ΑΤΕΙ ΠΑΤΡΩΝ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΥ ΣΧΕΔΙΑΣΜΟΥ ΚΑΙ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΤΙΤΛΟΣ ΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ:

ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

ΣΠΟΥΔΑΣΤΕΣ

ΑΡΑΒΑΝΤΙΝΟΣ ΔΗΜΗΤΡΙΟΣ

ΓΚΟΛΙΑ ΕΥΓΕΝΙΑ

ΓΚΟΡΛΑ ΧΡΥΣΑΝΘΗ

ΕΠΟΠΤΕΥΩΝ ΚΑΘΗΓΗΤΡΙΑ

ΜΠΟΥΜΠΟΥΛΗ ΑΘΑΝΑΣΙΑ

ΠΑΤΡΑ 2012

ΑΤΕΙ ΠΑΤΡΩΝ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΥ ΣΧΕΔΙΑΣΜΟΥ ΚΑΙ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΤΙΤΛΟΣ ΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ:

ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

ΣΠΟΥΔΑΣΤΕΣ

ΑΡΑΒΑΝΤΙΝΟΣ ΔΗΜΗΤΡΙΟΣ

ΓΚΟΛΙΑ ΕΥΓΕΝΙΑ

ΓΚΟΡΛΑ ΧΡΥΣΑΝΘΗ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΡΙΑ

ΜΠΟΥΜΠΟΥΛΗ ΑΘΑΝΑΣΙΑ

ΠΑΤΡΑ 2012

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ ΠΡΩΤΟ

ΕΙΣΑΓΩΓΗ

ΚΕΦΑΛΑΙΟ ΔΕΥΤΕΡΟ

ΣΥΛΛΟΓΗ ΣΤΑΤΙΣΤΙΚΩΝ ΣΤΟΙΧΕΙΩΝ

ΚΕΦΑΛΑΙΟ ΤΡΙΤΟ

ΔΕΙΓΜΑΤΟΛΗΨΙΑ

ΚΕΦΑΛΑΙΟ ΤΕΤΑΡΤΟ

ΕΜΠΕΙΡΙΚΕΣ ΚΑΤΑΝΟΜΕΣ ΣΥΧΝΟΤΗΤΩΝ

ΚΕΦΑΛΑΙΟ ΠΕΜΠΤΟ

ΠΑΡΑΜΕΤΡΟΙ ΜΟΝΟΜΕΤΑΒΛΗΤΩΝ ΠΛΗΘΥΣΜΩΝ

ΚΕΦΑΛΑΙΟ ΕΚΤΟ

ΔΙΑΣΠΟΡΑ

ΚΕΦΑΛΑΙΟ ΕΒΔΟΜΟ

ΑΣΣΥΜΕΤΡΙΑ – ΚΥΡΤΩΣΗ

ΚΕΦΑΛΑΙΟ ΟΓΔΟΟ

ΒΑΣΙΚΕΣ ΑΡΧΕΣ ΤΗΣ ΘΕΩΡΙΑΣ ΤΩΝ ΠΙΘΑΝΟΤΗΤΩΝ

1.ΕΙΣΑΓΩΓΗ

1.1 Τι είναι η Στατιστική

Στην καθομιλουμένη, Στατιστική σημαίνει συστηματική απαρίθμηση και παρουσίαση αριθμητικών δεδομένων ή στοιχείων, τα οποία προέρχονται από πολλές παρατηρήσεις ή μετρήσεις. Οι παρατηρήσεις αυτές ή οι μετρήσεις αναφέρονται σε συγκεκριμένο αντικείμενο ή γεγονός. Ανάλογα με το αντικείμενο ή το γεγονός στο οποίο αναφέρονται τα αριθμητικά δεδομένα, η Στατιστική παίρνει και ιδιαίτερη ονομασία. Έτσι, όταν μιλάμε για παράδειγμα για Γεωργική Στατιστική, Στατιστική επιχειρήσεων ή Στατιστική Εργατικού δυναμικού κ.λπ., εννοούμε αριθμητικά δεδομένα που αναφέρονται αντίστοιχα στη γεωργία, στις επιχειρήσεις ή στο εργατικό δυναμικό κ.λπ. Στην επιστημονική γλώσσα, η λέξη Στατιστική έχει ευρύτερη σημασία˙ σημαίνει την επιστήμη που έχει ως αντικείμενο όχι μόνο τη συγκέντρωση και παρουσίαση, αλλά και τη μελέτη και ανάλυση των παρατηρήσεων ή μετρήσεων που αναφέρονται σε ένα συγκεκριμένο αντικείμενο ή γεγονός , οποιαδήποτε και αν είναι η φύση του. Έτσι, η Στατιστική περιλαμβάνει τόσο τις μεθόδους συλλογής και επεξεργασίας στοιχείων, όσο και τις μεθόδους ανάλυσης και μελέτης τους, ανακαλύπτοντας έτσι σχέσεις που υπάρχουν ανάμεσα στα διάφορα φαινόμενα και διατυπώνοντας συμπεράσματα που είναι χρήσιμα για τη λήψη ορθών αποφάσεων. Μπορούμε λοιπόν να πούμε ότι:

Στατιστική είναι η επιστήμη που ασχολείται με τις επιστημονικές μεθόδους συλλογής, οργάνωσης, παρουσίασης και ανάλυσης των αριθμητικών εκείνων στοιχείων που αναφέρονται σε χαρακτηριστικές ιδιότητες διαφόρων οικονομικών, κοινωνικών, δημογραφικών, φυσικών κ.λπ. φαινομένων και έχει ως σκοπό τη συστηματική μελέτη αυτών των στοιχείων για την κατάληξη σε γενικά συμπεράσματα, που είναι χρήσιμα στη διαδικασία της λήψης ορθών αποφάσεων. Η Στατιστική δηλαδή επιχειρεί να εξαγάγει γνώση χρησιμοποιώντας εμπειρικά δεδομένα. 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 την 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.

Αναλύοντας τον ορισμό αυτό της Στατιστικής, παρατηρούμε ότι τα βασικά στάδια που ακολουθούμε για τη μελέτη των ιδιοτήτων των διαφόρων μονάδων μιας πολυπληθούς ομάδας είναι τα εξής:

(α) Η συγκέντρωση των στατιστικών στοιχείων που είναι αναγκαία για τη μελέτη του προβλήματος που θέλουμε να ερευνήσουμε.

(β) Η μεθοδική επεξεργασία και παρουσίαση των στατιστικών στοιχείων σε μορφή αριθμητικών πινάκων και γραφικών παραστάσεων.

(γ) Η ανάλυση των στοιχείων αυτών και η εξαγωγή χρήσιμων συμπερασμάτων για να ληφθούν σωστές αποφάσεις.

1.2 Ιστορία της Στατιστικής

Η λέξη Στατιστική προέρχεται από τη λατινική λέξη status (που σημαίνει κράτος) και δηλώνει αρχικά συλλογή στοιχείων για τις κρατικές ανάγκες (έκταση, παραγωγή, πληθυσμό κ.λπ.). Έχει εξακριβωθεί ότι η πρώτη απογραφή πληθυσμού έγινε στην Κίνα από τον αυτοκράτορα Υ-άο το έτος 2238 π. Χ. , ενώ στους Ρωμαίους η πρώτη απογραφή πληθυσμού έγινε επί Ρωμύλου (753-715 π. Χ.) και η τελευταία από τον αυτοκράτορα Βεσπασιανό το 73 μ. Χ.. Στην Αγγλία, η πρώτη καθολική απογραφή του πληθυσμού και του πλούτου γενικά έγινε το 1085 από τον Γουλιέλμο τον κατακτητή.

Το 1583 γράφεται από τον Fr. Sansovino το πρώτο βιβλίο στατιστικού πριεχομένου και λίγο αργότερα εισάγεται από τον Konring (1606-1681) η Στατιστική στην ανώτερη παιδεία.

Την ίδια εποχή εμφανίζεται το ενδιαφέρον για τις ασφάλειες ζωής και ο περίφημος αστρονόμος Halley, χρησιμοποιώντας τα ληξιαρχικά βιβλία γεννήσεων και θανάτων της πόλεως Breslaou, παρουσιάζει τον πρώτο πίνακα θνησιμότητας. Το ρεύμα αυτό των δημογραφικών μελετών επεκτείνεται και στη Γερμανία, όπου ο πάστορας Siissmilch (1707-1767) συγκεντρώνει στοιχεία από τα ληξιαρχικά βιβλία των εφημερίων της Πρωσίας και καταλήγει, το 1741, στο συμπέρασμα ότι το ποσοστό γέννησης των αγοριών είναι 51% και των κοριτσιών 49%, ενώ τα δύο φύλα έχουν ίσα ποσοστά κατά την εποχή του γάμου. Για το συγγραφέα το φαινόμενο αυτό δεν είναι τυχαίο γεγονός, αλλά νόμος θείας προέλευσης που αποσκοπεί στη διαιώνιση του είδους. Μέχρι την εποχή αυτή, η Στατιστική έχει περιγραφικό χαρακτήρα και ασχολείται κυρίως με θέματα Δημογραφίας.

Η Στατιστική θα ξεφύγει από τον περιγραφικό χαρακτήρα της με την ανάπτυξη ενός νέου κλάδου, του Λογισμού των Πιθανοτήτων, ο οποίος προήλθε από τη μελέτη των τυχερών παιγνίων (χαρακτηριστική μάλιστα είναι η αλληλογραφία ανάμεσα στους Γάλλους μαθηματικούς Pascal και Fermat, με αφορμή τα ερωτήματα που έθεσε στον Pascal ο Ιππότης DeMere για τα παίγνια του κύβου). Από τους θεμελιωτές του Λογισμού των Πιθανοτήτων αναφέρουμε τον Bernoulli, ο οποίος στο βιβλίο του ¨ Η τέχνη των προβλέψεων ¨ διατυπώνει τον περίφημο νόμο των μεγάλων αριθμών, και το Γάλλο μαθηματικό Laplace, στον οποίο οφείλεται η εφαρμογή του Λογισμού των Πιθανοτήτων στη σπουδή των φυσικών φαινομένων με πολυσύνθετες αιτίες. Στη νέα αυτή περίοδο της Στατιστικής, ο Βέλγος αστρονόμος Quetelet επεκτείνει την εφαρμογή της Στατιστικής στη σπουδή των φυσικών, διανοητικών και ηθικών ιδιοτήτων των ανθρώπου και παίρνει την πρωτοβουλία για τη σύγκληση του πρώτου Διεθνούς Συνεδρίου Στατιστικής που έγινε στις Βρυξέλλες το 1853, ενώ αργότερα ο F. Galton εφαρμόζει τη Στατιστική στη Βιολογία και, ειδικότερα, στα προβλήματα της κληρονομικότητας. Η προσπάθεια του Galton συνεχίστηκε από τον Άγγλο μαθηματικό Pearson, στον οποίο οφείλεται κατά πολύ η σημερινή ανάπτυξη και θέση της Στατιστικής.

1.3 Αντικείμενο της Στατιστικής

Για να ορισθεί με ακρίβεια το αντικείμενο της Στατιστικής, θα πρέπει πρώτα να γίνει η παρακάτω διάκριση:

Ο όρος Στατιστική ή Στατιστικές χρησιμοποιείται για να δηλώσει αριθμητικές πληροφορίες. Οι στατιστικές παρουσιάζουν τα αποτελέσματα της συστηματικής συλλογής ενός συνόλου πληροφοριών.

Σε κάθε χώρα έχουν δημιουργηθεί αυτοτελείς στατιστικοί οργανισμοί με σκοπό τον αποτελεσματικό συντονισμό όλων των στατιστικών εργασιών. Τέτοιος οργανισμός στην Ελλάδα είναι η Εθνική Στατιστική Υπηρεσία ή Ε. Σ. Υ. Ε. (https://www.statistics.gr) . Οι στατιστικές που καταρτίζει η Ε. Σ. Υ. Ε. (μηνιαίες, τριμηνιαίες, ανά 5ετία και ανά 10ετία), καλύπτουν όλους σχεδόν τους τομείς δραστηριότητας. Πληθυσμιακά στοιχεία (πληθυσμός κατά διάφορες διακρίσεις, φυσική κίνηση πληθυσμού – γάμοι – γεννήσεις – θάνατοι), στοιχεία απασχόλησης και ανεργίας, στοιχεία που αφορούν στην υγεία και στην κοινωνική ασφάλιση, την παιδεία, τη δικαιοσύνη, την παραγωγική διαδικασία, τα δημόσια οικονομικά, τις τιμές, το εθνικό εισόδημα και τέλος τις πολιτιστικές δραστηριότητες, συνιστούν το βασικό υλικό από το οποίο προκύπτουν οι στατιστικοί πίνακες και οι διάφοροι δείκτες που καταρτίζει η Ε. Σ. Υ. Ε. , σε βραχυχρόνια και μακροχρόνια δράση.

Βασικός χρήστης των στατιστικών και δεικτών που καταρτίζει η Ε. Σ. Υ. Ε. είναι το κράτος, που με βάση αυτά, σχεδιάζει, υλοποιεί και παρακολουθεί τις πολιτικές του στους διάφορους τομείς. Ακολουθεί η Ευρωπαϊκή Ένωση που χρειάζεται τα επιμέρους στοιχεία των κρατών – μελών της, για να συνθέσει τις ευρωπαϊκές στατιστικές με τη βοήθεια του ευρωπαϊκού στατιστικού γραφείου EUROSTAT (https://europa.eu.int/comm/eurosta/) , διεθνείς οργανισμοί, όπως η UNESCO με το στατιστικό της ινστιτούτο (https://www.uis.unesco.org/) με αντικείμενο τη συλλογή, παρουσίαση και την επεξεργασία αριθμητικών πληροφοριών για τα επιμέρους κράτη και τις μεταξύ τους σχέσεις. Τέλος, χρήστες των στοιχείων της Ε. Σ. Υ. Ε. είναι και ο επιχειρηματικός κόσμος, επιστήμονες, μελετητές και αναλυτές, όπως επίσης και ο απλός κόσμος.

Η Στατιστική είναι μια επιστήμη που αναφέρεται στις παρακάτω διαδικασίες:

- Συλλογή Δεδομένων
- Έλεγχος Δεδομένων (καταμέτρηση, διάταξη, διόρθωση λαθών και απομάκρυνση αμφίβολων περιπτώσεων, συμπλήρωση ελλιπών στοιχείων)
- Παρουσίαση δεδομένων (Πίνακες και Διαγράμματα)
- Επεξεργασία των δεδομένων με σκοπό τη διεξαγωγή χρήσιμων συμπερασμάτων.

Σήμερα από τους στατιστικούς αναμένονται πολλές πληροφορίες, καθώς και ανάλυση σύνθετων φαινομένων, με συνέπεια να πρέπει να χρησιμοποιηθούν πολλές και διαφορετικές στατιστικές μέθοδοι. Η χρησιμοποίηση κάθε φορά της πιο κατάλληλης ή των καταλληλότερων από αυτές, επιτρέπει στους στατιστικούς να ανταποκρίνονται στις μεγάλες απαιτήσεις των υπολοίπων επιστημονικών χώρων.

1.4 Χρησιμότητα και πεδία εφαρμογής της Στατιστικής

Μια απλή αρίθμηση των εφαρμογών της Στατιστικής, που είναι βασικά εφαρμοσμένη επιστήμη, δείχνει ότι αυτή χρησιμοποιείται σε όλους σχεδόν τους τομείς της ανθρώπινης δραστηριότητας. Η Στατιστική είναι απαραίτητη στη Διοίκηση γενικά, όπου η λήψη ορθών αποφάσεων έχει μεγάλη σημασία για την πρόοδο ενός κράτους, ενός οργανισμού, μιας βιομηχανίας ή μιας επιχείρησης. Γι ‘ αυτό και δεν υπάρχει σήμερα στις σύγχρονες επιχειρήσεις κανένας τομέας που να μην χρησιμοποιεί τις στατιστικές μεθόδους στη λήψη επιχειρηματικών αποφάσεων.

Μεγάλη σημασία έχει η εφαρμογή της Στατιστικής στη Δημογραφία. Δημογραφία είναι η επιστήμη που ασχολείται με τη στατιστική μελέτη του πληθυσμού για να διαπιστώσει το σύνολο των κατοίκων, τη σύνθεση του , τις μεταβολές του και όλα τα φαινόμενα που τις προκαλούν. Σκοπός της μελέτης είναι η εξαγωγή συμπερασμάτων για την υλική και ηθική βελτίωση των πληθυσμών και λήψη μέτρων για την αποτροπή επιζήμιων συνεπειών που ακολουθούν συνήθως τις διάφορες δημογραφικές μεταβολές.

Μια σπουδαία εργασία που γίνεται για το σκοπό αυτό είναι η απογραφή του πληθυσμού, με την οποία διαπιστώνεται η ποσοτική και ποιοτική κίνηση του και γίνεται μελέτη όλων των προβλημάτων που τον απασχολούν. Ειδικότερα, εξακριβώνεται ο αριθμός των κατοίκων σε κάθε περιφέρεια, η ηλικία τους, το φύλο, το θρήσκευμα, η γλώσσα, ο βαθμός μορφώσεως, ο τόπος γεννήσεως, το επάγγελμα κ.ά. Με την απογραφή επιτυγχάνεται και η στατιστική μελέτη της κινήσεως του πληθυσμού μιας χώρας σε σύγκριση με προηγούμενες στατιστικές. Μελετάται έτσι συγκριτικά η μετανάστευση, αστυφιλία, αύξηση ή μείωση γεννήσεων, γάμων, θανάτων, μείωση πληθυσμού ορισμένων περιφερειών κ.α. και εξάγονται συμπεράσματα για τη βελτίωση της ζωής του πληθυσμού και την αντιμετώπιση των διαφόρων προβλημάτων που δημιουργούνται με τις αλλαγές.

Η δημογραφία έκανε την εμφάνισή της κυρίως κατά τον 17ο αιώνα, συγχρόνως με τη γέννηση της στατιστικής και από τα μέσα του 17ου αιώνα έγιναν διάφορες αξιόλογες έρευνες με σημαντικότερη του Τόμας Ρόμπερτ Μάλθους [1798] σχετικά με τα προβλήματα της αύξησης του πληθυσμού, αλλά και του Αντόλφ Κετελέ [1796-1874].

Επίσης, η Στατιστική βρίσκει πολύ μεγάλη εφαρμογή και στον Οικονομικό τομέα (Οικονομική Στατιστική), όπου η παρακολούθηση του γενικού επιπέδου των τιμών, του εθνικού εισοδήματος, της νομισματικής ισοτιμίας και των οικονομικών διακυμάνσεων, της απασχόλησης, της παραγωγικότητας, της κατάρτισης δεικτών οικονομικής δραστηριότητας, των εθνικών πόρων και της εθνικής δαπάνης, είναι αντικείμενα στατιστικής επεξεργασίας.

Τέλος μερικοί ακόμη εξειδικευμένοι επιστημονικοί τομείς που χρησιμοποιούν την εφαρμοσμένη Στατιστική τόσο εκτεταμένα ώστε να έχουν ειδική ορολογία είναι:

- Βιοστατιστική
- Μηχανική Στατιστική
- Στατιστική Φυσική
- Ψυχολογική Στατιστική
- Κοινωνική Στατιστική (για όλες τις Κοινωνικές επιστήμες)
- Ανάλυση Διαδικασιών και Χημειομετρία (για την ανάλυση δεδομένων από την αναλυτική χημεία και τη χημική μηχανική)
- Οικονομετρία
- Τεχνομετρία

Η χρησιμότητα της Στατιστικής φαίνεται και από το γεγονός ότι η Στατιστική διδάσκεται σήμερα σχεδόν σε όλες τις Ανώτατες και Ανώτερες Σχολές της χώρας μας.

1.5 Κλίμακες μέτρησης

Οι κλίμακες μέτρησης είναι μια διαδικασία συνδυασμού δεικτών σε μια μετρήσιμη διάσταση, η οποία αντιπροσωπεύει την ευρύτερη έννοια. Οι κλίμακες δίνουν την δυνατότητα ανάπτυξης της έννοιας και βασίζονται στη συσσώρευση των δεικτών μιας μεταβλητής. Αφορούν την ίδια θεωρητική έννοια. Η γενική κλίμακα αποτελείται από μια σειρά ερωτήσεων οι οποίες ενοποιούνται και μετρώνται ως σύνολο. Ο συνολικός βαθμός που προκύπτει από τον ερωτώμενο αντιπροσωπεύει την γενική του στάση. Τα κύρια είδη κλιμάκων είναι:

ΚΑΤΗΓΟΡΙΑΚΗ

Βασίζεται στην τοποθέτηση ατόμων ή αντικειμένων σε κατηγορίες. Η κάθε κατηγορία είναι εντελώς διαφορετική από την άλλη, όπου γίνεται ταξινόμηση χωρίς ιεράρχηση.

Παράδειγμα:

Φύλο:	1=άντρας	2=γυναίκα	
Θρήσκευμα:	1=χριστιανός	2=μωαμεθανός	3=άλλο

ΔΙΑΤΑΚΤΙΚΗ

Αποσκοπεί στην κατάταξη ατόμων ή αντικειμένων σε σειρά ανάλογα με το βαθμό στον οποίο κατέχουν ένα χαρακτηριστικό.

Παράδειγμα:

Θέση:	1	στρατιώτης
	2	δεκανέας
	3	λοχίας
	4	επιλοχίας
	5	αξιωματικός

ΙΣΟΔΙΑΣΤΗΜΙΚΗ

Γίνεται χρήση του αυθαίρετου μηδέν. Οι διαδοχικές θέσεις τοποθετούνται σε ίσα διαστήματα, στις οποίες είναι δυνατές μόνο οι αριθμητικές πράξεις της πρόσθεσης και αφαίρεσης.

Παράδειγμα:

Θερμοκρασία σε ^oC ή ^oF, χρονολογία (π.χ. 2002)

ΑΝΑΛΟΓΙΚΗ

Στην αναλογική κλίμακα, αρχή μέτρησης είναι το πραγματικό μηδέν, όπου εδώ όλες οι αριθμητικές πράξεις είναι δυνατές.

Παράδειγμα:

Θερμοκρασία σε ^{o}K, απόσταση, βάρος, ηλικία

Οι σπουδαιότεροι κλίμακες που χρησιμοποιούνται στις έρευνες είναι:

- Κλίμακα Likert

Είναι μία ψυχομετρικός κλίμακα απάντησης χρησιμοποιούμενη συχνά μέσα σε ερωτηματολόγια, και είναι η ευρύτατα χρησιμοποιημένη κλίμακα στην έρευνα ερευνών. Στην κλίμακα τύπου Likert ο ερωτώμενος καλείται να εκφράσει τον βαθμό συμφωνίας του ως προς μια άποψη.

Στην προκειμένη περίπτωση το ερωτηματολόγιο αποτελείται από καταφατικές δηλώσεις και όχι από ερωτήσεις.
Η κλίμακα ονομάζεται κατόπιν Rensis Likert, ο οποίος δημοσίευσε μια έκθεση περιγράφοντας τη χρήση του. Η κλίμακα παρουσιάζεται σε ευθεία γραμμή με σταθερές διαιρέσεις (συνήθως 5). Το σχήμα ενός χαρακτηριστικού five-level στοιχείου Likert είναι:

1. Έντονα διαφωνείτε
2. Διαφωνείτε
3. Ούτε συμφωνείτε ούτε διαφωνείτε
4. Συμφωνείτε
5. Έντονα συμφωνείτε

Να σημειωθεί ότι η κλιμάκωση από 1 έως 5 είναι ενδεικτική. Πολλοί ερευνητές έχουν χρησιμοποιήσει έως και δεκαβάθμιες κλίμακες.

- Κλίμακα Guttman

Στη κλίμακα Guttman, μετράται η ένταση και η βαρύτητα της έννοιας και εξετάζεται ο διαφορετικός βαθμός όπου ένα άτομο συνδέεται με μια άλλη ομάδα. Οι ερωτήσεις τίθενται σε ιεραρχική διάταξη και εκφράζουν μειωμένη ή αυξανόμενη στήριξη. Κάθε ερώτηση συνοδεύεται από δύο απαντήσεις, ενώ οι απαντήσεις μπορούν να συνοψιστούν σε ένα κοινό αριθμό, ο οποίος αντιπροσωπεύει το πιο κοντινό αποδεκτό αποτέλεσμα.

- Κλίμακα Thurstone

Κατασκευάζει την κλίμακα με την συμμετοχή των ίδιων των ερωτώμενων. Εκφράζει την αντίληψη των ερωτώμενων περισσότερο από τις άλλες κλίμακες. Ο ερευνητής διατυπώνει 50-100 ερωτήσεις, τις διανέμει και παρατηρεί ποιες έχουν την μεγαλύτερη συχνότητα.

- Κλίμακα Semantic Differential

Στην ελληνική γλώσσα αποδίδεται ως σημασιολογικό διαφορικό και έγκειται στον καθορισμό των δύο άκρων ενός χαρακτηριστικού (π.χ. μορφωμένος – μη μορφωμένος), καθιστώντας εφικτή τη δυνατότητα διαβαθμίσεων ανάμεσα τους.

- Κλίμακα Borgadus

Είναι μια κοινωνική - ψυχολογικήκλίμακα δημιουργημένη από τον Emory S. Bogardus. Εμπειρικά μετρά την προθυμία των ανθρώπων να συμμετέχουν σε κοινωνικές επαφές των ποικίλων βαθμών στενότητας με τα μέλη των διαφορετικών κοινωνικών ομάδων, όπως άλλες φυλετικές και εθνικές ομάδες, παραβάτες φύλων, και ομοφυλόφιλους.

Η κλίμακα ρωτά τους ανθρώπους το βαθμό στον οποίο θα δέχονταν κάθε ομάδα (ένα αποτέλεσμα 1.00 για μια ομάδα λαμβάνεται για να μην δείξει καμία κοινωνική απόσταση):

- Σαν στενούς συγγενείς από το γάμο (αποτέλεσμα 1.00)
- Σαν τους στενούς προσωπικούς φίλους μου (2.00)
- Σαν γείτονες στην ίδια οδό (3.00)
- Σαν συναδέλφους στο ίδιο επάγγελμα (4.00)
- Σαν πολίτες στη χώρα μου (5.00)
- Σαν μόνο επισκέπτες στη χώρα μου (6.00)
- Θα απέκλειε από τη χώρα μου (7.00)

Η κοινωνική κλίμακα απόστασης Bogardus είναι μια συσσωρευτική κλίμακα (Κλίμακα Guttman), επειδή η συμφωνία με οποιοδήποτε στοιχείο υπονοεί τη συμφωνία με όλα τα προηγούμενα στοιχεία. Η κλίμακα έχει επικριθεί όπως πάρα πολύ απλή επειδή η κοινωνική απόσταση στις οικείες σχέσεις μπορεί να μην είναι στις τοποθετήσεις σχετικά με τις μακρινές επαφές, όπως οι πολίτες ή οι επισκέπτες στη χώρα μου.

Οι κλίμακες μετρήσεων θα πρέπει να ικανοποιούν δύο βασικές αρχές: την αξιοπιστία και την εγκυρότητα.

Πιο συγκεκριμένα:

Αξιοπιστία θεωρούμε:

- Τις επαναληπτικές ερωτήσεις οι οποίες να δίνουν το ίδιο αποτέλεσμα.
- Οι ερωτήσεις να πραγματοποιούνται σε διαφορετικές χρονικές περιόδους (μέθοδος ελέγχου – επανελέγχου).
- Εναλλασσόμενοι τύποι ερωτήσεων : δύο διαφορετικές αλλά ισοδύναμες κλίμακες, δηλαδή ίδιο δείγμα την ίδια στιγμή.
- Μέθοδος διχοτόμησης: δύο ίσα μέρη ερωτήσεων και έλεγχος ομοιογένειας των απαντήσεων.

Εγκυρότητα θεωρούμε:

- Απόλυτη εγκυρότητα: καλύπτονται τα λεγόμενα ή όχι.
- Εμπειρική εγκυρότητα: χρησιμοποιείται η θεωρία ως εργαλείο μέτρησης.
- Επιφανειακή εγκυρότητα: εάν τα κριτήρια ανταποκρίνονται στο θέμα.
- Εγκυρότητα περιεχομένου: όταν καλύπτονται όλες οι πτυχές του φαινομένου που εξετάζεται.
- Εγκυρότητα κατασκευής της έννοιας: εάν η έννοια συμφωνεί με την θεωρία.

1.6 Έννοια Στατιστικής Μεταβλητής – Διακρίσεις αυτής

Είναι μια καλά ορισμένη μετρήσιμη έκφραση ενός χαρακτηριστικού που μας ενδιαφέρει. Οι αριθμοί ή οι άλλες συμβολικές εκφράσεις που αντιπροσωπεύουν τις διάφορες καταστάσεις μιας μεταβλητής ονομάζονται τιμές της μεταβλητής. Κάθε μεταβλητή συμβολίζεται με ένα από τα κεφαλαία γράμματα X , Y, Z, ενώ οι τιμές, αν αυτή είναι ποσοτική, με τα αντίστοιχα μικρά γράμματα x_1, x_2, $x_3 \ldots x_\nu$ ή y_1, y_2, y_3, $..y_\kappa$. Αν η μεταβλητή είναι ποιοτική εκφράζεται με λέξεις.

Οι μεταβλητές χωρίζονται κυρίως σε δύο κατηγορίες:

1. Ποιοτικές μεταβλητές: Οι οποίες δεν επιδέχονται μέτρηση και οι τιμές τους εκφράζονται με λέξεις. Τέτοιες μεταβλητές είναι π.χ. «η οικογενειακή κατάσταση ενός υπαλλήλου», «η κατάσταση υγείας ενός μαθητή», «το επάγγελμα ενός ατόμου» κτλ.
2. Ποσοτικές μεταβλητές: Οι οποίες επιδέχονται μέτρηση και οι τιμές τους είναι αριθμοί αναφερόμενοι σε συγκεκριμένες μονάδες. Τέτοιες μεταβλητές είναι π.χ. το βάρος ή το ύψος ενός μαθητή, η ηλικία ή η το εισόδημα ενός ατόμου, η θερμοκρασία, οι εξαγωγές, ο αριθμός των δωματίων ενός διαμερίσματος κτλ. Αν μια ποσοτική μεταβλητή σημειωθεί με το γράμμα Χ, οι τιμές της θα σημειώνονται με χ_1, χ_2, $\chi_3 \ldots \chi_\nu$.

Άλλη τυπολογία για διάκριση μεταβλητών:

Συνεχείς και ασυνεχείς μεταβλητές ανάλογα με το αν η μονάδα μέτρησης μπορεί να διαιρεθεί. Ασυνεχείς ονομάζονται οι μεταβλητές εκείνες που μπορούν να λάβουν πεπερασμένο ή αριθμήσιμο πλήθος τιμών. Συνεχείς ονομάζονται οι μεταβλητές εκείνες που μπορούν να λάβουν όλες τις τιμές ενός διαστήματος.

Παράδειγμα ασυνεχών μεταβλητών: επαγγελματικό status, τόπος διαμονής, αριθμός επισκεπτών σε ένα μουσείο.

Παράδειγμα συνεχών μεταβλητών: ηλικία, εισόδημα, απόσταση.

2. ΣΥΛΛΟΓΗ ΣΤΑΤΙΣΤΙΚΩΝ ΣΤΟΙΧΕΙΩΝ

2.1 Πηγές συλλογής στατιστικών στοιχείων

Το πρώτο και σημαντικότερο στάδιο για τη μελέτη ενός φαινομένου, με τη βοήθεια της Στατιστικής, είναι η συγκέντρωση των στατιστικών στοιχείων για το φαινόμενο που μας ενδιαφέρει να εξετάσουμε και να αναλύσουμε.

Το στάδιο αυτό χρειάζεται ιδιαίτερη προσοχή και φροντίδα, γιατί από την αξία των στοιχείων που θα συγκεντρωθούν θα εξαρτηθεί η αξία των στατιστικών συμπερασμάτων. Αν τα στοιχεία είναι ψεύτικα ή λανθασμένα, είναι φανερό ότι και η αξία της στατιστικής τους ανάλυσης θα είναι κι αυτή ψεύτικη ή λανθασμένη.

Η συγκέντρωση των στατιστικών στοιχείων μπορεί να γίνει από πολλές πηγές, π.χ. από διάφορα κέντρα και Ινστιτούτα ερευνών, από Δημόσιους και Ιδιωτικούς Οργανισμούς, από Βιομηχανικά και Εμπορικά Επιμελητήρια, από Τράπεζες, Δημόσιες Υπηρεσίες, Διεθνείς Οργανισμούς κτλ.

Στη χώρα μας η μεγαλύτερη πηγή για παροχή στατιστικών στοιχείων είναι η Εθνική Στατιστική Υπηρεσία Ελλάδος (Ε.Σ.Υ.Ε.)

2.1.1 Μέθοδοι συλλογής στατιστικών στοιχείων

2.1.1.1 Πρωτογενής στατιστική έρευνα

Η πρωτογενής στατιστική έρευνα αναφέρεται στη συλλογή των στατιστικών δεδομένων απευθείας από τις μονάδες που διαθέτουν τα ερευνώμενα χαρακτηριστικά, δηλαδή τα νοικοκυριά, τις επιχειρήσεις, κλπ..

Οι μέθοδοι πρωτογενούς συλλογής στατιστικών στοιχείων διακρίνονται σε δύο βασικές κατηγορίες: στις απογραφές και στις δειγματοληπτικές έρευνες,

Το στάδιο της συλλογής στατιστικών δεδομένων πρωτογενώς καλείται διεθνώς “field work” που ερμηνεύεται ορθότερα ως «εργασίες υπαίθρου» αντί του «έρευνα πεδίου» που έχει ανεπιτυχώς καθιερωθεί στην Ελλάδα..

1. Απογραφή:

Η διαδικασία με την οποία συλλέγονται όλες οι παρατηρήσεις του πληθυσμού σε μία συγκεκριμένη χρονική στιγμή ονομάζεται απογραφή. Η απογραφή δεν είναι δυνατή παρά μόνο όταν ο πληθυσμός είναι καλά ορισμένος και πεπερασμένος. Σε μια τέτοια περίπτωση η γνώση μας για τον πληθυσμό είναι καθολική. Σύμφωνα με το αντικείμενο που ενδιαφερόμαστε, οι απογραφές παίρνουν και διαφορετικό χαρακτήρα όπως:

- Οι Γεωργικές απογραφές, στις οποίες συγκεντρώνονται πληροφορίες για τις εκτάσεις που καλλιεργούνται, το είδος της γεωργικής παραγωγής, τον αριθμό των γεωργικών μηχανημάτων, το είδος των λιπασμάτων κλπ.
- Οι Δημογραφικές απογραφές στις οποίες συλλέγονται στοιχεία σχετικά με το φύλο, την ηλικία, το επάγγελμα κλπ.
- Οι Οικονομικές απογραφές, όπου συγκεντρώνονται στοιχεία σχετικά με την οικονομική κατάσταση (δαπάνες, εισοδήματα,…) των στατιστικών μονάδων που απογράφονται.
- Οι Βιομηχανικές απογραφές, κατά τις οποίες δεν περιοριζόμαστε στην απλή καταγραφή των υπαρχόντων βιομηχανικών καταστημάτων, αλλά συγκεντρώνουμε πληροφορίες που αφορούν το κλάδο της οικονομική τους δραστηριότητας, τον αριθμό απασχολούμενων, το επίπεδο μηχανοργάνωσης κτλ.

Τα σπουδαιότερα χαρακτηριστικά ενός πληθυσμού που μελετάμε με τη 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 κλπ.

Είναι δυνατό ακόμη και όταν ο πληθυσμός είναι πεπερασμένος, να μην προτιμηθεί η μέθοδος της απογραφής. Αυτό οφείλεται στα μειονεκτήματα των απογραφών τα οποία είναι:

i. Το μεγάλο κόστος. Για να πραγματοποιηθεί μία απογραφή χρειάζεται ειδική προεργασία καθώς και μεγάλος αριθμός εκπαιδευμένων απογραφέων (τα άτομα που θα συγκεντρώνουν τα στοιχεία). Γι' αυτή την προεργασία αλλά και την ίδια ημέρα της απογραφής απαιτούνται μεγάλες δαπάνες. Αυτός είναι και ο λόγος που η απογραφή πληθυσμού γίνεται κάθε δεκαετία, ενώ η οικονομικότερη απογραφή βιομηχανιών και βιοτεχνών κάθε πενταετία.
ii. Τα εξειδικευμένα άτομα. Για να πραγματοποιηθεί μια απογραφή είναι απαραίτητο να υπάρχει ένας μεγάλος αριθμός ειδικευμένων ατόμων, των απογραφέων, ώστε να αποφευχθούν τα προσωπικά σφάλματα. Επειδή όμως πολλές φορές δεν είναι δυνατό, λόγω της έκτασης των απογραφών, να διενεργούνται μόνο από εξειδικευμένα άτομα, των οποίων η αυστηρή και πλήρης εκπαίδευση αποκλείει προσωπικά σφάλματα, η χρήση μη ειδικευμένων απογραφέων έχει ως συνέπεια τη συγκέντρωση πληροφοριών που λόγω των προσωπικών σφαλμάτων είναι δυνατό να δώσουν απατηλή εικόνα της διάρθρωσης του πληθυσμού.
iii. Επικαιρότητα αποτελεσμάτων. Ο μεγάλος αριθμός τόσο των πληροφοριών όσο και των ατόμων που αποτελούν τον πληθυσμό στις απογραφές δεν επιτρέπουν να έχουμε τη δυνατότητα να δημοσιεύσουμε σύντομα τα αποτελέσματα. Έτσι πολλές φορές, παρά τη μηχανογραφική επεξεργασία των στοιχείων, τα αποτελέσματα χάνουν την επικαιρότητα τους και περιορίζονται στην ιστορική τους μόνο αξία.

Τέλος σημειώνεται ότι σε ορισμένες περιπτώσεις η παρατήρηση ενός χαρακτηριστικού μίας στατιστικής μονάδας συνεπάγεται καταστροφή της. Για παράδειγμα η εξέταση της διάρκειας ζωής μιας παρτίδας ηλεκτρικών λαμπτήρων μπορεί να γίνει μόνον δειγματοληπτικά. Έτσι, ενώ θεωρητικά είναι επιθυμητό να έχουμε κάθε μέτρηση του πληθυσμού, στην πράξη θα έχουμε μόνον ένα, συνήθως μικρό, μέρος αυτών των μετρήσεων.

2. Δειγματοληψία

Δειγματοληψία είναι η απογραφή ορισμένων συγκεκριμένων χαρακτηριστικών ενός τμήματος του πληθυσμού. Το τμήμα του πληθυσμού που απογράφεται ονομάζεται δείγμα.

Η διαδικασία δειγματοληψίας περιλαμβάνει διάφορα στάδια:

- Καθορισμός του πληθυσμού της ανησυχίας
- Διευκρίνιση του πλαισίου δειγματοληψίας, του συνόλου από τα στοιχεία ή τα γεγονότα που πιθανά να μετρήσουν
- Διευκρίνιση της δειγματοληπτικής μεθόδου για την επιλογή των στοιχείων ή των γεγονότων από το πλαίσιο
- Καθορισμός του μεγέθους δειγμάτων
- Εφαρμογή του σχεδίου δειγματοληψίας
- Δειγματοληψία και συλλογή στοιχείων
- Αναθεώρηση της διαδικασίας δειγματοληψίας

Σκοπός των δειγματοληπτικών ερευνών είναι να προσδιορίσουμε όσο γίνεται ακριβέστερα ιδιότητες του πληθυσμού, μελετώντας τα στοιχεία του δείγματος. Οι εκτιμήσεις για τις ιδιότητες του πληθυσμού, σε αντίθεση με ότι συμβαίνει στις απογραφές, δεν είναι ακριβείς, αλλά αντιθέτως αποτελούν προσεγγίσεις που περιέχουν κάποιο σφάλμα. Το σύνολο των τεχνικών με τις οποίες εξάγονται συμπεράσματα για τον πληθυσμό χρησιμοποιώντας τις δειγματικές πληροφορίες αναφέρεται ως στατιστική συμπερασματολογία. Η διαδικασία αυτή κατά την οποία η γνώση για το δείγμα γενικεύεται στον πληθυσμό ονομάζεται επαγωγική στατιστική. Η επαγωγική στατιστική χρησιμοποιείται στις εμπειρικές επιστήμες όπου η διαδικασία παραγωγής νέας γνώσης συνίσταται στην εκτέλεση ενός πειράματος και στη γενίκευση των συμπερασμάτων από το πείραμα αυτό σε όλα τα όμοια πειράματα που θα μπορούσαν να γίνουν.

Η γενίκευση των πληροφοριών του δείγματος σε ολόκληρο τον πληθυσμό συνεπάγεται αβεβαιότητα η οποία μπορεί να μετρηθεί υπό την προϋπόθεση ότι το δείγμα είναι τυχαίο. Το τυχαίο δείγμα δεν είναι αυτό που έχει προκύψει στην τύχη αλλά αντίθετα, έχει επιλεγεί με καλά ορισμένες αρχές της θεωρίας πιθανοτήτων. Η τυχαία επιλογή των μονάδων του δείγματος από τον πληθυσμό σημαίνει ότι δεν υπεισέρχεται ο υποκειμενικός παράγοντας στην επιλογή ή όχι μιας στατιστικής μονάδας και κάθε μονάδα του πληθυσμού έχει γνωστή πιθανότητα να επιλεγεί σαν μονάδα του δείγματος.

Τυχαίες δειγματοληψίες είναι εκείνες, που κατά τη δειγματοληπτική συγκέντρωση των ατόμων του δείγματος, όλα τα άτομα του πληθυσμού έχουν την ίδια πιθανότητα τυχαίας επιλογής. Για να εφαρμόσουμε αυτές τις μεθόδους θα πρέπει:

- Να έχουμε στη διάθεση μας όλες τις μονάδες του πληθυσμού που θέλουμε δειγματοληπτικά να μελετήσουμε (σε μορφή κατάστασης ή σε μορφή αρχείου ηλεκτρονικού υπολογιστή).
- 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. Ανάλογα με τον τρόπο που επιτυγχάνεται η ταχύτητα οι τυχαίες δειγματοληψίες διακρίνονται σε:

- Απλές τυχαίες δειγματοληψίες.
- Συστηματικές δειγματοληψίες.
- Τυχαίες δειγματοληψίες κατά στρώματα.

Όταν κάθε στατιστική μονάδα του πληθυσμού που μελετούμε έχει την ίδια πιθανότητα να επιλεγεί σαν μονάδα του δείγματος, τότε έχουμε απλά τυχαία δειγματοληψία.

Η πιθανότητα επιλογής μιας μονάδας του πληθυσμού στο δείγμα πρέπει να είναι ανεξάρτητη από την επιλογή κάποιας άλλης μονάδας. Έτσι, η επιλογή n στατιστικών μονάδων από πληθυσμό μεγέθους N θα πρέπει να είναι ίδια με την εξαγωγή n σφαιρών από μια κάλπη που περιέχει N σφαίρες αριθμημένες από το 1 ως το N.

Η απλά τυχαία δειγματοληψία μπορεί να πραγματοποιηθεί και με τη χρήση των πινάκων τυχαίων αριθμών. Οι τυχαίοι αριθμοί δημιουργούνται με ειδικά προγράμματα Η/Υ και είναι τέτοιοι ώστε ανεξάρτητα από το που ξεκινάμε, ανεξάρτητα από την κατεύθυνση στην οποία κινούμαστε, τα 10 ψηφία 0, 1, 2, …, 8, 9, εμφανίζονται τυχαία με ίση πιθανότητα.

Πολλές φορές, παρά το ότι οι στατιστικές μονάδες του πληθυσμού είναι αριθμημένες ή διατεταγμένες έτσι ώστε να θεωρούνται αριθμημένες, δεν είναι εύκολο να επιτύχουμε αντιπροσωπευτικό δείγμα εφαρμόζοντας την απλή τυχαία δειγματοληψία. Εάν π.χ. θέλουμε να επιλέξουμε δείγμα 100 πελατών από τους 1000 ενός ξενοδοχείου, που είναι αριθμημένοι με τη σειρά άφιξής τους σ’ αυτό από την 1^{η} Ιανουαρίου μέχρι την 31^{η} Δεκεμβρίου και εφαρμόσουμε την απλή τυχαία δειγματοληψία, δεν είμαστε σίγουροι για την κατανομή των ατόμων του δείγματος μέσα στο χρόνο. Σε τέτοιες περιπτώσεις επιλέγουμε τυχαία την πρώτη μονάδα του δείγματος και στη συνέχεια μέχρι να συμπληρωθούν οι n μονάδες, επιλέγουμε αυτές που ο αριθμός τους συμπίπτει με τους όρους μιας αριθμητικής προόδου, με λόγο της αριθμητικής προόδου N/n (όπου N το πλήθος των ατόμων του πληθυσμού και n το μέγεθος του δείγματος).

Μια διαδικασία της μορφής αυτής ονομάζεται συστηματική δειγματοληψία και το δείγμα συστηματικό δείγμα. Σε ένα τέτοιο σχήμα δειγματοληψίας κάθε στοιχείο του πληθυσμού έχει τη ίδια πιθανότητα να περιληφθεί στο δείγμα όμως κάθε δυνατό δείγμα ίσου μεγέθους δεν έχει την ίδια πιθανότητα να επιλεγεί. Έτσι π.χ. το 20^{ο} και 21^{ο} στοιχείο του έχουν μηδενική πιθανότητα να περιληφθούν στο ίδιο δείγμα. Η συστηματική δειγματοληψία είναι χρήσιμη όχι μόνο για καταγραμμένα σε ένα πλαίσιο στοιχεία αλλά και για παραγόμενα εν σειρά προϊόντα, ή για μονάδες διατεταγμένες στο χώρο. Έτσι, π.χ. για να ελέγξουμε την ποιότητα ενός προϊόντος που παράγεται μαζικά μπορούμε να παίρνουμε για δειγματοληπτικό έλεγχο κάθε 30^{ο} παραγόμενο προϊόν ή ένα προϊόν ανά τέταρτο της ώρας από τη σειρά παραγωγής.

Η συστηματική δειγματοληψία μπορεί να δημιουργήσει μεροληψία στις ακόλουθες περιπτώσεις:

A. Όταν τα στοιχεία του πληθυσμού είναι διατεταγμένα σύμφωνα με κάποιο κριτήριο οπότε το σημείο εκκίνησης επιδρά συστηματικά στο αποτέλεσμα. Έτσι, π.χ. αν τα άτομα είναι διατεταγμένα σύμφωνα με το εισόδημα τους το δείγμα που θα ξεκινήσει από το 1^{ο} άτομο και θα περιλαμβάνει το 51^{ο}, 101^{ο} κοκ. θα δώσει χαμηλότερη εκτίμηση του μέσου εισοδήματος από το δείγμα που θα ξεκινήσει από το 50^{ο}. Στη περίπτωση αυτή η συστηματική δειγματοληψία θα πρέπει να αρχίζει από την αρχική μεσαία τιμή ή θα πρέπει να επιχειρείται άλλο σχήμα δειγματοληψίας.
B. Όταν το πλαίσιο έχει περιοδικότητα υποπολλαπλάσια του μεγέθους του δείγματος. Έτσι π.χ. αν θελήσουμε να πάρουμε τυχαίο δείγμα 30 ημερών με συστηματική δειγματοληψία τότε θα καταλήξουμε να παρατηρούμε την ίδια μέρα κάθε μήνα.

Όταν ο πληθυσμός μπορεί να διαιρεθεί σε στρώματα ομοιογενή ως προς το χαρακτηριστικό που μας ενδιαφέρει τότε μπορούμε να πάρουμε ένα τυχαίο δείγμα από κάθε στρώμα, οπότε στο συνολικό δείγμα που προκύπτει αντιπροσωπεύονται όλα τα στρώματα. Η μέθοδος αυτή ονομάζεται στρωματοποιημένη δειγματοληψία.

Η στρωματοποίηση μπορεί να γίνει κατά φύλο, κατά ηλικία, κατά επάγγελμα, κατά το οικογενειακό εισόδημα κτλ. Η στρωματοποιημένη δειγματοληψία συνιστάται ιδιαίτερα στις περιπτώσεις που υπάρχει διαφορετικό πλαίσιο για κάθε στρώμα. Τέτοιες είναι οι περιπτώσεις των τμημάτων ενός πανεπιστημίου, των σχολείων μιας εκπαιδευτικής περιφέρειας ή των εργοστασίων μιας παραγωγικής επιχείρησης.

Να σημειωθεί ότι όταν οι διαφορές μεταξύ των στρωμάτων είναι μικρές συγκρινόμενες με τις διαφορές μέσα στα στρώματα τότε η στρωματοποιημένη δειγματοληψία δεν αναμένεται να μας δώσει ακριβέστερη εικόνα του πληθυσμού από την απλή τυχαία δειγματοληψία.

Κατά τη διεξαγωγή των στατιστικών μελετών επιδιώκουμε να μειώνουμε στο ελάχιστο τις δυσκολίες που συναντάμε κατά τη συγκέντρωση των στατιστικών δεδομένων και να οδηγούμαστε σε όσο το δυνατό πιο αξιόπιστα συμπεράσματα. Για να μειωθούν αυτές οι δυσκολίες και να αυξηθεί η αξιοπιστία των συμπερασμάτων θα πρέπει να λαμβάνονται ορισμένες προφυλάξεις. Στο στάδιο της δειγματοληψίας θα πρέπει να ληφθούν υπ όψιν τα παρακάτω σφάλματα στη επιλογή του δείγματος:

- Το αυτοεπιλεγόμενο δείγμα έχει την τάση να περιλαμβάνει κυρίως άτομα με φανατισμό ή τουλάχιστον με πολύ ισχυρή γνώμη για το ερευνούμενο θέμα. Έτσι π.χ. σε τηλεοπτικές εκπομπές λόγου, οι τηλεθεατές έχουν την δυνατότητα να «ψηφίζουν» για ένα θέμα, συνήθως με ένα «ναι» ή με ένα «όχι» τηλεφωνώντας με δική τους χρέωση. Είναι προφανές ότι ένα τέτοιο δείγμα περιλαμβάνει μόνον αυτούς που έχουν ισχυρή άποψη και συγχρόνως δεν υπολογίζουν το οικονομικό κόστος του τηλεφωνήματος.
- Δειγματοληψία σε ορισμένο σημείο του δρόμου και σε ορισμένη ώρα της ημέρας αποκλείει όλους εκείνους, οι οποίοι τη συγκεκριμένη ώρα έχουν λόγους να μην βρίσκονται στο δρόμο, είτε διότι εργάζονται, είτε διότι έχουν μειωμένη κινητικότητα, είτε επειδή το καταναλωτικό προφίλ τους δεν ταιριάζει με αυτό του συγκεκριμένου δρόμου.

- Ομοίως δειγματοληψία οικιών μέσω τηλεφώνου 8-2 το πρωί έχει ως αποτέλεσμα, μόνο οι οικογένειες με ένα τουλάχιστον μη εργαζόμενο μέλος να αντιπροσωπεύονται στο δείγμα.

2.1.2 Επιλογή μεθόδου πρωτογενούς συλλογής στατιστικών δεδομένων

Η επιλογή της μεθόδου, με την οποία θα συλλέξουν πρωτογενώς τα δεδομένα της έρευνας, θα γίνει αφού συνεκτιμηθούν τα πλεονεκτήματα και τα μειονεκτήματα κάθε μιας όπως αυτά παρουσιάζονται στη συνέχεια.

Βασική πάντως παράμετρος, όπως και σε όλες τις μεθοδολογικές επιλογές της έρευνας, είναι το είδος και οι επιμέρους στόχοι αυτής, όπως και το κόστος, το διαθέσιμο προσωπικό και ο χρόνος.

Οι βασικές μέθοδοι διενέργειας των ερευνών, δειγματοληπτικών ή απογραφικών, που εφαρμόζονται διεθνώς, είναι οι ακόλουθες:

Ø Με προσωπική συνέντευξη

Ø Με συμπλήρωση των ερωτηματολογίων από τους ερευνώμενους

Ø Με το ταχυδρομείο

Ø Με τηλεφωνική συνέντευξη

Με προσωπική συνέντευξη

Σύμφωνα με τη μέθοδο αυτή, τα στατιστικά δεδομένα συλλέγονται από ερευνητή, ο οποίος συμπληρώνει ειδικό ερωτηματολόγιο κατά τη διάρκεια συνέντευξης που παίρνει από κατάλληλο πρόσωπο των ερευνώμενων μονάδων (υπεύθυνο νοικοκυριού, καταστήματος, επιχείρησης, κλπ.).

Η προσωπική επικοινωνία του ερευνητή με τον ερευνώμενο, δίνει τη δυνατότητα συλλογής του μεγαλύτερου όγκου στοιχείων. Επίσης, με τη μέθοδο αυτή είναι δυνατόν να χρησιμοποιηθεί μεγαλύτερος σχετικά αριθμός τεχνικών συνέντευξης, ανάλογα με τις ιδιαίτερες ανάγκες της έρευνας.

Στα μειονεκτήματα της μεθόδου περιλαμβάνονται το υψηλό κόστος και ο μεγάλος χρόνος που απαιτείται για τη συλλογή των δεδομένων. Επίσης, δίνει τη δυνατότητα για περισσότερα λάθη καθώς είναι η πλέον σύνθετη στην εκτέλεσή της και βασίζεται σχεδόν αποκλειστικά στις ικανότητες και στο ενδιαφέρον του ερευνητή.

Τα προβλήματα των ερευνητών μεγαλώνουν όταν η έρευνα αναφέρεται σε επαγγελματικά ή επιστημονικά θέματα που απαιτεί ειδικές γνώσεις και τη χρησιμοποίηση ειδικής ορολογίας κατά τη συνέντευξη, όπως για παράδειγμα, όταν ο ερευνητής παίρνει συνέντευξη από ιατρούς για ιατρικά θέματα

Τα μειονεκτήματα της μεθόδου με προσωπική συνέντευξη και κυρίως η ανάγκη για μείωση του κόστους, οδηγεί σε παρεμφερείς τεχνικές όπως αυτή στην οποία ο ερευνητής συναντά τους ερευνώμενους κατά ομάδες.

Συμπλήρωση των ερωτηματολογίων από τους ερευνώμενους

Σύμφωνα με τη μέθοδο αυτή, ο ερευνητής επισκέπτεται τις προς έρευνα μονάδες μόνο για να τους παραδώσει τα ερωτηματολόγια, τα οποία συμπληρώνονται από τους ίδιους τους ερευνώμενους κατά την ημέρα αναφοράς της έρευνας. Μαζί με τα ερωτηματολόγια δίνεται συνήθως και το σχετικό ενημερωτικό υλικό με βάση το οποίο θα ενεργήσει ο ερευνώμενος.

Ακολουθεί νέα επίσκεψη του ερευνητή για τη συγκέντρωση των συμπληρωθέντων ερωτηματολογίων, αφού προηγουμένως ελέγξει την πληρότητα αυτών και την ορθότητα των στοιχείων που έχουν δοθεί.

Με τη μέθοδο αυτή, που χρησιμοποιείται συνήθως σε απογραφές ή σε μεγάλες έρευνες, περιορίζεται το κόστος της έρευνας και ο χρόνος διενέργειάς της, ενώ παράλληλα περιορίζεται ο ρόλος του ερευνητή και αποφεύγονται, σε ένα βαθμό, τα αμερόληπτα λάθη που είναι δυνατόν να προκύψουν από αδυναμίες της επικοινωνίας μεταξύ ερευνώμενου και ερευνητή.

Από την άλλη μεριά για την επιτυχία της μεθόδου απαιτείται υψηλό μορφωτικό επίπεδο του πληθυσμού που ερευνάται, προϋπόθεση που δεν εξασφαλίζεται εύκολα.

Η μέθοδος αυτή αποτελεί μια ενδιάμεση λύση μεταξύ της μεθόδου με προσωπική συνέντευξη και αυτής που γίνεται μέσω αποστολής των ερωτηματολογίων με το ταχυδρομείο, συγκεντρώνοντας σημαντικά πλεονεκτήματα τόσο της πρώτης μεθόδου (δυνατότητα παρέμβασης του ερευνητή σε τυχόν ελλείψεις, χρήση σχετικά εκτεταμένου ερωτηματολογίου, κλπ.) όσο και της δεύτερης (περιορισμός της κυρίαρχης θέσης του ερευνητή, περιορισμός του κόστους, του χρόνου, κλπ.).

Με το ταχυδρομείο

Στις ταχυδρομικές έρευνες τα ερωτηματολόγια στέλνονται ταχυδρομικά στους ερευνώμενους, οι οποίοι αφού τα συμπληρώσουν τα επιστρέφουν και πάλι ταχυδρομικώς. Για αυτούς που δεν απαντούν στο προβλεπόμενο διάστημα, προβλέπεται σειρά υπομνήσεων και πιθανώς νέα αποστολή του υλικού της έρευνας.

Η μέθοδος με το ταχυδρομείο, που χρησιμοποιείται προ πολλού στις ΗΠΑ, θεωρείται γενικά λιγότερο δαπανηρή και επιτρέπει την άμεση και απλή επικοινωνία του σχεδιαστή της έρευνας με το κοινό, αλλά απαιτεί υψηλό μορφωτικό επίπεδο και «στατιστική συνείδηση» του πληθυσμού, άριστα οργανωμένες ταχυδρομικές υπηρεσίες ως και άλλες προϋποθέσεις όπως:

- Εξασφάλιση πλήρως ενημερωμένων καταλόγων – πλαισίων, δεδομένου ότι λόγω απουσίας του απογραφέα δεν είναι εύκολη η αντικατάσταση των μη ανταποκρινόμενων.
- Στενή εποπτεία των υπεύθυνων της έρευνας στο προσωπικό που ετοιμάζει το προς αποστολή υλικό, προς αποφυγή αποστολής ελλιπούς υλικού (π.χ. ερωτηματολογίων, επιστολής, οδηγιών, απαντητικού φακέλου, κλπ.).
- Κατάρτιση ερωτηματολογίων με πλήρη δυνατότητα μεταφοράς των μηνυμάτων που πρέπει να λάβει ο ερευνώμενος για να δώσει ορθές και πλήρης απαντήσεις, δεδομένου και πάλι ότι δεν υπάρχει δυνατότητα διορθωτικής παρέμβασης του ερευνητή.

Γενικώς, στις ταχυδρομικές έρευνες παρατηρείται μειωμένος δείκτης ανταπόκρισης του κοινού, καθώς και προβλήματα πληρότητας των ερωτηματολογίων, για τους προαναφερθέντες λόγους.

Με τηλεφωνική συνέντευξη

Τέλος, στην τηλεφωνική έρευνα, που τα τελευταία 20 χρόνια έχει αναδειχθεί σε δημοφιλή, ιδιαίτερα στις έρευνες αγοράς και διερεύνησης της κοινής γνώμης, η επαφή μεταξύ του ερευνητή και του ερευνώμενου είναι φωνητική.

Η μέθοδος έγινε δημοφιλής λόγω της μικρής προσπάθειας που απαιτεί για την εύρεση των ερευνώμενων, η ευκολία νέας επαφής και σε οποιαδήποτε ώρα της ημέρας, η τάση του πληθυσμού να μιλά στο τηλέφωνο ευκολότερα από το να ανοίγει την πόρτα του σε ένα ξένο και βεβαίως το χαμηλότερο κόστος που απαιτείται σε σχέση με τις άλλες μεθόδους.

Η μέθοδος πάντως απαιτεί κατάλληλα προετοιμασμένους ερευνητές γιατί τα πάντα εξαρτώνται από τη φιλικότητα της φωνής και την εντύπωση που δίνει η ομιλία του και είναι εύκολο στον ερευνώμενο να διακόψει τη συνομιλία ή να μη δώσει ιδιαίτερη σημασία σ' αυτή.

Μεγάλες δυνατότητες στις τηλεφωνικές έρευνες έδωσαν οι τεχνολογικές εξελίξεις στην πληροφορική και στις τηλεπικοινωνίες. Με τη νέα τεχνολογία είναι δυνατή η κλήση όλων των εν ενεργεία τηλεφωνικών αριθμών ανεξάρτητα από το αν έχουν συμπεριληφθεί στους τηλεφωνικούς καταλόγους.

Με τη διαδικασία μάλιστα που αναφέρεται ως Random–digit dialing επιλέγονται οι αριθμοί οι οποίοι θα κληθούν, χωρίς να υπάρχει η ανάγκη καταγραφής ή αρίθμησης αυτών πριν την επιλογή του δείγματος, σύμφωνα με τις παραδοσιακές μεθόδους δειγματοληψίας.

2.2 Η συλλογή στατιστικών δεδομένων από δευτερογενείς πηγές

Στη δευτερογενή στατιστική έρευνα τα δεδομένα συλλέγονται από υπάρχοντα στοιχεία που τηρούν διάφοροι δημόσιοι και ιδιωτικοί φορείς στα διοικητικά αρχεία τους (administrative records).

Η μεθοδολογία της δευτερογενούς έρευνας ακολουθεί τα στάδια που απαιτούνται από τη γενικότερη μεθοδολογία της στατιστικής έρευνας, όπως αυτή έχει αναλυθεί. Συνήθως, όμως, στη δευτερογενή έρευνα αξιοποιείται η πληροφορία από το σύνολο των στατιστικών μονάδων, δηλαδή συνήθως αποτελούν απογραφικές έρευνες.

Η κατάρτιση του ερωτηματολογίου, ορισμένες φορές, γίνεται ανεξάρτητα από οποιαδήποτε άλλη διοικητική ενέργεια του φορέα που έχει την πληροφόρηση και ο τελευταίος αναλαμβάνει την υποχρέωση να το δώσει στο κοινό για συμπλήρωση κατά την επίσκεψή του στον φορέα.

Πολλές φορές, αν η παραπάνω διαδικασία είναι αδύνατη ή πρακτικά ανέφικτη, η πληροφορία από τις στατιστικές μονάδες συλλέγεται μέσω των εντύπων που συμπληρώνει το κοινό κατά τις συναλλαγές του με τους διάφορους δημόσιους και ιδιωτικούς φορείς (φορολογική δήλωση, κλπ.).

Η συλλογή τέτοιας μορφής στατιστικών δεδομένων είναι επίμονη και απαιτεί πλήρη συνεργασία του φορέα παραγωγής της στατιστικής πληροφόρησης και του φορέα που κατέχει ή μέσω του οποίου δύναται να συλλεχθεί η πληροφορία για το κοινό. Αφορά δε τον τρόπο συλλογής δεδομένων από τους κρατικούς φορείς, παραγωγής, στατιστικής πληροφόρησης, δηλαδή τις επίσημες στατιστικές (official statistics).

Η διαδικασία αυτή είναι δύσκολο να επιτευχθεί από τον ιδιώτη– ερευνητή έστω και αν η μελέτη του γίνεται για λογαριασμό κρατικών φορέων. Ο ιδιώτης μελετητής έχει τη δυνατότητα να αξιοποιήσει υπάρχοντα δεδομένα που έχουν μάλλον συλλέγει για στατιστικούς σκοπούς. Στοιχεία που έχουν συλλέγει για διοικητικούς σκοπούς είναι δύσκολο να μετατραπούν σε στατιστική πληροφόρηση, είτε για μεθοδολογικούς λόγους (ακατάλληλη μορφή των δεδομένων για τις ανάγκες της μελέτης) είτε για πρακτικούς (αδυναμία πρόσβασης λόγω εμπιστευτικότητας, κλπ.).

Για τους λόγους αυτούς, ο ερευνητής αποφεύγει να αξιοποιεί τέτοια στοιχεία και διενεργεί πρωτογενή έρευνα.

Πάντως, η επιλογή μεταξύ πρωτογενούς ή δευτερογενούς στατιστικής έρευνας θα πρέπει να γίνει αφού διερευνηθεί η διαθεσιμότητα και καταλληλότητα των υπαρχόντων στοιχείων, είτε από στατιστικές πηγές (Στατιστικές Υπηρεσίες, κλπ.) είτε από διοικητικές πηγές, δεδομένου ότι με τον τρόπο αυτό εξοικονομείται κόστος

2.3 Συλλογή στατιστικών δεδομένων με τη μέθοδο των μητρώων

Η μέθοδος αυτή αφορά κυρίως στη συλλογή στατιστικών δεδομένων των επίσημων φορέων στατιστικής πληροφόρησης (official statistics).

Η βασική ιδέα του πληροφοριακού συστήματος βασισμένου στα πληθυσμιακά μητρώα (population registers) είναι ότι η συλλογή των στατιστικών στοιχείων που είναι απαραίτητα για τη παραγωγή στατιστικών πληροφοριών είναι δυνατόν να προέλθει από τα στοιχεία που έχουν κατά καιρούς συλλεχθεί από τις διοικητικές υπηρεσίες στα πλαίσια της λειτουργίας τους.

Τη βάση του συστήματος στο διεθνή χώρο αποτελούν τα οργανωμένα διοικητικά μητρώα που λειτουργούν στους Δήμους και τις Κοινότητες και αφορούν στον πληθυσμό, στις οικονομικές μονάδες, στις κατοικίες, κλπ. Με βάση τα στοιχεία των τοπικών μητρώων συγκροτούνται στις Στατιστικές Υπηρεσίες γενικά πληθυσμιακά, κλπ. μητρώα, με περιεχόμενο αντίστοιχα στοιχεία που καλύπτουν το σύνολο του πληθυσμού και των δραστηριοτήτων της χώρας.

Η συγκρότηση ενός στατιστικού συστήματος που βασίζεται στην κατάρτιση μητρώων προϋποθέτει (Jensen P., 1983):

(α) Την ύπαρξη ενός οργανωμένου συστήματος διοίκησης σε εθνικό,

περιφερειακό και τοπικό επίπεδο.

(β) Τη δημιουργία κατάλληλου νομικού πλαισίου που επιτρέπει:

- Τη δυνατότητα προσπέλασης των φορέων των αρμοδίων για την παραγωγή της στατιστικής πληροφόρησης στα διοικητικά μητρώα των Δημοσίων Υπηρεσιών και Οργανισμών και των Οργανισμών Τοπικής Αυτοδιοίκησης και τη στατιστική αξιοποίηση των στοιχείων που διαθέτουν οι τελευταίοι.
- Τη δυνατότητα συμμετοχής των Στατιστικών Υπηρεσιών στη διαδικασία κατάρτισης των διοικητικών μητρώων των δημόσιων και δημοτικών αρχών.
- Τη διασφάλιση της εξυπηρέτησης αλλά και της προστασίας της ιδιωτικής ζωής του πολίτη.

(γ) Την οργάνωση και τήρηση συστηματικών μητρώων πληθυσμού, κατοικιών, επιχειρήσεων, κλπ. στους Οργανισμούς Τοπικής Αυτοδιοίκησης.

Ο βαθμός συμμετοχής των πληθυσμιακών μητρώων στη διαδικασία συλλογής στατιστικών δεδομένων ποικίλλει σημαντικά ανάλογα με το βαθμό εξασφάλισης των ανωτέρω προϋποθέσεων.

Τα πλεονεκτήματα της μεθόδου συνοψίζονται στα ακόλουθα:

1. Μείωση του κόστους της στατιστικής πληροφόρησης, επειδή τα στατιστικά στοιχεία δεν συλλέγονται από τις επιμέρους μονάδες του πληθυσμού, αλλά από έτοιμα μητρώα.
2. Αυξημένη δυνατότητα συγκέντρωσης στατιστικών στοιχείων σε ετήσια βάση.
3. Δυνατότητα συγκέντρωσης στατιστικής πληροφόρησης σε επίπεδο Δήμου ή Κοινότητας, επαρκούς για τις ανάγκες του περιφερειακού και τοπικού προγραμματισμού.
4. Περιορισμός της όχλησης του κοινού στη συλλογή των στατιστικών στοιχείων.

Τα πλεονεκτήματα αυτά έναντι των παραδοσιακών μεθόδων συλλογής στατιστικών δεδομένων, που χαρακτηρίζουν τη μέθοδο με μητρώα, έχουν οδηγήσει στη συμμετοχή της στα στατιστικά συστήματα πολλών χωρών.

Η Δανία αρχικά, που θεωρείται πρωτοπόρος στην ανάπτυξη στατιστικού συστήματος βασισμένου σε μητρώα, έχει εγκαταλείψει τις παραδοσιακές μεθόδους συλλογής στατιστικών στοιχείων ήδη από το 1976.

Επίσης, οι σκανδιναβικές χώρες (Νορβηγία, Φινλανδία, Σουηδία) κάνουν εκτεταμένη χρήση των μητρώων στη διαδικασία συλλογής στατιστικών στοιχείων και στο άμεσο μέλλον προσανατολίζονται να αντικαταστήσουν πλήρως τις παραδοσιακές μεθόδους με την οργάνωση ολοκληρωμένου πληροφοριακού συστήματος βασισμένου σε μητρώα.

Στη Νορβηγία ειδικότερα, ο συνδυασμός των μητρώων και μιας δειγματοληπτικής έρευνας, έχει επιλεγεί ως η εναλλακτική προοπτική για τη λήψη των απογραφικών στοιχείων κατά το έτος 1990.

Με τη μέθοδο αυτή, τα στατιστικά στοιχεία θα συλλεχθούν αποκλειστικά από τα μητρώα και τα διοικητικά αρχεία των διαφόρων Υπηρεσιών, τα οποία και θα διορθωθούν με τη βοήθεια ειδικής δειγματοληπτικής έρευνας, προκειμένου να βελτιωθεί η ποιότητά τους.

Επιπλέον των ανωτέρω χωρών, αρκετές ακόμη (ΗΠΑ, Καναδάς, Γερμανία, Βέλγιο, Ισραήλ, κλπ.) χρησιμοποιούν τα μητρώα που προέρχονται από τα διοικητικά αρχεία, βοηθητικά ή συμπληρωματικά για τη συλλογή των στατιστικών δεδομένων.

Τέλος, τόσο εκ μέρους των αρμοδίων Υπηρεσιών της ΕΟΚ όσο και των επιμέρους χωρών – μελών της, υπάρχει σαφής προσανατολισμός για την υιοθέτηση της νέας μεθόδου και ήδη μελετάται σε βάθος η απαιτούμενη νομική βάση του συστήματος και η διασφάλιση των πολιτών από την καταχρηστική χρησιμοποίηση ατομικών στοιχείων για μη στατιστικούς σκοπούς.

2.4Επεξεργασία στατιστικών στοιχείων

Μετά τη συγκέντρωση των στατιστικών στοιχείων από τα πολυάριθμα συμπληρωμένα ερωτηματολόγια, ακολουθεί το στάδιο της επεξεργασίας και της μηχανογραφικής οργάνωσης των δημογραφικών στοιχείων. Το στάδιο αυτό της επεξεργασίας είναι πολύ κουραστικό και δαπανηρό, γιατί περιλαμβάνει τον έλεγχο όλων των ερωτηματολογίων, για τυχόν ασυμπλήρωτες, ασαφείς, δυσανάγνωστες ή ασυμβίβαστες απαντήσεις. Ακολουθεί η διαλογή των πληροφοριών των διαφόρων χαρακτηριστικών των στατιστικών μονάδων και η εμφάνιση αυτών των χαρακτηριστικών σε κατάλληλους αριθμητικούς πίνακες με βάση διάφορα κριτήρια. Η διαλογή των στατιστικών πληροφοριών 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 με το χέρι γίνεται όταν ο αριθμός των ερωτηματολογίων και των χαρακτηριστικών είναι περιορισμένος, ενώ αν ο αριθμός των ερωτηματολογίων και των χαρακτηριστικών είναι μεγάλος, όπως συμβαίνει στα δημογραφικά προβλήματα, χρησιμοποιούμε μηχανογραφικά συστήματα (ηλεκτρονικούς υπολογιστές, ηλεκτρονικές μηχανές κλπ.)

3. ΔΕΙΓΜΑΤΟΛΗΨΙΑ

3.1 Οι κυριότερες Μέθοδοι Συλλογής Δεδομένων

Συχνά βλέπουμε όλοι μας στον ημερήσιο και περιοδικό τύπο τα αποτελέσματα ερευνών που αναφέρονται σε οικονομικά, κοινωνικά και πολιτικά φαινόμενα, όπως για παράδειγμα το μέσο μηνιαίο εισόδημα των εργαζομένων σε διάφορες χώρες, την επίδραση της ανεργίας στο κοινωνικό σύνολο, τις προτιμήσεις των ψηφοφόρων για τους υποψήφιους και άλλα.

Για τη διεξαγωγή μιας έρευνας απαιτείται η συλλογή δεδομένων τα οποία αναφέρονται σε ορισμένες χαρακτηριστικές ιδιότητες του προς μελέτη πληθυσμού. Για τη συλλογή αυτών των στοιχείων εφαρμόζονται διάφορες τεχνικές συλλογής δεδομένων, οι κυριότερες από τις οποίες είναι η απογραφή και η δειγματοληψία.

Η απογραφή συνίσταται στη συγκέντρωση στοιχείων τα οποία αναφέρονται σε ορισμένα χαρακτηριστικά γνωρίσματα από όλα τα άτομα του προς μελέτη πληθυσμού. Γενικότερα οι απογραφές πραγματοποιούνται σε πανελλήνιο επίπεδο μια φορά κάθε δέκα χρόνια. Σημειώνεται ότι η γενική απογραφή που διενήργησε η Γ.Γ ΕΣΥΕ (Γενική Γραμματεία Εθνικής Στατιστικής Υπηρεσίας Ελλάδος) το 2001 προετοιμαζόταν από τα μέλη της για πολλά χρόνια και στοίχισε περίπου 46,955,250 ευρώ, πράγμα που σημαίνει ότι η απογραφή απαιτεί αρκετά χρήματα, προετοιμασία και ανθρώπινο δυναμικό. Επομένως, η απογραφή είναι πρακτικά μια ασύμφορη , αν όχι αδύνατη τεχνική. Για αυτό το λόγο συχνά καταφεύγουμε σε εναλλακτικές τεχνικές.

Στις περιπτώσεις αυτές εφαρμόζουμε δειγματοληπτικές έρευνες, οι οποίες καλύπτουν μέρος του ερευνώμενου πληθυσμού, δηλαδή του συνόλου των ατόμων ή αντικειμένων που μας ενδιαφέρει να μελετήσουμε. Μέσω της δειγματοληψίας επιλέγεται ένα υποσύνολο του πληθυσμού, το οποίο καλείται δείγμα. Το δείγμα επιλέγεται με τέτοιο τρόπο, ώστε να περιέχει όσο το δυνατόν περισσότερα χαρακτηριστικά του πληθυσμού στο οποίο ανήκει. Αυτό σημαίνει ότι οι πληροφορίες, οι εκτιμήσεις και τα συμπεράσματα που θα βγάλουμε από το δείγμα μπορούν να γενικευτούν στον πληθυσμό.

Τα βασικά πλεονεκτήματα της δειγματοληψίας είναι:

- Εξοικονόμηση χρόνου στη συλλογή και επεξεργασία στοιχείων
- Έχει μεγαλύτερη ταχύτητα διεξαγωγής με συνέπεια τα αποτελέσματα της να αποδίδουν την τρέχουσα κατάσταση.
- Σε ορισμένες περιπτώσεις είναι η μοναδική μέθοδος έρευνας (π.χ σε ποιοτικό έλεγχο προϊόντων, όπου απαιτείται καταστροφή των υλικών).
- Χαμηλό κόστος.

Είναι λογικό να απαιτούνται λιγότερες χρηματικές μονάδες όταν εξετάζεται ένα υποσύνολο του πληθυσμού παρά όταν εξετάζουμε ολόκληρο τον πληθυσμό.

3.2 Στάδια της Δειγματοληψίας

Η διαδικασία δειγματοληψίας μιας έρευνας περιλαμβάνει τα ακόλουθα στάδια, τα οποία αποτελούν βασική προϋπόθεση για την επιτυχία της:

- Προσδιορισμός του υπό μελέτη πληθυσμού
- Εξασφάλιση δειγματοληπτικού πλαισίου
- Επιλογή μεθόδου δείγματος
- Επιλογή κατάλληλης μεθόδου δειγματοληψίας.

3.2.1 Προσδιορισμός του υπό μελέτη πληθυσμού

Το πρώτο πράγμα που πρέπει να γίνει σε μια σχεδιαζόμενη δειγματοληψία είναι να ορισθούν ακριβώς όλες οι μονάδες που περιέχονται στον πληθυσμό.

Εάν η τιμή του χαρακτηριστικού που μετράμε παίρνει πολύ μικρές ή πολύ μεγάλες τιμές, θα πρέπει οι ακραίες περιπτώσεις να απομονωθούν και να μελετηθούν χωριστά γιατί αλλιώς τα συμπεράσματά μας θα είναι εντελώς λανθασμένα.

3.2.2 Εξασφάλιση Δειγματοληπτικού Πλαισίου

Το πλαίσιο δειγματοληψίας είναι ένας κατάλογος που περιέχει όλες τις μονάδες που έχουν τη χαρακτηριστική ιδιότητα που θέλουμε να μελετήσουμε, από τον οποίο με την εφαρμογή διαφόρων μεθόδων θα επιλέξουμε το επιθυμητό δείγμα.

Ως πλαίσια δειγματοληψίας χρησιμοποιούνται, συνήθως, αρχεία που διατηρούνται για διοικητικούς σκοπούς (μητρώα αρρένων, πελατολόγιο μιας εταιρείας κ.λ.π) ή μητρώα που προκύπτουν από τις γενικές απογραφές πληθυσμού. Δεν είναι πάντα εφικτό να υπάρχει ένας τέτοιος κατάλογος όπως για παράδειγμα στην περίπτωση κατά την οποία έχουμε ποιοτικό έλεγχο. Σε αυτή την περίπτωση ο στόχος μας είναι να τον κατασκευάσουμε. Έστω ότι μελετάμε τις αφίξεις ασθενών σε νοσοκομείο που έχουν ένα συγκεκριμένο πρόβλημα υγείας, το δειγματοληπτικό μας πλαίσιο είναι το βιβλίο εισερχόμενων ασθενών όπου αναγράφονται αναλυτικά τα στοιχεία του ασθενούς καθώς και τα συμπτώματα που παρουσιάζει.

Η εξασφάλιση ενός κατάλληλου καταλόγου-πλαισίου, είναι το πρώτο βήμα στη διαδικασία επιλογής δείγματος. Γι' αυτό και ο προσδιορισμός του πλαισίου παίζει σημαντικό ρόλο στο σχεδιασμό μιας δειγματοληπτικής έρευνας. Επηρεάζει τη συλλογή της πληροφορίας και επιδρά στην αποτελεσματικότητα με την οποία λαμβάνεται ένα δείγμα. Ένας κατάλογος- πλαίσιο θεωρείται ιδανικός όταν είναι πλήρης, δεν υπάρχουν παραλήψεις ή διπλές καταχωρήσεις και έχουν αφαιρεθεί οι μονάδες που δεν ανήκουν πλέον στον πληθυσμό της έρευνας. Διαφορετικά, τα συμπεράσματα που θα γενικευθούν επαγωγικά στο σύνολο του ερευνώμενου πληθυσμού δεν θα είναι αξιόπιστα.

3.2.3 Προσδιορισμός Μεγέθους Δείγματος

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

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

- τον επιθυμητό βαθμό ακρίβειας των αποτελεσμάτων μιας έρευνας. Η επάρκεια δείγματος εξαρτάται κυρίως από τον επιθυμητό βαθμό ακρίβειας, με άλλα λόγια τι περιθώρια σφάλματος πρόκειται να δεχθούμε σε μια δειγματοληπτική εκτίμηση. Ο βαθμός ακρίβειας της εκτίμησης είναι το τυπικό σφάλμα το οποίο ισούται με $\frac{\sigma}{\sqrt{n}}$ όπου σ η τυπική απόκλιση του χαρακτηριστικού στον πληθυσμό και n το μέγεθος του δείγματος, παραδείγματος χάριν, στην περίπτωση εκτίμησης του μέσου αριθμητικού ενός χαρακτηριστικού, το τυπικό σφάλμα εκφράζει τη διασπορά του πλήθους όλων των δυνατών εκτιμήσεων $\overline{\chi}$ από το πραγματικό μέσο του πληθυσμού μ. Όσο μικρότερο είναι το τυπικό σφάλμα, τόσο μικρότερο θα είναι το δειγματοληπτικό σφάλμα ***απόλυτο*** $|\chi - \mu|$ και άρα πιο ακριβής η εκτίμηση του μέσου αριθμητικού. Επίσης το τυπικό σφάλμα εξαρτάται από το μέγεθος του δείγματος και από τη διακύμανση του χαρακτηριστικού στον ερευνώμενο πληθυσμό. Όσο μικρότερη είναι η διακύμανση ενός πληθυσμού, τόσο μικρότερο δείγμα απαιτείται για κάποια προκαθορισμένη ακρίβεια. Η διακύμανση του ερευνώμενου πληθυσμού είναι συνήθως άγνωστη. Στην περίπτωση αυτή καταφεύγουμε σε αποτελέσματα προηγούμενης έρευνας ή στη διεξαγωγή μιας πιλοτικής έρευνας.
- Τον αριθμό των μεταβλητών που πρόκειται να ερευνηθούν. Μια άλλη δυσκολία που αντιμετωπίζεται κατά τον προσδιορισμό του μεγέθους του δείγματος, είναι ότι στην πράξη συγκεντρώνονται πληροφορίες για περισσότερες από μια μεταβλητές (χαρακτηριστικά) ταυτόχρονα. Έτσι ένα δείγμα το οποίο μπορεί να είναι ικανοποιητικό για μια μεταβλητή, ενδέχεται να μην είναι για τις υπόλοιπες. Ένας τρόπος προσδιορισμού του μεγέθους του δείγματος είναι να καθοριστούν περιθώρια σφάλματος για τις μεταβλητές που θεωρούνται ως πιο σπουδαίες για την έρευνα.

Στην συνέχεια προσδιορίζεται το μέγεθος του δείγματος που χρειάζεται, ξεχωριστά για κάθε μια από τις σπουδαιότερες μεταβλητές. Αν τα μεγέθη (n) δεν διαφέρουν σημαντικά μεταξύ τους και το μεγαλύτερο από αυτά περιέχεται εντός του καθορισμένου περιθωρίου σφάλματος, επιλέγεται ως μέγεθος δείγματος. Συνήθως τα μεγέθη (n) των μεταβλητών διαφέρουν μεταξύ τους οπότε επιλέγεται το μεγαλύτερο (n) γιατί δίνει τη μεγαλύτερη ακρίβεια.

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

3.2.4 Επιλογή Μεθόδου Δειγματοληψίας

Οι δειγματοληπτικές έρευνες διακρίνονται σε δύο κατηγορίες:

1. Την τυχαία δειγματοληψία ή δειγματοληψία με πιθανότητα, όπου κάθε μονάδα του πληθυσμού έχει κάποια πιθανότητα να επιλεγεί στο δείγμα.
2. Τη μη τυχαία δειγματοληψία ή δειγματοληψία χωρίς πιθανότητα, όπου η επιλογή των δειγματοληπτικών μονάδων γίνεται κατά τρόπο μη τυχαίο.

Λόγω του ότι στη μη τυχαία δειγματοληψία η επιλογή του δείγματος γίνεται με ντετερμινιστικό τρόπο, τα αποτελέσματά της δεν είναι αντιπροσωπευτικά και άρα είναι αναξιόπιστα. Γι' αυτό έχει καθιερωθεί στην πράξη η εφαρμογή της τυχαίας δειγματοληψίας ή δειγματοληψίας με πιθανότητα, η οποία δίνει περισσότερο αντικειμενικά αποτελέσματα καθώς και τη δυνατότητα μέτρησης του δειγματοληπτικού σφάλματος.

Υπάρχουν πάρα πολλές τεχνικές ***τυχαίας*** δειγματοληψίας. Οι πιο συνηθισμένες από αυτές είναι οι ακόλουθες:

- Απλή τυχαία δειγματοληψία
- Συστηματική δειγματοληψία
- Στρωματοποιημένη δειγματοληψία
- Κατά ομάδες δειγματοληψία
- Δειγματοληψία ποσοστών

Προτού αναλύσουμε διεξοδικά τις παραπάνω τεχνικές, είναι απαραίτητο να αναφερθούμε στα σφάλματα δειγματοληψίας, τα οποία επηρεάζουν την ακρίβεια των αποτελεσμάτων μιας δειγματοληπτικής έρευνας, τα οποία ταξινομούνται σε δύο βασικές κατηγορίες:

3.3 Δειγματοληπτικά σφάλματα

Μια καλή δειγματοληπτική έρευνα πρέπει να περιλαμβάνει τον υπολογισμό των δειγματοληπτικών σφαλμάτων. Το δειγματοληπτικό σφάλμα, προκύπτει στην περίπτωση που δεν ερευνάται ολόκληρος ο πληθυσμός, αλλά ένα δείγμα αυτού. Οφείλεται στις τυχαίες κυμάνσεις της δειγματοληψίας, δηλαδή στο γεγονός ότι το δείγμα, δεν είναι δυνατόν να είναι απολύτως αντιπροσωπευτικό του πληθυσμού από τον οποίο προέρχεται και επομένως οι μετρήσεις που προκύπτουν από το δείγμα θα αποκλίνουν από τις αντίστοιχες πραγματικές του πληθυσμού. Το δειγματοληπτικό σφάλμα γίνεται μικρότερο όσο αυξάνει το μέγεθος του δείγματος.

Γίνεται επομένως σαφές ότι η ύπαρξη του δειγματοληπτικού σφάλματος είναι εγγενής της διαδικασίας της δειγματοληψίας και ενώ είναι δυνατό να μειωθεί, δεν δύναται να μηδενιστεί παρά μόνο αν η έρευνα είναι απογραφική. Θα πρέπει να γίνεται σωστή επιλογή της μεθόδου δειγματοληψίας και κατάλληλου δείγματος, ώστε τα δειγματοληπτικά σφάλματα να είναι μειωμένα.

3.4Μη δειγματοληπτικά σφάλματα

Ο όρος « μη δειγματοληπτικά σφάλματα » 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 οφείλονται συνήθως σε:

- Ατέλειες του σχεδιασμού και της οργάνωσης της έρευνας για λόγους ταχύτητας και περιορισμού του κόστους
- Λάθη των ερευνητών , κωδικογράφων, κ.λ.π
- Άλλες αντικειμενικές δυσκολίες που ανακύπτουν κατά την εκτέλεση μιας στατιστικής έρευνας.

Τα μη δειγματοληπτικά σφάλματα περιορίζουν την ποιότητα των συλλεγόμενων στοιχείων και επομένως την αξία των αποτελεσμάτων που προκύπτουν από αυτά.

3.5 Συμβολισμοί

N : αριθμός μονάδων πληθυσμού ή ανάλογα ο αριθμός των ομάδων στον πληθυσμό.

n : αριθμός μονάδων δείγματος ή ανάλογα ο αριθμός των ομάδων που έχουν επιλεγεί στο δείγμα.

X_i : η τιμή του υπό μελέτη χαρακτηριστικού της μονάδας *i*

X= $\sum_{i=1}^{n} X_i$: Άθροισμα των τιμών του χαρακτηριστικού *X*στον πληθυσμό

x= $\sum_{i=1}^{n} x_i$: Άθροισμα των τιμών του χαρακτηριστικού *x*στον πληθυσμό

μ : η μέση τιμή του πληθυσμού

$\bar{X}$: η μέση τιμή του δείγματος

σ^2 : η διακύμανση του πληθυσμού

s^2 : η διακύμανση του δείγματος

$V(\bar{X})$: η διακύμανση της δειγματικής μέσης τιμής

f = n/N : κλάσμα δειγματοληψίας

σ : τυπικό σφάλμα ή τυπική απόκλιση του πληθυσμού

S : τυπικό σφάλμα ή τυπική απόκλιση του δείγματος

SE : η εκτίμηση του τυπικού σφάλματος

$W_i = \dfrac{N_i}{N}$: στάθμιση στρώματος *i*

A : ο αριθμός των μονάδων του πληθυσμού με μια συγκεκριμένη ιδιότητα

P: ποσοστό ενός χαρακτηριστικού στον πληθυσμό

a: ο αριθμός των μονάδων του δείγματος με την ιδιότητα αυτή

p : ποσοστό ενός χαρακτηριστικού στο δείγμα

Q : ποσοστό στον πληθυσμό που δεν παρουσιάζει την ιδιότητα (Q = 1-P)

q : ποσοστό στο δείγμα που δεν παρουσιάζει την ιδιότητα (q = 1-p)

λ : βήμα ή διάστημα δειγματοληψίας (το αντίγραφο του κλάσματος δειγματοληψίας) σ_w^2 : διακύμανση μεταξύ των μονάδων που βρίσκονται στο ίδιο συστημικό δείγμα

κ : συνολικός αριθμός των στρωμάτων

N_i : μέγεθος του στρώματος *i*

$\bar{X}_i$: ο μέσος όρος των N_i μονάδων στο στρώμα *i* (μέσος στρώματος)

n_i : μέγεθος του δείγματος για το στρώμα *i*

x_{ij} : η τιμή της μεταβλητής Χ της j μονάδας (j =1,2,...N) του συστημικού δείγματος ή στρώματος *i*

M : ο αριθμός των στοιχείων στον πληθυσμό στην κατά Ομάδες Δειγματοληψία

m_i : ο αριθμός των στοιχείων στη μονάδα *i*

$\overline{M} = \frac{N}{M}$.: το μέσο μέγεθος ομάδων στον πληθυσμό

a_i : ο αριθμός των στοιχείων στην ομάδα *i* που έχουν το συγκεκριμένο χαρακτηριστικό που μας ενδιαφέρει.

3.6 ΑΠΛΗ ΤΥΧΑΙΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

3.6.1 Εισαγωγή

Η απλή τυχαία δειγματοληψία είναι μια από τις μεθόδους δειγματοληψίας. Ονομάζεται δε και επιστημονική δειγματοληψία. Είναι η μέθοδος επιλογής n – μονάδων από ένα πληθυσμό Ν – μονάδων, με τέτοιο τρόπο ώστε κάθε δυνατό δείγμα μεγέθους n να έχει την πιθανότητα να επιλεγεί.

Με βάση τον παραπάνω ορισμό, αν ο πληθυσμός αποτελείται από Ν μονάδες κι εμείς επιθυμούμε ένα δείγμα μεγέθους n, το πλήθος των δυνατών διακεκριμένων δειγμάτων θα είναι :

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

Όπου n!=1x2x3x…xn

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

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

Η μέθοδος αυτή περιλαμβάνει τα εξής 2 βήματα :

1. Αντιστοιχούμε σε κάθε μονάδα του πληθυσμού έναν αριθμό από 1 -> Ν και κατόπιν
2. Διαλέγουμε μια σειρά n τυχαίων αριθμών απο 1->Ν με τη βοήθεια πινάκων τυχαίων αριθμών.Οι πίνακες τυχαίων αριθμών ειναι πίνακες ψηφίων 0,1,2,3,4,...,9 στους οποίους η πιθανότητα επιλογής σε οποιαδήποτε δοκιμή είναι η ίδια (1/10) για κάθε ψηφίο.

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

Άρα, η πιθανότητα με την οποία και οι μονάδες του επιθυμούμενου δείγματος θα επιλεγούν είναι :

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

Παρατηρούμε ότι κάθε φορά που απομακρύνεται μια μονάδα δείγματος από τον πληθυσμό δεν επανατοποθετείται. Για τον λόγο αυτό ,η δειγματοληψία αυτή ονομάζεται και απλή τυχαία δειγματοληψία χωρίς επανάθεση.

Παράδειγμα : Σχηματική παράσταση

Δυνατοί τρόποι επιλογής μιας μονάδας	N-1	N-2	...N-j+1	1	N-j	… .	N-n+1
Δοκιμή	1	2	…j-1	j	J+1	… .	N
Διαθέσιμες μονάδες του πληθυσμού	N	N-1	...N-j+2	N-j+1	N-j	… .	N-n+1

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

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

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

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

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

δ1 = { Ε , Γ}

δ2 = { Ε , Κ}

δ3 = { Ε , Μ}

δ4 = { Γ , Κ}

δ5 = { Γ , Μ}

και

δ6 = { Κ , Μ}

Αν λοιπόν το κάθε ένα από τα παραπάνω δείγματα επιλεγεί με πιθανότητα $\frac{1}{6}$ δηλαδή

$$P_1 = P_2 = \ldots = P_6 = \frac{1}{\binom{4}{2}} = \frac{1}{6}$$

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

3.6.2 Εκτιμήσεις στην απλή τυχαία δειγματοληψία

3.6.2.1 Εκτιμήσεις του μέσου όρου του πληθυσμού

Από απλό τυχαίο δείγμα τιμών της μεταβλητής Χ, μεγέθους n, λαμβάνονται οι εξής τιμές:

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

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

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

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

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

3.6.2.2 Εκτίμηση αναλογίας πληθυσμού

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

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

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

3.6.2.3 Εκτίμηση του τυπικού σφάλματος

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

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

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

Α) Δίνει τη δυνατότητα μέτρησης του βαθμού ακριβείας της εκτίμησης της μέσης τιμής του πληθυσμού και σύγκρισης του με τον βαθμό ακριβείας που παρέχει οποιαδήποτε άλλη μέθοδος δειγματοληψίας.

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

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

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

$$\hat{\sigma}^2 = S^{*2} = \frac{\sum_{i=1}^{n}(X_i - \overline{x}_n)^2}{n-1}$$

στην περίπτωση απλού δείγματος.

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

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

είναι αμερόληπτη εκτιμήτρια της διασποράς σ^2 του πληθυσμού.

Οι στατιστικές συναρτήσεις :

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

και

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

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

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

3.6.3 Πλεονεκτήματα-Μειονεκτήματα απλής τυχαίας δειγματοληψίας

3.6.3.1 Πλεονεκτήματα

1. Είναι μια τεχνική εύκολη στην κατανόηση δεδομένου ότι οι μέθοδοι (τρόποι) επιλογής δεδομένων όπως ο πίνακας των τυχαίων αριθμών, ο κλήρος κ.τ.λ. έχουν απλό εννοιολογικό σχεδιασμό, με αποτέλεσμα να είναι η μέθοδος αυτή προσιτή ακόμα και σε εκείνους που έχουν χαμηλό στατιστικό υπόβαθρο.

2. Τα αποτελέσματα που προκύπτουν κατά την τυχαία δειγματοληψία είναι αντικειμενικά. Το κάθε δείγμα λαμβάνεται με κλήρωση, μειώνοντας έτσι την πιθανότητα της μεροληψίας και καθιστώντας την υποκειμενική φύση της επιλογής των ερωτηθέντων ασήμαντη.

3. Το δείγμα είναι τυχαίο και επομένως μπορεί να υπολογιστεί το μέγεθος του δειγματοληπτικού σφάλματος.

4. Τα αποτελέσματα της απλής τυχαίας δειγματοληψίας είναι σε μεγάλο ποσοστό σωστά στην περίπτωση που ο πληθυσμός είναι σχετικά μικρός.

3.6.3.2 Μειονεκτήματα

1. Η μέθοδος αυτή είναι αρκετά χρονοβόρα όταν ο υπό μελέτη πληθυσμός είναι μεγάλος και γεωγραφικά διασκορπισμένος.

2. Είναι δαπανηρή, επειδή απαιτεί πλήρως ενημερωμένα πλαίσια. Είναι γενικά απαραίτητο να κατασκευασθεί κατάλογος και να αριθμηθεί κάθε μονάδα του συνόλου του πληθυσμού.

3. Το κόστος που απαιτείται για τη συλλογή δεδομένων είναι μεγάλο.

4. Το δείγμα είναι μη αντιπροσωπευτικό, αν ο πληθυσμός που εξετάζεται δεν είναι αρκετά ομοιογενής. Οι εκτιμήσεις δηλαδή του δειγματοληπτικού σχεδίου έχουν μειωμένη ακρίβεια.

3.7 ΣΥΣΤΗΜΑΤΙΚΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

3.7.1Εισαγωγή

Η συστηματική δειγματοληψία χρησιμοποιείται συχνά αντί της απλής τυχαίας δειγματοληψίας, ιδίως σε μεγάλους πληθυσμούς, επειδή η διαδικασία επιλογής του δείγματος είναι πολύ απλή. Τα *N* μέλη του πληθυσμού πρέπει να είναι αριθμημένα και καταγεγραμμένα στο πλαίσιο δειγματοληψίας.

Για τη συστηματική δειγματοληψία χρειαζόμαστε δύο πράγματα:

1. το διάστημα ή βήμα δειγματοληψίας, το οποίο θα παριστούμε με *λ* και
2. ένα τυχαίο ξεκίνημα *ι.*

Εάν *N* είναι το μέγεθος του πληθυσμού, *n* το απαιτούμενο μέγεθος δείγματος και υποθέτοντας ότι $N = n*\lambda$, τότε το βήμα δειγματοληψίας είναι ακέραιος αριθμός και δίνεται από τη σχέση: $\lambda = \frac{\mathrm{N}}{n} = \frac{1}{f}$.

Εκλέγουμε τώρα τυχαία (π.χ με τη βοήθεια των πινάκων των τυχαίων αριθμών) ένα αριθμό μεταξύ των πρώτων *λ* αριθμών της λίστας. Έστω *i* ($1 \le i \le \lambda$) ο επιλεγμένος αριθμός, δηλαδή το τυχαίο ξεκίνημα. Τότε οι μονάδες του πληθυσμού που έχουν αύξοντες αριθμούς i, $i + \lambda$, $i + 2\lambda, \ldots, i + (n-1)\lambda$ απαρτίζουν το επιλεγόμενο συστηματικό δείγμα μεγέθους *n*. Η πιθανότητα επιλογής κάθε μονάδας του πληθυσμού είναι ίση με την πιθανότητα επιλογής του τυχαίου αριθμού *i* , δηλαδή ίση με $\frac{1}{\lambda} = \frac{1}{\mathrm{N}/n} = \frac{n}{\mathrm{N}} = f$ όπως δηλαδή και στην περίπτωση της απλής τυχαίας δειγματοληψίας.

Στην περίπτωση που το μέγεθος *N* του πληθυσμού είναι άγνωστο ή πολύ μεγάλο, το μέγεθος *n* του δείγματος δεν μπορεί να προσδιορισθεί. Εκείνο που προσδιορίζεται είναι το κλάσμα δειγματοληψίας *f*. Έστω για παράδειγμα ότι θέλουμε να εξετάσουμε την γνώμη των πελατών ενός υποκαταστήματος μιας τράπεζας.

Προκαθορίζουμε ότι $f = 0.1$, οπότε το βήμα της συστηματικής δειγματοληψίας είναι λ = 1/f = 10 και ερευνάμε κάθε 10 πελάτες που βγαίνουν από το υποκατάστημα, μετά από μια αρχή *i* ($1 \le i \le \lambda$) .

3.7.2 Παράδειγμα

Ας πάρουμε για παράδειγμα τη Νομική Σχολή Αθηνών. Σ'αυτή φοιτούν 9.200 φοιτητές και θέλουμε δείγμα από 400. Τότε το διάστημα δειγματοληψίας θα είναι:

$$N/n = 9.200/400 = 23,$$

δηλαδή το λ =23.

Μπορούμε τώρα να ενοποιήσουμε τα μητρώα των ετών της σχολής και, αφού βρούμε έναν αριθμό μεταξύ 1 και 23, π.χ το 12, να πάρουμε τους φοιτητές με τη σειρά αρχίζοντας από τον 12:

12, 35, 58, 81, ..., 9.189.

Αυτοί οι αριθμοί ανταποκρίνονται στη συνεχή αρίθμηση των φοιτητών που κάναμε μετά την ενοποίηση των καταλόγων.

Στην πραγματικότητα κατά τη συστηματική δειγματοληψία, ο πληθυσμός χωρίζεται σε *λ* αποπλυθείς ομάδες των *n* μονάδων η κάθε μια. Δηλαδή έχουμε $N = n*\lambda$. Τα δυνατά συστηματικά δείγματα δίνονται στον παρακάτω πίνακα.

Συστηματικά Δείγματα	Επιλεγόμενες Ομάδες
1	$x_1\ x_{1+\lambda}\ x_{1+2\lambda} \ldots.\ x_{1+(n-1)\lambda}$
2	$x_2\ x_{2+\lambda}\ x_{2+2\lambda} \ldots.\ x_{2+(n-1)\lambda}$
...	$\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$
i	$x_i\ x_{i+\lambda}\ x_{i+2\lambda} \ldots.\ x_{i+(n-1)\lambda}$
…	$\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$
Λ	$x_\lambda\ x_{2\lambda}\ x_{3\lambda} \ldots.\ x_{n\lambda}$

Η πιθανότητα να επιλεγεί ένα από τα λ συστηματικά δείγματα είναι 1/λ.

Επομένως, κάθε μονάδα του πληθυσμού έχει την ίδια πιθανότητα να περιληφθεί στο δείγμα κατά τη μέθοδο της συστηματικής δειγματοληψίας.

3.7.3 Διακύμανση της Εκτίμησης του Μέσου

Για την αποφυγή των "πολύπλοκων" δεικτών στις τιμές των λ συστηματικών δειγμάτων συμβολίζουμε με x_{ij} την *j* μονάδα του *i* συστηματικού δείγματος, όπου $j=1,2,\dots n$ και $i = 1,2,\dots\lambda$. Για την εκτίμηση $\bar{x}$ στη συστηματική δειγματοληψία, ισχύει για $N = n*\lambda$ το εξής:

$$E(\overline{x}\,) = \mu$$
$$E(\overline{x_i}) = \frac{1}{n\lambda}\sum_{i=1}^{\lambda}\ \sum_{j=1}^{n} x_{ij} = \frac{\mathrm{X}}{n\lambda} = \frac{\mathrm{X}}{\mathrm{N}} = \mu$$

Η εκτίμηση του μέσου πληθυσμού μ, γίνεται από τον μέσο $\bar{x}$ του συστηματικού δείγματος, δηλαδή έχουμε:

$$\bar{x} = \frac{\sum_{j=1}^{n} x_{ij}}{n}$$

Οι δυνατοί δειγματικοί μέσοι είναι $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_i, \dots, \bar{x}_\lambda$ οπότε η διακύμανση του μέσου $\bar{x}$ είναι εξ ορισμού:

$$V(\bar{x}) = \frac{1}{\lambda}\sum_{i=1}^{\lambda}(\bar{x}_i - \mu)^2$$

Από τα στοιχεία ενός μόνο συστηματικού δείγματος δεν μπορεί να βρεθεί αμερόληπτη εκτίμηση της διακύμανσης του δειγματικού μέσου $\bar{x}$. Επομένως η διακύμανση της εκτίμησης του μέσου $\bar{x}$, δίνεται από τη σχέση:

$$V(\bar{x}) = \frac{N-1}{N}\sigma^2 - \frac{\lambda(n-1)}{N}\sigma_W^2$$

Και επειδή : $\lambda = \frac{\mathrm{N}}{n} \Leftrightarrow n\lambda = \mathrm{N} \Leftrightarrow n = \frac{N}{\lambda} \Leftrightarrow \frac{1}{n} = \frac{1}{\mathrm{N}/\lambda} \Leftrightarrow \frac{1}{n} = \frac{\lambda}{\mathrm{N}}$

Προκύπτει : $V(\bar{x}) = \frac{N-1}{N}\sigma^2 - \frac{\lambda(n-1)}{N}\sigma_W^2$

Και

$$SE(\bar{x}) = \sqrt{V(\bar{x})}$$

Όπου

$$\sigma^2 = \frac{\sum_{i=1}^{\lambda} \sum_{j=1}^{n} (x_{ij} - \mu)^2}{N-1}$$

είναι η διακύμανση του συνολικού πληθυσμού

$$\sigma_w^2 = \frac{1}{\lambda(n-1)} \sum_{i-1}^{\lambda} \sum_{j=1}^{n} (x_{ij} - \bar{x})^2$$

είναι η διακύμανση μεταξύ των μονάδων που βρίσκονται στο ίδιο συστηματικό δείγμα, δηλαδή ο μέσος όρος των λ διακυμάνσεων, αφού η διακύμανση εντός του iσυστηματικού δείγματος είναι:

$$\sigma^2 = \frac{\sum_{j=1}^{n} (x_{ij} - \bar{x})^2}{n-1}$$

Ο μέσος ενός συστηματικού δείγματος είναι πιο ακριβής από τον μέσο ενός απλού τυχαίου δείγματος, του ίδιου μεγέθους εάν και μόνο εάν $\sigma^2 < \sigma_w^2$.

Αυτό δείχνει ότι η συστηματική δειγματοληψία είναι περισσότερο ακριβής από την απλή τυχαία δειγματοληψία, εάν η διακύμανση μέσα στα συστηματικά δείγματα είναι μεγαλύτερη από τη διακύμανση του πληθυσμού.

Δηλαδή η συστηματική δειγματοληψία είναι περισσότερο ακριβής όταν οι μονάδες των συστηματικών δειγμάτων είναι ανομοιογενείς.

Σε περίπτωση που εντός των συστηματικών δειγμάτων η διακύμανση είναι μικρή ή μηδέν τότε πρέπει να αποφεύγεται η συστηματική δειγματοληψία, διότι οι εκτιμήσεις δεν θα είναι αξιόπιστες.

Μια περίπτωση που η διακύμανση των συστηματικών δειγμάτων ισούται με μηδέν, είναι όταν οι μονάδες του πληθυσμού παρουσιάζουν περιοδικότητα, ως προς το μέγεθος του ερευνώμενου χαρακτηριστικού, με συνέπεια κάθε παρατήρηση να είναι ίδια εντός του συστηματικού δείγματος.

Έτσι το σύνολο των παρατηρήσεων δίνει την ίδια πληροφορία που θα έδινε και μια μόνο παρατήρηση.

3.7.4 Εκτίμηση του Μέσου Πληθυσμού

Όπως αναφέρθηκε και προηγούμενα, η εκτίμηση του μέσου πληθυσμού μ από ένα συστηματικό δείγμα δίνεται από τον τύπο:

$$\bar{x} = \frac{\sum_{j=1}^{n} x_{ij}}{n}$$

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

$$V(\bar{x}) = \frac{s^2}{n}\left(\frac{N-n}{N}\right)$$

Η διακύμανση s^2 είναι μια αμερόληπτη εκτίμηση του σ^2, δηλαδή:

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

3.7.5 Εκτίμηση Ποσοστού ενός Πληθυσμού

Η εκτίμηση ποσοστού ενός πληθυσμού από ένα συστηματικό δείγμα, όταν ο πληθυσμός είναι τυχαίος, γίνεται όπως και στην περίπτωση της απλής τυχαίας δειγματοληψίας, δηλαδή έχουμε:

$$p = \frac{\sum_{i=1}^{n} x_i}{n}$$

Η διακύμανση του ποσοστού p όταν κάνουμε επιλογή μονάδων, χωρίς επανατοποθέτηση είναι:

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

Όταν η διακύμανση του πληθυσμού $\sigma^2 = PQ$ είναι άγνωστη, μια αμερόληπτη εκτίμηση της διακύμανσης του ποσοστού p, προκύπτει από το δείγμα, δηλαδή:

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

3.7.6 Πλεονεκτήματα- Μειονεκτήματα Συστηματικής Δειγματοληψίας

Η συστηματική δειγματοληψία χρησιμοποιείται συχνά αντί της απλής τυχαίας δειγματοληψίας, ιδίως σε μεγάλους πληθυσμούς, επειδή η διαδικασία επιλογής του δείγματος είναι πολύ απλή.
Η ευκολία στην εφαρμογή της συνεπάγεται *μείωση σφαλμάτων συλλογής* καθώς επίσης και *περιορισμό της μεροληπτικής επιλογής των δειγματοληπτικών μονάδων*, με αποτέλεσμα τη μεγαλύτερη δυνατότητα για αντιπροσωπευτικό δείγμα.
Επίσης, η συστηματική δειγματοληψία μπορεί να δώσει πληροφορίες, σε δεδομένο κατά μονάδα κόστος, διότι ένα συστηματικό δείγμα κατά κανόνα είναι διεσπαρμένο με μεγαλύτερη ομοιογένεια στον ερευνώμενο πληθυσμό.
Επιπλέον, η συστηματική δειγματοληψία είναι πιο αποτελεσματική από την απλή τυχαία δειγματοληψία, εάν η διακύμανση μέσα στα συστηματικά δείγματα είναι μεγαλύτερη από τη διακύμανση του πληθυσμού. Δηλαδή η συστηματική δειγματοληψία είναι περισσότερο ακριβής όταν οι μονάδες των συστηματικών δειγμάτων είναι ανομοιογενείς. Όταν η διακύμανση των συστηματικών δειγμάτων είναι πολύ μικρή ή μηδέν, η εφαρμογή της συστηματικής δειγματοληψίας για την επιλογή του δείγματος πρέπει να αποφεύγεται. Μια τέτοια περίπτωση είναι όταν οι μονάδες του ερευνώμενου πληθυσμού παρουσιάζουν περιοδικότητα.
Για παράδειγμα αν οι μονάδες που έχουν επιλεγεί σε μια συστηματική δειγματοληψία

$x_2, x_{12}, x_{22,} x_{32}...$

παρουσιάζουν περιοδικότητα, τότε θα παρέχουν την ίδια πληροφορία και θα έχουν:

$$x_2 = x_{12} = x_{22} = ... = \bar{x}$$

δηλαδή η εκτίμηση του μέσου μ του πληθυσμού από ένα δείγμα μεγέθους n συνεπάγεται την ίδια ακρίβεια με μια εκτίμηση που θα βασίζεται σε δείγμα μεγέθους $n = 1$. Επομένως θα έχουμε υπερεκτίμηση ή υποεκτίμηση των υπό μελέτη χαρακτηριστικών.

3.7.7 Παράδειγμα Συστηματικής Δειγματοληψίας

Ο στόχος της συγκεκριμένης έρευνας είναι να εκτιμήσουμε κατά μέσο όρο τον αριθμό των ατόμων που εισέρχονται ημερησίως σε ένα κατάστημα κατά τη διάρκεια 120 ημερών. Επειδή η επιλογή τυχαίων αριθμών απαιτεί πολύ χρόνο θα εφαρμόσουμε την συστηματική δειγματοληψία η οποία είναι πιο εύκολη όσον αφορά την επιλογή του δείγματος. Λόγω της ευκολίας στην εφαρμογή της, αποφεύγονται τα σφάλματα και η μεροληψία, με αποτέλεσμα να έχουμε πιο αντιπροσωπευτικό δείγμα σε σχέση με την απλή τυχαία δειγματοληψία.

Έτσι λαμβάνοντας υπόψη και το κόστος αποφασίσαμε να καταμετρήσουμε τα άτομα που εισέρχονται στο κατάστημα κάθε 10 ημέρες, δηλαδή το διάστημα δειγματοληψίας είναι:

$\lambda = N/n = 120/12 = 10$.

Επομένως, ο αριθμός των ατόμων που εισέρχονται στο κατάστημα κάθε δέκατη μέρα, με τυχαίο ξεκίνημα τη δεύτερη μέρα δίνεται στον παρακάτω πίνακα:

Ημέρα	Αριθμός ατόμων που εισέρχονται στο κατάστημα	Ημέρα	Αριθμός ατόμων που εισέρχονται στο κατάστημα
2	20	62	20
12	18	72	22
22	24	82	25
32	23	92	28
42	26	102	19
52	17	112	22

Η εκτίμηση του μέσου όρου των ατόμων που εισέρχονται ημερησίως στο κατάστημα είναι:

$$\bar{x} = \frac{\sum_{j=1}^{n} x_{ij}}{n} = \frac{20+18+\ldots+19+22}{12} = \frac{264}{12} = 22$$

Η διακύμανση του $\bar{x}$ είναι:

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

Όπου:

$$s^2 = \frac{\sum_{j=1}^{n}(x_i j - \bar{x})^2}{n-1} = \frac{(20-22)^2+(18-22)^2+\ldots+(22-22)^2}{12-1} = 11,3$$

και

$$V(\bar{x}) = \frac{s^2}{n}\left(\frac{N-n}{N}\right) = \frac{11,3}{12}\left(\frac{120-12}{120}\right) = 0,85$$

Επομένως το τυπικό σφάλμα της εκτίμησης είναι:

$$SE(\bar{x}) = \sqrt{V(\bar{x})} = \sqrt{0,85} = 0.92$$

Το διάστημα εμπιστοσύνης στο οποίο βρίσκεται ο μέσος όρος των ατόμων μ που εισέρχονται στο κατάστημα με πιθανότητα 95% είναι:

$$\bar{x} \pm z_{a/2} \cdot SE(\bar{x})$$

$$\bar{x} \pm z_{a/2} \cdot SE(\bar{x})$$

$$22 \pm 1,8$$

Δηλαδή ο μέσος αριθμός ατόμων που εισέρχονται στο κατάστημα ημερησίως βρίσκεται στο διάστημα 20- 24, με πιθανότητα 95%.

3.8 ΣΤΡΩΜΑΤΟΠΟΙΗΜΕΝΗ ΤΥΧΑΙΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

3.8.1 Εισαγωγή

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

Ορισμός 1

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

Ορισμός 2

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

Παράδειγμα :

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

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

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

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

Με βάση τα παραπάνω, παρατηρούμε ότι το δείγμα αυτό των 400 σπουδαστών αντιπροσωπεύει τους σπουδαστές όλων των ετών σπουδών. Αντίστοιχο δείγμα απλό τυχαίο από τους 400 σπουδαστές θα ήταν δυνατό να αποτελείται μόνο από τους σπουδαστές του Α έτους ή μόνο από τους σπουδαστές του Β έτους κ.ο.κ.

Η στρωματοποιημένη δειγματοληψία χωρίζεται σε 2 κατηγορίες:

Α) Ανάλογη δειγματοληψία κατά στρώματα

Β) Δυσανάλογη δειγματοληψία κατά στρώματα

3.8.2 Εκτίμηση παραμέτρων στρωματοποιημένης δειγματοληψίας

3.8.2.1 Εκτίμηση του μέσου όρου του πληθυσμού

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

$$E\left[\frac{N_1}{N}\cdot\overline{X_i}+\frac{N_2}{N}\cdot X_2+\ldots+\frac{N_k}{N}\cdot\overline{X_k}\right]=E\left[\overline{X}\right]=\mu$$

Στην πιο πάνω σχέση με τα σύμβολα $\mathrm{N}_1, \mathrm{N}_2, \ldots, \mathrm{N}_\kappa$ εκφράζονται το μέγεθος κάθε στρώματος στον πληθυσμό με, N το μέγεθος ολόκληρου του πληθυσμού και με $\overline{\mathrm{X}_1}, \overline{\mathrm{X}_2}, \ldots, \overline{\mathrm{X}_\kappa}$ οι μέσοι όροι της μεταβλητής X από τις τιμές του δείγματος μέσα σε κάθε στρώμα (οι υποδείκτες 1,2,...,K συμβολίζουν κατά σειρά το πρώτο, το δεύτερο , ... , το τελευταίο από τα K στρώματα του πληθυσμού).

Η διακύμανση του σταθμικού αθροίσματος των μέσων όρων των δειγμάτων (ένα δείγμα από κάθε στρώμα) δηλαδή της X, δίνεται από τη σχέση:

$$S_{\overline{X}}{}^2=\left[\frac{N_1}{N}\right]^2\cdot\sigma^2_{\overline{x_1}}+\left[\frac{N_2}{N}\right]^2\cdot\sigma^2_{\overline{x_2}}+\ldots+\left[\frac{N_K}{N}\right]^2\cdot\sigma^2_{\overline{x_K}}$$

$$=\left[\frac{N_1}{N}\right]^2\cdot\frac{\sigma^2_{x_1}}{n_1}+\left[\frac{N_2}{N}\right]^2\cdot\frac{\sigma^2_{x_2}}{n_2}+\ldots+\left[\frac{N_K}{N}\right]^2\cdot\frac{\sigma^2_{x_K}}{n_K}$$

Στις πιο πάνω σχέσεις με σ^2 συμβολίζεται η διακύμανση της $\overline{\mathrm{X}}$ ή της Χ κάθε στρώματος με $n_1, n_2, \ldots, n_k$,το μέγεθος του δείγματος από κάθε στρώμα. Αν σ^2 για τη Χ κάθε στρώματος είναι άγνωστη (και αυτό συμβαίνει συνήθως), στη θέση της τοποθετούμε S^2, όπως την υπολογίζουμε από τα στοιχεία των δειγμάτων. Επίσης, παρατηρούμε ότι από τις πιο πάνω σχέσεις έχει παραλειφθεί ο γνωστός διορθωτικός παράγοντας. Αν συνεπώς λάβουμε υπόψη τις τιμές S^2 και τους διορθωτικούς παράγοντες, η $S^2_{\bar{x}}$ παίρνει την ακόλουθη διατύπωση:

$$S^2_{\bar{X}}=\left[\frac{N_1}{N}\right]^2\cdot\frac{Sx_1^2}{n_1}\cdot\frac{N_i-n_1}{N_1}+\ldots+\left[\frac{N_K}{N}\right]^2\cdot\frac{S^2_{x_k}}{n_k}\cdot\frac{N_K-n_K}{N_K}$$

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

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

3.8.2.2 Εκτίμηση της αναλογίας του πληθυσμού

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

$$\Sigma\left[\frac{\mathrm{N}_1}{\mathrm{N}}\cdot P_1+\frac{\mathrm{N}_2}{\mathrm{N}}\cdot P_2+\ldots+\frac{\mathrm{N}_K}{\mathrm{N}}\cdot P_K\right]=\Sigma(P)=N$$

Η διακύμανση του σταθμικού αθροίσματος των $\mathrm{P}_1, \mathrm{P}_2, .., \mathrm{P}_k$ υπολογίζεται ως εξής :

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

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

3.8.3 Πλεονεκτήματα-Μειονεκτήματα στρωματοποιημένης δειγματοληψίας

3.8.3.1 Πλεονεκτήματα

1. Εξασφαλίζει μεγαλύτερη ακρίβεια, αφού τα δεδομένα σε κάθε στρώμα είναι περισσότερο ομοιογενή (σε σχέση με την απλή τυχαία δειγματοληψία) απ' ότι σε ολόκληρο τον πληθυσμό. Αυτό σημαίνει μικρότερη διασπορά στην εκτίμηση των παραμέτρων.

2. Κάθε στρώμα έχει μικρότερο μέγεθος σε σχέση με τον πληθυσμό, με αποτέλεσμα να είναι πιο απλό να επιλέγει αντιπροσωπευτικό δείγμα και να γίνει η συλλογή πληροφοριών.

3. Εξασφαλίζει την αύξηση της ακρίβειας του σχεδίου, διατηρώντας σταθερό το μέγεθος του δείγματος που οδηγεί στην μείωση της τιμής δειγματοληπτικού σφάλματος. Σκοπός εδώ είναι η βελτίωση της αντιπροσωπευτικότητας του δείγματος, με αποτέλεσμα το κόστος της δειγματοληψίας να μειωθεί σημαντικά.

4. Η στρωματοποίηση διευκολύνει την αντιμετώπιση κάποιων διαφορών που μπορεί να υπάρχουν μεταξύ των πληθυσμιακών ομάδων π.χ υπάρχει διαφορά στο σύνολο των οικονομικών δραστηριοτήτων μεταξύ των μεγάλων και μικρών επιχειρήσεων.

3.8.3.2 Μειονεκτήματα

1. Η διαδικασία που ακολουθείται για την επιλογή ενός δείγματος είναι πιο περίπλοκη.

2. Απαιτείται περισσότερη προκαταρκτική έρευνα για το σύνολο του πληθυσμού που πρόκειται να μελετηθεί.

3. Το κόστος δειγματοληπτικού σφάλματος είναι υψηλότερο, εάν το δείγμα δεν παραμείνει σταθερό.

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

3.9 ΚΑΤΑ ΟΜΑΔΕΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

3.9.1 Εισαγωγή

Η δειγματοληψία κατά ομάδες είναι μια απλή τυχαία δειγματοληψία στην οποία κάθε δειγματοληπτική μονάδα είναι μια ομάδα στοιχείων του ερευνώμενου πληθυσμού. Για παράδειγμα αν θέλουμε να εκτιμήσουμε το μέσο εισόδημα των νοικοκυριών μιας πόλης, μπορούμε να ομαδοποιήσουμε όλα τα νοικοκυριά κατά οικοδομικό τετράγωνο. Έτσι το σύνολο των νοικοκυριών ισοδυναμεί με το σύνολο των οικοδομικών τετραγώνων, οπότε παίρνουμε τυχαία έναν αριθμό οικοδομικών τετραγώνων και ερευνούμε το σύνολο των νοικοκυριών που βρίσκονται στα οικοδομικά τετράγωνα του δείγματος.

Με τον τρόπο αυτό συντομεύεται ο χρόνος διεξαγωγής της έρευνας αλλά κυρίως με τη δειγματοληψία κατά ομάδες μειώνεται το κόστος. Συνήθως οι ομάδες αποτελούνται από μονάδες του πληθυσμού που είναι γεωγραφικά εντοπισμένες με αποτέλεσμα να απαιτούνται λιγότερες μετακινήσεις και ερευνητές για τη διερεύνησή τους.

Με την κατά ομάδες δειγματοληψία λύνουμε πιο εύκολα το πρόβλημα της κατασκευής πλαισίου. Δηλαδή μπορεί να εφαρμοστεί όταν δεν υπάρχουν τα απαιτούμενα δειγματοληπτικά πλαίσια (π.χ κατάλογοι νοικοκυριών μιας πόλης), ενώ αντίθετα υπάρχουν πλαίσια για ομάδες στοιχείων (π.χ οικοδομικά τετράγωνα μιας πόλης).

Ομάδες στοιχείων, μπορεί να αποτελέσουν τα οικοδομικά τετράγωνα, όταν ως μονάδα του ερευνώμενου πληθυσμού είναι το νοικοκυριό, τα σχολεία, η κατοικία, το άτομο κ.ο.κ.

Τα στοιχεία μέσα σε μια ομάδα πρέπει ιδανικά να είναι όσο το δυνατόν πιο ετερογενή. Αυτό έχει σαν αποτέλεσμα τη μείωση της διακύμανσης και του τυπικού σφάλματος της μέσης τιμής. Στην πράξη όμως, μεταξύ των στοιχείων των ομάδων παρουσιάζεται ομοιογένεια ως προς τα χαρακτηριστικά που ερευνώνται. Επίσης, οι ομάδες θα πρέπει να είναι κατά το δυνατόν μικρού μεγέθους. Είναι προτιμότερο δηλαδή να έχουμε, για το ίδιο μέγεθος δείγματος, πολλές μικρές ομάδες, αντί για λίγες και μεγάλες. Τέλος, οι ομάδες θα πρέπει να είναι όσο το δυνατόν πιο ισομεγέθεις. Σε αντίθετη περίπτωση, ο εκτιμητής της μέσης τιμής είναι μεροληπτικός.

Είναι σημαντικό να μη γίνεται σύγχυση της διαίρεσης του πληθυσμού σε στρώματα με τη διαίρεσή του σε ομάδες. Ένα στρώμα αποτελεί έναν υποπληθυσμό από τον οποίο θα πάρουμε με τυχαίο τρόπο ορισμένες μονάδες. Επιπλέον πρέπει να χρησιμοποιήσουμε όλα τα στρώματα στα οποία διαιρέσαμε τον πληθυσμό. Στην κατά ομάδες δειγματοληψία θα κληρώσουμε ορισμένες από τις ομάδες και θα διερευνήσουμε όλες τις μονάδες τους. Επίσης το στρώμα προϋποθέτει ομοιογένεια των μονάδων που περιλαμβάνονται σ΄αυτό, ενώ αντίθετα στη δειγματοληψία κατά ομάδες, πρέπει να υπάρχει μέσα σε κάθε ομάδα η ανομοιογένεια που υπάρχει στον ερευνώμενο πληθυσμό, ώστε να προκύπτουν αξιόπιστα αποτελέσματα.

Τέλος, στην κατά ομάδες δειγματοληψία σκοπός είναι να πετύχουμε την μεγαλύτερη ακρίβεια σε σχέση με το κόστος ανά στοιχειώδη μονάδα δείγματος.

3.9.2 Εκτίμηση του Μέσου Πληθυσμού

Στην ουσία η δειγματοληψία κατά ομάδες είναι μια απλή τυχαία δειγματοληψία στην οποία κάθε δειγματοληπτική μονάδα αποτελείται από μια ομάδα στοιχείων του πληθυσμού, έτσι, η εκτίμηση του μέσου πληθυσμού γίνεται με τον ίδιο τρόπο που γίνεται και στην απλή τυχαία δειγματοληψία. Επομένως μια αμερόληπτη εκτίμηση του μέσου μ του πληθυσμού είναι ο μέσος του δείγματος και συμβολίζεται με $\bar{x}$

Δηλαδή έχουμε :

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

Η διακύμανση της εκτίμησης του $\bar{x}$ είναι:

$$V(\bar{x}) = \frac{1-f}{n\overline{M}^2}\sigma^2 = \frac{1-n/N}{n\overline{M}^2}\sigma^2 = \frac{\frac{N-n}{N}}{n\overline{M}^2}\sigma^2 \Rightarrow$$

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

Όπου

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu m_i)^2}{N-1},$$

σε περίπτωση που το σ^2 είναι άγνωστο τότε χρησιμοποιείται η εκτιμήτρια που είναι:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x} m_i)^2}{n-1}$$

Η εκτίμηση του τυπικού σφάλματος είναι:

$$SE(\bar{x}) = \sqrt{V(\bar{x})}$$

3.9.3 Εκτίμηση Ποσοστού ενός Πληθυσμού

Η εκτίμηση ποσοστού ενός πληθυσμού με τη μέθοδο της δειγματοληψίας κατά ομάδες, γίνεται όπως και στην απλή τυχαία δειγματοληψία από το ποσοστό των στοιχείων του δείγματος που ανήκουν σε μια συγκεκριμένη κατηγορία, η οποία μας ενδιαφέρει να μελετήσουμε, δηλαδή έχουμε:

$$p = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} m_i}$$

Στην συνέχεια η διακύμανση της εκτίμησης του pείναι:

$$V(p) = \frac{(N-n)}{Nn\bar{M}^2} s^2$$

Όπου

$$s^2 = \frac{\sum_{i=1}^{n}(a_i - p m_i)^2}{n-1}$$

Η εκτίμηση του τυπικού σφάλματος είναι:

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

3.9.4 Πλεονεκτήματα- Μειονεκτήματα Δειγματοληψίας κατά Ομάδες

Η δειγματοληψία κατά ομάδες είναι μια αποτελεσματική μέθοδος η οποία συνήθως εφαρμόζεται όταν το πλήθος των στοιχείων του πληθυσμού είναι γεωγραφικά εντοπισμένο μεν, αλλά άγνωστο και άρα οι άλλες μέθοδοι που περιγράψαμε παραπάνω δεν μπορούν να χρησιμοποιηθούν. Επίσης δεν είναι απαραίτητο να γνωρίζουμε το πλήθος των στοιχείων των ομάδων, παρά μόνο αυτών που θα επιλεγούν τελικά στο δείγμα. Μπορεί επίσης να εφαρμοστεί όταν δεν υπάρχουν τα απαιτούμενα δειγματοληπτικά πλαίσια για ομάδες στοιχείων. Με αυτόν τον τρόπο, ο χρόνος διεξαγωγής της έρευνας περιορίζεται σημαντικά αλλά κυρίως μειώνεται το κόστος σε σχέση με τις άλλες μεθόδους.

Το μειονέκτημα που παρουσιάζει η κατά ομάδες δειγματοληψία είναι ότι η ομαδοποίηση του πληθυσμού έχει σαν αποτέλεσμα οι μονάδες κάθε ομάδας να παρουσιάζουν μεγάλη ομοιογένεια, με αποτέλεσμα να αυξάνεται η διασπορά των εκτιμήσεων και έτσι το δείγμα μας να είναι λιγότερο αντιπροσωπευτικό του πληθυσμού απ’όσο θα ήταν να εφαρμόζαμε οποιαδήποτε άλλη μέθοδο με το ίδιο μέγεθος δείγματος.

Για ένα δεδομένο όμως προϋπολογισμό κόστους, με την κατά ομάδες δειγματοληψία έχουμε τη δυνατότητα να μελετήσουμε περισσότερες μονάδες του πληθυσμού. Από αυτή την άποψη, αν η μέθοδος χρησιμοποιηθεί σωστά, μπορούμε να έχουμε καλύτερα αποτελέσματα, από ότι αν εφαρμόζαμε, για παράδειγμα απλή τυχαία δειγματοληψία με αυτό το κόστος.

3.9.5 Παράδειγμα Δειγματοληψίας κατά Ομάδες

Έστω ότι θέλουμε να εκτιμήσουμε τη μέση μηνιαία κατανάλωση νερού σε κυβικά μέτρα (m^3) των νοικοκυριών μιας πόλης, όπου διαμένουν 3.600 νοικοκυριά. Συνήθως είναι αδύνατο να βρεθεί ένας κατάλογος με καταχωρημένα και αριθμημένα όλα τα νοικοκυριά. Σε αυτή την περίπτωση, εφαρμόζουμε δειγματοληψία κατά ομάδες. Έτσι η συγκεκριμένη πόλη διαιρείται σε οικοδομικά τετράγωνα και καθένα από αυτά αποτελεί μια ομάδα νοικοκυριών. Με αυτόν τον τρόπο, ο χρόνος διεξαγωγής της έρευνας περιορίζεται σημαντικά αλλά κυρίως μειώνεται το κόστος σε σχέση με τις άλλες μεθόδους.

Το σύνολο των οικοδομικών τετραγώνων της πόλης είναι 200. Από το σύνολο των οικοδομικών τετραγώνων επελέγησαν 10 με τυχαίο τρόπο. Από την έρευνα προέκυψαν τα αποτελέσματα που παρουσιάζονται στον παρακάτω πίνακα.

Ομάδα I (οικοδομικό τετράγωνο)	Αριθμός νοικοκυριών n_i	Συνολική κατανάλωση νερού (σε m^3) κατά ομάδα x_i
1	10	1.200
2	15	1.900
3	20	2.500
4	12	1.300
5	25	3.000
6	18	2.400
7	16	1.980
8	21	2.700
9	24	2.900
10	17	2.350
Σύνολο	178	22.230

Η εκτίμηση της μέσης κατανάλωσης νερού των νοικοκυριών της πόλης είναι:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} n_i} = \frac{22.230}{178} = 124,88 m^3$$

Η διακύμανση του $\bar{x}$ είναι:

$$V(\bar{x}) = \frac{N-n}{Nn\bar{M}^2}\sigma^2$$

Επειδή M είναι γνωστό έχουμε

$$\overline{M} = \frac{N}{M} = \frac{3.600}{200} = 18$$

Καταρτίζουμε τον παρακάτω πίνακα υπολογισμών:

Ομάδα ι	x_i	$\overline{x}n_i$	$(x_i - \overline{x}n_i)^2$
1	1.200	1249	2.401
2	1.900	1873	729
3	2.500	2498	4
4	1.300	1498	39.204
5	3.000	3122	14.884
6	2.400	2248	23.104
7	1.980	1998	324
8	2.700	2622	6.084
9	2.900	2997	9.409
10	2.350	2123	51.529
Σύνολο	22.230		147.672

Οπότε

$$V(\overline{x}) = \frac{N-n}{Nn\overline{M}^2} \frac{\sum_{i=1}^{n}(x_i - \overline{x}n_i)^2}{n-1} = \frac{200-10}{200 \cdot 10 \cdot 18^2} \frac{147.672}{10-1} = 4,8$$

Το τυπικό σφάλμα είναι:

$$SE(\overline{x}) = \sqrt{V(\overline{x})} = \sqrt{4,8} = 2,2$$

Ο συντελεστής μεταβλητότητας είναι:

$$CV(\bar{x}) = \frac{SE(\bar{x})}{\bar{x}} = \frac{2,2}{124,88} = 0.0188$$ ή 18%

Το διάστημα εμπιστοσύνης στο οποίο βρίσκεται η μέση μηνιαία κατανάλωση νερού των νοικοκυριών της πόλης με πιθανότητα 95% είναι:

$$\bar{x} \pm Z_{a/2} \cdot SE(\bar{x})$$

Ή

$$124,88 \pm 1,96 \cdot 2,2$$

Ή

$$124,88 \pm 4,3$$

Δηλαδή το μ βρίσκεται στο διάστημα από 120.58 έως 129,18 με πιθανότητα 95% και με την προϋπόθεση ότι η εκτίμηση $\bar{x}$ ακολουθεί την κανονική κατανομή.

3.10 Δειγματοληψία ποσοστών

3.10.1 Εισαγωγή

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

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

Παράδειγμα:

Έστω ότι θέλουμε να εξετάσουμε το ποσοστό των Ελλήνων που είναι ικανοποιημένοι από την είσοδο μας στη ΕΟΚ.
Για το σκοπό αυτό σχηματίζουμε τα εξής στρώματα που είναι απαραίτητο να έχουν τα εξής χαρακτηριστικά για τη μέγιστη δυνατή αντιπροσωπευτικότητα του δείγματος:

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

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

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

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

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

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

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

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

3.10.2 Πλεονεκτήματα-Μειονεκτήματα δειγματοληψίας ποσοστών

3.10.2.1 Πλεονεκτήματα

1. Είναι σχετικά μη δαπανηρή. Αυτό οφείλεται στο γεγονός ότι σε περίπτωση εμφάνισης τυχόν προβλήματος ο συνεντευκτής έχει τη δυνατότητα σε μικρό χρονικό διάστημα να αντικαταστήσει τον **Χ** ερωτώμενο με κάποιον **Υ** ερωτώμενο. Επιπλέον, για να είναι το κόστος σε μειωμένο επίπεδο, οι συνεντευκτές φροντίζουν να εντάσσουν το δείγμα τους σ' εκείνα τα πρόσωπα με τα επιθυμητά χαρακτηριστικά, ώστε να μην χρειάζονται μετακινήσεις.

2. Είναι δυνατόν να διεξαχθεί ακόμα και σε περιπτώσεις που η τυχαία λήψη του δείγματος δεν μπορεί να γίνει επειδή το αναγκαίο δειγματοληπτικό πλαίσιο είναι ανύπαρκτο.

3. Δίνει τη δυνατότητα αξιολόγησης των πληροφοριών που περιέχονται στο δείγμα με στατιστικές μεθόδους.

4. Δίνει τη δυνατότητα υπολογισμού του σφάλματος που είναι συνδεδεμένο με τη χρησιμοποιούμενη εκτιμήτρια.

5. Τα διοικητικά προβλήματα σ' αυτήν την περίπτωση είναι λιγότερα. Αυτό οφείλεται στην έλλειψη των προβλημάτων που προκύπτουν από τις αρνήσεις για συνεργασία και από την επανάληψη των συνεντεύξεων. Συνήθως τα πρόσωπα που παρέχουν αναγκαίες πληροφορίες στον συνεντευκτή κατά κανόνα, παραμένουν άγνωστα και έτσι δεν είναι δυνατόν να γίνει κλασικός έλεγχος του αν έγινε και πως έγινε η συνέντευξη (μπορεί βεβαίως να γίνει μαγνητοφώνηση).

6. Η μέθοδος αυτή απαιτεί λιγότερο χρόνο για να ολοκληρωθεί, γι' αυτό είναι κατάλληλη για έρευνες που πρέπει να διεξαχθούν σε μικρό χρονικό διάστημα. Τέτοιες έρευνες είναι όσες αναφέρονται σε ζητήματα που δεν είναι δυνατόν να κρατηθούν στη μνήμη για μακρύ χρονικό διάστημα ή που έχουν επείγοντα χαρακτήρα (π.χ σφυγμομέτρηση της κοινής γνώμης για διαμόρφωση στάσης που θα ακολουθήσει μια παράταξη για ένα πολιτικό θέμα).

3.10.2.2 Μειονεκτήματα

1. Ο υπολογισμός των τυπικών σφαλμάτων είναι δύσκολος, λόγω έλλειψης τυχαιότητας κατά την επιλογή των μονάδων του δείγματος. Δηλαδή τα αποτελέσματα ενός μη τυχαίου δείγματος ισχύουν μόνο για το δείγμα από το οποίο υπολογίσθηκαν και δεν είναι δυνατόν να ενταχθούν με στατιστικό τρόπο στον αντίστοιχο πληθυσμό. Είναι δυνατόν όμως να υπολογισθούν, εάν το συνολικό δείγμα διαιρείται σε μικρότερα, άλλα ανεξάρτητα δείγματα που αντλούνται από τον ίδιο πληθυσμό. Κατόπιν, τα μικρότερα αυτά δείγματα προσεγγίζονται ως δείγματα ποσοστών από τους συνεντευκτές και παρέχουν τόσες ανεξάρτητες εκτιμήσεις για το συγκεκριμένο χαρακτηριστικό όσα είναι τα μικρότερα ανεξάρτητα δείγματα. Η μεταβλητικότητα των εκτιμήσεων αυτών χρησιμοποιείται για να υπολογισθεί η μεταβλητικότητα της δειγματοληψίας ποσοστών.

2. Η προσωπική επιλογή μπορεί να είναι σοβαρά μεροληπτική.

3. Η δειγματοληψία ποσοστών δεν μπορεί να απαλλαγεί από τη μεροληψία επιλογής των μονάδων του δείγματος που ίσως δημιουργούν οι συνεντευκτές.

4. Δεν μπορούμε να ελέγξουμε σοβαρά τον συνεντευκτή κατά το χρόνο διεξαγωγής της έρευνας των μονάδων του δείγματος και συλλογής πληροφοριών.

4. ΕΜΠΕΙΡΙΚΕΣ ΚΑΤΑΝΟΜΕΣΣΥΧΝΟΤΗΤΩΝ

4.1 Ασυνεχείς και Συνεχείς κατανομές συχνοτήτων

Μετά τησυγκέντρωση των στατιστικών στοιχείων που αναφέρονται σε μια μόνο ιδιότηταενός πληθυσμού, το πρώτο βασικό βήμα είναι η κατάλληλη κατάταξη και η συστηματικη ομαδοποίηση των τιμών της μεταβλητής, με αντικειμενικό σκοπό την παρουσίαση των στοιχείων κατά τρόπο που ναδίνει πληροφορίες για τη δομή του πληθυσμού που ερευνάται και ναδιευκολύνει το έργο της στατιστικής ανάλυσηςκαι εξαγωγής χρήσιμων συμπερασμάτων. Για την ομαδοποίηση των τιμών της μεταβλητήςχρησιμοποιούνταιειδικές κατατάξεις που τις ονομάζουμε *κατανομές συχνοτήτων* ή *πίνακες συχνοτήτων*. 0 τρόπος κατασκευής των κατανομών συχνοτήτων εξαρτάται από το είδος της μεταβλητής.

1. Όταν η μεταβλητή είναι ασυνεχής

Αν η μεταβλητή είναι ασυνεχής και το εύρος των τιμών της μεταβλητής είναι μικρό, τότε αθροίζουμε τις παρατηρήσεις (συχνοτήτων) που αναφέρονται στην ίδια τιμή της μεταβλητής και παρουσιάζουμε τα δεδομένα σε έναν πίνακα απλήςεισόδου**,** όπου στην πρώτη στήλητοποθετούμε τις τιμές της μεταβλητής και στη δεύτερη στήλη τον αριθμό των παρατηρήσεων που αντιστοιχεί σε κάθε τιμή της μεταβλητής. Στην περίπτωση αυτή της ασυνεχούς μεταβλητής, ο πίνακας συχνοτήτωνπαίρνει τη μορφή του πίνακα Α. Στον πίνακα αυτό:

(1) $x_1 \ldots, x_2 \ldots, x_i \ldots, x_\nu$ είναι οι τιμές της ασυνεχούς μεταβλητής,τις οποίες έχουμε τοποθετήσεικατα φυσική αύξουσα τάξη μεγέθους, από τη μικρότερη ως τη μεγαλύτερη.

(2) $f_1, f_2 ..., f_i ..., f_\nu$ είναι οι αντίστοιχες συχνότητες της μεταβλητής $\chi,$ δηλαδή πόσες

φορές εμφανίζεται, στο συνολικό πληθυσμό, κάθε τιμή τηςμεταβλητής χ

Πίνακας Α

Τιμές μεταβλητής (xi)	Αριθμός παρατηρήσεων ή συχνότητες (f,)
Xι	f_1
X2	f_2
X3	f_3
X4	f_4
.	.
.	.
.	.
Xi	f_i
.	.
.	.
.	.
Xν	fv
Σύνολο	$\sum f_i$

Όταν ητιμη x_i της μεταβλητής X εμφανίζεται f_i φορές, λέμε ότι το f_i είναι η απόλυτη συχνότητα της τιμής x_i ενώ η τιμή του πηλίκου $f_i / \sum f_i$ ονομάζεται σχετική συχνότητα της x_i .
Ο συμβολισμός $\sum f_i$ σημαίνειτο αθροισμα ολων των f_i δηλαδη:

$$\sum_{i=1}^{\nu} f_i = f_1 + f_2 + f_3 + \ldots + f_\nu$$

Για την καλύτερη κατανόηση του τρόπου παρουσίασης των αριθμητικών δεδομένων ενός πληθυσμού με μορφή ασυνεχούς κατανομής συχνοτήτων, δίνουμε το παρακάτω παράδειγμα. Καθένας από τους 30 υπαλλήλους της επιχείρησης Α έλαβε για θερινή άδεια το 1977 τον παρακάτω αριθμό ημερών:

25,26,24,22,21,25,26,26,24,22,21,26,26,25,27
24,24,27,22,23,23,26,25,25,24,25,25,27,24,25

Ζητείτα ο πίνακας κατανομής των πιο πάνω παρατηρήσεων κατά συχνότητες.

Λύση:

Τοποθετούμε πρώτα τις παρατηρήσεις κατά φυσικά αύξουσα τάξη μεγέθους:

21,21,22,22,22,23,23,24,24,24,24,24,24,25,25

25,25,25,25,25,25,26,26,26,26,26,26,27,27,27

Μετά υπολογίζουμε το εύρος της μεταβολής των παρατηρήσεων, δηλαδή τη διαφορά ανάμεσα στη μεγαλύτερη και μικρότερη τιμή των παρατηρήσεων.Επειδή η διαφορά αυτή είναι πολύ μικρή, R=27-21=6, η παρουσίαση γίνεται με μορφή ασυνεχούς κατανομής, όπως δείχνει ο παρακάτω πίνακας.

Πίνακας

Τιμές μεταβλητης (Xi)	Συχνότητες (f_i)
21	2
22	3
23	2
24	6
25	8
26	6
27	3
Σύνολο	30

2. Όταν η μεταβλητή είναι συνεχής

Στην περίπτωση των συνεχών μεταβλητών που η μεταβλητή x μπορεί να πάρει οποιαδήποτε τιμή σε ένα διάστημα μεταξύ δύο πραγματικών αριθμών [α, β], καταφεύγουμε στην ομαδοποίηση των παρατηρήσεων σε μορφή κατανομής συχνοτήτων.Για το σκοπό αυτό, διαιρείται όλη η έκταση της μεταβολής των τιμών των στατιστικών δεδομένων σε έναν ορισμένο αριθμό διαδοχικών διαστημάτων (ίσων ή και άνισων τάξεων) και τοποθετούμε σε κάθε τάξη (διάστημα) το πλήθος των παρατηρήσεων (f_i) που περιέχονται σε αυτή. Αυτό φαίνεται και καλύτερα στο πιο κάτω σχήμα:

$\underline{f_1\ f_2}$ ___...___ $\underline{f_i}$ ___...___ $\underline{f_n}$

$a_0\ a_1\ a_2 ... a_{i-1}\ a_i ... a_{n-1}\ a_n$

Κάθε τάξη διακρίνεται από τα όριά της, δηλαδή από το κατώτερο όριο α,-ι και απο το ανώτερο όριο (a_i).

Σε κάθε τάξη (a_{i-1}, a_i) αντιστοιχούμε την ονομαζόμενη κεντρική τιμή x_i αυτής, γύρω από την οποία θεωρούμε ότι συγκεντρώνονται οι παρατηρήσεις κάθε τάξης. Η κεντρική τιμή κάθε τάξης υπολογίζεται από τη σχέση:

$$x_i = \frac{a_i + a_{i-1}}{2}$$

Οι κεντρικές τιμές των τάξεων μιας συνεχους μεταβλητης έχουν τηνίδια έννοια και χρήση που έχουν οι ασυνεχείς τιμές μιας μεταβλητής.

Μια μεγάλη δυσκολία που παρουσιάζεται στην ομαδοποίηση των παρατηρήσεων μιας συνεχούς μεταβλητής σε μορφή κατανομής συχνοτήτων,είναι ο προσδιορισμός του αριθμού των τάξεων. Γενικός κανόνας για τον αριθμό των τάξεων κάθε κατανομής δεν υπάρχει, μπορούμε όμως να χρησιμοποιήσουμε έναν εμπειρικό τύπο που δίνει κατά προσέγγιση τον αριθμό των τάξεων αυτός είναι:

K= 1 + 3,322 $\log_{10}$ N (κανόνας του Sturges)

όπου *k* ο αριθμός των τάξεων και N ο αριθμός των παρατηρήσεων.

Στην πράξη όμως ο αριθμός των τάξεων προσδιορίζεται με βαση το κριτήριο της ομοιογένειαςκαι το κριτήριο της απλότητας.

Το πρώτο κριτήριο χρειάζεται τη διαίρεση του συνολικού εύρους της μεταβολής σε πολλές τάξεις μικρού πλάτους.

Αντίθετα, με κριτήριο την απλότητα της εμφάνισης των δεδομένων, διαιρούμε το συνολικό εύρος σε λίγες μόνο τάξεις ίσου ή άνισου πλάτους.

Πολλοί συγγραφείς δέχονται ότι ο αριθμός των τάξεων για την ομαδοποίηση των δεδομένων δεν πρέπει, ανάλογα με το σκοπο που επιδιώκουμε, να είναι μικρότερος από 5 ή μεγαλύτερος από 20.

Η διαφορά μεταξύ κατώτερου και ανώτερου ορίου κάθε τάξης ονομάζεται *πλάτος* αυτής και συμβολίζεται με δ. Στις περισσότερες περιπτώσεις το πλάτος *(δ)* όλων των τάξεων της κατανομής είναι το ίδιο.

Οι κατανομές συχνοτήτων ίσου πλάτους εφαρμόζονται συνήθως στις περιπτώσεις εκείνες που το εύρος μεταβολής της μεταβλητής είναι μικρό.

Σε ορισμένες όμως περιπτώσεις, είναι αναγκαίο το πλάτος των τάξεων να είναι άνισο. Αυτό συμβαίνει όταν το συνολικό εύρος είναι αρκετά μεγάλο ή άπειρο.

Συνηθισμένη περίπτωση στην πράξη είναι οι κατανομές εισοδημάτων ή δαπανών.

Γενικά, είτε πρόκειται για κατανομή ίσου πλάτους, είτε άνισου πλάτους, ο πίνακας συχνοτήτων στην περίπτωση μιας συνεχούς μεταβλητής έχει τη μορφή του παρακάτω πίνακα.

Τάξεις	Συχνότητες (fj)	Κεντρικες τιμές των τάξεων (Xι)
$a_0 - a_1$	f_1	Xι
$a_1 - a_2$	f_2	X2
$a_2 - a_3$ . . .	f_3 . . .	X3 . . .
$a_{i-1} - a_i$ . . . $a_{n-1} - a_n$	f_i . . . f_n	Xi . . . x_n
Σύνολο	Σf_i	

Η κατανομή του παραπάνω πίνακα ονομάζεται *κλειστή,* γιατίδενλείπειούτε το κατώτερο όριο της πρώτης τάξης ούτε το ανώτερο της τελευταίας τάξης. Αν η κατανομή συχνοτήτωνστερείται είτε του κατώτερου ορίου της πρώτης τάξης, είτε του ανώτερου ορίου της τελευταίας τάξης ή και των δύο μαζί, ονομάζεται *ανοικτή κατανομή*. Οι ανοικτές κατανομές παρουσιάζουν ορισμένα προβλήματα κατά τον υπολογισμό των στατιστικών παραμέτρων, όπως θα δούμε αργότερα.

Για την κατανόηση του τρόπου κατασκευής μιας κατανομής κατά συχνότητες δίνουμε το παρακάτω παράδειγμα Β.

Παράδειγμα Β

Η ωριαία αποζημίωση (σε ευρώ) του καθενός αποτους40 υπάλληλους μιας επιχείρησης έχει ως εξής:

31 34 45 46 42 44 50 53

54 55 56 57 60 61 62 63

64 65 66 67 68 75 74 76

59 85 85 84 86 90 99 88

78 87 92 93 94 95 96 89

Να γίνει ομαδοποίηση των παραπάνω παρατηρήσεων με μορφή κατανομής συχνοτήτων σε τάξεις και να υπολογιστούν οι κεντρικές τιμές των τάξεων.

Λύση

Για να διευκολύνουμε την ταξινόμησητων παρατηρήσεων τις κατατάσσουμε κατά Φυσική αύξουσατάξη μεγέθους, όπως παρακάτω:

31 50 59 65 76 87 94

34 53 60 66 78 88 95

42 54 61 67 84 89 96

44 55 62 68 85 90 99

45 56 63 74 85 92

46 57 64 75 86 93

Στη συνέχεια, βρίσκουμε τη διαφορά ανάμεσα στη μεγαλύτερη *(M = 99)* και τη μικρότερη τιμη *(E = 31),* η οποία ονομάζεται *εύρος της κατανομής,* δηλαδη$R = M - E = 99 - 31 = 68$. Επειδή το εύροςR είναι σχετικά μικρό, επτά π.χ. τάξεις είναι αρκετές για να ικανοποιήσουν το κριτήριο της απλότηταςκαιτης ομοιογένειας, το εύρος κάθε τάξης θα είναι: $\delta = 68 : 10$.

Κατά συνέπεια, θα έχουμε τις 40 παρατηρήσεις που μας δόθηκαν, με μορφή κατανομής κατά συχνότητες, οπως φαινονιαι στον πίνακα.

Τάξεις	Συχνότητες(f_i)	Κεντρικές τιμές των τάξεων (x_i)
30-40	2	35
40-50	4	45
50-60	7	55
60-70	9	65
70-80	4	75
80-90	7	85
90-100	7	95
Σύνολο	40	

Σημείωση:Τα ανώτερα όρια των τάξεων δεν περιέχονται σε αυτές.

4.2 Ιστόγραμμα συχνοτήτων

Το ιστόγραμμα χρησιμοποιείται κατά κανόνα για τη γραφική απεικόνιση των κατανομών συχνοτήτων όταν έχουμε διαστήματα τάξεων ίσου ή άνισου πλάτους.
Το διάγραμμα αυτό είναι μια σειρά εφαπτόμενων ορθογωνίων παραλληλογράμμων που έχει βάση τον οριζόντιο άξονα των ορθογωνίων συντεταγμένων, το δε ύψος των ορθογωνίων είναι ανάλογο με τη συχνότητα την οποία αντιπροσωπεύουν.

Παράδειγμα

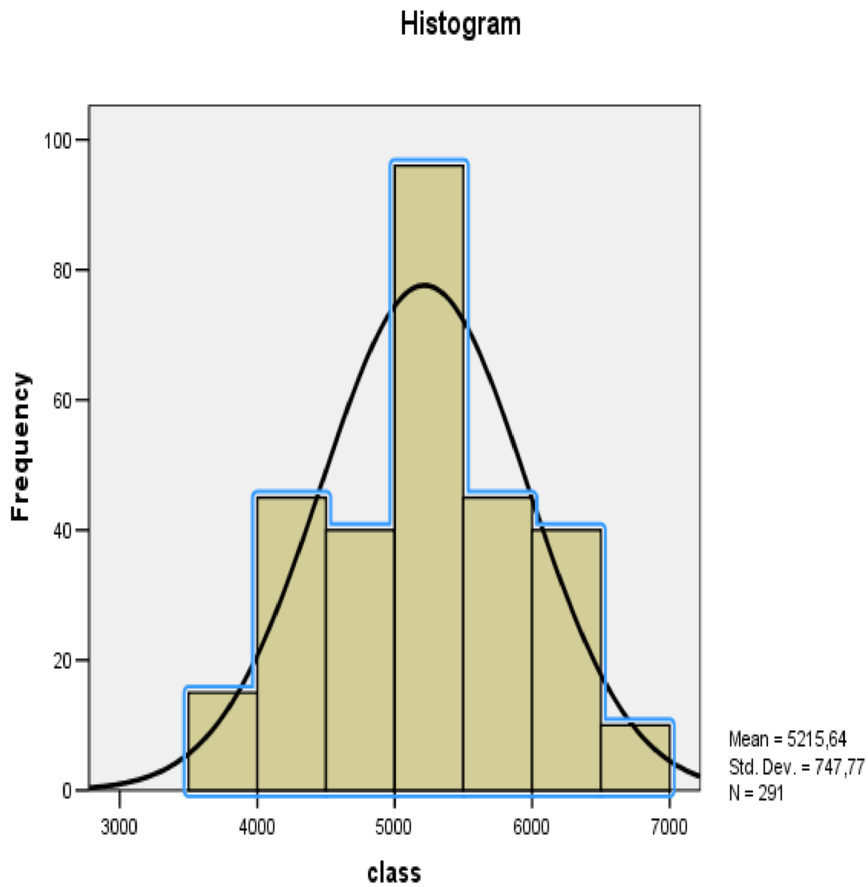
Στην περίπτωση αυτή, παίρνουμε το σύστημα των ορθογωνίων συντεταγμένων και τοποθετούμε στον άξονα των χ (τετμημένων) τα διαστήματα τάξεων της κατανομής, αφού διαιρέσουμε τον άξονα των χ σε ίσα τμημαια, ενώ στον κάθετο άξονα (τεταγμένων) τοποθετούμε τις συχνότητες (f_i).

Μετά, από τα όρια των τάξεων ($a_i - a_{i-1}$) φέρνουμε κατακορύφως ύψη ίσα προς τη συχνότητά τους (f_i) και προκύπτει έτσι ένας αριθμός αλληλοεφαπτόμενων ορθογωνίων, τα οποία μας δίνουν το ιστόγραμμα τάξεων με ίσο πλάτος που θέλουμε να κατασκευάσουμε.

Στον παρακάτω πίνακα δίνεται η κατανομή των ημερήσιων αποδοχών 290 εργαζομένων σε μια επιχείρηση.

Τάξεις αποδοχών	Συχνότητες (f_i)
3500-4000	15
4000-4500	35
4500-5000	50
5000-5500	95
5500-6000	45
6000-6500	40
6500-7000	10
Σύνολο	290

Να παρασταθούν τα δεδομένα του πίνακα αυτού σε μορφή ιστογράμματος.



Με την ένωση των μέσων των άνω βάσεων των ορθογωνίων του ιστογράμματος σχηματίζεται μια τεθλασμένη γραμμή που ονομάζεται πολυγωνική γραμμή ή πολύγωνο συχνοτήτων.

4.3 Αθροιστικές κατανομές συχνοτήτων

Πολλές φορές χρειάζεται να γνωρίζουμε πόσες ή τι ποσοστό των περιπτωσεων μιας μεταβλητής περιλαμβάνεται μέχρι ενός ορισμένου διαστήματος τάξης, ή ενδιαφερόμαστε να γνωρίσουμε το πλήθος των παρατηρήσεων που οι τιμές τους είναι μικρότερες ή ίσες με ορισμένη τιμή χ, της μεταβλητής Χ. Απάντηση στο παραπάνω ερώτημα δίνουν οι καλούμενες αθροιστικές συχνότητες F_i Επιπλέον, η αθροιστική συχνότητα ονομάζεται και *δεξιόστροφη αθροιστική σειρά*.

Αν ο πίνακας συχνοτήτων των τιμών της μεταβλητής X είναι όπως δείχνει ο κάτωθι πίνακας, η αθροιστική συχνότητα της τιμής x_i, θα είναι:

$$F_i = f_1 + f_2 + f_3 + \ldots + f_i$$

Απο τον πίνακα κατανομής συχνοτήτων μπορούμε να κατασκαυασουμε τον πίνακα αθροιστικών συχνοτήτων.

Πίνακας Κατανομής Συχνοτήτων

X_i	X_1	X_2	X_3	…	X_i	…	X_n
f_i	f_1	f_2	f_3	…	f_i	…	f_n
F_i	F_1	F_2	F_3	…	F_i	…	F_n

Για τη γραφική απεικόνιση αθροιστικών συχνοτήτων χρησιμοποιούνται τα καλούμενα *αθροιστικά διαγράμματα*. Η κατασκευή των διαγραμμάτων αυτών δεν παρουσιάζει καμιά δυσκολία, απαιτείται όμως η διάκριση μεταξύ συνεχών και ασυνεχών κατανομών.

Παράδειγμα

Δίνεται η κατανομή της ηλικίας 70 ατόμων, όπως φαίνεται στον πίνακα.

Πίνακας

Τάξεις	Αριθμός ατόμων (f_i)
20-25	4
25-30	5
30-35	10
35-40	28
40-45	12
45-50	6
50-55	3
55-60	2
Σύνολο	70

Ζητείται:

Α)Η αθροιστικη σειρά συχνοτήτων

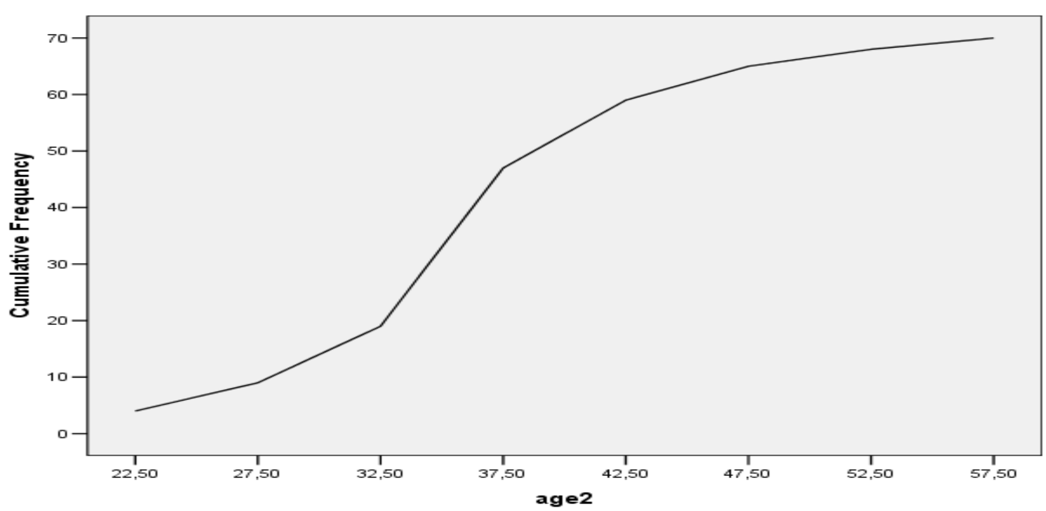
Β)Το αθροιστικό διάγραμμα

Λύση

α) Όταν η κατανομή μας εμφανίζεται σε μορφή τάξεων, για να σχηματισουμε την αθροιστική σειρά εργαζόμαστε όπως φαίνεται στον πίνακα.

Πίνακας

Τάξεις	fi	Περιοχή ηλικιών	Δεξιόστροφη αθροιστική σειρά (F_i)
20 25	4	20-25	4
25-30	5	20-30	9
30-35	10	20-35	19
35-40	28	20-40	47
40-45	12	20-45	59
45 - 50	6	20-50	65
50-55	3	20-55	68
55 - 60	2	20-60	70
Σύνολο	70		



5. ΠΑΡΑΜΕΤΡΟΙ ΜΟΝΟΜΕΤΑΒΛΗΤΩΝ ΠΛΗΘΥΣΜΩΝ

Σκοπός της εμφάνισης των στατιστικών δεδομένων με μορφή συνοπτικών πινάκων συχνοτήτων είναι ο περιορισμός του όγκου των στοιχείων που συγκεντρώθηκαν και η εύκολη μελέτη και περιγραφή της δομής του πληθυσμού που ερευνούμε. Για το λόγο αυτό, θεωρείται πολλές φορές αναγκαία μια παραπέρα συμπύκνωση και συγκεφαλαίωση των πινάκων συχνοτήτων.

5.1 Στατιστικά Περιγραφικά Μέτρα

- Χρησιμοποιούνται για ποσοτικά δεδομένα
- Ένα περιγραφικό μέτρο είναι ένας αριθμός ο οποίος μας δίνει κάποια πληροφορία για τη μορφή και τη δομή των δεδομένων.
- Χρησιμοποιούνται ως βάση για τη στατιστική συμπερασματολογία (εκτιμητική, έλεγχος υποθέσεων).

5.1.1 Μέτρα

- **Μέτρα κεντρικής τάσης** (αριθμητικός, γεωμετρικός, αρμονικός μέσος).
- **Μέτρα θέσης** (διάμεσος, μέση τιμή, επικρατούσα τιμή, ποσοστημόρια).

5.1.1.1 Μέτρα κεντρικής τάσης

I. Μέσος αριθμός.

Συμβολισμοί:

Α. Πληθυσμιακός Μέσος

Έστω ένας πληθυσμός με N μονάδες $x_1, x_2, ..., x_N$

Ο πληθυσμιακός μέσος είναι:

$f(a)=\sum_{i=1}^{n}[(\log x_i)^2+(\log\alpha)^2-2\log x_1\log\alpha]=\sum_{i=1}^{n}(\log x_i)^2+N(\log\alpha)^2-2\sum_{i=1}^{n}\log x_i\log\alpha.$

B. Αν επιλέξουμε ένα δείγμα $X_1,\ldots,X_n$ Μεγέθους *n*,

Ο δειγματικός μέσος θα είναι: $\bar{x}=\dfrac{x_1+x_2+\ldots+x_n}{n}=\dfrac{\sum_{i=1}^{n}X_i}{n}$.

<u>ΙΔΙΟΤΗΤΕΣ</u>

- Αν X_1 =a για όλα τα i= 1,2,…,n, τότε $\bar{X}$ =a.

- Η τιμή του αριθμητικού μέσου βρίσκεται πάντα μεταξύ της ελάχιστης και της μέγιστης τιμής της μεταβλητής $X\min\le\bar{X}\le X\max$.

Πράγματι ισχύει για κάθε i, $X\min\le Xi$, δηλαδή $X\min\le X_1$, $X\min\le X_2,\ldots,$ $X\min\le X_n$.

Προσθέτοντας όλες αυτές τις ανισότητες κατά μέλη παίρνουμε $nx_{\min}\le\sum_{i=1}^{n}X_i$.

Και διαιρώντας δια *n* προκύπτει ότι: $x_{\min}\le\bar{X}$.

Παρόμοια αποδεικνύεται και το $\bar{X}\le x_{\max}$.

- Το άθροισμα των αποκλίσεων των τιμών της μεταβλητής είναι πάντα ίσο με μηδέν, δηλαδή $\sum_{i=1}^{n}(X_i-\bar{X})=0$.

Πράγματι με βάση τις ιδιότητες των αθροισμάτων έχουμε:

$$\sum_{i=1}^{n}(X_i-\bar{X})=\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}\bar{X} \;=\sum_{i=1}^{n}X_i-n\bar{X}=0 \;\; .$$

Από τον ορισμό του αριθμητικού μέσου.

- Το άθροισμα των τετραγώνων των αποκλίσεων των τιμών της μεταβλητής από μια ποσότητα *α*, δηλαδή το $\sum_{i=1}^{n}(X_i-a)^2$, γίνεται ελάχιστο ως προς *α*, όταν *α*= $\bar{X}$.

Για την απόδειξη θεωρώ τη συνάρτηση *f(α)*=$\sum_{i=1}^{n}(X_i-a)^2$ **.**Για να είναι ένα σημείο ακρότατο της *f*, πρέπει να μηδενίζει την παράγωγο της, δηλαδή να ισχύει *f'(α)*=0. Αλλά για την παράγωγο έχω:

$$f'(\alpha)=-2\sum_{i=1}^{n}(X_i-a)=-2\sum_{i=1}^{n}X_i+2na \;\; .$$

Θέτοντας την ποσότητα αυτή ίση με μηδέν, παίρνω α=$\dfrac{\sum_{i=1}^{n}X_i}{n}=\bar{X}$.

Για να αποδείξω ότι πρόκειται για ελάχιστο, βρίσκω το πρόσημο της $2^{ης}$ παραγώγου *f''(α)*= 2n>0, άρα πράγματι η συγκεκριμένη τιμή ελαχιστοποιεί τη συνάρτηση.

- Αν σε όλες τις τιμές ενός συνόλου δεδομένων προσθέσω μια ποσότητα *α*, ο νέος αριθμητικός μέσος αυξάνεται κατά *α*. Δηλαδή, αν οι τιμές $X_1, X_{2,}......, X_N$ έχουν $\bar{\alpha}\;=\bar{X}$, τότε οι τιμές $x_1+a, x_2+a,..., x_n+a$ έχουν μέσο $\bar{X}+a$ **.**Αυτό προκύπτει επειδή:**.**

$$\sum_{i=1}^{n}(x_i + a) = \frac{x_1 + x_2 + \ldots + x_n + na}{n} = \frac{\sum_{i=1}^{n} x_i}{n} + a = \bar{X} + a$$

- Αν πολλαπλασιάσω όλες τις τιμές ενός συνόλου δεδομένων με *α*, ο αριθμητικός μέσος πολλαπλασιάζεται επίσης επί *α*. Για την απόδειξη, έστω ότι τα αρχικά δεδομένα είναι $X_1, \ldots, X_n$, πολλαπλασιάζοντας επί *α*, τα νέα δεδομένα είναι: $ax_1, \ldots, ax_n$ και ο αριθμητικός μέσος για τα δεδομένα αυτά θα

  είναι: $\frac{\sum_{i=1}^{n}(a * X_i)}{n} = \frac{a\sum_{i=1}^{n} X_i}{n} = \bar{X} * a$.

- Αν έχω *k* σύνολα δεδομένων

$x_1, x_2, \ldots, x_{n1}$ με μέσο $\bar{X}_1$

$x_{n1+1}, x_{n2+2}, \ldots, x_{n1+n2}$ με μέσο $\bar{X}_2$

$x_{n1+1+1}, x_{n2+2+2}, \ldots, x_{n1+n2+n3}$ με μέσο $\bar{X}_3$

Τότε για τον αριθμητικό μέσο όλων των παρατηρήσεων έχω:

$$\bar{X} = \frac{x_1 + x_2 + \ldots + x_{n1} + \ldots + x_{n1+n2} + \ldots + x_{n1+n2+n3} + \ldots + x_{n1+n2+\ldots+nk}}{n_1 + n_2 + \ldots + n_k} =$$

$$\bar{X} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + \ldots + n_k\bar{X}_k}{n_1 + n_2 + \ldots + n_k}.$$

Ή για συντομία $\overline{X} = \dfrac{\sum_{i=1}^{n} n_i \overline{X}_i}{\sum_{i=1}^{n} n_i}$.

<u>Έμμεση μέθοδος υπολογισμού του μέσου αριθμητικού</u>

Ο υπολογισμός του μέσου αριθμητικού με την άμεση μέθοδο παρουσιάζει δυσκολία στις πράξεις, όταν οι κεντρικές τιμές των τάξεων και οι συχνότητες (Λ) είναι μεγάλοι αριθμοί. Στις περιπτώσεις αυτές, χρησιμοποιούμε αντί της άμεσης μεθόδου την έμμεση μέθοδο υπολογισμού του μέσου αριθμητικού.Για την εφαρμογή της μεθόδου αυτής, υποθέτουμε ότι μια από τις κεντρικές τιμές των τάξεων παίρνεται ως μέσος αριθμητικός και συμβολίζεται με x_0. Ως x_0 παίρνουμε μια οποιαδήποτε από τις κεντρικές τιμές των τάξεων, συνήθως παίρνουμε εκείνη που αντιστοιχεί στη μεγαλύτερη συχνότητα ή κατέχει την κεντρική θέση. Στη συνέχεια, υπολογίζουμε τις αποκλίσεις κάθε κεντρικής τιμής από τον προσωρινό μέσο αριθμητικό σε μονάδες διαστημάτων τάξεων. Οι αποκλίσεις αυτές συμβολίζονται με ξ_i και δίνονται από την παρακάτω σχέση:

$$\xi_i = \frac{x_i - x_0}{\delta}$$

Όπου δ το πλάτος της τάξης.

Στη συνέχεια, υπολογίζουμε το μέσο αριθμητικό με την έμμεση μέθοδο, χρησιμοποιώντας τον τύπο:

$$\mu = x_0 + \delta \frac{\sum f_i \xi_\iota}{\sum f_i}$$

Παράδειγμα

Οι ηλικίες μιας ομάδας 264 ατόμων παρουσιάζεται στον παρακάτω πίνακα.

Πίνακας

Ηλικία (σε χρόνια)	Αριθμός ατόμων (f_i)
15-25	34
25-35	38
35-45	53
45-55	55
55-65	46
65-75	27
75-85	11
Σύνολο	264

Να υπολογιστεί η μέση ηλικία με την έμμεση μέθοδο.

Λύση:

Για τον υπολογισμό της μέσης ηλικίας με την έμμεση μέθοδο καταρτίζουμε τον παρακάτω πίνακα

Πίνακας

Τάξεις	x_i	f_i	$\xi = \frac{x_i - x_0}{\delta}$	$f_i \xi_i$
15-25	20	34	-3	-102
25-35	30	38	-2	-76
35-45	40	53	-1	-53
45-55	50	55	0	0
55-65	60	46	1	46
65-75	70	27	2	54
75-85	80	11	3	33
Σύνολο		264		-98

Η μέση αριθμητική τιμή θα είναι:

$$\mu = x_0 + \delta \frac{\sum f_i \xi_\iota}{\sum f_i} = 50 + \frac{-98}{264} = 46,3\,\text{έτη}$$

II. Γεωμετρικός μέσος.

Συμβολισμοί:

Α. Πληθυσμιακός Μέσος

Έστω ένας πληθυσμός με *N* μονάδες $x_1, x_2, ..., x_N$.

Ο πληθυσμιακός μέσος είναι:

$G = (x_1 * x_2 * ... * x_N)^{1/N} = (\Pi_{i=1}^{N} x_1)^{1/N}$.

Β. Αν επιλέξουμε ένα δείγμα $x_1, x_2, ..., x_n$.Μεγέθους *n*, ο γεωμετρικός μέσος του δείγματος είναι: μέσος $g = (x_1 * x_2 * ... * x_n)^{1/n} = (\Pi_{i=1}^{n} x_1)^{1/n}$.

ΙΔΙΟΤΗΤΕΣ

- Αν πάρουμε λογαρίθμους στη σχέση $G = (x_1 * x_2 * ... * x_N)^{1/N} = (\Pi_{i=1}^{N} x_1)^{1/N}$, βλέπουμε ότι $\log G = \log(\Pi_{i=1}^{N} x_i)^{1/N} = \frac{1}{N} \log(\Pi_{i=1}^{N} x_i) = \frac{1}{N} \sum_{i=1}^{n} \log(x_i)$, δηλαδή ο λογάριθμος του γεωμετρικού μέσου είναι ο αριθμητικός μέσος των λογαρίθμων των τιμών της μεταβλητής.

- Αν θέσω $u_i = \frac{x_i}{G}$, τότε ισχύει πάντα $\Pi_{i=1}^{N} u_i = u_1 u_2 ... u_N = 1$. Η απόδειξη είναι απλή, εφόσον $\Pi_{i=1}^{N} u_i = \frac{\Pi_{i=1}^{N}}{G^N} = \frac{x_1 * x_2 * ... * x_N}{[(x_1 * x_2 * ... * x_N)^{1/N}]^N} = 1$, χρησιμοποιώντας τον ορισμό του γεωμετρικού μέσου.

- Το άθροισμα τετραγώνων $\sum_{i=1}^{n} [\log(\frac{x_i}{a})^2]$, γίνεται ελάχιστο ως προς *α*, όταν *α*=*G*.

Για την απόδειξη ορίζω τη συνάρτηση $f(\alpha) = \sum_{i=1}^{n} [\log(\frac{x_i}{a})^2] = \sum_{i=1}^{n} [(\log x_i - \log \alpha)^2]$.

Παίρνοντας το ανάπτυγμα

έχω: $f(\alpha) = \sum_{i=1}^{n} [\log(x_i)^2 + (\log \alpha)^2 - 2\log x_i \log \alpha] = \sum_{i=1}^{n} (\log x_i)^2 + N(\log \alpha)^2 - 2\sum_{i=1}^{n} \log x_i \log \alpha$.

Η παράγωγος της συνάρτησης είναι: $f'(\alpha) = \frac{2N\log \alpha}{a} - \frac{\sum_{i=1}^{n} \log x_i}{a}$.

Η παράγωγος μηδενίζεται όταν ισχύει: $N \log \alpha = \sum_{i=1}^{n} \log x_i$, δηλαδή $\log \alpha = \frac{\sum_{i=1}^{n} \log x_i}{N}$ το οποίο σύμφωνα με την ιδιότητα *1*, συμβαίνει όταν το *α* συμπίπτει με το γεωμετρικό μέσο, δηλαδή όταν *α*=*G*.

(a) Οι παραπάνω ιδιότητες αναφέρονται στον πληθυσμό επειδή χρησιμοποιούν το *G*.

(b) Οι ίδιες ισχύουν ακόμα και για το δειγματικό μέσο *g*.

III. Αρμονικός μέσος.

Ο αρμονικός μέσος H για τον πληθυσμό ορίζεται ως εξής: $\dfrac{1}{H} = \dfrac{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_N}}{N} = \dfrac{\sum_{i=1}^{n} \frac{1}{x_i}}{N}$ ή

Ισοδύναμα, $H = \dfrac{N}{\sum_{i=1}^{n} \frac{1}{x_i}}$, δηλαδή ο αντίστροφος του αρμονικού μέσου είναι ο αριθμητικός μέσος των αντίστροφων των τιμών της μεταβλητής.

Αντίστοιχα για ένα δείγμα μεγέθους n ο αρμονικός μέσος είναι: $\bar{h} = \dfrac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$,ο αντίστροφος του αριθμητικού μέσου των αντίστροφων των τιμών στο δείγμα.

Ιδιότητες του αρμονικού μέσου.

1. Αν θέσω $u_i = \dfrac{1}{x_i} - \dfrac{1}{H}$, τότε ισχύει πάντα $\sum_{i=1}^{N} u_i = 0$.

Αυτό προκύπτει εύκολα, επειδή $\sum_{i=1}^{N} u_i = \sum_{i=1}^{N} (\frac{1}{x_i} - \frac{1}{H}) = \sum_{i=1}^{N} \frac{1}{x_i}) - \frac{N}{H}$, με βάση τις ιδιότητες των αθροισμάτων. Το παραπάνω ισούται με μηδέν, αφού από τον ορισμό του H έχουμε:

$$\frac{1}{H} = \frac{\sum_{i=1}^{N} \frac{1}{x_i}}{N} \Rightarrow \frac{N}{H} = \sum_{i=1}^{N} \frac{1}{x_i}.$$

2. Το άθροισμα τετραγώνων $\sum_{i=1}^{N} (\frac{1}{x_i} - \frac{1}{\alpha})^2$. Γίνεται ελάχιστο ως προς α όταν $\alpha = H$.

- Η απόδειξη είναι παρόμοια με αντίστοιχες για τον αριθμητικό και το γεωμετρικό μέσο.

IV. Σταθμικός μέσος αριθμητικός.

Όταν τα στατιστικά δεδομένα δίνονται σε μορφή κατανομής συχνοτήτων, διακρίνουμε δύο περιπτώσεις υπολογισμού του μέσου αριθμητικού.

a. Όταν η μεταβλητή είναι ασυνεχής.

Στην περίπτωση κατά την οποία η μεταβλητή είναι ασυνεχής και κάθε τιμή της εμφανίζεται πολλές φορές, δηλαδή παρουσιάζει συχνότητα, για τον υπολογισμό του μέσου αριθμητικού πολλαπλασιάζουμε κάθε τιμή της μεταβλητής με το αντίστοιχο αριθμό συχνοτήτων και διαιρούμε το άθροισμα των γινομένων με το συνολικό αριθμό των συχνοτήτων. Αυτό φαίνεται αν εφαρμόσουμε τον τύπο του μέσου αριθμητικού στα δεδομένα του πίνακα *5.1*

Τιμές μεταβλητής (x_i)	Συχνότητες (f_i)	$f_i x_i$
x_1	f_1	$f_1 x_1$
x_2	f_2	$f_2 x_2$
x_3	f_3	$f_3 x_3$
.	.	.
.	.	.
.	.	.
x_k	f_k	$f_k x_k$
Σύνολο	$\sum f_i$	$\sum f_i x_i$

Ο μέσος αριθμητικός θα δίνεται από τον τύπο: $\mu = \dfrac{x_1 f_1 + x_2 f_2 + x_3 f_3 + ... + x_i f_i + x_k f_k}{f_1 + f_2 + f_3 + ... + f_i + ... + f_k}$

Αν συμβολίσουμε με:

$$\sum f_i x_i = x_1 f_1 + x_2 f_2 + x_3 f_3 + ... + x_i f_i + ... + x_k f_k$$

Και με: $\sum f_i = f_1 + f_2 + ... + f_i + ... + f_k$ τότε ο τύπος του μέσου αριθμητικού μπορεί να

γράφει: $\mu = \dfrac{\sum f_i x_i}{\sum f_i}$

Η παραπάνω μέθοδος υπολογισμού του σταθμικού μέσου αριθμητικού ονομάζεται *άμεση μέθοδος*.

Παράδειγμα

Η ποσοστιαία κατανομή των νοικοκυριών των αστικών περιοχών της χώρας, ως προς τον αριθμό των δωματίων, έχει όπως δείχνει ο παρακάτω πίνακας, Ζητείται ο μέσος αριθμητικός.

Πίνακας

Αριθμός δωματίων(x_i)	Αναλογία νοικοκυριών(f_i%)	$f_i x_i$
1	17	17
2	35	70
3	26	78
4	14	56
5	5	25
6	2	12
7	1	7
Σύνολο	100	265

Ο μέσος αριθμητικός θα είναι:

$$\mu = \frac{\sum f_i x_i}{\sum f_i} = \frac{265}{100} = 2.65$$

Ο αριθμός 2,65 σημαίνει ότι κάθε νοικοκυριό που κατοικεί σε μια αστική πόλη διαθέτει κατά μέσο όρο 2,65 δωμάτια.

b. Όταν τα δεδομένα εμφανίζονται με μορφή κατανομής συχνοτήτων κατά τάξεις (περίπτωση συνεχούς μεταβολής)

Στην περίπτωση αυτή, για τον υπολογισμό του μέσου αριθμητικού με την άμεση μέθοδο, βρίσκουμε τις κεντρικές τιμές όλων των τάξεων στη συνέχεια πολλαπλασιάζουμε τις κεντρικές τιμές με τις αντίστοιχες συχνότητες κάθε τάξης, προσθέτουμε τα γινόμενα και διαιρούμε το άθροισμα τους με το άθροισμα των συχνοτήτων. Έτσι, από τα δεδομένα του πίνακα συχνοτήτων της γενικής μορφής (βλ. πίνακα), ο μέσος αριθμητικός υπολογίζεται με την άμεση μέθοδο από τον τύπο: $\mu = \dfrac{\sum f_i x_i}{\sum f_i}$

Πίνακας

Τάξεις	Συχνότητες (f_i) (απόλυτες ή σχετικές)	Κεντρικές τιμές των τάξεων (x_i)	$f_i x_i$
$\alpha_0 - \alpha_1$	f_1	x_1	$f_1 x_1$
$\alpha_1 - \alpha_2$	f_2	x_2	$f_2 x_2$

$\alpha_2 - \alpha_3$	f_3	x_3	$f_3 x_3$
.	.	.	.
.	.	.	.
.	.	.	.
$\alpha_{i-1} - \alpha_i$	f_i	x_i	$f_i x_i$
.	.	.	.
.	.	.	.
.	.	.	.
$\alpha_{\kappa-1} - \alpha_\kappa$	f_k	x_k	$f_k x_k$
Σύνολο	$\sum f_i$		$\sum f_i x_i$

Παράδειγμα

Ένα εργοστάσιο απασχολεί 80 εργάτες. Η κατανομή των ημερομησθίων έχει ως εξής:

Πίνακας

Τάξεις ημερομισθίων	Αριθμός εργατών(f_i)	Κεντρικές τιμές(x_i)	$f_i x_i$
70-80	4	75	300
80-90	8	85	680
90-100	18	95	1,710
100-110	20	105	2,100
110-120	30	115	3,450
Σύνολο	80		8,240

Να υπολογιστεί το μέσο ημερομίσθιο με την άμεση μέθοδο.

Λύση:

Αν οι εργάτες κάθε τάξης έπαιρναν ως ημερομίσθιο τον κεντρικό όρο της, τότε το συνολικό ποσό της ημερήσιας μισθοδοσίας θα ήταν 8,240 ευρώ. Οπότε, αν κάθε ένας έπαιρνε το ίδιο ακριβώς ημερομίσθιο, αυτό θα ήταν :

$$\mu = \frac{\sum f_i x_i}{\sum f_i} = \frac{8.240}{80} = 103\,\text{ευρώ.}$$

Επομένως, το μέσο ημερομίσθιο των 103 ευρώ ημερησίως είναι κατά κάποιο τρόπο το ημερομήσθιο που θα έπαιρνε καθένας από τους 80 εργάτες, αν η αμοιβή ήταν για όλους η ίδια.

5.1.1.2 Μέτρα θέσης

I.Διάμεσος

Η διάμεσος αποτελεί τη χαρακτηριστικότερη παράμετρο θέσεως και χρησιμοποιείται συχνότατα, μαζί με τον αριθμητικό μέσο, προκειμένου να εξαχθούν συμπεράσματα σε μια στατιστική μελέτη. Αν οι τιμές μιας μεταβλητής τοποθετηθούν κατά φυσική αύξουσα τάξη μεγέθους, τότε *διάμεσος τιμή ονομάζεται η στατιστική εκείνη παράμετρος η οποία χωρίζει τις τιμές της* *μεταβλητής σε δυο ίσες ομάδες,* δηλαδή ο όρος, πριν και μετά τον οποίο ισοκατανέμονται οι υπόλοιποι όροι της τάξης.

Η διάμεσος συμβολίζεται με το γράμμα *Μ*.

Για τον υπολογισμό της τιμής της διαμέσου διακρίνουμε δύο περιπτώσεις.

a) Περίπτωση αταξινόμητων παρατηρήσεων

Στην περίπτωση των αταξινόμητων παρατηρήσεων, δηλαδή όταν το πλήθος των παρατηρήσεων δεν εμφανίζεται σε μορφή κατανομών συχνοτήτων, για τον υπολογισμό της διαμέσου διακρίνουμε δύο περιπτώσεις.

1. *Όταν το πλήθος των παρατηρήσεων είναι αριθμός περιττός.*

Τότε ως διάμεσο παίρνουμε την τιμή εκείνη της μεταβλητής που βρίσκεται ακριβώς στο κέντρο, αφού προηγουμένως οι τιμές της μεταβλητής τοποθετούν κατά φυσική αύξουσα τάξη μεγέθους.

Π.χ. μια μεταβλητή *X* παίρνει τις τιμές:

5,2,3,13,7

Τις τιμές αυτές τις τοποθετούμε πρώτα κατά φυσική αύξουσα τάξη: 2,3,5,7,13. Η θέση της διαμέσου καθορίζεται από τον αριθμό:

$$\frac{N+1}{2} = \frac{5+1}{2} = 3$$

Όπου *N* το πλήθος των παρατηρήσεων, δηλαδή ο τρίτος όρος είναι η διάμεσος. Επομένως :*M=5*.

2. *Όταν το πλήθος των τιμών της μεταβλητής είναι αριθμός άρτιος.*

Τότε η διάμεσος ορίζεται ως ο μέσος αριθμητικός των τιμών των δύο κεντρικών όρων.

Π.χ. Η διάμεσος των παρακάτω παρατηρήσεων:

6,8,12,10,19,15 υπολογίζεται ως εξής:

Τοποθετούμε πρώτα τις παρατηρήσεις κατά φυσική αύξουσα σειρά:6,8,10,12,15,19.

Η θέση της διαμέσου καθορίζεται από τον αριθμό:

$$\frac{N+1}{2}=\frac{6+1}{2}=3,5.$$

Δηλαδή η διάμεσος θα είναι η τιμή εκείνη που έχει την *3,5* θέση. Επομένως, η διάμεσος περιλαμβάνεται μεταξύ του τρίτου και του τετάρτου όρου της σειράς των παρατηρήσεων και δίνεται από το μέσο αριθμητικό των κεντρικών τιμών, δηλαδή:

$$\frac{10+12}{2}=11.$$

Επομένως: *M=11*.

b) Περίπτωση ταξινομημένων δεδομένων

Στην περίπτωση αυτή, οι τιμές της μεταβλητής εμφανίζονται με μορφή κατανομής συχνοτήτων. Για τον υπολογισμό της διαμέσου διακρίνουμε δυο περιπτώσεις.

- Όταν η κατανομή είναι συνεχής.

Για τον υπολογισμό της διαμέσου σχηματίζουμε τη δεξιόστροφη αθροιστική σειρά *F1*,F2,**…**FNτων συχνοτήτων. Μετά διαιρούμε το σύνολο των παρατηρήσεων με το *2*, δηλαδή *N/2,* και βρίσκουμε έτσι το μέσο της συνολικής συχνότητας, το οποίο αντιστοιχεί σε κάποια τάξη της κατανομής. Ύστερα χρησιμοποιούμε τον τύπο:

$$M = \alpha_{i+1} + \frac{\delta}{f_i}(\frac{N}{2} - F_{i-1}).$$

Όπου M η διάμεσος που ζητούμε

α_{i-1} το κατώτερο όριο της τάξης στην οποία εντοπίζεται η διάμεσος

f_i. Η συχνότητα της τάξης στην οποία εντοπίζεται η διάμεσος

F_{i-1} η δεξιόστροφη αθροιστική συχνότητα της τάξης που προηγείται εκείνης στην οποία εντοπίζεται η διάμεσος

δ το πλάτος του διαστήματος τάξης στην οποία εντοπίζεται η διάμεσος

Ν ο συνολικός αριθμός συχνοτήτων της κατανομής.

Παράδειγμα

Στον παρακάτω πίνακα δίνεται η ταχύτητα 100 αυτοκινήτων με την οποία πέρασαν από μια διασταύρωση

Πίνακας

Τάξεις	f_i	F_i
30-40	5	5
40-50	12	17
50-60	19	36
60-70	30	66
70-80	17	83
80-90	10	93
90-100	7	100
Σύνολο	100	

Να υπολογιστεί η διάμεσος ταχύτητα.

Λύση:

Για τον υπολογισμό της διαμέσου όταν η κατανομή παρουσιάζει συχνότητα, κάνουμε τις παρακάτω εργασίες:

1. Σχηματίζουμε τη δεξιόστροφη αθροιστική σειρά των συχνοτήτων (F_i).
2. Προσδιορίζουμε την τιμή Ν/2, όπου Ν το σύνολο των συχνοτήτων, δηλαδή 100/2=50.
3. Η τιμή Ν/2=50 βρίσκεται ανάμεσα σε διαδοχικους όρους της αθροιστικής σειράς F_i (εδώ ανάμεσα στο 36 και στο 66). Ο προηγούμενος όρος, δηλαδή ο 36, είναι ο F_{i-1}.
4. Παρατηρούμε ότι ο επόμενος όρος, δηλαδή ο 66, ανήκει στο ταξικό διάστημα 60-70, το κατώτατο όριο του οποίου το συμβολίζουμε με α_{i-1}, δηλαδή α_{i-1}=60.
5. Πηγαίνουμε στην τάξη από την οποία προσδιορίσαμε την τιμή α_{i-1} και παρατηρούμε πόσες συχνότητες έχει. Αυτή είναι η τιμή του f_i, εδώ f_i=30
6. δ είναι το πλάτος της τάξης στην οποία ανήκει το α_{i-1}, οπότε έχουμε:

$$M = a_{i-1} + \frac{\delta}{f_i}\left(\frac{N}{2} - F_{i-1}\right) = 60 + \frac{10}{30}(50-36) = 60 + \frac{14}{3} = 64.6$$

- Όταν η κατανομή συχνοτήτων είναι ασυνεχής.

Για τον υπολογισμό της διαμέσου όταν η κατανομή παρουσιάζει συχνότητες αλλά όχι τάξεις, κάνουμε τις παρακάτω ενέργειες:

- Σχηματίζουμε τη δεξιόστροφη αθροιστική σειρά συχνοτήτων *(Fi)*
- Προσδιορίζουμε την τιμή *N/2,* όπου *N* το σύνολο των συχνοτήτων.
- Η τιμή *N/2* περιέχεται ανάμεσα σε δυο διαδοχικούς όρους της αθροιστικής σειράς *(Fi),* δηλαδή μεταξύ F_{i-1} και $F_i (F_{i-1} < N/2 < F_i$. Η τιμή της μεταβλητής που αντιστοιχεί στην τιμή *(Fi)* είναι η τιμή της διαμέσου, δηλαδή $M = x_i$.

Ένα κείμενο υπαγορεύτηκε σε 100 μαθητές. Ο αριθμός των ορθογραφικών σφαλμάτων δίνεται στο πίνακα που ακολουθεί.

Πίνακας

Αριθμος σφαλμάτων(x_i)	Αριθμός μαθητών(f_i)	Δεξιόστροφη αθροιστικη σειρα(Fi)
0	12	12
1	27	39
2	29	68
3	19	87
4	8	95
5	4	99
6	1	100
Σύνολο	100	

Ζητείται η διάμεσος.

Λύση:

Για τον υπολογισμό της διαμέσου, γνωρίζοντας ότι η κατανομή είναι ασυνεχής: α) σχηματίζουμε την αθροιστική σειρά *Fi,* β) υπολογίζουμε την τιμή *N*/2=100/2=50 και γ) παρατηρούμε ότι η τιμή *N*/2=50 βρίσκεται ανάμεσα στους αριθμούς 39 και 68 της αθροιστικής σειράς. Ο μεγαλύτερος αριθμός (68) αντιστοιχεί στην τιμή της μεταβλητής x=2, επομένως Μ=2.

<u>ΙΙΙ Επικρατούσα τιμή</u>

Επικρατούσα τιμή ονομάζεται εκείνη η τιμή της μεταβλητής που παρουσιάζει τη μεγαλύτερη συχνότητα, χρησιμοποιείται για να δηλώσει την κεντρική τάση των δεδομένων και ορίζεται ως η τιμή που εμφανίζεται με τη μεγαλύτερη συχνότητα και συμβολίζεται με M_o. Αν υπάρχουν πάνω από μια τέτοιες τιμές, τότε όλες θεωρούνται επικρατούσες τιμές. Είναι φανερό πως η επικρατούσα τιμή δεν έχει νόημα όταν το δείγμα δεν αποτελείται από διακεκριμένες επαναλαμβανόμενες τιμές.

Η επικρατούσα τιμή υπολογίζεται από τον τύπο :

$$M_o = \alpha_{i-1} + \frac{\Delta_1 * \delta}{\Delta_1 + \Delta_2}$$

Όπου :

α_{i-1} το κατώτερο όριο της τάξης στην οποία ανήκει ο μεγαλύτερος αριθμός συχνοτήτων στο πλάτος της τάξης.

Δ_1 η διαφορά μεταξύ της μέγιστης συχνότητας και της προηγούμενης

Δ_2 η διαφορά μεταξύ της μέγιστης συχνότητας και της επομενης

Παράδειγμα

Δίνεται η κατανομή ημερομισθίων σε ευρώ 100 υπαλλήλων μιας επιχείρησης.

Πίνακας

Ημερομίσθια (σε ευρώ)	Συχνότητες(f_i)
30-40	5
40-50	12
50-60	19
60-70	30
70-80	17
80-90	10
90-100	7
Σύνολο	100

Να υπολογιστεί η επικρατούσα τιμή και να ερμηνευτεί η σημασία της.

Λύση:

Παρατηρούμε ότι η μεγαλύτερη συχνότητα (30) αντιστοιχεί στην τάξη 60-70, επομένως μπορούμε να θέσουμε δ=10, α_{i-1}=60, Δ_1=11, Δ_2=13

Οπότε:

$$M_o = a_{i-1} + \delta \frac{\Delta_1}{\Delta_1 + \Delta_2} = 60 + 10 \frac{11}{11+13} = 64.58 \text{ ευρώ.}$$

Λέγοντας ότι επικρατέστερο ημερομίσθιο είναι 64,58 ευρώ, εννοούμε ότι, ανάμεσα στους εργάτες που εργάζονται στην επιχείρηση, αυτοί που παίρνουν το ημερομίσθιο των 64,58 ευρώ είναι και οι περισσότεροι.

Δύο πλεονεκτήματα έχει η επικρατούσα τιμή σε σχέση με τα άλλα περιγραφικά μέτρα:

- Μπορεί να υπολογιστεί και για κατανομές που είναι ανοιχτές προς τα πάνω ή προς τα κάτω.
- Μπορεί να χρησιμοποιηθεί και σε ονομαστικά δεδομένα.

Μειονέκτημα επικρατούσας τιμής:

- Μπορεί να βρεθεί μόνο για μονοκόρυφες κατανομές.

<u>Σχέση μεταξύ αριθμητικού μέσου, διαμέσου και επικρατούσας τιμής.</u>

Για έναν πληθυσμό, ο συμβολισμός είναι:

- Αριθμητικός Μέσος μ
- Διάμεσος M
- Επικρατούσα Τιμή M_O

Για το δείγμα ο αντίστοιχος συμβολισμός είναι:

- Αριθμητικός Μέσος $\overline{X}$
- Διάμεσος m
- Επικρατούσα Τιμή t

Γενικά για έναν πληθυσμό ισχύει:

- Όταν $\mu = M_O = T$, τα δεδομένα προέρχονται από μια συμμετρική κατανομή.
- Όταν έχω θετική ασυμμετρία, ισχύει $\mu > M_O > T$.
- Όταν έχω αρνητική ασυμμετρία, ισχύει αντίστροφα $\mu < M_O < T$.

Για μια κατανομή με μικρή ή μέτρια ασυμμετρία ισχύει ο εμπειρικός τύπος:

$\bar{X} - \tau = 3(\bar{X} - m) \Rightarrow 3m - 2\bar{X} = \tau$.

Μπορεί να χρησιμοποιηθεί, αν γνωρίζω το μέσο και τη διάμεσο.

Διαφορά μέσης τιμής και διαμέσου

Τόσο η μέση τιμή όσο και η διάμεσος αποτελούν μέτρα θέσης του κέντρου της κατανομής. Η μέση τιμή εκφράζει το σημείο ισορροπίας του ιστογράμματος, ενώ η διάμεσος είναι η τιμή που αφήνει μισές παρατηρήσεις αριστερά της (μικρότερες από αυτήν) και μισές δεξιά της (μεγαλύτερες από αυτήν).

Στον υπολογισμό της μέσης τιμής συνεισφέρουν με την ίδια βαρύτητα όλες οι παρατηρήσεις, ενώ στον υπολογισμό της διάμεσου χρησιμοποιούμε την παρατήρηση (ή τις παρατηρήσεις αν έχουμε άρτιο αριθμό παρατηρήσεων) που βρίσκονται ακριβώς στο μέσο των διατεταγμένων δεδομένων. Κατά συνέπεια αν υπάρχουν ακραίες τιμές αυτές επηρεάζουν κατά πολύ την μέση τιμή ενώ αφήνουν αμετάβλητη την διάμεσο.

Τέλος, η σχετική θέση μέσης τιμής $(\bar{X})$ και διαμέσου *(M)* μπορεί να μας πληροφορήσει για την συμμετρία (όταν $\bar{X} = \mathrm{M}$) δεξιά ασυμμετρία (όταν $\bar{X} > \mathrm{M}$) και αριστερή ασυμμετρία (όταν $\bar{X} < \mathrm{M}$) της κατανομής των δεδομένων.

IV Ποσοστημόρια

Για μια ολοκληρωμένη περιγραφή της κατανομής της συνολικής μοναδιαίας πιθανότητας κατά μήκος του άξονα των*x*, χρειάζονται και ορισμένα άλλα σημεία του πραγματικού άξονα (παράμετροι) γνωστά ως ποσοστημόρια.

Ορισμός.

Τιμή τάξεως *k* ή *k*-ποσοστημόριο με $0<k<1$ της τυχαίας μεταβλητής *X* ονομάζεται ένας πραγματικός αριθμός x_k (σημείο του πραγματικού άξονα) τέτοιος ώστε:

$$P\{X < x_K\} \le k \le P\{X \le x_k\}.$$

Αν η τυχαία μεταβλητή *X* είναι διακριτή (αντίστοιχα συνεχής), η τιμή τάξεως *k*αυτής είναι η τιμή x_i της *ΤΜΧ*, για την οποία ισχύει:

$$F(x_{i-1}) < k < F(x_i).$$

(αντίστοιχα η τιμή x_k είναι ρίζα της εξίσωσης $\int_{-\infty}^{x} f(x)dx = k$ ή $F(x_K) = k$ **).**

Γενικά, η τιμή τάξεως *k* δεν ορίζεται μονοσήμαντα και στη διακριτή *ΤΜΧ* δεν υπάρχει πάντοτε. Στη συνεχή *ΤΜΧ* ποσοστημόρια κάθε τάξεως υπάρχουν εφόσον η F(x) είναι γνησίως αύξουσα.

Τα ποσοστημόρια για k=25%, 50%, 75% ονομάζονται τεταρτημόρια και συμβολίζονται αντίστοιχα ως:

Q_1 :(πρώτο τεταρτημόριο)

Q_2 : (δεύτερο τεταρτημόριο ή διάμεσος)

Q_3 : (τρίτο τεταρτημόριο). Ομοίως ορίζονται δεκατημόρια, εκατοστημόρια κ.ο.κ.

Ορισμός.

Η διαφορά $Q_3 - Q_1$ (αντίστοιχα ο λόγος $\frac{Q_3 - Q_1}{2}$) ονομάζεται **ενδοτεταρτημοριακόπλάτος** (αντίστοιχα τεταρτημοριακή απόκλιση). Και τα δυο αυτά ποσοτικά μεγέθη χρησιμοποιούνται ως μέτρα της μεταβλητότητας της κατανομής.

Τεταρτημόρια

Όπως εξηγήσαμε, η διάμεσος χωρίζει το σύνολο των δεδομένων σε δυο υποσύνολα, το καθένα εκ των οποίων περιλαμβάνει το 50% των παρατηρήσεων.

Εάν, με την ίδια λογική χωρίσουμε το σύνολο των στατιστικών δεδομένων σε τέσσερα, ισομεγέθη υποσύνολα, τότε το καθένα από τα υποσύνολα αυτά θα περιλαμβάνει το 25% των στατιστικών μονάδων.

1^{ο} τεταρτημόριο (Q_1)**.**

Είναι η τιμή εκείνη της μεταβλητής κάτω από την οποία βρίσκεται το 25% του συνόλου των παρατηρήσεων και επάνω από αυτή το 75% των παρατηρήσεων.

Για τον υπολογισμό του πρώτου τεταρτημορίου διακρίνουμε, όπως και στη διάμεσο, δυο περιπτώσεις:

- Όταν το πλήθος των παρατηρήσεων δεν εμφανίζεται σε μορφή κατανομής συχνοτήτων, τότε η θέση του πρώτου τεταρτημορίου καθορίζεται από τον αριθμό $\frac{N+1}{4}$.

- Όταν οι τιμές των παρατηρήσεων ξεπερνούν τις *30*.

Τοποθετούμε τις τιμές των παρατηρήσεων σε μορφή κατανομής συχνοτήτων και με παρεμβολή υπολογίζουμε τη διάμεσο με τη βοήθεια του παρακάτω τύπου, εφόσον η μεταβλητή είναι συνεχής.

$$Q_1 = \alpha_{i-1} + \frac{\delta}{f_i}\left(\frac{N}{4} - F_{i-1}\right)$$

Παράδειγμα

Τα ημερομήσθια των εργατών μιας επιχείρησης κατανέμονται όπως δείχνει ο κάτωθι πίνακας:

Πίνακας

Τάξεις	f_i	F_i
100-110	30	30
110-120	80	110
120-130	120	230
130-140	150	380
140-150	75	455
150-160	65	520
Σύνολο	520	

Να υπολογιστεί το πρώτο τεταρτημόριο.

Λύση:

$$Q_1 = a_{i-1} + \frac{\delta}{f_i}\left(\frac{N}{4} - F_{i-1}\right) = 120 + \frac{10}{120}\left(\frac{520}{4} - 110\right) = 121,66$$

3ο τεταρτημόριο (Q_3).

Ονομάζεται η τιμή εκείνη της μεταβλητής κάτω από την οποία βρίσκεται το 75% του συνόλου των παρατηρησεων και επάνω από αυτή το 25%. Διακρίνουμε δυο περιπτώσεις:

- Όταν το πλήθος των παρατηρήσεων είναι μικρό. Τοτε η θέση του τρίτου τεταρτημορίου καθορίζεται από τον αριθμό: $\frac{3(N+1)}{4}$.
- Όταν οι τιμές των παρατηρήσεων ξεπερνούν τις *30*.

Τοποθετούμε αυτές τις τιμές σε μορφή κατανομής συχνοτήτων και εφαρμόζουμε τον παρακάτω τύπο: $Q_3 = \alpha_{i-1} + \frac{\delta}{f_i}(\frac{3N}{4} - F_{i-1})$.

Παράδειγμα

Το ύψος σε εκατοστά της βροχόπτωσης που παρατηρήθηκε κατά το μήνα Δεκέμβριο του 1976 σε 132 βαθμούς είχε όπως δείχνει ο παρακάτω πίνακας.

Τάξεις	f_i	F_i
20-25	2	2
25-30	10	12
30-35	25	37
35-40	50	87
40-45	20	107
45-50	12	119
50-55	8	127
55-60	5	132
Σύνολο	132	

Να υπολογιστεί το τρίτο τεταρτημόριο.

Λύση:

$$Q_3 = a_{i-1} + \frac{\delta}{f_i}\left(\frac{3N}{4} - F_{i-1}\right) = 40 + \frac{5}{20}\left(\frac{132*3}{4} - 87\right) = 43$$

2^o τεταρτημόριο (Q_2).

Επειδή ο τύπος της διαμέσου είναι ίδιος με τον τύπο του $2^{ου}$ τεταρτημορίου :

$$M = a_{i-1} + \frac{\delta}{f_i}\left(\frac{N}{2} - F_{i-1}\right) = Q_2$$

Άρα :

Το 2^o τεταρτημόριο συμπίπτει με τη διάμεσο.

Δεκατημόρια-Εκατοστημόρια.

Εάν το πλήθος των στατιστικών δεδομένων είναι μεγάλο, ενδιαφέρον παρουσιάζει πολλές φορές ο υπολογισμός και άλλων παραμέτρων θέσεως εκτός από τη διάμεσο *M* και τα δυο τεταρτημόρια Q_1 και Q_3. Οι παράμετροι αυτές είναι τα δεκατημόρια D_k και τα εκατοστημόρια C_k.

Εάν χωρίσουμε το πλήθος των (διατεταγμένων κατά τάξη αυξανόμενου μεγέθους) δεδομένων, σε 10 διαδοχικά υποσύνολα, το κάθε ένα από τα υποσύνολα αυτά θα περιλαμβάνει το 10% των όρων της σειράς (ή κατανομής).

Το πρώτο δεκατημόριο D_1 βρίσκεται στην θέση κάτω από την οποία υπάρχει το 10% των όρων και πάνω από την οποία, το υπόλοιπο 90% των όρων. Ομοίως το D_2 βρίσκεται στην θέση κάτω από την οποία υπάρχει το 20% των όρων και πάνω από την οποία το υπόλοιπο 80% κ.ο.κ.

Τα δεκατημόρια είναι συνολικά εννέα, με τελευταίο (μεγαλύτερο) το D_9. Δηλαδή D_k, k=1,2,3,…,9.

Σύμφωνα με τα παραπάνω, ο γενικός τύπος υπολογισμού των δεκατημορίων μιας κατανομης, είναι:

$$D_k = d_{i-1} + \frac{\delta}{f_i}\left(\frac{k*N}{10} - F_{i-1}\right)$$

Όπου k=1,2,3,…,9.

Παράδειγμα

Σαράντα μαθητές έλαβαν τους βαθμούς που δείχνει ο παρακάτω πίνακας σε ένα test εξυπνάδας.

Πίνακας

Τάξεις	(f_i)	(F_i)
30-40	2	2
40-50	5	7
50-60	7	14
60-70	8	22
70-80	4	26
80-90	8	34
90-100	6	40
Σύνολο	40	

Να υπολογιστεί το 6^o δεκατημόριο.

Λύση:

Εφαρμόζουμε τον τύπο:

$$D_6 = d_{i-1} + \frac{\delta}{f_i}\left(\frac{6N}{10} - F_{i-1}\right) = 70 + \frac{10}{4}(24 - 22) = 75$$

Με εντελώς ανάλογο τρόπο υπολογίζουμε τα εκατοστημόρια μιας κατανομής, χωρίζοντας το πλήθος των όρων σε 100 ισοπληθή υποσύνολα.

Ο γενικός τύπος υπολογισμού του κάθε εκατοστημορίου είναι:

$$C_k = d_{i-1} + \frac{\delta}{f_i}\left(\frac{k*N}{100} - F_{i-1}\right).$$

Παράδειγμα

Το ύψος σε εκατοστά της βροχόπτωσης που παρατηρήθηκε κατά το μήνα Νοέμβριο του 1973 σε 100 σταθμούς είχε ως εξής:

Πίνακας

Τάξεις	(f_i)	(F_i)
0-10	17	17
10-20	16	33
20-30	14	47
30-40	15	62
40-50	12	74
50-60	10	84
60-70	9	93
70-80	4	97
80-90	2	99
90-100	1	100
Σύνολο	100	

Να υπολογιστεί το 5^o εκατοστημόριο.

Λύση:

$$C_5 = d_{i-1} + \frac{\delta}{f_i}\left(\frac{5N}{100} - F_{i-1}\right) = 0 + \frac{10}{17}(5-0) = 2.94$$

Παρατήρηση

Όλοι οι τύποι υπολογισμού της διάμεσου και των τεταρτημορίων, δεκατημορίων και εκατοστημορίων μιας κατανομής, έχουν γενικά, την ίδια «δομή». Το μόνο σημείο στο οποίο διαφέρουν είναι ο μειωτέος στην παρένθεση, ο οποίος καθορίζει και την θέση που έχει η ζητούμενη παράμετρος, μέσα στο πλήθος των στατιστικών δεδομένων, που μελετάμε.

Η παρατήρηση αυτή, μας οδηγεί στο συμπέρασμα, ότι μπορούμε να υπολογίσουμε την τιμή οποιουδήποτε όρου μιας κατανομής, χωρίς ο όρος αυτός υποχρεωτικά να είναι η διάμεσος ή κάποιο τεταρτημόριο δεκατημόριο ή εκατοστημόριο.

6. ΔΙΑΣΠΟΡΑ

6.1 ΕΝΝΟΙΑ ΤΗΣ ΔΙΑΣΠΟΡΑΣ

Ο μέσος αριθμητικός , ο μέσος γεωμετρικός, η διάμεσος, η επικρατούσα τιμή και τα υπόλοιπα άλλα μέτρα τα οποία εξετάζουμε έχουν ως αντικειμενικό σκοπό να αντιπροσωπεύουν ένα πληθυσμό με μια μόνο παράμετρο, η οποία μας δίνει το σημείο στο οποίο τείνουν να συγκεντρωθούν οι τιμές της μεταβλητής του πληθυσμού που ερευνούμε.

Η αντιπροσώπευση όμως ενός πληθυσμού με μια από τις πιο πάνω παραμέτρους έχει αξία εφόσον ο πληθυσμός παρουσιάζει μεγάλη *ομοιογένεια*. Αντίθετα, αν ο πληθυσμός παρουσιάζει μεγάλη *ανομοιογένεια,* τότε τα μέτρα της κεντρικής τάσης και θέσης θα πρέπει να μη χρησιμοποιούνται ως αντιπροσωπευτικοί αριθμοί ενός πληθυσμού.

Ο βαθμός κατά τον οποίο οι διάφορες τιμές ενός πληθυσμού τείνουν να είναι διεσπαρμένες γύρω από τον μέσο αριθμητικό ονομάζεται *διασπορά*.

Υποθέτουμε ότι οι μεταβλητές Χ και Υ παίρνουν τις τιμές:

x_i :**10,40,43,46,47,48,50,50,52,52,54**

y_i **:7,14,15,23,38,48,50,50,75,85,90**

Παρατηρούμε ότι και στις δυο μεταβλητές ο μέσος αριθμητικός είναι 45 και η διάμεσος 48, η μια όμως μεταβλητή έχει διάφορη μορφή από την άλλη.

Στη μεταβλητή Χ οι τιμές των παρατηρήσεων κυμαίνονται μέσα στους αριθμούς 10 και 45, ενώ στη μεταβλητή Υ οι τιμές κυμαίνονται ανάμεσα στους αριθμούς 7 και 90.

Επομένως, οι πληροφορίες που μας δίνουν οι παράμετροι που χαρακτηρίζουν την τάση και οι παράμετροι που χαρακτηρίζουν τη θέση μιας κατανομής είναι ανεπαρκείς, γιατί δε μας δίνουν ενδείξεις για τον τρόπο συγκέντρωσης των τιμών της μεταβλητής γύρω από τους κεντρικούς μέσους όρους και, για το λόγο αυτό, είναι αναγκαία η χρησιμοποίηση ενός δείκτη που μας δίνει το *βαθμό συγκέντρωσης ήδιασποράς* των τιμών της μεταβλητής από το μέσο αριθμητικό. Η παράμετρος που μας πληροφορεί αν οι τιμές των παρατηρήσεων είναι συγκεντρωμένες ή διασκορπισμένες σε σχέση με το μέσο αριθμητικό ονομάζεται *διασπορά ήδιακύμανση*.

6.2 Μέτρα Διασποράς

Τα μέτρα διασποράς που συνήθως χρησιμοποιούμε στην Στατιστική είναι: το εύρος μεταβολής, το ημιενδοτεταρτημοριακό εύρος, η μέση απόκλιση, η τυπική απόκλιση και ο συντελεστής μεταβλητότητας.

6.2.1 ΤΟ ΕΥΡΟΣ ΜΕΤΑΒΟΛΗΣ

Είναι μια παράμετρος που μας δίνει τη διαφορά ανάμεσα στη μεγαλύτερη και μικρότερη τιμή μιας σειράς παρατηρήσεων. Πχ η μέση ημερήσια θερμοκρασία σε 18 πόλεις της χώρας στις 25 Αυγούστου του 1978 είχε ως εξής: 19,20,20,23,24,24,25,25,26,28,29,30,31,32,33,34,35. Το εύρος μεταβολής των πιο πάνω παρατηρήσεων θα είναι η διαφορά ανάμεσα στη μεγαλύτερη και μικρότερη θερμοκρασία, δηλαδή:

R = M-E = 35-19 = 16

Η παράμετρος όμως αυτή δεν είναι ικανοποιητική, γιατί εξαρτάται μόνο από τις ακραίες τιμές της σειράς των παρατηρήσεων που μας δόθηκε. Αν όμως οι δυο ακραίες τιμές είναι *αρκετά απομακρυσμένες* από τα σύνολα των άλλων παρατηρήσεων, μας δίνουν ψεύτικη εικόνα της έντασης της διασποράς.

Το εύρος μεταβολής χρησιμοποιείται κυρίως στη Μετεωρολογία και στα χρηματιστήρια.

6.2.2Ενδοτεταρτημοριακό εύρος.

Είναι η διαφορά του πρώτου τεταρτημορίου Q_1 από το τρίτο τεταρτημόριο Q_3. Στο μεταξύ τους διάστημα περιλαμβάνεται το 50% του δείγματος. Επομένως, όσο μικρότερο θα είναι αυτό το διάστημα, τόσο μεγαλύτερη θα είναι η συγκέντρωση των τιμών και άρα μικρότερη η διασπορά των τιμών της μεταβλητής. Το μισό της διαφοράς $Q_3 - Q_1$ είναι το λεγόμενο ημιενδοτεταρτημοριακό εύρος και συμβολίζεται μεQ, δηλαδή:

$$Q = \frac{Q_3 - Q_1}{2}$$

Το Q μετριέται με τις ίδιες μονάδες της μεταβλητής και δεν εξαρτάται από όλες τις τιμές, αλλά μόνο από εκείνες που περιλαμβάνονται στον υπολογισμό των Q_1 και Q_3.

Παράδειγμα

Στον παρακάτω πίνακα δίνεται η κατανομή κατά ηλικία του ελληνικού πληθυσμού σε 10ετείς ομάδες ηλικιών, σύμφωνα με την απογραφή του 1971.

Ζητείται να υπολογιστεί το ημιενδοτεταρτημοριακό εύρος.

Πίνακας

Τάξεις ηλικιών	Ποσοστό (f_i)	F_i
0-10	17	17
10-20	16	33
20-30	14	47
30-40	15	62
40-50	12	74
50-60	10	84
60-70	9	93
70-80	4	97
80-90	2	99
90-100	1	100
Σύνολο	100	

Λύση:

1^o τεταρτημόριο:

$$Q_1 = \alpha_{i-1} + \frac{\delta}{f_i}(\frac{N}{4} - F_{i-1}) = 10 + \frac{10}{16}(25-17) = 15$$

3^o τεταρτημόριο:

$$Q_3 = \alpha_{i-1} + \frac{\delta}{f_i}(\frac{3N}{4} - F_{i-1}) = 50 + \frac{10}{10}(75-74) = 51$$

Το ημιενδοτεταρτημοριακό εύρος θα είναι:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{51-15}{2} = 18$$

6.2.3 ΜΕΣΗ ΑΠΟΚΛΙΣΗ

Η μέση απόκλιση ορίζεται ως ο μέσος αριθμητικός όλων των διαφορών των τιμών μιας μεταβλητής από το μέσο αριθμητικό της μεταβλητής αυτής.

α)Αταξινόμητες παρατηρήσεις

Στην περίπτωση αταξινόμητων στοιχείων, η μέση απόκλιση δίνεται από τον τύπο:

$$\text{M * A} = \frac{\sum |X_i - \mu|}{N}$$

Όπου οι κάθετες γραμμές σημαίνουν ότι παίρνουμε την απόλυτη διαφορά ανάμεσα στις τιμές της μεταβλητής και του μέσου αριθμητικού και N το πλήθος των παρατηρήσεων της μεταβλητής X.

Π.χ δίνονται οι παρατηρήσεις :3,5,7,9,16.

Ο μέσος αριθμητικός θα είναι:

μ = 3+5+7+9+16/5 = 40/5 = 8

Η μέση απόκλιση θα είναι :

$$\text{M*A} = \frac{\sum |X_i - \mu|}{N} = \frac{|3-8|+|5-8|+|7-8|+|9-8|+|16-8|}{5} = \frac{18}{5} = 3.6$$

β)Ταξινομημένα στοιχεία σε μορφή κατανομών συχνοτήτων

Στην περίπτωση ταξινομημένων στοιχείων, η μέση απόκλιση δίνεται από τον τύπο:

$$\text{M*A} = \frac{\sum f_i |X_i - \mu|}{\sum f_i}$$

ΠΑΡΑΔΕΙΓΜΑ

Στον παρακάτω πίνακα δίνεται η κατανομή των ημερών απουσίας 21 εργαζομένων λογω ασθενείας.

Τάξεις	fi	Κεντρικές τιμές	Αποκλίσεις	Αποκλίσεις	Αποκλίσεις
			fixi	|xi-μ|	fi|xi-μ|
1 - 3	2	2	4	4	8
3 - 5	5	4	20	2	10
5 - 7	7	6	42	0	0
7 - 9	5	8	40	2	10
9 - 11	2	10	20	4	8
Σύνολο	21		126		36

Να υπολογιστεί η μέση απόκλιση.

Λύση:

Ο μέσος αριθμητικός είναι μ = $\dfrac{\sum f_i X_i}{\sum f_i} = \dfrac{126}{21} = 6$

Η μέση απόκλιση θα είναι Μ.Α = $\dfrac{\sum f_i \,|\, X_i - \mu \,|}{\sum f_i} = \dfrac{36}{21} = 1.7$

Παρατηρούμε ότι για τον υπολογισμό της μέσης απόκλισης δεν λαμβάνουμε υπόψη μας τα σημεία των αποκλίσεων αλλά μονό τις απόλυτες τιμές των διαφορών. Αυτό γίνεται από ανάγκη, γιατί αν λάβουμε το άθροισμα των αποκλίσεων μεταξύ των τιμών μιας μεταβλητής και του μέσου αριθμητικού, αυτό είναι πάντοτε μηδέν, δηλαδή:

$$\sum (X_i - \mu) = \sum X_i - \mu * \mathrm{N} = \mathrm{N} * \mu - \mu * \mathrm{N} = 0$$

6.2.4 ΔΙΑΚΥΜΑΝΣΗ ΚΑΙ ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ

Για να αποφύγουμε τη διαδικασία εκείνη κατά την οποία δεν λαμβάνουμε υπόψη σημεία των αρνητικών διαφορών μεταξύ του μέσου αριθμητικού και των τιμών της μεταβλητής, μπορούμε να υψώσουμε τις διαφορές στο τετράγωνο για να μην έχουμε να απασχοληθούμε με θετικά ή αρνητικά σημεία των αποκλίσεων.

Ο μέσος αριθμητικός των τετραγώνων των αποκλίσεων είναι επίσης μέτρο διασποράς και ονομάζεται *διακύμανση*.

Επομένως, διακύμανση ενός πλήθους παρατηρήσεων ονομάζεται ο μέσος αριθμητικός των τετραγώνων των αποκλίσεων των τιμών των παρατηρήσεων από τον αριθμητικό μέσο.

Η διακύμανση εκφράζεται σε μονάδες, οι οποίες είναι τα τετράγωνα των αρχικών μονάδων, π.χ αν το εισόδημα είναι σε ευρώ, η διακύμανση εκφράζεται με μονάδα το ευρώ στο τετράγωνο, ομοίως αν η μεταβλητή εκφράζεται σε εκατοστά, η διακύμανση εκφράζεται σε εκατοστά στο τετράγωνο κ.λ.π. Για να έχουμε όμως ένα δείκτη ο όποιος να μετρά τη διασπορά και να εκφράζεται στις ίδιες μονάδες που εκφράζεται η μεταβλητή μας, παίρνουμε την τετραγωνική ρίζα της διακύμανσης. Το μέτρο αυτό ονομάζεται *τυπική απόκλιση* και είναι το μέτρο διασποράς που χρησιμοποιούμε συνήθως στην πράξη.

Όσο μεγαλύτερη είναι η τυπική απόκλιση, τόσο μεγαλύτερη είναι η διασπορά των παρατηρήσεων από το μέσο αριθμητικό.

Διαφορά μεταξύ διακύμανσης και τυπικής απόκλισης.

Η διακύμανση είναι το τετράγωνο της τυπικής απόκλισης. Συνήθως χρησιμοποιούμε την τυπική απόκλιση διότι αυτή εκφράζεται στις μονάδες της μεταβλητής, ενώ η διακύμανση στα τετράγωνα των μονάδων αυτών και κατά συνέπεια δεν είναι ερμηνεύσιμη.

6.2.4.1 ΥΠΟΛΟΓΙΣΜΟΣ ΤΗΣ ΔΙΑΚΥΜΑΝΣΗΣ ΚΑΙ ΤΗΣ ΤΥΠΙΚΗΣ ΑΠΟΚΛΙΣΗΣ

Η διακύμανση συμβολίζεται με σ^2 και η τυπική απόκλιση με σ.

α) Αταξινόμητες παρατηρήσεις

Υποθέτουμε ότι έχουμε παρατηρήσεις $x_{1,}x_2,...,x_i,...,x_N$, που ο μέσος αριθμητικός τους είναι ο μ.

Η διακύμανση των παραπάνω παρατηρήσεων δίνεται από τον τύπο:

$$\sigma^2 = \frac{\sum (X_i - \mu)^2}{\mathrm{N}}$$

Και η τυπική απόκλιση από τον τύπο:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum (X_i - \mu)^2}{\mathrm{N}}}$$

ΠΑΡΑΔΕΙΓΜΑ

Να υπολογιστεί η διακύμανση και η τυπική απόκλιση στις παρακάτω παρατηρήσεις:

χ_ι :2,3,5,8,12

Λύση:

Επειδή οι παρατηρήσεις δεν παρουσιάζουν συχνότητα, θα χρησιμοποιήσουμε τον παρακάτω τύπο για τον υπολογισμό της διακύμανσης:

$$\sigma^2 = \frac{\sum (X_i - \mu)^2}{\mathrm{N}}$$

Υπολογίζουμε πρώτα το μέσο αριθμητικό:

$$\mu = \frac{\sum X_\iota}{\mathrm{N}} = \frac{2+3+5+8+12}{5} = 6$$

Μετά σχηματίζουμε τον πίνακα 6.2

Πίνακας 6.2

xi	xi-μ	(xi-μ)τετράγωνο
2 3 5 8 12	-4 -3 -1 2 6	16 9 1 4 36
Σύνολο		66

Οπότε έχουμε: $\sigma^2 = \dfrac{\sum (X_i - \mu)^2}{\mathrm{N}} = \dfrac{66}{5} = 13,2$

Η τυπική απόκλιση θα είναι:

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{\mathrm{N}}} = \frac{66}{5} = \sqrt{13,2} = 3,63$$

β)Ομαδοποιημένες παρατηρήσεις

Στην περίπτωση κατά την οποία οι παρατηρήσεις δίνονται σε μορφή κατανομής συχνοτήτων, η διακύμανση υπολογίζεται με τους παρακάτω τύπους:

$$\sigma^2 = \frac{\sum f_i (x_i - \mu)^2}{\sum f_i}$$

$$\sigma^2 = \frac{\sum f_i x_i^2}{\sum f_i} - \mu^2$$

$$\sigma^2 = \delta^2 \left[\frac{\sum f_i \xi_i^2}{\sum f_i} - \left(\frac{\sum f_i \xi_i}{\sum f_i} \right)^2 \right]$$

Παράδειγμα

Στον παρακάτω πίνακα δίνεται η ταχύτητα με την οποία 10 αυτοκίνητα πέρασαν από μια επικίνδυνη διασταύρωση.

Πίνακας

Τάξεις	f_i
0-10	4
10-20	3
20-30	2
30-40	1
Σύνολο	10

Ζητείται η διακύμανση και η τυπική απόκλιση με τρεις διαφορετικούς τρόπους.

Λύση:

1ος τρόπος

Υπολογισμός της διακύμανσης και τυπικής απόκλισης με τον τύπο:

$$\sigma^2 = \frac{\sum f_i (x_i - \mu)^2}{\sum f_i}$$

Σχηματίζουμε τον πίνακα

Πίνακας

Τάξεις	f_i	x_i	$x_i - \mu$	$(x_i - \mu)^2$	$f_i (x_i - \mu)^2$	$f_i \ x_i$
0-10	4	5	-10	100	400	20
10-20	3	15	0	0	0	45
20-30	2	25	10	100	200	50
30-40	1	35	20	400	400	35
Σύνολο	10				1000	150

Ο μέσος αριθμητικός είναι: $\mu = \frac{150}{10} = 15$

Η διακύμανση θα είναι:

$$\sigma^2 = \frac{\sum f_i (x_i - \mu)^2}{\sum f_i} = \frac{1000}{10} = 100$$

Και η τυπική απόκλιση:

$$\sigma = \sqrt{100} = 10$$

2^{ος} τρόπος

Υπολογισμός της διακύμανσης με τον τύπο:

$$\sigma^2 = \frac{\sum f_i x_i^2}{\sum f_i} - \mu^2$$

Πίνακας

Τάξεις	f_i	x_i	$f_i x_i$	x_i^2	$f_i x_i^2$
0-10	4	5	20	25	100
10-20	3	15	45	225	675
20-30	2	25	50	625	1250
30-40	1	35	35	1225	1225
Σύνολο	10		150		3250

$$\mu = \frac{\sum f_i x_i}{\sum f_i} = \frac{150}{10} = 15$$

Επομένως:

$$\sigma^2 = \frac{\sum f_i x_i^2}{\sum f_i} - \mu^2 = \frac{3250}{10} - 15^2 = 100$$

Και

$$\sigma = \sqrt{100} = 10$$

Και 3ος τρόπος

Υπολογισμός της διακύμανσης με τον τύπο:

$$\sigma^2 = \delta^2 \left[\frac{\sum f_i \xi_i^2}{\sum f_i} - \left(\frac{\sum f_i \xi_i}{\sum f_i} \right)^2 \right]$$

Σχηματίζουμε τον παρακάτω πίνακα:

Τάξεις	x_i	f_i	ξ_i	ξ_i^2	$f_i\xi_i$	$f_i\xi_i^2$
0-10	5	4	-1	1	-4	4
10-20	15	3	0	0	-	-
20-30	25	2	1	1	2	2
30-40	35	1	2	4	2	4
Σύνολο		10			0	10

Έτσι:

$$\sigma^2 = \delta^2 \left[\frac{\sum f_i \xi_i^2}{\sum f_i} - \left(\frac{\sum f_i \xi_i}{\sum f_i} \right)^2 \right] = 10^2 \left[\frac{10}{10} - \left(\frac{0}{10^2} \right) \right] = 100$$

Και

$\sigma = \sqrt{100} = 10$.

6.2.5 ΣΥΝΤΕΛΕΣΤΗΣ ΜΕΤΑΒΛΗΤΙΚΟΤΗΤΑΣ

Η τυπική απόκλιση, η οποία θεωρείται ως το κυρίως χρησιμοποιούμενο μέτρο για τη μέτρηση της διασποράς, εκφράζεται στις ίδιες μονάδες με τις οποίες εκφράζεται η τυχαία μεταβλητή και μας δίνει τη απόλυτη διασπορά των τιμών της τυχαίας μεταβλητής από το μέσο αριθμητικό.

Η χρησιμοποίηση όμως της τυπικής απόκλισης και των άλλων μέτρων διασποράς είναι σε αρκετές περιπτώσεις αδύνατη και σε άλλες περιπτώσεις πολύ περιορισμένη. Αυτό συμβαίνει όταν θέλουμε να συγκρίνουμε δυο κατανομές οι οποίες εκφράζονται σε διαφορετικές μονάδες (μέτρα, κιλά, ευρώ, εκατοστά ,κ.λ.π), ή όταν οι μέσοι αριθμητικοί δυο διαφορετικών τυχαίων μεταβλητών, έστω και αν εκφράζονται στις ίδιες μονάδες, διαφέρουν πάρα πολύ μεταξύ τους. Τότε τα μέτρα της απόλυτης διασποράς δεν μας εξυπηρετούν και χρησιμοποιούμε τη σχετική διασπορά. Το βασικό μέτρο της σχετικής διασποράς είναι συντελεστής μεταβλητικότητας.Ο συντελεστής αυτός είναι ανεξάρτητος από τις μονάδες μέτρησης που χρησιμοποιούμε και, επομένως, επιτρέπει τη σύγκριση τόσο των ομοειδών όσο και των ετεροειδών κατανομών.Ο συντελεστής μεταβλητικότητας δίνεται από τον τύπο:

$CV(X) = \dfrac{\sigma}{\mu}$ ή $CV(X) = \dfrac{\sigma}{\mu} * 100\%$

Παράδειγμα:

Δίνονται οι παρακάτω παρατηρήσεις των μεταβλητών X και Y,οι οποίες εκφράζονται στην ίδια μονάδα μέτρησης.

χ_i : 44,37,36,46,42

Y_i : 88,79,86,75,84

Ποιά απο τις δύο μεταβλητές παρουσιάζει μεγαλύτερη ομοιογένεια;

Λύση

Απάντηση θα δώσει ο συντελεστής μεταβλητικότητας.

Για το λόγο αυτό σχηματίζουμε τον παρακάτω πίνακα:

Πίνακας

	χ_i	Y_i	$\chi_i - \mu_\chi$	$(\chi_i - \mu_\chi)^2$	$y_i - \mu_y$	$(y_i - \mu_y)^2$
	44	88	3	9	6	36
	37	79	-4	16	-3	9
	36	86	-5	25	4	16
	46	75	5	25	-7	49
	42	82	1	1	0	0
Σύνολο	205	410		76		110

$\mu_x = \frac{205}{5}$ =41

$\mu_y = \frac{410}{5}$ =82

$$\sigma_{\chi}^{2} = \frac{\Sigma\ (\chi_i - \mu_{\chi})^{2}}{\mathrm{N}} = \frac{76}{5} = 15{,}2 \text{ και } \sigma_{\chi} = 3{,}9$$

$$\sigma_{y}^{2} = \frac{\Sigma\ (y_i - \mu_{y})^{2}}{\mathrm{N}} = \frac{110}{5} = 22 \text{ και } \sigma_{y} = 4{,}69$$

Επομένως, ο συντελεστής μεταβλητικότητας θα είναι :

$$CV(x) = \frac{\sigma_x}{\mu_x} * 100\% = \frac{3{,}9}{41} * 100\% = 9{,}5\%$$

$$CV(y) = \frac{\sigma_y}{\mu_y} * 100\% = \frac{4{,}69}{82} * 100\% = 5{,}7\%$$

Επομένως η μεταβλητή Υ παρουσιάζει τη μικρότερη διασπορά.

Άρα, ο συντελεστής μεταβλητικότητας είναι το πηλίκο της τυπικής απόκλισης μιας κατανομής προς τον αριθμητικό μέσο αυτής και εκφράζει την τυπική απόκλιση ως ποσοστό επί τις εκατό του μέσου αριθμητικού μ.

7. ΑΣΣΥΜΕΤΡΙΑ – ΚΥΡΤΩΣΗ

7.1 Μέτρα λοξότητας ή ασυμμετρίας

I. Συντελεστής ασυμμετρίας του Pearson

Η κατανομή ενός πληθυσμού μπορεί να είναι είτε συμμετρική είτε μη συμμετρική.

Στην πρώτη περίπτωση η κορυφή, η διάμεσος και η μέση τιμή συμπίπτουν. Στις άλλες περιπτώσεις ένα από τα τμήματα στα οποία χωρίζει την κατανομή η κορυφή περιέχει περισσότερες παρατηρήσεις, καθώς επίσης και η διάμεσος και η μέση τιμή, βρίσκονται δεξιά της κορυφής και στην περίπτωση αυτή μάλιστα ισχύει $M_0 < x_\delta < x$, και η αρνητική ασυμμετρία στην οποία οι περισσότερες παρατηρήσεις, όπως η διάμεσος και η μέση τιμή, βρίσκονται αριστερά της κορυφής και στην περίπτωση αυτή μάλιστα ισχύει $x < x_\delta < M_0$.

Σαν αριθμητικό μέτρο καθορισμού της ασυμμετρίας το συνηθέστερο είναι ο συντελεστής ασυμμετρίας του Pearson:

$$s_k = \frac{\mu - M_0}{\sigma}.$$

Άλλος δείκτης ασυμμετρίας δίνεται από τον τύπο:

$$s_k = \frac{Q_3 - M \;\; - \;\; M - Q_1}{Q_3 - M \;\; + \;\; M - Q_1} \quad \text{(Δείκτης του Bowley)}$$

Η τιμή του S_k κυμαίνεται ανάμεσα στο -1 και +1, δηλαδή:

$$-1 \leq S_k \leq +1$$

Με το S_k=0 η κατανομή είναι συμμετρική, όσο δε η τιμή του S_k απομακρύνεται από το μηδέν και τείνει προς το ± 1, τόσο η ασυμμετρία είναι εντονότερη. Αν S_k>0, έχουμε θετική ασυμμετρία, ενώ αν S_k<0, έχουμε αρνητική ασυμμετρία.

Αν η τιμή του S_k είναι μεταξύ του μηδενός και του ± 0.1 , η ασυμμετρία είναι μικρή.

Μεταξύ ± 0.1 και 0.3 η ασυμμετρία είναι μέτρια.

Πάνω από το ± 0.3 η ασυμμετρία είναι έντονη.

Επίσης, η ασυμμετρία διακρίνεται σε θετική και αρνητική.

Σε μια συμμετρική κατανομή, ο μέσος αριθμητικός ισούται με τη διάμεσο και την επικρατούσα τιμή και ισχύει η σχέση $Q_1 + Q_3 = 2M$

7.2 Μέτρα Κύρτωσης

<u>Ι. Συντελεστής Κύρτωσης.</u>

Η κύρτωση χαρακτηρίζει την αιχμηρότητα της καμπύλης μιας κατανομής.
Με βάση την κύρτωση, οι κατανομές διακρίνονται σε:

• λεπτόκυρτες

• μεσόκυρτες

• πλατύκυρτες

Ο συντελεστής κύρτωσης β2 ορίζεται από τη σχέση

$$\beta_2 = \frac{1}{n} \cdot \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^4 = \frac{1}{n} \cdot \frac{\sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^4}{s^4}$$

Παρατηρούμε και πάλι ότι αν

$$z_i = \frac{x_i - \bar{x}}{s}$$

είναι οι τυποποιημένες τιμές που αντιστοιχούν στα δεδομένα, τότε

$$\beta_2 = \frac{1}{n} \cdot \sum_{i=1}^{n} z_i^{\,4}$$

Για ομαδοποιημένα δεδομένα, ο συντελεστής κύρτωσης $\beta 2$ γράφεται

$$\beta_2 = \frac{1}{n} \cdot \frac{\sum_{i=1}^{k} f_i (w_i - \bar{x})^4}{s^4}$$

Ένας ακόμη τύπος για το βαθμό κύρτωσης δίνεται απο τον Pearson:

$$\beta_2 = \frac{\mu^4}{\mu_2^2} \text{ ή } \beta_2 = \frac{\mu_4}{\sigma^4}$$

Ο βαθμός κύρτωσης δίνεται επίσης και από τον τύπο του Fisher:

$$\gamma_2 = \beta_2 - 3 = \frac{\mu_4}{\sigma^4} - 3$$

Και ο οποίος για μεσόκυρτη καμπύλη είναι 0.

• Σε αντίθεση με το συντελεστή ασυμμετρίας, ο συντελεστής κύρτωσης παίρνει μόνο θετικές τιμές.

- Για μεσόκυρτες κατανομές, όπως π.χ. η κανονική, ισχύει β2 = 3.
- Όταν η τιμή του συντελεστή είναι β2 < 3, τότε η κατανομή είναι πλατύκυρτη.
- Όταν η τιμή του συντελεστή είναι β2 > 3, τότε η κατανομή είναι λεπτόκυρτη.

Παράδειγμα

Έστω ότι έχουμε τα δεδομένα 10, 12, 17, 21, 23, 27, 30

• Η μέση τιμή είναι $x = 20$.

• Για τη διακύμανση και τους συντελεστές ασυμμετρίας και κύρτωσης, κατασκευάζουμε τον πίνακα.

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$	$(x_i - \bar{x})^4$
10	-10	100	-1000	10000
12	-8	64	-512	4096
17	-3	9	-27	81
21	1	1	1	1
23	3	9	27	81
27	7	49	343	2401
30	10	100	1000	10000

Από τον πίνακα, βρίσκουμε ότι

$$\sum_{i=1}^{7} \left(x_i - \overline{x}\right)^2 = 332$$

$$\sum_{i=1}^{7} \left(x_i - \overline{x}\right)^3 = -168$$

$$\sum_{i=1}^{7} \left(x_i - \overline{x}\right)^4 = 26660$$

Συνεπώς έχουμε, πρώτα για τη διακύμανση

$$s^2 = \frac{\sum_{i=1}^{7}(x_i - \overline{x})^2}{6} = \frac{332}{6} = 55.33$$

άρα η τυπική απόκλιση είναι

$$s = \sqrt{s^2} = 7.44$$

Ο συντελεστής ασυμμετρίας είναι

$$\beta_1 = \frac{1}{n} \cdot \frac{\sum_{i=1}^{n}(x_i - \overline{x})^3}{s^3} = \frac{1}{7} \cdot \frac{(-168)}{(7.44)^3} = -0.06$$

- Τα δεδομένα είναι πιθανό να προέρχονται από κάποια συμμετρική κατανομή. Η μικρή αρνητική ασυμμετρία στο δείγμα φαίνεται να οφείλεται στην τυχαιότητα του δείγματος.

Ο συντελεστής κύρτωσης είναι

$$\beta_2 = \frac{1}{n} \cdot \frac{\sum_{i=1}^{n}(x_i - \overline{x})^4}{s^4} = \frac{1}{7} \cdot \frac{26660}{(7.44)^4} = 1.244$$

Τα δεδομένα προέρχονται από μια πλατύκυρτη κατανομή.

Οι συντελεστές ασυμμετρίας και κύρτωσης χρησιμοποιούν τα αθροίσματα

$$\sum_{i=1}^{n}(x_i - \overline{x})^3 \quad , \quad \sum_{i=1}^{n}(x_i - \overline{x})^4$$

Τα αθροίσματα αυτά είναι παραδείγματα ροπών για ένα σύνολο δεδομένων.

Ροπές κατανομής συχνοτήτων

Ο όρος ροπή έχει εισαχθεί από τη Μηχανική, όπου χρησιμοποιείται για τη μέτρηση της τάσης μιας δύναμης να παράγει περιστροφή.

Η δύναμη της τάσης αυτής εξαρτάται τόσο από το μέγεθος της δύναμης όσο και από την απόσταση από την αρχή του σημείου πάνω στο οποίο ασκείται η δύναμη. Έτσι, αν υποθέσουμε ότι έχουμε αναρτήσει μία ομοιογενή σιδερένια ράβδο από το κέντρο βάρους της, αυτή θα ισορροπεί. Αν τώρα, σε απόσταση δύο μονάδων μήκους, αναρτήσουμε βάρος π.χ 5kgr, θα αναπτυχθεί ροπή $2 \cdot 5 = 10$ μονάδων, η οποία θα τείνει να περιστρέφει τη ράβδο γύρω από το σημείο εξάρτησής της. Για να ισορροπήσει, θα πρέπει να αναπτυχθεί μια ροπή ίση και αντίθετη με την πρώτη.

Αυτό θα γίνει αν, σε απόσταση -1 μονάδας μήκους, τοποθετήσουμε βάρος 10kgr, οπότε θα αναπτυχθεί ροπή $-1 \cdot 10 = -10$ μονάδες.

Ανάλογη είναι η σημασία της ροπής και στη Στατιστική.

Το ιστόγραμμα μιας κατανομής μπορεί να θεωρηθεί ως κάτι το στέρεο και ότι κάθε στήλη του ασκεί πίεση πάνω στον άξονα των τετμημένων ίση με την αντίστοιχη της στήλης συχνότητας.

Η ροπή κάθε στήλης μετριέται από το γινόμενο της συχνότητας της τάξης και της αντίστοιχης απόστασης από το σημείο της μεταβλητής x_i , το οποίο πήραμε ως αρχή.

Ορισμός:Γενικά η ροπή τάξης *r* γύρω από το σημείο *x*0 μιας κατανομής είναι το άθροισμα

$$\frac{\sum_{i=1}^{n} (x_i - x_0)^r}{n}$$

Δύο ειδικές περιπτώσεις είναι:

Α. x0 = 0, τότε έχουμε ροπές γύρω από την αρχή (γύρω από το μηδέν) − η ροπή τάξης *r* είναι

$$v_r = \frac{\sum_{i=1}^{n} x_i^{\,r}}{n}$$

Η κατανομή που μας δόθηκε δεν παρουσιάζει συχνότητες:

Για r = 1, έχουμε: $v_1 = \dfrac{\sum_{i=1}^{n} x_i}{n}$ − ο αριθμητικός μέσος

Για r = 2, έχουμε: $v_2 = \dfrac{\sum_{i=1}^{n} x_i^{\,2}}{n}$

Η κατανομή παρουσιάζει συχνότητες:

Για r=1, έχουμε:

$$v_1 = \frac{\sum f_i x_i}{\sum f_i}$$

Για r=2, έχουμε:

$$v_2 = \frac{\sum f_i x^2{}_i}{\sum f_i}$$

Για r=3, έχουμε:

$$v_3 = \frac{\sum f_i x^3{}_i}{\sum f_i}$$

Για r=t, έχουμε: $v_t = \dfrac{\sum f_i x^t{}_i}{\sum f_i}$

Β. x0 = x, οπότε έχουμε την κεντρική ροπή (ροπή γύρω από το μέσο) τάξης r,

$$\mu_r = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \overline{x})^r$$

Αν η κατανομή δεν παρουσιάζει συχνότητες:

Για r = 1 παίρνουμε:

$$\mu_1 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \overline{x}) = \frac{1}{n} \cdot \left(\sum_{i=1}^{n} x_i - n\overline{x} \right) = 0$$

Για r = 2 παίρνουμε:

$$\mu_2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \overline{x})^2$$

Για r = 3 παίρνουμε:

$$\mu_3 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \overline{x})^3$$

Κτλ…

Αν η κατανομή παρουσιάζει συχνότητες:

Για r=1, έχουμε:

$$\mu_1 = \frac{\sum f_i \;\; x_i - \mu}{\sum f_i}$$

Για r=2, έχουμε:

$$\mu_2 = \frac{\sum f_i \;\; x_i - \mu \;\; ^{2}}{\sum f_i}$$

Για r=3, έχουμε:

$$\mu_3 = \frac{\sum f_i \;\; x_i - \mu \;\; ^{3}}{\sum f_i}$$

Για r=t, έχουμε:

$$\mu_t = \frac{\sum f_i \;\; x_i - \mu \;\; ^{t}}{\sum f_i}$$

Για τον πρακτικό υπολογισμό των ροπών, κάνουμε αλλαγή των τιμών x_iμε τις:

$$\xi_i = \frac{x_i - \mu}{\delta}$$

Οπότε αυτές εκφράζονται σε μονάδες τάξης διαστήματος δ.

$$\nu_1^{'} = \frac{\sum f_i \xi_i}{\sum f_i}$$

$$\nu_2^{'} = \frac{\sum f_i \xi^2{}_i}{\sum f_i}$$

8.ΒΑΣΙΚΕΣ ΑΡΧΕΣ ΤΗΣ ΘΕΩΡΙΑΣ ΤΩΝ ΠΙΘΑΝΟΤΗΤΩΝ

8.1 ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ :ΤΥΧΑΙΑ ΠΕΙΡΑΜΑΤΑ-ΔΕΙΓΜΑΤΟΧΩΡΟΙ-ΕΝΔΕΧΟΜΕΝΑ

Η έννοια του τυχαίου πειράματος είναι βασική στη θεωρία των πιθανοτήτων. Το τυχαίο πείραμα θα πρέπει να το αντιδιαστείλουμε μ’ αυτό που συνήθως εννοούμε σαν πείραμα. Διαισθητικά σαν πείραμα εννοούμε μια ενέργεια που εκτελείται κάτω από ορισμένες συνθήκες, μπορεί να αναλυθεί οσεσδήποτε φορές κάτω από τις ίδιες πάντοτε συνθήκες και μετά τη συμπλήρωση της παρατηρούνται κάποια αποτελέσματα. Έτσι 300 ευρώ κατατιθέμενα στην τράπεζα με επιτόκιο 100% που κεφαλαιοποιούνται ετήσια, γίνονται $100(1+α)^n$ ευρώ σε n χρόνια. Επίσης αν σπρώξουμε ένα αντικείμενο, αυτό θα αποκτήσει κάποια επιτάχυνση. Πειράματα όπου οι γνώσεις των συνθηκών που εκτελούνται μας επιτρέπουν να γνωρίζουμε πλήρως το αποτέλεσμα, ονομάζονται αιτιοκρατικά. Ενώ σ’ αντίθεση σαν τυχαία πειράματα (τ.π.) θα χαρακτηρίζονται εκείνα όπου οι γνώσεις των συνθηκών που εκτελούνται, μας καθορίζουν ένα σύνολο δυνατών αποτελεσμάτων για κάθε πείραμα. Τέτοια μπορεί να είναι:

1. Όταν παρατηρούμε το εισόδημα των εργαζομένων σε μια περιοχή.

2. Η τιμή μιας μετοχής κατά μια ορισμένη χρονική διάρκεια.

3. Η προτίμηση ενός αγοραστή για ένα προϊόν.

4. Η καταγραφή των μηνιαίων εξόδων ενός άτομου για τροφή ή ψυχαγωγία κ.λπ.

5. Η ρίψη ενός νομίσματος.

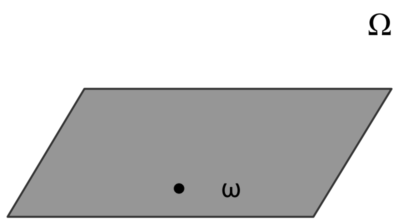
6. Οι αριθμοί του Λόττο που κληρώνονται κάθε Σάββατο.

7.Η καταγραφή των ελαττωματικών αντικειμένων που παράγονται από μια βιομηχανική μονάδα.

8.Το πλήθος των ατόμων που φθάνουν σ’ ένα κατάστημα κ.λπ.

9.Η ζήτηση ενός τύπου πορτοκαλάδας ανά εβδομάδα κ.λπ.

Το σύνολο όλων των δυνατών αποτελεσμάτων ενός τυχαίου πειράματος θα το ονομάζουμε Δειγματοχώρο (‘ή πληθυσμό)και θα συμβολίζεται με Ω. Τα στοιχειά του Ω θα τα ονομάζουμε δειγματοσημεία και γενικά ένα τέτοιο θα συμβολίζεται με το γράμμα ω.Καθε ω είναι το αποτέλεσμα ενός τυχαίου πειράματος που εκτελέστηκε.



Παράδειγμα 8.1.1

Να βρεθούν οι Δειγματοχώροι των προηγούμενων τυχαίων πειραμάτων.

1)Εδώ βασικά είναι ο Ω=[m,M], όπου m το ελάχιστο και εισόδημα που μπορεί να παρατηρηθεί σ’ αυτή την περιοχή, και M το μέγιστο.

2)Εδώ, Ω=[0,M], όπου το 0 θα σημαίνει ότι πτώχευσε η εταιρία που έδωσε αυτές τις μετοχές, M η μεγίστη τιμή που μπορεί να πάρει αυτή η μετοχή.

3)Εδώ, Ω={π,α}, όπου π σημαίνει προτιμώ το εν λόγω προϊόν, ενώ α απορρίπτω το εν λόγω προϊόν. Σ’ αυτό το παράδειγμα μπορούμε να σημειώσουμε ότι ένας Δειγματοχώρος μπορεί να παρασταθεί με διαφόρους τρόπους. Έτσι αν σημειώσουμε με 1 όταν ο αγοραστής προτίμα το εν λόγω προϊόν, και όταν το απορρίπτει με 0,ο Δειγματοχώρος γράφεται Ω={0,1}.

4) π.χ. μπορεί να είναι από 300 ευρώ έως 440.

5)Εδώ , Ω={Κ,Γ},όπου Κ σημαίνει έρχεται κορώνα, ενώ Γ έρχεται γράμμα.

Όπως στο 3)’ μπορούμε να αντιστοιχίσουμε στο Κ το 0,και στο Γ το 1, και να πάρουμε S={0,1}.

6)Εδώ Ω=’’το πλήθος όλων των εξάδων που σχηματίζονται από το 1 έως το 49’’={(χ1,…,χ6)’:όπου χ1,…χ6 αριθμοί από το 1 έως το 49}.

7)Ανάλογο του 3)’ και του 5)’.

8)Εδώ , Ω={0,1,2,3,…,M},όπου,0, σημαίνει κανένας δεν μπαίνει στο κατάστημα ,1, ένα άτομο μπαίνει στο κατάστημα.

9)Εδώ , Ω={0,1,2,3,…,M}, όπου,0,σημαίνει δεν ζητήθηκε καμία πορτοκαλάδα,1, ζητήθηκε μια πορτοκαλάδα κ.λπ.

Κάθε υποσύνολο του Δειγματικού χώρου Ω θα ονομάζεται **ενδεχόμενο,** και θα συμβολίζεται με κεφαλαία γράμματα. Τα δειγματοσημεία από μονά τους, θεωρούνται ενδεχόμενα και μάλιστα θα καλούνται **απλά.** Πολλά δειγματοσημεία μαζί μας δίνουν τα **σύνθετα ενδεχόμενα**.

Παράδειγμα 8.1.2

Να βρεθεί ο Δειγματοχώρος κατά τη ρίψη 3 νομισμάτων. Ποιό είναι το ενδεχόμενο ότι εμφανίζεται τουλάχιστον μια κορώνα (από μια και πάνω);Ποιό είναι το ενδεχόμενο ότι εμφανίζονται 2 κορώνες και ένα γραμμα; Ποιό είναι το ενδεχόμενο ότι εμφανίζονται 3 κορώνες;

Απάντηση:

Εδώ , Ω={(Κ,Κ,Κ),(Κ,Κ,Γ),(Κ,Γ,Κ,),(Γ,Κ,Κ),(Κ,Γ,Γ),(Γ,Κ,Γ),(Γ,Γ,Κ),(Γ,Γ,Γ)}.

Αν συμβολίσουμε το ενδεχόμενο ακριβώς 2 κορώνες με Α τότε

Α={(Κ,Κ,Γ),(Κ,Γ,Κ),(Γ,Κ,Κ)}

Ενώ αν Β συμβολίζει το ενδεχόμενο τουλάχιστον μια κορώνα τότε,

Β={(Κ,Κ,Κ),(Κ,Κ,Γ),(Κ,Γ,Κ,),(Γ,Κ,Κ),(Κ,Γ,Γ),(Γ,Κ,Γ),(Γ,Γ,Κ)}, επίσης αν το ενδεχόμενο εμφανίζονται 2 κορώνες και ένα γράμμα συμβολίζεται με Δ, τότε

Δ={(Κ,Κ,Γ),(Κ,Γ,Κ),(Γ,Κ,Κ)},και τέλος με

Ε={(Κ,Κ,Κ)} εμφανίζονται 3 κορώνες.

Το τελευταίο ενδεχόμενο είναι απλό, ενώ τα άλλα σύνθετα.

Θα λέμε ότι ένα **ενδεχόμενο Α συμβαίνει** αν η εκτέλεση του υπό συζήτηση τυχαίου πειράματος καταλήξει σε κάποιο δειγματοσημείο ω έτσι ώστε το ω να βρίσκεται στο Α. Στο προηγούμενο παράδειγμα θα λέμε ότι συμβαίνει το Α, αν π.χ. ω=(Κ,Κ,Γ), ή ω=(Γ,Κ,Κ) κ.λπ.

Θα λέμε ότι ένα **ενδεχόμενο Α δεν συμβαίνει** αν η εκτέλεση του υπό συζήτηση τυχαίου πειράματος καταλήξει σε κάποιο ω που δεν βρίσκεται στο Α. Έτσι αν π.χ. ω=(Γ,Γ,Γ),στο προηγούμενο παράδειγμα κανένα από τα ενδεχόμενα Α,Β,Γ,Δ,Ε δεν συμβαίνει. Αυτό που συμβαίνει στην προκειμένη περίπτωση, είναι ότι πραγματοποιούνται τα συμπληρωματικά ενδεχόμενα των Α,Β,Γ,Δ,Ε που αντίστοιχα είναι:

Α'={(Κ,Κ,Κ),(Κ,Γ,Γ),(Γ,Κ,Γ),(Γ,Γ,Κ),(Γ,Γ,Γ)},

Β'={(Γ,Γ,Γ)},

Δ'={(Κ,Κ,Κ),(Κ,Γ,Γ),(Γ,Κ,Γ),(Γ,Γ,Κ),(Γ,Γ,Γ)},

Ε'={(Κ,Κ,Γ),(Κ,Γ,Κ),(Γ,Κ,Κ),(Κ,Γ,Γ),(Γ,Κ,Γ),(Γ,Γ,Κ),(Γ,Γ,Γ)}.

Το ενδεχόμενο'' εμφανίζονται 4 γράμματα'' δεν πραγματοποιείται ποτέ, αν εκτελέσουμε το τυχαίο πείραμα του προηγούμενου παραδείγματος. Ένα τέτοιο ενδεχόμενο θα το ονομάζουμε **αδύνατο,** και θα συμβολίζεται με το Ø (σύμβολο κενού συνόλου), ενώ το ενδεχόμενο'' εμφανίζονται οι τρεις όψεις τριών νομισμάτων'' θα συμβαίνει πάντοτε και είναι το **βέβαιο ενδεχόμενο** που στο εξής θα συμβολίζεται όπως ο Δειγματοχώρος Ω.

Γενικά για τα ενδεχόμενα ισχύουν όλες οι πράξεις που έχουμε στα σύνολα. Ο παρακάτω πίνακας δίνει αναλυτικά αυτές τις ιδιότητες.

Ενδεχόμενα	Λεκτική περιγραφή	Συμβολισμοί με σύνολα
1.Το ενδεχόμενο Α συμβαίνει σίγουρα.	Α είναι αβέβαιο ενδεχόμενο.	Α=Ω
2.Το Α δεν συμβαίνει ποτέ.	Α είναι αδύνατο ενδεχόμενο.	Α=Ø
3.Οταν συμβαίνει το Α, συμβαίνει και το Β.	Α είναι μερικό ενδεχόμενο του Β.	$A \subset B$
4.Οταν συμβαίνει το Α, συμβαίνει και το Β, και αντίστροφα.	Α και Β είναι ισοδύναμα.	Α=Β
5.Οταν συμβαίνει το Α, δεν συμβαίνει και το Β.	Α και Β είναι ξένα.	Α∩Β=Ø
6.Οταν δεν συμβαίνει το Α αλλά συμβαίνει το Β.	Α και Β είναι συμπληρωματικά.	$B=A'$
7.Συμβαίνει το Α όταν τουλάχιστον ένα από τα ενδεχόμενα Α1,Α2, συμβαίνει(συμβαίνει το Α1 ή το Α2 ή το Α3 …).	Το Α είναι η ένωση των Α1,Α2,…	$A= A_1 \cup A_2 \cup A_3 \ldots$
8.Συμβαίνει το Α όταν όλα τα ενδεχόμενα Α1,Α2,συμβαίνουν(συμβαίνει και το Α1 και το Α2 και το Α3…).	Το Α είναι η τομή των Α1,Α2,…	$A= A_1 \cap A_2 \cap A_3 \cap \ldots$

8.2 ΟΡΙΣΜΟΣ ΤΗΣ ΠΙΘΑΝΟΤΗΤΑΣ

Όπως και στο προηγούμενο εδάφιο όταν εκτελείται ένα τυχαίο πείραμα δεν γνωρίζουμε επακριβώς το αποτέλεσμα αυτού, παρά ένα σύνολο δυνατών αποτελεσμάτων . Είναι φυσικό επόμενο,λοιπον, εφόσον έχουμε διάφορα ενδεχόμενα σ' αυτό να ρωτήσουμε με ποια βεβαιότητα μπορεί να συμβεί αυτό. Κάθε ενδεχόμενο, φυσικά, μπορεί να μην έχει τον αυτό βαθμό βεβαιότητας για να πραγματοποιηθεί. Θα προσπαθήσουμε να μετρήσουμε τον βαθμό βεβαιότητας κάθε ενδεχόμενου. Αυτό θα γίνει με τη βοήθεια μιας νέας έννοιας που θα εισάγουμε, ''την πιθανότητα '' να συμβεί ένα ενδεχόμενο, η οποία θα μας δίνει το ποσοστό πραγματοποίησης αυτού στο υπό συζήτηση τυχαίο πείραμα.

ΚΛΑΣΙΚΟΣ ΟΡΙΣΜΟΣ ΤΗΣ ΠΙΘΑΝΟΤΗΤΑΣ

Θεωρούμε ότι ο Δειγματοχώρος μας έχει πεπερασμένο το πλήθος Δειγματοσημεία,και ότι το κάθε ένα από αυτά έχει τον αυτό βαθμό βεβαιότητας για να πραγματοποιηθεί. Π.χ. κατά τη ρίψη ενός τίμιου ζαριού, ο Δειγματοχώρος είναι

Ω={1,2,3,4,5,6}, και εφόσον το ζάρι είναι τίμιο κάθε όψη του ζαριού για να εμφανιστεί έχει τον αυτό βαθμό βεβαιότητας.Σε τέτοιες περιπτώσεις **η πιθανότηταενός ενδεχόμενου Α** θα είναι ένας αριθμός, που θα συμβολίζεται με P(A), και θα ορίζεται ως εξής:

$$P(A)=\frac{\text{ευνοϊκές περιπτώσεις για να πραγματοποιηθεί το Α}}{\text{σύνολο όλων των δυνατών αποτελεσμάτων}}.$$

Έτσι στην περίπτωση της ρίψης ενός ζαριού αν θέλουμε να βρούμε την πιθανότητα του ενδεχόμενου Α= να έρθει ζυγός αριθμός ={2,4,6}, έχουμε

$$P(A)=\frac{3}{6}=\frac{1}{2}.$$

Παράδειγμα 8.2.1

Ο διευθυντής μιας εταιρείας θέλει να προσλάβει σ'αυτή δυο άτομα,απο τα τέσσερα που έχουν κάνει αίτηση. Αν θεωρήσουμε ότι δεν είναι σε θέση να αποφασίσει σωστά,και διαλέγει στην τύχη τα δύο άτομα,τοτε

Α) ποιά είναι η πιθανότητα να έχουν επιλεγεί οι δυο καλύτεροι;

Β) ποιά είναι η πιθανότητα ότι επιλέγεται τουλάχιστον μια φορά ο καλύτερος;

Λύση:Το τυχαίο πείραμα είναι να επιλεγούν οι δυο καλύτεροι από τους τέσσερις. Θεωρούμε ότι τα άτομα είναι διαφορετικής ικανότητας, και συμβολίζουμε τον καλύτερο με 1, τον επόμενο καλύτερο με 2, το μεθεπόμενο καλύτερο με 3, και το άτομο με τις λιγότερες ικανότητες με 4. Τότε ο Δειγματοχώρος μας είναι

Ω={(1,2), (1,3), (1,4), (2,3), (2,4), (3,4)} ή Ω={$\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6$},

Όπου το ω_1=(1,2), σ_I ω_2=(1,3),κλπ. Κάθε δειγματοσημείο για να πραγματοποιηθεί έχει πιθανότητα $\frac{1}{6}$.Συνοψίζοντας έχουμε τον ακόλουθο πίνακα.

Δειγματοσημεία	**Ζευγάρι**	**Πιθανότητα**
ω_1	(1,2)	$\frac{1}{6}$
ω_2	(1,3)	$\frac{1}{6}$
ω_3	(1,4)	$\frac{1}{6}$
ω_4	(2,3)	$\frac{1}{6}$
ω_5	(2,4)	$\frac{1}{6}$
ω_6	(3,4)	$\frac{1}{6}$

Α) Α= επιλέγονται οι δυο καλύτεροι = $\{\omega_1\}$, Ρ(Α)=ρ($\{\omega_1\}$) = $\frac{1}{6}$

Β) Β= επιλέγεται τουλάχιστον μια φορά ο ένας καλύτερος

=$\{\omega_1\} \cup \{\omega_2\} \cup \{\omega_3\} \cup \{\omega_4\} \cup \{\omega_5\}$, και Ρ(Β)= $\frac{5}{6}$.

Αυτός ο ορισμός όμως δεν μπορεί να εφαρμοστεί όταν ο Δειγματοχώρος δεν είναι πεπερασμένος, και τα σημεία δεν έχουν τον αυτό βαθμό βεβαιότητας .Π.χ αν το ζάρι είναι δουλευμένο για να φέρνει συχνότερα {6}. Αυτή η παρατήρηση και με το γεγονός ότι ένα τυχαίο πείραμα επαναλαμβάνεται οσεσδήποτε φορές κάτω από τις ίδιες συνθήκες, οδηγούν στον επόμενο ορισμό.

ΟΡΙΣΜΟΣ ΤΗΣ ΠΙΘΑΝΟΤΗΤΑΣ ΣΑΝ ΣΧΕΤΙΚΗ ΣΥΧΝΟΤΗΤΑ

Όταν επαναλαμβάνουμε ένα τυχαίο πείραμα n- φορές, τότε η σχετική συχνότητα του ενδεχομένου Α για να συμβεί, ορίζεται από τη σχέση

f_n(Α) = $\frac{1}{n}$(πλήθος εμφανίσεων του Α)

Όταν το n αυξάνει (τείνει στο άπειρο) αναμένει κανείς ότι η αριθμητική ακολουθία f_n(Α) σταθεροποιείται οριακά στο Ρ(Α), δηλ.

f_n(Α) τείνει στο Ρ(Α), όταν n τείνει στο άπειρο (n→∞).

Ιδιότητες της Σχετικής Συχνότητας

1. 0≤) f_n(Α) ≤1,για κάθε ενδεχόμενο Α.
2. f_n(Α∪Β) = f_n(Α) + f_n(Β) , αν Α και Β είναι ξένα μεταξύ τους.
3. f_n(Α)′=1- f_n(Α), Α′ το συμπληρωματικό του Α.
4. f_n(Α) ≤ f_n(Β) , αν Α⊂Β.

Παράδειγμα 8.2.2.

Οι αρμόδιες αρχές μιας χώρας διαπίστωσαν ότι από τα 400 εισαγόμενα ζώα, τα 50 ήταν θετικά στο test K_1 ότι περιέχουν ορμόνες, 40 ήταν θετικά στο test K_2 ότι περιέχουν αντιβιοτικά, ενώ 20 βρέθηκαν θετικά και στα δυο test. Να υπολογιστούν οι πιθανότητες των ενδεχομένων A_i:’’ το αποτέλεσμα είναι θετικό στο test K_i ’’, i=1,2.

Απάντηση:

$$P(A_1)=\frac{50}{400}=\frac{1}{80},P(A_2)=\frac{40}{400}=\frac{1}{10},P(A_1\cap A_2)=\frac{20}{400}=\frac{1}{20}$$

Γενικά όμως και αυτός ορισμός έχει αδυναμίες, ιδιαίτερα όταν κανείς θέλει να ασχοληθεί με την ένωση ενδεχομένων που είναι στο πλήθος μη πεπερασμένος. Γι’αυτό το λόγο αναζητήθηκε ένας πιο ισχυρός ορισμός της πιθανότητας, που είναι ο ακολουθος:

ΑΞΙΩΜΑΤΙΚΟΣ ΟΡΙΣΜΟΣ ΤΗΣ ΠΙΘΑΝΟΤΗΤΑΣ

Ο αριθμός P(A) που μετρά το βαθμό βεβαιότητας του ενδεχομένου A, δηλ. η πιθανότητα του A,είναι θετικός αριθμος,για το σχεδόν βέβαιο ενδεχόμενο Ω γίνεται 1, ενώ η πιθανότητα της ένωσης ξένων μεταξύ των ενδεχομένων ισούται με το άθροισμα των πιθανοτήτων αυτών. Έτσι έχουμε:

1.P(A)≥0, για κάθε A.

2.P(Ω)=1.

3.P($A_1\cup A_2\cup A_3\ldots$)= $P(A_1)+P(A_2)+P(A_3)+\ldots$

Όταν τα ενδεχόμενα είναι ξένα μεταξύ τους ανά δυο, δηλ.

$A_i\cap A_J=\varnothing$, για i≠j, και I,j=1,2,3,……..

Οι παραπάνω προϋποθέσεις που θέσαμε για τον ορισμό της πιθανότητας αποτελούν τα λεγόμενα Αξιώματα των Πιθανοτήτων.Σε κάθε Δειγματοχώρο Ω ενός τυχαίου πειράματος, μπορούμε να ορίσουμε μια συνάρτηση P, που για κάθε ενδεχόμενο A να μας δίνει την πιθανότητα P(A) έτσι ώστε τα τρία αξιώματα να ικανοποιούνται.Η συνάρτηση P(·) ονομάζεται μέτρο πιθανότητας, και έχει όλες τις ιδιότητες άλλων μέτρων, όπως του μήκους ℓ ενός διαστήματος [α,β] για το οποίο έχουμε ℓ

([α,β])=β-α. Π.χ. το μήκος ℓ του[5,7] είναι l ([5,7])=7-5=2 κ.λ.π.

ΚΑΝΟΝΕΣ ΠΙΘΑΝΟΤΗΤΩΝ

Θεωρούμε τα ενδεχόμενα A,B, A_1, A_2, A_3,….. του Ω, και την πιθανότητα P στον Ω.

Τότε ισχύουν:

1. 0≤P(A)≤1,Κάθε ενδεχόμενο πραγματοποιείται με πιθανότητα το πολύ 1.
2. P(Ø)=0, Η πιθανότητα του αδύνατου ενδεχομένου είναι 0.
3. Αν $A \subset B$, Τότε και P(A)≤P(B).
4. P (A′) =1-P (A), Η πιθανότητα του συμπηρωματικού ενός ενδεχομένου A ισούται με 1 μείον την πιθανότητα του A.

5α.P $(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$.

5β.P $(A_1 \cup A_2 \cup A_3) =$

$$P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$$

5γ. $$\begin{aligned} &P(A_1 \cup A_2 \cup ... \cup A_n) = \\ &P(A_1) + P(A'_1 \cap A_2) + P(A'_1 \cap A'_2 \cap A'_3) + ... + P(A'_1 \cap A'_{n-1} \cap A'_n) \end{aligned}$$

6. Αν τα ενδεχόμενα $A_1, A_2, A_3, ...$ αποτελούν μια διαμέριση του Ω τότε για κάθε άλλο ενδεχόμενο B του Ω, ισχύει P(B)=$P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) + ...$

8.3 ΒΑΣΙΚΕΣ ΑΡΧΕΣ ΑΠΑΡΙΘΜΗΣΗΣ

Έως τώρα οι Δειγματοχώροι που θεωρούσαμε ήταν αρκετά απλοί, και οι περιπτώσεις απαρίθμησης των δειγματοσημείων μας επιτυγχάνονται κατά έναν απλοϊκό τρόπο. Σε πολλές όμως περιπτώσεις η απαρίθμηση των δειγματοσημείων στο Δειγματοχώρο ή και στα διάφορα ενδεχόμενα γίνεται κατόπιν χρήσης κατάλληλων μεθόδων της συνδυαστικής ανάλυσης.

ΑΘΡΟΙΣΤΙΚΟΣ ΚΑΝΟΝΑΣ ΑΠΑΡΙΘΜΗΣΗΣ

Αν υπάρχουν n_1 τρόποι εκτέλεσης του Π_1 έργου, n_2 του Π_2 έργου,…και n_κ του Π_κ έργου, και εφόσον αυτοί είναι διαφορετικοί μεταξύ τους, τότε ο αριθμός των τρόπων εκτέλεσης του Π_1 ή Π_2 ….ή Π_κ είναι $n_1 + ... + n_k$

ΠΟΛΛΑΠΛΑΣΙΑΣΤΙΚΟΣ ΚΑΝΟΝΑΣ ΑΠΑΡΙΘΜΗΣΗΣ

Αν ένα έργο Π αποτελείται από κ-υποέργα $\Pi_1, \Pi_2, \ldots, \Pi_\kappa$ και κάθε υποέργο μπορεί να εκτελεστεί κατά $n_1, n_2, \ldots, n_\kappa$ τρόπους, τότε το έργο Π εχει $n_1 * n_2 * \ldots * n_\kappa$ τρόπους εκτέλεσης.

Παράδειγμα 8.3.1

α) Πόσα δειγματοσημεία λαμβάνουμε όταν ρίχνουμε 3 νομίσματα ή 4 ζάρια;

β)Ένας οδηγός για να πάει από την πόλη Α στη Β εχει 3 δυνατότητες, ενώ από τη Β στη Γ έχουν 4, και από τη Γ στη Δ έχουν 3. Πόσες δυνατότητες εχει για να πάει από την Α στη Δ;

γ)Αν σε ένα κατάστημα ρούχων υπάρχουν 3 κόκκινα πουκάμισα, 2 άσπρα και 4 γαλάζια, με πόσους τρόπους μπορούμε να αγοράσουμε ένα;

Απάντηση:

α)2*2*2=8 και 6*6*6*6=216

β)3*4*3=36

γ)3+2+4=9

Σ’αυτό το σημείο θεωρούμε ότι κάθε σημείο του Δειγματοχώρου μας αντιστοιχεί σ’ένα σφαιρίδιο. Αν ο Δειγματοχώρος εχει n σημεία, τότε θεωρούμε ένα δοχείο που να περιέχει n σφαιρίδια αριθμημένα από το 1 έως το n. Ο σχηματισμός ενός **δείγματος** μεγέθους m θα αντιστοιχεί με την ανάσυρση m σφαιριδιων από το δοχείο μας. Το δείγμα μας μπορεί να μετασχηματιστεί με δυο τρόπους, με **επανατοποθετήσεις ή χωρίς επανατοποθετήσεις**. Επίσης έχουμε τις εξής δυνατότητες : να μας ενδιαφέρει η σειρά που τα σφαιρίδια εμφανίζονται, οπότε μιλάμε για **διατεταγμένα δείγματα**, ή να μην μας ενδιαφέρει η σειρά που τα σφαιρίδια εμφανίζονται, οπότε μιλάμε για **μη διατεταγμένα δείγματα**.

Διατάξεις

Χωρίς επανατοποθετήσεις: Οι τρόποι για να ανασύρουμε m σφαιρίδια από τα n, όταν μας ενδιαφέρει η σειρά εμφάνισης αυτών και με χωρίς επανατοποθετήσεις είναι

$$\Delta_m^n = n(n-1)...(n-m+1) = \frac{n!}{(n-m)!}$$

Όπου n!=n(n-1)…3.2.1 και 0!=1.

Με επανατοποθετήσεις: Οι τρόποι για να ανασύρουμε m σφαιρίδια από τα n, όταν μας ενδιαφέρει η σειρά εμφάνισης αυτών και με επανατοποθετήσεις είναι n^m.

Συνδυασμοί

Χωρίς επανατοποθετήσεις: Οι τρόποι για να ανασύρουμε m σφαιρίδια από τα n, όταν δεν μας ενδιαφέρει η σειρά εμφάνισης αυτών και με χωρίς επανατοποθετήσεις είναι

$$\binom{n}{m} = \frac{n!}{(n-m)!m!}, (\binom{n}{m} = \Delta_m^n \frac{1}{m!})$$

Με επανατοποθετήσεις: Οι τρόποι για να ανασύρουμε m σφαιρίδια από τα n, όταν δεν μας ενδιαφέρει η σειρά εμφάνισης αυτών και με επανατοποθετήσεις είναι

$$\binom{n+m-1}{m}.$$

8.4 ΔΕΣΜΕΥΜΕΝΕΣ ΠΙΘΑΝΟΤΗΤΕΣ

Συχνά παρουσιάζει ενδιαφέρον πως ορισμένα ενδεχόμενα σχετίζονται με άλλα,ειδικότερα μας ενδιαφέρει να υπολογίσουμε την πιθανότητα όπως συμβεί ένα ενδεχόμενο όταν γνωρίζουμε ότι συμβαίνει κάποιο άλλο. Σ'αυτή την περίπτωση θα μιλάμε για την δεσμευμένη πιθανότητα του ενδεχομένου, τότε η δεσμευμένη πιθανότητα του Α όταν δίνεται ότι συμβαίνει το Β, ορίζεται από την

P(A/B)= $\frac{P(A\cap B)}{P(B)}$, αν P(B)>0.

Παράδειγμα 8.4.1.

Ο παρακάτω πίνακας δείχνει την προτίμηση ενός καταναλωτή ανάλογα με το φύλο του ως προς δυο προϊόντα Π_1, Π_2 .

Φύλο	Π_1	Π_2	Σύνολο
A_1 άνδρες	200	300	500
A_2 γυναίκες	100	400	500
	300	700	1000

Ποιά είναι η πιθανότητα ότι προτιμάται το προϊόν Π_1;

Ποιά είναι η πιθανότητα για έναν άνδρα καταναλωτή να προτίμα το προϊόν Π_1;

Απάντηση:

Συμφωνά με τον κανόνα 6,

$\mathrm{P}(\Pi_1) = \mathrm{P}(\Pi_1 \cap \mathrm{A}_1) + \mathrm{P}(\Pi_1 \cap \mathrm{A}_2) = 0,20 + 0,10 = 0,30$.

Ενώ $\mathrm{P}(\mathrm{A}_1 / \Pi_1) = \dfrac{\mathrm{P}(\mathrm{A}_1 \cap \Pi_1)}{\mathrm{P}(\Pi_1 \cap \mathrm{A}_1) + \mathrm{P}(\Pi_2 \cap \mathrm{A}_1)} = \dfrac{0,20}{0,20+0,30} = \dfrac{0,20}{0,50} = 0,4$

8.5 ΠΟΛΛΑΠΛΑΣΙΑΣΤΙΚΟΣ ΚΑΝΟΝΑΣ ΠΙΘΑΝΟΤΗΤΩΝ

Πολλές φόρες όταν θέλουμε να βρούμε την πιθανότητα για να συμβαίνουν συγχρόνως δυο ενδεχόμενα Α και Β, χρησιμοποιούμε κατάλληλα τη δεσμευμένη πιθανότητα του ενός όταν δίνεται ότι συμβαίνει το άλλο.Ετσι

$\mathrm{P}(\mathrm{A} \cap \mathrm{B}) = \mathrm{P}(\mathrm{A}) * \mathrm{P}(\mathrm{A} / \mathrm{B})$.

Γενικά αν έχουμε τα ενδεχόμενα $\mathrm{A}_1, ..., \mathrm{A}_n \subset \Omega$,τότε

$\mathrm{P}(\mathrm{A}_1 \cap \mathrm{A}_{2...} \cap \mathrm{A}_n) = \mathrm{P}(\mathrm{A}_1) * \mathrm{P}(\mathrm{A}_2 / \mathrm{A}_1) * \mathrm{P}(\mathrm{A}_3 / \mathrm{A}_1 \cap \mathrm{A}_2) ... \mathrm{P}(\mathrm{A}_n / \mathrm{A}_1 \cap ... \cap \mathrm{A}_{n-1})$,

Αν $\mathrm{P}(\mathrm{A}_1 \cap \mathrm{A}_{2...} \cap \mathrm{A}_{n-1}) > 0$.

Παράδειγμα

Μια κάλπη περιέχει 10 σφαιρίδια, από τα οποία 6 είναι άσπρα και 4 μαύρα. Εξάγουμε ένα σφαιρίδιο, χωρίς να το ξανατοποθετήσουμε στην κάλπη. Επίσης εξάγουμε και ένα δεύτερο. Ποιά είναι η πιθανότητα το πρώτο σφαιρίδιο να είναι άσπρο και το δεύτερο μαύρο;

Λύση:

Ονομάζουμε Α το ενδεχόμενο <<το πρώτο σφαιρίδιο να είναι άσπρο>> και Β το ενδεχόμενο <<το δεύτερο σφαιρίδιο να είναι μαύρο>>.

Η πιθανότητα <<το πρώτο σφαιρίδιο να είναι άσπρο>> είναι:

$$P \;\; A \; = \frac{6}{10}$$

Αφού το πρώτο σφαιρίδιο που έχει εξαχθεί και δεν έχει ξανατοποθετηθεί στην κάλπη είναι άσπρο, στην κάλπη θα παραμείνουν 9 σφαιρίδια, από τα οποία 5 θα είναι άσπρα και 4 μαύρα και, κατά συνέπεια, η πιθανότητα να εξαχθεί ένα μαύρο σφαιρίδιο όταν το πρώτο θα είναι άσπρο, θα είναι:

$$P \;\; B / A \; = \frac{4}{9}$$

Άρα, η πιθανότητα να εξαχθεί διαδοχικά ένα άσπρο και ένα μαύρο σφαιρίδιο θα είναι:

$$P\ A\cap B\ = P\ A\ P\ B\ = \frac{6}{10}.\frac{4}{9} = \frac{24}{90}$$

8.6 ΑΝΕΞΑΡΤΗΤΑ ΕΝΔΕΧΟΜΕΝΑ

Δυο ενδεχόμενα Α και Β θα είναι ανεξάρτητα, αν όταν συμβαίνει το ένα δεν μας δίνει καμία πληροφορία για το άλλο.Δηλ.

P(A/B)= P(A/B)= $\frac{P(A\cap B)}{P(B)}$ ή P(A∩B)=P(A)*P(B).

Ο παραπάνω ορισμός γενικεύεται και για τρία ενδεχόμενα Α,Β,Γ ή και περισσότερα. Μονό που σε αυτή την περίπτωση, π.χ για 3 ενδεχόμενα εξετάζουμε τις πιθανότητες των ενδεχομένων ανά 2, και ανά 3. Τα ενδεχόμενα Α,Β,Γ θα είναι ανεξάρτητα αν

P(A∩B)=P(A)P(B), P(A∩Γ)=P(A)P(Γ), P(B∩Γ)=P(B)P(Γ)

Και

P(A∩B∩Γ)=P(A)P(B)P(Γ).

Παράδειγμα

Δύο κάλπες έχουν 10 σφαιρίδια με τους αριθμούς 1,2,3,...,10. Εξάγεται από κάθε κάλπη ένα σφαιρίδιο και εξετάζουμε τα παρακάτω δεδομένα:

Α το ενδεχόμενο να εμφανιστεί από την πρώτη κάλπη ζυγός αριθμός.

Β το ενδεχόμενο να εμφανιστεί από τη δεύτερη κάλπη μονός αριθμός.

Γ το ενδεχόμενο το άθροισμα των δύο αριθμών που εξάχθηκαν να είναι ζυγός αριθμός.

Να εξεταστεί αν τα ενδεχόμενα αυτά είναι πλήρως ανεξάρτητα.

Λύση:

$$P\ A\ = \frac{1}{2},\ P\ B\ = \frac{1}{2},\ P\ \Gamma\ = \frac{1}{2}$$

$$P(A\cap B)=P(A)\cdot P(B)=\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}$$

$$P(A\cap \Gamma)=P(A)\cdot P(\Gamma)=\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}$$

$$P(B\cap \Gamma)=P(B)\cdot P(\Gamma)=\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}$$

Από τις παραπάνω σχέσεις συμπεραίνουμε ότι τα ενδεχόμενα *A,B* και *Γ* είναι ανά δύο ανεξάρτητα. Όμως, τα παραπάνω ενδεχόμενα δεν είναι πλήρως ανεξάρτητα, γιατι:

$$(A\cap B\cap \Gamma)=\varnothing \Rightarrow P(A\cap B\cap \Gamma)=0\neq \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=P(A)P(B)P(\Gamma)=P(\varnothing)$$

8.7 ΝΟΜΟΣ ΤΟΥ BAYES-ΥΠΟΚΕΙΜΕΝΙΚΕΣ ΠΙΘΑΝΟΤΗΤΕΣ

Όπως είδαμε στον κανόνα 6, αν $A_1, A_2, ...$ αποτελούν μια διαμέριση του Δειγματοχώρου Ω, τότε για κάθε άλλο ενδεχόμενο Β του Ω ισχύει:

$$\text{P(B)}=\sum_i \text{P}(\text{B}\cap \text{A}_i)=\text{P}(\text{B}\cap \text{A}_1)+\text{P}(\text{B}\cap \text{A}_2)+...$$

Η τελευταία σχέση με τη βοήθεια της δεσμευμένης πιθανότητας γράφεται ισοδύναμα

$$\text{P(B)}=\sum_i \text{P}(\text{A}_i)*\text{P}(\text{B}/\text{A}_i)=\text{P}(\text{A}_1)*\text{P}(\text{B}/\text{A}_1)+\text{P}(\text{A}_2)*\text{P}(\text{B}/\text{A}_2)+...$$

Αν $\text{P}(\text{A}_i)>0, i=1,2,\ldots$

Αν και επιπλέον είναι P(B)>0, τότε έχουμε τον τύπο του Bayes

$$\text{P}(\text{A}_i/\text{B})=\frac{\text{P}(\text{B}/\text{A}_i)*\text{P}(\text{A}_i)}{\sum_i \text{P}(\text{A}_i)*\text{P}(\text{B}/\text{A}_i)},\ \text{για κάθε i.}$$

Παράδειγμα

Σε ένα εργοστάσιο οι μηχανές Α,Β και Γ παράγουν το 25%, το 35% και το 40% της συνολικής παραγωγής. Μετά από έλεγχο βρέθηκε ότι το 5% των παραγόμενων τεμαχίων από το μηχάνημα Α είναι ελαττωματικό, επίσης το 4% και το 2% των παραγόμενων τεμαχίων από τις μηχανές Β και Γ είναι ελαττωματικό.Η συνολική παραγωγή φτάνει τα 5.000 τεμάχια. Ένα τεμάχιο εξάγεται τυχαία και βρίσκεται ελαττωματικό.

Ποιά η πιθανότητα κατασκευής του από τη μηχανή Β;

Λύση:

Έστω:

E το ενδεχόμενο ότι το τεμάχιο που έχει εξαχθεί είναι ελαττωματικό.

A το ενδεχόμενο ότι το τεμάχιο αυτό έχει εξαχθεί από το Α.

B το ενδεχόμενο ότι το τεμάχιο αυτό έχει εξαχθεί από το Β.

Γ το ενδεχόμενο ότι το τεμάχιο αυτό έχει εξαχθεί από το Γ.

Θα έχουμε:

$$P(A)=\frac{25}{100},\ P(B)=\frac{35}{100},\ P(\Gamma)=\frac{40}{100}$$

$$P(E/A)=0,05,\ P(E/B)=0,04,\ P(E/\Gamma)=0,02$$

Επομένως:

$$P(B/E)=\frac{P(B)\cdot P(E/B)}{P(A)P(E/A)+P(B)P(E/B)+P(\Gamma)P(E/\Gamma)}\Rightarrow$$

$$\Rightarrow P(B/E)=\frac{0,35\cdot 0,04}{0,25\cdot 0,05+0,35\cdot 0,04+0,40\cdot 0,02}=0,4058$$

ΕΦΑΡΜΟΓΗ

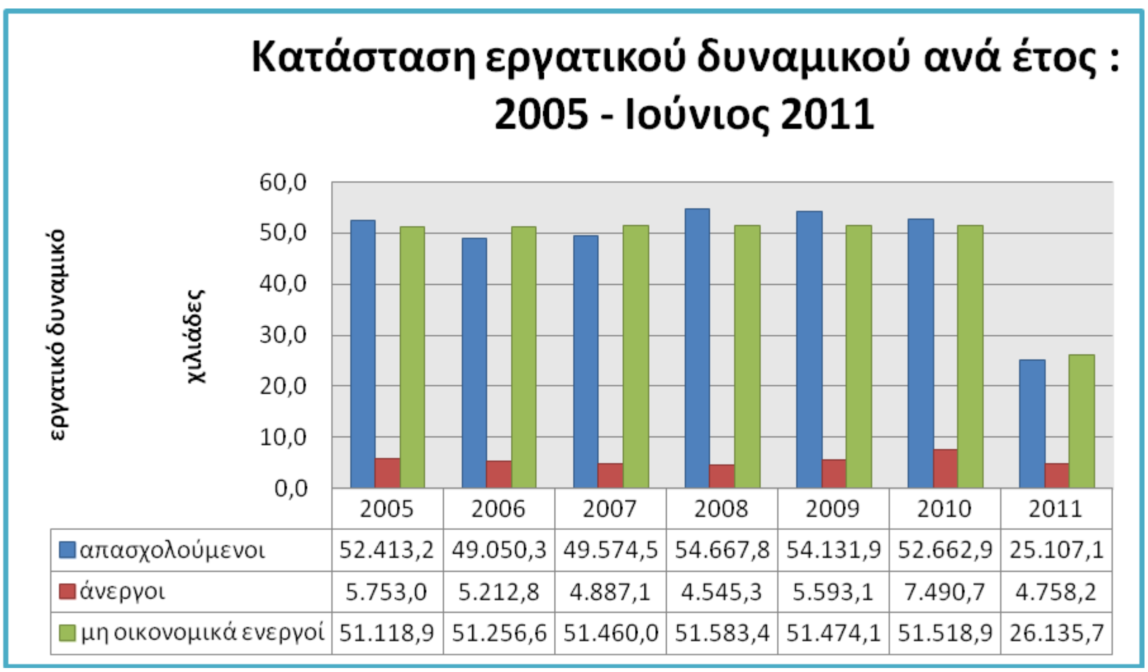
Θέλοντας να δώσουμε μια πιο σαφή εικόνα όλων των παραπάνω, θα χρησιμοποιήσουμε μια εφαρμογή στην οποία θα αναλύουμε εκτενέστερα τα στατιστικά μέτρα, καταλήγοντας στα συμπεράσματα στα οποία αναφερθήκαμε στο θεωρητικό τμήμα της εργασίας.Για το λόγο αυτό, λάβαμε στοιχεία από μηνιαία έρευνα εργατικού δυναμικού για την κατάσταση απασχόλησής τους κατά φύλο και κατά ποσοστό ανεργίας.Ειδικότερα, τα στοιχεία που θα εξετάσουμε, αναφέρονται σε άνδρες και γυναίκες οι οποίοι διακρίνονται σε:

1. Απασχολούμενοι
2. Άνεργοι
3. Μη οικονομικά ενεργοί

Τα δεδομένα θα εισαχθούν στο στατιστικό πρόγραμμα SPSS, όπου θα γίνει και η επίλυσή τους. Αναφορικά, θα παρουσιαστούν τα διαγράμματα στο EXCELL και θα διεξαχθούν τα συμπεράσματα που μας ενδιαφέρουν.

ΜΕΡΟΣ 1ο:

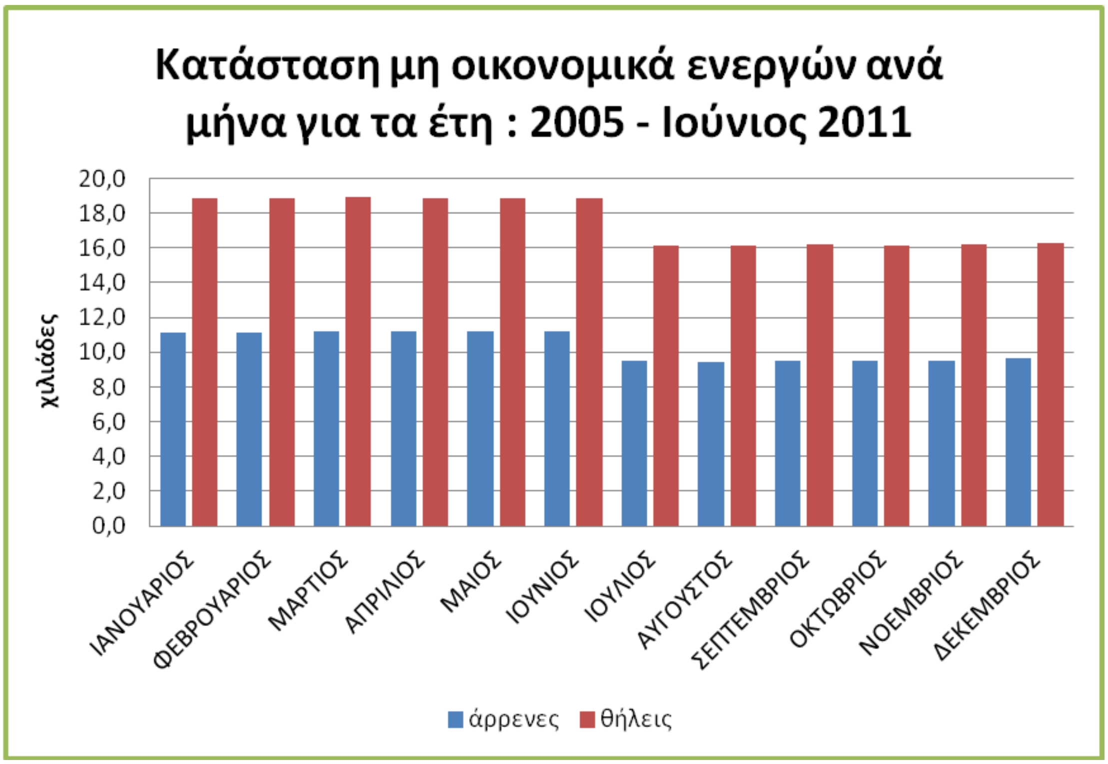
<u>ΔΙΑΓΡΑΜΜΑ 1</u>



	2005	2006	2007	2008	2009	2010	2011
απασχολούμενοι	52.413,2	49.050,3	49.574,5	54.667,8	54.131,9	52.662,9	25.107,1
άνεργοι	5.753,0	5.212,8	4.887,1	4.545,3	5.593,1	7.490,7	4.758,2
μη οικονομικά ενεργοί	51.118,9	51.256,6	51.460,0	51.583,4	51.474,1	51.518,9	26.135,7

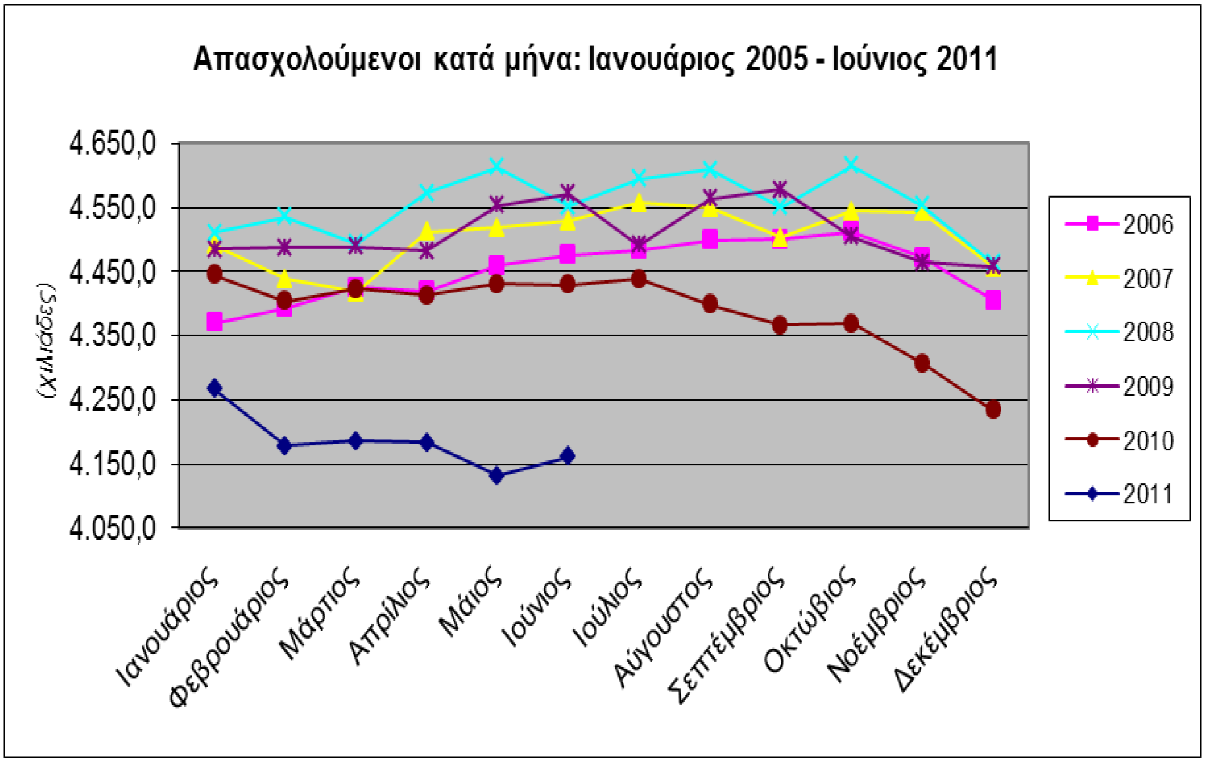
Στο διάγραμμα 1 παρατίθεται η κατάσταση του εργατικού δυναμικού ανά έτος από το 2005 έως τον Ιούνιο του 2011. Αναφορικά, ερμηνεύοντας την 4η στήλη η οποία παρουσιάζει τη μεγαλύτερη αύξηση, θα λέγαμε πως το έτος 2008 ο αριθμός τον απασχολούμενων ανδρών-γυναικών ανέρχεται στις 54.667,8 χιλιάδες. Αντίστοιχα ο αριθμός των ανέργων παρουσιάζει μεγάλη αύξηση το έτος 2010 και ανέρχεται στις 7.490,7 χιλιάδες ενώ το έτος 2008 παρατηρείται επίσης αυξημένος αριθμός (51.583,4) και των μη οικονομικά ενεργών ανδρών- γυναικών.

ΔΙΑΓΡΑΜΜΑ 2



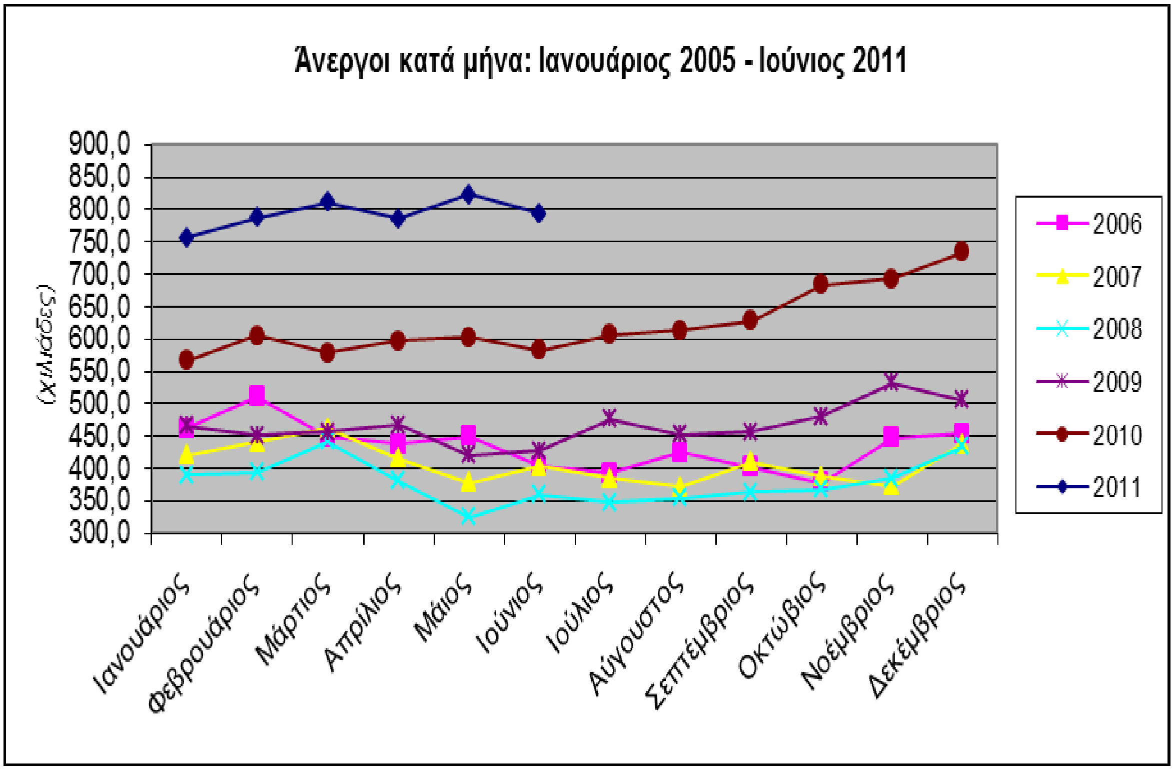
Στοπαραπάνω διάγραμμα απεικονίζεται η κατάσταση των μη οικονομικά ενεργών ανά μήνα για τα έτη 2005 έως τον Ιούνιο του έτους 2011. Παρατηρώντας το γράφημα, ένα γενικό συμπέρασμα στο οποίο καταλήγουμε είναι πως ο αριθμός των μη οικονομικά ενεργών γυναικών παρουσιάζει μεγαλύτερη αύξηση από τον αριθμό των μη οικονομικά ενεργών ανδρών στο σύνολο όλων των ετών. Επίσης, παρατηρούμε πως από τον μήνα Ιούλιο έως τον Δεκέμβριο υπάρχει μείωση του αριθμού των μη οικονομικά ενεργών και για τα δύο φύλα.

ΔΙΑΓΡΑΜΜΑ 3



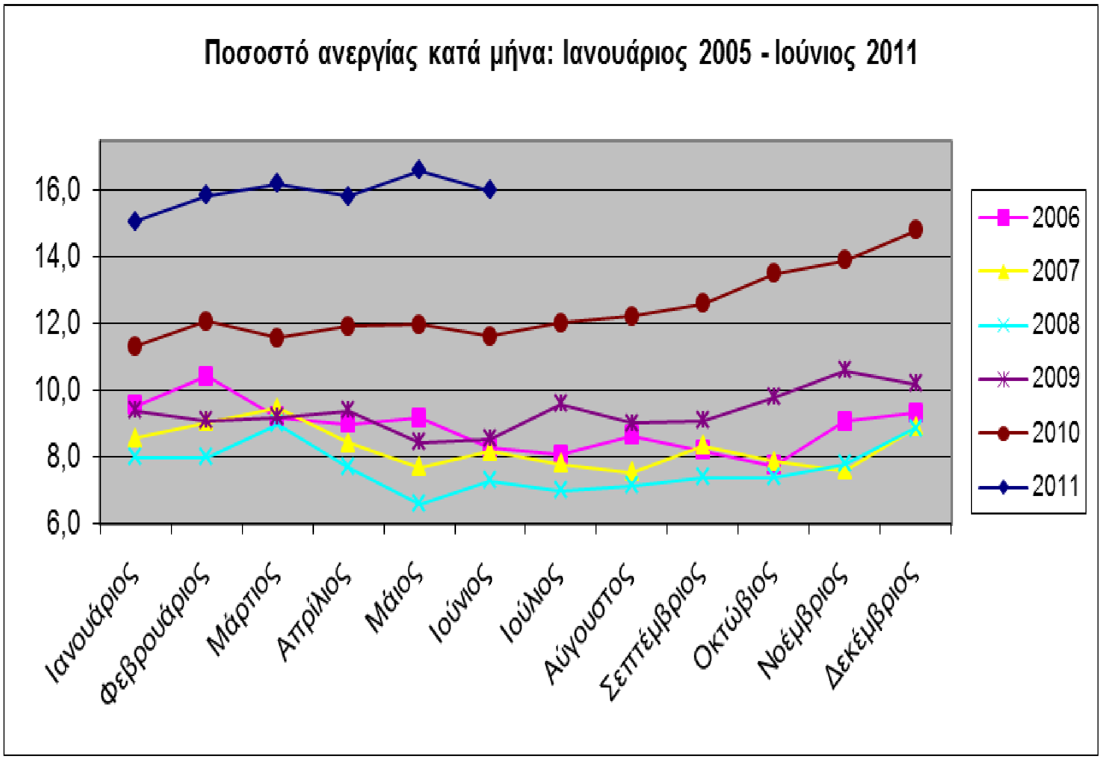
Όσον αφορά το διάγραμμα 3, παρατηρούμε πως το έτος 2011 σε αντίθεση με τα προηγούμενα έτη υπάρχει αισθητή μείωση του αριθμού των απασχολούμενων. Πιο συγκεκριμένα το έτος 2006 ο αριθμός των απασχολούμενων παρουσιάζει μια ανοδική τάση με τη μεγαλύτερη αύξηση τον Οκτώβριο, ενώ για τους μήνες Νοέμβριο και Δεκέμβριο παρατηρείται μείωση. Αντίθετα, το έτος 2007 ξεκινά με μια καθοδική τάση μέχρι τον Μάρτιο, όπου σταδιακά παρουσιάζει άνοδο μέχρι τον Ιούλιο, έπειτα υπάρχει μείωση έως τον Σεπτέμβριο, αύξηση έως το Νοέμβριο και ξανά μείωση έως τον Δεκέμβριο. Το έτος 2008 παρατηρούνται οι μεγαλύτερες αυξήσεις, εν συνεχεία το 2009 υπάρχουν αυξομειώσεις και το 2010 παρατηρείται μεγάλη μείωση από τον Ιούλιο και έπειτα. Τέλος το 2011, ξεκινά με μια μείωση έως τον Φεβρουάριο, συνεχίζει με μια μικρή αύξηση έως τον Απρίλιο, έπειτα μέχρι τον Μάιο παρατηρείται μείωση, και ξανά αύξηση μέχρι τον Ιούνιο.

ΔΙΑΓΡΑΜΜΑ 4



Σε αντίθεση με τοδιάγραμμα 3 όπου το έτος 2011 παρατηρούνταν μείωση του αριθμού των απασχολούμενων, στο διάγραμμα 4 ο αριθμός των ανέργων ξεπερνά τα προηγούμενη έτη με αισθητή διαφορά. Παρόμοιο φαινόμενο παρατηρείται και για το έτος 2010 και συγκεκριμένα από τον μήνα Ιούνιο και ύστερα υπάρχει έντονη αύξηση του αριθμού των ανέργων. Ειδικότερα, από το γράφημα μπορούμε να παρατηρήσουμε πως τα έτη 2006 έως 2009 ο αριθμός των ανέργων κυμαίνεται περίπου στα ίδια επίπεδα με αρκετές αλλά μικρές αυξομειώσεις.

ΔΙΑΓΡΑΜΜΑ 5



Στο διάγραμμα 5 παρατίθεται το ποσοστό ανεργίας για τα έτη 2005 έως τον Ιούνιο του έτους 2011. Όπως μπορούμε να παρατηρήσουμε το υψηλότερο ποσοστό ανεργίας παρουσιάζεται το έτος 2011. Επιπλέον, παρόμοια ανοδική πορεία παρουσίαζε και το έτος 2010. Συμπερασματικά θα μπορούσαμε να πούμε πως τα δύο τελευταία έτη το επίπεδο του ποσοστού ανεργίας αυξάνεται ραγδαία.

ΜΕΡΟΣ 2ο :

Στο περιγραφικό μέρος της εφαρμογής δώσαμε έμφαση στην κατανομή της συχνότητας κάθε μεταβλητής, με σκοπό να παρουσιάσουμε όσο σαφέστερα γίνεται τα αποτελέσματα της έρευνας.

Στο αμέσως επόμενο μέρος θα ασχοληθούμε με τα μέτρα θέσης. Πιο συγκεκριμένα, θα συγκρίνουμε τα αποτελέσματα για κάθε φύλο, ώστε να καταλήξουμε στα συμπεράσματα που θα μας βοηθήσουν να καταλάβουμε την κατάσταση του εργατικού δυναμικού σύμφωνα με τα στοιχεία της έρευνας.

Επίσης, θα πρέπει να επισημάνουμε πως τα σημαντικότερα μέτρα με τα οποία θα ασχοληθούμε είναι:

- ΑΡΙΘΜΗΤΙΚΟΣ ΜΕΣΟΣ: Μας υποδεικνύει το μέσο όρο της κατάστασης του εργατικού δυναμικού ανά έτος.

- ΔΙΑΜΕΣΟΣ: Διαπιστώνουμε σε ποιο ύψος ανήκει το ήμισυ των παρατηρήσεων και σε ποιο το υπόλοιπο 50%.

- ΣΥΝΤΕΛΕΣΤΗΣ ΑΣΥΜΜΕΤΡΙΑΣ : Μας υποδεικνύει την ύπαρξη θετικής ή αρνητικής ασυμμετρίας ή συμμετρικής κατανομής.

- ΣΥΝΤΕΛΕΣΤΗΣ ΚΥΡΤΩΣΗΣ: Ορίζει αν υπάρχει μεσόκυρτη ή κανονική κατανομή, λεπτόκυρτη ή πλατύκυρτη κατανομή

- ΔΙΑΚΥΜΑΝΣΗ : Η διακύμανση ή διασπορά s^2 είναι ένα μέτρο διασποράς των τιμών του δείγματος.Όταν χρησιμοποιούμε όλον τον στατιστικό πληθυσμό,τότε συμβολίζουμε τη διακύμανση με σ^2,ενώ όταν αναφερόμαστε σε δείγμα την συμβολίζουμε με s^2.

- ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ : Η θετική τετραγωνική ρίζα της διασποράς.Χρησιμοποιείται πιο συχνά απ’ τη διασπορά,γιατί έχει τις ίδιες μονάδες μέτρησης με το μέσο.Συμβολίζεται με “ΔΧ” ή με “σ” ή με “σ(Χ)”.

- ΕΥΡΟΣ : Είναι το μέτρο που μας δίνει τη διαφορά μεταξύ της μέγιστης και της ελάχιστης τιμής ενός πληθυσμού.

- ΕΛΑΧΙΣΤΟ : Ελάχιστο είναι η μικρότερη τιμή του δείγματος.

- ΜΕΓΙΣΤΟ : Μέγιστο είναι η μεγαλύτερη τιμή του δείγματος.

ΜΕΤΡΑ ΘΕΣΗΣ	**ΕΤΗ**	**ΑΠΑΣΧΟΛΟΥΜΕΝΟΙ**	**ΑΝΕΡΓΟΙ**	**ΜΗ ΟΙΚΟΝΟΜΙΚΑ ΕΝΕΡΓΟΙ**
ΜΕΣΟΣ	2005	4367,767	479,417	4259,908
	2006	4451,683	434,383	4271,358
	2007	4505,500	407,250	4288,333
	2008	4555,641	378,783	4298,625
	2009	4511,000	466,091	4289,525
	2010	4388,566	624,225	4293,258
	2011	4184,516	793,033	4355,966
ΔΙΑΜΕΣΟΣ	2005	4375,850	475,500	4267,350
	2006	4466,300	442,200	4271,000
	2007	4515,550	406,850	4289,900
	2008	4553,330	374,000	4292,950
	2009	4490,950	461,350	4293,400
	2010	4408,500	606,150	4290,500
	2011	4180,800	790,450	4364,500
ΔΙΑΚΥΜΑΝΣΗ	2005	1526,906	276,972	844,310
	2006	2217,109	1319,680	736,426
	2007	2099,647	861,730	345,437
	2008	2377,384	1154,520	658,486
	2009	1893,107	940,172	627,033
	2010	3908,459	2664,837	729,146
	2011	2060,338	520,759	1015,215

ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ	2005	39,0756	16,6425	29,0570
	2006	47,0861	36,3274	27,1371
	2007	45,8219	29,3552	18,5859
	2008	48,7604	33,9782	25,6609
	2009	43,5098	30,6622	25,0406
	2010	62,5176	51,6120	27,0027
	2011	45,3931	21,8201	31,8624
ΕΛΑΧΙΣΤΟ	2005	4292,1	458,7	4207,5
	2006	4369,9	378,6	4230,7
	2007	4418,5	372,1	4254,5
	2008	4463,4	324,8	4260,1
	2009	4457,7	420,4	4241,0
	2010	4233,8	567,1	4261,5
	2011	4131,5	756,8	4302,3
ΜΕΓΙΣΤΟ	2005	4423,0	508,1	4294,6
	2006	4512,8	511,3	4325,1
	2007	4558,1	462,0	4322,6
	2008	4615,8	442,6	4350,9
	2009	4577,8	532,0	4323,1
	2010	4445,7	733,6	4353,1
	2011	4267,6	822,7	4385,6

ΕΥΡΟΣ	2005	130,9	49,40	87,1
	2006	142,9	132,70	94,4
	2007	139,6	89,90	68,1
	2008	152,4	117,80	90,8
	2009	120,1	111,60	82,1
	2010	211,9	166,50	91,6
	2011	136,1	65,90	83,3
ΑΣΣΥΜΕΤΡΙΑ	2005	-1,135	0,628	-0,651
	2006	-0,392	0,420	0,292
	2007	-0,806	0,472	0,016
	2008	-0,457	0,618	0,544
	2009	-0,575	0,722	-0,676
	2010	-1,668	1,145	0,968
	2011	1,320	-0,400	-1,068
ΚΥΡΤΩΣΗ	2005	0,915	-0,876	-0,479
	2006	-1,201	0,522	-0,002
	2007	-0,486	-0,812	0,157
	2008	-0,569	0,217	0,037
	2009	-1,391	0,936	-0,333
	2010	2,606	0,333	0,764
	2011	2,958	0,493	0,419

ΜΕΤΡΑ ΘΕΣΗΣ	**ΕΤΗ**	**ΑΝΕΡΓΟΙ ΑΝΔΡΕΣ**	**ΑΝΕΡΓΕΣ ΓΥΝΑΙΚΕΣ**
ΜΕΣΟΣ	2005	176,9333	302,492
	2006	162,6250	271,766
	2007	150,5917	256,650
	2008	148,4583	230,300
	2009	198,7750	268,158
	2010	286,4417	337,775
	2011	384,5667	408,466
ΔΙΑΜΕΣΟΣ	2005	174,4000	304,900
	2006	163,5000	276,200
	2007	144,3500	257,900
	2008	146,5000	233,450
	2009	194,6500	271,600
	2010	281,6000	327,500
	2011	383,1500	410,100
ΔΙΑΚΥΜΑΝΣΗ	2005	92,121	152,154
	2006	352,662	387,168
	2007	251,546	274,506
	2008	412,779	265,562
	2009	497,237	408,337
	2010	884,188	626,368
	2011	242,087	77,687
ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ	2005	9,59795	12,3351
	2006	18,77930	19,6765
	2007	15,86021	16,5682
	2008	20,31696	16,2960
	2009	22,29880	20,2073

	2010	29,73530	25,0273
	2011	15,55913	8,8140
ΕΛΑΧΙΣΤΟ	2005	165,50	282,2
	2006	130,30	244,3
	2007	135,90	232,6
	2008	124,40	200,3
	2009	168,60	234,7
	2010	249,00	312,3
	2011	361,10	395,7
ΜΕΓΙΣΤΟ	2005	190,10	319,4
	2006	196,30	315,0
	2007	180,40	281,7
	2008	188,60	257,6
	2009	257,70	303,2
	2010	342,50	391,1
	2011	405,50	417,3
ΕΥΡΟΣ	2005	24,60	37,2
	2006	66,00	70,7
	2007	44,50	49,1
	2008	64,20	57,3
	2009	89,10	68,5
	2010	93,50	78,8
	2011	44,40	21,6
ΑΣΣΥΜΕΤΡΙΑ	2005	0,246	-0,331
	2006	-0,066	0,510
	2007	1,078	-0,093
	2008	1,086	-0,221
	2009	1,639	-0,184

	2010	0,790	1,123
	2011	-0,179	-0,483
ΚΥΡΤΩΣΗ	2005	-1,857	-0,869
	2006	0,124	0,928
	2007	-0,230	-1,202
	2008	0,561	-0,475
	2009	4,382	-0,428
	2010	-0,122	0,334
	2011	-0,016	-1,656

ΜΕΤΡΑ ΘΕΣΗΣ	**ΕΤΗ**	**ΑΠΑΣΧΟΛΟΥΜΕΝΟΙ ΑΝΔΡΕΣ**	**ΑΠΑΣΧΟΛΟΥΜΕΝΕΣ ΓΥΝΑΙΚΕΣ**
ΜΕΣΟΣ	2005	2694,992	1672,800
	2006	2727,175	1724,568
	2007	2756,991	1748,58
	2008	2771,833	1789,283
	2009	2720,725	1790,191
	2010	2627,150	1761,416
	2011	2494,533	1690,000
ΔΙΑΜΕΣΟΣ	2005	2694,450	1676,300
	2006	2737,100	1731,000
	2007	2757,650	1753,550
	2008	2780,700	1792,000
	2009	2711,850	1783,450
	2010	2634,550	1770,750
	2011	2481,150	1691,350
ΔΙΑΚΥΜΑΝΣΗ	2005	230,032	1115,827
	2006	650,404	603,446
	2007	676,679	566,034
	2008	1261,697	1185,196
	2009	595,138	545,108
	2010	1670,243	701,262
	2011	1286,719	404,448
ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ	2005	15,1668	33,4040
	2006	25,5030	24,5651
	2007	26,0130	23,7914
	2008	35,5205	34,4266
	2009	24,3954	23,3475
	2010	40,8686	26,4813

	2011	53,8708	20,1108
ΕΛΑΧΙΣΤΟ	2005	2663,2	1604,5
	2006	2683,3	1686,6
	2007	2711,7	1702,0
	2008	2694,9	1751,0
	2009	2680,8	1750,0
	2010	2535,3	1698,5
	2011	2460,0	1667,6
ΜΕΓΙΣΤΟ	2005	2715,9	1721,5
	2006	2762,2	1757,3
	2007	2793,3	1779,8
	2008	2819,0	1877,4
	2009	2763,6	1829,5
	2010	2676,7	1796,6
	2011	2557,0	1710,5
ΕΥΡΟΣ	2005	52,7	117,0
	2006	78,9	70,7
	2007	81,6	77,8
	2008	124,1	126,4
	2009	82,8	79,5
	2010	141,4	98,1
	2011	97,0	42,9
ΑΣΣΥΜΕΤΡΙΑ	2005	-0,521	-0,626
	2006	-0,499	-0,276
	2007	-0,357	-0,972
	2008	-0,772	1,521
	2009	0,410	0,145
	2010	-1,289	-1,151
	2011	1,268	-0,060

ΚΥΡΤΩΣΗ	2005	0,273	0,257
	2006	-0,901	-1,567
	2007	-0,551	0,504
	2008	0,773	3,354
	2009	-0,624	-0,642
	2010	1,423	1,719
	2011	1,021	-3,059

ΜΕΤΡΑ ΘΕΣΗΣ	ΕΤΗ	ΜΗ ΟΙΚΟΝΟΜΙΚΑ ΕΝΕΡΓΟΙ ΑΝΔΡΕΣ	ΜΗ ΟΙΚΟΝΟΜΙΚΑ ΕΝΕΡΓΕΣ ΓΥΝΑΙΚΕΣ
ΜΕΣΟΣ	2005	1558,500	2701,392
	2006	1567,491	2703,891
	2007	1579,091	2709,075
	2008	1583,8367	2714,783
	2009	1601,133	2688,566
	2010	1628,275	2664,975
	2011	1677,900	2694,733
ΔΙΑΜΕΣΟΣ	2005	1556,450	2714,900
	2006	1567,550	2708,900
	2007	1579,400	2711,300
	2008	1584,920	2720,520
	2009	1603,850	2694,650
	2010	1619,950	2663,300
	2011	1687,700	2678,400
ΔΙΑΚΥΜΑΝΣΗ	2005	111,304	1115,827
	2006	164,888	603,446
	2007	141,423	566,034
	2008	434,592	1185,196
	2009	533,921	545,108
	2010	421,173	701,262
	2011	632,016	404,448
ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ	2005	10,5501	33,5891
	2006	12,8408	22,3379
	2007	11,8921	18,7686
	2008	20,8468	18,4710
	2009	18,2734	20,0288
	2010	20,5225	10,0028
	2011	25,1399	48,6096

ΕΛΑΧΙΣΤΟ	2005	1545,2	2648,1
	2006	1547,1	2668,9
	2007	1562,0	2683,1
	2008	1555,2	2680,0
	2009	1564,8	2656,3
	2010	1608,3	2652,0
	2011	1635,1	2653,9
ΜΕΓΙΣΤΟ	2005	1577,8	2748,7
	2006	1588,3	2736,8
	2007	1599,8	2737,5
	2008	1629,0	2734,3
	2009	1622,0	2718,7
	2010	1672,2	2682,5
	2011	1698,6	2787,0
ΕΥΡΟΣ	2005	32,6	100,6
	2006	41,2	67,9
	2007	37,8	54,4
	2008	73,8	54,3
	2009	57,2	62,4
	2010	63,9	30,3
	2011	63,5	133,1
ΑΣΣΥΜΕΤΡΙΑ	2005	0,595	-0,534
	2006	-0,011	-0,511
	2007	0,119	-0,096
	2008	0,625	-0,708
	2009	-0,629	-0,517
	2010	1,046	0,397
	2011	-1,191	1,765
ΚΥΡΤΩΣΗ	2005	-0,506	-1,247
	2006	-0,992	-0,782

	2007	-0,689	-1,485
	2008	0,699	-0,790
	2009	-0,496	-0,742
	2010	0,159	-0,406
	2011	0,423	3,32

Τα συμπεράσματα που προκύπτουν βάση των ετησίων αποτελεσμάτων που εξαγάγαμε,μας βοηθούν να συγκρίνουμε σε πιο συνολικό επίπεδο τους ανέργους,τους απασχολούμενους και τους μη οικονομικά ενεργούς άνδρες-γυναίκες,ώστε να καταλήξουμε ποιά ήταν τα έτη στα οποία είχαμε το μεγαλύτερο ποσοστό ανέργων,απασχολούμενων και μη οικονομικά ενεργών ανδρών-γυναικών.

Εκ νέου,λοιπόν και πιο συνοπτικά,θα μπορούσαμε να πούμε ότι:
Για τους απασχολόυμενος παρατηρούμε οτι το 2008 ήταν το έτος με το μεγαλύτερο ποσοστό απασχόλησης.Πιο συγκεκριμένα,το 2008 οι απασχολούμενοι άνδρες κατείχαν μεγαλύτερο ποσοστό απασχόλησης απ’τις γυναίκες με 2.771.833 απασχολούμενους άνδρες ενώ οι απασχολούμενες γυναίκες ήταν 1.789.283.

Για τους ανέργους βλέπουμε ότι το 2011 ήταν το έτος που παρατηρήθηκε το υψηλότερο ποσοστό ανεργίας.Πιο συγκεκριμένα,το 2011 ο αριθμός των ανέργων ανδρών ανέρχεται στις 384.566 σε αντίθεση με τις γυναίκες οι οποίες είναι περισσότερες και ανέρχονται στις 408.466.

Για τους μη οικονομικά ενεργούς το έτος 2011 είχαμε το μεγαλύτερο ποσοστό.Πιο συγκεκριμένα οι μη οικονομικά ενεργοί άνδρες ήταν 1.677.900 και οι γυναίκες ήταν 2.694.733.

Επίσης,αν εξετάσουμε πιο σφαιρικά τα αποτελέσματα,δηλαδή τα ποσοστά ανεργίας,απασχόλησης και μη οικονομικά ενεργής απασχόλησης για όλα τα έτη με βάση τους μήνες,παρατηρούμε ότι :

Από το διάγραμμα 2 οι μη οικονομικά ενεργοί για τα έτη 2005 εως 2011 από τον μήνα Ιανουάριο εως το μήνα Ιούνιο έχουν υψηλότερα ποσοστά σε σχέση με τους μήνες Ιούλιο εως Δεκέμβριο.

Πιο συγκεκριμένα,για τις γυναίκες το ποσοστό είναι περίπου 2,5 χιλιάδες υψηλότερο,ενώγια τους άνδρες είναι περίπου 1,5 χιλιάδες υψηλότερο.

Από το διάγραμμα 3 παρατηρούμε ότι τα ποσοστά απασχόλησης ανδρών και γυναικών για τα έτη 2005 εως 2009 είναι υψηλότερα κατά τους μήνες Απρίλιο εως Οκτώβριο,ενώ στα έτη 2010 και 2011 το ποσοστό απασχόλησης βλέπουμε να μειώνεται από το μήνα Ιούλιο.

Και τέλος, από το διάγραμμα 4 παρατηρούμε ότι το ποσοστό ανεργίας ανδρών και γυναικών για τα έτη 2005 εως 2009 είναι αυξημένα τους μήνες Οκτώβριο εως Απρίλιο,ενώ για τα έτη 2010 και 2011 έχουμε αύξηση του ποσοστού ανεργίας από τον μήνα Ιούνιο.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Α. ΕΛΛΗΝΙΚΗ

1. Κιόχος Α.Π.: «Στατιστική» , εκδόσεις Interbooks, Αθήνα 1993 «

2.Κικιλιάς Π.-Παλαμούρδας Δ.-Πετράκης Α.-Τσουκαλάς Δ.: «Στατιστική- Πιθανότητες», εκδόσεις Δήρος, Αθήνα 20001

3.Ρόντος Κ.-Παπανής Ε.: «Μέθοδοι συλλογής στατιστικών δεδομένων», εκδόσεις Σιδέρη, Αθήνα 2007

4. Κιόχος Α.Π.: «Περιγραφική Στατιστική», εκδόσεις Interbooks, Αθήνα 1993

5. Κουρουκλή Σ.: «Στατιστική Ι», εκδόσεις Πανεπιστημίου Πατρών, Πάτρα 1999

6. Ιωαννίδης Α.Δ.: «Στατιστικές Μέθοδοι», εκδόσεις Ζήτη, Θεσσαλονίκη 1999

7. Τσερπές Ν.Α.- Αλεβίζος Φ.: «Εισαγωγή στην Θεωρία της δειγματοληψίας», εκδόσεις Πανεπιστημίου Πατρών, Πάτρα 1999

Β. ΗΛΕΚΤΡΟΝΙΚΗ

1.www.statistics.scientist.gr

2. www.el.wikipedia.org

3. www.nnipi.gr

4 www.math.ntua.gr

5 www.cc.uoq.gr