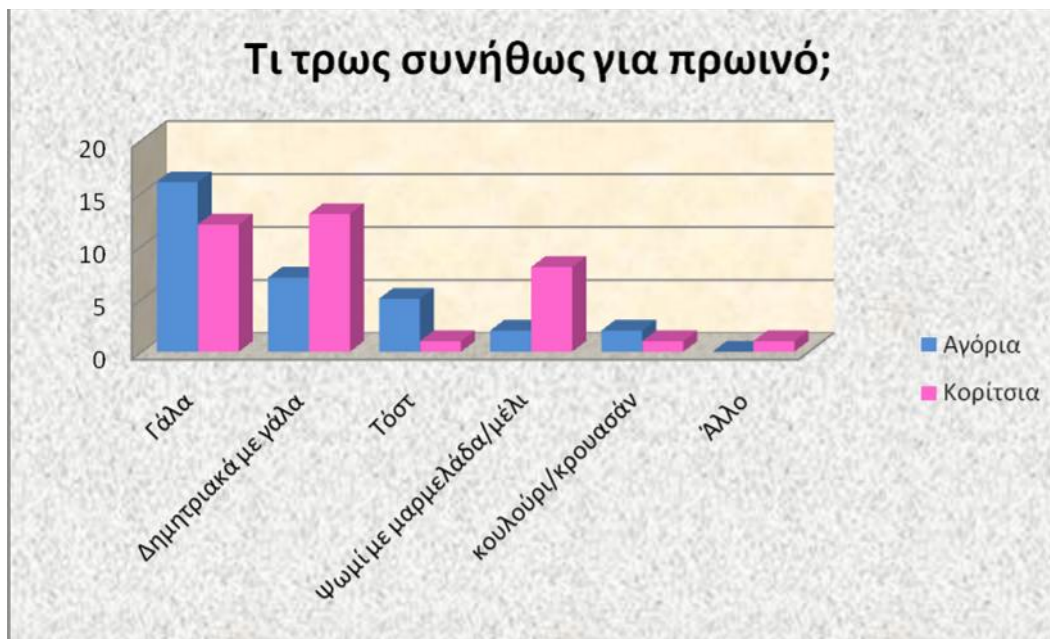
Γράφημα 3 - Αποτελέσματα ερώτησης "Τρώς καθημερινά πρωινό; "

Από το γράφημα, προκύπτει διαφοροποίηση των προτιμήσεων στα δυο φύλλα. Συγκεκριμένα, παρατηρούμε ότι τα αγόρια προτιμούν να τρώνε καθημερινά το πρωινό τους σε σχέση με τα κορίτσια.



Γράφημα 4 - Αποτελέσματα ερώτησης "Πόσο συχνά τρώς πρωινό";

Είναι εμφανές ότι τα κορίτσια δεν τρώνε συχνά πρωινό, όπως έδειξαν και τα αποτελέσματα στο παραπάνω διάγραμμα. Μεγάλο ποσοστό κοριτσιών, τρώνε μόνο μια φορά την εβδομάδα πρωινό, σε αντίθεση με τα αγόρια, που προτιμούν να τρώνε περισσότερες από τρεις φορές την εβδομάδα.



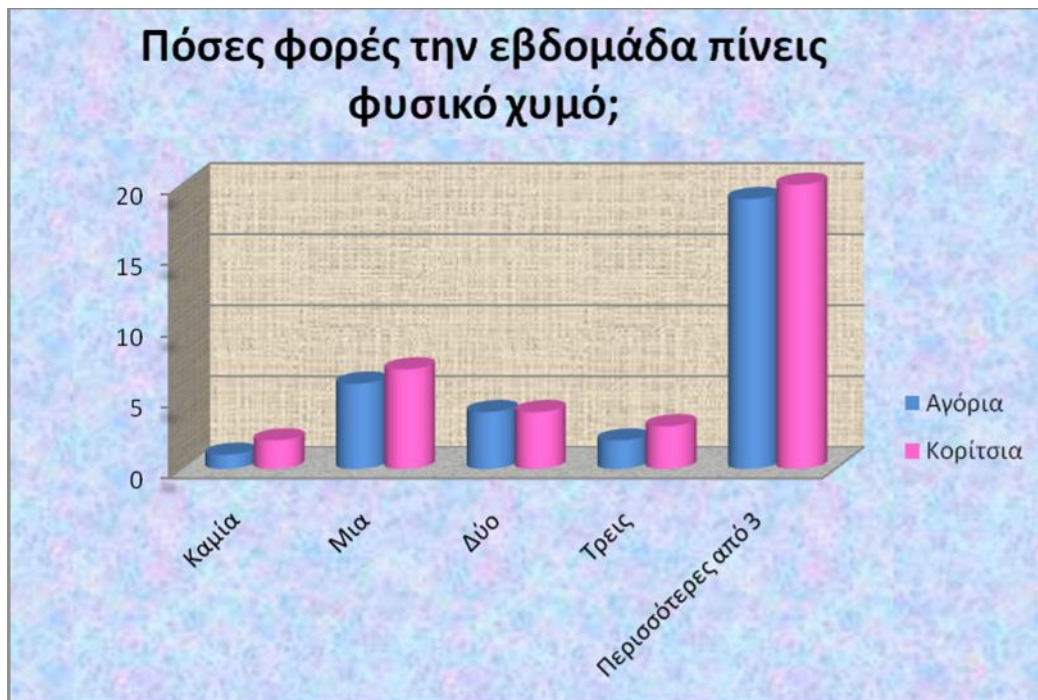
Γράφημα 5 - Αποτελέσματα ερώτησης "Τι τρώς συνήθως για πρωινό";

Στο γράφημα αυτό παρατηρούμε ότι τα κορίτσια προτιμούν γάλα με δημητριακά ή σκέτο γάλα σε αντίθεση τα αγόρια προτιμούν περισσότερο σκέτο γάλα.



Γράφημα 6 - Αποτελέσματα ερώτησης "Τι τρώς συνήθως στο διάλλειμα";

Στο διάγραμμα αυτό παρατηρούμε ότι και τα δύο φύλα προτιμούν να τρώνε στο διάλλειμα. Δείχνουν προτίμηση στο «κουλούρι/κρουασάν», με μικρή όμως διαφορά μεταξύ τους. Να σχολιάσουμε με όμως ότι στο ερώτημα «τρώς καθημερινά πρωινό» παρατηρήθηκε ότι κορίτσια δεν προτιμούν να παίρνουν πρωινό, σε αντίθεση όμως με το διάλλειμα.



Γράφημα 7 - Αποτελέσματα ερώτησης "Πόσες φορές την εβδομάδα πίνεις φυσικό χυμό";

Τα αποτελέσματα σε αυτό το διάγραμμα είναι αρκετά καλά παρατηρούμε ότι το μεγαλύτερο ποσοστό από αγόρια και κορίτσια πίνουν περισσότερο από 3 φορές την εβδομάδα φυσικό χυμό.



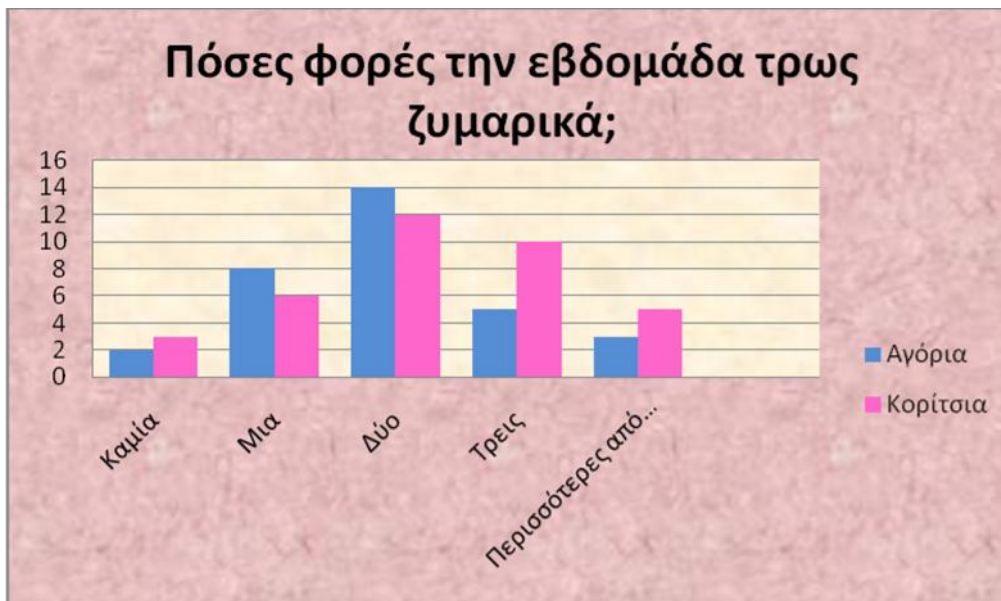
Γράφημα 8 - Αποτελέσματα ερώτησης "Πόσες φορές την εβδομάδα πίνεις αναψυκτικό";

Στο παραπάνω διάγραμμα βλέπουμε ότι ο μεγαλύτερος αριθμός των κοριτσιών πίνουν από μια έως καμία φορά την εβδομάδα αναψυκτικό τα αποτελέσματα είναι σχετικά καλά σε αντίθεση με τα αγόρια τα αποτελέσματα δεν είναι τόσο ικανοποιητικά, παρατηρείτε αύξηση στην απάντηση περισσότερες από τρεις φορές την εβδομάδα .



Γράφημα 9 - Αποτελέσματα ερώτησης " Πόσες φορές την εβδομάδα τρως δημητριακά";

Σε αυτό το διάγραμμα παρατηρούμε, αρκετά μεγάλο αριθμό κοριτσιών τα οποία προτιμούν να τρώνε δημητριακά, περισσότερο από 3 φορές την εβδομάδα. Στα αγόρια παρατηρούμε ότι μεγάλο ποσοστό δεν τρώνε δημητριακά. Σε αυτό το διάγραμμα τα αποτελέσματα των κοριτσιών είναι καλύτερα σε σχέση με τον αγοριών. .



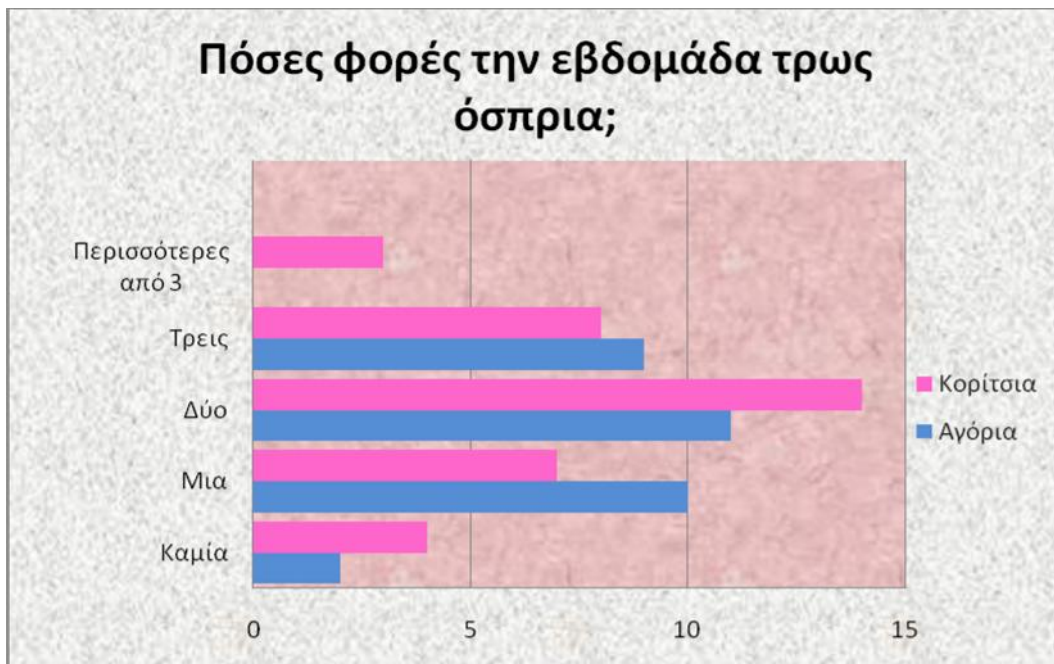
Γράφημα 10 - Αποτελέσματα ερώτησης "Πόσες φορές την εβδομάδα τρως ζυμαρικά";

Στο διάγραμμα αυτό παρατηρούμε ότι τα αποτελέσματα και των δύο φύλων είναι καλά, ο μεγαλύτερος αριθμός κοριτσιών και αγοριών τρώνε δυο φορές την εβδομάδα που θεωρούμε ότι είναι το ιδανικό για μια ισορροπημένη διατροφή.



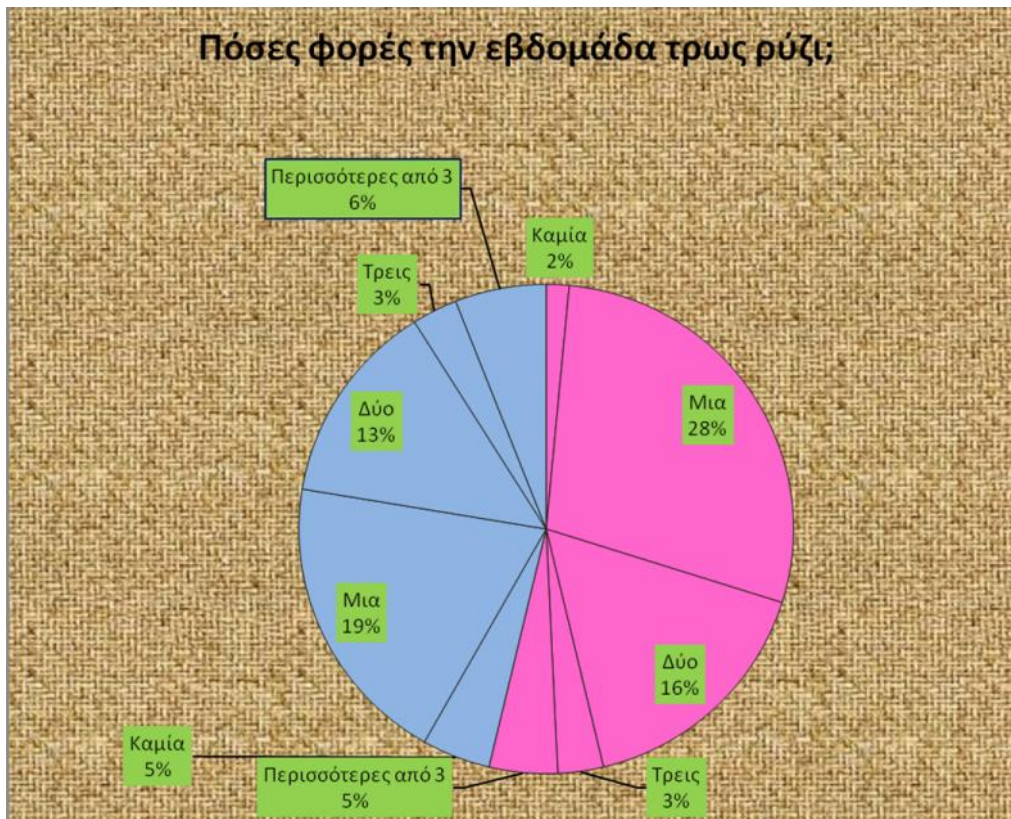
Γράφημα 11 - Αποτελέσματα ερώτησης "Πόσες φορές την εβδομάδα τρως ψάρι";

Στο παραπάνω διάγραμμα τα αποτελέσματα είναι καλά βλέπουμε και στα δύο φύλλα μεγαλύτερο αριθμό παιδιών προτιμούν να τρώνε μια φορά την εβδομάδα ψάρι. Σύμφωνα με μελέτες ένα παιδί θα πρέπει να τρώει τουλάχιστον δύο φορές την εβδομάδα ψάρι.



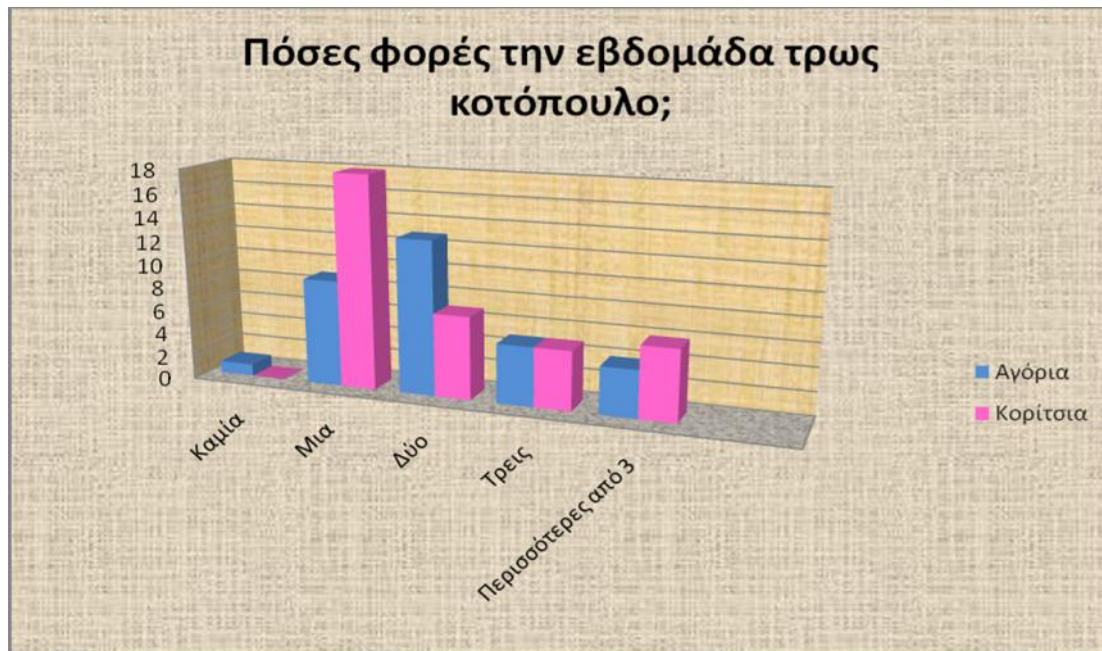
Γράφημα 12 - Αποτελέσματα ερώτησης "Πόσες φορές την εβδομάδα τρως όσπρια";

Στο παραπάνω διάγραμμα παρατηρούμε ότι μεγάλο ποσοστό κοριτσιών τρώνε δύο φορές την εβδομάδα όσπρια το οποίο είναι ικανοποιητικό με βάση την διατροφική πυραμίδα, σε αντίθεση με τα αγόρια που δεν είναι τόσο ικανοποιητικά τα αποτελέσματα .



Γράφημα 13 - Αποτελέσματα ερώτησης "Πόσες φορές την εβδομάδα τρώς ρύζι";

Στο διάγραμμα αυτό παρατηρούμε ότι τα αποτελέσματα στα δύο φύλλα είναι καλά, μεγάλο ποσοστό παιδιών τρώνε μια με δύο φορές την εβδομάδα ρύζι. Σύμφωνα με τις έρευνες είναι ικανοποιητικά τα αποτελέσματα.



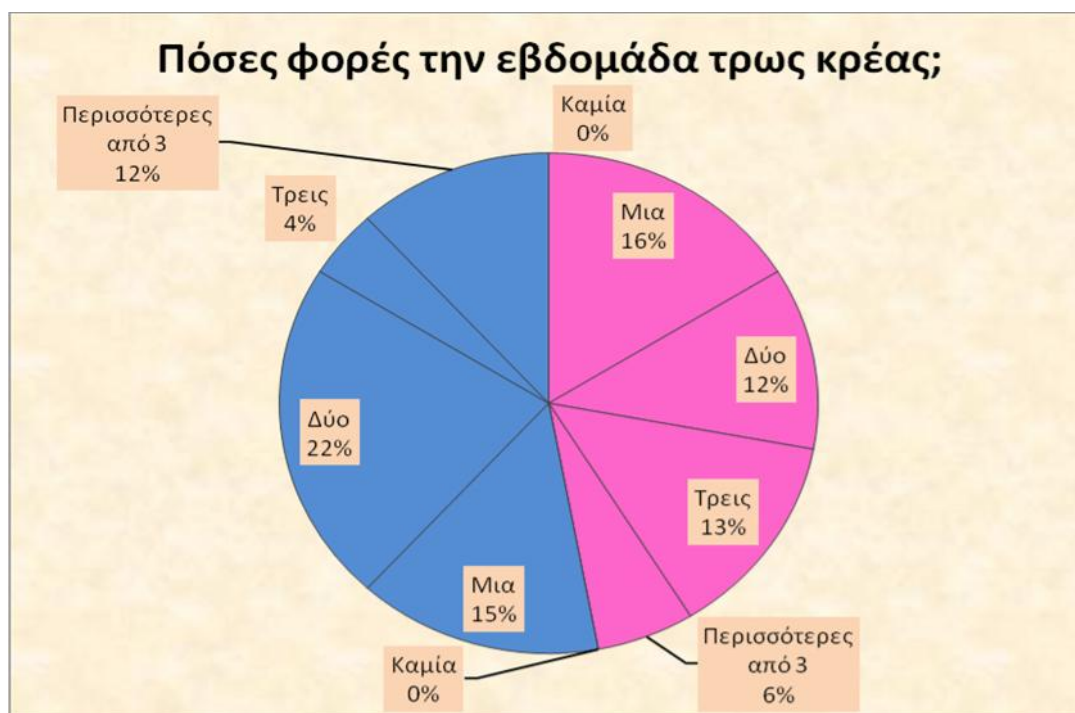
Γράφημα 14 - Αποτελέσματα ερώτησης "Πόσες φορές την εβδομάδα τρως κοτόπουλο";

Στο παραπάνω διάγραμμα παρατηρούμε μεγάλο αριθμό κοριτσιών να τρώνε μία φορά την εβδομάδα κοτόπουλο. Σε αντίθεση στα αγόρια βλέπουμε μια μικρή αύξηση τρώνε δύο φορές την εβδομάδα κοτόπουλο και αυτό είναι πιο ιδανικό.



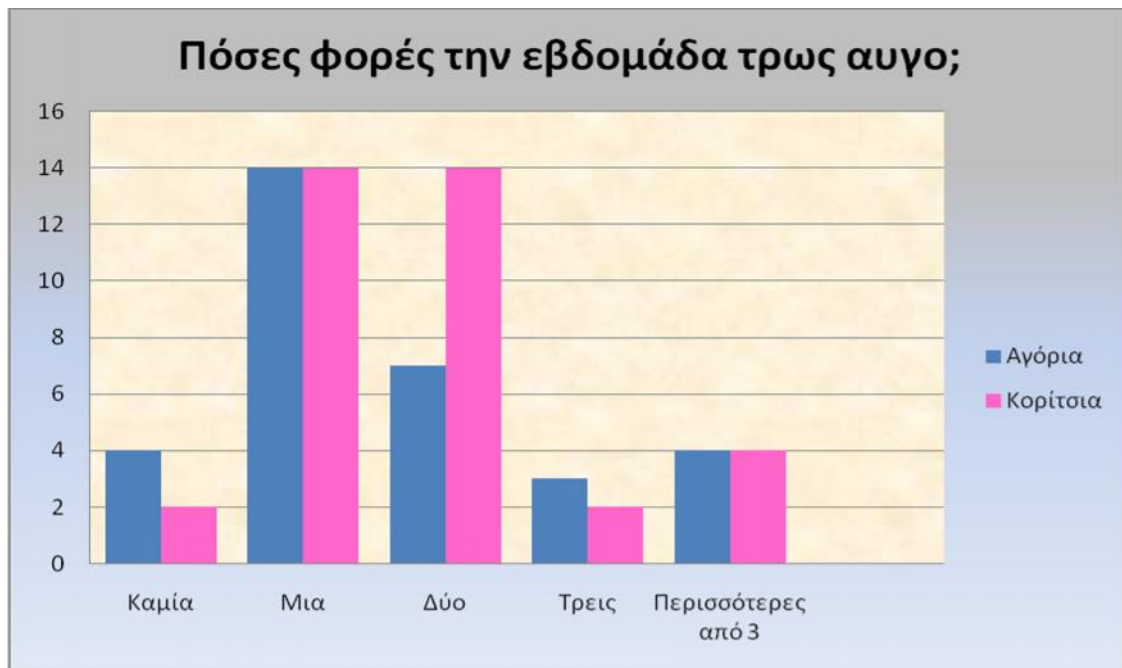
Γράφημα 15 - Αποτελέσματα ερώτησης "Πόσες φορές την εβδομάδα τρως λαχανικά";

Στο παραπάνω διάγραμμα τα αποτελέσματα των κοριτσιών είναι πολύ καλά, περισσότερες από τρεις φορές την εβδομάδα τρώνε λαχανικά. Σε αντίθεση, τα αγόρια τρώνε λιγότερες φορές την εβδομάδα λαχανικά σε σχέση με τα κορίτσια. Αυτό δεν είναι καλό διότι τα λαχανικά πρέπει να υπάρχουν καθημερινά στο γεύμα τους.



Γράφημα 16 - Αποτελέσματα ερώτησης "Πόσες φορές την εβδομάδα τρως κρέας";

Στο παραπάνω διάγραμμα παρατηρούμε ότι τα αποτελέσματα είναι καλά. Όλα τα παιδιά συμπεριλαμβάνουν κρέας στο εβδομαδιαίο διατροφολόγιό τους. Το μεγαλύτερο ποσοστό θτρώνε μια με δύο φορές την εβδομάδα. Μια μικρή αύξηση βλέπουμε στα αγόρια περισσότερο από 3 φορές την εβδομάδα σε σχέση με τα κορίτσια. Το οποίο δεν το θεωρούμε τόσο καλό, διότι το ιδανικό είναι μία δύο φορές .



Γράφημα 17 Αποτελέσματα ερώτησης "Πόσες φορές την εβδομάδα τρως αυγό";

Στο παραπάνω διάγραμμα παρατηρούμε ότι το ποσοστό των κοριτσιών είναι καλύτερο από τον αγοριών. Τα κορίτσια παρατηρούμε να τρώνε μια με δύο φορές την εβδομάδα αυγό σε σχέση με τα αγόρια που βλέπουμε ότι προτιμούνε μόνο μια φορά την εβδομάδα. Σύμφωνα με τους διατροφολόγους το σωστό είναι να τρώνε δύο.



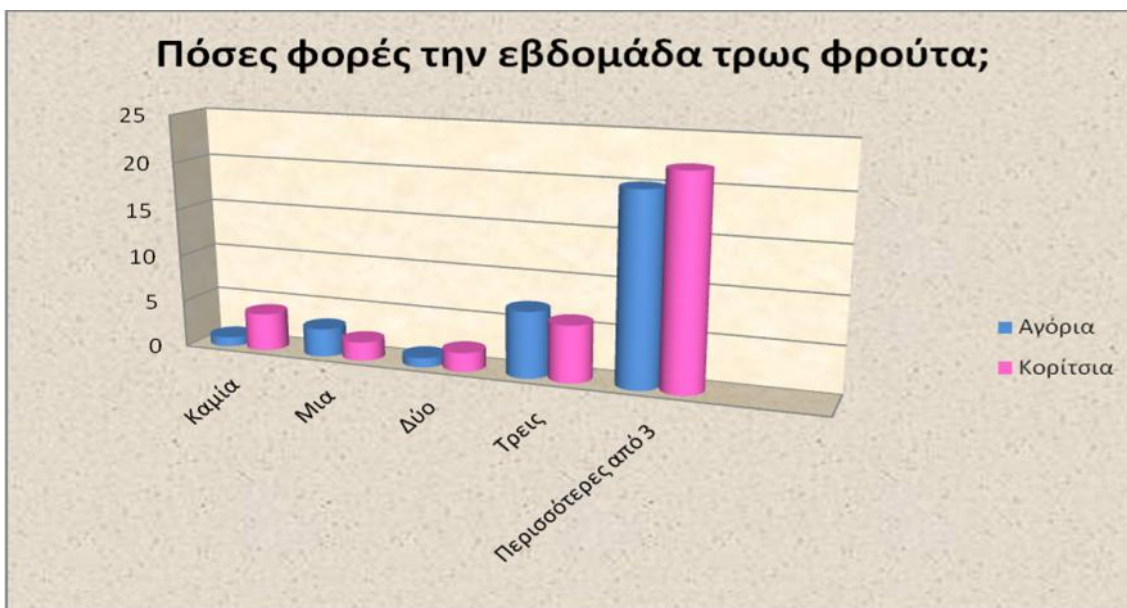
Γράφημα 18 - Αποτελέσματα ερώτησης "Πόσες φορές την εβδομάδα τρως πατάτες";

Παρατηρούμε ότι μεγάλο ποσοστό κοριτσιών προτιμούν μια φορά την εβδομάδα σε αντίθεση το μεγαλύτερο ποσοστό των αγοριών προτιμούν δύο φορές την εβδομάδα πατάτες. Αλλά παράλληλα εμφανίζονται αυξημένα τα αποτελέσματα περισσότερο από τρεις φορές την εβδομάδα και στα δύο φύλλα το οποίο δεν είναι πολύ ικανοποιητικό για διατροφική συνήθεια ενός παιδιού.



Γράφημα 19 - Αποτελέσματα ερώτησης "Πόσες φορές την εβδομάδα τρως γαλακτοκομικά";

Στο παραπάνω διάγραμμα παρατηρούμε τα κορίτσια προτιμούν να τρώνε περισσότερο από τρεις φορές την εβδομάδα γαλακτοκομικά και δεν υπάρχει κάποιο κορίτσι που να μην τρώει. Είναι αρκετά καλά τα αποτελέσματα στα κορίτσια. Στα αγόρια βλέπουμε μία μείωση στο αντίστοιχο αποτέλεσμα περισσότερο από τρεις φορές την εβδομάδα και υπάρχουν αγόρια τα οποία δεν τρώνε καθόλου γαλακτοκομικά.



Γράφημα 20 - Αποτελέσματα ερώτησης "Πόσες φορές την εβδομάδα τρως φρούτα";

Τα αποτελέσματα και στα δύο φύλλα είναι ιδανικά παρατηρούμε ότι μεγάλο ποσοστό και αγοριών και κοριτσιών τρώνε περισσότερο από τρεις φορές την εβδομάδα φρούτα. Το ικανοποιητικό είναι καθημερινά.



Γράφημα 21 - Αποτελέσματα ερώτησης "Πόσες φορές την εβδομάδα τρως γλυκά";

Στο παραπάνω διάγραμμα παρατηρούμε ότι μεγάλος αριθμός κοριτσιών τρώνε μια φορά την εβδομάδα σε σχέση με τα αγόρια που τρώνε μια με δύο φορές την εβδομάδα. Δεν είναι άσχημα τα αποτελέσματα των αγοριών αλλά θεωρούμε ότι των κοριτσιών είναι πιο ικανοποιητικά. Το ιδανικό σύμφωνα με τους διατροφολόγους είναι ότι πρέπει να καταναλώνονται μία φορά την εβδομάδα.

9.4.2 Δεύτερο μέρος της έρευνας

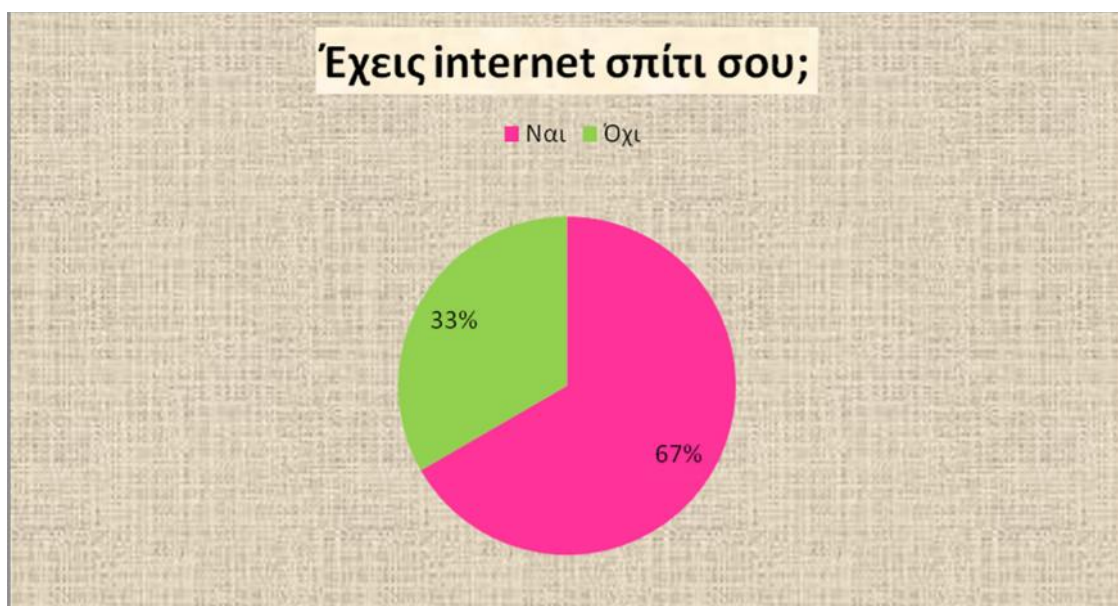
Τα αποτελέσματα που θα δούμε παρακάτω αφορούν τη δεύτερη έρευνα που κάναμε για τις εξωσχολικές δραστηριότητες των παιδιών και πόσο χρόνο αφιερώνουν γι' αυτές .

Η έρευνα έδειξε ότι οι μαθητές προτιμούν τα ομαδικά παιχνίδια και τις αθλοπαιδιές. Μικρό ποσοστό των μαθητών προτιμά να περνά τον ελεύθερο χρόνο του χωρίς τους φίλους του ή στο internet café και σε παιχνίδια στο internet και στον υπολογιστή.

Είναι σημαντικό να πούμε ότι οι περισσότεροι μαθητές χρησιμοποιούν το internet για ενημέρωση.

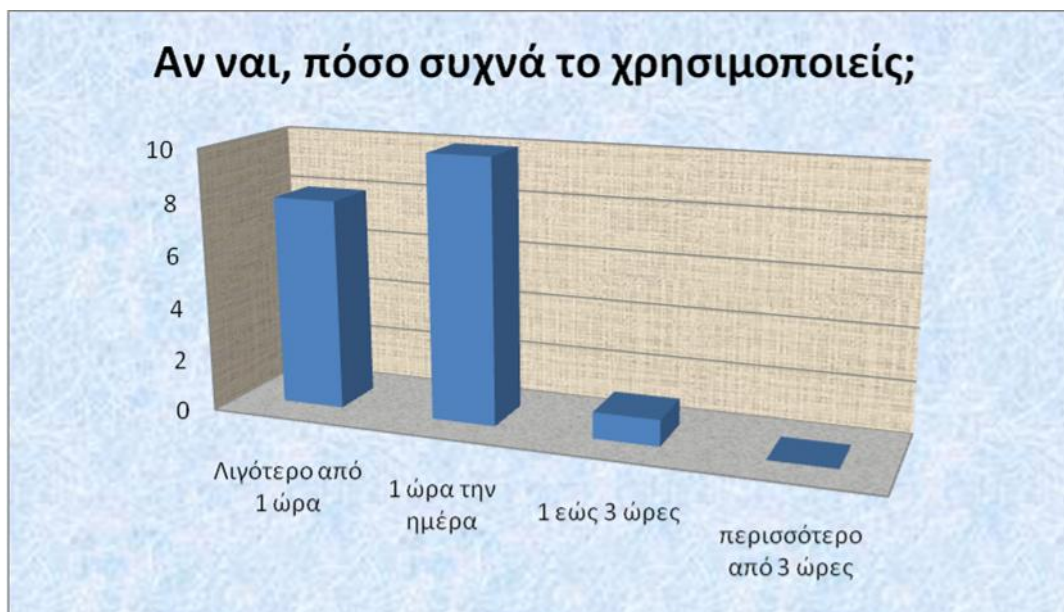
Όσον αφορά τον τρόπο ψυχαγωγίας και παιχνιδιών στην περιοχή τους, οι μαθητές φαίνονται ικανοποιημένοι με βάση τα αποτελέσματα.

Το γράφημα που ακολουθεί παρουσιάζει το ποσοστό των παιδιών που έχουν internet στο σπίτι τους.



Γράφημα 1 - Αποτελέσματα ερώτησης " Έχεις internet στο σπίτι σου; "

Στο γράφημα, φαίνεται ότι το μεγαλύτερο ποσοστό των μαθητών, έχουν internet στο σπίτι. Το 67% απάντησε «ναι» ενώ το 33% απάντησε «όχι». Θεωρούμε το ποσοστό αυτό είναι ιδανικό, αφού τα παιδιά φαίνονται εξεικωμένα με την τεχνολογία.



Γράφημα 2 - Αποτελέσματα ερώτησης " Αν ναι, πόσο συχνά το χρησιμοποιείς;"

Στο γράφημα φαίνεται ο χρόνος που αφιερώνουν οι μαθητές στο internet είναι μία ώρα την ημέρα ή λιγότερο από μία ώρα την ημέρα. Τα αποτελέσματα είναι λογικά και ιδανικά για έναν μαθητή.



Γράφημα 3 - Αποτελέσματα ερώτησης "Πόσο χρόνο αφιερώνεις καθημερινά σε παιχνίδια και ψυχαγωγία;"

Τα αποτελέσματα της έρευνας δείχνουν ότι οι μαθητές αφιερώνουν χρόνο σε ψυχαγωγία και παιχνίδια 1 έως 3 ώρες ή πάνω από τρεις ώρες καθημερινά. Δεν είναι καλά τα αποτελέσματα των μαθητών που αφιερώνουν πάνω από τρεις ώρες καθημερινά, σε παιχνίδια και ψυχαγωγία, αλλά το ιδανικό, σύμφωνα με το προηγούμενο διάγραμμα, είναι ότι στο διάστημα αυτό δεν απασχολούνται στο internet πάνω από μία ώρα.



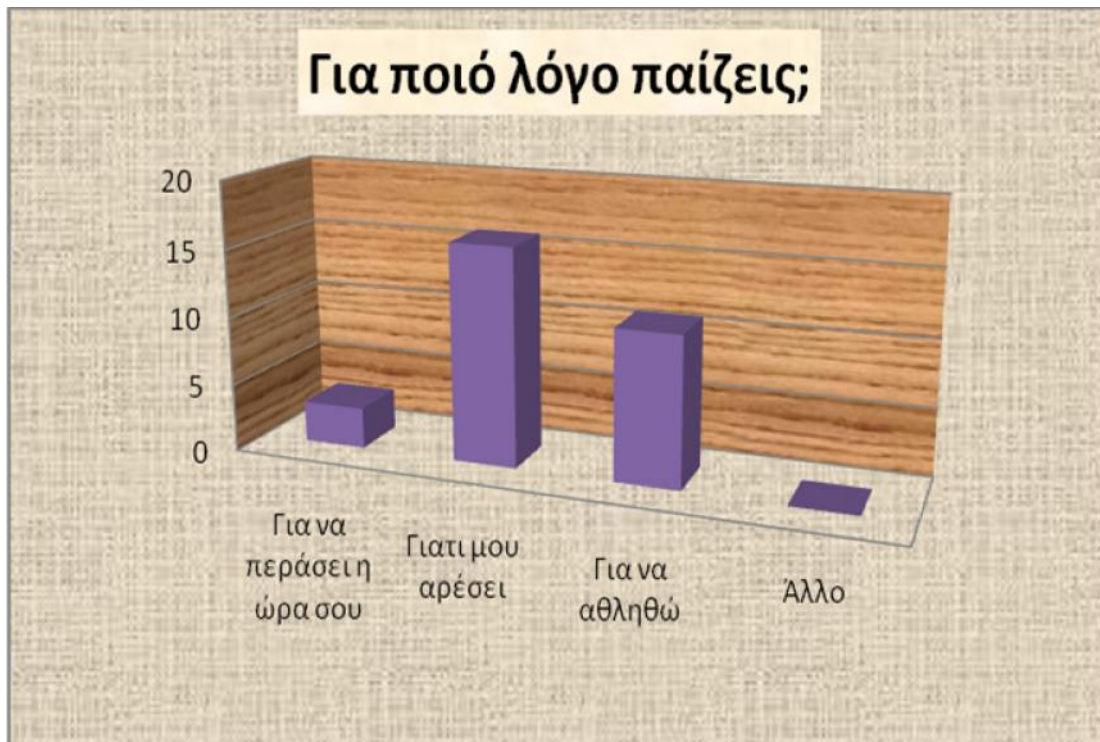
Γράφημα 4 - Αποτελέσματα ερώτησης " Σου αρέσει να περνάς τον ελεύθερο σου χρόνο μόνος σου ή με τους φίλους σου; "

Είναι προφανές ότι οι περισσότεροι μαθητές προτιμούν να περνούν τον ελεύθερο χρόνο τους με τους φίλους τους, ενώ λίγοι είναι εκείνοι που προτιμούν να περνάνε τον χρόνο τους μόνοι τους. Τα αποτελέσματα είναι ιδανικά, αφού στα προηγούμενα διάγραμμα δείχνουν ότι τα παιδιά ασχολούνται λιγότερο από μία ώρα με το internet, που σημαίνει ότι αφιερώνουν τον υπόλοιπο χρόνο ψυχαγωγίας με τους φίλους τους. Το μικρό ποσοστό των παιδιών που προτιμούν να παίζουν μόνα τους, τα οποία μπορούν να σπαταλούν όλο τον ελεύθερο χρόνο τους στο internet, σύμφωνα με έρευνες παρατηρούνται λιγότερο κοινωνικά και αυτό επηρεάζει τη πνευματική ανάπτυξή τους.



Γράφημα 5 - Αποτελέσματα ερώτησης "Ποιο είδος παιχνιδιού προτιμάς; "

Η πλειοψηφία των μαθητών προτιμά τις αθλοπαιδιές, ενώ πολύ μικρό ποσοστό μαθητών προτιμά την τηλεόραση και τα DVD. Τα αποτελέσματα είναι ικανοποιητικά γιατί οι περισσότεροι μαθητές προτιμούν παιχνίδια, τα οποία βοηθούν στην πνευματική αλλά και τη σωματική τους άσκηση.



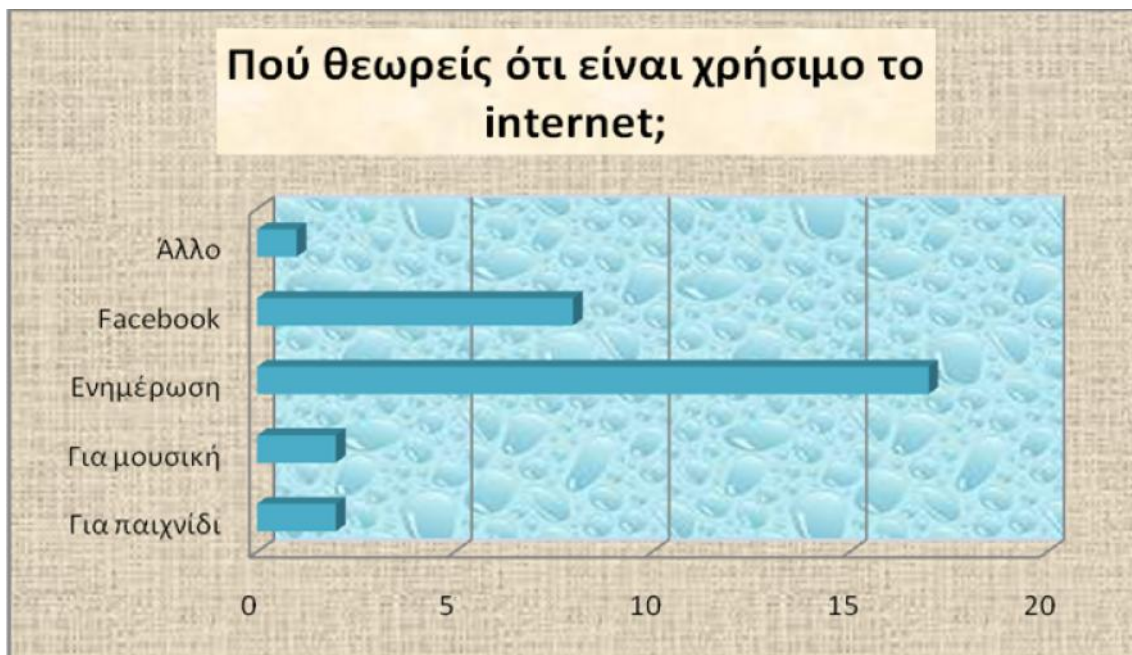
Γράφημα 6 - Αποτελέσματα ερώτησης " Για ποιο λόγο παίζεις;

Πολλοί είναι οι μαθητές που απάντησαν "παίζω γιατί μου αρέσει" σε αντίθεση με τους μαθητές που παίζουν "για να περάσει η ώρα τους". Τα αποτελέσματα είναι ικανοποιητικά, διότι τα παιδιά παίζουν επειδή τους αρέσει. Σε αντίθεση, με τα παιδιά που παίζουν για να περάσει η ώρα τους. Είναι μικρό το ποσοστό αυτών των παιδιών, αλλά είναι αρνητικό επειδή δείχνει ένα παιδί χωρίς ενδιαφέροντα.



Γράφημα 7 - Αποτελέσματα ερώτησης "Είσαι ικανοποιημένος με τους τρόπους παιχνιδιού ψυχαγωγίας της περιοχής σου;"

Εδώ οι μαθητές φαίνονται αρκετά ικανοποιημένοι από τους τρόπους ψυχαγωγίας της περιοχής τους. Αυτό μπορεί να είναι αθλητικές εγκαταστάσεις, πάρκα, εμπορικά κέντρα(τα οποία μπορεί να περιέχουν ηλεκτρονικά ή επιτραπέζια παιχνίδια), σινεμά.



Γράφημα 8 - Αποτελέσματα ερώτησης " Πού θεωρείς ότι είναι χρήσιμο το internet; "

Στο παραπάνω διάγραμμα έχουμε μεγάλο αριθμό μαθητών που χρησιμοποιούν το internet για ενημέρωση. Αυτό είναι αρκετά ικανοποιητικό αν η ενημέρωση αποσκοπεί στην εκπόνηση μιας εργασίας. Το ποσοστό για τη χρήση του facebook δεν είναι ικανοποιητικό, διότι το παιδί σε αυτή την ηλικία, δεν σε θέση να το χρησιμοποιεί. Λίγοι είναι οι μαθητές που χρησιμοποιούν το internet για παιχνίδια, μουσική ή κάτι άλλο.



Γράφημα 9 - Αποτελέσματα ερώτησης " Πηγαίνεις σε internet cafe; "

Παρατηρούμε ότι μεγάλο ποσοστό μαθητών δεν πηγαίνει σε internet café. Το αποτέλεσμα φαίνεται λογικό, αφού όπως έχουμε αναφέρει και σε παραπάνω διάγραμμα το 67% των μαθητών έχει internet στο σπίτι.



Γράφημα 10 - Αποτελέσματα ερώτησης " Αν ναι, πόσο συχνά; "

Από το 23% των μαθητών που πηγαίνουν σε internet cafe , απάντησαν ότι πηγαίνουν σπάνια ή μια έως τρεις φορές την εβδομάδα. Η απάντηση «σπάνια» δεν θεωρείται ικανοποιητική διότι τα παιδιά πρέπει να έχουν σχέση με το internet και την τεχνολογία.

Κεφάλαιο 10: Συμπεράσματα

Σε αυτήν την εργασία ασχοληθήκαμε με τέσσερις μεθόδους δειγματοληψίας και πιο συγκεκριμένα, την απλή τυχαία δειγματοληψία, την συστηματική δειγματοληψία, τη δειγματοληψία κατά στρώματα και τη δειγματοληψία κατά ομάδες. Η επιλογή της κατάλληλης μεθόδου δειγματοληψίας που θα εφαρμοστεί σε κάθε περίπτωση, είναι μια αρκετά σημαντική απόφαση και εξαρτάται από πολλούς παράγοντες. Πιο συγκεκριμένα, εξαρτάται από το ζητούμενο βαθμό ακρίβειας των αποτελεσμάτων, από τα χρονικά και χρηματικά περιθώρια της έρευνας, από τη μεταβλητότητα των μονάδων του πληθυσμού και από τη δυνατότητα που έχει ο πληθυσμός να μπορεί να διαιρεθεί σε υποπληθυσμούς με μεγάλη ομοιογένεια. Με βάση τα παραπάνω θα πρέπει να ληφθεί η απόφαση ποια μέθοδο δειγματοληψίας θα εφαρμοστεί στην έρευνα.

Στην απλή τυχαία δειγματοληψία, κάθε μέλος του πληθυσμού έχει την ίδια πιθανότητα να επιλεγεί και κάθε μία ακριβώς επιλογή είναι τελείως ανεξάρτητη από την άλλη. Υπάρχουν δύο τρόποι επιλογής του τυχαίου δείγματος, ο πρώτος είναι η μέθοδος της κλήρωσης και ο δεύτερος είναι η μέθοδος των τυχαίων αριθμών. Η μέθοδος της κλήρωσης εφαρμόζεται σπάνια γιατί είναι δύσκολο να εφαρμοστεί σε μεγάλο πληθυσμό, ενώ η μέθοδος των τυχαίων αριθμών είναι πιο απλή καθώς και αποτελεί έναν πιο αξιόπιστο τρόπο επιλογής του τυχαίου δείγματος. Σε αυτή τη μέθοδο δειγματοληψίας παρατηρούμε ότι η διαδικασία επιλογής του δείγματος είναι δίκαιη και αμερόληπτη όμως είναι δύσκολο να εφαρμοστεί αυτή η μέθοδος καθώς πολλές φορές δεν είναι γνωστός ολόκληρος ο πληθυσμός.

Η συστηματική δειγματοληψία αποτελεί μια εναλλακτική μέθοδο περισσότερο εφαρμόσιμη στην πράξη. Σύμφωνα με αυτή επιλέγουμε με τυχαίο τρόπο έναν αριθμό k και οι ομάδες του δείγματος προκύπτουν με βάση αυτόν. Σε αρκετές περιπτώσεις, ο πληθυσμός είναι τόσο μεγάλος που η απαρίθμηση όλων των μελών του, η καταχώρησή τους σε πίνακες και η επιλογή δείγματος με την απλή τυχαία δειγματοληψία καθίσταται εξαιρετικά επίπονη και πρακτικά αδύνατη διαδικασία. Όταν όμως υπάρχει μια ενημερωμένη λίστα με τα μέλη ενός πληθυσμού και ο πληθυσμός παρουσιάζει ομοιογένεια στη σύστασή του, τότε συχνά χρησιμοποιείται η απλή μέθοδος της συστηματικής δειγματοληψίας. Αυτή η μέθοδος είναι απλή και εύχρηστη, αλλά η επιλογή δεν είναι πραγματικά τυχαία και η ανεξαρτησία δεν εξασφαλίζεται.

Στη δειγματοληψία κατά στρώματα, τα μέλη του πληθυσμού χωρίζονται σε δύο ή περισσότερες ομοιογενείς ομάδες, έτσι ώστε κάθε μέλος του πληθυσμού να ανήκει σε μία μόνο ομάδα. Για να μπορέσει να εφαρμοστεί η δειγματοληψία κατά στρώματα, θα πρέπει να είναι από πριν γνωστή η κατανομή των χαρακτηριστικών του πληθυσμού που σχετίζονται με τη μεταβλητή που αποτελεί το πεδίο της έρευνας. Από αυτό προκύπτει ότι αν δεν υπάρχουν οι απαιτούμενες πληροφορίες σχετικά με τη διαστρωμάτωση του πληθυσμού, τότε πρέπει να αποφεύγεται η επιλογή αυτής της μεθόδου, καθώς σε αυτή την περίπτωση, η απλή τυχαία δειγματοληψία θα μπορέσει να επιτύχει έναν αντίστοιχο βαθμό ακρίβειας. Σε αυτή τη μέθοδο το δείγμα επιλέγεται με τη διαίρεση του πληθυσμού σε στρώματα και την τυχαία επιλογή

ενός αριθμού μελών, ανάλογου του ποσοστού τους στον αντίστοιχο πληθυσμό. Επίσης μας εξασφαλίζει ότι η σύνθεση του δείγματος θα είναι αντιπροσωπευτική του πληθυσμού και μας δίνει μια ικανοποιητική προσέγγιση στη μεταβλητή του πεδίου έρευνας.

Ενώ, οι προηγούμενες μέθοδοι δειγματοληψίας βασίζονται στην επιλογή των μελών ενός πληθυσμού ένα-προς-ένα, συχνά οι συμμετέχοντες ανήκουν ήδη σε κάποιες ομάδες – συστάδες.

Συνεπώς ο ερευνητής, μπορεί να επιλέξει με τυχαίο τρόπο ομάδες του πληθυσμού αντί για μέλη του πληθυσμού. Σε αυτή την περίπτωση είναι ευνοϊκό να χρησιμοποιηθεί η μέθοδος της δειγματοληψία κατά ομάδες. Η μέθοδος αυτή είναι ένας γρήγορος τρόπος να επιλέξουμε μεγάλο δείγμα και οι μετρήσεις μπορούν να γίνουν κατά ομάδες. Πρόκειται για μια απλή και εύχρηστη μέθοδο δειγματοληψίας που ευνοεί την επιλογή μεγάλου, σχετικά τυχαίου δείγματος, όμως η επιλογή δεν είναι πραγματικά τυχαία και δεν εξασφαλίζεται η ανεξαρτησία.

Θα πρέπει να αναφερθούμε και στη διαδικασία παρουσίασης των δειγματοληπτικών στοιχείων, όπου γίνεται είτε με μορφή πινάκων είτε με διαγράμματα. Είναι πάρα πολύ σημαντική αυτή η διαδικασία γιατί βοηθάει στο να γίνουν αντιληπτά τα αποτελέσματα της έρευνας και μας καθοδηγούν στην ερμηνεία τους αλλά και στη εξαγωγή συμπερασμάτων.

Συμπερασματικά, καταλήγουμε στο ότι για τη διεξαγωγή μιας δειγματοληπτικής έρευνας είναι σημαντικό, τόσο η επιλογή της κατάλληλης μεθόδου δειγματοληψίας για να έχουμε ακριβή αποτελέσματα, στο βαθμό που επιθυμούμε, όσο και ο τρόπος παρουσίασης των αποτελεσμάτων για να υπάρξει μια ολοκληρωμένη εικόνα της έρευνας.

Βιβλιογραφία

- .Castillo, Joseph Juan (2009)<<Stratified Sampling Methods>>
- Cochran W.,(New York 1997), <<Sampling Techniques>>
- Lindqvist M., 1993
- Hoinville G. et al,(London 1983) << Survey Research Practice>>
- Javeau, (ΑΘΗΝΑ 2000), <<Η έρευνα με ερωτηματολόγιο>>
- Yates, Daniel S. David S. Moore, Daren S. Starnes (3rd Ed. Freeman 2008)<<The Practice of Statistics>>
- ΑΠΟΣΤΟΛΟΠΟΥΛΟΣ Η. ΘΕΟΔΩΡΟΣ - (ΑΘΗΝΑ 1988), <<Στατιστική Επιχειρήσεων- Περιγραφική & Επαγωγική Στατιστική>>
- ΔΑΜΙΑΝΟΥ Χ.- (2η Έκδοση-1992 Αθήνα),<<Μεθοδολογία δειγματοληψίας- Τεχνικές κ εφαρμογές>>
- ΔΗΜΗΤΡΟΠΟΥΛΟΣ, Ε. Γ. (1994)<< Εισαγωγή στη μεθοδολογία της επιστημονικής έρευνας>>. Αθήνα
- ΖΑΦΕΙΡΙΟΥ, 2003
- ΚΙΟΧΟΣ Α. ΠΕΤΡΟΣ,(INTERBOOKS 1993),<<Στατιστική>>
- ΜΠΕΝΟΥ Β. ,(Εκδόσεις Σταμούλη 1991).<<Μέθοδοι και Τεχνικές Δειγματοληψίας>>
- ΜΠΕΝΟΥ Β.,(1997), <<Περιγραφική Στατιστική>>
- ΡΟΝΤΟΣ Κ.-Ε.ΠΑΠΑΝΗΣ (Εκδόσεις Σιδέρη 2007),<< Οι τεχνικές του καλού ερωτηματολογίου>>
- ΤΣΕΡΠΕΣ Ν.Α.- Φ.ΑΛΕΒΙΖΟΣ (ΠΑΤΡΑ 1996).<< Εισαγωγή στη θεωρία δειγματοληψίας>>
- ΧΑΡΙΣΗΣ-ΚΙΟΧΟΣ (INTERBOOKS 1997), <<Θεωρία Δειγματοληψίας και Εφαρμογές>>.

Παραρτήματα

Παράρτημα Α: Πίνακας Τυχαίων Αριθμών 2500 ψηφίων

54463	22662	65905	70639	79365	67382	29085	69831	47058	08186
15389	85205	18850	39226	42249	90669	96325	23248	60933	26927
85941	40756	82414	02015	13858	78030	16269	65978	01385	15345
61149	69440	11286	88218	58925	03638	52862	62733	33451	77455
05219	81619	10651	67079	92511	59888	84502	72095	83463	75577
41417	98326	87719	92294	46614	50948	64886	20002	97365	30976
28357	94070	20652	35774	16249	75019	21145	05217	47286	76305
17783	00015	10806	83091	91530	36466	39981	62481	49177	75779
40950	84820	29881	85966	62800	70326	84740	62660	77379	90279
82995	64157	66164	41180	10089	41757	78258	96488	88629	37231
96754	17676	55659	44105	47361	34833	86679	23930	53249	27083
34357	88040	53364	71726	45690	66334	60332	22554	90600	71113
06318	37403	49927	57715	50423	67372	63116	48888	21505	80182
62111	52820	07243	79931	89292	84767	85693	73947	22278	11551
47534	09243	67879	00544	23410	12740	02540	54440	32949	13491
98614	75993	84460	62846	59844	14922	48730	73443	48167	34770
24856	03648	44898	09351	98795	18644	39765	71058	90368	44104
96887	12479	80621	66223	86085	78285	02432	53342	42846	94771
90801	21472	42815	77408	37390	76766	52615	32141	30268	18106
55165	77312	83666	36028	28420	70219	81369	41943	47366	41067
75884	12952	84318	95108	72305	64620	91318	89872	45375	85436
16777	37116	58550	42958	21460	43910	01175	87894	81378	10620
46230	43877	80207	88877	89380	32992	91380	03164	98656	59337
42902	66892	46134	01432	94710	23474	20423	60137	60609	13119
81007	00333	39693	28039	10154	95425	39220	19774	31782	49037
68089	01122	51111	72373	06902	74373	96199	97017	41273	21546
20411	67081	89950	16944	93054	87687	96693	87236	77054	33848
58212	13160	06468	15718	82627	76999	05999	58680	96739	63700
70577	42866	24969	61210	76046	67699	42054	12696	93758	03283
94522	74358	71659	62038	79643	79169	44741	05437	39038	13163
42626	86819	85651	88678	17401	03252	99547	32404	17918	62880
16051	33763	57194	16752	54450	19031	58580	47629	54132	60631
08244	27647	33851	44705	94211	46716	11738	55784	95374	72655
59497	04392	09419	89964	51211	04894	72882	17805	21896	83864
97155	13428	40293	09985	58434	01412	69124	82171	59058	82859
98409	66162	95763	47420	20792	61527	20441	39435	11859	41567
45476	84882	65109	96597	25930	66790	65706	61203	53634	22557
89300	69700	50741	30329	11658	23166	05400	66669	48708	03887
50051	95137	91631	66315	91428	12275	24816	68091	71710	33258
31753	85178	31310	89642	98364	02306	24617	09609	83942	22716
79152	53829	77250	20190	56535	18760	69942	77448	33278	48805
44560	38750	83635	56540	64900	42912	13953	79149	18710	68618
68328	83378	63369	71381	39564	05615	42451	64559	97501	65747
46939	38689	58625	08342	30459	85863	20781	09284	26333	91777
83544	86141	15707	96256	23068	13782	08467	89469	93842	55349
91621	00881	04900	54224	46177	55309	17852	27491	89415	23466
91896	67126	04151	03795	59077	11848	12630	98375	52068	60142
55751	62515	21108	80830	02263	29303	37204	96926	30506	09808
85156	87689	95493	88842	00664	55017	55539	17771	69448	87530
07521	56898	12236	60277	39102	62315	12239	07105	11844	01117

Παράρτημα Β:

Ερωτηματολόγιο Α

1. Πόσο χρόνο αφιερώνεις καθημερινά σε παιχνίδια και ψυχαγωγία;

Α. καθόλου Β. Λιγότερο από 1 ώρα Γ. 1 έως 3 ώρες Δ. Πάνω από 3 ώρες

2. Σου αρέσει να περνάς τον ελεύθερο χρόνο σου;

Α. Μόνος μου Β. Με τους φίλους μου

3. Ποιο είδος παιχνιδιού προτιμάς;

Α. Ηλεκτρονικά παιχνίδια Β. Τηλεόραση, DVD

Γ. Αθλοπαιδιές (ποδόσφαιρο-μπάσκετ)

Δ. Παραδοσιακά παιχνίδια (κρυφτό, σχοινάκι)

Ε. Άλλο(.....)

4. Για ποιο λόγο παίζεις;

Α. Για να περάσει η ώρα Β. Γιατί μου αρέσει Γ. Για να αθληθώ

Ε. Άλλο(.....)

5. Είσαι ικανοποιημένος με τους τρόπους παιχνιδιού-ψυχαγωγίας της περιοχής σου;

A. Καθόλου B. Λίγο Γ. Αρκετά Δ. Πολύ Ε. Πάρα πολύ

6. Έχεις internet σπίτι σου;

A. Ναι B. Όχι

7. Αν ναι, πόσο συχνά το χρησιμοποιείς;

A. Λιγότερο από 1 ώρα B. 1 ώρα την ημέρα Γ. 1 έως 3 ώρες
Δ. Περισσότερο από 3 ώρες

8. Που θεωρείς ότι είναι χρήσιμο το internet;

A. Για παιχνίδια B. Για μουσική Γ. Ενημέρωση Δ. Facebook
E. Άλλο(.....)

9. Πηγαίνεις σε internet cafe;

A. Ναι B. Όχι

10. Αν ναι, πόσο συχνά;

A. Σπάνια B. 1 με 3 φορές την εβδομάδα Γ. Κάθε μέρα
Δ. Άλλο(.....)

Παράρτημα Γ :

Ερωτηματολόγιο Β

Έρευνα για τις διατροφικές συνήθειες των μαθητών του Δημοτικού Σχολείου Πατρών

Ερωτηματολόγιο Εβδομαδιαίας διατροφικής συνήθειας μαθητών.

1. Ηλικία:

A. 6-7 B. 7-8 Γ. 8-9 Δ. 9-10 E. 10-11 ΣΤ. 11-12

2. Φύλο:

A. Αγόρι B. Κορίτσι

3. Τρως καθημερινά Πρωινό;

A. Ναι B. Όχι

4. Αν όχι, πόσο συχνά τρως:

A. 1 Φορά την εβδομάδα B. 2-3 Φορές την εβδομάδα
Γ. Περισσότερο από 3 φορές την εβδομάδα

5. Τι τρως συνήθως για Πρωινό;

A. Γάλα B. Δημητριακά με γάλα Γ. Τοστ Δ. Ψωμί με μαρμελάδα/μέλι
E. Κουλούρι/Κρουασάν ΣΤ. Άλλο

6. Τι τρως συνήθως στο Διάλειμα;

A. Σάντουιτς B. Τυρόπιτα Γ. Κουλούρι/Κρουασάν Δ. Πατατάκια
E. Τίποτα ΣΤ. Άλλο

7. Πόσες φορές την εβδομάδα πίνεις Φυσικό χυμό;

A. Καμία B. Μία Γ. Δύο Δ. Τρεις E. Περισσότερες από τρεις

8. Πόσες φορές την εβδομάδα πίνεις Αναψυκτικό;

A. Καμία B. Μία Γ. Δύο Δ. Τρεις E. Περισσότερες από τρεις

9. Πόσες φορές την εβδομάδα τρως Δημητριακά;

A. Καμία B. Μία Γ. Δύο Δ. Τρεις E. Περισσότερες από τρεις

10. Πόσες φορές την εβδομάδα τρως Ζυμαρικά;

A. Καμία B. Μία Γ. Δύο Δ. Τρεις E. Περισσότερες από τρεις

11. Πόσες φορές την εβδομάδα τρως Ψάρι;

A. Καμία B. Μία Γ. Δύο Δ. Τρεις E. Περισσότερες από τρεις

12. Πόσες φορές την εβδομάδα τρως Όσπρια;

A. Καμία B. Μία Γ. Δύο Δ. Τρεις E. Περισσότερες από τρεις

13. Πόσες φορές την εβδομάδα τρως Ρύζι;

A. Καμία B. Μία Γ. Δύο Δ. Τρεις E. Περισσότερες από τρεις

14. Πόσες φορές την εβδομάδα τρως Κοτόπουλο;

A. Καμία B. Μία Γ. Δύο Δ. Τρεις E. Περισσότερες από τρεις

15. Πόσες φορές την εβδομάδα τρως Λαχανικά;

A. Καμία B. Μία Γ. Δύο Δ. Τρεις E. Περισσότερες από τρεις

16. Πόσες φορές την εβδομάδα τρως Κρέας;

A. Καμία B. Μία Γ. Δύο Δ. Τρεις E. Περισσότερες από τρεις

17. Πόσες φορές την εβδομάδα τρως Αυγό;

A. Καμία B. Μία Γ. Δύο Δ. Τρεις E. Περισσότερες από τρεις

18. Πόσες φορές την εβδομάδα τρως Πατάτες;

A. Καμία B. Μία Γ. Δύο Δ. Τρεις E. Περισσότερες από τρεις

19. Πόσες φορές την εβδομάδα τρως Γαλακτοκομικά;

A. Καμία B. Μία Γ. Δύο Δ. Τρεις E. Περισσότερες από τρεις

20. Πόσες φορές την εβδομάδα τρως Φρούτα;

A. Καμία B. Μία Γ. Δύο Δ. Τρεις E. Περισσότερες από τρεις

21. Πόσες φορές την εβδομάδα τρως Γλυκά;

A. Καμία B. Μία Γ. Δύο Δ. Τρεις E. Περισσότερες από τρεις