ΤΕΙ ΠΑΤΡΩΝ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΥ ΣΧΕΔΙΑΣΜΟΥ ΚΑΙ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΤΙΤΛΟΣ ΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ:
ΕΡΕΥΝΗΤΙΚΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

ΣΠΟΥΔΑΣΤΕΣ
ΔΑΜΙΡΗΣ ΚΩΝΣΤΑΝΤΙΝΟΣ
ΜΟΥΝΔΡΟΥ ΚΑΛΛΙΟΠΗ
ΠΙΠΕΡΑ ΚΩΝΣΤΑΝΤΙΝΑ

ΕΠΟΠΤΕΥΩΝ ΚΑΘΗΓΗΤΡΙΑ
ΜΠΟΥΜΠΟΥΛΗ ΑΘΑΝΑΣΙΑ

ΠΑΤΡΑ, 2009

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ ΠΡΩΤΟ

ΕΙΣΑΓΩΓΗ

1.1 Τι είναι η στατιστική ……………………………………………………………….Σελ.6
1.2 Ιστορία της Στατιστικής ………… ……………………………………………………Σελ.7
1.3 Αντικείμενο της Στατιστικής ………………………………………………………Σελ.8
1.4 Χρησιμότητα και πεδία εφαρμογής της Στατιστικής……………………………….Σελ.9
1.5 Κλίμακες μέτρησης …………………………………………………………………Σελ.11
1.6 Έννοια Στατιστικής μεταβλητής – Διακρίσεις αυτής ……………………………...Σελ.14

ΚΕΦΑΛΑΙΟ ΔΕΥΤΕΡΟ

ΣΥΛΛΟΓΗ ΣΤΑΤΙΣΤΙΚΩΝ ΣΤΟΙΧΕΙΩΝ

2.1 Επιλογή μεθόδου πρωτογενούς συλλογής στατιστικών δεδομένων…………...Σελ.16
2.2 Συλλογή στατιστικών δεδομένων από δευτερογενείς πηγές …..………………….Σελ.20
2.3 Συλλογή στατιστικών δεδομένων με τη μέθοδο των μητρώων………………….Σελ.21
2.4 Στατιστική μονάδα - Στατιστικός πληθυσμός ………………………………….Σελ.24
2.5 Πηγές συλλογής στατιστικών στοιχείων ……………………………………...Σελ.29
2.6 Μέθοδοι συγκέντρωσης στατιστικών δεδομένων ……………………………….Σελ.30
2.6.1 Απογραφή ……………………………………………………………………....Σελ.30
2.6.2 Δειγματοληπτική μέθοδος ……………………………………………………….Σελ.32

ΚΕΦΑΛΑΙΟ ΤΡΙΤΟ

ΔΕΙΓΜΑΤΟΛΗΨΙΑ

3.1 Γενικές έννοιες της δειγματοληψίας ……………………………………………...Σελ.33
3.2 Τεχνική των στατιστικών δειγματοληπτικών ερευνών ………………………….Σελ.34
3.3 Κατάρτιση πλαισίου δειγματοληψίας ……………………………………………...Σελ. 35
3.4 Μέθοδος δείγματος ……………………………………………………………….Σελ.35

3.4.1 Δειγματοληπτικά – Μη δειγματοληπτικά λάθη …………………………….....Σελ.36
3.4.2 Ερωτηματολόγιο ……………………………………………………………..Σελ.36
3.5 Μέθοδοι συλλογής στοιχείων στις δειγματοληπτικές έρευνες …………………..Σελ.39
3.5.1 Παρατήρηση ……………………………………………………………....Σελ.39
3.5.2 Συνέντευξη ……………………………………………………………….....Σελ.40
3.6 Μέθοδοι διενέργειας της δειγματοληψίας ……………………………………Σελ.40
3.6.1 Απλή τυχαία δειγματοληψία ………………………………………………..Σελ.41
3.6.2 Δειγματοληψία κατά στρώματα ……………………………………………Σελ.44
3.6.3 Κατά συστοιχίες δειγματοληψία ……………………………………….....Σελ.47
3.6.4 Επιφανειακή δειγματοληψία ………………………………………………..Σελ.47
3.6.5 Δειγματοληψία κατά ομάδες ………………………………………………..Σελ.47
3.6.6 Μεροληπτική δειγματοληψία ………………………………………………..Σελ.48
3.6.7 Συστηματική δειγματοληψία ………………………………………………..Σελ.48
3.7 Ορισμός τυχαίου δείγματος …………………………………………………………Σελ.49
3.8 Στατιστικά δειγματοληπτικά σφάλματα …………………………………………Σελ.52
3.9 Επεξεργασία δειγματοληπτικών στατιστικών στοιχείων …………………………Σελ.53
3.9.1 Παρουσίαση δειγματοληπτικών στοιχείων σε μορφή πινάκων …………...Σελ.54

ΚΕΦΑΛΑΙΟ ΤΕΤΑΡΤΟ

ΠΑΡΟΥΣΙΑΣΗ ΣΤΑΤΙΣΤΙΚΩΝ ΣΤΟΙΧΕΙΩΝ

4.1 Γραφήματα – Διαγράμματα ……………………………………………………….Σελ.58

ΚΕΦΑΛΑΙΟ ΠΕΜΠΤΟ

ΕΜΠΕΙΡΙΚΕΣ ΚΑΤΑΝΟΜΕΣ ΣΥΧΝΟΤΗΤΩΝ

5.1 Ασυνεχείς και Συνεχείς κατανομές ……………………………………………….Σελ.70
5.2 Ιστόγραμμα συχνοτήτων ………………………………………………………….Σελ.77
5.3 Αθροιστικές κατανομές συχνοτήτων ……………………………………………Σελ.81

ΚΕΦΑΛΑΙΟ ΕΚΤΟ

ΠΑΡΑΜΕΤΡΟΙ ΜΟΝΟΜΕΤΑΒΛΗΤΩΝ ΠΛΗΘΥΣΜΩΝ

6.1 Στατιστικά περιγραφικά μέτρα ……………………………………………………..Σελ.84
6.1.1 Είδη στατιστικών περιγραφικών μέτρων ………………………………..Σελ.85

ΚΕΦΑΛΑΙΟ ΕΒΔΟΜΟ

ΔΙΑΣΠΟΡΑ

7.1 Έννοια της διασποράς ……………………………………………………………Σελ.106
7.2 Μέτρα διασποράς ………………………………………………………………..Σελ.107
7.2.1Εύρος μεταβολής …………………………………………………………..Σελ.107
7.2.2 Ημιενδοτεταρτομοριακό εύρος ………………………………………...Σελ.108
7.2.3 Μέση απόκλιση …………………………………………………………Σελ.108
7.2.4 Διακύμανση και τυπική απόκλιση ……………………………………Σελ.109
7.2.5 Συντελεστής μεταβλητικότητας …………………………………………Σελ.114

ΚΕΦΑΛΑΙΟ ΟΓΔΟΟ

ΑΣΥΜΜΕΤΡΙΑ – ΚΥΡΤΩΣΗ

8.1 Ασυμμετρία ……………………………………………………………………Σελ.116
8.2 Κύρτωση ………………………………………………………………………Σελ. 119

ΚΕΦΑΛΑΙΟ ΕΝΑΤΟ

ΘΕΩΡΙΑ ΠΙΘΑΝΟΤΗΤΩΝ

9.1 Βασικές έννοιες ……………………………………………………………………Σελ.122
9.2 Συνοπτική ιστορική εξέλιξη της θεωρίας των πιθανοτήτων ……………………Σελ.123
9.3 Στατιστικός ορισμός της πιθανότητας ………………………………………Σελ.125
9.4 Ενδεχόμενα ……………………………………………………………………Σελ.125
9.4.1 Πράξεις με ενδεχόμενα ……………………………………………………..Σελ.126
9.5 Αξιωματική θεμελίωση των πιθανοτήτων ……………………………………..Σελ.132
9.6 Βασικά θεωρήματα των πιθανοτήτων …………………………………………Σελ.133

9.7 Δεσμευμένη πιθανότητα …………………………………………………………………Σελ.134
9.8 Ολική πιθανότητα ……………………………………………………………………...Σελ.135
9.8.1 Θεώρημα Bayes …………………………………………………………….Σελ.135
9.9 Συνδυαστική ……………………………………………………………………………Σελ.136
9.9.1 Αρχές απαρίθμησης ………………………………………………………...Σελ.136
9.10 Τυχαίες μεταβλητές …………………………………………………………………Σελ.141
9.10.1 Συνάρτηση πιθανότητας ………………………………………………Σελ.146
9.10.2 Συνάρτηση κατανομής ασυνεχούς τυχαίας μεταβλητής ………………Σελ.147
9.11 Διακύμανση …………………………………………………………………………..Σελ.148
9.11.1 Ιδιότητες της διακύμανσης …………………………………………….Σελ.148
9.12 Διάμεσος – Τεταρτημόρια …………………………………………………..Σελ.149

ΚΕΦΑΛΑΙΟ ΔΕΚΑΤΟ

ΘΕΩΡΗΤΙΚΕΣ ΚΑΤΑΝΟΜΕΣ

10.1 Γενικά …………………………………………………………………………Σελ.151
10.2 Ασυνεχείς κατανομές …………………………………………………………..Σελ.152
10.2.1 Διωνυμική κατανομή ………………………………………………………Σελ.153
10.2.2 Κατανομή Poisson ………………………………………………………..Σελ.158
10.3 Συνεχείς κατανομές …………………………………………………………….Σελ.161
10.3.1 Κανονική κατανομή ……………………………………………………….Σελ.162
10.3.2. Τυποποιημένη κανονική κατανομή ………………………………….Σελ.166
ΕΦΑΡΜΟΓΗ……………………………………………………………………………..Σελ.167
ΠΑΡΑΡΤΗΜΑ……………………………………………………………………………Σελ.190
ΒΙΒΛΙΟΓΡΑΦΙΑ…………………………………………………………………………Σελ.228

1. <u>Εισαγωγή</u>

1.1 **Τι είναι η Στατιστική**

Η Στατιστική είναι η επιστήμη που 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 των πιθανοτήτων. Η πρακτική της περιλαμβάνει τη σχεδίαση, συλλογή και ερμηνεία στατιστικών δεδομένων, τα οποία προκύπτουν από αβέβαιες παρατηρήσεις. Επειδή, λοιπόν, η στατιστική σαν επιστήμη αποσκοπεί στην εξαγωγή των πιο αξιόπιστων πληροφοριών, κατατάσσεται από μερικούς σαν κλάδος της θεωρίας των αποφάσεων. Μπορούμε λοιπόν να πούμε ότι:

<u>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. φαινόμενων και έχει ως σκοπό τη συστηματική μελέτη αυτών των στοιχείων για την κατάληξη σε γενικά συμπεράσματα, που είναι χρήσιμα στη διαδικασία λήψη ορθών αποφάσεων.</u>

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 περιεχομένου 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 Άγγλος αστρονόμος Hal1ey, 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 παιγνιδιών (χαρακτηριστική μάλιστα είναι η αλληλογραφία ανάμεσα στους Γάλλους μαθηματικούς Pascal και Fermat, με αφορμή τα ερωτήματα που έθεσε στον Pascal ο Ιππότης De Mere για τα παιγνίδια 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 φαινόμενων 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 του 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 είναι στην Ελλάδα η Εθνική Στατιστική Υπηρεσία ή Ε.Σ.Υ.Ε. (http://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 (http://europa.eu.int/comm/eurostat/) , διεθνείς οργανισμοί, όπως η UNESCO με το στατιστικό της ινστιτούτο (http://www.uis.unesco.org/) με αντικείμενο τη συλλογή, παρουσίαση και την επεξεργασία αριθμητικών πληροφοριών για τα επιμέρους κράτη και τις μεταξύ τους οικονομικές 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

Αρχικά, η Στατιστική είχε εμπειρικό (περιγραφικό) χαρακτήρα (Περιγραφική Στατιστική). Από τις αρχές του 20ου αιώνα, όμως, με την ανάπτυξη και μαθηματική θεμελίωση της Θεωρίας των Πιθανοτήτων, η Στατιστική εκτός από την αρχική περιγραφική της μορφή (που εξακολουθεί να διατηρεί) άρχισε παράλληλα να προσλαμβάνει και αυστηρή, μαθηματική μορφή (Μαθηματική Στατιστική).

Σήμερα η Στατιστική (Περιγραφική ή Μαθηματική) έχει εφαρμογές σε πάρα πολλούς τομείς της ανθρώπινης δραστηριότητας, όπως οι οικονομικές, φυσικές, κοινωνικές επιστήμες, η ιατρική (Βιοστατιστική), η οικολογία (Βιομετρία), η γλωσσολογία κλπ.

Στη Μαθηματική Στατιστική, τα προς ανάλυση δεδομένα θεωρούνται παρατηθείσες τιμές τυχαίων μεταβλητών. Ο πυρήνας της Μαθηματικής Στατιστικής είναι η Στατιστική Συμπερασματολογία που έχει σαν αντικείμενο την εξαγωγή συμπερασμάτων από τα δεδομένα, για άγνωστες παραμέτρους της κατανομής ή των κατανομών των παραπάνω τυχαίων μεταβλητών. Τα συμπεράσματα αυτά μπορούν εν συνεχεία να χρησιμοποιηθούν για τη λήψη αποφάσεων. 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 Συμπερασματολλογία (Σ.Σ.) περιλαμβάνει κυρίως τρεις κλάδους:

α. Εκτιμητική ή Σημειοληπτική (εκτίμηση με μία τιμή, ένα σημείο). Είναι ο κλάδος της Σ.Σ που ασχολείται με μεθόδους εκτίμησης (της τιμής) αγνώστων παραμέτρων κάποιας κατανομής. Η εκτίμηση γίνεται με τον προσδιορισμό των συναρτήσεων, των εκτιμητών, η τιμή των οποίων "τείνει να είναι κοντά" στην τιμή των αγνώστων παραμέτρων.

β. Διαστήματα Εμπιστοσύνης. Είναι ο κλάδος της Σ.Σ. που ασχολείται με τον προσδιορισμό Διαστημάτων (περιοχών γενικότερα) που περιέχουν με "μεγάλη πιθανότητα" άγνωστες παραμέτρους κάποιας κατανομής.

γ. Έλεγχος Στατιστικών Υποθέσεων. Είναι ο κλάδος της Σ.Σ. που ασχολείται με τον έλεγχο ισχυρισμών (υποθέσεων) για την τιμή αγνώστων παραμέτρων κάποιας κατανομής.

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, όπου η μελέτη της γαμηλιότητας, της γεννητικότητας, της θνησιμότητας, της μετανάστευσης κ.λπ., απαιτεί μακροχρόνιες στατιστικές παρατηρήσεις και επίπονες αναλύσεις. Επίσης, η

Στατιστική εφαρμόζεται σήμερα στην Ιατρική, Φυσική, Γενετική, Αστρονομία, Βιολογία, Μετεωρολογία, Γεωργία, Βιομηχανία, στη μελέτη του φυσικού περιβάλλοντος, στη μελέτη των ανθρωπίνων ιδεών και προθέσεων, στη Θεωρία των αποφάσεων, στον έλεγχο ποιότητας των προϊόντων κ.λπ. Τέλος, η Στατιστική βρίσκει πολύ μεγάλη εφαρμογή και στον Οικονομικό τομέα, 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. Η χρησιμότητα της Στατιστικής φαίνεται και από το γεγονός ότι η Στατιστική διδάσκεται σήμερα σχεδόν σε όλες της 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 στην 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.

Παραδείγματα

Φύλο: 1 = άντρας 2 = γυναίκα

Θρήσκευμα: 1 = χριστιανός 2 = μωαμεθανός 3 = άλλο

<u>ΔΙΑΤΑΚΤΙΚΗ</u>

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

<u>Παράδειγμα</u> Θέση:
1 στρατιώτης
2 δεκανέας
3 λοχίας
4 επιλοχίας
5 αξιωματικός

<u>ΙΣΟΔΙΑΣΤΗΜΙΚΗ</u>

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

<u>Παράδειγμα</u>

Θερμοκρασία σε °C ή °F, χρονολογία (π.χ. 2002)

<u>ΑΝΑΛΟΓΙΚΗ</u>

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

<u>Παράδειγμα</u>

Θερμοκρασία σε °K, απόσταση, βάρος, ηλικία

Οι σπουδαιότεροι κλίμακες που χρησιμοποιούνται στις έρευνες είναι:

- **Κλίμακα Likert**

Εδώ εξετάζεται ο βαθμός συμφωνίας ή διαφωνίας σε κάποιο θέμα που διατυπώνεται από το ερωτηματολόγιο. Η κλίμακα παρουσιάζεται σε ευθεία γραμμή με σταθερές διαιρέσεις (συνήθως 5), όπου ο ερωτώμενος καλείται να απαντήσει εάν συμφωνεί ή διαφωνεί, σε μία ερώτηση σαν σύνολο. Χάρη στην ευκολία που παρέχει, γίνεται και ευρεία χρήση της.

- **Κλίμακα Guttman**

Στην 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 αυξημένη 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**

Η Κλίμακα κατασκευάζεται με την συμμετοχή των ίδιων των ερωτώμενων, η οποία εκφράζει την αντίληψη των ερωτώμενων σε μεγαλύτερο βαθμό από τις υπόλοιπες κλίμακες. Ο ερευνητής διατυπώνει 50-100 ερωτήσεις, τις διανέμει, για να διαπιστώσει ποιες έχουν την μεγαλύτερη συχνότητα.

- **Κλίμακα Semantic Differential**

Στην ελληνική γλώσσα αποδίδεται ως κλίμακα σημασιολογικής διαφοροποίησης και έγκειται στον καθορισμό των δύο άκρων ενός χαρακτηριστικού (λ.χ. μορφωμένος μη μορφωμένος), καθιστώντας εφικτή τη δυνατότητα διαβαθμίσεων ανάμεσά τους.

- **Κλίμακα Borgadus**

Η κλίμακα Borgadus (social distance scale) είναι μία ψυχολογική δοκιμαστική κλίμακα η οποία δημιουργήθηκε από τον Emory S. Bogardus. Σχεδιάστηκε για να μπορεί να εξετάζει εμπειρικά την θέληση των ατόμων να συμμετέχουν σε διάφορες κοινωνικές επαφές με ποικίλους βαθμούς οικειότητας, έχοντας μέλη από διαφορετικές κοινωνικές ομάδες, όπως ρατσιστικές και εθνικές, αντίθετου φίλου και ομοφυλοφιλικές.

Οι Κλίμακες Μετρήσεων θα πρέπει να ικανοποιούν δύο βασικές αρχές: **αξιοπιστία και εγκυρότητα.**

Λέγοντας **αξιοπιστία,** εννοούμε τρόπους ελέγχου, όπως:

- Επαναληπτικές μετρήσεις, οι οποίες να δίνουν το ίδιο αποτέλεσμα.
- Οι ερωτήσεις να πραγματοποιούνται σε διαφορετικές χρονικές περιόδους.
- Εναλλασσόμενοι τύποι ερωτήσεων: δύο διαφορετικές αλλά ισοδύναμες κλίμακες, δηλαδή ίδιο δείγμα την ίδια στιγμή.
- Μέθοδος διχοτόμησης: δύο ίσα μέρη ερωτήσεων και ελέγχος ομοιογένειας των απαντήσεων.

Σε ότι αφορά την **εγκυρότητα,** υπάρχουν οι εξής κατηγορίες:

- 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

Οι χαρακτηριστικές ιδιότητες των στατιστικών μονάδων ενός πληθυσμού, με τη μελέτη των οποίων ασχολείται η Στατιστική, ονομάζονται μεταβλητές. 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,..., $x_ν$ ή y_1, y_2, y_3, . . . , $y_κ$. Αν η μεταβλητή είναι ποιοτική, εκφράζεται με λέξεις.

Για την καλύτερη κατανόηση της έννοιας της μεταβλητής, δίνονται στον πίνακα διάφορες μεταβλητές και οι αντίστοιχες τιμές τους.

Πίνακας

Μεταβλητές	*Τιμές*
Αριθμός αγοριών σε μια οικογένεια	0,1,2,3,4,5, ...
Ένδειξη ενός ζαριού	1,2,3,4,5,6
Ένδειξη ενός νομίσματος	«πρόσωπο», «γράμματα»
Αριθμός δωματίων ενός διαμερίσματος	1,2,3,4, ...
Φυλή ατόμων	λευκός, μαύρος, κίτρινος, ερυθρόδερμος
Ύψος ατόμων σε εκατοστά	145, 146, 147, "'
Υγεία ατόμων	άριστη, καλή, μέτρια, κακή
Φύλο ατόμων	αρσενικό, θηλυκό

Οι μεταβλητές χωρίζονται σε δύο κυρίως κατηγορίες:

- Στις *ποιοτικές μεταβλητές,* που 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» κ.λπ.

- Στις *ποσοτικές μεταβλητές που* 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 το εισόδημα ενός ατόμου, η θερμοκρασία, οι εξαγωγές, ο αριθμός των δωματίων ενός διαμερίσματος κ.λπ. Αν μια ποσοτική μεταβλητή σημειωθεί με το γράμμα X, οι τιμές της θα σημειώνονται με $X_1, X_2, X_3,...$

Οι ποσοτικές μεταβλητές διακρίνονται σε ασυνεχείς και συνεχείς.

Ασυνεχείς ονομάζονται οι μεταβλητές εκείνες που μπορούν να λάβουν πεπερασμένο ή αριθμήσιμο πλήθος τιμών. Έτσι, π.χ. η ένδειξη ενός ζαριού είναι μια ασυνεχής τυχαία μεταβλητή, γιατί το σύνολο των τιμών της {1, 2, 3, 4, 5, 6} είναι *πεπερασμένο.* Επίσης, ο αριθμός ρίψεων ενός νομίσματος μέχρι να εμφανισθεί για πρώτη φορά η όψη {προσώπου} είναι μια ασυνεχής μεταβλητή, γιατί το σύνολο τιμών της {1, 2, 3, …, *ν,* ...} είναι αριθμήσιμο.

Συνεχείς ονομάζονται οι μεταβλητές εκείνες που μπορούν να λάβουν όλες τις τιμές ενός διαστήματος. Έτσι, π.χ. το βάρος ή το ύψος ενός μαθητή, το εισόδημα ή η ηλικία ενός ατόμου, η ταχύτητα, η θερμοκρασία κ.λπ., είναι συνεχείς μεταβλητές.

Στα επόμενα κεφάλαια, θα αναλύσουμε σαφέστερα τις έννοιες και τη χρησιμότητα των μεταβλητών.

2. Συλλογή στατιστικών στοιχείων

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, για την έννοια και τα χαρακτηριστικά των οποίων θα αναφερθούμε και στις επόμενες σελίδες.

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.

2.1 Επιλογή μεθόδου πρωτογενούς συλλογής στατιστικών δεδομένων

Η επιλογή της μεθόδου, με την οποία θα συλλεχθούν πρωτογενώς τα δεδομένα της έρευνας, θα γίνει αφού συνεκτιμηθούν τα πλεονεκτήματα και τα μειονεκτήματα κάθε μιας. Βασική πάντως παράμετρος, όπως και σε όλες τις μεθοδολογικές επιλογές της έρευνας, είναι το είδος και οι επιμέρους στόχοι αυτής, όπως και το i) κόστος, ii) το διαθέσιμο προσωπικό και iii) ο χρόνος.

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

- **Με προσωπική συνέντευξη.**
- **Με συμπλήρωση των ερωτηματολογίων από τους ερευνώμενους.**
- **Με το ταχυδρομείο.**
- **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 σύνθετος 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 (Luck D. And Rubin R., 1987). Μια λύση στο πρόβλημα, όπου φυσικά είναι εφικτό, είναι η χρησιμοποίηση ως ερευνητών, φοιτητών του επιστημονικού κλάδου ή του επαγγελματικού τομέα που ερευνάται.

Όπως διαπιστώνουμε, είναι ιδιαίτερα σημαντική η σημασία της επιλογής των κατάλληλων ερευνητών ως και τα βασικά χαρακτηριστικά που πρέπει να συγκεντρώνουν, ιδιαίτερα όταν έχει επιλεγεί η μέθοδος της προσωπικής συνέντευξης.

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.

Η μέθοδος της προσωπικής συνέντευξης περιλαμβάνει έναν ελεύθερο διάλογο μάλλον (που παρομοιάζεται, ως προς το χαρακτηριστικό αυτό, με τη συνέντευξη που παίρνει ο ψυχίατρος από τον ασθενή ή από συγγενικό του πρόσωπο), παρά απαντήσεις σε δομημένο ερωτηματολόγιο. Επίσης, ομαδικές συνεντεύξεις της μορφής αυτής χρησιμοποιούνται σε ένα αρχικό στάδιο του σχεδιασμού μιας έρευνας προκειμένου να προκύψουν χρήσιμα συμπεράσματα για τη δομή και το περιεχόμενο του ερωτηματολογίου που θα καταρτιστεί (Hoinville G. et al, 1983).

Σύμφωνα με τη μέθοδο συμπλήρωση των ερωτηματολογίων από τους ερευνώμενους, 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 μεροληπτικά 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, ιδιαίτερα σε κοινωνίες όπως αυτή της Ελλάδος. Η αδυναμία αυτή απαιτεί την κατάρτιση ιδιαίτερα απλών ερωτηματολογίων και τη χρήση ειδικών τεχνικών σ’ αυτό ώστε ο ερωτώμενος να οδηγείται ευχερώς στον τρόπο με τον οποίο θα συμπληρώσει τα ερωτηματολόγια.

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.).

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.

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 τηλεφωνική επικοινωνία, 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.

Με μια άλλη επίσης τεχνική, που καλείται **Computer–assisted telephone interviewing (CATI),** πέραν της επιλογής του δείγματος, το σύστημα παρέχει τη δυνατότητα εισαγωγής των απαντήσεων (που εμφανίζονται κωδικοποιημένες σε οθόνη) απευθείας στον υπολογιστή, αποφεύγοντας έτσι τις διαδικασίες εγγραφής, διόρθωσης, κωδικογράφησης των ερωτηματολογίων και εισαγωγής των απαντήσεων στον υπολογιστή (Luck D. And Rubin R., 1987).

2.2 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 συλλεγεί 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 συλλεχθεί για στατιστικούς σκοπούς. Στοιχεία που έχουν συλλεχθεί 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).
Θα τη συμπεριλάβουμε, όμως, εδώ γιατί η εργασία καλύπτει και τη μεθοδολογία παραγωγής των στατιστικών αυτών.

Επιπλέον, ο ιδιώτης ερευνητής είναι δυνατόν να εξασφαλίζει κατά περίπτωση δεδομένα, αρκετά αναλυτικά, που προέρχονται από την οργάνωση και τη λειτουργία των Μητρώων.

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 (Redfern P., 1983).

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):

§ Την ύπαρξη ενός οργανωμένου συστήματος διοίκησης σε εθνικό, περιφερειακό και τοπικό επίπεδο.
§ Τη δημιουργία κατάλληλου νομικού πλαισίου που επιτρέπει:
    - Τη δυνατότητα προσπέλασης των αρμοδίων φορέων 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.

- 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.

Στην περίπτωση που έχει εξασφαλιστεί το κατάλληλο νομικό πλαίσιο και η απαιτούμενη υλικοτεχνική υποδομή, οι παραδοσιακές μέθοδοι συλλογής στατιστικών δεδομένων (με χρήση ερωτηματολογίων, απογραφέων, κλπ.) είναι δυνατόν να υποκατασταθούν πλήρως από ένα ολοκληρωμένο σύστημα μητρώων, με μικρές διαφορές στη ποιότητα των στατιστικών στοιχείων.

Στην περίπτωση που η συλλογή των στατιστικών στοιχείων δεν είναι δυνατόν να γίνει αποκλειστικά μέσω του συστήματος των μητρώων, η μέθοδος αυτή είναι δυνατό να χρησιμοποιηθεί ως συμπληρωματική της παραδοσιακής μεθόδου (συλλογή στοιχείων για ορισμένα χαρακτηριστικά του πληθυσμού) ή βοηθητικά αυτής (χρησιμοποίηση των υπαρχόντων μητρώων για έλεγχο των πρωτογενώς συλλεχθέντων στοιχείων).

Τα πλεονεκτήματα της μεθόδου συνοψίζονται στα ακόλουθα:

- Μείωση του κόστους της στατιστικής πληροφόρησης, επειδή τα στατιστικά στοιχεία δεν συλλέγονται από τις επιμέρους μονάδες του πληθυσμού, αλλά από έτοιμα μητρώα.
- Αυξημένη δυνατότητα συγκέντρωσης στατιστικών στοιχείων σε ετήσια βάση.
- Δυνατότητα συγκέντρωσης στατιστικής πληροφόρησης σε επίπεδο Δήμου ή Κοινότητας, επαρκούς για τις ανάγκες του περιφερειακού και τοπικού προγραμματισμού.
- 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 (Heldal J. et al., 1987).

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.

2.4 Στατιστική μονάδα – Στατιστικός Πληθυσμός

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

Ως παράδειγμα, σημειώνεται το ότι, όταν στις διάφορες απογραφές αναφερόμαστε σε νοικοκυριά, πρέπει να προσδιορίζεται με σαφήνεια ότι, άλλη είναι η έννοια του νοικοκυριού και άλλη αυτή της οικογένειας.

Οι στατιστικές μονάδες μπορεί να είναι **απλές** όπως π.χ. ένα πρόσωπο, ένα αντικείμενο ή μια μέρα, είναι δυνατό, όμως, να είναι **σύνθετες** και να αποτελούνται από περισσότερα αντικείμενα ή πρόσωπα, όπως η οικογένεια, η ημερήσια ή ετήσια παραγωγή ενός εργοστασίου κ.λπ.

Το σύνολο των στατιστικών μονάδων, των οποίων επιθυμούμε τη μελέτη ενός ή περισσοτέρων συγκεκριμένων χαρακτηριστικών ονομάζεται **πληθυσμός ή στατιστικός πληθυσμός.** Όπως η στατιστική μονάδα, έτσι και ο στατιστικός πληθυσμός, πρέπει να ορίζεται με ακρίβεια και σαφήνεια.

Ένα από τα βασικά στοιχεία που πρέπει να γνωρίζουμε για το στατιστικό πληθυσμό είναι τα όριά του, δηλαδή ποιες ακριβώς είναι οι στατιστικές μονάδες αυτού του συγκεκριμένου πληθυσμού που θα μελετηθεί.

Ο στατιστικός πληθυσμός μπορεί να είναι **άπειρος,** όπως η παραγωγή ενός προϊόντος, οι γεννήσεις βρεφών σε μία πόλη, οπότε είμαστε υποχρεωμένοι να περιορίσουμε τη μελέτη μας σε κάποιο χρονικό διάστημα, **ή πεπερασμένος,** όπως οι θεατές μιας θεατρικής παράστασης ή ενός ποδοσφαιρικού αγώνα.

Παράδειγμα

ΧΡΗΣΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΟ ΕΜΠΟΡΙΟ ΣΤΙΣ ΕΠΙΧΕΙΡΗΣΕΙΣ

ΜΕΘΟΔΟΛΟΓΙΑ ΕΡΕΥΝΑΣ

Σκοπός

Σκοπός της ετήσιας έρευνας είναι η εξακρίβωση στοιχείων για τη χρήση τεχνολογιών πληροφορικής και επικοινωνιών, για τη χρήση του Διαδικτύου και για το ηλεκτρονικό εμπόριο στις επιχειρήσεις με στόχο την υποβοήθηση εφαρμογής προγραμμάτων πολιτικής από το Δημόσιο και τον Ιδιωτικό τομέα.

Κάποιοι δείκτες από αυτήν την έρευνα χρησιμοποιούνται για να επιτύχουν τους στόχους του Προγράμματος Δράσης e-Europe 2005, το οποίο καθορίστηκε από τον Κανονισμό Ευρωπαϊκού Συμβουλίου 5197/03 , όπου ένα καθορισμένο σύνολο δεικτών συγκριτικής αξιολόγησης θα παρασχεθεί μέσω του Ευρωπαϊκού Στατιστικού Συστήματος. Η έρευνα αυτή, παρέχει επίσης διάφορους Διαρθρωτικούς Δείκτες.

Η έρευνα συμμορφώνεται με τον Κανονισμό No.808/2004 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 21ης Απριλίου 2004 που αφορά τις Κοινοτικές στατιστικές της κοινωνίας της πληροφορίας. Ο στόχος του παρόντος κανονισμού είναι η θέσπιση κοινού πλαισίου για τη συστηματική παραγωγή Κοινοτικών στατιστικών της κοινωνίας της πληροφορίας.

Γεωγραφική κάλυψη

Η έρευνα καλύπτει όλες της ελεύθερες περιοχές της Κυπριακής Δημοκρατίας.

Στατιστικές μονάδες

Επιχειρήσεις που απασχολούν τουλάχιστον 10 υπαλλήλους στις ακόλουθες κατηγορίες οικονομικής δραστηριότητας:

1. Μεταποιητικές Βιομηχανίες

2. Κατασκευές

3. Χονδρικό και Λιανικό Εμπόριο. Επισκευή αυτοκινήτων, μοτοσικλετών και ειδών προσωπικής ή οικιακής χρήσης.

4. Ξενοδοχεία και Εστιατόρια

5. Μεταφορές, Αποθήκευση και Επικοινωνίες

6. Διαχείριση Ακίνητης Περιουσίας, Εκμίσθωση και Επιχειρηματικές δραστηριότητες

7. Ψυχαγωγικές, Πολιτιστικές και Αθλητικές δραστηριότητες

<u>Περίοδος αναφοράς</u>

Ιανουάριος

<u>Περίοδος έρευνας</u>

2004: 18/02/2004 - 08/04/2004

2005: 28/01/2005 – 18/03/2005

<u>Στατιστικός πληθυσμός</u>

<u>Πληθυσμός στόχος (ο αριθμός επιχειρήσεων και απασχολουμένων στον πληθυσμό)</u>

2004: Επιχειρήσεις = 2.381

2005: Επιχειρήσεις = 2.427

Δείγμα

Δειγματοληπτική μέθοδος

Χρησιμοποιήθηκε η τυχαία στρωματοποιημένη δειγματοληψία. Για τη στρωματοποίηση χρησιμοποιήθηκαν δύο μεταβλητές, το μέγεθος και η οικονομική δραστηριότητα της επιχείρησης. Συνολικά περιλήφθηκαν 14 κατηγορίες οικονομικών δραστηριοτήτων και 3 κατηγορίες μεγέθους.

Οι 14 κατηγορίες οικονομικών δραστηριοτήτων που χρησιμοποιήθηκαν είναι: 15-21, 23-25, 26-28, 29-37, 45, 50, 51, 52, 55.1+55.2, 60-63, 64, 70+71+73+74, 72, 92.1+92.2. Οι 3 κατηγορίες μεγέθους που χρησιμοποιήθηκαν είναι:
Μικρές επιχειρήσεις (10-49 απασχολούμενοι),
Μεσαίες επιχειρήσεις (50-249 απασχολούμενοι),
Μεγάλες επιχειρήσεις (250+ απασχολούμενοι)

Λόγω του μικρού αριθμού μεγάλων και μεσαίων επιχειρήσεων, αποφασίστηκε όπως όλες οι μεγάλες και μεσαίες επιχειρήσεις συμπεριληφθούν δείγμα. Από τις μικρές επιχειρήσεις επιλέγηκε το 50% με τυχαία συστηματική παρακολούθηση.

Το μέγεθος του δείγματος ήταν 1329 επιχειρήσεις το 2004 και 1240 επιχειρήσεις το 2005.

Δειγματοληπτικό πλαίσιο

Αρχείο Επιχειρήσεων 2000

Τελικό δείγμα

2004: Επιχειρήσεις = 1.198

2005: Επιχειρήσεις = 1.088

Τεχνική συλλογής δεδομένων

Προσωπική συνέντευξη με το Διευθυντή Πληροφορικής της επιχείρησης, στο χώρο εργασίας.
Η έρευνα ήταν υποχρεωτική.

Ρυθμίσεις
Τα αποτελέσματα προσαρμόστηκαν σε εκτιμήσεις του πληθυσμού των επιχειρήσεων και των απασχολουμένων.

2.5 Πηγές συλλογής στατιστικών στοιχείων

Το πρώτο και βασικότερο στάδιο για τη μελέτη ενός φαινόμενου, 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 λαθεμένα, είναι φανερό ότι και η αξία της στατιστικής τους ανάλυσης θα είναι κι αυτή ψεύτικη ή λαθεμένη.

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

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

2.6 Μέθοδοι συγκέντρωσης στατιστικών δεδομένων

Για τη συλλογή στατιστικών στοιχείων εφαρμόζονται διάφορες μέθοδοι, από τις οποίες σπουδαιότερες είναι η απογραφή, η δειγματοληπτική μέθοδος και η μέθοδος των συνεχών εγγραφών.

2.6.1. Απογραφή

Η απογραφή συνίσταται στη συγκέντρωση στοιχείων από όλες τις στατιστικές μονάδες του πληθυσμού που επιθυμούμε να μελετήσουμε. Η απογραφή δεν είναι δυνατή, παρά μόνο όταν ο πληθυσμός είναι καλά ορισμένος και πεπερασμένος. Σε μία τέτοια περίπτωση η γνώση μας για αυτόν 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, κ.λπ.
- Οι Δημογραφικές απογραφές, στις οποίες συλλέγονται στοιχεία σχετικά με το φύλο, την ηλικία, το επάγγελμα, κ.λπ.
- Οι Οικονομικές απογραφές, όπου συγκεντρώνονται στοιχεία σχετικά με την οικονομική κατάσταση (δαπάνες, εισοδήματα,….) των πληθυσμών που απογράφονται.
- Οι Βιομηχανικές απογραφές, κατά τις οποίες δεν περιοριζόμαστε στην απλή καταγραφή βιομηχανικών καταστημάτων, αλλά συγκεντρώνουμε πληροφορίες που αφορούν τον κλάδο της οικονομικής τους δραστηριότητας, τον αριθμό των απασχολούμενων, το επίπεδο μηχανοργάνωσης κ.λπ.

Από όλες τις μορφές των απογραφών, *η απογραφή του πληθυσμού είναι η σημαντικότερη, γιατί αποτελεί την κύρια πηγή πληροφοριών πάνω στην άποψη των δημογραφικών, οικονομικών και κοινωνικών χαρακτηριστικών.*

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*
(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 κ.λπ.*

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

2.6.1.1. Μειονεκτήματα της απογραφής

Η μέθοδος της απογραφής παρουσιάζει τα ακόλουθα μειονεκτήματα:

(1) Απαιτεί μεγάλο κόστος. Για να πραγματοποιηθεί μια απογραφή, χρειάζεται ειδική προεργασία, καθώς και μεγάλο αριθμό εκπαιδευμένων απογραφέων. Για αυτήν την προεργασία, αλλά και την ημέρα της απογραφής απατιούνται μεγάλες δαπάνες. Αυτός είναι και ο λόγος που η απογραφή πληθυσμού γίνεται ανά δεκαετία, ενώ η οικονομικότερη απογραφή βιομηχανιών και βιοτεχνών ανά πενταετία.

(2) Η απογραφή δε γίνεται πάντα από ειδικευμένο προσωπικό.

(3) Επικαιρότητα αποτελεσμάτων. Ο μεγάλος αριθμός τόσο των πληροφοριών, όσο και των ατόμων που αποτελούν τον πληθυσμό στην απογραφή, δεν επιτρέπουν να έχουμε τη δυνατότητα να δημοσιεύσουμε έγκαιρα τα αποτελέσματα. Έτσι,

πολλές φορές, παρά τη μηχανογραφική επεξεργασία των στοιχείων, τα αποτελέσματα χάνουν την επικαιρότητά τους και περιορίζονται μόνο στην ιστορική τους αξία.

2.6.2. Δειγματοληπτική μέθοδος

Αντίθετα με τη γενική απογραφή, αν ο πληθυσμός που θέλουμε να μελετήσουμε από την άποψη ορισμένων ιδιοτήτων, *αποτελείται από μεγάλο πλήθος στατιστικών μονάδων* και αν *η μελέτη των ιδιοτήτων καταστρέφει τις μονάδες* του πληθυσμού που εξετάζουμε, τότε η γενική απογραφή είναι *πρακτικά αδύνατη οικονομικά και χρονικά ασύμφορη.*

Στις περιπτώσεις αυτές, εφαρμόζουμε τη μέθοδο της *δειγματοληψίας*, που συνίσταται στην προσπάθεια να γνωρίσουμε τις ιδιότητες ενός πληθυσμού εξετάζοντας από αυτόν μόνο ένα *δείγμα*.

Σκοπός των δειγματοληπτικών ερευνών, είναι να προσδιορίσουμε όσο γίνεται ακριβέστερα ιδιότητες του πληθυσμού, μελετώντας τα στοιχεία του δείγματος.

2.6.2.1. α) Πλεονεκτήματα της δειγματοληψίας

1. *Μεγαλύτερη ταχύτητα πληροφοριών.*
2. *Μεγαλύτερη ακρίβεια.*
3. *Μεγαλύτερη ευχέρεια εφαρμογής.*
4. *Χαμηλότερο κόστος.*
5. *Ολοκληρωτική δύναμη εφαρμογής της γενικής απογραφής*

2.6.2.2. β) Μειονεκτήματα της δειγματοληψίας

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

Αν οι μονάδες του πληθυσμού που ερευνούμε εμφανίζονται πολύ σπάνια, πρέπει να μελετήσουμε ένα σημαντικά μεγάλο *δείγμα,* αν θέλουμε να έχουμε αξιόπιστη εκτίμηση των χαρακτηριστικών ιδιοτήτων του πληθυσμού. Π.χ. αν θέλουμε με τη βοήθεια δείγματος να μελετήσουμε το ποσοστό καπνιστών από άτομα μιας πόλης που έχουν ηλικία πάνω από 80 χρόνια, πρέπει να πάρουμε ένα αρκετά μεγάλο δείγμα ώστε σε αυτό να περιλαμβάνεται ικανοποιητικός αριθμός ατόμων με ηλικία 80 ετών και πάνω, γιατί τα άτομα με τέτοια ηλικία αποτελούν μικρό ποσοστό στο συνολικό πληθυσμό της πόλης.

Ο σχεδιασμός και η εκτέλεση της δειγματοληψίας χρειάζονται *ιδιαίτερη προσοχή* και θα πρέπει να ακολουθηθεί αυστηρά η θεωρητική διαδικασία που επιβάλλεται για την επιλογή του δείγματος και η στατιστική ανάλυση των αποτελεσμάτων της δειγματοληψίας,

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

Άλλο βασικό μειονέκτημα της δειγματοληψίας είναι τα *δειγματοληπτικά σφάλματα,* για τα οποία μιλήσαμε παραπάνω.

Στο επόμενο κεφάλαιο θα αναφερθούμε εκτενέστερα στη μέθοδο της δειγματοληψίας.

3. Δειγματοληψία

3.1 Γενικές έννοιες της δειγματοληψίας

Για τη συλλογή στατιστικών στοιχείων που αναφέρονται σε ορισμένες χαρακτηριστικές ιδιότητες ενός πληθυσμού, όπως αναφέραμε σε προηγούμενο κεφάλαιο, εφαρμόζουμε διάφορες στατιστικές μεθόδους, οι σπουδαιότερες από τις οποίες είναι *η απογραφή, η* *δειγματοληψία* και *οι συνεχείς καταγραφές στατιστικών στοιχείων*.

Θέλοντας να τονίσουμε τη διαφορά μεταξύ δειγματοληψίας και απογραφής, εξηγήσαμε ότι η απογραφή συνίσταται στη συγκέντρωση στοιχείων τα οποία αναφέρονται σε ορισμένα χαρακτηριστικά γνωρίσματα από όλες τις στατιστικές μονάδες του πληθυσμού που επιθυμούμε να μελετήσουμε. Όταν, όμως, ο πληθυσμός που θέλουμε να μελετήσουμε από την άποψη ορισμένων χαρακτηριστικών ιδιοτήτων αποτελείται από μεγάλο πλήθος στατιστικών μονάδων, όπως συνήθως συμβαίνει στην πράξη, και αν η μελέτη των χαρακτηριστικών ιδιοτήτων καταστρέφει τις μονάδες του πληθυσμού που εξετάζουμε, τότε η γενική απογραφή είναι πρακτικά αδύνατη ή οικονομικά και χρονικά ασύμφορη.

Στις περιπτώσεις αυτές εφαρμόζουμε τη μέθοδο της *δειγματοληψίας,* που συνίσταται στην προσπάθεια να γνωρίσουμε τις ιδιότητες ενός πληθυσμού εξετάζοντας μόνο ένα μικρό αριθμό στατιστικών μονάδων του, το καλούμενο *δείγμα*, το οποίο επιλέγουμε κατά τέτοιο τρόπο, ώστε οι πληροφορίες, οι εκτιμήσεις και τα συμπεράσματα που θα πάρουμε από αυτό να ισχύουν για το συνολικό πληθυσμό από τον οποίο πήραμε το δείγμα. Το δείγμα θεωρείται ότι είναι *αντιπροσωπευτικό* του πληθυσμού, εφόσον είναι *τυχαίο*.

Το σύνολο των δυνατών αποτελεσμάτων της δειγματοληψίας ονομάζεται *δειγματικός χώρος*.

Η θεωρία της Δειγματοληψίας από πεπερασμένους πληθυσμούς χρησιμοποιείται στις <u>επισκοπήσεις</u> (ή <u>δειγματοληπτικές έρευνες</u>), που έχουν σκοπό την εκτίμηση ορισμένων χαρακτηριστικών ενός πεπερασμένου πληθυσμού (του <u>πληθυσμού στόχου</u>, όπως ονομάζεται), με τη χρήση ενός καταλλήλου επιλεγμένου (<u>πιθανοτικού</u>) δείγματος από αυτόν τον πληθυσμό.

Βέβαια, εκτός από εκτίμηση παραμέτρων και κατασκευή διαστημάτων εμπιστοσύνης και δειγματικών κατανομών, ασχολείται επιπλέον με το σχεδιασμό πιθανοτικών δειγμάτων (σχεδιασμός δειγματοληψίας).

3.2 Τεχνική των στατιστικών δειγματοληπτικών ερευνών

Όταν αποφασιστεί μια δειγματοληπτική έρευνα, χρειάζονται ορισμένες αποφάσεις, κυρίως τεχνικής φύσης, και προκαταρκτικές ενέργειες. Για το σκοπό αυτό, είναι αναγκαία η συνεργασία ενός στατιστικού, με μεγάλη εμπειρία στη διεξαγωγή των ερευνών, με ειδικούς επιστήμονες που είναι σχετικοί με το εξεταζόμενο πρόβλημα. Το σύνολο αυτό των επιστημόνων αποτελεί τη λεγόμενη *ομάδα εργασίας.* Αυτοί είναι και οι υπεύθυνοι για το σχεδιασμό, την εκτέλεση και ανάλυση των αποτελεσμάτων της έρευνας.

Οι υπεύθυνοι της έρευνας θα πρέπει να προσδιορίσουν τη φύση του προς εξέταση προβλήματος και τη διατύπωση των αντικειμενικών σκοπών της.

Η ομάδα εργασίας πρέπει να μελετήσει παρόμοιες έρευνες που έχουν γίνει στο παρελθόν, τόσο στο εσωτερικό της χώρας όσο και στο εξωτερικό.

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

Στο σχεδιασμό της δειγματοληπτικής έρευνας βασική θέση έχουν η κατάρτιση του πλαισίου, το ερωτηματολόγιο, το μέγεθος του δείγματος και η μέθοδος δειγματοληψίας.

3.3. Κατάρτιση πλαισίου δειγματοληψίας

Για τη διενέργεια μιας δειγματοληψίας είναι απαραίτητος ένας κατάλογος που θα περιέχει όλες τις μονάδες που αποτελούν τον προς εξέταση πληθυσμό. Ο κατάλογος αυτός, που ονομάζεται *πλαίσιο,* χρησιμεύει για την επιλογή του δείγματος. Το πλαίσιο για να χρησιμοποιηθεί θα πρέπει να πληρεί ορισμένους όρους, π.χ. να είναι ενημερωμένο, να μην έχει παραλείψεις, διπλοεγγραφές κ.λπ.

3.4. Μέγεθος δείγματος

Υπάρχουν αριθμητικές τεχνικές για να καθορίσουμε τα μεγέθη του δειγμάτων για όλες τις δειγματοληπτικές τεχνικές, π.χ. στρωματοποιημένη ή κατά συστοιχίες, και διδάσκονται σε μαθήματα για δειγματοληπτικές έρευνες (survey sampling). Γενικά όσο πιο μεγάλο το μέγεθος του δείγματος τόσο πιο ακριβή αναμένονται οι εκτιμητές του δείγματος να είναι.

3.4.1 Δειγματοληπτικά – Μη Δειγματοληπτικά λάθη

Δύο βασικοί τύποι λαθών μπορούν να εμφανισθούν όταν το δείγμα των παρατηρήσεων παίρνεται από έναν πληθυσμό: δειγματοληπτικό λάθος και μη-δειγματοληπτικό λάθος.

Το **δειγματοληπτικό λάθος** αναφέρεται σε διαφορές μεταξύ του δείγματος και του πληθυσμού οι οποίες υπάρχουν επειδή αυτές οι συγκεκριμένες παρατηρήσεις έτυχε να επιλεχθούν.

Ένας άλλος τρόπος για να δούμε αυτό είναι: οι διαφορές αποτελεσμάτων για διαφορετικά δείγματα (ιδίου μεγέθους) οφειλόμενη καθαρά δειγματοληπτικό λάθος:

Π.χ. Δύο δείγματα μεγέθους 10 από 1.000 εταιριών. Εάν συνέβη να πάρουμε τα δεδομένα με το υψηλότερα εισοδήματα στο πρώτο μας δείγμα και όλα τα χαμηλότερα στο δεύτερο, αυτή η διαφορά οφείλεται καθαρά σε δειγματοληπτικό λάθος.

Αύξηση του δείγματος, μειώνει το δειγματοληπτικό λάθος.

Τα **μη-δειγματοληπτικά λάθη** είναι πιο σοβαρά και οφείλονται σε λάθη κατά την απόκτηση των δεδομένων ή οφείλεται στην ακατάλληλη επιλογή των δειγματοληπτικών παρατηρήσεων.

Τρεις τύποι μη-δειγματοληπτικών λαθών:

- Λάθη κατά την απόκτηση των δεδομένων.
- Λάθη από αναπάντητα ερωτηματολόγια.
- Μεροληψία στην επιλογή του δείγματος.

Σημειώστε**:** Αύξηση του δείγματος, **δε** μειώνει το μη-δειγματοληπτικό λάθος.

3.4.2 Ερωτηματολόγιο

Το *ερωτηματολόγιο* είναι ένα έντυπο στο οποίο καταχωρούνται οι λαμβανόμενες πληροφορίες από τις ερευνώμενες δειγματοληπτικές μονάδες. Το ερωτηματολόγιο πρέπει να συνταχτεί με μεγάλη προσοχή από το τεχνικό προσωπικό της έρευνας.

Με τον καιρό, μεγάλη επιστημονική μελέτη έχει γίνει για τον σχεδιασμό δειγματοληπτικών ερωτήσεων. Βασικές αρχές σχεδιασμού:

1. *Να είναι σύντομο.*
2. *Να είναι εύκολο στην απάντηση και να μπορεί εύκολα να ταχυδρομηθεί.*
3. *Τα διάφορα ερωτήματα να έχουν μια λογική ακολουθία.*
4. *Να είναι δυνατή η επεξεργασία των στοιχείων.*
5. *Να έχει κατάλληλο σχηματικό μέγεθος.*
6. *Οι ερωτήσεις να είναι σύντομες, απλές και ξεκάθαρες.*
7. *Αρχίζουμε με δημογραφικές ερωτήσεις, ώστε να βοηθήσουμε τους ερωτηθέντες να εξοικειωθούν και να αισθάνονται άνετα.*
8. *Χρησιμοποιούμε δυαδικές (ναι/όχι) και ερωτήσεις πολλαπλών επιλογών.*
9. *Να χρησιμοποιούνται ελεύθερες (open – ended) και όχι ερωτήσεις που πιέζουν ή καθοδηγούν (leading) το συμμετέχοντα.*
10. *Να έχει σχεδιαστεί καλά.*

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

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

3.4.2.1 Λάθη κατά την Απόκτηση των Δεδομένων.

Εμφανίζονται από την καταγραφή λανθασμένων απαντήσεων, οφειλόμενη σε:

- λανθασμένες μετρήσεις παίρνονται εξαιτίας ελαττωματικού εξοπλισμού
- λάθη γίνονται κατά την μεταγραφή από βασικές πηγές
- λανθασμένη καταγραφή των δεδομένων οφειλόμενη σε παρερμηνεία των όρων
- λανθασμένες απαντήσεις σε ερωτήσεις οφειλόμενες σε ευαίσθητα θέματα.

3.4.2.2 Λάθη από αναπάντητα ερωτηματολόγια

Αναφέρεται σε λάθος (ή *μεροληψία*) που εισάγεται όταν απαντήσεις δεν πετυχαίνονται από κάποια μέλη του δείγματος, π.χ. οι παρατηρήσεις του δείγματος που συλλέγονται ενδεχομένως δεν αντιπροσωπεύουν τον πληθυσμό τον οποίο εξετάζουμε. Το *Ποσοστό Ανταπόκρισης* (η αναλογία των ανθρώπων που συμμετέχει στην σφυγμομέτρηση) είναι πολύ βασική παράμετρος και βοηθάει στην κατανόηση της εγκυρότητας της σφυγμομέτρησης και στην κατανόηση πηγών με λάθη από αναπάντητα ερωτηματολόγια.

<u>Μεροληψία</u> συμβαίνει όταν το δειγματοληπτικό σχέδιο είναι τέτοιο που κάποια μέλη του πληθυσμού δεν μπορούν να επιλεχθούν και δεν συμπεριλαμβάνονται μέσα στο δείγμα.

3.4.2.3 Διατύπωση των ερωτήσεων

Υπάρχουν ορισμένες καθοδηγητικές αρχές σχετικά με τη διατύπωση των ερωτήσεων.
Μερικές τέτοιες αρχές διατύπωσης είναι:

- *Απλότητα της γλώσσας.* Ένα μεγάλο ποσοστό του πληθυσμού δεν έχει συνεχίσει την εκπαίδευσή του πέρα από τη στοιχειώδη εκπαίδευση και μερικοί ίσως είναι τελείως αγράμματοι.
- *Σαφήνεια.* Να αποφεύγουμε τις αμφιβολίες ως προς το τι εννοούμε. Γι' αυτό πρέπει τα ερωτήματα να είναι απλά διατυπωμένα.
- *Αποφυγή ερωτήσεων που οδηγούν σε αποκρύψεις.*
- *Αποφυγή διατύπωσης που προκαλεί ψυχολογικές αντιδράσεις.* Δηλαδή οι ερωτήσεις πρέπει να είναι κατά τέτοιο τρόπο διατυπωμένες ώστε να μην προσβάλλουν τον ερωτώμενο.
- *Αποφυγή κατευθυνόμενων ερωτήσεων.* Η κατευθυνόμενη ερώτηση σκοπό έχει να

οδηγήσει τον ερωτώμενο σε επιθυμητή απάντηση.

3.4.2.4 Πρόσληψη προσωπικού και εκπαίδευση

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

Οι κατηγορίες του προσωπικού που θα προσληφθεί πρέπει να είναι:

1. *Ειδικευμένο προσωπικό,* που θα ασχοληθεί με το σχεδιασμό, την έρευνα, την εποπτεία και την ανάλυση των αποτελεσμάτων.
2. *Υπάλληλοι γραφείου,* που θα ασχοληθούν με την επεξεργασία κ.λπ.
3. *Ερευνητές,* που θα ασχοληθούν με τη συμπλήρωση του ερωτηματολογίου.

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

3.5 Μέθοδοι συλλογής στοιχείων στις δειγματοληπτικές έρευνες

Αυτές είναι η παρατήρηση, οι προσωπικές συνεντεύξεις και το ταχυδρομείο.

3.5.1. Παρατήρηση

Η παρατήρηση ως μέθοδος συγκέντρωσης στοιχείων χρησιμοποιείται αρκετά στις φυσικές επιστήμες, ενώ στις στατιστικές έρευνες είναι πολύ λίγο διαδεδομένη. Χρησιμοποιείται κυρίως από τους ανθρωπολόγους για τη μελέτη υπανάπτυκτων κοινωνιών, στην ανάλυση διατροφής σε υποανάπτυκτο κράτος, στην έρευνα συμπεριφοράς του πληθυσμού μιας χώρας, εάν εφαρμόζονται ορισμένες οδηγίες, π.χ. διαφώτιση για την πρόληψη των τροχαίων ατυχημάτων και αξιολόγησή της, με παρατηρήσεις στους δρόμους για τυχόν παραβάσεις του κώδικα οδικής κυκλοφορίας από τους οδηγούς και τους πεζούς.

3.5.2 Συνέντευξη

Η συνέντευξη είναι ο καλύτερος τρόπος συλλογής στατιστικών στοιχείων και χρησιμοποιείται σήμερα πάρα πολύ στις δειγματοληπτικές έρευνες. Σύμφωνα με τη μέθοδο αυτή, η συμπλήρωση του ερωτηματολογίου γίνεται από ειδικά εκπαιδευμένα άτομα που λέγονται ερευνητές ή ερευνήτριες. Το έργο των ερευνητών συνίσταται:

- Στον εντοπισμό των μονάδων του δείγματος.
- Στην πραγματοποίηση των συνεντεύξεων.
- Στον τρόπο υποβολής των ερωτήσεων.
- Στην καταγραφή των απαιτήσεων. Η καταγραφή φαίνεται απλή εργασία αλλά παρουσιάζονται λάθη που οφείλονται, κυρίως, στο ότι το έργο είναι κουραστικό και προκαλεί διάσπαση προσοχής, ενώ ο τρόπος συμπλήρωσης του ερωτηματολογίου πολλές φορές είναι πολύπλοκος κ.λπ.

3.6 Μέθοδοι διενέργειας της δειγματοληψίας

Για τη διενέργεια των δειγματοληπτικών ερευνών χρησιμοποιούμε πολλές μεθόδους.
Η εκλογή της μεθόδου εξαρτάται:

1. Από τη ζητούμενο βαθμό ακριβείας των αποτελεσμάτων.
2. Από τα χρονικά και χρηματικά περιθώρια της έρευνας.
3. Από τη μεταβλητικότητα των μονάδων του πληθυσμού που ερευνάμε.
4. Από την ύπαρξη καταλόγων των μονάδων του πληθυσμού.
5. Από τη δυνατότητα που έχει ο πληθυσμός να μπορεί να διαιρεθεί σε υποπληθυσμούς με μεγάλη ομοιογένεια.

Έτσι, με βάση τα παραπάνω κριτήρια, θα αποφασιστεί αν θα εφαρμοστεί *τυχαία δειγματοληψία, η στρωματοποιημένη δειγματοληψία* ή κάποια άλλη μορφή δειγματοληψίας.

Πάντως, για κάθε μέθοδο έχει αναπτυχθεί και η αντίστοιχη θεωρία για τη συλλογή και επεξεργασία των στοιχείων που παίρνουμε με την παρατήρηση. Η μέθοδος όμως που χρησιμοποιούμε περισσότερο στην πράξη είναι η τυχαία δειγματοληψία, η οποία υπακούει στους νόμους της τύχης.

3.6.1. Απλή τυχαία δειγματοληψία

Όταν κάθε στατιστική μονάδα του πληθυσμού που μελετούμε έχει την ίδια πιθανότητα να επιλεγεί σα μονάδα του δείγματος, τότε έχουμε **απλή τυχαία δειγματοληψία.** Η πιθανότητα επιλογής μιας μονάδας του πληθυσμού στο δείγμα, πρέπει να είναι ανεξάρτητη από την επιλογή κάποιας άλλης μονάδας.

Προκειμένου να επιλέξουμε ένα δείγμα με n μονάδες από ένα πληθυσμό, οποίος αποτελείται από N μονάδες, η τυχαία επιλογή γίνεται με μια από τις ακόλουθες μεθόδους:

1. *Μέθοδος της κλήρωσης.*

Υποθέτουμε ότι έχουμε στη διάθεσή μας ένα φάκελο με καρτέλες στις οποίες περιέχονται γραμμένα τα ονόματα π.χ. 1.000 ασφαλιστικών εταιριών. Για να μελετήσουμε τα κέρδη τους, παίρνουμε δείγμα από 50 ονόματα εταιριών ως ακολούθως:

Τα ατομικά στοιχεία των 1.000 ασφαλιστικών εταιριών είναι δυνατό να αντιγραφούν σε ισομεγέθη τεμάχια από χαρτόνι ή σε ομοιόμορφα σφαιρίδια και ύστερα να ανασυρθούν από ένα άτομο τυχαία 50 καρτέλες ή 50 σφαιρίδια από μια κάλπη στην οποία θα έχουμε τοποθετήσει το σύνολο των 1.000 καρτελών ή σφαιριδίων.

Για να μην αλλοιώνεται σε κάθε κλήρωση η σύνθεση της κάλπης, πρέπει οι αριθμοί οι οποίοι έχουν κληρωθεί, αφού σημειωθούν, να ξαναριχτούν στην κάλπη πριν από κάθε νέα κλήρωση, γιατί τότε και μόνο τότε κάθε καρτέλα ή κάθε σφαιρίδιο έχει την ίδια πιθανότητα να συμπεριληφθεί στο δείγμα. Με αυτή την έννοια, το δείγμα των 50 εταιριών προέρχεται από ολόκληρο τον πληθυσμό των 1.000 μονάδων και κάθε εταιρία από τους1.000 έχει την ίδια πιθανότητα να επιλεγεί.

Το δείγμα που παίρνουμε με αυτό τον τρόπο λέγεται *δείγμα με επανατοποθέτηση.*

Το δείγμα των 50 μονάδων είναι δυνατόν να επιλεγεί και ως εξής:

Εξάγουμε μια σφαίρα ή μια καρτέλα από την κάλπη και σύμφωνα με την ένδειξή της εντοπίζουμε την αντίστοιχη μονάδα του πληθυσμού. Ύστερα, χωρίς επανατοποθέτηση της σφαίρας ή της καρτέλας που βγάλαμε από την κάλπη, εξάγουμε μια νέα σφαίρα ή καρτέλα για τον εντοπισμό της δεύτερης μονάδας του πληθυσμού που επιλέξαμε κ.λπ. Εκτελώντας n = 50 τέτοιες εξαγωγές, έχουμε τους ακόλουθους αριθμούς που έχουν εκλεγεί από τον πληθυσμό: $x_1,x_2,\dots ,x_{50}$. Αυτοί θα αποτελέσουν το τυχαίο δείγμα το οποίο λέγεται *δείγμα χωρίς επανατοποθέτηση.*

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

2. Μέθοδος τυχαίων αριθμών

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

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

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

Ο πίνακας τυχαίων αριθμών κατασκευάζεται ως εξής:

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

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

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

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

Κατά την επιλογή των 250 αριθμών, δεν λαμβάνουμε υπόψη μας τους τετραψήφιους αριθμούς που είναι μεγαλύτεροι από το πλήθος των μονάδων του πληθυσμού (στην περίπτωσή μας από τον αριθμό 5.000) ούτε και τους αριθμούς που εμφανίζονται για δεύτερη φορά.

Υποθέτουμε ότι επιλέγεται με αμερόληπτο τρόπο η πρώτη στήλη που έχει ως πρώτο αριθμό το 1528, ως δεύτερο τον αριθμό 4864, ως τρίτο τον αριθμό 8609, ο οποίος και παραλείπεται, γιατί 8609 > 5000 κ.λπ. και συνεχίζουμε μέχρι να πάρουμε και τους 250 αριθμούς.

Όταν εξαντληθούν όλες οι στήλες, επαναλαμβάνουμε την εκλογή των αριθμών, αρχίζοντας π.χ. από το δεύτερο ψηφίο κάθε στήλης και καταλήγοντας στο πρώτο ψηφίο της άλλης στήλης κ.λπ. Επειδή καθένας από τους παραπάνω 250 αριθμούς αντιστοιχεί σε κάποια μονάδα του πληθυσμού, μπορούμε να πάρουμε τις μονάδες αυτές ως δείγμα που, λόγω της τυχαίας επιλογής των μονάδων του, ονομάζεται *τυχαίο.*

Το πλεονέκτημα της απλής τυχαίας δειγματοληψίας είναι η ευκολία και η ταχύτητά της, τόσο στο σχεδιασμό όσο και στην εκτέλεση και εφαρμόζεται σε έρευνες που η καταμέτρηση του *N* είναι εύκολη.

3.6.2 Δειγματοληψία κατά στρώματα

Όταν οι χαρακτηριστικές ιδιότητες ενός πληθυσμού που θέλουμε να εξετάσουμε με τη βοήθεια ενός δείγματος παρουσιάζουν μεγάλη ανομοιογένεια, διαιρούμε το συνολικό πληθυσμό σε ένα πλήθος ομάδων (στρωμάτων), που κάθε μια παρουσιάζει όσο το δυνατό μεγαλύτερη ομοιογένεια στατιστικών μονάδων, και παίρνουμε από κάθε στρώμα με τυχαίο τρόπο χωρίς καμιά εξαίρεση ορισμένες μονάδες προκειμένου να σχηματίσουμε το ενιαίο συνολικό δείγμα.

Η μέθοδος αυτή ονομάζεται *στρωματοποιημένη δειγματοληψία* και μας δίνει, με δεδομένο μέγεθος δείγματος, περισσότερο αξιόπιστα αποτελέσματα *από* εκείνα της απλής τυχαίας δειγματοληψίας. Στην απλή τυχαία δειγματοληψία οι ακραίες τιμές μπορούν να αποκλειστούν, ενώ στη στρωματοποιημένη δειγματοληψία θα αντιπροσωπεύονται.

Αν π.χ. θέλουμε να μελετήσουμε με τη βοήθεια ενός δείγματος το ύψος των ενοικίων στην πόλη των Αθηνών, θα πρέπει να χωρίσουμε την Αθήνα σε περιοχές ή ζώνες που καθεμιά να παρουσιάζει όσο το δυνατό μεγαλύτερη ομοιογένεια. Κατόπιν, από κάθε περιοχή να πάρουμε τυχαίο δείγμα, γιατί η διαφορά των ενοικίων από περιοχή σε περιοχή είναι πάρα πολύ μεγάλη. Όσον αφορά τον τρόπο επιλογής του δείγματος από κάθε στρώμα, έχουμε να παρατηρήσουμε τα εξής:

Έστω ότι ο συνολικός πληθυσμός, ο οποίος αποτελείται από *N* μονάδες, διαιρείται σε m στρώματα και κάθε στρώμα περιέχει $N_1, N_2, \ldots, N_m$ μονάδες.
Στη συνέχεια, παίρνουμε από κάθε στρώμα ένα τυχαίο δείγμα n_i και έχουμε:

$$n_r + n_2 + n_3 + \ldots + n_m = n$$

Το συνολικό μέγεθος του δείγματος *n* κατανέμεται στα m στρώματα κατά τους εξής τρόπους:

1) Αναλογικά ως προς το μέγεθος του κάθε στρώματος με τη βοήθεια της παρακάτω σχέσης:

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

Ένα *στρωματοποιημένο τυχαίο δείγμα* πετυχαίνεται αν χωρίσουμε τον πληθυσμό σε <u>αμοιβαία αποκλειόμενα σύνολα</u>, ή <u>στρώμα</u>, και μετά επιλέγουμε απλά τυχαία δείγματα από κάθε στρώμα.

Στρώμα 1: Φύλο	Στρώμα 2: Ηλικία	Στρώμα 3: Επάγγελμα
Αρσενικό	< 20	Επαγγελματίας
Θηλυκό	20-30	Υπάλληλος
	31-40	Εργάτης
	41-50	Λοιπά
	51-60	
	> 60	

* Μπορούμε να θέλουμε σχετικά με τον συνολικό πληθυσμό, να βγάλουμε συμπεράσματα <u>μέσα σε ένα στρώμα</u> ή να βγάλουμε συμπεράσματα <u>διασταυρώνοντας στρώματα</u>.

Αφού στρωματοποιήσουμε τον πληθυσμό, μπορούμε να χρησιμοποιήσουμε ***απλή τυχαία δειγματοληψία*** για να παράγουμε το πλήρες δείγμα:

Income Category	Population Proportion	Sample Size n = 400	n = 1000
under $25,000	25%	100	250
$25,000 - $39,999	40%	160	400
$40,000 – $60,000	30%	120	300
over $60,000	5%	20	50

Εάν μόνο έχουμε την δυνατότητα να επιλέξουμε 400 ανθρώπους συνολικά, θα επιλέγαμε 100 από αυτούς από την ομάδα των χαμηλών εισοδημάτων. Εάν επιλέξουμε 1000 ανθρώπους, θα επιλέγαμε 50 από αυτούς από την κατηγορία με τα υψηλότερα εισοδήματα.

2) Σύμφωνα με τη λεγόμενη *δυσανάλογη δειγματοληψία.* Αν το κόστος επιλογής των μονάδων του δείγματος δεν διαφέρει στα διάφορα στρώματα και εάν οι διακυμάνσεις (S_r) των στρωμάτων διαφέρουν μεταξύ τους, τότε παίρνουμε περισσότερες μονάδες από εκείνα τα στρώματα που έχουν τη μεγαλύτερη διασπορά.

Στην περίπτωση αυτή, το μέγεθος του δείγματος σε κάθε στρώμα δίνεται από τη σχέση:

$$ni = \frac{NiBi}{\Sigma NiSi} * n$$

Δηλαδή το μέγεθος n_i πρέπει να είναι ανάλογο προς το μέγεθος του στρώματος N_i και προς την τυπική απόκλιση B_i του στρώματος i.

3) Σύμφωνα με τη λεγόμενη *άριστη κατανομή.* Όταν και το κόστος επιλογής των μονάδων του δείγματος στα διάφορα στρώματα και η διασπορά μέσα στα στρώματα διαφέρει, τότε η κατανομή του δείγματος *n* στα διάφορα στρώματα δίνεται από τη σχέση:

$$ni = \frac{\frac{NiSi}{\sqrt{Ci}}}{\frac{NiSi}{\Sigma\sqrt{Ci}}}$$

όπου C_i το κατά μονάδα κόστος στο στρώμα i.

3.6.3. Κατά συστοιχίες δειγματοληψία

Μία *κατά συστοιχίες δειγματοληψία* είναι ένα απλό τυχαίο δείγμα ομάδων ή συστοιχιών (σε αντίθεση με την απλή τυχαία δειγματοληψία από μεμονωμένα άτομα). Αυτή η μέθοδος είναι χρήσιμη όταν είναι δύσκολο ή κοστίζει να έχουμε μία πλήρης λίστα των μελών του πληθυσμού ή όταν τα στοιχεία του πληθυσμού είναι ευρέως διάσπαρτα γεωγραφικός. Η κατά συστοιχίες δειγματοληψία ενδέχεται να αυξήσει το λάθος του δείγματος οφειλόμενο στις ομοιότητες μεταξύ των μελών των ομάδων.

3.6.4. Επιφανειακή δειγματοληψία

Σύμφωνα με την επιφανειακή δειγματοληψία, μια επιφάνεια, π.χ. η περιφέρεια της πρωτεύουσας, υποδιαιρείται σε μικρότερες περιοχές κάθε περιοχή υποδιαιρείται σε μικρότερα τμήματα που ονομάζονται *οικοδομικά τετράγωνα.* Με τη βοήθεια του πλαισίου και με τυχαίο τρόπο επιλέγονται ορισμένα οικοδομικά τετράγωνα.

Καταγράφουμε τα νοικοκυριά που ανήκουν σε αυτά και επιλέγουμε με τυχαίο τρόπο ορισμένα από αυτά τα νοικοκυριά, τα οποία αποτελούν το λεγόμενο επιφανειακό δείγμα. Η επιφανειακή δειγματοληψία διεξάγεται συνήθως σε δύο, τρία ή τέσσερα στάδια.

Στη δισταδιακή δειγματοληψία, πρωτογενής μονάδα είναι το οικοδομικό τετράγωνο (μονάδα επιφανείας) και δευτερογενής η κατοικία.

Στη τρισταδιακή δειγματοληψία, πρωτογενής μονάδα είναι ο Δήμος ή η Κοινότητα, δευτερογενής το οικοδομικό τετράγωνο και τελική η κατοικία.

3.6.5. Δειγματοληψία κατά ομάδες

Σύμφωνα με τη μέθοδο αυτή, ο συνολικός πληθυσμός διαιρείται σε ομάδες και με τυχαίο τρόπο επιλέγουμε μόνο ορισμένες ομάδες και στο δείγμα περιλαμβάνουμε όλες τις δειγματοληπτικές μονάδες των ομάδων που έχουν επιλεγεί.

Η μέθοδος αυτή είναι πολύ χρήσιμη στην περίπτωση που δεν υπάρχουν πλαίσια των μονάδων του πληθυσμού που θέλουμε να μελετήσουμε, υπάρχουν όμως πλαίσια των ομάδων.

Με τη μέθοδο αυτή περιορίζεται το κόστος και ο χρόνος της έρευνας. Αν π.χ. θέλουμε να επιλέξουμε ένα δείγμα στην περιφέρεια της πρωτεύουσας για να μελετήσουμε τη ζήτηση ηλεκτρικού ρεύματος νοικοκυριών, μπορούμε να ομαδοποιήσουμε τις κατοικίες σε οικοδομικά τετράγωνα, επιλέγουμε με τυχαίο τρόπο ορισμένα οικοδομικά τετράγωνα και στη συνέχεια εξετάζουμε όλα τα νοικοκυριά που ανήκουν σε όλες τις ομάδες που έχουν επιλεγεί.

3.6.6. Μεροληπτική δειγματοληψία

Στην περίπτωση αυτή, εμφανίζονται συντελεστές μεροληπτικής λήψης ενός δείγματος, δηλαδή σε μία ή περισσότερες κατηγορίες δειγματοληπτικών μονάδων δίνεται συνειδητά μεγαλύτερη ευκαιρία να συμπεριληφθούν στο επιλεγόμενο δείγμα σε σύγκριση με άλλες κατηγορίες.

Αν π.χ. πρόκειται να ερευνηθεί η εξέλιξη του αναστήματος και του βάρους των ελληνοπαίδων σε συνάρτηση με την ηλικία τους, θα ήταν μεροληπτική η λήψη δείγματος μόνο από τους μαθητές του Αμερικανικού Κολλεγίου ή άλλου σχολείου στο οποίο σπουδάζουν συνήθως μαθητές πλουσίων οικογενειών. Επίσης παράδειγμα μεροληπτικής δειγματοληψίας είναι η διεξαγωγή τιμοληψίας σε καταστήματα πολυτελείας ή σε καταστήματα πολύ λαϊκών συνοικιών.

Κάθε έμπειρος ερευνητής πρέπει να γνωρίζει τις πηγές μεροληψίας που υπάρχουν στην επιλογή του δείγματος. Μια από τις πρώτες προσπάθειες της ομάδας έρευνας είναι η βαθιά γνώση του προς εξέταση θέματος, για να μπορέσει να εξαλείψει όλες τις πηγές μεροληπτικής επιλογής του δείγματος.

3.6.7. Συστηματική δειγματοληψία

Η μέθοδος αυτή συνίσταται στην καταχώριση όλων των στατιστικών μονάδων του πληθυσμού σε καταλόγους (μητρώα) κατά αριθμητική ή αλφαβητική σειρά από 1 μέχρι *N* και στην επιλογή *n* μονάδων του δείγματος.

Αν π.χ. από ένα κατάλογο (μητρώο) που περιέχει 1.000 μονάδες θέλουμε να πάρουμε δείγμα από 50 μονάδες, ενεργούμε με τον παρακάτω τρόπο:

Υπολογίζουμε το πηλίκο:

$$l = \frac{N}{n} = \frac{1000}{50} = 20$$

Η πρώτη μονάδα του δείγματος βρίσκεται από τυχαία επιλογή ενός αριθμού από το 1 μέχρι το 20.

Αν Π.χ. ο αριθμός 8 είναι τυχαίος αριθμός, οι μονάδες του πληθυσμού που έχουν αύξοντα αριθμό 8.20 + 8.40 + 8.60 + 8 κ.λπ. αποτελούν το ζητούμενο δείγμα. Δηλαδή, αν *χ* είναι ο πρώτος τυχαίος αριθμός του δείγματος, οι μονάδες του μητρώου που έχουν αριθμό:

$$x,\ x + \lambda,\ x + 2\lambda,\ x + 3\lambda,\ \ldots,\ x + (n - 1)\lambda$$

μας δίνουν το ζητούμενο δείγμα με τη μέθοδο της συστηματικής δειγματοληψίας.

3.7 Ορισμός τυχαίου δείγματος

Την έννοια του τυχαίου δείγματος με μαθηματική έκφραση μπορούμε να τη διατυπώσουμε ως εξής:

Έστω ότι έχουμε ένα πληθυσμό που περιγράφεται από την τυχαία μεταβλητή *X* και έχει συνάρτηση πιθανότητας ή συνάρτηση πυκνότητας πιθανότητας *f(x),* ανάλογα αν η μεταβλητή *X* είναι ασυνεχής ή συνεχής.

Από τον παραπάνω πληθυσμό, τον οποίο θέλουμε να εξετάσουμε, παίρνουμε με μια μέθοδο ένα δείγμα μεγέθους *n* παρατηρήσεων:

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

Έστω ότι ο μηχανισμός της επιλογής των *n* μονάδων του δείγματος επαναλαμβάνεται με τον ίδιο τρόπο για δεύτερη φορά. Τότε οι μονάδες του δείγματος που θα επιλεγούν θα είναι:

$x'_1, x'_2\, x'_3 \, . \, . \, . \, , x'_i \, . \, . \, . \, , x'_n$

Εάν ο μηχανισμός αυτός επιλογής των *n* μονάδων του δείγματος επαναληφθεί για τρίτη, τέταρτη και γενικά για μιοστή φορά, θα έχουμε μια ακολουθία δειγμάτων ως εξής:

1ο δείγμα $x_1, \; x_2, \; x_3, \; \ldots, \; x_i, \; \ldots, \; x_n$

2ο δείγμα $x'_1, \; x'_2, \; x'_3, \; \ldots, \; x'_i, \; \ldots, \; x'_n$

3ο δείγμα $x''_1, \; x''_2, \; x''_3, \; \ldots, \; x''_i, \; \ldots, \; x''_n$

4ο δείγμα $x'''_1, \; x'''_2, \; x'''_3, \; \ldots, \; x'''_i, \; \ldots, \; x'''_n$

……

……

……

μιοστό δείγμα $\overset{(m)}{x_1}, \; \overset{(m)}{x_2}, \; \overset{(m)}{x_3}, \; \ldots, \; \overset{(m)}{x_i}, \; \ldots, \; \overset{(m)}{x_n}$

Το σύνολο των τιμών της πρώτης μονάδας του δείγματος που επιλέγουμε, δηλαδή οι τιμές της πρώτης στήλης:

$$x_1, x_1'\, x_1'', \ldots ,x_1''', \overset{(m)}{x_1}$$

ορίζουν μια τυχαία μεταβλητή, την x_1, της οποίας η συνάρτηση πιθανότητας ή πυκνότητας πιθανότητας είναι f_1 (x_1).

Επίσης, οι τιμές της δεύτερης στήλης, δηλαδή της δεύτερης μονάδας δείγματος που θα επιλεγεί:

$$x_2,\ x_2'\ x_2'',\dots\ ,x_2''',\ x_2^{(m)}$$

θεωρούνται σαν τιμές μιας τυχαίας μεταβλητής X_2, που έχει συνάρτηση πιθανότητας ή πυκνότητας πιθανότητας την $f_2(x_2)$, και γενικά οι τιμές στήλης ορίζουν μια τυχαία μεταβλητή X_i, που έχει συνάρτηση πιθανότητας την $f_i(x_i)$.

Το δείγμα των τυχαίων μεταβλητών

$x_1,x_2,x_3,\dots\ ,x_i,\dots\ ,x_n$

λέγεται τυχαίο, αν η μέθοδος επιλογής είναι τέτοια, ώστε:

$$f_1(x_1) = f_2(x_2) = f_3(x_3) = \dots = f_n(x_n) = f(x)$$

και επιπλέον:

$$f(x_1,\ x_2,\ x_3,\dots,\ x_n) = f_1(x_1) * f_2(x_2) * \dots\dots * f_n(x)$$

όπου $f(x_1,\ x_2,\dots,\ x_n)$ είναι η από κοινού συνάρτηση πιθανότητας ή συνάρτηση πυκνότητας πιθανότητας των τυχαίων μεταβλητών x_1, x_2,..., x_n.

Αν ισχύουν οι σχέσεις, λέμε ότι έχουμε ένα τυχαίο δείγμα ανεξαρτήτων μεταβλητών.

Την έννοια του τυχαίου δείγματος μπορούμε να τη διατυπώσουμε και ως εξής: Έστω ότι έχουμε ένα πληθυσμό που αποτελείται από Ν μονάδες. Έστω , ακόμη ότι από τον πληθυσμό αυτό θέλουμε να επιλέξουμε ένα δείγμα n μονάδων. Πόσα δυνατά δείγματα των n μονάδων μπορούμε να επιλέξουμε από τον πληθυσμό που έχει Ν μονάδες;

Η απάντηση είναι: όσοι είναι οι συνδυασμοί των $\binom{N}{n}$.

Αν κάθε δείγμα του παραπάνω συνδυασμού έχει την ίδια πιθανότητα να επιλεγεί, τότε το δείγμα αυτό είναι τυχαίο και επομένως, αντιπροσωπευτικό του συνολικού πληθυσμού από τον οποίο το πήραμε. Π.χ., αν ένας πληθυσμός αποτελείται από 10 μονάδες και θέλουμε να πάρουμε δείγμα 4 μονάδων, τότε οι δυνατές τετράδες θα είναι όσοι είναι οι συνδυασμοί των 10 ανά 4, δηλαδή:

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

Δηλαδή μπορούν να επιλεγούν 210 δείγματα των 4 μονάδων. Καθένα από τα 210 δείγματα έχει την ίδια πιθανότητα να επιλεχθεί:

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

3.8 Στατιστικά δειγματοληπτικά σφάλματα

Ο όρος *σφάλμα* σημαίνει τη διαφορά που υπάρχει μεταξύ μιας εκτίμησης που προκύπτει από ένα δείγμα από την αντίστοιχη παράμετρο του πληθυσμού που προκύπτει από την απογραφή. Τα δειγματοληπτικά σφάλματα οφείλονται στην τυχαία επιλογή των μονάδων του δείγματος και τείνουν να αλληλοεξουδετερωθούν στο μέσο όρο.

Το δειγματοληπτικό σφάλμα γίνεται μικρότερο, όταν αυξάνει το μέγεθος του δείγματος, και τείνει να εξαφανιστεί, αν πλησιάζει τη γενική απογραφή.

Π.χ. έστω ότι η μέση ηλικία όλων των σπουδαστών του 4ου εξαμήνου της Σχολής μας είναι 20 έτη. Από το σύνολο των σπουδαστών παίρνουμε ένα δείγμα 1% και βρίσκουμε μέση ηλικία 19 έτη.

Η διαφορά $|\mu - \bar{x}| = |20 - 19| = 1$ έτος λέγεται *δειγματοληπτικό σφάλμα.*

Μια εκτίμηση π.χ. του μέσου του πληθυσμού που προκύπτει από ένα δείγμα, όπως στο προηγούμενο παράδειγμα, μπορεί να διαφέρει από τον αντίστοιχο μέσο αριθμητικό του πληθυσμού, αλλά ο μέσος της κατανομής δειγματοληψίας ($\bar{x}$) ισούται με το μέσο του πληθυσμού ($\bar{x} = \mu$).

Τα μη δειγματοληπτικά σφάλματα οφείλονται στην απροσεξία των ατόμων της ομάδας που ασχολούνται με την έρευνα και συγκεκριμένα:

- σε λανθασμένες πληροφορίες,
- σε λανθασμένες μετρήσεις και καταχωρήσεις,
- στο γεγονός ότι το πλαίσιο που χρησιμοποιούμε για την επιλογή του δείγματος δεν είναι πλήρως ενημερωμένο και σωστό,
- στο ότι τα ερωτήματα του ερωτηματολογίου δεν είναι σαφή,
- στο ότι τα ερωτήματα είναι πολλά και κουραστικά και
- στη μη ορθή επεξεργασία των ερωτηματολογίων και μη σωστή ανάλυση των δεδομένων.

Το σύνολο των δειγματοληπτικών και μη δειγματοληπτικών σφαλμάτων μας δίνουν το συνολικό σφάλμα της έρευνας, το οποίο θα πρέπει κάθε φορά να επιδιώκουμε να είναι ελάχιστο.

3.9 Επεξεργασία δειγματοληπτικών στατιστικών στοιχείων

Μετά τη συγκέντρωση των στατιστικών στοιχείων από τα πολυάριθμα (συμπληρωμένα ερωτηματολόγια ακολουθεί το στάδιο της επεξεργασίας και της μηχανογραφικής οργάνωσης των δειγματοληπτικών στατιστικών στοιχείων. Το στάδιο αυτό της επεξεργασίας περιλαμβάνει τον *έλεγχο όλων των ερωτηματολογίων*, μήπως αυτά περιέχουν ασυμπλήρωτες, ασαφείς, δυσανάγνωστες ή ασυμβίβαστες απαντήσεις. Μετά ακολουθεί *η διαλογή των πληροφοριών των διάφορων χαρακτηριστικών των στατιστικών μονάδων και η εμφάνισή τους, 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, υπολογιστικές μηχανές κ.λπ.).

3.9.1 Παρουσίαση δειγματοληπτικών στοιχείων σε μορφή πινάκων

Ύστερα από την επεξεργασία, τη μηχανογραφική οργάνωση και την ταξινόμηση των δειγματοληπτικών στοιχείων ακολουθεί το στάδιο της συνοπτικής παρουσίασης των συγκεντρωθέντων στοιχείων, κατά τρόπο που να διευκολύνει την ανάλυση των στοιχείων αυτών και την εξαγωγή χρησίμων συμπερασμάτων.

Η παρουσίαση των δειγματοληπτικών στοιχείων γίνεται συνήθως σε μορφή:

1. πινάκων
2. γραφικών παραστάσεων
3. εκθέσεων ή αναφορών.

Η παρουσίαση των δειγματοληπτικών στοιχείων σε πίνακες γίνεται με την κατάλληλη τοποθέτηση των δειγματοληπτικών πληροφοριών σε στήλες και γραμμές, κατά τέτοιο τρόπο που να διευκολύνεται η σύγκριση των στοιχείων και η καλύτερη ενημέρωση του αναγνώστη για τη δομή του πληθυσμού που ερευνούμε. Τους βασικούς τύπους των δειγματοληπτικών πινάκων, που μας ενδιαφέρουν περισσότερο, τους διακρίνουμε κυρίως σε δύο κατηγορίες:

3.9.1.1 Πίνακες απλής εισόδου

Οι πίνακες απλής εισόδου αναφέρονται στην παρουσίαση ενός φαινομένου από την άποψη μόνο ενός χαρακτηριστικού και χρησιμοποιούνται συνήθως για συγκρίσεις και εξαγωγή συμπερασμάτων. Π.χ. στον πίνακα Α δίνεται ένα δείγμα 8.530 οικογενειών ως προς τον αριθμό των τέκνων τους.

Η γενική μορφή ενός πίνακα απλής εισόδου, ανάλογα αν η μεταβλητή είναι ασυνεχής ή συνεχής, έχει όπως φαίνεται στους πίνακες Β και Γ:

Πίνακας Α

Αριθμός τέκνων (xi)	Αριθμός οικογενειών
0	200
1	1.000
2	2.500
3	2.000
4	1.600
5	700
6	300
7	150
8	80
Σύνολο	8.530

Πίνακας Β

Τιμές μεταβλητής (x_i)	*Συχνότητες* (f_i)
x_1	f_1
x_2	f_2
x_3	f_3
x_4	f_4
.	.
.	.
.	.
x_i	f_i
.	.
.	.
.	.
x_n	f_n

Σύνολο	Σf_i

Πίνακας Γ

Τάξεις	Συχνότητες (f_i)
α_0-α_1 α_1-α_2 α_2-α_3 . . . α_i-1 – α_i . . α_n-1 -α_n	f_1 f_2 f_3 . . . f_i . . . f_n
Σύνολο	Σf_i

3.9.1.2 Πίνακες διπλής εισόδου

Οι πίνακες αυτοί μας δίνουν πληροφορίες ενός δείγματος από την άποψη δύο ποσοτικών ή ποιοτικών χαρακτηριστικών. Γενικά, αν πρόκειται για τη μελέτη ενός πληθυσμού από την άποψη δύο ποσοτικών μεταβλητών (βάρος, ύψος), η γενική μορφή του πίνακα διπλής εισόδου θα είναι όπως δείχνει ο πίνακας:

Πίνακας Δ

	Τάξεις

		$\beta_0 - \beta_1$	$\beta_1 - \beta_2$	…	$\beta_{j-1} - \beta_j$	…	$\beta_{\lambda-1} - \beta_\lambda$	
τάξεις	χ_i							
		Y_j						
		y_1	y_2	…	y_j	…	y_λ	f_i
$\alpha_0 - \alpha_1$	χ_1	f_{11}	f_{12}	…	f_{ij}	…	$f_{1\lambda}$	f_1
$\alpha_1 - \alpha_2$	χ_2	f_{21}	f_{22}	…	f_{2j}	…	$f_{2\lambda}$	f_2
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
α_i-1 – α_κ	χ_i	f_{i1}	f_{i2}	…	f_j	…	$f_{i\lambda}$	$f_{i\lambda}$
	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	χ_κ	$f_{\kappa 1}$	$f_{\kappa 2}$	…	$f_{\kappa j}$	…	$f_{\kappa\lambda}$	f_κ
$\alpha_{\kappa-l} - \alpha_\kappa$								
	$f_{.j}$	$f_{.1}$	$f_{.2}$	…	$f_{.j}$	…	$f_{.\lambda}$	N

Αν όμως πρόκειται για ποιοτικά χαρακτηριστικά (εθνικότητα, εκπαίδευση, οικογενειακή κατάσταση, χρώμα μαλλιών και ματιών), η γενική μορφή του πίνακα διπλής εισόδου είναι αυτή του πίνακα Ε.

Πίνακας Ε

Χαρακτηριστικό Β	Χαρακτηριστικό Α					
	A_i	A_2	…	A_j	…	A_κ

B_1	f_{11}	f_{12}	…	f_{1j}	…	$f_{1\kappa}$
B_2	f_{21}	f_{22}	…	f_{2j}	…	f_{2k}
B_i	f_{i1}	f_{i2}	…	f_{ij}	…	f_{ik}
B_n	f_{nl}	f_{n2}	…	f_{nj}	…	$f_{n\kappa}$

4. ΠΑΡΟΥΣΙΑΣΗ ΣΤΑΤΙΣΤΙΚΩΝ ΣΤΟΙΧΕΙΩΝ

Ύστερα από την επεξεργασία, την οργάνωση και την ταξινόμηση των στατιστικών στοιχείων, ακολουθεί το στάδιο της συνοπτικής παρουσίασης των στοιχείων, που έχουμε συγκεντρώσει με τέτοιο τρόπο ώστε να διευκολύνει την ανάλυση των στοιχείων αυτών και την εξαγωγή χρήσιμων συμπερασμάτων.

Η παρουσίαση των στατιστικών στοιχείων μπορεί να γίνει με τρείς τρόπους:

- Με μορφή πινάκων
- Με μορφή γραφημάτων και
- Με μορφή εκθέσεων

Παρακάτω θα αναλύσουμε τα γραφήματα όπου αποτελούν την μέθοδο που διασφαλίζει πιο ξεκάθαρα αποτελέσματα.

4.1 ΓΡΑΦΗΜΑΤΑ - ΔΙΑΓΡΑΜΜΑΤΑ

Γράφημα ονομάζεται μια γραφική αναπαράσταση μίας ή περισσοτέρων μεταβλητών. Τα γραφήματα είναι χρήσιμα για να βλέπουμε και να καταλαβαίνουμε το σχήμα της κατανομής μίας μεταβλητής. Είναι χρήσιμα επίσης για να δούμε οπτικά τη σχέση ανάμεσα σε δύο ή περισσότερες μεταβλητές.

Κύριος <u>στόχος</u> της στατιστικής ανάλυσης είναι να αντληθούν όσο το δυνατό περισσότερες πληροφορίες από τα δεδομένα. Θα πρέπει να εξηγήσουμε και όχι να δώσουμε ερμηνείες στα στοιχεία. Βασικά χαρακτηριστικά για την δημιουργία ενός γραφήματος είναι:

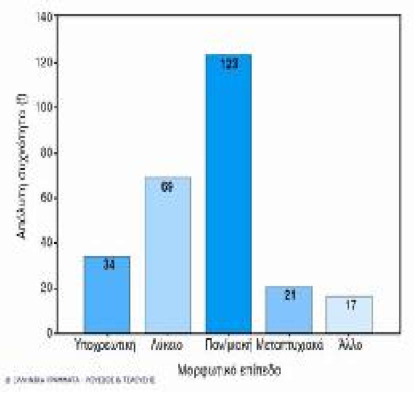
- Να ξεκαθαρίσουμε τους στόχους και τις προτεραιότητες σε ότι αφορά το μήνυμα που θέλουμε να δώσουμε.
- Να επιλέξουμε το κατάλληλο είδος γραφικής παράστασης.
- Να ενημερώσουμε τον αναγνώστη σχετικά με την φύση των απεικονιζόμενων πληροφοριών με σαφή τίτλο.
- Να κατασκευάσουμε ένα σχεδιάγραμμα το οποίο να είναι: παραστατικό, σαφές και ακριβές.

Θα πρέπει να δοθεί **<u>προσοχή</u>**:

- Στις διαστάσεις του διαγράμματος.
- Στις χρησιμοποιούμενες γραμμοσκιάσεις και χρωματισμούς.
- Στην ορθολογική χρήση των εργαλείων λογισμικού.
- Στις τυχόν διευκρινήσεις που είναι απαραίτητες να δοθούν.

4.1.1 ΑΚΙΔΩΤΟ ΔΙΑΓΡΑΜΜΑ

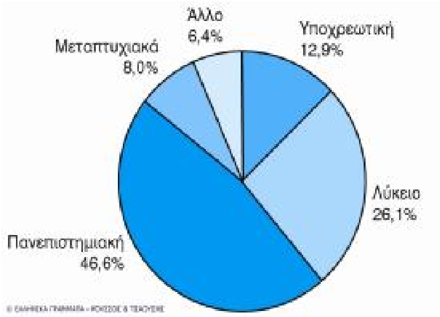
Περιγράφει τις κατηγορίες μιας ποιοτικής μεταβλητής με ράβδους. Το ύψος της κάθε ράβδου συνήθως είναι ανάλογο του



πραγματικού αριθμού ή του ποσοστού που αντιστοιχεί σε κάθε κατηγορία.

4.1.2 ΚΥΚΛΙΚΟ ΔΙΑΓΡΑΜΜΑ (Pie chart)

Περιγράφει τις κατηγορίες μιας ποιοτικής μεταβλητής με κομμάτια μιας πίτας (κύκλου). Τοκομμάτι κάθε κατηγορίας είναι ανάλογο του αριθμού των αντικειμένων που ανήκουν σε κάθε κατηγορία. Παρουσιάζουν μία στατική εικόνα ενός δείγματος ή πληθυσμού.

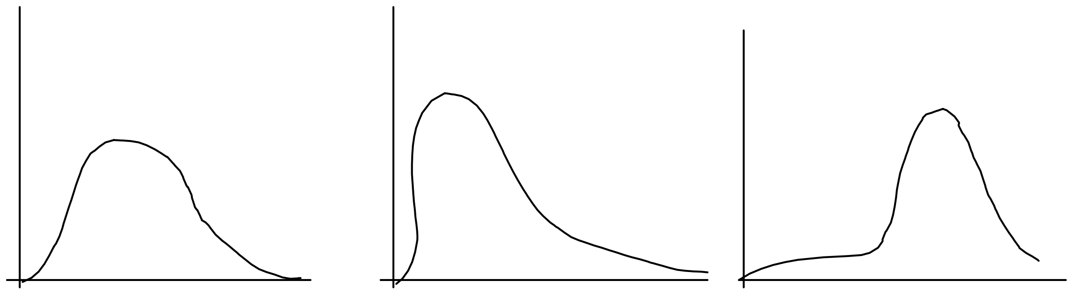


4.1.3 ΙΣΤΟΓΡΑΜΜΑ (Histogram)

Απεικονίζει την κατανομή μιας ποσοτικής μεταβλητής με την βοήθεια ράβδων. Κάθε ράβδος αντιστοιχεί σε ένα διάστημα τιμών και το ύψος είναι ανάλογο των αντικειμένων που ανήκουν σε αυτό το διάστημα Συχνά απεικονίζουμε και την γραμμή της κανονικής κατανομής για σύγκριση. Στον άξονα x εμφανίζονται είτε μεμονωμένες τιμές είτε σύνολο τιμών. Το ύψος κάθε τμήματος αναπαριστά τη συχνότητα με την οποία εμφανίζεται αυτή η τιμή ή το διάστημα. Αντίθετα με το διάγραμμα μπάρας στο οποίο ο άξονας x παριστάνει κατηγορίες, ο οριζόντιος άξονας ενός ιστογράμματος δείχνει μια αριθμητική κλίμακα. Επίσης χρησιμοποιείται για να δείξει το σχήμα μιας μεταβλητής.

Το ιστόγραμμα κατασκευάζεται με βάση το πίνακα συχνοτήτων ομαδοποιημένων μετρήσεων. Αν η πρώτη ή η τελευταία ομάδα είναι ανοικτή παραλείπεται στην κατασκευή του ιστογράμματος. Το μήκος των ομάδων επηρεάζει την εικόνα του ιστογράμματος. Όσο πιο μεγάλο είναι το μήκος των ομάδων τόσο πιο ασαφές και ακαθόριστο είναι το σχήμα της κατανομής των μετρήσεων. Αν οι ομάδες είναι πολλές και έχουν μικρό μήκος το ιστόγραμμα παρουσιάζει ανωμαλίες που αντανακλούν την μεταβλητικότητα της δειγματοληψίας.

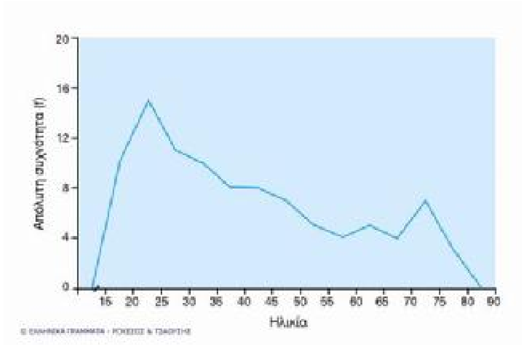
Επίσης το ιστόγραμμα δείχνει την ασυμμετρία (skewness) μιας κατανομής που είναι μέτρο της συμμετρίας της. Αν μία κατανομή είναι συμμετρική, η αριστερή πλευρά της είναι ένα κατοπτρικό είδωλο της δεξιάς. Αν οι τιμές μίας κατανομής συσσωρεύονται στις μικρές τιμές η κατανομή είναι θετικά ασύμμετρη. Αν οι τιμές της κατανομής βρίσκονται στο άλλο άκρο της κλίμακας η κατανομή είναι αρνητικά συμμετρική.



Συμμετρική Θετικά Συμμετρική Αρνητικά Συμμετρική

4.1.4 ΠΟΛΥΓΩΝΟ ΣΥΧΝΟΤΗΤΩΝ

Το πολύγωνο συχνοτήτων αποτελεί το δεύτερο τρόπο γραφικής σύμπτυξης και παρουσίασης των μετρήσεων. Κατασκευάζεται ενώνοντας με ευθείες γραμμές τα μέσα των πάνω πλευρών του ορθογωνίων του ιστογράμματος. Παρουσιάζει με πιο ευδιάκριτο τρόπο απ' ότι το ιστόγραμμα τη μορφή των στοιχείων. Χρησιμοποιείται μόνο για την διαγραμματική απεικόνιση συνεχών διδομένων και ενδείκνυται κυρίως για την ταυτόχρονη σύγκριση δύο ή περισσοτέρων κατανομών. Η κατασκευή του σχεδιάζεται είτε με βάση το ιστόγραμμα της εξεταζόμενης κατανομής είτε ανεξάρτητα από αυτό.



4.1.5 ΕΜΒΑΔΟΓΡΑΜΜΑ

Σύγκριση με βάση το εμβαδόν κάποιας περιοχής. Τα γραφήματα μπορεί να είναι απλά ή συσσωρευμένα τα οποία συγκρίνουν την ίδια μεταβλητή, ή μεταβλητές μεταξύ τους ή περιπτώσεις μεταξύ τους.

4.1.6 ΔΙΑΓΡΑΜΜΑ ΜΙΣΧΟΥ-ΦΥΛΛΟΥ (Steam-and-leaf)

Η τεχνική αυτή οδηγεί σε σχηματική αναπαράσταση παρόμοια με εκείνη του ιστογράμματος χωρίς να οδηγεί σε απώλεια πληροφοριών. Το διάγραμμα αυτό δίνει την δυνατότητα ανασύστασης και ανάκλησης των μετρήσεων των αρχικών δεδομένων του δείγματος με ακρίβεια, πράγμα το οποίο δεν επιτυγχάνεται με το ιστόγραμμα ή τους πίνακες συχνοτήτων. Χρησιμοποιείται για την επεξεργασία μέτριου αριθμού παρατηρήσεων (περίπου 150). Η παρουσίαση του σχήματος μοιάζει με εκείνου του ιστογράμματος αλλά η τεχνική κατάρτισης δεν είναι η ίδια.

Το φυλλόγραμμα:

- Εμφανίζει τα δεδομένα σε όλο το εύρος των παρατηρηθέντων μετρήσεων.
- Παρουσιάζει την συγκέντρωση των παρατηρήσεων (συχνότητες).
- Δείχνει την μορφή της κατανομής.
- Εμφανίζει τυχόν ακραίες και εκτροπές παρατηρήσεις.
- Επιτρέπει την επισήμανση της απουσίας συγκεκριμένων τιμών ή μετρήσεων.

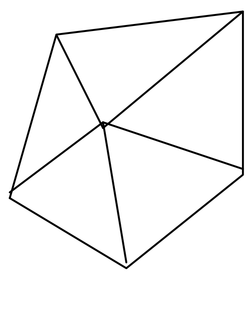
4.1.7 ΑΣΤΕΡΟΕΙΔΗ ΔΙΑΓΡΑΜΜΑΤΑ

Κάθε παρατήρηση απεικονίζεται σαν ένα αστέρι το οποίο αποτελείται από τόσες ακτίνες όσες και οι μεταβλητές που χρησιμοποιούμε. Το μήκος του είναι ανάλογο των τιμών των μεταβλητών. Τα αστεροειδή διαγράμματα μας βοηθούν να κατατάξουμε τις παρατηρήσεις σε ομοειδή ομάδες με το μμέγεθος και το σχήμα των αστέρων. Τα μήκη των αστέρων καθορίζονται με δύο τρόπους:

α . Από τις πραγματικές τιμές

β. Από τις αναδιαβαθμισμένες τιμές.

Στην περίπτωση αυτή οι τιμές αλλάζουν έτσι ώστε η ελάχιστη τιμή να γίνει 0 (όχι ακτίνα) και η μέγιστη 1 (μεγάλη ακτίνα).

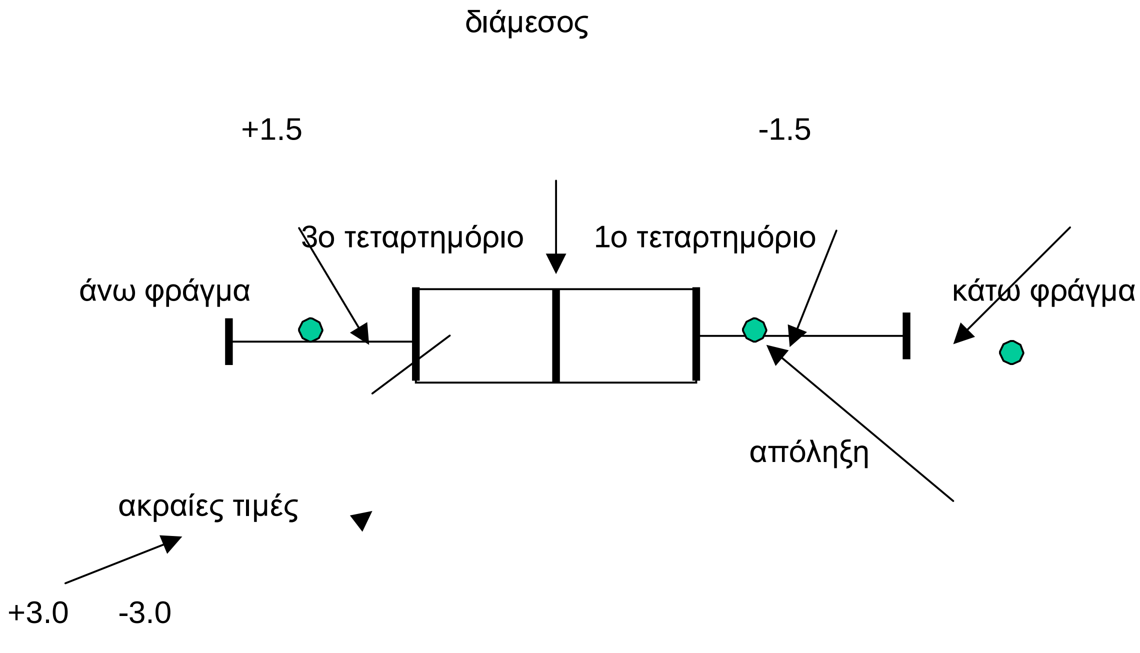


4.1.8 ΠΡΟΣΩΠΑ ΤΟΥ CHERNOFF

Είναι γραφήματα ανάλογα των αστεροειδών μόνο που αντί η κάθε παρατήρηση να απεικονίζεται σαν ακτίνα απεικονίζεται σαν πρόσωπο. Κάθε μεταβλητή καθορίζει ένα συγκεκριμένο χαρακτηριστικό του προσώπου. Η απεικόνιση αυτή είναι διαδεδομένη στις κοινωνικές επιστήμες και την ψυχολογία.

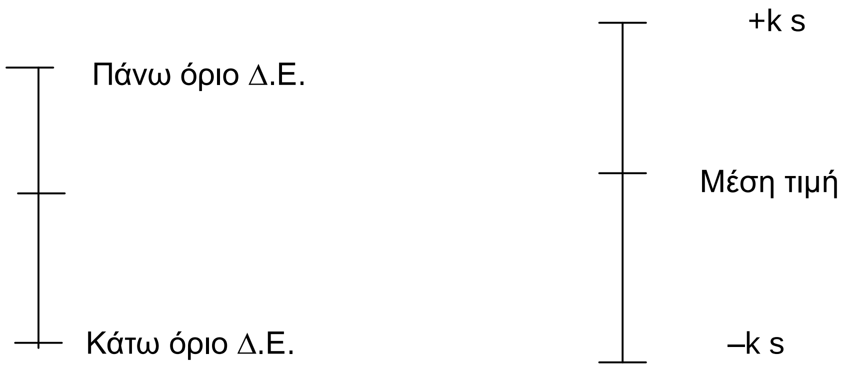
4.1.9 ΔΙΑΓΡΑΜΜΑΤΑ ΠΛΑΙΣΙΟΥ-ΑΠΟΛΗΞΕΩΝ (Box plot)

Περιλαμβάνουν περιληπτικά την κατανομή των ποσοτικών μεταβλητών. Κάθε πλαίσιο-κουτί απεικονίζει το 1ο τεταρτημόριο, την διάμεσο και το 3ο τεταρτημόριο. Οι απολήξεις υποδεικνύουν τα όρια των ακραίων τιμών. Οι τιμές εκτός των φραγμάτων των απολήξεων θεωρούνται ακραίες και υποδεικνύονται στο γράφημα με ξεχωριστά σημεία. Συμμετρικά διαγράμματα πλησιάζουν την κανονική κατανομή.



4.1.10 ΔΙΑΓΡΑΜΜΑΤΑ ΣΦΑΛΜΑΤΩΝ (Error-bars)

Είναι διαγράμματα που περιγράφουν περιληπτικά την κατανομή μιας ποσοτικής μεταβλητής σε διάφορα επίπεδα. Συνήθως αναπαριστά διαστήματα εμπιστοσύνης για το μέσο, αλλά εναλλακτικά μπορεί να χρησιμοποιηθεί και για την τυπική απόκλιση. Μοιάζουν με τα θηκογράμματα, αλλά συγκρίνουν διαστήματα εμπιστοσύνης και όχι κατανομές. Το σχήμα τους είναι μια ράβδος κατανεμημένη ισομερώς γύρω από το τη μέση τιμή των τιμών. Το μήκος κάθε ράβδου ισούται με το 100(1-α)% διαστήμα εμπιστοσύνης γύρω από τον μέσο.



όπου k μπορεί να πάρει τιμές 1(για 70%), 2(για 95%) και 3(για 99%).

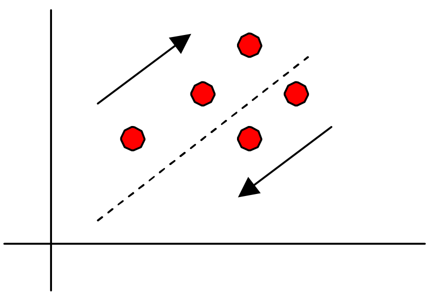
4.1.11 ΔΙΑΓΡΑΜΜΑΤΑ ΣΗΜΕΙΩΝ (Scatter plots)

Ονομάζονται και *διαγράμματα* διασποράς. Αποτελούν το γραφικό τρόπο αναζήτησης σχέσεων μεταξύ μεταβλητών. Περιγράφουν τη δυσδιάστατη κατανομή δύο ποσοτικών μεταβλητών. Κάθε σημείο απεικονίζει ένα ζευγάρι τιμών των υπό εξέταση μεταβλητών (συσχέτιση μεταβλητών). Εντοπίζονται εύκολα συσχετίσεις και ακραίες τιμές.

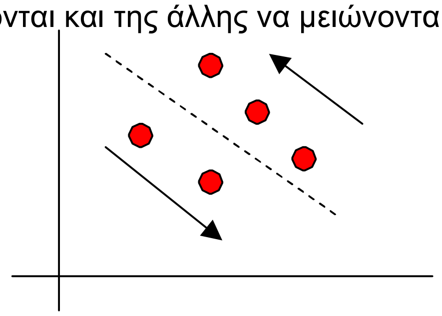
Συσχέτιση αναφέρεται ο βαθμός με τον οποίο σχετίζονται (συμμεταβάλονται) δύο μεταβλητές. Ο βαθμός σχέσης με την οποία σχετίζονται δύο μεταβλητές ονομάζεται *απλή συσχέτιση* και ο βαθμός σχέσης με την οποία σχετίζονται περισσότερες μεταβλητές ονομάζεται *πολλαπλή συσχέτιση*. Η απλή συσχέτιση ασχολείται με το βαθμό το οποίο τα σημεία συγκεντρώνονται γύρω από μια ευθεία χωρίς να προσδιορίζεται ποιά είναι ακριβώς αυτή η γραμμή που διέρχεται μέσα από το νέφος των σημείων.

- **Γραμμική συσχέτιση** εμφανίζεται όταν στο διάγραμμα διασκορπισμού τα σημεία όλων των παρατηρήσεων τείνουν να συγκεντρώνονται γύρο από μια ευθεία και **μη γραμμική** όταν τα σημεία τείνουν να συγκεντρωθούν γύρο από μια καμπύλη.

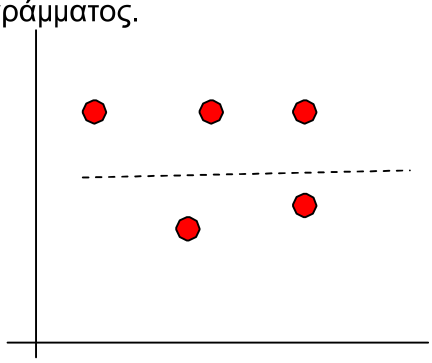
Θετική συσχέτιση όταν δύο μεταβλητές τείνουν να μεταβάλλονται προς την ίδια κατεύθυνση. Στην περίπτωση αυτή οι τιμές τείνουν να αυξάνονται ή να μειώνονται.



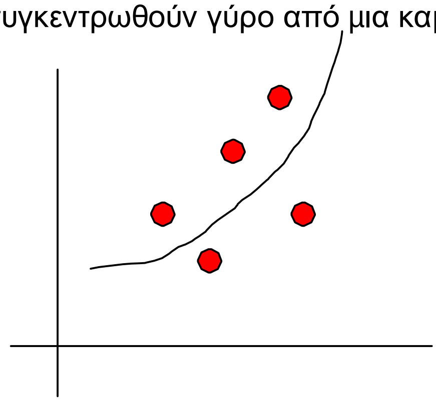
Αρνητική συσχέτιση όταν δύο μεταβλητές τείνουν να μμεταβάλλονται προς αντίθετη κατεύθυνση. Στην περίπτωση αυτή οι τιμές της μιας μεταβλητής τείνουν να αυξάνονται και της άλλης να μειώνονται.



Μηδενική συσχέτιση όταν οι μεταβολές των τιμών της μίας μεταβλητής δεν συνδέονται με τις μεταβολές της άλλης. Τα σημεία του νέφους είναι διασκορπισμένα σε όλο το μήκος του διαγράμματος.



- **Μη γραμμική συσχέτιση** όταν οι μεταβολές των τιμών της μίας μεταβλητής συνδέονται με τις μεταβολές της άλλης με μη γραμμική μορφή. Τα σημεία του νέφους τείνουν να συγκεντρωθούν γύρο από μια καμπύλη.

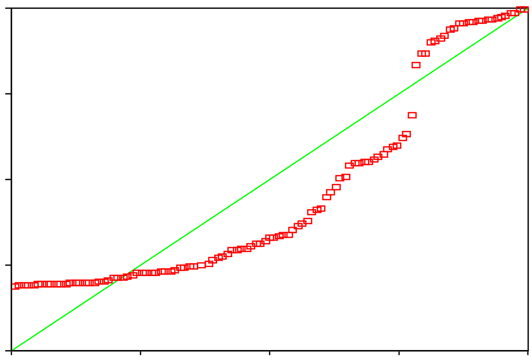


4.1.12 ΠΙΘΑΝΟΘΕΩΡΗΤΙΚΑ ΔΙΑΓΡΑΜΜΑΤΑ (P-P, Q-Q)

Προσπαθούμε να διαπιστώσουμε (με γραφικούς τρόπους) πόσο κοντά σε κάποια κατανομή που προσδιορίζουμε είναι τα δεδομένα. Συνήθως ενδιαφερόμαστε για την κανονική κατανομή. Όσο πιο κοντά στην διχοτόμο της γωνίας των αξόνων είναι συγκεντρωμένα τα σημεία τόσο περισσότερο ενισχύεται η υπόθεση ότι τα δεδομένα ακολουθούν την κανονική κατανομή. Πιθανοθεωρητικά διαγράμματα είναι διαγράμματα ενός δείγματος σε σχέση με τα δεδομένα που περιμέναμε να πάρουμε αν θεωρούσαμε από μια κανονική κατανομή.

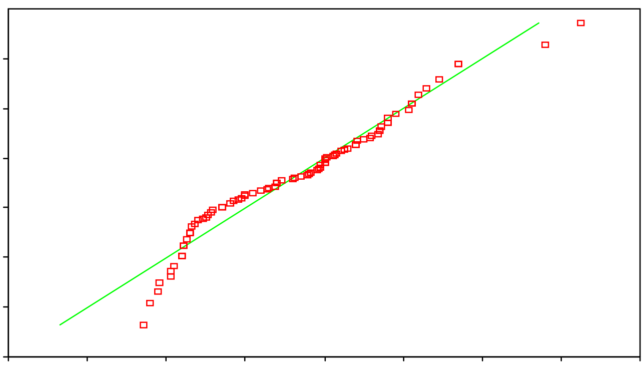
- **Normal P-P**

Τα γραφήματα αυτά παριστάνουν σε σύστημα οριζοντίων και καθέτων αξόνων τις τιμές της παρατηρούμενης αθροιστικής συχνότητας (άξονας x) και της υποτιθέμενης κανονικής κατανομής (άξονας y) που ακολουθεί η μεταβλητή που εξετάζεται. Όσο πιο κοντά στην διχοτόμο της γωνίας των αξόνων είναι συγκεντρωμένα τα σημεία τόσο περισσότερο ενισχύεται η υπόθεση ότι τα δεδομένα ακολουθούν την κανονική κατανομή.



- **Normal Q-Q**

Τα γραφήματα αυτά παριστάνουν τα εκατοστημόρια της παρατηρούμενης ως προς την αναμενόμενη κανονική κατανομή. Τα εκατοστημόρια υπολογίζονται με βάση διαφορετικούς αλγόριθμους και τοποθετούνται στον κάθετο άξονα y, ενώ α αντίστοιχα παρατηρούμενα στον οριζόντια άξονα x.



Τέλος, τα γραφήματα αποτελούν το καλύτερο μέσο στατιστικής παρουσίασης γιατί δίνουν στους αφηρημένους αριθμούς μια συγκεκριμένη μορφή που μας διευκολύνει να έχουμε με τη βοήθεια ενός γεωμετρικού σχήματος, μία άμεση αντίληψη της μορφής του φαινομένου το οποίο θέλουμε να μελετήσουμε.

Ένα καλά σχεδιασμένο γράφημα μπορεί να γίνει κατανοητό και να διατηρηθεί στη μνήμη του αναγνώστη πιο εύκολα από έναν αριθμητικό πίνακα.

5. ΕΜΠΕΙΡΙΚΕΣ ΚΑΤΑΝΟΜΕΣ ΣΥΧΝΟΤΗΤΩΝ

5.1 Ασυνεχείς και Συνεχείς κατανομές συχνοτήτων

Μετά τη συγκέντρωση των στατιστικών στοιχείων που αναφέρονται σε μια μόνο ιδιότητα ενός πληθυσμού, το πρώτο βασικό βήμα είναι η κατάλληλη κατάταξη και η συστηματική ομαδοποίηση των τιμών της μεταβλητής, με αντικειμενικό σκοπό την παρουσίαση των στοιχείων κατά τρόπο που να δίνει πληροφορίες για τη δομή του πληθυσμού που ερευνάται και να διευκολύνει το έργο της στατιστικής ανάλυσης και εξαγωγής χρήσιμων συμπερασμάτων. Για την ομαδοποίηση των τιμών της μεταβλητής χρησιμοποιούνται ειδικές κατατάξεις που τις ονομάζουμε *κατανομές συχνοτήτων* ή *πίνακες συχνοτήτων.*
Ο 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 απλής εισόδου, όπου στην πρώτη στήλη τοποθετούμε 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 συχνοτήτων παίρνει τη μορφή του πίνακα Α.
Στον πίνακα αυτό:

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

(2) $f_1, f_2,\ldots, f_i,\ldots, f_v$ είναι οι αντίστοιχες συχνότητες της μεταβλητής x, δηλαδή πόσες

φορές εμφανίζεται, στο συνολικό πληθυσμό, κάθε τιμή της μεταβλητής *x.*

Πίνακας Α

Τιμές μεταβλητής (x_i)	*Αριθμός παρατηρήσεων ή συχνότητες (f_i)*
x_1 x_2 x_3 x_4 . . . x_i . . . x_v	f_1 f_2 f_3 f_4 . . . f_i . . . f_v
Σύνολο	Σf_i

Όταν η τιμή 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 $f_i / \Sigma f_i$ ονομάζεται σχετική συχνότητα της x_i.

Ο συμβολισμός Σf_i σημαίνει το άθροισμα όλων των f_i, δηλαδή:

$$\sum_{i=1}^{n} f_i = f_1 + f_2 + f_3 + \ldots + f_n$$

Για την καλύτερη κατανόηση του τρόπου παρουσίασης των αριθμητικών δεδομένων ενός πληθυσμού με μορφή ασυνεχούς κατανομής συχνοτήτων, δίνουμε το παρακάτω παράδειγμα.

Πίνακας

Τιμές μεταβλητής (x_i)	*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:

f_1 f_2 … f_i … f_n

α_0 α_1 $\alpha_2 \ldots \alpha_{i-1}$ $\alpha_i \ldots \alpha_{n-1}$ α_n

Κάθε τάξη διακρίνεται από τα όριά της, δηλαδή από το κατώτερο όριο α_{i-1} και από το ανώτερο όριο (α_i).

Σε κάθε τάξη (α_{i-1}, α_i) αντιστοιχούμε την ονομαζόμενη κεντρική τιμή xi 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 συνεχούς μεταβλητής έχουν την ίδια 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 βάση το κριτήριο της

ομοιογένειας και το κριτήριο της απλότητας.
Το πρώτο κριτήριο χρειάζεται τη διαίρεση του συνολικού εύρους της μεταβολής σε πολλές τάξεις μικρού πλάτους. Αντίθετα, με κριτήριο τη ν 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 σκοπό 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.

Πίνακας

Τάξεις	*Συχνότητες (f_i)*	*Κεντρικές τιμές των τάξεων (x_i)*
α_0-α_1	f_1	x_1
α_1-α_2	f_2	x_2
α_2-α_3	f_3	x_3
.	.	.
.	.	.
.	.	.
$\alpha_{i-1} – \alpha_i$	f_i	x_i

. . α_{n-1}- α_n	. . . f_n	. . . x_n
Σύνολο	Σf_i	

Η κατανομή του παραπάνω πίνακα ονομάζεται *κλειστή,* γιατί δεν λείπει ούτε το κατώτερο όριο της πρώτης τάξης ούτε το ανώτερο της τελευταίας τάξης. Αν η κατανομή συχνοτήτων στερείται 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 δραχμές) του καθενός από τους 40 υπαλλήλους 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 ταξινόμηση των παρατηρήσεων τις κατατάσσουμε κατά φυσική αύξουσα τάξη 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 *(Μ = 99)* και τη μικρότερη τιμή *(Ε = 31),* η οποία ονομάζεται *εύρος της κατανομής*, δηλαδή *R = M - E = 99 - 31 = 68*. Επειδή το εύρος *R* είναι σχετικά μικρό, επτά π.χ. τάξεις είναι αρκετές για να ικανοποιήσουν το κριτήριο της απλότητας και της ομοιογένειας, το εύρος κάθε τάξης θα είναι: *δ = 68 : 7 ≈ 10*.

Κατά συνέπεια, θα έχουμε τις 40 παρατηρήσεις που μας δόθηκαν, με μορφή κατανομής κατά συχνότητες, όπως φαίνονται 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.

5.2 Ιστόγραμμα συχνοτήτων

Το ιστόγραμμα χρησιμοποιείται κατά κανόνα για τη γραφική απεικόνιση των κατανομών συχνοτήτων όταν έχουμε διαστήματα τάξεων.
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. Για την κατασκευή του ιστογράμματος διακρίνουμε δύο περιπτώσεις:

α) Όταν τα διαστήματα τάξεων είναι ίσα.

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 τμήματα, ενώ στον κάθετο άξονα (τεταγμένων) τοποθετούμε τις συχνότητες (f_i).

Μετά, από τα όρια των τάξεων ($α_i$ - $α_{i-1}$) 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.

Παράδειγμα

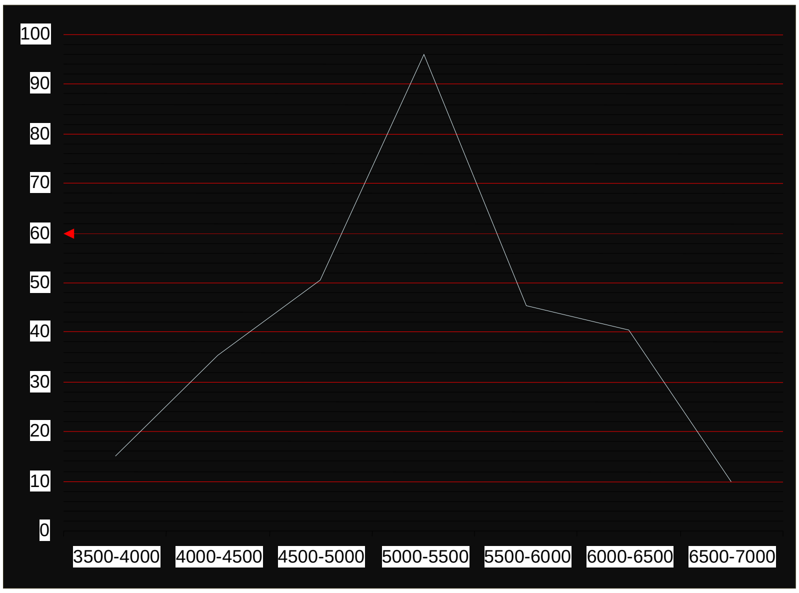
Στον παρακάτω πίνακα δίνεται η κατανομή των ημερήσιων αποδοχών 290 εργαζόμενων σε μια επιχείρηση.

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.

Λύση



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

β) Όταν το πλάτος των τάξεων είναι άνισο

Στην περίπτωση των κατανομών άνισου πλάτους, παίρνουμε στον οριζόντιο άξονα άνισα διαστήματα, ανάλογα με το πλάτος των αντίστοιχων τάξεων και με βάση τα άνισα αυτά διαστήματα κατασκευάζουμε ορθογώνια παραλληλόγραμμα, τα ύψη των οποίων είναι ίσα προς το πηλίκο της συχνότητας δια του πλάτους κάθε τάξης *(f_i/δ_i).*

Παράδειγμα

Δίνεται η κατανομή του παρακάτω πίνακα με άνισα διαστήματα τάξεων.

Πίνακας

Τάξεις	*Συχνότητες* (f_i)
0-10	8
10-15	6
15-20	13
20-25	22
25-40	15
40-70	6
Σύνολο	70

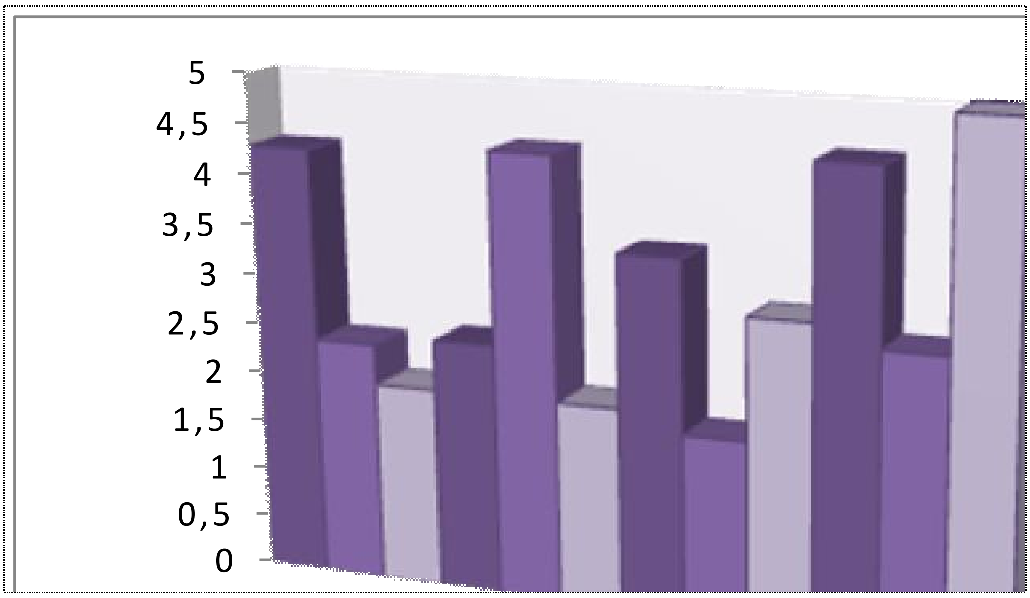
Να σχηματιστεί το ιστόγραμμα συχνοτήτων.

Λύση

Σε περίπτωση κατά την οποία η κατανομή παρουσιάζει άνισα ταξικά διαστήματα, σχηματίζουμε μια νέα στήλη, η οποία ονομάζεται πυκνότητα συχνοτήτων f_i/δ_i, και με βάση αυτή κατασκευάζουμε το ιστόγραμμα.

Πίνακας

Διάστημα τάξεων	*Συχνότητες* (f_i)	*Πυκνότητα συχνοτήτων* (f_i/δ_i)
0-10	8	0,8
10 - 15	6	1,2
15 - 20	13	2,6
20 - 25	22	4,4
25 - 40	15	1,0
40 - 70	6	0,2
Σύνολο	70	



5.3 Αθροιστικές κατανομές συχνοτήτων

Πολλές φορές χρειάζεται να γνωρίζουμε πόσες ή τι ποσοστό των περιπτώσεων 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 x_i της μεταβλητής X. 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:

Από τον πίνακα κατανομής συχνοτήτων μπορούμε να κατασκευάσουμε 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 των 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 αθροιστική σειρά συχνοτήτων και β) Το αθροιστικό διάγραμμα.

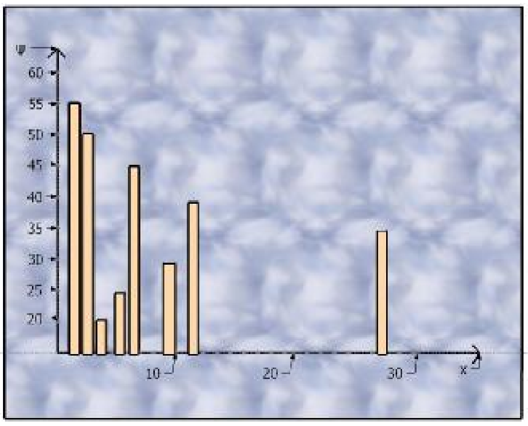
<u>Λύση</u>

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

Πίνακας

Τάξεις	f_i	*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		

β) Το αθροιστικό διάγραμμα έχει ως εξής:



6. Παράμετροι μονομεταβλητών πληθυσμών

Σκοπός της εμφάνισης των στατιστικών δεδομένων με μορφή συνοπτικών πινάκων συχνοτήτων είναι ο περιορισμός του όγκου των στοιχείων που συγκεντρωθήκαν 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. Η συμπύκνωση αυτή γίνεται με την αντικατάσταση των πινάκων συχνοτήτων από ορισμένους αντιπροσωπευτικούς αριθμούς, οι οποίοι ονομάζονται *στατιστικές παράμετροι*, χαρακτηρίζουν τη *θέση*, τη *διασπορά* και τη *μορφολογία* του πληθυσμού που ερευνούμε και διευκολύνουν επιπλέον, πάρα πολύ στις συγκρίσεις ομοειδών ομάδων.

6.1 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 τα στατιστική συμπερασματολογία (εκτιμητική, έλεγχος υποθέσεων).

6.1.1 Είδη στατιστικών περιγραφικών μέτρων

- **Μέτρα κεντρικής τάσης** (αριθμητικός, γεωμετρικός, αρμονικός μέσος).
- **Μέτρα θέσης** (διάμεσος, μέση τιμή, επικρατούσα τιμή, ποσοστημόρια).
- **Μέτρα διασποράς** (διακύμανση, τυπική απόκλιση, ενδοτεταρτημοριακό εύρος, εύρος μεταβολής).
- **Μέτρα λοξότητας ή ασυμμετρίας** (συντελεστής ασυμμετρίας του Pearson).
- **Μέτρα κύρτωσης** (συντελεστής κύρτωσης).

6.1.1.1 Μέτρα κεντρικής τάσης.

I. Μέσος αριθμητικός

Συμβολισμοί:

Α. Πληθυσμιακός Μέσος.

Έστω ένας πληθυσμός με Ν μονάδες $x_1, x_2, \ldots, x_N$.

Ο πληθυσμιακός μέσος είναι:

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

Β. Αν επιλέξουμε ένα δείγμα $x_1, x_2, \ldots, x_n$, μεγέθους n,

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

Ιδιότητες.

1. Αν x_i = a για όλα τα i=1,2,…,n, τότε $\overline{X}=a$.

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

Πράγματι ισχύει για κάθε i, $x_{\min}\le x_i$, 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 διά n προκύπτει ότι: $x_{\min}\le\overline{X}$.

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

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

Πράγματι με βάση της ιδιότητες των αθροισμάτων έχουμε:

$$\sum_{i=1}^{n}(x_i-\overline{X})=\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}\overline{X}=\sum_{i=1}^{n}X_i-n\overline{X}=0$$

από τον ορισμό του αριθμητικού μέσου.

4. 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 $a=\overline{X}$.

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

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

Θέτοντας την ποσότητα αυτή ίση με μηδέν, παίρνω $a=\frac{\sum_{i=1}^{n}x_i}{n}=\overline{X}$.

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.

5. Αν σε όλες τις τιμές ενός συνόλου δεδομένων προσθέσω μία ποσότητα α, ο νέος αριθμητικός μέσος αυξάνεται κατά α. Δηλαδή, αν οι τιμές $x_1,x_2,\ldots,x_n$ έχουν μέσο $a=\overline{X}$, τότε οι τιμές $x_1+α,x_2+α,\ldots,x_n+α$ έχουν μέσο $\overline{X}+a$. Αυτό προκύπτει επειδή:

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

6. 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 $x_1,x_2,\ldots,x_n$, πολλαπλασιάζοντας επί α, τα νέα δεδομένα είναι $αx_1,αx_2,\ldots,αx_n$ και ο αριθμητικός μέσος για τα δεδομένα αυτά θα είναι:

$$\frac{\sum_{i=1}^{n}(a*x_i)}{n} = \frac{a\sum_{i=1}^{n}x_i}{n} = \overline{X}*a .$$

7. Αν έχω k σύνολα δεδομένων

$x_1,x_2,\ldots,x_{n1}$	με μέσο $\overline{X}_1$
$x_{n_1+1},x_{n_1+2,\ldots,}x_{n_1+n_2}$	με μέσο $\overline{X}_2$
$x_{n_1+n_2+1,}x_{n_1+n_2+2,\ldots,}x_{n_1+n_2+n_3}$	με μέσο $\overline{X}_3$

κ.ο.κ.

$x_{n_1+n_2+\ldots+n_{k-1}+1,}x_{n_1+n_2+n_{k-1}+2,\ldots,}x_{n_1+n_2+\ldots+n_{k-1}+\ldots+n_k}$	με μέσο $\overline{X}_k$

τότε για τον αριθμητικό μέσο όλων των παρατηρήσεων έχω:

$$\overline{X} = \frac{x_1+x_2+\ldots+x_{n_1}+\ldots+x_{n_1+n_2}+\ldots+x_{n_1+n_2+n_3}+\ldots+x_{n_1+n_2+\ldots+n_k}}{n_1+n_2+\ldots+n_k} = \frac{n_1\overline{X}_1+n_2\overline{X}_2+\ldots+n_k\overline{X}_k}{n_1+n_2+\ldots+n_k}$$

ή για συντομία $\overline{X} = \frac{\sum_{i=1}^{k}n_i\overline{X}_i}{\sum_{i=1}^{k}n_i}$.

Έμμεση μέθοδος υπολογισμού του μέσου αριθμητικού.

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.

II. Γεωμετρικός μέσος.

Συμβολισμοί:

Α. Πληθυσμιακός Μέσος.

Έστω ένας πληθυσμός με Ν μονάδες $x_1, x_2, \ldots, x_N$.

Ο πληθυσμιακός μέσος είναι: $G = (x_1 * x_2 * \ldots * x_N)^{1/N} = (\Pi_{i=1}^{N} x_i)^{1/N}$.

Β. Αν επιλέξουμε ένα δείγμα $x_1, x_2, \ldots, x_n$ μεγέθους n, ο γεωμετρικός μέσος του δείγματος είναι:

μέσος $g = (x_1 * x_2 * \ldots * x_n)^{1/n} = (\Pi_{i=1 x_i}^{n})^{1/n}$.

Ιδιότητες.

1. Αν πάρουμε λογαρίθμους στη σχέση $G = (x_1 * x_2 * \ldots * x_N)^{1/N} = (\Pi_{i=1}^{N} x_i)^{1/N}$, 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.

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

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

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

Παίρνοντας το ανάπτυγμα έχω:

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

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

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

- Οι παραπάνω ιδιότητες αναφέρονται στον πληθυσμό επειδή χρησιμοποιούν το G.
- Οι ίδιες ισχύουν ακόμα και για το δειγματικό μέσο g.

III. Αρμονικός μέσος

Ο αρμονικός μέσος *H* για τον πληθυσμό ορίζεται ως εξής: $\frac{1}{H}=\frac{\frac{1}{x_1}+\frac{1}{x_2}+\ldots+\frac{1}{x_N}}{N}=\frac{\sum_{i=1}^{n}\frac{1}{x_i}}{N}$ ή ισοδύναμα, $H=\frac{N}{\sum_{i=1}^{n}\frac{1}{x_i}}$, δηλαδή ο αντίστροφος του αρμονικού μέσου είναι ο αριθμητικός μέσος των αντιστρόφων των τιμών της μεταβλητής.

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

Ιδιότητες του αρμονικού μέσου.

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

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

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

2. Το άθροισμα τετραγώνων $\sum_{i=1}^{N}\left(\frac{1}{x_i} - \frac{1}{a}\right)^2$ γίνεται ελάχιστο ως προς *α* όταν *α=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.

α) 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 τον 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.

Πίνακας 5.1

Τιμές μεταβλητής (x_i)	*Συχνότητες (f_i)*	$f_i x_i$
x_1	f_1	f_1x_1
x_2	f_2	f_2x_2
x_3	f_3	f_3x_3
.	.	.
.	.	.
.	.	.
x_k	f_k	f_kx_k
Σύνολο	Σf_i	$\Sigma f_i x_i$

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

Αν συμβολίσουμε με: $\sum f_i x_i = x_1 f_1 + x_2 f_2 + x_3 f_3 + \ldots + x_i f_i + \ldots + x_k f_k$

και με: $\Sigma f_i = f_1 + f_2 + \ldots + f_i + \ldots + f_k$, τότε ο τύπος του μέσου αριθμητικού μπορεί να γραφεί: $m = \dfrac{\sum f_i x_i}{\sum f_i}$

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

β) Όταν τα δεδομένα εμφανίζονται με μορφή κατανομής συχνοτήτων κατά τάξεις (περίπτωση συνεχούς μεταβλητής)

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 άθροισμά 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: $m = \dfrac{\sum f_i x_i}{\sum f_i}$

Πίνακας

Τάξεις	*Συχνότητες (f_i)* *(απόλυτες ή σχετικές)*	*Κεντρικές τιμές των τάξεων (xi)*	$f_i x_i$
$\alpha_o - \alpha_l$	f_1	x_1	$f_1 x_1$
$\alpha_l - \alpha_2$	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	$f_i x_i$
.	.	.	.
.	.	.	.
.	.	.	.
$\alpha_{k-1} - \alpha_k$	f_k	x_k	$f_k x_k$
Σύνολο	Σf_i		$\Sigma f_i x_i$

6.1.1.2 Μέτρα θέσης

I. <u>Διάμεσος.</u>

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 δύο ίσες ομάδες,* δηλαδή ο όρος, πριν και μετά τον οποίο ισοκατανέμονται οι υπόλοιποι όροι της τάξης.

Η διάμεσος συμβολίζεται με το γράμμα *M.*

Για τον υπολογισμό της τιμής της διαμέσου διακρίνουμε δύο περιπτώσεις.

α) 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 υποπεριπτώσεις:

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 τοποθετηθούν κατά φυσική αύξουσα τάξη μεγέθους.
Π.χ. μια μεταβλητή Χ 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, δηλ. 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:

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

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

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

Επομένως: M = 11.

β) Περίπτωση ταξινομημένων δεδομένων

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

1. *Όταν η κατανομή είναι συνεχής.*

Για τον υπολογισμό της διαμέσου σχηματίζουμε τη δεξιόστροφη αθροιστική σειρά *F1, F2,... FN* των 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:

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

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.

2. Όταν η κατανομή συχνοτήτων είναι ασυνεχής

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

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

II. <u>Μέση τιμή.</u>

Η μέση ή αναμενόμενη τιμή μιας τυχαίας μεταβλητής X είναι μια από τις βασικότερες παραμέτρους στη συνοπτική περιγραφή αυτής.

III. <u>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. Αν υπάρχουν πάνω από μία 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.

Έστω f_i η συχνότητα ενός διαστήματος και f_{i-1}, f_{i+1} οι συχνότητες των διαστημάτων που βρίσκονται αριστερά και δεξιά από αυτό.

Τότε η επικρατούσα τιμή είναι:

$$m = \frac{1}{n}$$

$$X \sim N\left(m, s^2\right)$$

$$\sum_{i=1}^{n}(x_i - \bar{X})^2 = \frac{1}{n}\cdot\left(\sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n} x_i\bar{X} + \sum_{i=1}^{n}\bar{X}^2\right) = \frac{1}{n}\cdot\left(\sum_{i=1}^{n} x_i^2 - 2\bar{X}n\bar{X} + \sum_{i=1}^{n}\bar{X}^2\right) = \frac{1}{n}\cdot\left(\sum_{i=1}^{n} x_i^2 - nx^2\right) = v_2 - v_1^2$$

$$t = L_i + d_i\frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})}$$

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

1. Μπορεί να υπολογιστεί και για κατανομές που είναι ανοιχτές προς τα πάνω ή προς τα κάτω.
2. Μπορεί να χρησιμοποιηθεί και 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 T

Για το δείγμα ο αντίστοιχος συμβολισμός είναι:

- Αριθμητικός Μέσος $\bar{X}$
- Διάμεσος m
- Επικρατούσα τιμή t

Γενικά για έναν πληθυσμό ισχύει:

- Όταν μ = M = T, τα δεδομένα προέρχονται από μία συμμετρική κατανομή.
- Όταν έχω θετική ασυμμετρία, ισχύει μ > M > T.
- Όταν έχω αρνητική ασυμμετρία, ισχύει αντίστροφα μ < M < T.

Για μια κατανομή με μικρή ή μέτρια ασυμμετρία ισχύει ο εμπειρικός τύπος: $\bar{X} - t = 3(\bar{X} - m) \Rightarrow 3m - 2\bar{X} = t$.

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

<u>Διαφορά μέσης τιμής και διαμέσου.</u>

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 διαμέσου 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 ($\overline{X}$) και διαμέσου (M) μπορεί να μας πληροφορήσει για την συμμετρία (όταν $\overline{X} = M$), δεξιά ασυμμετρία (όταν $\overline{X} > M$) και αριστερή ασυμμετρία (όταν $\overline{X} < M$) της κατανομής των δεδομένων.

IV. Ποσοστημόρια.

Για μια ολοκληρωμένη περιγραφή της κατανομής της συνολικής μοναδιαίας πιθανότητας κατά μήκος του άξονα τω 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:

$$\mathbf{P\{X < x_k\} \leq k \leq P\{ X \leq x_k\}} \text{ .}$$

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:

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

(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 **ενδοτεταρτομοριακό πλάτος** (αντίστοιχα τεταρτημοριακή απόκλιση). Και τα δύο αυτά ποσοτικά μεγέθη χρησιμοποιούνται ως μέτρα της μεταβλητότητας της κατανομής.

Τεταρτημόρια.

Όπως εξηγήσαμε, η διάμεσος χωρίζει το σύνολο των δεδομένων σε δύο 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.

1ο 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 τεταρτημόριου διακρίνουμε, όπως και στη διάμεσο, δύο περιπτώσεις:

α) Όταν το πλήθος των παρατηρήσεων δεν εμφανίζεται σε μορφή κατανομής συχνοτήτων, τότε η θέση του πρώτου

τεταρτημορίου καθορίζεται από τον αριθμό: $\frac{N+1}{4}$.

β) 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.

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

3ο τεταρτημόριο (Q_3).

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

α) Όταν το πλήθος των παρατηρήσεων είναι μικρό

Τότε η θέση του τρίτου τεταρτημορίου καθορίζεται από τον αριθμό: $\frac{3(N+1)}{4}$.

β) Όταν οι τιμές των παρατηρήσεων ξεπερνούν τις 30

Τοποθετούμε αυτές τις τιμές σε μορφή κατανομής συχνοτήτων και εφαρμόζουμε τον παρακάτω τύπο:

$$Q_3 = a_{i-1} + \frac{d}{f_i}\left(\frac{3N}{4} - F_{i-1}\right)$$

- ***Το 2^o τεταρτημόριο συμπίπτει με τη διάμεσο.***

<u>Δεκατημόρια – Εκατοστημόρια.</u>

Εάν το πλήθος των στατιστικών δεδομένων είναι μεγάλο, ενδιαφέρον παρουσιάζει πολλές φορές ο υπολογισμός και άλλων παραμέτρων θέσεως εκτός από τη διάμεσο Μ και τα δύο 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 αυξανομένου μεγέθους) δεδομένων, σε 10 διαδοχικά υποσύνολα, το κάθε ένα από τα υποσύνολα αυτά θα περιλαμβάνει το 10% των όρων της σειράς (ή κατανομής). Τα πρώτο δεκατημόριο D_1 βρίσκεται στη 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 στη 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 κατανομής, είναι:

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

όπου k = 1,2,…,9.

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:

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

Παρατήρηση.

Όλοι οι τύποι υπολογισμού της διαμέσου και των τεταρτημορίων, δεκατημορίων ναι εκατοστημορίων μιας κατανομής, έχουν, γενικά, την ίδια «δομή». Το μόνο σημείο στο οποίο διαφέρουν είναι ο μειωτέος στην παρένθεση, ο οποίος καθορίζει και τη 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.

6.1.1.3 Μέτρα Διασποράς.

I. Διακύμανση.

Ορισμός.

Η Διακύμανση ή Διασπορά s^2 (Second Moment or Variation) 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 τη συμβολίζουμε με s^2. Ορίζεται από την σχέση:

$$s^2 = \frac{1}{v}\sum_{i=1}^{v} x_i^2 \left(x_i - \overline{X}\right)^2$$

Ο τύπος αυτός μπορεί να μετασχηματισθεί κάνοντας τις πράξεις και ως εξής:

$$s^2 = \frac{1}{v}\left[\sum_{i=1}^{v} x_i^2 - \frac{1}{v}\left(\sum_{i=1}^{v} x_i\right)^2\right]$$

ή $s^2 = \frac{1}{v}\sum_{i=1}^{v} x_i^2 - \overline{X}^2 \quad b_2 = \frac{1}{n} * \frac{\sum_{i=1}^{k} f_i \left(w_i - \overline{X}\right)^4}{s^4}$

Εάν οι παρατηρήσεις έχουν συχνότητες v_i οι τύποι γράφονται ως εξής:

$$s^2 = \frac{1}{v}\sum_{i=1}^{m} \left(x_i - \overline{X}\right)^2 * v_i$$

$$s^2 = \frac{1}{v}\left[\sum_{i=1}^{v} x_i^2 * v_i - \frac{1}{v} * \left(\sum_{i=1}^{v} x_i * v_i\right)^2\right]$$

$$s^2 = \frac{1}{v}\sum_{i=1}^{v} x_i^2 * v_i - \overline{X}^2$$

Ανάλυση διακύμανσης κατά ένα παράγοντα. Αυτό το εργαλείο εκτελεί μια απλή ανάλυση διακύμανσης, επαληθεύοντας την υπόθεση ότι οι μέσες τιμές δύο ή περισσοτέρων δειγμάτων είναι ίσες (εφόσον λαμβάνονται από πληθυσμούς με την ίδια μέση τιμή). Η τεχνική αυτή επεκτείνεται στις δοκιμές δύο μέσων τιμών, όπως ο έλεγχος t.

Ανάλυση διακύμανσης δύο παραγόντων με αλληλεπίδραση. Αυτό το εργαλείο ανάλυσης εκτελεί μια παραλλαγή, δύο παραγόντων με αναπαραγωγή, της ανάλυσης διακύμανσης ενός παράγοντα, που περιλαμβάνει περισσότερα από ένα δείγματα για κάθε ομάδα δεδομένων.

Ανάλυση διακύμανσης δύο παραγόντων χωρίς αλληλεπίδραση. Αυτό το εργαλείο ανάλυσης εκτελεί μια ανάλυση διακύμανσης δύο παραγόντων, η οποία δεν περιλαμβάνει περισσότερες από μία δειγματοληψίες ανά ομάδα, κάνοντας δοκιμή της υπόθεσης ότι οι μέσες τιμές δύο ή περισσοτέρων δειγμάτων είναι ίσες (εφόσον λαμβάνονται από πληθυσμούς με την ίδια μέση τιμή). Η τεχνική αυτή επεκτείνεται σε δοκιμές για δύο μέσες τιμές, όπως ο έλεγχος t.

II. Τυπική απόκλιση.

Η θετική τετραγωνική ρίζα της διασποράς. Χρησιμοποιείται πιο συχνά από τη διασπορά, γιατί έχει τις ίδιες μονάδες μέτρησης με το μέσο.

Συμβολίζεται με ΔX ή με σ(X) ή σ.

***** Σημείωση: Εάν η X εκφράζεται σε κάποιες φυσικές μονάδες, τότε η τυπική απόκλιση σ εκφράζεται στις ίδιες μονάδες.

Ο γενικός τύπος είναι:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum (x_i - m)^2}{\mathrm{N}}}$$

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.

III. Ενδοτεταρτομοριακό 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 ημιεδοτεταρτημοριακό εύρος και συμβολίζεται με Q, 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.

IV. <u>Εύρος μεταβολής</u>**.**

Είναι το μέτρο που μας δίνει τη διαφορά μεταξύ της μέγιστης και της ελάχιστης τιμής ενός πληθυσμού.

$R = M – E$

*Παρατήρηση: Επειδή εξαρτάται μόνο από τις ακραίες τιμές μιας σειράς παρατηρήσεων, το εύρος μεταβολής δεν παρέχει την απαιτούμενη αξιοπιστία ως μέτρο διασποράς. Για το λόγο αυτό, η χρησιμότητά του περιορίζεται σε ελάχιστες εφαρμογές.

6.1.1.4 Μέτρα λοξότητας ή ασυμμετρίας.

I. <u>Συντελεστής ασυμμετρίας του Pearson.</u>

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 από το άλλο. Υπάρχουν δύο ειδών ασυμμετρίες, η θετική ασυμμετρία στην οποία οι 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:

$$S_k = \frac{m - \mathrm{M}_0}{s}$$

Με βάση τις ροπές, ο συντελεστής ορίζεται ως:

$$g = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^3}{\left\{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2}\right\}^3}$$

Όταν γ > 0 έχουμε θετική ασυμμετρία, όταν γ < 0 έχουμε αρνητική ασυμμετρία, ενώ για γ = 0 έχουμε συμμετρία.

6.1.1.5 Μέτρα Κύρτωσης.

I. Συντελεστής Κύρτωσης.

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

- λεπτόκυρτες
- μεσόκυρτες
- πλατύκυρτες

Ο συντελεστής κύρτωσης β_2 ορίζεται από τη σχέση: $b_1 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{X}}{s}\right)^4 = \frac{1}{n} * \frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^4}{s^4}$

Για ομαδοποιημένα δεδομένα ο συντελεστής γράφεται: $b_2 = \frac{1}{n} * \frac{\sum_{i=1}^{k} f_i\left(w_i - \bar{X}\right)^4}{s^4}$

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

- Για μεσόκυρτες κατανομές, όπως η κανονική, ισχύει $β_2 = 3$.
- Όταν η τιμή του συντελεστή είναι $β_2 < 3$, τότε η κατανομή είναι πλατύκυρτη.
- Όταν η τιμή του συντελεστή είναι $β_2 > 3$, τότε η κατανομή είναι λεπτόκυρτη.

7. Διασπορά

7.1 Έννοια της διασποράς

Ο μέσος αριθμητικός, ο μέσος γεωμετρικός, η διάμεσος, η επικρατούσα τιμή και τα υπόλοιπα άλλα μέτρα τα οποία έχουμε εξετάσει στο προηγούμενο κεφάλαιο, έχουν ως αντικειμενικό σκοπό να αντιπροσωπεύσουν έναν πληθυσμό με μία 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 το μέσο αριθμητικό ονομάζεται *διασπορά.*

Υποθέτουμε ότι οι μεταβλητές *X* και Y 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 δύο μεταβλητές ο μέσος αριθμητικός είναι 45 και η διάμεσος 48, η μια όμως μεταβλητή έχει διάφορη μορφή από την άλλη.

Στη μεταβλητή *X* οι τιμές των παρατηρήσεων κυμαίνονται μέσα στους αριθμούς 10 και 45, ενώ στη μεταβλητή Y 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 ή διασποράς 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 ή διακύμανση.

7.2 Μέτρα διασποράς

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

7.2.1 Εύρος μεταβολής

Είναι μια παράμετρος που μας δίνει τη διαφορά ανάμεσα στη μεγαλύτερη και μικρότερη τιμή μιας σειράς παρατηρήσεων.

Π.χ. η μέση ημερήσια θερμοκρασία σε 18 πόλεις της χώρας στις 25 Αυγούστου του 2008 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:

$R = M - E = 35 - 19 = 16$

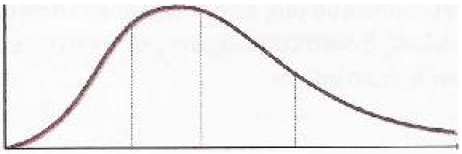
Η παράμετρος όμως αυτή δεν είναι ικανοποιητική, γιατί εξαρτάται μόνο από τις ακραίες τιμές της σειράς των παρατηρήσεων που μας δόθηκε. Αν όμως οι δύο ακραίες τιμές είναι αρκετά απομακρυσμένες από τα σύνολα των άλλων παρατηρήσεων, μας δίνουν ψεύτικη εικόνα της έντασης της διασποράς.

Το εύρος μεταβολής χρησιμοποιείται κυρίως στη Μετεωρολογία και στα χρηματιστήρια.

7.2.2 Ημιενδοτεταρτημοριακό εύρος

Η παράμετρος αυτή ορίζεται ως το ημιάθροισμα της διαφοράς μεταξύ τρίτου και πρώτου τεταρτημόριου (βλ. διάγραμμα), δηλαδή:

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



Διάγραμμα

Όσο μικρότερη είναι η τιμή του Q, τόσο μεγαλύτερη είναι η συγκέντρωση των τιμών της μεταβλητής. Το ημιενδοτεταρτημοριακό εύρος δεν επηρεάζεται από τις ακραίες τιμές, όπως συμβαίνει με το εύρος μεταβολής, και υπολογίζεται και όταν η κατανομή είναι και ανοικτή.

7.2.3 Μέση απόκλιση

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

α) Αταξινόμητες παρατηρήσεις

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

$$M.A = \frac{\Sigma \left| x_i - m \right|}{\mathrm{N}}$$

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:

$$m = \frac{3+5+7+9+16}{5} = \frac{40}{5} = 8$$

Η μέση απόκλιση θα είναι:

$$\mathrm{M.A.} = \frac{\Sigma \left| x_i - m \right|}{\mathrm{N}} = \frac{\left|3-8\right| + \left|5-8\right| + \left|7-8\right| + \left|9-8\right| + \left|16-8\right|}{5} = \frac{18}{5} = 3,6$$

β) Ταξινομημένα στοιχεία σε μορφή κατανομών συχνοτήτων

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

$$\mathrm{M.A.} = \frac{\sum f_i \left| x_i - m \right|}{\sum f_i}$$

7.2.4 Διακύμανση και τυπική απόκλιση.

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 ασχοληθούμε 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 δραχμές, η διακύμανση εκφράζεται με μονάδα τη δραχμή στο τετράγωνο, ομοίως αν η μεταβλητή εκφράζεται σε εκατοστά, η διακύμανση εκφράζεται σε εκατοστά στο τετράγωνο κ.λπ. Για να έχουμε όμως ένα δείκτη ο οποίος να μετράει 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.

7.2.4.1 Υπολογισμός της διακύμανσης και της τυπικής απόκλισης.

Η διακύμανση συμβολίζεται με σ^2 ή *Var(x)* και η τυπική απόκλιση με *σ.*

α) Αταξινόμητες παρατηρήσεις

Υποθέτουμε ότι έχουμε τις παρατηρήσεις x_1, x_2,..., x_i,.., x_N, που ο μέσος αριθμητικός τους είναι μ.

Η διακύμανση των παραπάνω παρατηρήσεων δίνεται από τον τύπο:

$$s^2 = \frac{\sum (x_i - m)^2}{\mathrm{N}}$$

και η τυπική απόκλιση από τον τύπο:

$$s^2 = \sqrt{s^2} = \frac{\sqrt{\Sigma (x_i - m)^2}}{N}$$

β) Ομαδοποιημένες παρατηρήσεις

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

$$s^2 = \frac{\sum f_i (x_i - m)^2}{\sum f_i}$$

$$s^2 = \frac{\sum f_i x_i^2}{\sum f_i} - m^2$$

$$s^2 = d^2 \left[\frac{\sum f_i x_i^2}{\sum f_i} - \left(\frac{\sum f_i x_i}{\sum f_i} \right)^2 \right]$$

7.2.4.2. Ιδιότητες της διακύμανσης

1. *Αν* $x_1 = x_2 = x_3 = \ldots = x_v = \alpha$, *τότε η διακύμανση είναι μηδέν, δηλαδή η διακύμανση μιας σταθεράς είναι μηδέν.*

2. *Αν όλες οι τιμές μιας μεταβλητής* X *πολλαπλασιαστούν με τον ίδιο αριθμό* α, *η διακύμανση πολλαπλασιάζεται με το τετράγωνο του αριθμού αυτού, δηλαδή:*

Var(aX) = α². Var(X)

3. *Αν όλες οι τιμές μιας μεταβλητής* Χ *αυξηθούν κατά μια σταθερά* α, *τότε η διακύμανση δεν μεταβάλλεται, δηλαδή:*

Var(X +α) = Var(X)

Παράδειγμα

Αν οι τιμές μιας μεταβλητής αυξηθούν κατά 8 μονάδες ή κατά 8%, ποια μεταβολή επέρχεται στο μέσο αριθμητικό και στη διακύμανση;

<u>Λύση</u>

Αν X_i είναι οποιαδήποτε τιμή της μεταβλητής Χ πριν από την αύξηση, μετά την αύξηση θα είναι $Y_i = x_i + 8$ αν κάθε τιμή αυξηθεί κατά 8 μονάδες, και $Y_i = x_i + 0.08x_i$ αν κάθε τιμή αυξηθεί κατά 8%. Αν κάθε τιμή αυξηθεί κατά 8 μονάδες, ο μέσος αριθμητικός θα είναι:

$$m_y = \frac{\sum y_i}{N} = \frac{\sum (x_i + 8)}{N} = \frac{\sum x_i}{N} + \frac{N*8}{N} = m_x + 8$$

Αν στη συνέχεια κάθε τιμή αυξηθεί κατά 8%, ο μέσος αριθμητικός θα είναι:

$$m_y = \frac{\sum y_i}{N} = \frac{\sum (x_i + 0.08x_i)}{N} = \frac{\sum x_i}{N} + 0.08\frac{N*8}{N} = m_x + 0,08m_x$$

δηλαδή ο μέσος αριθμητικός αυξήθηκε κατά 8%. Επίσης, η διακύμανση της μεταβλητής *Y* θα είναι:

α) Αν κάθε τιμή αυξηθεί κατά 8 μονάδες:

$$s\,y^2 = \frac{\sum (y_i - m_y)^2}{N} = \frac{\sum \left[x_i + 8 - \left(m_x + 8\right)\right]^2}{N} = \frac{\sum (x_i - m_x)^2}{N} = s\,_x^2$$

Επομένως, καμιά μεταβολή δεν επέρχεται στη διακύμανση.

β) Αν οι τιμές της μεταβλητής X αυξηθούν κατά 8%:

$$s\,y^2 = \frac{\sum (y_i - m_y)^2}{N} = \frac{\sum \left[x_i + 0,08x_i - \left(m_x + 0,08m_x\right)\right]^2}{N} = 1,08^2\,\frac{\sum (x_i - m_x)^2}{N} = 1,08s\,_x^2$$

Επομένως, η διακύμανση έχει μεταβληθεί.

4. *Έστω ότι έχουμε κ κατανομές με πλήθος παρατηρήσεων N_1, N_2, N_3,..., $N_κ$, με διακυμάνσεις $σ_1^2$, $σ_2^2$, $σ_3^2$, . . . ,$σ_κ^2$, και μέσους $μ_1$, $μ_2$, . . . , $μ_i$, . . . , $μ_κ$. Η συνολική διακύμανση όλων των υποπληθυσμών δίνεται από την παρακάτω σχέση:*

$$s\,_2 = \frac{1}{\mathrm{N}}\sum \mathrm{N}_i s\,_i^2 + \frac{1}{N}\sum N_i\left(m_i - m\right)^2$$

όπου:

$$m = \frac{\mathrm{N}_1 m_1 + \mathrm{N}_2 m_2 + \ldots + \mathrm{N}_k m_k}{\mathrm{N}_1 + \mathrm{N}_2 + \ldots + \mathrm{N}_\mathrm{K}}$$

Η διακύμανση όλων των υποπληθυσμών δίνεται από την παρακάτω σχέση, αν οι υποπληθυσμοί είναι *κ:*

$$s_2 = \frac{1}{\mathrm{N}} \sum_{i=1}^{k} \mathrm{N}_i s_i^{2} + \frac{1}{N} \sum_{i=1}^{k} \mathrm{N}_i \left(m_i - m \right)^2$$

7.2.5 Συντελεστής μεταβλητικότητας

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 την 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 δύο κατανομές οι οποίες εκφράζονται σε διαφορετικές μονάδες (μέτρα, κιλά, δραχμές, εκατοστά κ.λπ.), ή όταν οι μέσοι αριθμητικοί δύο 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 ο *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:

$$CV(\mathrm{X}) = \frac{s}{m} \quad \text{ή} \quad CV(\mathrm{X}) = \frac{s}{m} 100\%$$

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

Παράδειγμα

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

X_i : 44,37,36,46,42

Y_i : 88,79,86,75,84

Ποια από τις δύο μεταβλητές παρουσιάζει μεγαλύτερη ομοιογένεια;

<u>Λύση</u>

Απάντηση θα δώσει ο συντελεστής μεταβλητικότητας. Για το λόγο αυτό σχηματίζουμε τον πίνακα.

Πίνακας

	x_i	y_i	$x_i - \mu_x$	$(x_i - \mu_x)^2$	$y_i - \mu_y$	$(y_i - \mu_y)^2$
	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

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

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

$$s_{x^2} = \frac{\sum (x_i - m_X)^2}{N} = \frac{76}{5} = 15,2$$ και $σ_x$= 3,9

$$s_{y^2} = \frac{\sum (y_i - m_y)^2}{N} = \frac{110}{5} = 22$$ και $σ_x$= 4,69

Επομένως, οι συντελεστές μεταβλητικότητας θα είναι:

$$CV(X) = \frac{s_X}{m_X} 100\% = \frac{3,9}{41} * 100\% = 9,5\%$$

$$CV(Y) = \frac{s_y}{m_y} 100\% = \frac{4,69}{82} * 100\% = 5,7\%$$

Επομένως, η μεταβλητή Υ παρουσιάζει τη μικρότερη διασπορά.

8. Ασυμμετρία – Κύρτωση

8.1 Ασυμμετρία

Τα μέτρα τάσης και διασποράς μιας κατανομής, δίνουν μια πρώτη εικόνα της μορφής της. Βέβαια, η εικόνα αυτή βελτιώνεται αν προσδιορίσουμε και ένα μέτρο ασυμμετρίας της. Με άλλα λόγια, αν προσδιορίσουμε πόσο και προς ποια κατεύθυνση αποκλίνει η κατανομή μας από την πλήρως συμμετρική κατανομή. Δύο ή περισσότερες κατανομές συχνοτήτων είναι δυνατό να έχουν την ίδια μέση τιμή και την ίδια διασπορά και να μη συμπίπτουν, αν δεν παρουσιάζουν τον ίδιο βαθμό ασυμμετρίας.

Για να εννοήσουμε την ασυμμετρία, πρέπει να έχουμε υπόψη τι εννοούμε λέγοντας ότι μια κατανομή είναι συμμετρική.

Συμμετρική είναι μια κατανομή όταν οι τιμές της τοποθετούνται συμμετρικά γύρω από τη μέση αριθμητική τιμή, οπότε 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 από το άλλο.

Η επιδίωξή μας είναι να προσδιορίσουμε ένα μέτρο, που να μας πληροφορεί για το αν οι τιμές της μεταβλητής τοποθετούνται συμμετρικά γύρω από το μέσο ή όχι και πόσο. Αυτό το κατορθώνουμε με το συντελεστή ασυμμετρίας, ο οποίος συμβολίζεται με το Sκ, η δε τιμή του είναι μέτρο αφηρημένο και προσδιορίζει το βαθμό ασυμμετρίας μιας κατανομής συχνοτήτων.

Το μέτρο, το οποίο προτάθηκε από τον K. Pearson και βασίζεται στη σχέση που υπάρχει μεταξύ του αριθμητικού μέσου (μ) και της επικρατούσας τιμής (M_0) για το ως μέτρο ασυμμετρίας δίνεται από τον τύπο:

$$S_k = \frac{m - \mathrm{M}_0}{s}$$

Άλλος δείκτης ασυμμετρίας δίνεται από τον τύπο:

$$S_k = \frac{(Q_3 - M) - (M - Q_1)}{(Q_3 - M) + (M - Q_1)}$$ (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 Sκ 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.

Υπάρχουν δύο ειδών ασυμμετρίες, η θετική ασυμμετρία και η αρνητική. Θετικά ασυμμετρική είναι μια κατανομή όταν παρουσιάζει εξόγκωση προς τα αριστερά και επιμήκυνση του άκρου της που αντιστοιχεί στις μεγαλύτερες τιμές του χαρακτηριστικού. Αυτό σημαίνει πως η μεγάλη συγκέντρωση των παρατηρήσεων βρίσκεται στις μικρές τιμές της μεταβλητής. Στην περίπτωση αυτή ισχύει:

$$M_0 < x_d < \bar{X}$$

Αντίθετα, στην αρνητική ασυμμετρία η κατανομή παρουσιάζει εξόγκωση προς τα δεξιά και επιμήκυνση του άκρου της που αντιστοιχεί στις μικρότερες τιμές του χαρακτηριστικού. Με άλλα λόγια, η συγκέντρωση των παρατηρήσεων βρίσκεται στις μεγάλες τιμές της μεταβλητής. Στην περίπτωση αυτή ισχύει:

$$\bar{X} < x_d < M_0$$

Σαν αριθμητικό μέτρο καθορισμού της ασυμμετρίας συνηθέστερο είναι ο συντελεστής ασυμμετρίας με βάση τις ροπές, ο οποίος ορίζεται ως:

$$g = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^3}{\left\{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2}\right\}}$$

Όταν γ > 0 έχουμε θετική ασυμμετρία, όταν γ < 0 έχουμε αρνητική, ενώ όταν γ = 0 υπάρχει συμμετρία.

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

8.2 Κύρτωση

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:

$$b_1 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \overline{X}}{s}\right)^4 = \frac{1}{n}\cdot\frac{\sum_{i=1}^{n}(X_i - \overline{X})^4}{s^4}$$

Από τον τύπο παρατηρούμε ότι αν $z_i = \frac{x_i - \overline{X}}{s}$ είναι οι τυποποιημένες τιμές που αντιστοιχούν στα δεδομένα, τότε

$$b_2 = \frac{1}{n}\cdot\frac{\sum_{i=1}^{k} f_i(z_i - \overline{X})^4}{s^4}$$

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

- Για μεσόκυρτες κατανομές, όπως π.χ. η κανονική, ισχύει $\beta_2 = 3$.
- Όταν η τιμή του συντελεστή είναι $\beta_2 < 3$, τότε η κατανομή είναι πλατύκυρτη.
- Όταν η τιμή του συντελεστή είναι $\beta_2 > 3$, τότε η κατανομή είναι λεπτόκυρτη.

Οι συντελεστές ασυμμετρίας και κύρτωσης χρησιμοποιούν τα αθροίσματα $\sum_{i=1}^{n}(X_i - \overline{X})^3$ και $\sum_{i=1}^{n}(X_i - \overline{X})^4$ τα οποία είναι παραδείγματα **ροπών** για ένα σύνολο δεδομένων.

Ορισμός.

Γενικά η ροπή τάξης r γύρω από το σημείο x0 μιας κατανομής είναι το άθροισμα $\frac{\sum_{i=1}^{n}(x_i - x_0)^r}{n}$.

§ Δύο ειδικές περιπτώσεις είναι:

Α. $x_0 = 0$, οπότε σε αυτήν την περίπτωση έχω ροπές γύρω από την αρχή (γύρω από το μηδέν) και η ροπή τάξης r θα είναι: $v_r = \frac{\sum_{i=1}^{n} x_i^r}{n}$.

Για r = 1, ισχύει $v_1 = \frac{\sum_{i=1}^{n} x_i}{n}$, για r = 2, ισχύει $v_2 = \frac{\sum_{i=1}^{n} x_i^2}{n}$ κ.ο.κ.

Β. $x_0 = \overline{X}$, οπότε έχω την κεντρική ροπή (ροπή γύρω από το μέσο) τάξης r, $m_r = \frac{1}{n} \cdot \sum_{i=1}^{n}(X_i - \overline{X})^r$.

Για r = 1, έχουμε $m_1 = \frac{1}{n} \cdot \sum_{i=1}^{n}(X_i - \overline{X}) = \frac{1}{n} \cdot \left(\sum_{i=1}^{n} X_i - n\overline{X} \right) = 0$.

Για r = 2, έχουμε $m_2 = \frac{1}{n} \cdot \sum_{i=1}^{n}(X_i - \overline{X})^2$.

Για r = 3, έχουμε $m_3 = \frac{1}{n} \cdot \sum_{i=1}^{n}(X_i - \overline{X})^3$ κ.ο.κ.

Αντίστοιχα, ορίζονται και οι ροπές για έναν **πληθυσμό**.

Π.χ. η ροπή 1ης τάξης γύρω από το μηδέν θα είναι: $m = \frac{\sum_{i=1}^{N} X_i}{N}$.

Η ροπή 2ης τάξης γύρω από τη μέση τιμή θα υπολογίζεται από τον τύπο:

$\frac{1}{N} \cdot \sum_{i=1}^{N} (x_i - m)^2 = s^2$, δηλαδή την πληθυσμιακή διακύμανση.

Αν έχουμε ομαδοποιημένα δεδομένα, κ διαστήματα και f_i είναι η απόλυτη συχνότητα του διαστήματος i, τότε η ροπή τάξης r γύρω από το σημείο x_0 είναι: $\frac{1}{n} \cdot \sum_{i=1}^{k} f_i (w_i - x_0)^r$

όπου w_i είναι η κεντρική τιμή του διαστήματος.

Για $x_0 = 0$, έχω τη ροπή τάξης r γύρω από την αρχή $\frac{1}{n} \cdot \sum_{i=1}^{k} f_i \, w_i^r$.

Ο λόγος $\frac{f_i}{n}$ είναι η σχετική συχνότητα για το διάστημα i και συχνά συμβολίζεται με p_i, αφού εκτιμά την πιθανότητα η μεταβλητή που εξετάζεται να πάρει τιμή στο διάστημα i.

Επομένως, γράφοντας $p_i = \frac{f_i}{n}, i = 1, 2, ..., k$, η ροπή της τάξης r γύρω από την αρχή διαμορφώνεται ως: $v_r = \sum_{i=1}^{k} p_i w_i^r$

Για ομαδοποιημένα δεδομένα, η κεντρική ροπή r τάξης είναι: $m_r = \frac{1}{n} \cdot \sum_{i=1}^{k} f_i (w_i - \bar{X})^r = \sum_{i=1}^{k} p_i (w_i - \bar{X})^r$.

<u>Σχέση μεταξύ κεντρικών ροπών και ροπών γύρω από το μηδέν.</u>

Όταν γνωρίζουμε τις ροπές γύρω από την αρχή, v_k, για k = 1,2,…,r, μπορούμε εύκολα να υπολογίσουμε τις ροπές μ_k γύρω από το μέσο.

- Για r = 1,

$$m = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{X}) = \frac{1}{n} \cdot \left(\sum_{i=1}^{n} x_i - n\bar{X} \right) = 0$$

- Για r = 2,

$$m = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{X})^2 = \frac{1}{n} \cdot \left(\sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n} x_i \bar{X} + \sum_{i=1}^{n} \bar{X}^2 \right) = \frac{1}{n} \cdot \left(\sum_{i=1}^{n} x_i^2 - 2\bar{X}n\bar{X} + \sum_{i=1}^{n} \bar{X}^2 \right) = \frac{1}{n} \cdot \left(\sum_{i=1}^{n} x_i^2 - nx^2 \right) = v_2 - v$$

9. ΘΕΩΡΙΑ ΠΙΘΑΝΟΤΗΤΩΝ

9.1 Βασικές έννοιες

Κάθε ανθρώπινη ενέργεια οσοδήποτε προγραμματισμένη και αν είναι περικλείει πάντοτε ένα στοιχείο αβεβαιότητας, έτσι ώστε το αποτέλεσμά της να εξαρτάται άλλοτε λιγότερο και άλλοτε περισσότερο από την τύχη, δηλαδή από τους παράγοντες που δε μπορούν να προβλεφθούν/ελεχθούν εκ των προτέρων.

Σε περιπτώσεις που δε μπορούμε να κάνουμε ακριβή πρόβλεψη του αποτελέσματος μιας ενέργειας, έχουμε τη δυνατότητα να μετρήσουμε το βαθμό αβεβαιότητας ενός ορισμένου αποτελέσματος της, αφού σε μια ενέργεια που έχει ακριβή έκβαση, τελικά θα εμφανιστεί ένα μόνο από τα δυνατά ενδεχόμενα της.

Γεγονός, όπως δηλώνει και ο όρος, σημαίνει κάθε τι του οποίου η πραγματοποίηση λαμβάνει χώρα σε κάποια στιγμή ή σε κάποιο διάστημα (είτε του παρελθόντος, είτε του παρόντος, είτε του μέλλοντος) του χρόνου.

Η λέξη **συμβάν** εκφράζει κάθε τι το οποίο συνέβη. Όπως, όμως, και η έννοια του γεγονότος γενικά εκφράζει κάθε τι το οποίο είτε συνέβη, είτε συμβαίνει, είτε θα συμβεί. Η έννοια του συμβάντος είναι έννοια μερικότερη της έννοιας του γεγονότος και χαρακτηρίζεται από την αίσθηση της κινήσεως.

Τυχαίο φαινόμενο λέγεται κάθε μεταβολή (αντιληπτή από τις αισθήσεις των όντων), η οποία συμβαίνει στη φύση ή στην κοινωνία και της οποίας η κατάληξη δεν είναι γνωστή εκ των προτέρων.

Πείραμα τύχης είναι μια διαδικασία όπου η γνώση των συνθηκών, κάτω από τις οποίες εκτελούνται, απλώς καθορίζει ένα σύνολο δυνατών αποτελεσμάτων για το κάθε πείραμα. Ένα τέτοιο πείραμα μπορεί να επαναλαμβάνεται υπό αυτές συνθήκες πάντοτε και του οποίου το αποτέλεσμα δε μπορεί να προβλεφθεί a priori.

Απλά ή στοιχειώδη ενδεχόμενα ονομάζονται τα δυνατά αποτελέσματα ενός πειράματος. (Το i ενδεχόμενο, για παράδειγμα, συμβολίζεται με A_i).

Δειγματικός χώρος είναι το σύνολο των δυνατών αποτελεσμάτων ενός πειράματος. Ο δειγματικός χώρος συμβολίζεται συνήθως με το γράμμα S. Ο δειγματικός χώρος S καλείται συνήθως και **βέβαιο ενδεχόμενο** και είναι ένα σύνολο πεπερασμένο ή απέραντο.

Ενδεχόμενο καλείται κάθε υποσύνολο του δειγματικού χώρου.

- Δύο ενδεχόμενα E_1 και E_2 χαρακτηρίζονται ως ασυμβίβαστα, όταν δε μπορούν να πραγματοποιηθούν και τα δύο. Βέβαια, η μη πραγματοποίηση του ενός, δεν εμποδίζει την πραγματοποίηση του άλλου.
- Δύο ενδεχόμενα E_1 και E_2 λέγονται αντίθετα, όταν η πραγματοποίηση του ενός αποκλείει τη πραγματοποίηση του άλλου ή αντίστροφα την επιβάλλει.

Το ενδεχόμενο $\varnothing$ ονομάζεται αδύνατο ενδεχόμενο.

9.2 Συνοπτική Ιστορική εξέλιξη της Θεωρίας των Πιθανοτήτων

Η έννοια της πιθανότητας εμφανίζεται για πρώτη φορά στις διδασκαλίες του Καρνεάδου, διευθυντού της Ακαδημίας του Πλάτωνος (2ος Π.Χ. αιώνας), ο οποίος εξετάζει την έννοια αυτή από τη φιλοσοφική της υπόσταση. Στοιχεία της Συνδυαστικής Αναλύσεως η οποία αποτελεί τη βάση στον προσδιορισμό της πιθανότητας εμφανίζονται, επίσης, στα έργα του Αμμωνίου, Αριστοτέλους, Διοφάντου, Ευκλέιδου, Πάππου, Πλούταρχου.

Η γένεση του Λογισμού των Πιθανοτήτων οφείλεται στην ενασχόληση του ανθρώπου με τα τυχερά παιχνίδια (ρίψη κύβων, νομίσματος, εξαγωγή μίας ή περισσότερων καρτών από μία τράπουλα κ.λ.π.).

Οι Ιταλοί μαθηματικοί Faccioli (1494), Tartaglia (1556), Peverone (1558), προσπάθησαν να προσδιορίσουν το πλήθος των δυνατών αποτελεσμάτων και από αυτό αόριστα την πιθανότητα εμφανίσεως κάποιου αποτελέσματος.

Ο Laplace (1749-1827) θεωρεί τους Fermat (1608-1665) και Pascal (1623-1662) ως θεμελιωτές των Πιθανοτήτων. Ο ίδιος στο έργο του Theorie Analytique de probalitites (1887) υπογραμμίζει τη χρήση της θεωρίας των πιθανοτήτων και διαμορφώνει τη μαθηματική πλευρά στις εφαρμογές.

Ο Γάλλος Cournot στο έργο του Exposition de la theorie des chances et des porbabilites (1843) συμβάλλει στην καλύτερη θεμελίωση της θεωρίας των Πιθανοτήτων.

Ο Bernoulli στο έργο του Ars Conjectandi (1913) υποδεικνύει συστηματικό τρόπο διδασκαλίας της Συνδυαστικής Αναλύσεως και προσπαθεί να δείξει ότι η σημασία του Λογισμού των Πιθανοτήτων δεν περιορίζεται στα παιχνίδια της τύχης, αλλά υπεισέρχεται σε πολλές πλευρές της καθημερινότητας.

Ανάλογα με την οπτική γωνία με την οποία αντιμετωπίζουν οι διάφορες σχολές τη Θεωρία των Πιθανοτήτων, διακρίνονται τρία είδη της Θεωρίας των Πιθανοτήτων.

I. **Λογική Θεωρία των Πιθανοτήτων.**

Κατά την άποψη αυτή ο Λογισμός των Πιθανοτήτων στηρίζεται σε διαίσθηση και πίστη στο συγκεκριμένο ενδεχόμενο, σε σχέση με κάποιες άλλες ενδείξεις που υπάρχουν.

II. **Εμπειρική Θεωρία των Πιθανοτήτων.**

Εδώ η Θεωρία στηρίζεται στην πίστη σε παρατηρούμενες καταστάσεις των πραγμάτων.

III. **Υποκειμενική Θεωρία των Πιθανοτήτων.**

Η πιθανότητα ενός ενδεχομένου επηρεάζεται από τις προσωπικές εμπειρίες του κάθε ατόμου και συνεπώς, δεν υπάρχει μία αντικειμενική πιθανότητα για το ενδεχόμενο αυτό.

Γενικά, σκοπός του Λογισμού των Πιθανοτήτων είναι η δημιουργία επιστήμης από την άγνοια η οποία χαρακτηρίζει τα διάφορα φαινόμενα.

Κατά πολλούς οικονομολόγους, η θεωρία των πιθανοτήτων προσδιορίζει και εξετάζει τους νόμους της τύχης και μελετά τυχαία γεγονότα. Χρησιμοποιείται ευρέως σε όλο τα τα πεδία της ανθρώπινης δραστηριότητας. Στη μελέτη τυχαίων φαινομένων και πειραμάτων το πρόβλημα βρίσκεται στον προσδιορισμό του δειγματικού χώρου, στην επιλογή των ενδεχομένων που μας ενδιαφέρουν και στον υπολογισμό των πιθανοτήτων που έχουν αυτά να συμβούν.

Πριν ορίσουμε την πιθανότητα θα πρέπει να σημειώσουμε ότι οι διάφοροι ερευνητές έχουν προσεγγίσει την έννοια της πιθανότητας με διαφορετικούς τρόπους και κατά συνέπεια υπάρχουν διαφορετικοί ορισμοί.

Ο κλασικός ορισμός της πιθανότητας διατυπώθηκε κατά πρώτον από το θεμελιωτή του

Λογισμού των Πιθανοτήτων Laplace (1812) και έχει ως εξής:

Η πιθανότητα P(E) ενός γεγονότος (ή συμβάντος ή ενδεχομένου) ορίζεται ως ο λόγος του ρυθμού n_E δι' αυτό περιπτώσεων προς το πλήθος N_E των δυνατών περιπτώσεων, με την προϋπόθεση ότι όλες οι περιπτώσεις είναι εξίσου πιθανές.

Είναι δηλαδή:

$$F_n\left(E_i, s\right) = \frac{f_n\left(E_i, s\right)}{n} \quad P(E) = \frac{n_E}{N_E}$$

Οι αριθμοί n_E και N_E προσδιορίζονται εκ των προτέρων (a priori) εκ της φύσεως του πειράματος και όλες οι δυνατές και ευνοϊκές περιπτώσεις θεωρούνται ισοδύναμες ή ισοπίθανες. Αν, όμως, το πείραμα δεν είναι αυστηρώς καθορισμένο, τότε οι αριθμοί n_E και N_E (και κατ' επέκταση η πιθανότητα P(E)) δεν προσδιορίζονται μονοσήμαντα.

9.3 Στατιστικός ορισμός της Πιθανότητας

Το 1919 ο R. Von Mises διατυπώνει ένα νέο ορισμό με βασική σκέψη την εναρμόνιση του νέου αυτού ορισμού προς τα στατιστικά κυρίως δεδομένα, των όποιων οι παρατηρήσεις δείχνουν μία τάση σταθεροποίησης των συχνοτήτων επαναλήψεως.

Αναχωρεί από μία απείρως εκτεταμένη σειρά ομοιογενών παρατηρήσεων $E_1, E_2, \ldots, E_n, \ldots$ σε κάθε στοιχείο της οποίας αντιστοιχεί ένα σήμα $\sigma_i / i=1,2,\ldots n,\ldots$ Το σύνολο Σ όλων των δυνατών σημάτων ονομάζεται χώρος σημάτων. Αν μεταξύ των n πρώτων ενδεχομένων $E_1, E_2, \ldots, E_n, \ldots$ το πλήθος των εμφανιζόντων το συγκεκριμένο σήμα σ είναι f_n (E_i , σ), τότε το πηλίκο

$$F_n\left(E_i, s\right) = \frac{f_n\left(E_i, s\right)}{n}$$

Ονομάζεται μέση συχνότητα εμφανίσεως του σ. Αν, τώρα, το n τείνει στο άπειρο και το πηλίκο τείνει σε κάποιο πεπερασμένο όριο P_n (E_i , σ), το όριο αυτό ονομάζεται στατιστική πιθανότητα εμφανίσεως του σήματος σ.

9.4 Ενδεχόμενα

Δύο ενδεχόμενα Α και Β ενός χώρου πιθανοτήτων (Ρ (Ω), Ω, Ρ) ονομάζονται *στατιστικά ή στοχαστικά ανεξάρτητα* όταν και μόνο όταν ισχύει η σχέση: $P\{A \cap B\} = P\{A\} P\{B\}$.

Αν αληθεύει η παραπάνω σχέση και $P\{B\} \neq 0$, θα έχουμε:

$$P\{A/B\} = \frac{P\{A \cap B\}}{P\{B\}} = \frac{P\{A\} P\{B\}}{P\{B\}} = P\{A\}$$

Δηλαδή το ενδεχόμενο Α είναι ανεξάρτητο του Β.

Επομένως, αν Ρ{Α / Β }= Ρ{Α} και Ρ{Β / Α}= Ρ{Β}, τα ενδεχόμενα Α και Β είναι ανεξάρτητα μεταξύ τους και η πραγματοποίηση του ενός δεν επηρεάζει την πιθανότητα πραγματοποίησης του άλλου ενδεχομένου.

Η έννοια της ανεξαρτησίας ισχύει και για περισσότερο από 2 ενδεχόμενα.

Τα ενδεχόμενα $A_1, A_2, A_3, \ldots, A_v$ είναι ανεξάρτητα μεταξύ τους αν και μόνο αν για κάθε υποσύνολο κ $(2 \leq k \leq n)$ προκύπτει:

$$\mathrm{P}\{\mathrm{A}_1 \cap \mathrm{A}_2 \cap \mathrm{A}_3 \cap \ldots \cap \mathrm{A}_k\} = \mathrm{P}\{\mathrm{A}_1\}\mathrm{P}\{\mathrm{A}_2\}\mathrm{P}\{\mathrm{A}_3\} \ldots \mathrm{P}\{\mathrm{A}_k\}$$

Αν τα ενδεχόμενα είναι ανεξάρτητα ανά δύο μεταξύ τους, $\mathrm{P}\{\mathrm{A}_i \cap \mathrm{A}_j\} = \mathrm{P}\{\mathrm{A}_i\} * \mathrm{P}\{\mathrm{A}_j\} \ (i \neq j)$, δε σημαίνει ότι υπάρχει πλήρης ανεξαρτησία.

Ο αριθμός των σχέσεων, για να έχουμε πλήρη ανεξαρτησία των ν ενδεχομένων, δίνεται από τη σχέση:

$$\sum_{k=2}^{n} C_k^n = \sum_{k=0}^{n} C_k^n - C_0^n - C_1^n = 2^n - (n+1)$$

9.4.1 Πράξεις με ενδεχόμενα

Η πιθανότητα μπορεί να θεωρηθεί ως ένα αριθμητικό μέτρο που εκφράζει τη τύχη του να πραγματοποιηθεί ένα **ενδεχόμενο**, μέσα από ένα σύνολο εναλλακτικών δυνατών ενδεχομένων.

Έτσι, το πρώτο απαραίτητο στοιχείο στη διατύπωση ενός πιθανοκρατικού προβλήματος είναι ο προσδιορισμός του συνόλου των δυνατών ενδεχομένων ενός φαινόμενου καθώς και του ενδεχομένου που ενδιαφέρει. Στη συνέχεια, κάθε ενδεχόμενο μέσα στο συγκεκριμένο χώρο δυνατότητας συσχετίζεται με μια τιμή: **την πιθανότητα ότι το ενδεχόμενο θα συμβεί**.

Για τα πιθανοκρατικά προβλήματα ισχύει:

-Κάθε πρόβλημα ορίζεται αναφορικά με ένα συγκεκριμένο χώρο δυνατότητας (που περιέχει πάνω από μια δυνατότητα).Ένα ενδεχόμενο αποτελείται από ένα ή περισσότερα δυνατά αποτελέσματα μέσα από αυτό το χώρο δυνατότητας.

-Η πιθανότητα ενός ενδεχομένου εξαρτάται από τις πιθανότητες των ξεχωριστών αποτελεσμάτων μέσα σε ένα δοσμένο χώρο δυνατότητας.

Για τον υπολογισμό της **πιθανότητας** ενός ενδεχομένου είναι απαραίτητος ένας βασικός κανόνας, σύμφωνα με τον οποίο να καθορίζονται οι πιθανότητες των διαφόρων δυνατών αποτελεσμάτων. Ο κανόνας αυτός μπορεί να βασίζεται σε ένα ή και τα δύο από τα παρακάτω:

-Προηγούμενες συνθήκες (ή συμπεράσματα προκαθορισμένων υποθέσεων)

-Αποτελέσματα εμπειρικών παρατηρήσεων

Συμβολισμοί:

$\cap$ Ένωση

$\cup$ Τομή ή Γινόμενο

$\supset$ Ανήκει/Περιέχεται

$\subset$ Περιέχει

Τομή δύο ενδεχομένων A_1 και A_2 καλείται το ενδεχόμενο $A_1 \cap A_2$ το οποίο συμβαίνει όταν συμβαίνουν και τα δύο ενδεχόμενα A_1, A_2.

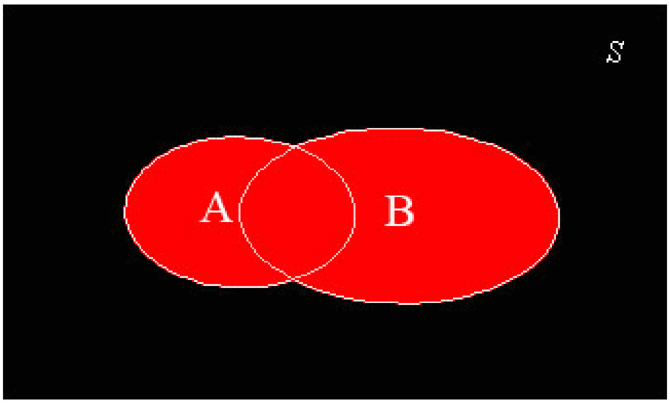
Αν η τομή των δύο ενδεχομένων είναι το αδύνατο ενδεχόμενο τότε αυτά καλούνται ασυμβίβαστα ή ξένα ενδεχόμενα.

Συμπλήρωμα του ενδεχομένου A καλείται το ενδεχόμενο A΄, το οποίο περιέχει τα αποτελέσματα του δειγματικού χώρου S τα οποία δεν περιέχονται στο A.

<u>Συνδυασμός ενδεχομένων:</u>

1. Ένωση ενδεχομένων.

$A \cup B$



Ένωση δύο ενδεχομένων A_1 και A_2 καλείται το ενδεχόμενο $A_1 \cup A_2$ το οποίο συμβαίνει όταν συμβαίνει τουλάχιστον ένα από τα ενδεχόμενα A_1 , A_2.

Ιδιότητες:

$A \cup B = B \cup A$

$A \cup \varnothing = A$

$A \cup A = A$

$(A \cup B) \cup C = A \cup (B \cup C)$

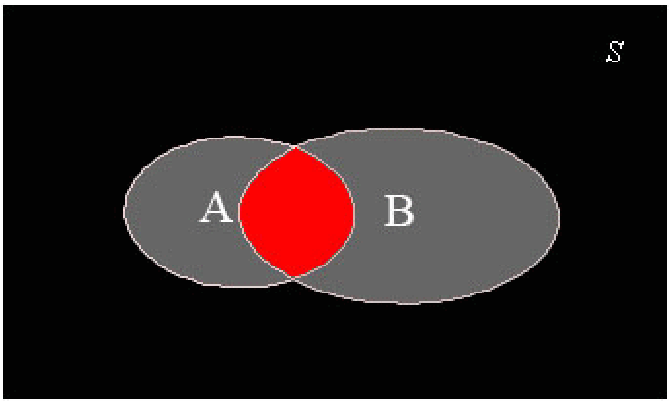
$A \cup S = S$

$A \subset B \Leftrightarrow A \cup B = B$

2. Τομή (ή Γινόμενο Ενδεχομένων):

$A \cap B$

Πραγματοποιούνται ταυτόχρονα τα ενδεχόμενα Α, Β.



Ιδιότητες:

$A \cap B = B \cap A$

$(A \cap B) \cap \Gamma = A \cap (B \cap \Gamma)$

$A \subset B \Leftrightarrow A \cap B = A$

$A \cap B = \varnothing \Leftrightarrow A \subset B^C$

$A \cap \varnothing = \varnothing$

$A \cap S = A$

$A \cap A = A$

$B \cup (A \cap \Gamma) = (\mathrm{B} \cup \mathrm{A}) \cap (\mathrm{B} \cup \Gamma)$

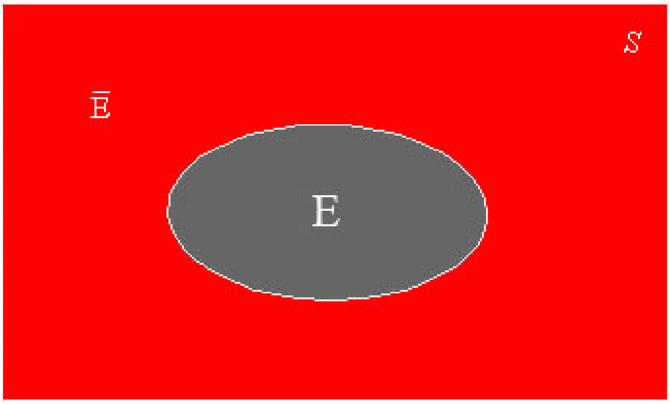
$\mathrm{B} \cap (\mathrm{A} \cup \Gamma) = (\mathrm{B} \cap \mathrm{A}) \cup (\mathrm{B} \cap \Gamma)$

Αν $A \cap B = A \cap \Gamma$ τότε το B δεν είναι κατ' ανάγκη ίσο με το Γ.

3. Συμπλήρωμα ενδεχομένου:

E^c ή $\overline{E}$

Το ενδεχόμενο που πραγματοποιείται όταν το ενδεχόμενο Ε δεν πραγματοποιείται.



Ιδιότητες:

$A \cup A^{c} = S$

$A \cap A^{c} = \varnothing$

$(A^{c})^{c} = A$

Κανόνας του de Morgan:

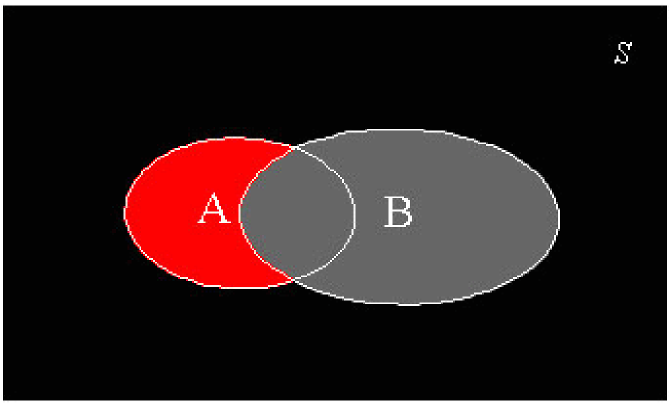
$(A \cup B)^{c} = A^{c} \cap B^{c}$

και γενικότερα:

Το συμπλήρωμα ενώσεων και τομών είναι ίσο με τις τομές και ενώσεις των αντιστοίχων συμπληρωμάτων.

4. **Διαφορά Ενδεχομένων:**

$A - B$



Ιδιότητες:

$A - B = A \cap B^C$

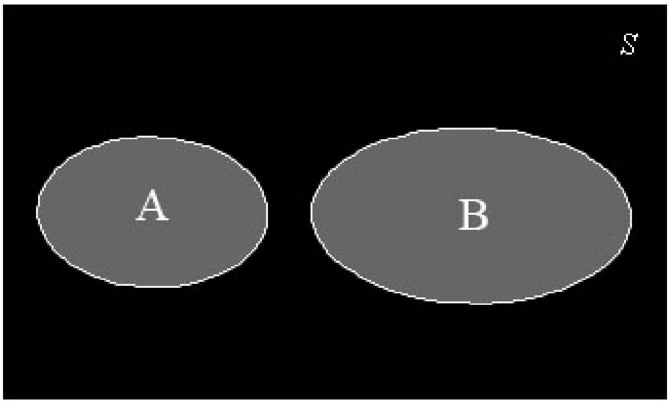
Ισχύει:

$A - (A - B) = A \cap B$

Απόδειξη:

$$A - (A - B) =$$
$$= A - (A \cap B^C) = A \cap (A \cap B^C)^C = A \cap (A^C \cup B) =$$
$$= (A \cap A^C) \cup (A \cap B) = \varnothing \cup (A \cap B) = A \cap B$$

Ασυμβίβαστα (ή Ξένα μεταξύ τους) Ενδεχόμενα:



Συλλεκτικά Εξαντλημένα Ενδεχόμενα:

Δυο ή περισσότερα ενδεχόμενα είναι **συλλεκτικά εξαντλημένα** αν η ένωση όλων αυτών ταυτίζεται με το δειγματοχώρο πάνω στον οποίο έχουν οριστεί.

Ανεξάρτητα ενδεχόμενα.

Για δύο γεγονότα Α και Β, με $P(A) > 0$, ορίσαμε τη δεσμευμένη πιθανότητα του Β δεδομένου του Α, $P(B/A)$. Συγκρίνοντας τώρα τις πιθανότητες $P(B/A)$ και $P(B)$, είναι δυνατόν να ισχύει μία από τις τρεις σχέσεις:

$P(B/A) > P(B)$, $P(B/A) = P(B)$, $P(B/A) < P(B)$

Το ποια από αυτές θα ισχύει, καθορίζεται από τις επιλογές των A και B.

Στην περίπτωση που είναι $P(B/A) = P(B)$, λέμε ότι το γεγονός B είναι **ανεξάρτητο** (**στοχαστικά** ή **στατιστικά** ή **υπό την έννοια της πιθανότητας**) από το γεγονός A. Δηλαδή, η γνώση του ότι το γεγονός A πραγματοποιήθηκε δεν δίνει καινούριες πληροφορίες για την επανεκτίμηση της πιθανότητας του γεγονότος B. Αν τώρα υποθέσουμε ότι και $P(B) > 0$, τότε το ότι το B είναι ανεξάρτητο του A συνεπάγεται και το ότι το A είναι ανεξάρτητο του B.

Πράγματι,

$$P(A/B) = \frac{P(B \cap A)}{P(B)} = \frac{P(B/A)P(A)}{P(B)} = \frac{P(B)P(A)}{P(B)} = P(A)$$

Λόγω της συμμετρίας αυτής, λέμε ότι τα γεγονότα A και B είναι ανεξάρτητα. Από την προηγούμενη σχέση παίρνουμε $P(A \cap B) = P(A)P(B)$ που έχει έννοια ακόμα κι αν $P(A) = 0$ ή $P(B) = 0$. Έτσι οδηγούμαστε στον ακόλουθο ορισμό της ανεξαρτησίας γεγονότων:

Ορισμός: Δύο γεγονότα A_1 και A_2 λέγονται (στοχαστικά ή στατιστικά ή υπό την έννοια της πιθανότητας) ανεξάρτητα, αν $P(A_1 \cap A_2) = P(A_1)P(A_2)$. Πιο γενικά, λέμε ότι $n \geq 3$ γεγονότα $A_1, A_2, ..., A_n$ είναι ανεξάρτητα, αν

$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1) * P(A_2) ... P(A_n)$$

και οποιοδήποτε υποσύνολό τους που περιέχει τουλάχιστον δύο αλλά λιγότερα από n ενδεχόμενα, αποτελείται από ανεξάρτητα ενδεχόμενα.

9.5. Αξιωματική Θεμελίωση των Πιθανοτήτων

Η αξιωματική θεμελίωση των Πιθανοτήτων προτάθηκε το 1933 από το μαθηματικό A.N. Kolmogorov και έγινε κοινά αποδεκτή.
Σύμφωνα με τη θεωρία του Kolmogorov η πιθανότητα είναι μια συνάρτηση που ορίζεται ως εξής :

Έστω S ένας δειγματικός χώρος και έστω B το σύνολο όλων των ενδεχομένων του S. Ορίζουμε ως συνάρτηση πιθανότητας μια συνάρτηση P:

P: B $\rightarrow$ R

η οποία σε κάθε ενδεχόμενο A αντιστοιχεί έναν πραγματικό αριθμό P(A) έτσι ώστε να ικανοποιούνται τα παρακάτω αξιώματα:

i. $P(A) \geq 0$
ii. $P(S) = 1$
iii. $P\left[\bigcup_{i=1}^{n} A_i\right] = P(A_1) + P(A_2) + ... + P(A_n) = \sum_{i=1}^{n} P(A_i), A_i \cap A_j = \emptyset, i \neq j$

9.6 Βασικά Θεωρήματα των Πιθανοτήτων

Έστω ο δειγματικός χώρος S, B το σύνολο όλων των ενδεχομένων του S και P η συνάρτηση πιθανότητας που ικανοποιεί τα τρία αξιώματα του Kolmogorov. Αποδεικνύεται ότι ισχύουν τα παρακάτω θεωρήματα:

Θεώρημα 1: Για κάθε ενδεχόμενο A ισχύει $P(A') = 1 - P(A)$.

Θεώρημα 2: Ισχύει ότι $P(\emptyset) = 0$.

Θεώρημα 3: Για κάθε ενδεχόμενο A ισχύει $P(A) \leq 1$.

Θεώρημα 4: Για οποιαδήποτε ενδεχόμενα A_1 και A_2 ισχύει ότι:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

Το θεώρημα 4 γενικεύεται για την περίπτωση n ενδεχομένων. Στην περίπτωση που n

= 3 γίνεται:

$$P\left(A_1 \cup A_2 \cup A_3\right) = P(A_1) + P(A_2) + P(A_3) - P\left(A_1 \cap A_2\right) - P\left(A_1 \cap A_3\right) - P\left(A_2 \cap A_3\right) + P\left(A_1 \cap A_2 \cap A_3\right)$$

Για δύο ενδεχόμενα τα οποία είναι ασυμβίβαστα, το θεώρημα οδηγείται στο συμπέρασμα:

$$P\left(A_1 \cup A_2\right) = P(A_1) + P(A_2), A_1 \cap A_2 = \varnothing$$

το οποίο είναι ειδική περίπτωση του $3^{ου}$ αξιώματος.

9.7 Δεσμευμένη πιθανότητα

Θεωρούμε δύο γεγονότα Α και Β με $P(A) > 0$, και θέτουμε το εξής ερώτημα: Ποια είναι η πιθανότητα να συμβεί το Β, δεδομένου ότι συνέβη (πραγματοποιήθηκε) το Α. Για να απαντήσουμε, πρέπει πρώτα να δώσουμε τον ακριβή ορισμό της πιθανότητας ενός ενδεχομένου με δεδομένο κάποιο άλλο.

Ορισμός

Έστω Α και Β δύο ενδεχόμενα τέτοια ώστε $P(A) > 0$. Τότε η δεσμευμένη πιθανότητα του Β με δεδομένο το Α, η οποία συμβολίζεται με $P(B/A)$, ορίζεται από τη σχέση

$$P(B/A) = \frac{P(B \cap A)}{P(A)}.$$

Αν $P(A) = 0$, τότε η $P(B/A)$ δεν ορίζεται. Η σχέση γραμμένη ισοδύναμα ως

$P(B \cap A) = P(B/A)P(A)$, μπορεί να μεταφραστεί ως η πιθανότητα να συμβεί και το Α και το Β είναι ίση με την πιθανότητα να συμβεί το Α επί την πιθανότητα να συμβεί το Β δεδομένου του ότι συνέβη το Α.

Η $P(B/A)$ λέγεται επίσης και πιθανότητα υπό συνθήκη του Β δεδομένου του Α. Το νόημα της εισαγωγής της δεσμευμένης πιθανότητας είναι η παροχή δυνατότητας ενσωμάτωσης τυχόν διαθέσιμων πληροφοριών κατά τον υπολογισμό της πιθανότητας ενός γεγονότος.

Στην πράξη, αν γνωρίζουμε ότι το Α έχει πραγματοποιηθεί, τότε αυτό αντικαθιστά το Ω στον υπολογισμό της πιθανότητας του Β, δηλαδή η δεσμευμένη πιθανότητα του Β $P(B/A)$ είναι στην ουσία η πιθανότητα του Β στον δειγματοχώρο Α. Αυτό μπορεί να χρησιμοποιηθεί για τον υπολογισμό του $P(B/A)$.

<u>Συνοψίζοντας:</u> Ο υπολογισμός της δεσμευμένης πιθανότητας $P(B/A)$ μπορεί να γίνει είτε χρησιμοποιώντας τον ορισμό είτε υπολογίζοντας την πιθανότητα του Β στον νέο δειγματοχώρο Α.

9.8 Ολική Πιθανότητα

9.8.1 Θεώρημα Bayes

Έστω το ενδεχόμενο Ε του οποίου η εμφάνιση εξαρτάται από την επενέργεια ενός οπουδήποτε των n ασυμβίβαστων ενδεχομένων (αιτιών και τρόπων, υποθέσεων) αυτών $E_1, E_2, \ldots, E_n$ (δηλαδή $E_i \cap E_j = \varnothing$ για κάθε $i \neq j$).

Είναι προφανές ότι

$E = E \cap (E_1 \cup E_2 \cup E_n)$ ή $E = (E \cap E_1) \cup (E \cap E_2) \cup \ldots \cup (E \cap E_n)$.

Επειδή $E_i \cap E_j = \varnothing$ για κάθε $i \neq j$ θα είναι και τα ενδεχόμενα $E \cap E_i \,/\, i = 1, 2, \ldots, n$ ξένα μεταξύ τους. Κατά το αξίωμα $P(E_1 \cup E_2) = P(E_1) + P(E_2)$ θα είναι:

$$P(E) = P(E \cup E_1) + P(E \cap E_2) + \ldots + P(E \cap E_n)$$

Επειδή

$$P(E/E_i) = \frac{(E \cap E_i)}{P(E_i)}$$

$$\boxed{P(E \cap E_i) = P(E_i)P(E/E_i)}$$

η προηγούμενη σχέση γράφεται ως:

$$\boxed{P(E) = P(E_1)P(E/E_1) + P(E_2)P(E/E_2) + \ldots + P(E_n)P(E/E_n)}$$

Ο τύπος αυτός είναι γνωστός ως **τύπος της ολικής πιθανότητας**. Εκ του ορισμού της δεσμευμένης πιθανότητας είναι γνωστό ότι:

$$P(E_i / E) = \frac{P(E_i \cap E)}{P(E)}$$

ή (λόγω των των παραπάνω σχέσεων)

$$\boxed{P(E_i / E) = \frac{P(E_i)P(E / E_i)}{P(E_1)P(E / E_1) + ... + P(E_n)P(E / E_n)}}$$

Ο τύπος αυτός εκφράζει το **θεώρημα του Bayes**.

Η πιθανότητα $P(E_i)$ ονομάζεται a priori πιθανότητα (εκ των προτέρων πιθανότητα) και η δεσμευμένη πιθανότητα $P(E_i / E)$ ονομάζεται a posteriori πιθανότητα (εκ των υστέρων πιθανότητα).

Ο τύπος του Bayes παρέχει την πιθανότητα πραγματοποιήσεως της επενέργειας E_i με δεδομένο ότι πραγματοποιήθηκε το ενδεχόμενο Ε.

9.9 Συνδυαστική

Η συνδυαστική ανάλυση είναι δημιούργημα των Leibniz, Pascol, Fermat, Euler σε μια προσπάθειά τους να δώσουν λύση σε προβλήματα τυχερών παιχνιδιών.

Από τότε και μετά, η συνδυαστική χρησιμοποιείται εκτεταμένα στην επίλυση προβλημάτων του Λογισμού Πιθανοτήτων, καθώς και σε προβλήματα επιλογών και εκλογών.

Αντικείμενο της συνδυαστικής ανάλυσης είναι η περιγραφή τρόπων και μεθόδων που βοηθούν στον υπολογισμό του αριθμού των συμπλεγμάτων, τα οποία μπορούν να κατασκευαστούν από n αντικείμενα συνδυαζόμενα μεταξύ τους, σύμφωνα με ορισμένους κανόνες.

Κάθε ένα από τα συμπλέγματα αυτά διαφέρει από τα υπόλοιπα ή κατά τη θέση των αντικειμένων τους ή κατά τη φύση τους.

9.9.1 Αρχές Απαρίθμησης

Σύμφωνα με τον κλασικό ορισμό η πιθανότητα εμφάνισης ενός ενδεχομένου A εκφράζεται ως ο λόγος του αριθμού των ευνοϊκών για το ενδεχόμενο αποτελεσμάτων προς τον αριθμό του συνόλου των δυνατών αποτελεσμάτων του:

$$P(A) = \frac{n(A)}{N}$$

Σε απλά προβλήματα τόσο ο αριθμός των ευνοϊκών όσο και ο αριθμός των δυνατών αποτελεσμάτων μπορούν εύκολα να μετρηθούν. Σε πιο σύνθετα, όμως, προβλήματα η απαρίθμηση των ευνοϊκών και των δυνατών αποτελεσμάτων είναι εξαιρετικά δύσκολη αν όχι αδύνατη. Για το λόγο αυτό χρησιμοποιούμε τις Αρχές Απαρίθμησης.

- **Μεταθέσεις και Διατάξεις.**

Έστω E={α_1,α_2,…,α_n} ένα σύνολο n αντικειμένων. **Απλή μετάθεση** n δοθέντων αντικειμένων ονομάζεται κάθε σύμπλεγμα που προκύπτει από κάθε τοποθέτηση επί ανοικτής γραμμής των n αυτών αντικειμένων. Είναι δηλαδή η απλή μετάθεση των n αντικειμένων μια αμφιμονοσήμαντη απεικόνιση του συνόλου E των n αντικειμένων του συνόλου E των n αντικειμένων επί του εαυτού του). Αν τα n αντικείμενα τοποθετηθούν επί ανοικτής γραμμής καθ' όλους τους δυνατούς τρόπους, τότε προκύπτουν όλες οι μεταθέσεις των n αντικειμένων, το πλήθος των οποίων συμβολίζεται με το σύμβολο P_n.

Με πόσους τρόπους μπορούμε να τοποθετήσουμε n διαφορετικά αντικείμενα σε n διαφορετικές αριθμημένες θέσεις με την προϋπόθεση ότι μία θέση μπορεί να χωρέσει μόνο ένα αντικείμενο;

Για το πρώτο αντικείμενο έχουμε n επιλογές θέσεων, για το δεύτερο αντικείμενο υπάρχουν n-1 επιλογές ελεύθερων θέσεων, … , και για το τελευταίο αντικείμενο υπάρχει 1 τελευταία κενή θέση. Συνεπώς υπάρχουν n(n-1)(n-2)…1 διαφορετικές **ΜΕΤΑΘΕΣΕΙΣ**. Το γινόμενο το ονομάζουμε n-παραγοντικό και το συμβολίζουμε με το σύμβολο n!

Άρα το πλήθος των μεταθέσεων n αντικειμένων είναι:

$n! = 1*2*3*…*n$

Με πόσους τρόπους μπορούμε να τοποθετήσουμε k διαφορετικά αντικείμενα σε n διαφορετικά αριθμημένες θέσεις με την προϋπόθεση ότι μία θέση μπορεί να χωρέσει μόνο ένα αντικείμενο;

Για το πρώτο αντικείμενο έχουμε n επιλογές θέσεων, για το δεύτερο αντικείμενο υπάρχουν n-1 επιλογές ελεύθερων θέσεων, … , και για το k αντικείμενο υπάρχουν (n-k+1) ελεύθερες θέσεις. Συνεπώς υπάρχουν n(n-1)(n-2)…(n-k + 1) διαφορετικές **ΔΙΑΤΑΞΕΙΣ** των k από n αντικείμενα. Για το γινόμενο αυτό χρησιμοποιούμε το σύμβολο P(n, k).

Γενικότερα τα προβλήματα τέτοιας μορφής μπορούμε να τα θεωρήσουμε ως παραλλαγές του παρακάτω προβλήματος:

Έστω ότι θέλουμε να τοποθετήσουμε k διαφορετικά βαμμένες μπάλες σε n διαφορετικά αριθμημένα κουτιά με την προϋπόθεση ότι ένα κουτί μπορεί να χωρέσει μόνο μία μπάλα. Η απάντηση σε αυτό το πρόβλημα δίνεται από το τύπο:

$P(n, k) = n(n-1)(n-2)…(n-k+1)$

Σημειώνουμε ότι για κάθε αριθμό $b \neq 0$ ισχύει $a= axb/b$. Άρα ισχύουν τα παρακάτω:

$P(n, k)=n(n-1)(n-2)…(n-k+1) = n(n-1)(n-2)…(n-k+1)(n-k)(n-k-1) …2*1/(n-k)(n-k-1) …2*1 = n!/(n-k)!$

Άρα το πλήθος των διατάξεων των k από n αντικείμενα είναι:

$P(n, k)= n!/(n-k)!$

ν Διατάξεις

- *Διατάξεις χωρίς επανάληψη ν διαφορετικών πραγμάτων* $α_1,α_2,α_3,…,α_ν$ ανά μ (όπου $μ \leq ν$), ονομάζουμε όλες τις ομάδες που σχηματίζονται με τα ν αυτά δεδομένα και που η κάθε ομάδα περιέχει μ στοιχεία και διαφέρει από τις άλλες, είτε κατά τη φύση των πραγμάτων της, είτε κατά τη θέση τους και το ίδιο αντικείμενο εμφανίζεται σε μια ομάδα το πολύ μια φορά.

Το πλήθος των διατάξεων των ν διαφορετικών αντικειμένων ανά μ το συμβολίζουμε με το σύμβολο A_m^n και το υπολογίζουμε με τον τύπο:

$$A_m^n = \frac{n!}{(n-m)!}$$

- *Επαναληπτικές διατάξεις των ν διαφορετικών πραγμάτων* $α_1,α_2,α_3,\dots,α_ν$ ανά μ, ονομάζουμε το πλήθος των ομάδων που μπορούμε να σχηματίσουμε με τα ν αυτά πράγματα, έτσι που κάθε ομάδα να διαφέρει από τις άλλες ή κατά τη φύση ή κατά τη θέση των πραγμάτων της και κάθε αντικείμενο να μπορεί να επαναληφθεί σε κάθε ομάδας περισσότερες από μία φορά.

Οι διατάξεις με επανάληψη συμβολίζονται με A_m^n και δίνονται από τον τύπο:

$$A_m^n = n_m$$

v Μεταθέσεις

- *Μεταθέσεις ν διαφορετικών αντικειμένων* $α_1,α_2,α_3,\dots,α_ν$ ονομάζουμε τις διάφορες ομάδες που μπορούμε να σχηματίσουμε τοποθετώντας τα ν αυτά αντικείμενα το ένα μετά το άλλο σε μια σειρά επί μιας γραμμής (ανοικτής) με όλους τους δυνατούς τρόπους.

Συνεπώς, δύο μεταθέσεις ν διαφορετικών αντικειμένων θα θεωρούνται διαφορετικές εφόσον διαφέρουν κατά τη θέση ενός τουλάχιστον πράγματος.

Τις μεταθέσεις ν διαφορετικών πραγμάτων τις συμβολίζουμε με $P_ν$ και υπολογίζουμε το πλήθος τους με τον παρακάτω τύπο:

$$P_n = 1*2*3*\dots*(n-1)*n$$

- *Μεταθέσεις με επαναλήψεις (ή συμπτωματικές) ν διαφορετικών πραγμάτων*, από τα οποία $α_1$ είναι τα ίδια (εντελώς όμοια μεταξύ τους)και διάφορα των άλλων, επίσης $α_2$ το ίδιο έως ότου ισχύει για ως και $α_ρ$:

$α_1 + α_2 + \dots + α_ρ = ν$

ονομάζουμε τις διάφορες ομάδες όπου μπορούμε να σχηματίσουμε τοποθετώντας τα ν αυτά αντικείμενα το ένα ύστερα από το άλλο σε μια σειρά επί μιας γραμμής (ανοικτής) με όλους τους δυνατούς τρόπους. Το πλήθος των μεταθέσεων των ν πραγμάτων με επανάληψη δίνεται από τον τύπο:

$$(a_1, a_2, ..., a_n)\mathrm{P}_n = \frac{n!}{a_1! a_2! * ... * a_r}$$

ν Συνδυασμοί

- Συνδυασμοί των ν πραγμάτων ανά μ χωρίς επανάληψη ονομάζονται όλες οι ομάδες που μπορούμε να σχηματίσουμε με τα ν αυτά στοιχεία, έτσι ώστε κάθε ομάδα να περιέχει μ στοιχεία και να διαφέρει από τις άλλες τουλάχιστον κατά τη φύση ενός στοιχείου της.

Υποθέτουμε ότι έχουμε τρία αντικείμενα, τα $\alpha_1, \alpha_2, \alpha_3$. Οι συνδυασμοί αυτών των τριών αντικειμένων ανά δύο θα είναι:

$\alpha_1\alpha_2, \alpha_1\alpha_3, \alpha_2\alpha_3$

Παρατηρούμε ότι κάθε μία από τις τρεις αυτές ομάδες διαφέρει από τις άλλες κατά ένα στοιχείο. Το πλήθος των συνδυασμών των ν πραγμάτων ανά μ το συμβολίζουμε μ $\binom{n}{m}$ ή C_m^n και το υπολογίζουμε με τον τύπο:

$$C_m^n = \binom{n}{m} = \frac{\mathrm{A}_m^n}{\mathrm{P}_m^n} = \frac{n*(n-1)*(n-2)*(n-3)*...*[n-(m-1)]}{1*2*3*4*...*m}$$

Αν πολλαπλασιάσουμε αριθμητή και παρονομαστή της σχέσης με την παράσταση $(n-m)(n-m-1)*...*3*2*1$, θα έχουμε:

$$C_m^n = \binom{n}{m} = \frac{n*(n-1)*(n-2)*(n-3)*...*(n-m+1)*(n-m)*(n-m-1)*...*3*2*1}{1*2*3...(n-2)(n-1)*n = \mathrm{P}_n = n!}$$

Άρα, ο τύπος που δίνει το πλήθος των συνδυασμών γράφεται με την εξής μορφή, (η οποία χρησιμοποιείται και στην πράξη):

$$C_m^n = \binom{n}{m} = \frac{n!}{m!(n-m)!}$$

- *Επαναληπτικοί συνδυασμοί ν διαφορετικών αντικειμένων ανά μ* ονομάζονται οι διάφορες ομάδες τις οποίες μπορούμε να σχηματίσουμε αν πάρουμε μ στοιχεία από τα ν και κάθε στοιχείο μπορεί να επαναλαμβάνεται στην κάθε ομάδα περισσότερο από μ φορές και κατά τρόπο, ώστε οι δύο ομάδες να θεωρούνται διαφορετικές, αν διαφέρουν στη φύση ενός τουλάχιστον στοιχείου.

Τα πλήθος των συνδυασμών με επανάληψη το συμβολίζουμε με $C_{n,m}^{(e)}$ και υπολογίζεται με τη βοήθεια του τύπου:

$$C_{n,m}^{(e)} = \frac{n\,(n+1)*(n+2)*...*9n+m-1)}{1*2*3*...*m} = C_m^{n+m-1}$$

9.10. Τυχαίες μεταβλητές

Η έννοια της τυχαίας μεταβλητής

Αν η φύση των στοιχείων των τυχαίων ενδεχομένων είναι ποσοτική, τότε τα στοιχειώδη ενδεχόμενα του δειγματικού χώρου ενός πειράματος τύχης εκφράζονται με πραγματικούς αριθμούς. Π.χ. κατά τη ρίψη ενός ζαριού τα στοιχειώδη ενδεχόμενα θα είναι: 0, 1, 2, 3, 4, 5, 6 ˙ επίσης, κατά την περιστροφή της ρουλέτας τα στοιχειώδη ενδεχόμενα θα είναι: 0, 1,2,3, ..., 36.

Αν όμως η φύση των στοιχείων των ενδεχομένων είναι ποιοτική, τότε τα στοιχειώδη ενδεχόμενα του δειγματικού χώρου εκφράζονται με τις αντίστοιχες τιμές των ποιοτικών χαρακτηριστικών. Π.χ. κατά τη ρίψη ενός νομίσματος τα στοιχειώδη ενδεχόμενα «πρόσωπο - γράμματα» δεν εκφράζονται με πραγματικούς αριθμούς. Επίσης, τα ενδεχόμενα «οικογενειακή κατάσταση», «το κέρδος του παίκτη», «σφαιρίδιο άσπρο» κ.λπ. δεν εκφράζονται με αριθμούς. Είναι όμως χρήσιμο να παρασταθούν συμβατικά τα στοιχεία του δειγματικού χώρου με πραγματικούς αριθμούς. Επίσης, και στην περίπτωση κατά την οποία τα στοιχειώδη ενδεχόμενα του δειγματικού χώρου εκφράζονται με αριθμούς, μπορούμε σε καθέναν από τους αριθμούς αυτούς να αντιστοιχίσουμε ένα νέο αριθμό.

Εξετάζουμε ένα δειγματικό χώρο (Ω) που αναφέρεται σε ένα πείραμα και αντιστοιχίζουμε σε κάθε αποτέλεσμα του δειγματικού χώρου έναν πραγματικό αριθμό *x(ωi),* Ορίζουμε έτσι μια συνάρτηση *x(ωi),* η οποία έχει πεδίο ορισμού *το* δειγματικό χώρο Ω και πεδίο τιμών ένα υποσύνολο *(Ex)* του συνόλου των πραγματικών αριθμών. Η συνάρτηση αυτή ονομάζεται *τυχαία μεταβλητή.* Σε κάθε στοιχειώδες ενδεχόμενο *ωi* του δειγματικού χώρου θα πρέπει να αντιστοιχεί μία μόνο τιμή της τυχαίας μεταβλητής.

Η τυχαία μεταβλητή συμβολίζεται συνήθως με τα κεφαλαία γράμματα X, *Y, Z,...* και οι δυνατές τιμές τους με τα αντίστοιχα μικρά γράμματα: x_1, x_2, x_3, . . . , xn ή y_1, y_2, y_3, . . . , y_n,. . . Ο συμβολισμός X = $x(\omega_i)$, μας δείχνει τον αριθμό που αντιστοιχεί στο αποτέλεσμα ωi του δειγματικού χώρο.

Για να κατανοήσουμε καλύτερα την έννοια της τυχαίας μεταβλητής και την αντιστοιχία σε κάθε στοιχειώδες ενδεχόμενο ενός μόνο πραγματικού αριθμού, δίνουμε μερικά συγκεκριμένα παραδείγματα πειραμάτων τύχης.

Παραδείγματα πειραμάτων τύχης

1. Μια εξαγωγή σφαιριδίου από μια κάλπη μπορεί να αναφέρεται στα πιο κάτω ενδεχόμενα:

ω_1: σφαιρίδιο άσπρο ω_2: σφαιρίδιο κόκκινο
ω_3: σφαιρίδιο πράσινο ω_4: σφαιρίδιο μαύρο

Στα παραπάνω ενδεχόμενα μπορούμε να αντιστοιχίσουμε τη συνολική ζημιά ή το συνολικό κέρδος που αναφέρεται σε καθένα από αυτά. Π.χ. υποθέτουμε μια ζημιά 200 δρχ. στην περίπτωση εξαγωγής άσπρου σφαιριδίου, μια ζημιά 100 δρχ. στην περίπτωση εξαγωγής κόκκινου σφαιριδίου, ένα κέρδος100 δρχ. στην περίπτωση εξαγωγής πράσινου σφαιριδίου, ένα κέρδος 200 δρχ. στην περίπτωση εξαγωγής μαύρου σφαιριδίου. Τότε θα έχουμε:

$x(\omega_1)$ = -200, *$x(\omega_2)$* = -100, *$x(\omega_3)$* = 100, *$x(\omega_4)$* = 200

2. Κατά τη ρίψη ζαριού τα δυνατά στοιχειώδη ενδεχόμενα είναι:

ω1 = {l}, ω2 = {2}, ω3 = {3}, ω4 = {4}, ω5 = {5}, ω6 = {6}

Καταβάλλοντας ένα ποσό 3.500 δρχ. έχουμε δικαίωμα να εισπράξουμε, ανάλογα με τον αριθμό που θα εμφανιστεί, το παρακάτω ποσό:

x(ωι) = 1.000, x(ω2) = 2.000, x(ω3) = 3.000, x(ω4) = 4.000, x(ω5) =5.000, x(ω6) = 6.000, δηλαδή x(ωi) = 1.000i

3. Υποθέτουμε ότι ρίχνουμε 3 κανονικά νομίσματα. Τα στοιχειώδη ενδεχόμενα του δειγματικού χώρου θα είναι:

ω1 = (ΓΓΓ) ω2 = (ΓΓΠ) ω3 = (ΓΠΓ) ω4 = (ΠΓΓ) ω5 = (ΠΠΓ) ω6 = (ΠΓΠ) ω7 = (ΓΠΠ) ωB = (ΠΠΠ)

Το πρόβλημα που μπορεί να τεθεί είναι: Πόσα γράμματα στις 3 ρίψεις μπορούν να εμφανιστούν; Αν εκφράσουμε τον αριθμό της εμφάνισης της όψης «πρόσωπο» με την τυχαία μεταβλητή X, τότε οι τιμές της τυχαίας μεταβλητής θα είναι: x = 0, l, 2, 3, και επειδή σε κάθε στοιχειώδες ενδεχόμενο του δειγματικού χώρου Ω αντιστοιχεί ένας μόνο πραγματικός αριθμός, παρατηρούμε ότι στα ίδια στοιχειώδη ενδεχόμενα *ω2, ω3, ω4* αντιστοιχεί η ίδια τιμή x=1, δηλαδή:
x(ω2) = x(ω3) = x(ω4) =1.

Επίσης, στα στοιχειώδη ενδεχόμενα ω5,ωβ,ω7 αντιστοιχεί η ίδια τιμή x=2 της τυχαίας μεταβλητής X, δηλαδή:
x(ω5)= x(ω6) = x(ω7) = 2 x(ω1) = 0 x(ω8) = 3

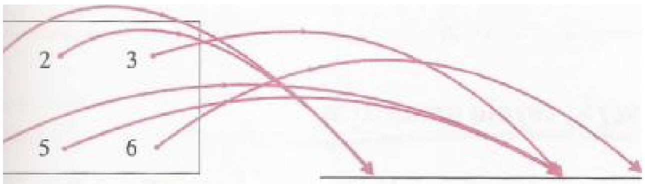
4. Ένα παιχνίδι που αναφέρεται στη ρίψη ενός ζαριού έχει ως εξής:

Ο Α χάνει 50 δρχ., αν εμφανιστεί ο αριθμός 1 ή 2, κερδίζει 20 δρχ., αν το ζάρι εμφανίσει 3 ή 4 ή 5, και κερδίζει 40 δρχ., αν εμφανιστεί ο αριθμός 6. Οι τιμές της τυχαίας μεταβλητής Χ θα είναι:

xι = -50, αν ω1= 1, ω2 = 2

x2 = 20, αν ω3 = 3, ω4 = 4, ω5 = 5

x3 = 40, αν ω6 = 6



-50 20 40

Από τα παραπάνω παραδείγματα φαίνεται καθαρά ότι η τυχαία μεταβλητή είναι μια συνάρτηση με πεδίο ορισμού το δειγματικό χώρο Ω και πεδίο τιμών ένα υποσύνολο του συνόλου των πραγματικών αριθμών.

Παράδειγμα

Ποια είναι η συνάρτηση πιθανότητας και η κατανομή πιθανοτήτων της τυχαίας μεταβλητής «αριθμός προσώπων» κατά τη ρίψη 2 νομισμάτων;

<u>Λύση</u>

Ο δειγματικός χώρος θα είναι:

Ω = {ΠΠ,ΠΓ,ΓΠ,ΓΓ}

Οι τιμές της τυχαίας μεταβλητής «πρόσωπο» θα είναι x : 0,1,2.

Οι τιμές της συνάρτησης πιθανότητας θα είναι:

Για x = 0 =} P{X = 0} = 1/4

Για x = 1 =} P {X = 1} = 1/4 + 1/4 = 2/4

Για x = 2 =} P{X = 2} = 1/4

xi	Pi
0 1 2	¼ 2/4 ¼
Σύνολο	1

Ονομάζουμε ασυνεχή τυχαία μεταβλητή μια τυχαία μεταβλητή X η οποία μπορεί να λάβει πεπερασμένο ή άπειρο πλήθος τιμών, αλλά αριθμήσιμο.

Στη ρίψη ενός ζαριού η τυχαία μεταβλητή X είναι ασυνεχής και παίρνει πεπερασμένο πλήθος τιμών, δηλαδή:

X_i: 1,2,3,4,5,6

Επίσης, στην περίπτωση κατά την οποία ρίχνουμε 2 νομίσματα, η τυχαία μεταβλητή «αριθμός γραμμάτων» θα λάβει τις τιμές:

x: 0,1,2

Η τυχαία μεταβλητή που αναφέρεται στον αριθμό «πόσες φορές πρέπει να ριχτεί ένα ζάρι μέχρι να εμφανιστεί για πρώτη φορά ο αριθμός 5», είναι μια ασυνεχής μεταβλητή τιμών που μπορεί να λάβει άπειρο πλήθος τιμών, αλλά αριθμήσιμο, δηλαδή:

x: 1,2,3,..., n

9.10.1 Συνάρτηση πιθανότητας

Υποθέτουμε ότι *x* είναι μια τυχαία ασυνεχής μεταβλητή, που μπορεί να λάβει τιμές x_1, x_2, x_3,..., x_N.

Η πιθανότητα με την οποία η τυχαία μεταβλητή *x* μπορεί να λάβει την τιμή x_i ονομάζεται *συνάρτηση πιθανότητας* και συμβολίζεται $P\{ X = x_i\}$,

Γενικά, μια συνάρτηση με πεδίο ορισμού το σύνολο των τιμών της τυχαίας μεταβλητής *x* και πεδίο τιμών το διάστημα [0, 1] του συνόλου των πραγματικών αριθμών ονομάζεται συνάρτηση πιθανότητας. Η συνάρτηση πιθανότητας θα πρέπει να ικανοποιεί τις παρακάτω σχέσεις:

(1) $P\{X = x_i\} \geq 0$ για όλα τα $x \in Ex$

(2) $\Sigma^{P\{X=x^i\}} = 1$

Πίνακας

Τιμές μεταβλητής (x_i)	*Πιθανότητες (P_i)*
x_1	P_1
x_2	P_2
x_3	P_3
x_4	P_4
.	.
.	.
.	.
x_i	P_i
.	.
.	.

. x_n	. P_n
Σύνολο	$\Sigma P_i=1$

Το σύνολο των ζευγών (x_i, $P\{X = x_i\}$) ή συντομότερα (x_i,P_i) ονομάζεται κατανομή πιθανοτήτων της τυχαίας μεταβλητής X και δίνεται από τον παραπάνω πίνακα.

9.10.2 Συνάρτηση αθροιστικής κατανομής ασυνεχούς τυχαίας μεταβλητής

Ονομάζεται συνάρτηση κατανομής μιας ασυνεχούς τυχαίας μεταβλητής X η πιθανότητα να πάρει η τυχαία μεταβλητή X τιμή μικρότερη ή ίση από την τιμή X. Η συνάρτηση κατανομής συμβολίζεται με F(x) και δίνεται από τη σχέση:

$F(x) = P\{X \leq x\}$

Παράδειγμα

Αναφερόμενοι στο παράδειγμα 1 της προηγούμενης ενότητας, όπου:

$P\{\omega_1\} = P\{\omega_2\} = P\{\omega_3\} = P\{\omega_4\} = 1/4$

και

$\omega_1 = -200,\ \omega_2 = -100,\ \omega_3 = 100,\ \omega_4 = 200$

η συνάρτηση κατανομής της τυχαίας μεταβλητής X θα είναι:

$$F(x)=\begin{cases} 0gia\ -\infty\ \mathbf{p}\ x\ \mathbf{p}\ -200 \\ \frac{1}{4}gia\ -200\ \mathbf{p}\ x\ \mathbf{p}\ -100 \\ \frac{2}{4}gia\ -100\ \mathbf{p}\ x\ \mathbf{p}\ 100 \\ \frac{3}{4}gia\ 100\ \mathbf{p}\ x\ \mathbf{p}\ 200 \\ 1gia\ 200\ \mathbf{p}\ x\ \mathbf{p}\ \infty \end{cases}$$

9.11 Διακύμανση

Η διακύμανση μιας τυχαίας μεταβλητής X είναι η μέση τιμή της τυχαίας μεταβλητής $[X - E(X)]^2$, δηλαδή η μέση τιμή που προέρχεται από το τετράγωνο της διαφοράς μεταξύ της τυχαίας μεταβλητής X και της μαθηματικής ελπίδας αυτής, E(X).

Αν η τυχαία μεταβλητή X είναι ασυνεχής, η διακύμανση δίνεται από τον τύπο:

$Var(X) = \sigma_x^2 = E[X - E(X)]^2 = E(X - \mu)^2 = \Sigma(x_i - \mu)^2 P_i$

ενώ αν η τυχαία μεταβλητή X είναι συνεχής, τότε η διακύμανση δίνεται από τον πιο κάτω τύπο:

$Var(X) = \sigma_x^2 = E[X - E\, F(M) = \int_{-\infty}^{\infty} Mf(x)dx = 1/2\ (X)]^2 = \int^{\infty}_{-\infty} (x - \mu)^2 = f(x)dX$

όπου μ = E(X).

9.11.1 Ιδιότητες της διακύμανσης

1. Η διακύμανση μιας σταθερής ποσότητας α είναι 0.

2. Αν όλες οι τιμές μιας τυχαίας μεταβλητής *X* πολλαπλασιαστούν με τον ίδιο αριθμό α, τότε η διακύμανση πολλαπλασιάζεται με το τετράγωνο του αριθμού αυτού, δηλαδή:

$Var(\alpha X) = a^2 Var(X)$

3. Αν X είναι μια τυχαία μεταβλητή και α ένας ορισμένος πραγματικός αριθμός, τότε θα έχουμε:

$Var(X \pm \alpha) = Var(X)$

4. Αν η μεταβλητή *Y* είναι γραμμικός μετασχηματισμός της μεταβλητής *X,* δηλαδή αν $Y = \alpha + \beta X$, όπου α και *β* είναι 2 πραγματικοί αριθμοί, θα είναι:

$Var(Y) = \beta^2 Var(X)$

9.12 Διάμεσος - Τεταρτημόρια

α) Ασυνεχής τυχαία μεταβλητή

Έστω η θεωρητική κατανομή πιθανοτήτων του πίνακα .

Πίνακας

x_i	P_i	$F(x)$
X_1 X_2 . . . X_κ	P_1 P_2 . . . P_κ	F_1 F_2 . . . F_κ
Σύνολο	1	

Βρίσκουμε πρώτα την τιμή N/2. Στη συνέχεια βρίσκουμε το ζεύγος των τιμών της συναρτήσεως κατανομής που ικανοποιούν τη σχέση:

$F_{i-1} < N/2 < F_i$

Η τιμή της τυχαίας μεταβλητής που αντιστοιχεί στο F_i είναι η διάμεσος. Επίσης, για το πρώτο τεταρτημόριο:

$F_{i-1} < N/4 < F_i$

και το τρίτο τεταρτημόρια είναι:

$F_{i-1} < 3N/4 < F_i$

β) Συνεχής τυχαία μεταβλητή

Για να βρούμε τη διάμεσο στην περίπτωση που η τυχαία μεταβλητή είναι συνεχής, βρίσκουμε τη συνάρτηση κατανομής και την εξισώνουμε με το ½. Η τιμή που θα προκύψει από αυτήν την εξίσωση είναι η διάμεσος:

$$F(M) = \int_{-\infty}^{\infty} Mf(x)dx = 1/2$$

Για το πρώτο και τρίτο τεταρτημόρια είναι αντίστοιχα:

F(x) = 1/4 και F(x) = 3/4

10. ΘΕΩΡΗΤΙΚΕΣ ΚΑΤΑΝΟΜΕΣ

10.1 Γενικά

Όπως ο ερευνητής στις πειραματικές επιστήμες προσπαθεί να ανακαλύψει τους φυσικούς νόμους οι οποίοι θα περιγράφουν τόσο την τωρινή όσο και την μελλοντική εξέλιξη των διαφόρων φυσικών φαινομένων, έτσι και ο ερευνητής των οικονομικών, δημογραφικών, κοινωνικών και λοιπών φαινομένων προσπαθεί να ανακαλύψει τους νόμους πιθανοτήτων που ακολουθούν τα διάφορα φαινόμενα τύχης και να προσαρμόσει τους νόμους αυτούς στις εμπειρικές κατανομές.

Τους νόμους πιθανοτήτων τους ονομάζουμε συνήθως *θεωρητικά μαθηματικά υποδείγματα συναρτήσεων πιθανότητας και πυκνότητας πιθανότητας.* Με τα *θεωρητικά* αυτά *υποδείγματα* μπορούμε να περιγράψουμε και να μελετήσουμε τα τυχαία φαινόμενα τα οποία εμφανίζονται περισσότερο στην πράξη.

Οι βασικές θεωρητικές κατανομές που εξετάζονται συνηθέστερα είναι:

- Κατανομή Bernoulli.
- Διωνυμική Κατανομή.
- Γεωμετρική Κατανομή
- Αρνητική Διωνυμική Κατανομή
- Υπεργεωμετρική Κατανομή
- Κατανομή Poisson
- Διακριτική Ομοιόμορφη Κατανομή
- Συνεχής Ομοιόμορφη Κατανομή
- Εκθετική Κατανομή
- Κανονική Κατανομή
- Γάμμα Κατανομή
- x_v^2 Κατανομή
- t_v Κατανομή ή Κατανομή Student

- $F(v_1, v_2)$ Κατανομή ή Κατανομή Fisher

Τα πλεονεκτήματα της προσαρμογής σε μια σειρά παρατηρήσεων (εμπειρική κατανομή) είναι:

(α) Το πλήθος των n παρατηρήσεων αντικαθίσταται μόνο από λίγες παραμέτρους και γίνεται έτσι ευκολότερη η σύγκριση των διαφόρων παρατηρήσεων.

(β) Η θεωρητική κατανομή έχει γνωστές ιδιότητες.

Τις σπουδαιότερες κατανομές που θα μελετήσουμε τις διακρίνουμε σε ασυνεχείς και συνεχείς.

10.2 Ασυνεχείς κατανομές

Οι βασικότερες ασυνεχείς κατανομές (μονοδιάστατες διακριτές κατανομές) τις οποίες και θα εξετάσουμε είναι α) Κατανομή Bernoulli, β) Διωνυμική Κατανομή και γ) Κατανομή Poisson.

Ας θεωρήσουμε μία ακολουθία (σειρά) πειραμάτων στην οποία ισχύουν τα επόμενα:

- Σε κάθε επανάληψη του πειράματος μπορούν να εμφανισθούν μόνο δύο δυνατά αποτελέσματα τα οποία θα χαρακτηρίζουμε σαν επιτυχία (ε) ή σαν αποτυχία (α).
- Τα πειράματα είναι ανεξάρτητα μεταξύ τους, έτσι ώστε το αποτέλεσμα οποιουδήποτε πειράματος να μην επηρεάζει τα αποτελέσματα των υπολοίπων.
- Η πιθανότητα επιτυχίας ή (αποτυχίας) δε μεταβάλλεται από πείραμα σε πείραμα.

Μια ακολουθία πειραμάτων με τις παραπάνω ιδιότητες λέγεται ακολουθία ανεξάρτητων δοκιμών Bernoulli. Η κοινή πιθανότητα επιτυχίας (αποτυχίας) των δοκιμών συμβολίζεται συνήθως με p (1 – p).

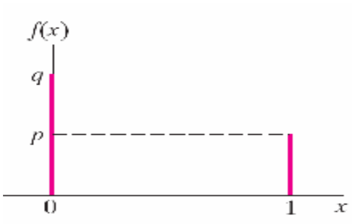
ΟΡΙΣΜΟΣ.

Έστω X ο αριθμός των επιτυχιών σε μια δοκιμή Bernoulli με πιθανότητα επιτυχίας p (και αποτυχίας q = 1 – p). Η κατανομή της δίτιμης (μηδέν – ένα) τυχαίας μεταβλητής X καλείται κατανομή Bernoulli με παράμετρο p.

Αν συμβολίσουμε με p, q τις πιθανότητες εμφάνισης των στοιχειωδών ενδεχομένων {ε}, {α} αντίστοιχα, είναι φανερό ότι θα έχουμε $p+q=1,\ 0\le p\le 1, 0\le q\le 1$.

<u>Συνάρτηση πιθανότητας.</u>

$f_X = P(X=x) = \begin{cases} p \\ q=1-p \end{cases}$ p για x = 1, q = 1 – p για x = 0



ή ισοδύναμα $f_X = p^2 s^{1-x}$, x = 0,1

Η μέση τιμή της κατανομής θα είναι: $m = \sum_{x=0}^{1} xf(x) = 0\cdot f(0) + 1\cdot f(1) = f(1) = p$

$$m_2^{'} = E(X^2) = \sum_{x=0}^{1} x^2 f(x) = 0^2\cdot f(0) + 1^2\cdot f(1) = f(1) = p$$

Ενώ, η διακύμανση της κατανομής θα είναι:

$$s^2 = V(X) = E(X^2) - [(X)]^2 = p - p^2 = p(1-p) = pq$$

10.2.1 Διωνυμική Κατανομή

Από τις ασυνεχείς κατανομές η πιο σπουδαία είναι η διωνυμική, γιατί τα περισσότερα φαινόμενα μπορούν να μελετηθούν με τη βοήθεια αυτής της κατανομής. Τη διωνυμική κατανομή ανακάλυψε ο Bernoulli το 1700.

Στη διωνυμική κατανομή ισχύει το εξής: όταν ένα πείραμα με δύο δυνατά αποτελέσματα (δοκιμή Bernoulli) εξαγόμενα, επαναλαμβάνεται συγκεκριμένο αριθμό φορών, αυτό το οποίο μας ενδιαφέρει είναι συνήθως, ο αριθμός των επιτυχιών ή αποτυχιών που εμφανίστηκαν.

Η τυχαία μεταβλητή Χ που δίνει τον αριθμό επιτυχιών σε ν δοκιμές Bernoulli με κοινή πιθανότητα επιτυχίας p λέγεται Διωνυμική τυχαία μεταβλητή.

ΟΡΙΣΜΟΣ.

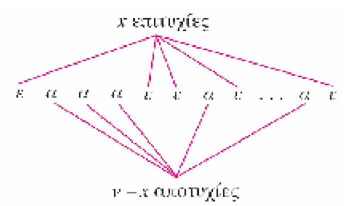
Έστω Χ ο αριθμός των επιτυχιών σε μια ακολουθία ν (ανεξάρτητων) δοκιμών Bernoulli με σταθερή πιθανότητα επιτυχίας p και αποτυχίας q = 1 – p σε όλες τις δοκιμές. Η κατανομή της τυχαίας μεταβλητής Χ καλείται διωνυμική κατανομή με παραμέτρους ν και p και συμβολίζεται με b (ν, p).

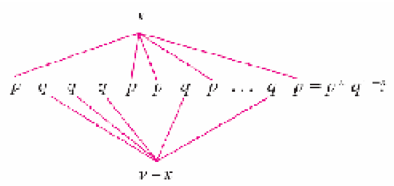
Η συνάρτηση πιθανότητας της διωνυμικής κατανομής με παραμέτρους ν και p δίδεται από τον τύπο:

$$f(x) = P(X = x) = \binom{\nu}{x} p^x q^{\nu - x}, \qquad x = 0, 1, 2, \ldots, \nu$$

Απόδειξη

Έστω x ένας συγκεκριμένος αριθμός του συνόλου των τιμών R_X, δηλαδή $0 \le x \le \nu$. Το ενδεχόμενο {X = x} θα αποτελείται από εκείνα τα στοιχεία του Ω, τα οποία περιέχουν x αποτελέσματα ίσα με ε και ν – x αποτελέσματα ίσα με α, δηλαδή είναι της μορφής:





f (*x*) = *P*(*X* = *x*)

$$f(x) = P(X = x) = \binom{\nu}{x} p^x q^{\nu - x}$$

<u>Σχέση της διωνυμικής κατανομής με την κατανομή Bernoulli.</u>

$$f(x) = P(X = x) = \binom{v}{x} p^x q^{v-x} \rightarrow v = 1 \rightarrow f(x) = P(X = x) = \binom{1}{x} p^x q^{1-x} = p^x q^{1-x}$$

$x_i = 1$, αν στην i δοκιμή εμφανίστηκε ε

$x_i = 0$, αν στην i δοκιμή εμφανίστηκε α

Από τα παραπάνω συμπεραίνουμε ότι: $x_1 + x_2 + ... + x_v \sim b(1, p)$

<u>Εφαρμογές της διωνυμικής κατανομής.</u>

Η διωνυμική κατανομή εφαρμόζεται σε πειράματα τύχης που χαρακτηρίζονται από τις ακόλουθες συνθήκες:

- Υπάρχουν δύο <u>μόνο</u> λογικά αποτελέσματα στα πειράματα τύχης: να πραγματοποιηθεί για παράδειγμα το ενδεχόμενο Α, ή να μην πραγματοποιηθεί. Οπότε, αν οι αντίστοιχες πιθανότητες είναι p και q, τότε ισχύει ότι $p + q = 1$.
- Το πείραμα τύχης εφαρμόζεται ν ανεξάρτητες φορές ή εκτελούνται ν ανεξάρτητα, αλλά ίδια πειράματα τύχης κάθε φορά.
- Η πιθανότητα p της πραγματοποίησης του ενδεχομένου Α παραμένει σταθερή στις ν επαναλήψεις του πειράματος τύχης ή είναι ίδια και στα ν πειράματα που θα εκτελεστούν κάθε φορά.

<u>Πίνακες διωνυμικής κατανομής.</u>

Στην πράξη οι πιθανότητες της διωνυμικής κατανομής δίνονται από ειδικούς πίνακες. Οι πίνακες αυτοί περιέχουν τις πιθανότητες για διάφορες τιμές των n και p. Δίνουμε παρακάτω απόσπασμα από πίνακα διωνυμικής κατανομής.

ΔΙΩΝΥΜΙΚΗ ΚΑΤΑΝΟΜΗ

Τιμές της συνάρτησης $P\{X=x\}=C^m_x p^x q^{n-x}$

	x	p=1%	p=2%	p=3%	p=4%	p=5%	p=6%	p=7%	p=8%	p=9%	p=10%
	0	0,9150	0,9039	0,8587	0,8153	0,7738	0,7339	0,6957	0,6591	0,6240	0,5905
n=5	1	0,0480	0,0922	0,1328	0,1699	0,2036	0,2342	0,2618	0,2865	0,3086	0,3280
	2	0,0010	0,0038	0,0082	0,0142	0,0214	0,0299	0,0394	0,0498	0,0610	0,0729
	3			0,0003	0,0006	0,0011	0,0019	0,0030	0,0043	0,0060	0,0081
	4							0,0001	0,0002	0,0003	0,0005
	x	p=1%	p=2%	p=3%	p=4%	p=5%	p=6%	p=7%	p=8%	p=9%	p=10%
	0	0,9044	0,8171	0,737	0,6648	0,5987	0,5386	0,4840	0,4344	0,3894	0,3487
	1	0,0914	0,1667	0,2281	0,2770	0,3151	0,3438	0,3643	0,3777	0,3851	0,3874
n=10	2	0,0042	0,016	0,0317	0,0519	0,0746	0,0988	0,1234	0,1478	0,1714	0,1937
	3	0,0001	0,0008	0,0026	0,0058	0,0105	0,0168	0,0248	0,0343	0,0452	0,0574
	4			0,0001	0,0004	0,0010	0,0019	0,0033	0,0052	0,0078	0,0112
	5					0,0001	0,0001	0,0003	0,0005	0,0009	0,0015
	6									0,0001	0,0001

Χαρακτηριστικές τιμές της διωνυμικής κατανομής.

α) Ο αριθμητικός μέσος ή η μαθηματική ελπίδα μιας διωνυμικής κατανομής είναι:

$$E(X)=\sum_{x=0}^{n} xC_x^m p^x q^{n-x}=\sum_{0}^{n} x\frac{n!}{x!(n-x!)}p^x q^{n-x}=\sum_{0}^{n} x\frac{n(n-1)!}{x(x-1)!(n-x!)}pp^{x-1}q^{n-x}=$$

$$np\sum_{0}^{n} x\frac{(n-1)!}{(x-1)!(n-x!)}p^{x-1}q^{n-1-(x-1)}=np(p+q)^{n-1}=np$$

Άρα: Ε (Χ) = np

β) Η διακύμανση είναι:

$$E\left[X\left(X-1\right)\right]=\sum x\left(x-1\right)C_x^m p^x q^{n-x}=\sum x\left(x-1\right)\frac{n!}{x!\left(n-x\right)!}p^x q^{n-x}=\sum x\left(x-1\right)\frac{n\left(n-1\right)\left(n-2\right)}{x\left(x-1\right)\left(x-2\right)!\left(n-x\right)!}p^2 p$$

$$=n\left(n-1\right)p^2\sum\frac{\left(n-2\right)!}{\left(x-2\right)!\left(n-x\right)!}p^{x-2}q^{n-x}=n\left(n-1\right)p^2\sum C_{x-2}^{m-2}p^{x-2}q^{n-2}=n\left(n-1\right)p^2\left(p+q\right)^{n-2}=n\left(n-1\right)p^2$$

Η τυπική απόκλιση θα είναι: $s=\sqrt{npq}$

γ) Ο συντελεστής ασυμμετρίας αποδεικνύεται ότι είναι:

$$b_1=\frac{\left(q-p\right)^2}{npq}$$

Όσο το n αυξάνει, τόσο ο συντελεστής ασυμμετρίας β_1 τείνει στο μηδέν ($\beta_1 \rightarrow 0$).

δ) Ο συντελεστής κυρτότητας αποδεικνύεται ότι είναι:

$$b_2=3+\frac{1-6pq}{npq}$$

Αν αυξάνει το n, τότε το $\beta_2 \rightarrow 3$.

Αν σε μια διωνυμική κατανομή p = q = ½, τότε η διωνυμική κατανομή είναι συμμετρική.

Αν, όμως p ≠ q, τότε η διωνυμική κατανομή είναι ασυμμετρική. Όσο, όμως, το n αυξάνει, τόσο τείνει να γίνει συμμετρική. Π.χ. για p = 0,9, q = 0,1 και n = 50 ή n = 100 η κατανομή είναι συμμετρική.

10.2.2 Κατανομή Poisson.

Πολλές φορές στην εφαρμογή της διωνυμικής κατανομής παρατηρούμε ότι η τιμή p είναι μικρή (μικρότερη του 0,10), ενώ το μέγεθος του δείγματος n είναι αριθμός αρκετά μεγάλος (μεγαλύτερος του 50), έτσι που ο αναμενόμενος μέσος αριθμητικός E(X) = np να είναι ένας μικρός αριθμός που περιλαμβάνεται μεταξύ του 0 και 10. Με άλλα λόγια, αν η συνεχής τυχαία μεταβλητή X ακολουθεί τη διωνυμική κατανομή, όπου n → ∞ και p → 0, τότε αντί της διωνυμικής κατανομής:

$$P\{X = x\} = P_X = C_x^m p^x q^{n-x}$$

για τον υπολογισμό των πιθανοτήτων *P{X = x},* όπου x = 0, 1, 2, 3, . . ., n, εφαρμόζεται μια άλλη ασυνεχής θεωρητική κατανομή, γνωστή σαν *κατανομή Poisson* και της οποίας η συνάρτηση πιθανότητας δίνεται από τη σχέση:

$$P\{X = x\} = P_X = \frac{e^{-l}\, l^{x}}{x!}$$

όπου x = 0, 1, 2, 3, . . ., n

λ ο μέσος της κατανομής

e = 2,71828

P_X η πιθανότητα σε ένα δείγμα μεγέθους n μονάδων να έχουμε x επιτυχίες.

Η κατανομή Poisson χρησιμοποιείται όταν θέλουμε να μετρήσουμε τον αριθμό των <<συμβάντων>> που εμφανίζονται μέσα σε ένα διάστημα (0, t).

Απαραίτητες προϋποθέσεις για τη χρησιμοποίηση της κατανομής Poisson είναι να ισχύουν τα εξής:

1) Η πιθανότητα να πραγματοποιηθεί ένα συμβάν σε διάστημα μήκους t είναι ανάλογη προς το μήκος του διαστήματος.

2) Η πιθανότητα να πραγματοποιηθούν δύο ή περισσότερα <<συμβάντα>> σε διάστημα μήκους t, είναι πολύ μικρή σε σχέση με την πιθανότητα εμφάνισης μόνο ενός <<συμβάντος>>, όταν το μήκος του διαστήματος είναι μικρό.

3) Η πιθανότητα εμφάνισης k <<συμβάντων>> σε ένα διάστημα είναι ανεξάρτητη από την αντίστοιχη πιθανότητα σε ένα δεύτερο διάστημα ξένο προς το πρώτο.

Αν λοιπόν ισχύουν τα παραπάνω, λέμε ότι η τυχαία μεταβλητή Χ ακολουθεί την κατανομή Poisson.

Η κατανομή Poisson ανακαλύφθηκε από το Γάλλο S.D. Poisson στις αρχές του $19^{ου}$ αιώνα.

Η κατανομή Poisson εφαρμόζεται σε εκείνες τις περιπτώσεις στις οποίες το ενδεχόμενο «επιτυχία» εμφανίζεται σπάνια, ενώ το μέγεθος του δείγματος είναι αρκετά μεγάλο, ώστε ο αναμενόμενος μέσος αριθμητικός $E(X) = np$ να είναι αριθμός μεταξύ του 0 και 10. Στην περίπτωση αυτή οι τιμές της τυχαίας μεταβλητής Χ που αναφέρονται στο ενδεχόμενο «επιτυχία» τείνουν να συγκεντρωθούν στις πρώτες μικρές τιμές της Χ, ενώ οι μεγάλες τιμές δεν έχουν συχνότητες.

Εφαρμογές της κατανομής Poisson.

- Ο αριθμός συγκεκριμένων τηλεφωνικών κλήσεων που φτάνουν σε ένα τηλεφωνικό κέντρο κατά τη διάρκεια ενός διαστήματος, π.χ. ενός δεκαλέπτου.
- Ο αριθμός των αφίξεων μεταφορικών μέσων, π.χ. αεροπλάνων, κατά τη διάρκεια ενός διαστήματος.
- Ο αριθμός των απεργιών σε μια επιχείρηση στο διάστημα ενός χρόνου, κλπ.

Πίνακες κατανομή Poisson.

Για να αποφύγουμε τους υπολογισμούς που απαιτεί ο τύπος (16.2) και επειδή η κατανομή Poisson περιέχει μόνο μία παράμετρο *(λ)*, χρησιμοποιούμε ειδικούς πίνακες που δίνουν τις πιθανότητες για διάφορες τιμές του *λ* = *np*.

Τιμές της συνάρτησης πιθανότητας

$$P\{X = x\} = P_x = \frac{e^{-\lambda}\lambda^x}{x!}$$

X	λ= 0,1	λ = 0,2	λ = 0,3	λ = 0,4	λ = 0,5	λ = 0,6	λ = 0,7	λ = 0,8	λ = 0,9
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066
1	0,0905	0,1637	0,2222	0,2681	0,3033	0,3293	0,3476	0,3595	0,3659
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1217	0,1438	0,1647
3	0,0002	0,0011	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0194
4		0,0001	0,0003	0,0007	0,0016	0,0030	0,0050	0,0077	0,0111
5				0,0001	0,0002	0,0004	0,0007	0,0012	0,0020
6							0,0001	0,0002	0,0003

Χαρακτηριστικές τιμές της κατανομής Poisson

α) Μαθηματική ελπίδα ή μέσος αριθμητικός:

$$E(X) = \sum_{x=0}^{\infty} x e^{-l} \frac{l^x}{x!} = e^{-l}\, l \sum_{x=1}^{\infty} x \frac{l^{x-1}}{x(x-1)!} = e^{-l}\, l \sum_{x=1}^{\infty} \frac{l^{x-1}}{(x-1)!} = e^{-l}\, l\, e^{l} = l = np$$

β) Η διακύμανση είναι:

$$Var(X) = E(X^2) - [E(X)]^2 = \sum_{x=0}^{\infty} x^2 \frac{e^{-l} l^x}{x!} - l^2 = \sum_{x=0}^{\infty} (x^2 - x + x) \frac{e^{-l} l^x}{x!} - l^2$$

$$= \sum_{x=0}^{\infty} \frac{x(x-1)e^{-l} l^x}{x!} + \sum_{x=0}^{\infty} \frac{xe^{-l} l^x}{x!} - l^2 = l^2 \sum_{x=2}^{\infty} \frac{e^{-l} l^{x-2}}{(x-2)!} + l \sum_{x=1}^{\infty} \frac{e^{-l} l^{x-1}}{(x-1)!} - l^2 = l^2 \cdot 1 + l \cdot 1 - l^2 = l = np$$

και η τυπική απόκλιση θα είναι: $s = \sqrt{l}$.

γ) Ο συντελεστής θα δίνεται από τον τύπο:

$$b_1 = \frac{1}{l}$$

Άρα, η κατανομή Poisson τείνει να γίνει συμμετρική, όταν το λ αυξάνει.

δ) Ο συντελεστής κυρτότητας δίνεται από την παρακάτω σχέση:

$$b_2 = 3 + \frac{1}{l}$$

και η κατανομή είναι μεσόκυρτη αν $β_2$ = 3.

10.3 Συνεχείς κατανομές.

Οι κυριότερες συνεχείς κατανομές είναι:

α. Κανονική κατανομή.

β. t-student

γ. x^2

δ. F

Στο κεφάλαιο αυτό θα αναφερθούμε ειδικότερα στην κανονική κατανομή και την τυποποιημένη κανονική κατανομή (με τυποποιημένες τιμές).

10.3.1 Κανονική κατανομή.

Για πρώτη φορά η κανονική κατανομή χρησιμοποιήθηκε το 1733 από το Γάλλο μαθηματικό Abraham de Moivre σαν μια οριακή μορφή της διωνυμικής κατανομής. Όταν το μέγεθος του δείγματος *n* αυξάνει *(n → ∞),* τότε το διωνυμικό ανάπτυγμα $(\rho + q)^n$ προσεγγίζει μια συμμετρική ομαλή καμπύλη που ονομάζεται κανονική κατανομή.

Η εφαρμογή όμως της κανονικής κατανομής στη Στατιστική άρχισε μετά από τις εργασίες του αστρονόμου και μαθηματικού Pierres Laplace και του μαθηματικού Karl Gauss, οι οποίοι τη μελέτησαν ανεξάρτητα ο ένας από τον άλλο και χωρίς να έχουν υπόψη τους την εργασία του De Moivre, και αφορούσε τη θεωρία σφαλμάτων στις διάφορες παρατηρήσεις, γι' αυτό και ονομάζεται και *κατανομή σφαλμάτων* ή *κατανομή των Gauss-Laplace.*

Η κανονική κατανομή είναι η σπουδαιότερη από όλες τις θεωρητικές κατανομές, γιατί έχει πολύ μεγάλη εφαρμογή στη θεωρία των πιθανοτήτων, τη δειγματοληψία, τη Στατιστική, κ.λπ.

Χαρακτηριστικά γνωρίσματα της κανονικής κατανομής

Η μαθηματική μορφή της κανονικής κατανομής δίνεται από την παρακάτω συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \frac{1}{s\sqrt{2p}} e^{\frac{1}{2}\left(\frac{x-m}{s}\right)^2}$$

όπου *μ* ο μέσος αριθμητικός της τυχαίας μεταβλητής *X,* που προσδιορίζει τη θέση της καμπύλης στον άξονα των *x* και μπορεί να πάρει τιμές από -∞ μέχρι +∞, δηλαδή:

$$-\infty \leq \mu \leq \infty$$

σ η τυπική απόκλιση της τυχαίας μεταβλητής X, που παίρνει μόνο θετικές τιμές (σ > 0) και προσδιορίζει το σχήμα της κανονικής κατανομής (καμπύλης)

e = 2,71828, η βάση των ν επέρειων λογαρίθμων *π=3,14159,* ο λόγος της περιφέρειας ενός κύκλου ως προς τη διάμετρό του.

Η καμπύλη που παριστάνει τη συνάρτηση πυκνότητας πιθανότητας της σχέσης έχει τη μορφή καμπάνας, είναι συμμετρική και έχει ασύμπτωτο τον άξονα των τετμημένων (άξονα των *x).* Η συνάρτηση πυκνότητας πιθανότητας κανονικής κατανομής για x = μ παρουσιάζει μέγιστο $\left(\frac{1}{s\sqrt{2p}}\right)$ενώ για τις τιμές x = μ ± σ παρουσιάζει σημεία καμπής.

Η κανονική κατανομή έχει συντελεστή ασυμμετρίας του Pearson $\beta_1 = 0$ και συντελεστή κυρτότητας β_2 = 3. Επίσης, ο μέσος αριθμητικός, η διάμεσος και η επικρατούσα τιμή συμπίπτουν $(\mu = m = m_0)$. Αν διαπιστώσουμε ότι μια κατανομή συχνοτήτων έχει $\beta_1 = 0$ και $\beta_2 = 3$, τότε συμπεραίνουμε ότι η κατανομή συχνοτήτων ακολουθεί κανονική κατανομή.

Παράμετροι της κανονικής κατανομής.

1. Μέση τιμή: Ε(Χ) = μ.
2. Διακύμανση: Var (X) = σ^2 .
3. Τυπική Απόκλιση: Std(X) = σ.
4. Επικρατούσα τιμή: $\mathrm{T} = \left(\frac{1}{s\sqrt{2p}}\right)$
5. Συντελεστής Ασυμμετρίας: $\gamma_1 = 0$.
6. Συντελεστής Κυρτότητας: $\beta_2 = 3$.

Σημείωση:

Στην κανονική κατανομή η μέση τιμή, η επικρατούσα τιμή και η διάμεσος συμπίπτουν.

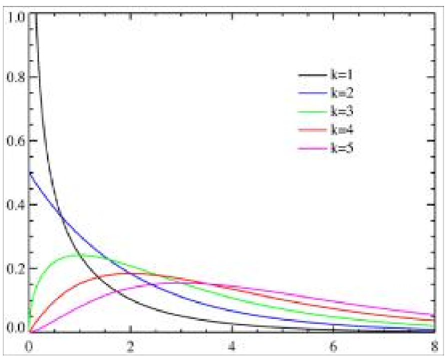
Η κανονική κατανομή συμβολίζεται με $\mathrm{N}\left(m, s^2\right)$και η κανονικοποιημένη με $\mathrm{N}(0,1)$.

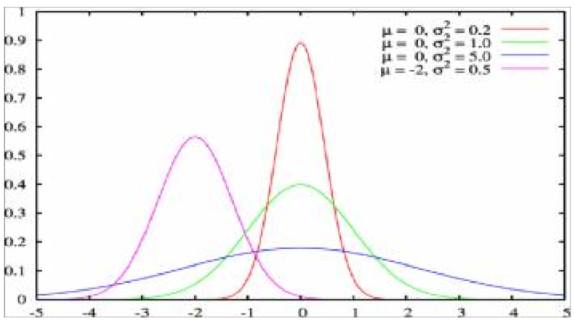
Ιδιότητες της κανονικής κατανομής.

- Εάν $X \sim N\left(m, s^2\right)$, τότε $aX + b \sim N\left(am + b, a^2 s^2\right)$, $a \neq 0$
- Εάν $X \sim N\left(m, s^2\right)$, τότε $\frac{x-m}{s} \sim N(0,1)$ (τυπική κατανομή).
- Εάν $X \sim N(0,1)$, τότε X^2 έχει χι- τετράγωνο κατανομή με ένα βαθμό ελευθερίας.
- Εάν $X \sim N\left(m, s^2\right)$, τότε $\left(\frac{x-m}{s}\right)^2$ έχει χι- τετράγωνο κατανομή με ένα βαθμό ελευθερίας.

Η κανονική κατανομή είναι η πιο χρήσιμη κατανομή στη στατιστική, τόσο γιατί πολλές ποσότητες που συναντάμε στη φύση ακολουθούν τη συγκεκριμένη κατανομή, όσο και γιατί κάτω από ορισμένες συνθήκες είναι δυνατό να προσεγγίζουμε την κατανομή κάποιων τυχαίων μεταβλητών με την κανονική κατανομή λόγω του κεντρικού οριακού θεωρήματος. Ενδεικτικά, μπορεί να αναφερθεί ότι, όταν υπάρχει ένας μεγάλος αριθμός παραγόντων που επιδρούν πάνω σε μια φυσική ποσότητα και οι παράγοντες αυτοί διαμορφώνουν τις τιμές που παίρνει, τότε μπορούμε να ισχυριστούμε ότι η ποσότητα που εξετάζουμε ακολουθεί την κανονική κατανομή. Για παράδειγμα τα ετήσια κέρδη μιας επιχείρησης, τις πωλήσεις προϊόντων της, κλπ.

Το γράφημα της κανονικής κατανομής είναι:





Πίνακας κανονικής κατανομής

z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84850	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92786	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520

2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99897	0.99900

α	**0.0005**	**0.001**	**0.005**	**0.01**	**0.025**	**0.05**	**0.10**
z_{α}	**3.29**	**3.09**	**2.576**	**2.326**	**1.960**	**1.645**	**1.282**

10.3.2. Τυποποιημένη κανονική κατανομή.

Έστω η μεταβλητή $X \sim N\left(m, s^2\right)$. Η μεταβλητή $X = \left(\frac{x-m}{s}\right) \sim N(0,1)$. Η συνάρτηση της μεταβλητής Z δίνεται από τον τύπο $f(z) = \frac{1}{\sqrt{2p}} e^{-\frac{z^2}{2}}$ και η συνάρτηση αθροιστικής κατανομής από $F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2p}} e^{-\frac{x^2}{2}} dx$.

Η συνάρτηση αθροιστικής κατανομής της τυπικής κανονικής κατανομής δεν εξαρτάται από καμία παράμετρο. Συνεπώς, γνωρίζοντας μόνο το Z μπορούμε να βρούμε το F(Z).

Αν γνωρίζουμε την κατανομή που ακολουθεί μια μεταβλητή, μπορούμε να βρούμε ποια πιθανότητα έχουν οι τιμές της να ανήκουν σε κάποιο διάστημα. Επίσης, έχουμε τη δυνατότητα να αντιστρέψουμε τη διαδικασία, δηλαδή μπορούμε να υπολογίσουμε το διάστημα που περιέχει ένα συγκεκριμένο ποσοστό παρατηρήσεων της μεταβλητής.

Οπότε, μπορούμε να συμπεράνουμε πως η κανονική κατανομή που έχει μέσο αριθμητικό 0 και διακύμανση 1, λέγεται τυποποιημένη κανονική κατανομή.

ΕΦΑΡΜΟΓΗ

Θέλοντας να δώσουμε μία πιο σαφή εικόνα όλων των παραπάνω, θα χρησιμοποιήσουμε μία εφαρμογή στην οποία θα αναλύουμε εκτενέστερα τα στατιστικά μέτρα, καταλήγοντας στα συμπεράσματα τα οποία αναφερθήκαμε στο θεωρητικό τμήμα της εργασίας.
Για το λόγο αυτό, λάβαμε στοιχεία των ετήσιων για το 2008 συναλλαγών δύο τραπεζών:

- ΑΓΡΟΤΙΚΗ ΤΡΑΠΕΖΑ ΕΛΛΑΔΟΣ
- ΕΘΝΙΚΗ ΤΡΑΠΕΖΑ ΕΛΛΑΔΟΣ

Ειδικότερα, τα στοιχεία που θα εξετάσουμε, αναφέρονται σε κοινά δανειακά προϊόντα των δύο χρηματοπιστωτικών ιδρυμάτων. Τα προϊόντα αυτά είναι:

I. Δάνεια Καταναλωτικά
II. Δάνεια Στεγαστικά
III. Δάνεια Κεφαλαίων Κίνησης
IV. Δάνεια Επαγγελματικά

Θέτοντας τις μεταβλητές, κάνουμε γνωστή την επεξήγησή τους ως εξής:

- όπου 1: αναφερόμαστε σε δάνεια της Α.Τ.Ε.
- όπου 2: αναφερόμαστε σε δάνεια της Ε.Τ.Ε.

Οι μεταβλητές και τα δεδομένα των μεταβλητών, θα εισαχθούν στο στατιστικό πρόγραμμα SPSS 16.0, όπου θα γίνει και η επίλυσή τους. Αναφορικά, θα παρουσιαστούν τα διαγράμματα και θα διεξαχθούν τα συμπεράσματα που μας ενδιαφέρουν.

Σημείωση: Μεγάλη σημασία δίνεται στο γεγονός ότι οι δύο μεταβλητές που αφορούν καταναλωτικά δάνεια είναι διακριτές με μεγάλο πλήθος τιμών, οπότε και καθίσταται απαραίτητο να δημιουργήσουμε τάξεις, ώστε να έχουμε καλύτερη εικόνα των αποτελεσμάτων.

ΜΕΡΟΣ 1^{Ο}

1. ΠΙΝΑΚΕΣ ΣΥΧΝΟΤΗΤΩΝ ΚΑΙ ΔΙΑΓΡΑΜΜΑΤΑ ΓΙΑ ΤΗ ΜΕΤΑΒΛΗΤΗ ΚΑΤΑΝΑΛΩΤΙΚΑ1 ΔΑΝΕΙΑ.

ΠΙΝΑΚΑΣ 1

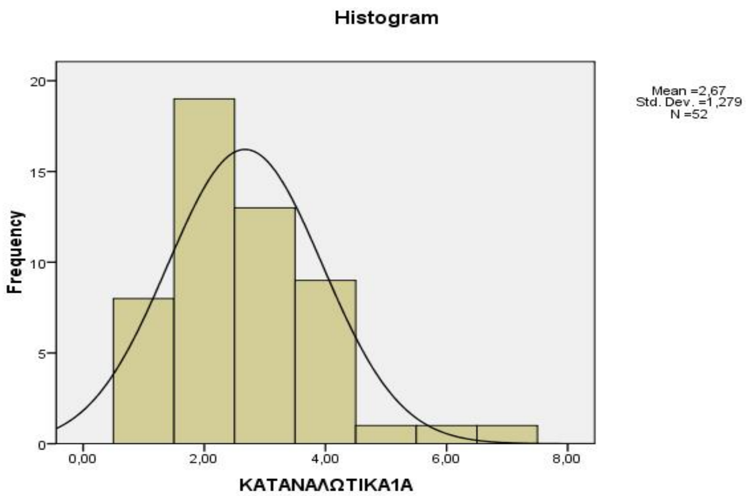
Valid	52
Missing	0
Range	17
Minimum	1
Maximum	18

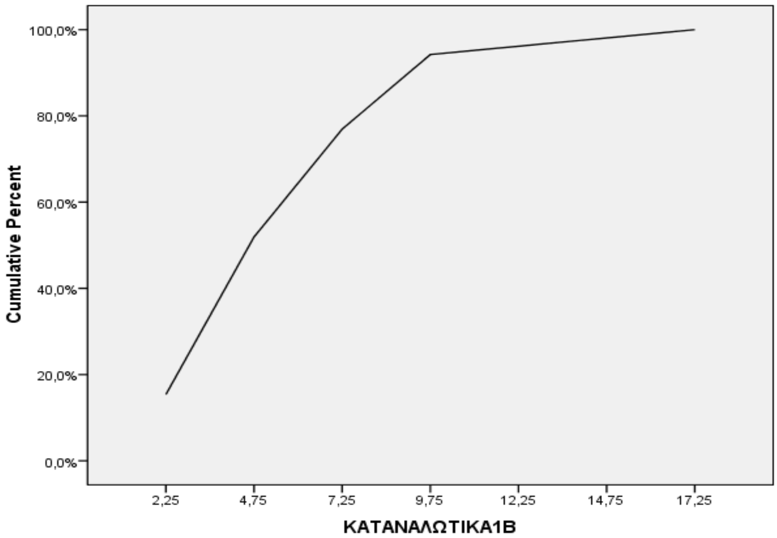
Σαν πρώτη εκτίμηση αναφέρουμε το χαμηλότερο και τον υψηλότερο αριθμό αιτήσεων, αντίστοιχα, όπου μας επεξηγούν πως σε διάστημα 52 εβδομάδων (valid = 52), το εύρος των αιτήσεων ήταν από μία (minimum = 1), έως 18 (maximum = 18).

ΠΙΝΑΚΑΣ 2

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
1-3,5	8	15,4	15,4	15,4
3,5-6,0	19	36,5	36,5	51,9
6,0-8,5	13	25,0	25,0	76,9
8,5-11	9	17,3	17,3	94,2
11-13,5	1	1,9	1,9	96,2
13,5-16	1	1,9	1,9	98,1
16-18,5	1	1,9	1,9	100,0
Total	52	100,0	100,0	

Στον πίνακα 2 παρατίθενται τα εβδομαδιαία ποσοστά των αιτήσεων σύμφωνα με τις τάξεις τις οποίες δημιουργήσαμε.
Αναφορικά, ερμηνεύοντας την $2^{η}$ τάξη η οποία παρουσιάζει τη μεγαλύτερη συχνότητα, θα λέγαμε πως από τρία (3) έως και έξι (6) δάνεια αιτήθηκαν σε δέκα εννέα (19) από τις συνολικά πενήντα δύο (52) εβδομάδες, καλύπτοντας ένα ποσοστό 36,5% επί του συνολικού ποσοστού αιτήσεων.
Παρακάτω, παρουσιάζονται το ιστόγραμμα συχνότητας και το αθροιστικό διάγραμμα.





2. ΠΙΝΑΚΕΣ ΣΥΧΝΟΤΗΤΩΝ ΚΑΙ ΔΙΑΓΡΑΜΜΑΤΑ ΓΙΑ ΤΗ ΜΕΤΑΒΛΗΤΗ ΚΑΤΑΝΑΛΩΤΙΚΑ 2 ΔΑΝΕΙΑ.

ΠΙΝΑΚΑΣ 1

Valid	52
Missing	0
Range	15
Minimum	1
Maximum	16

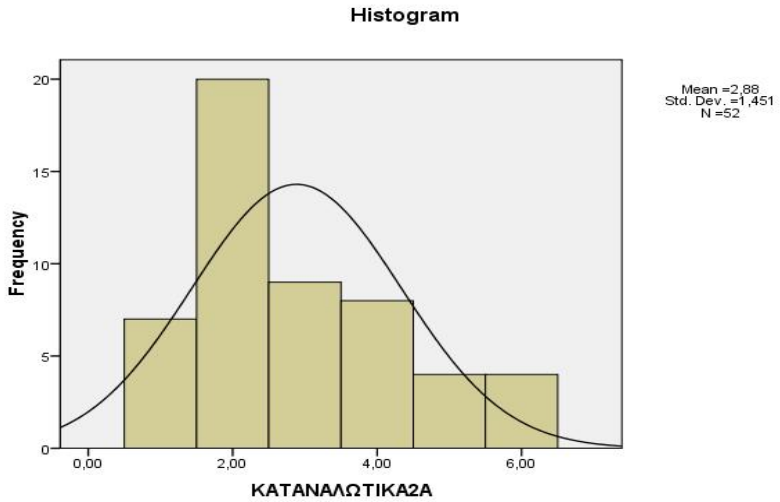
Όσο αφορά τα καταναλωτικά δάνεια της Εθνικής Τράπεζας, παρατηρούμε πως η ελάχιστη εμφάνιση αιτήσεων ήταν ένα δάνειο, ενώ η μεγαλύτερη εμφάνιση ήταν 16 αιτήσεις.

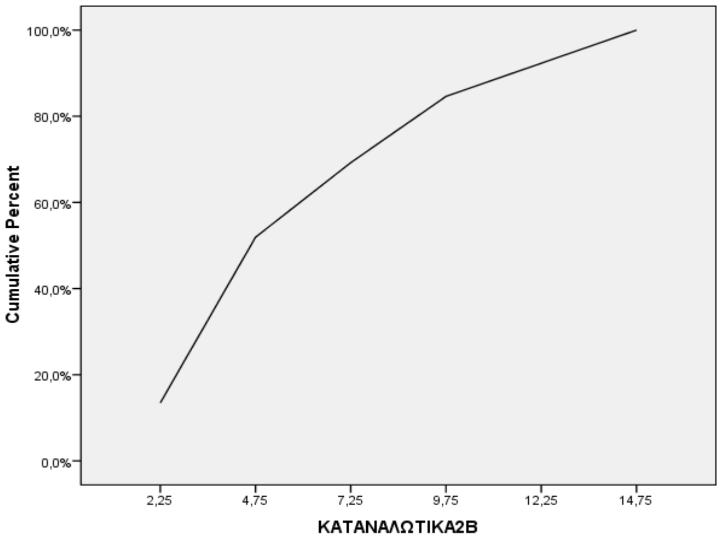
ΠΙΝΑΚΑΣ 2

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
1-4	7	13,5	13,5	15,4
4-7	20	38,5	38,5	51,9
7-10	9	17,3	17,3	69,2
10-13	8	15,4	15,4	84,6
13-16	4	7,7	7,7	92,3
16-19	4	7,7	7,7	100,0
Total	52	100,0	100,0	

Ερμηνεύοντας τις αριθμητικές ενδείξεις που αναφέρονται στον πίνακα και λαμβάνοντας υπ' όψιν τα στοιχεία της προηγούμενης εξέτασης, συγκριτικά αναφέρουμε πως και στα καταναλωτικά δάνεια της Ε.Τ.Ε. μεγαλύτερη εμφάνιση παρουσιάζει η $2^{η}$ τάξη, η οποία περιλαμβάνει από τρεις (3) έως έξι (6) αιτήσεις, με τη διαφορά ότι εδώ ζητήθηκαν σε είκοσι (20) εβδομάδες αποτελώντας έτσι το 38,5% και εν συνεχεία το 51,9% του αθροιστικού ποσοστού το οποίο και αναφέρεται από ένα (1) έως και επτά (7) δάνεια την εβδομάδα.

Το ιστόγραμμα συχνότητας και το αθροιστικό διάγραμμα, διαμορφώνονται ως εξής:





3. ΠΙΝΑΚΕΣ ΣΥΧΝΟΤΗΤΩΝ ΚΑΙ ΔΙΑΓΡΑΜΜΑΤΑ ΓΙΑ ΤΗ ΜΕΤΑΒΛΗΤΗ ΕΠΑΓΓΕΛΜΑΤΙΚΑ 1 ΔΑΝΕΙΑ.

ΠΙΝΑΚΑΣ 1

Valid	52
Missing	0
Range	4
Minimum	0
Maximum	4

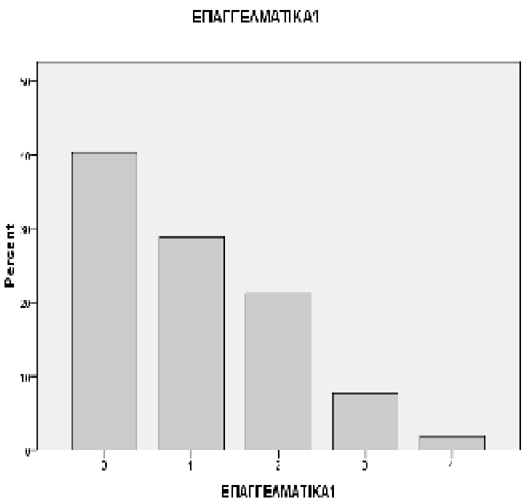
Στα επαγγελματικά δάνεια της Α.Τ.Ε., παρατηρήθηκε πως ζητήθηκαν από κανένα δάνειο (0) έως τέσσερα (4) μέσα σε ολόκληρη τη διάρκεια του έτους.

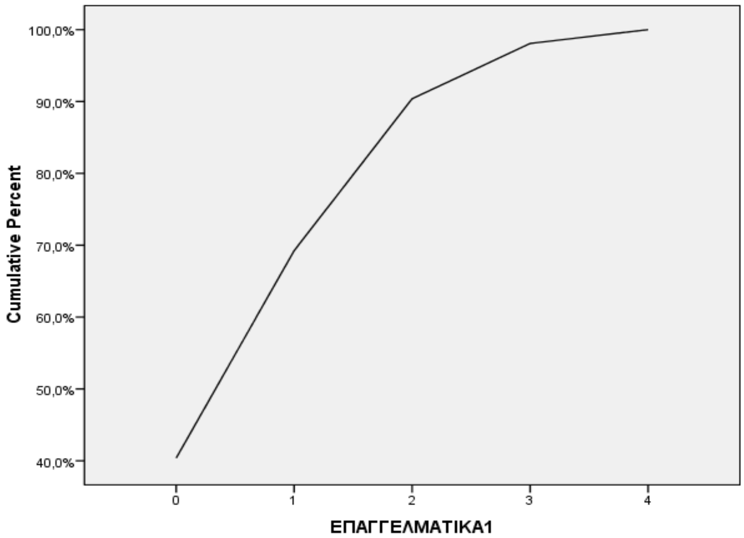
ΠΙΝΑΚΑΣ 2

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
0	21	40,4	40,4	40,4
1	15	28,8	28,8	69,2
2	11	21,2	21,2	90,4
3	4	7,7	7,7	98,1
4	1	1,9	1,9	100,0
Total	52	100,0	100,0	

Εδώ, πλέον, δεν ισχύουν οι τάξεις, αφού τα δεδομένα κυμάνθηκαν μεταξύ μικρού εύρους τιμών. Μία αναφορική εκτίμηση είναι πως το μεγαλύτερο ποσοστό καλύφθηκε από κανένα δάνειο. Αυτό σημαίνει πως σε είκοσι μία (21) εβδομάδες δε ζητήθηκε επαγγελματικό δάνειο.

Τα διαγράμματα θα είναι:





4. ΠΙΝΑΚΕΣ ΣΥΧΝΟΤΗΤΩΝ ΚΑΙ ΔΙΑΓΡΑΜΜΑΤΑ ΓΙΑ ΤΗ ΜΕΤΑΒΛΗΤΗ ΕΠΑΓΓΕΛΜΑΤΙΚΑ 2 ΔΑΝΕΙΑ.

ΠΙΝΑΚΑΣ 1

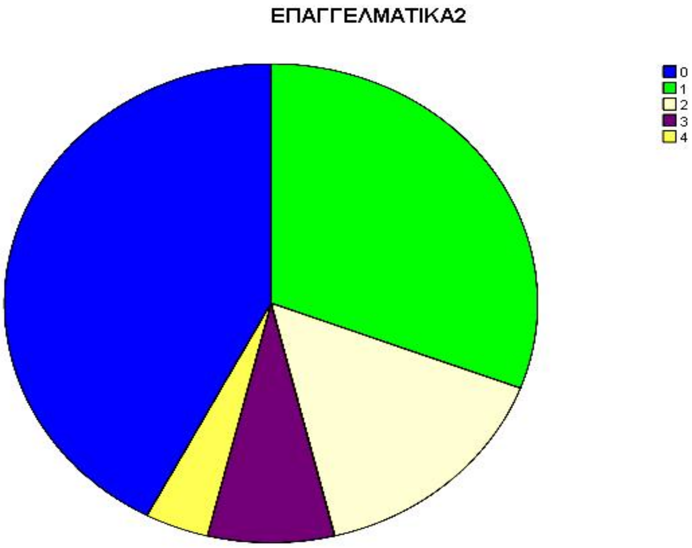
Valid	52
Missing	0
Range	4
Minimum	0
Maximum	4

Στα ίδια επίπεδα αιτήσεων κινήθηκε και η ΕΘΝΙΚΗ ΤΡΑΠΕΖΑ με ελάχιστο όριο τις μηδενικές αιτήσεις και μέγιστο τις τέσσερις (4).

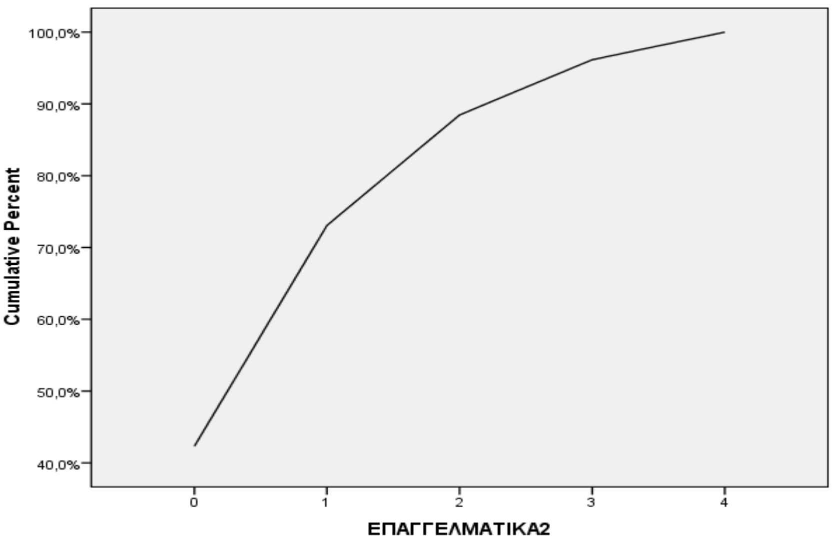
ΠΙΝΑΚΑΣ 2

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
0	22	42,4	42,4	42,3
1	16	30,8	30,8	73,1
2	8	15,4	15,4	88,5
3	4	7,7	7,7	96,2
4	2	3,8	3,8	100,0
Total	52	100,0	100,0	

Επιζητώντας σε ένα βαθμό, να συγκρίνουμε τις χαμηλότερες αιτήσεις, παρατηρούμε πως δεν αιτήθηκε κανένα επαγγελματικό δάνειο σε συνολικά είκοσι δύο (22) εβδομάδες, με συνολικό ποσοστό 42,3%, λίγο υψηλότερο από την προηγούμενη τράπεζα που άγγιζε το 40,4%.
Το διάγραμμα συχνότητας διαμορφώθηκε ως:



Ενώ το αθροιστικό διάγραμμα ως:



5. ΠΙΝΑΚΕΣ ΣΥΧΝΟΤΗΤΩΝ ΚΑΙ ΔΙΑΓΡΑΜΜΑΤΑ ΓΙΑ ΤΗ ΜΕΤΑΒΛΗΤΗ ΣΤΕΓΑΣΤΙΚΑ 1 ΔΑΝΕΙΑ.

ΠΙΝΑΚΑΣ 1

Valid	52
Missing	0
Range	11
Minimum	0
Maximum	11

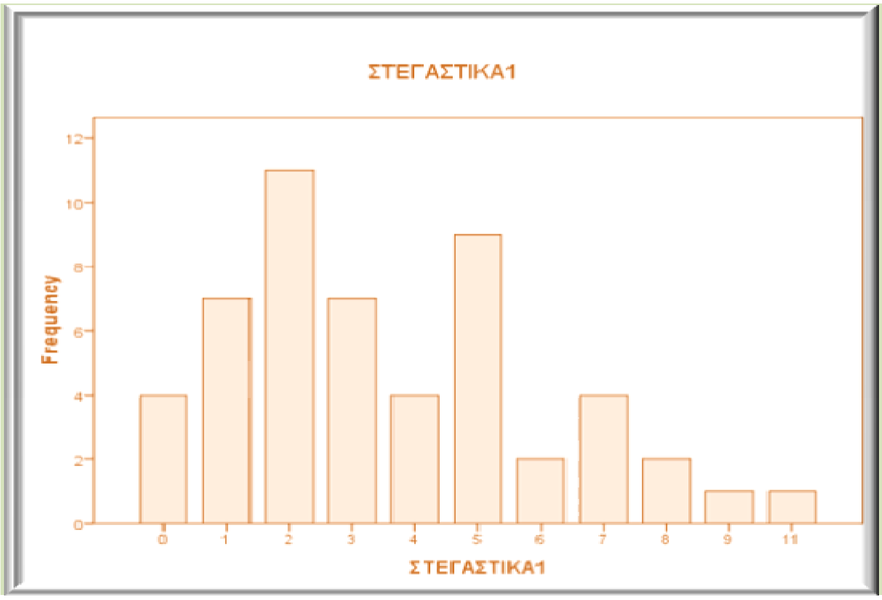
Στα στεγαστικά δάνεια της ΑΓΡΟΤΙΚΗΣ ΤΡΑΠΕΖΑΣ οι περισσότερες αιτήσεις μέσα σε μία εβδομάδα ανήλθαν στις έντεκα (11), εν αντιθέσει με τις χαμηλότερες που ήταν πάλι μηδενικές.

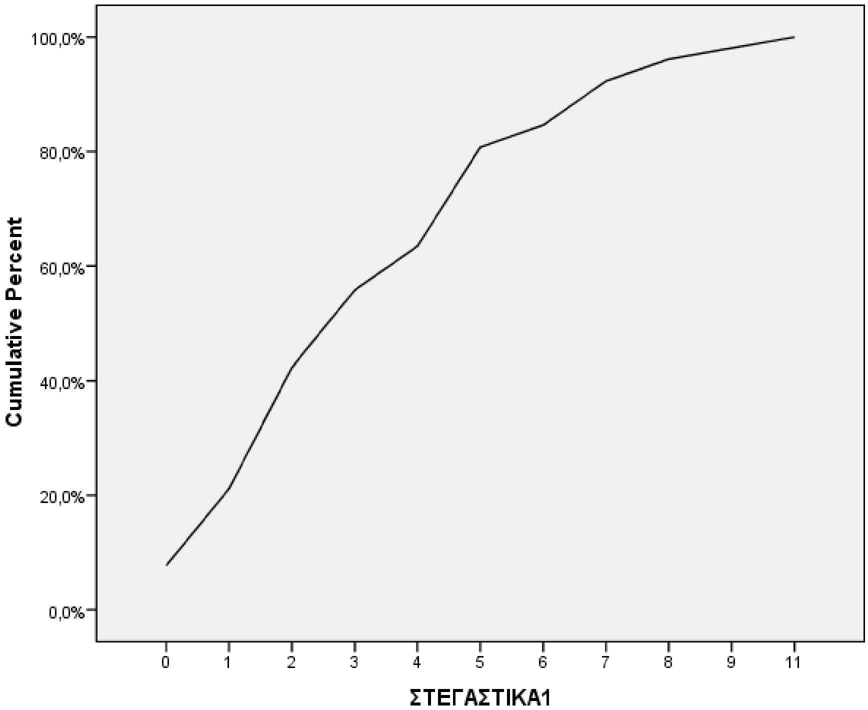
ΠΙΝΑΚΑΣ 2

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
0	4	7,7	7,7	7,7
1	7	13,5	13,5	21,2
2	11	21,2	21,2	42,3
3	7	13,5	13,5	55,8
4	4	7,7	7,7	63,5
5	9	17,3	17,3	80,8
6	2	3,8	3,8	84,6
7	4	7,7	7,7	92,3
8	2	3,8	3,8	96,2
9	1	1,9	1,9	98,1
11	1	1,9	1,9	100,0
Total	52	100,0	100,0	

Αξίζει εδώ να αναφέρουμε πως τη συχνότερη εμφάνιση αιτήσεων κατείχαν οι δύο (2) αιτήσεις, οι οποίες ζητήθηκαν σε έντεκα εβδομάδες από το σύνολο με ποσοστό 21,2% και αντίστοιχα 42,3% επί του αθροιστικού ποσοστό που εμπεριέχει από κανένα έως και δύο δάνεια την κάθε εβδομάδα.

Παραθέτοντας και τα διαγράμματα:





6. ΠΙΝΑΚΕΣ ΣΥΧΝΟΤΗΤΩΝ ΚΑΙ ΔΙΑΓΡΑΜΜΑΤΑ ΓΙΑ ΤΗ ΜΕΤΑΒΛΗΤΗ ΣΤΕΓΑΣΤΙΚΑ 2 ΔΑΝΕΙΑ.

ΠΙΝΑΚΑΣ 1

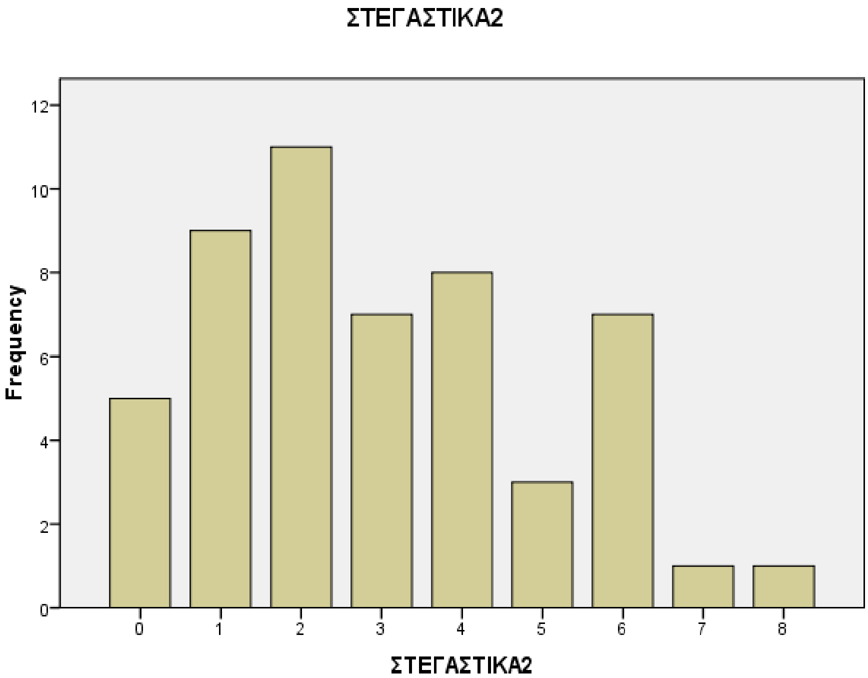
Valid	52
Missing	0
Range	8
Minimum	0
Maximum	8

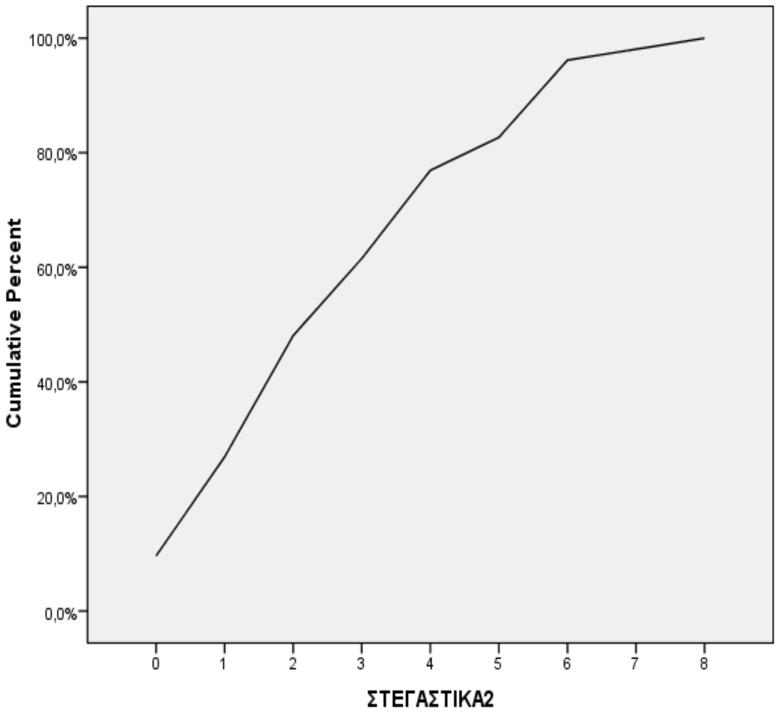
Σε αντίθεση με το μέγιστο αριθμό αιτήσεων στεγαστικών δανείων στην Α.Τ.Ε. ο οποίος άγγιζε τις έντεκα (11), στην Ε.Τ.Ε. το υψηλότερο όριο έφτασε τις οκτώ (8) αιτήσεις, ενώ το χαμηλότερο ήταν καμία αίτηση.

ΠΙΝΑΚΑΣ 2

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
0	5	9,6	9,6	9,6
1	9	17,3	17,3	26,9
2	11	21,2	21,2	48,1
3	7	13,5	13,5	61,5
4	8	15,4	15,4	76,9
5	3	5,8	5,8	82,7
6	7	13,5	13,5	96,2
7	1	1,9	1,9	98,1
8	1	1,9	1,9	100,0
Total	52	100,0	100,0	

Διαπιστώνουμε και στα στεγαστικά της ΕΘΝΙΚΗΣ πως δύο (2) ήταν οι αιτήσεις με τη μεγαλύτερη συχνότητα εμφάνισης σε ποσοστό 48,1% επί του συνολικού αθροίσματος από μηδέν (0) μέχρι δύο (2) δάνεια. Θα δώσουμε, όμως, λίγη έμφαση στο γεγονός ότι στην αμέσως επόμενη θέση βρίσκεται η μία (1) αίτηση με συχνότητα τις εννέα (9) εβδομάδες μέσα στο έτος, σε αντίθεση με τις αιτήσεις της ΑΓΡΟΤΙΚΗΣ όπου στην ίδια θέση ήταν επτά (7) αιτήματα.
Παρακάτω παραβάλλονται και τα διαγράμματα:





7. ΠΙΝΑΚΕΣ ΣΥΧΝΟΤΗΤΩΝ ΚΑΙ ΔΙΑΓΡΑΜΜΑΤΑ ΓΙΑ ΤΗ ΜΕΤΑΒΛΗΤΗ ΔΑΝΕΙΑ ΚΕΦΑΛΑΙΩΝ ΚΙΝΗΣΗΣ 1.

ΠΙΝΑΚΑΣ 1

Valid	52
Missing	0
Range	5
Minimum	0
Maximum	5

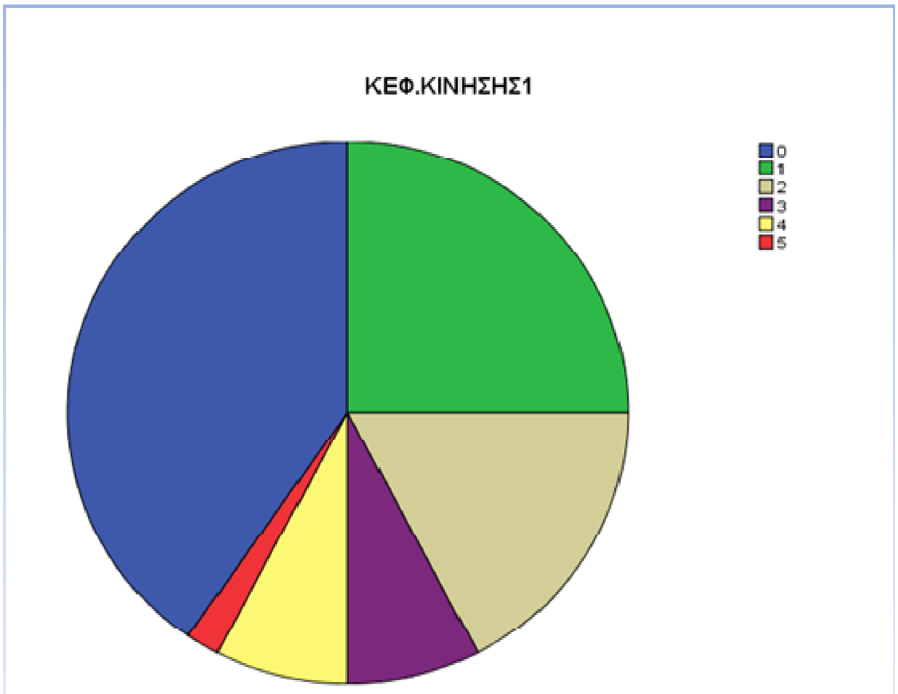
Σε αυτήν την κατηγορία δανείων ο μέγιστος αριθμός αιτήσεων ήταν μόλις πέντε (5), δεδομένο το οποίο μας επιτρέπει να αντιληφθούμε τη σπανιότητα και την εξεζητημένη φύση αυτών των δανείων.

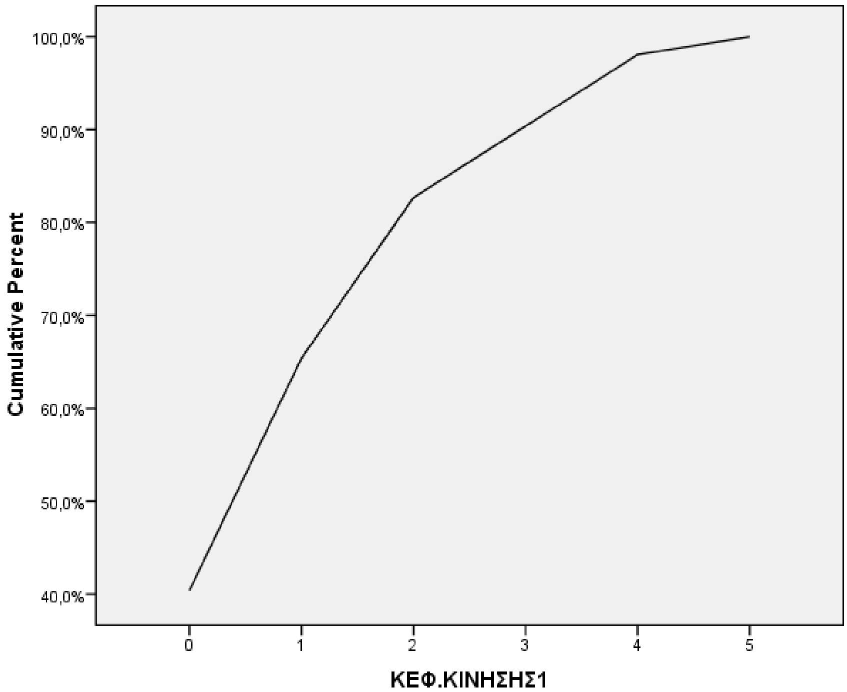
ΠΙΝΑΚΑΣ 2

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
0	5	9,6	9,6	9,6
1	9	17,3	17,3	26,9
2	11	21,2	21,2	48,1
3	7	13,5	13,5	61,5
4	8	15,4	15,4	76,9
5	3	5,8	5,8	100,0
Total	52	100,0	100,0	

Από τον πίνακα κατανομής συχνοτήτων έχουμε να παρατηρήσουμε πως από το σύνολο των εβδομάδων σε έντεκα από αυτές ζητήθηκαν δύο (2) δάνεια. Επιπλέον,

αντιλαμβανόμαστε πως εξαιρουμένων πέντε (5) εβδομάδων μέσα στη χρονιά όπου η κίνησή τους παρέμεινε μηδενική, κάθε εβδομάδα υπήρχε τουλάχιστον ένα (1) αίτημα.
Διαγράμματα:





8. ΠΙΝΑΚΕΣ ΣΥΧΝΟΤΗΤΩΝ ΚΑΙ ΔΙΑΓΡΑΜΜΑΤΑ ΓΙΑ ΤΗ ΜΕΤΑΒΛΗΤΗ ΔΑΝΕΙΑ ΚΕΦΑΛΑΙΩΝ ΚΙΝΗΣΗΣ 2.

ΠΙΝΑΚΑΣ 1

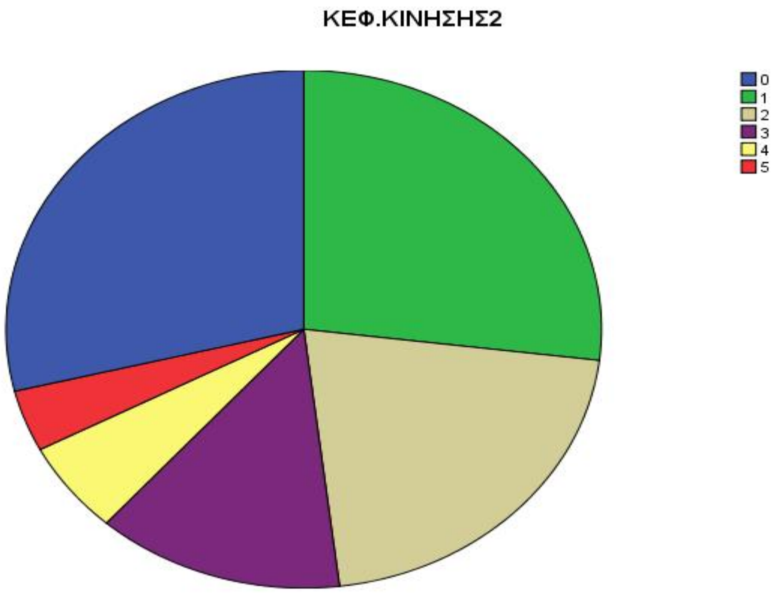
Valid	52
Missing	0
Range	5
Minimum	0
Maximum	5

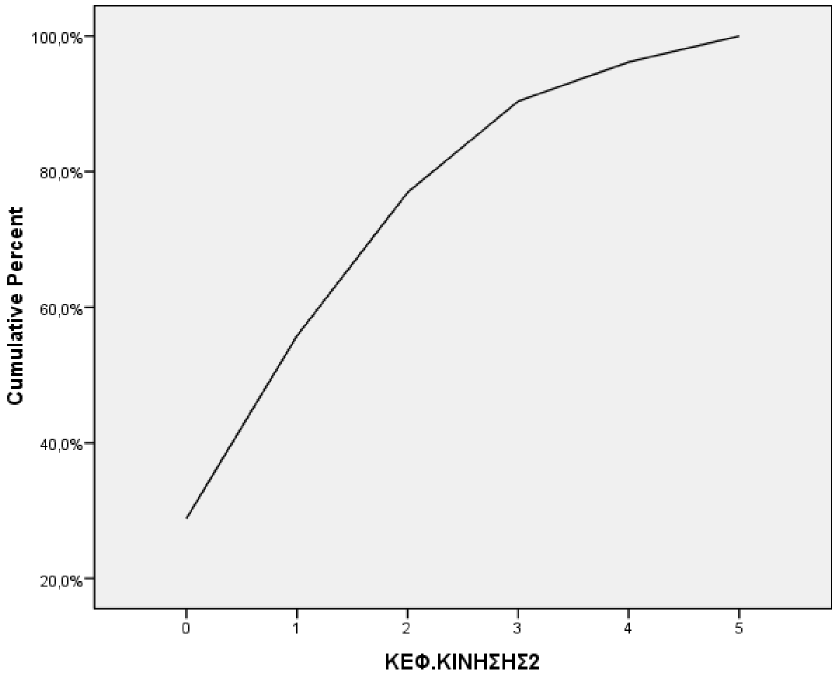
Με πανομοιότυπη συμπεριφορά κινήθηκαν και τα κεφάλαια κίνησης στην ΕΘΝΙΚΗ φτάνοντας σε μέγιστο βαθμό τις πέντε (5) αιτήσεις.

ΠΙΝΑΚΑΣ 2

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
0	15	28,8	28,8	28,8
1	14	26,9	26,9	55,8
2	11	21,2	21,2	76,9
3	7	13,5	13,5	90,4
4	3	5,8	5,8	96,2
5	2	3,8	3,8	100,0
Total	52	100,0	100,0	

Ερμηνεύοντας τα στοιχεία του πίνακα, θα λέγαμε πως μέσα στο χρόνο το συγκεκριμένο είδος δανείων δε ζητήθηκε ιδιαίτερα, αγγίζοντας τη μεγαλύτερη συχνότητα των δέκα τεσσάρων (14) εβδομάδων με μόνο μία (1) αίτηση, ποσοστό παραπάνω από το μισό (55,8%) του συνολικού αθροιστικού ποσοστού.
Έτσι, τα διαγράμματα θα διαμορφωθούν ως εξής:





ΜΕΡΟΣ 2^{ο}

Στο περιγραφικό μέρος της εφαρμογής δώσαμε έμφαση στην κατανομή συχνότητας κάθε μεταβλητής, με σκοπό να παρουσιάσουμε όσο σαφέστερα γίνεται τα αποτελέσματα της έρευνας.

Στο αμέσως επόμενο μέρος θα ασχοληθούμε με τα μέτρα θέσης των μεταβλητών. Πιο συγκεκριμένα, θα συγκρίνουμε τα αποτελέσματα για κάθε είδος δανείων των τραπεζών, ώστε να καταλήξουμε στα συμπεράσματα που θα μας βοηθήσουν να καταλάβουμε σε τι βαθμό η μία τράπεζα ξεπέρασε την άλλη όσο αφορά τα ύψη των αιτήσεων που πραγματοποιήθηκαν.

Επίσης, θα πρέπει να επισημάνουμε πως τα σημαντικότερα μέτρα με τα οποία θα ασχοληθούμε είναι:

- ΑΡΙΘΜΗΤΙΚΟΣ ΜΕΣΟΣ: Μας υποδεικνύει το μέσο όρο των αιτήσεων που πραγματοποιηθήκαν ανά εβδομάδα.

- 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.

- 5% TRIMMED MEAN: Δείχνει τον αριθμητικό μέσο που προκύπτει αν αφαιρέσουμε τις πέντε μεγαλύτερες και πέντε μικρότερες παρατηρήσεις από το δείγμα που εξετάζουμε.

Παραθέτουμε τους πίνακες οι οποίοι περιλαμβάνουν τα μέτρα θέσης για κάθε προϊόν.

ΜΕΤΡΑ ΘΕΣΗΣ	ΚΑΤΑΝΑΛΩΤΙΚΑ1	ΚΑΤΑΝΑΛΩΤΙΚΑ 2
ΜΕΣΟΣ	6,73	7,23
Δ.Ε. 95% (ΚΑΤΩΤΑΤΟ ΟΡΙΟ, ΑΝΩΤΑΤΟ ΟΡΙΟ)	5,89 7,59	6,22 8,24
TRIMMED MEAN 5%	6,52	7,10
ΔΙΑΜΕΣΟΣ	6,00	6,00
ΔΙΑΚΥΜΑΝΣΗ	9,612	13,201
ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ	3,100	3,633
ΕΛΑΧΙΣΤΟ	1	1
ΜΕΓΙΣΤΟ	18	16
ΕΥΡΟΣ	17	15
ΑΣΥΜΜΕΤΡΙΑ	1,137	0,603
ΚΥΡΤΩΣΗ	2,858	-0,367

Όσο αφορά τα καταναλωτικά δάνεια, στην Α.Τ.Ε. ζητήθηκαν κατά μέσο όρο ανά εβδομάδα 6,73, ενώ στην Ε.Τ.Ε. 7,23. Σύμφωνα με τη διάμεσο, το ήμισυ των αιτήσεων ήταν πάνω από έξι δάνεια την εβδομάδα και το υπόλοιπο 50% κάτω από αυτό το όριο, παραδοχή που ισχύει και για τις δύο τράπεζες.
Λαμβάνοντας τώρα υπ' όψιν το 5%, στην Α.Τ.Ε. αιτήθηκαν 6,50 δάνεια, ενώ στην Ε.Τ.Ε. 7,10.
Τέλος, ερμηνεύοντας και τους συντελεστές, έχουμε:

ΑΓΡΟΤΙΚΗ ΤΡΑΠΕΖΑ

- <u>συντελεστής ασυμμετρίας</u> = $\frac{1.137}{0.33} = 3.45$, άρα θετική ασυμμετρία
- <u>συντελεστής κύρτωσης</u> = $\frac{2.858}{0.65} = 4.40$, άρα λεπτόκυρτη κατανομή

ΕΘΝΙΚΗ ΤΡΑΠΕΖΑ

- <u>συντελεστής ασυμμετρίας</u> = $\frac{0.603}{0.33} = 1.83$, άρα συμμετρική κατανομή
- <u>συντελεστής κύρτωσης</u> = $\frac{-0.367}{0.65} = -0.56$, άρα μεσόκυρτη κατανομή

ΜΕΤΡΑ ΘΕΣΗΣ	ΣΤΕΓΑΣΤΙΚΑ1	ΣΤΕΓΑΣΤΙΚΑ 2
ΜΕΣΟΣ	3,60	3,00
Δ.Ε. 95% (ΚΑΤΩΤΑΤΟ ΟΡΙΟ, ΑΝΩΤΑΤΟ ΟΡΙΟ)	2,89 4,30	2,43 3,57
TRIMMED MEAN 5%	3,47	2,94
ΔΙΑΜΕΣΟΣ	3,00	3,00
ΔΙΑΚΥΜΑΝΣΗ	6,442	4,235
ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ	2,538	2,058
ΕΛΑΧΙΣΤΟ	0	0
ΜΕΓΙΣΤΟ	11	8
ΕΥΡΟΣ	11	8
ΑΣΥΜΜΕΤΡΙΑ	0,734	0,449
ΚΥΡΤΩΣΗ	0,168	-0,647

Στα στεγαστικά δάνεια η κίνηση είχε ως εξής: 3,6 αιτήσεις την εβδομάδα για την Α.Τ.Ε. και 3 για την Ε.Τ.Ε. Και στις δύο τράπεζες οι μισές εβδομαδιαίες αιτήσεις ήταν πάνω από 3, ενώ το 5% μας παραπέμπει στο γεγονός ότι οι αιτήσεις άγγιξαν κατά μέσο όρο τις 3,47 για την ΑΓΡΟΤΙΚΗ και τις 2,94 για την ΕΘΝΙΚΗ.
Τέλος, ερμηνεύοντας και τους συντελεστές, έχουμε:

ΑΓΡΟΤΙΚΗ ΤΡΑΠΕΖΑ

- συντελεστής ασυμμετρίας = $\frac{0.734}{0.33} = 2.23$, άρα θετική ασυμμετρία
- συντελεστής κύρτωσης = $\frac{0.168}{0.65} = 0.26$, άρα μεσόκυρτη κατανομή

ΕΘΝΙΚΗ ΤΡΑΠΕΖΑ

- συντελεστής ασυμμετρίας = $\frac{0.449}{0.33} = 1.36$, άρα συμμετρική κατανομή
- συντελεστής κύρτωσης = $\frac{-0.647}{0.65} = -0.99$, άρα μεσόκυρτη κατανομή

ΜΕΤΡΑ ΘΕΣΗΣ	ΚΕΦΑΛΑΙΑ ΚΙΝΗΣΗΣ 1	ΚΕΦΑΛΑΙΑ ΚΙΝΗΣΗΣ 2
ΜΕΣΟΣ	1,23	1,52
Δ.Ε. 95% (ΚΑΤΩΤΑΤΟ ΟΡΙΟ, ΑΝΩΤΑΤΟ ΟΡΙΟ)	0,85 1,61	1,13 1,91
TRIMMED MEAN	1,12	1,42

5%		
ΔΙΑΜΕΣΟΣ	1,00	1,00
ΔΙΑΚΥΜΑΝΣΗ	1,867	1,941
ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ	1,366	1,393
ΕΛΑΧΙΣΤΟ	0	0
ΜΕΓΙΣΤΟ	5	5
ΕΥΡΟΣ	5	5
ΑΣΥΜΜΕΤΡΙΑ	1,002	0,751
ΚΥΡΤΩΣΗ	0,153	-0,126

Τα κεφάλαια κίνησης όπως διαπιστώνουμε και από τον πίνακα μέτρων, δεν παρουσίασαν ιδιαίτερη ζήτηση με μόλις 1,23 αιτήσεις για την $1^{η}$ τράπεζα και 1,52 για τη $2^{η}$. Οπότε οι μισές αιτήσεις ήταν πάνω από μία την εβδομάδα, ενώ υπήρξαν και εβδομάδες μέσα στο έτος, όπου δεν πραγματοποιήθηκε καμία αίτηση.
Τέλος, ερμηνεύοντας και τους συντελεστές, έχουμε:

ΑΓΡΟΤΙΚΗ ΤΡΑΠΕΖΑ

- <u>συντελεστής ασυμμετρίας</u> = $\frac{1.002}{0.33} = 3.04$, άρα θετική ασυμμετρία
- <u>συντελεστής κύρτωσης</u> = $\frac{0.153}{0.65} = 0.24$, άρα μεσόκυρτη κατανομή

ΕΘΝΙΚΗ ΤΡΑΠΕΖΑ

- <u>συντελεστής ασυμμετρίας</u> = $\frac{0.751}{0.33} = 2.28$, άρα θετική ασυμμετρία
- <u>συντελεστής κύρτωσης</u> = $\frac{-0.126}{0.65} = -0.19$, άρα μεσόκυρτη κατανομή.

ΜΕΤΡΑ ΘΕΣΗΣ	ΕΠΑΓΓΕΛΜΑΤΙΚΑ 1	ΕΠΑΓΓΕΛΜΑΤΙΚΑ 2
ΜΕΣΟΣ	1,02	1,00
Δ.Ε. 95% (ΚΑΤΩΤΑΤΟ ΟΡΙΟ, ΑΝΩΤΑΤΟ ΟΡΙΟ)	0,72 1,31	0,69 1,31
TRIMMED MEAN 5%	0,94	0,90
ΔΙΑΜΕΣΟΣ	1,00	1,00
ΔΙΑΚΥΜΑΝΣΗ	1,117	1,255
ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ	1,057	1,120
ΕΛΑΧΙΣΤΟ	0	0
ΜΕΓΙΣΤΟ	4	4
ΕΥΡΟΣ	4	4
ΑΣΥΜΜΕΤΡΙΑ	0,789	1,044

ΚΥΡΤΩΣΗ	-0,122	0,401

Τέλος, για τα επαγγελματικά δάνεια, αιτήθηκαν 1,02 στην ΑΓΡΟΤΙΚΗ ΤΡΑΠΕΖΑ και 1 στην ΕΘΝΙΚΗ. Το 50% των εβδομαδιαίων αιτήσεων ήταν πάνω από μία και το υπόλοιπο 50% κάτω από μία δηλαδή κανένα δάνειο και στα δύο ιδρύματα.
Τέλος, ερμηνεύοντας και τους συντελεστές, έχουμε:

ΑΓΡΟΤΙΚΗ ΤΡΑΠΕΖΑ

- <u>συντελεστής ασυμμετρίας</u> = $\frac{0.789}{0.33} = 2.39$, άρα θετική ασυμμετρία
- <u>συντελεστής κύρτωσης</u> = $\frac{-0.122}{0.65} = -0.19$, άρα μεσόκυρτη κατανομή

ΕΘΝΙΚΗ ΤΡΑΠΕΖΑ

- <u>συντελεστής ασυμμετρίας</u> = $\frac{1.044}{0.33} = 3.16$, άρα θετική ασυμμετρία
- <u>συντελεστής κύρτωσης</u> = $\frac{0.401}{0.65} = 0.62$, άρα μεσόκυρτη κατανομή.

.

<u>ΜΕΤΡΑ ΘΕΣΗΣ ΜΕΤΑΒΛΗΤΩΝ ΓΙΑ ΚΑΘΕ ΤΡΙΜΗΝΟ ΕΤΟΥΣ 2008</u>

ΜΕΤΡΑ ΘΕΣΗΣ	ΤΡΙΜΗΝΑ	ΚΑΤΑΝΑΛΩΤΙΚΑ 1	ΚΑΤΑΝΑΛΩΤΙΚΑ 2
ΜΕΣΟΣ	1^{O}	4,85	5,31
	2^{O}	6,00	6,08
	3^{O}	9,46	10,31
	4^{O}	6,62	7,23
Δ.Ε. 95% (ΚΑΤΩΤΑΤΟ ΟΡΙΟ, ΑΝΩΤΑΤΟ ΟΡΙΟ)	1ο	3,59 6,10	3,76 6,86
	2ο	4,50 7,50	3,98 8,18
	3ο	7,21 11,72	8,46 12,16
	4ο	5,42 7,81	5,11 9,36
TRIMMED MEAN 5%	1^{O}	4,71	5,12
	2^{O}	6,06	5,97
	3^{O}	9,24	10,29
	4^{O}	6,63	6,98

ΔΙΑΜΕΣΟΣ	$1^{Ο}$	5,00	5,00
	$2^{Ο}$	6,00	5,00
	$3^{Ο}$	9,00	11,00
	$4^{Ο}$	6,00	7,00
ΔΙΑΚΥΜΑΝΣΗ	$1^{Ο}$	4,308	6,564
	$2^{Ο}$	6,67	12,077
	$3^{Ο}$	13,936	9,397
	$4^{Ο}$	3,923	12,359
ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ	$1^{Ο}$	2,075	2,562
	$2^{Ο}$	2,483	3,475
	$3^{Ο}$	3,733	3,066
	$4^{Ο}$	1,981	3,516
ΕΛΑΧΙΣΤΟ	$1^{Ο}$	2	2
	$2^{Ο}$	1	1
	$3^{Ο}$	5	6
	$4^{Ο}$	3	3
ΜΕΓΙΣΤΟ	$1^{Ο}$	9	12
	$2^{Ο}$	10	13
	$3^{Ο}$	18	15
	$4^{Ο}$	10	16
ΕΥΡΟΣ	$1^{Ο}$	7	10
	$2^{Ο}$	9	12
	$3^{Ο}$	13	9
	$4^{Ο}$	7	13
ΑΣΥΜΜΕΤΡΙΑ	$1^{Ο}$	0,82	2,42
	$2^{Ο}$	-0,69	0,90
	$3^{Ο}$	1,90	0,01
	$4^{Ο}$	-0,20	2,11
ΚΥΡΤΩΣΗ	$1^{Ο}$	-0,45	2,72
	$2^{Ο}$	0,03	-0,28
	$3^{Ο}$	0,90	-1,05
	$4^{Ο}$	-0,29	1,91

ΜΕΤΡΑ ΘΕΣΗΣ	ΤΡΙΜΗΝΑ	ΣΤΕΓΑΣΤΙΚΑ 1	ΣΤΕΓΑΣΤΙΚΑ 2
ΜΕΣΟΣ	$1^{Ο}$	3,62	3,62
	$2^{Ο}$	3,46	3,92
	$3^{Ο}$	5,23	2,85
	$4^{Ο}$	2,08	1,62
Δ.Ε. 95% (ΚΑΤΩΤΑΤΟ ΟΡΙΟ,	1ο	2,34 4,89	2,08 5,15
	2ο	2,34 4,58	2,78 5,07
		3,22	1,78

ΑΝΩΤΑΤΟ ΟΡΙΟ)	3ο 4ο	7,24 1,02 3,14	3,92 0,89 2,34
TRIMMED MEAN 5%	1^{O} 2^{O} 3^{O} 4^{O}	3,52 3,40 5,20 2,03	3,57 3,91 2,83 1,57
ΔΙΑΜΕΣΟΣ	1^{O} 2^{O} 3^{O} 4^{O}	3,00 3,00 5,00 2,00	4,00 4,00 2,00 1,00
ΔΙΑΚΥΜΑΝΣΗ	1^{O} 2^{O} 3^{O} 4^{O}	6,564 3,436 11,026 3,077	6,423 3,577 3,141 1,423
ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ	1^{O} 2^{O} 3^{O} 4^{O}	2,103 1,854 3,320 1,754	2,534 1,891 1,772 1,193
ΕΛΑΧΙΣΤΟ	1^{O} 2^{O} 3^{O} 4^{O}	1 1 0 0	0 1 0 0
ΜΕΓΙΣΤΟ	1^{O} 2^{O} 3^{O} 4^{O}	8 7 11 11	8 7 6 4
ΕΥΡΟΣ	1^{O} 2^{O} 3^{O} 4^{O}	7 6 11 5	8 6 6 4
ΑΣΥΜΜΕΤΡΙΑ	1^{O} 2^{O} 3^{O} 4^{O}	0,98 0,79 -0,02 0,84	0,65 -0,21 0,96 0,89
ΚΥΡΤΩΣΗ	1^{O} 2^{O} 3^{O} 4^{O}	-0,18 -0,43 -0,74 -0,62	-0,85 -0,67 0,003 -0,21

ΜΕΤΡΑ ΘΕΣΗΣ	ΤΡΙΜΗΝΑ	ΚΕΦΑΛΑΙΑ ΚΙΝΗΣΗΣ 1	ΚΕΦΑΛΑΙΑ ΚΙΝΗΣΗΣ 2
	1^{O}	0,92	0,92

ΜΕΣΟΣ	2^{o}	2,38	2,62
	3^{o}	0,92	1,15
	4^{o}	0,69	1,38
Δ.Ε. 95% (ΚΑΤΩΤΑΤΟ ΟΡΙΟ, ΑΝΩΤΑΤΟ ΟΡΙΟ)	1ο	0,09 1,76	0,35 1,50
	2ο	1,41 3,36	1,64 3,59
	3ο	0,35 1,50	0,56 1,75
	4ο	0,24 1,15	0,55 2,22
TRIMMED MEAN 5%	1^{o}	0,80	0,86
	2^{o}	2,37	2,63
	3^{o}	2,83	0,86
	4^{o}	1,57	0,66
ΔΙΑΜΕΣΟΣ	1^{o}	0,00	1,00
	2^{o}	2,00	3,00
	3^{o}	1,00	1,00
	4^{o}	1,00	1,00
ΔΙΑΚΥΜΑΝΣΗ	1^{o}	1,91	0,91
	2^{o}	2,59	2,59
	3^{o}	0,974	1,308
	4^{o}	0,564	1,923
ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ	1^{o}	1,382	0,954
	2^{o}	1,609	1,609
	3^{o}	0,954	0,987
	4^{o}	0,751	1,387
ΕΛΑΧΙΣΤΟ	1^{o}	0	0
	2^{o}	0	0
	3^{o}	0	0
	4^{o}	0	0
ΜΕΓΙΣΤΟ	1^{o}	4	3
	2^{o}	5	5
	3^{o}	3	3
	4^{o}	2	4
ΕΥΡΟΣ	1^{o}	4	3
	2^{o}	5	5
	3^{o}	3	3
	4^{o}	2	4
ΑΣΥΜΜΕΤΡΙΑ	1^{o}	2,08	1,39
	2^{o}	-0,06	0,06
	3^{o}	1,39	0,43

	4^{O}	0,99	0,82
ΚΥΡΤΩΣΗ	1^{O} 2^{O} 3^{O} 4^{O}	0,41 -0,89 0,19 -0,65	0,19 0,89 -0,77 -0,82

ΜΕΤΡΑ ΘΕΣΗΣ	ΤΡΙΜΗΝΑ	ΕΠΑΓΓΕΛΜΑΤΙΚΑ 1	ΕΠΑΓΓΕΛΜΑΤΙΚΑ 2
ΜΕΣΟΣ	1^{O} 2^{O} 3^{O} 4^{O}	0,31 1,46 1,15 1,15	1,62 0,85 0,69 0,85
Δ.Ε. 95% (ΚΑΤΩΤΑΤΟ ΟΡΙΟ, ΑΝΩΤΑΤΟ ΟΡΙΟ)	1ο 2ο 3ο 4ο	-0,07 0,69 0,78 2,14 0,46 1,84 0,56 1,75	0,85 2,38 0,30 1,39 -0,02 1,41 0,25 1,44
TRIMMED MEAN 5%	1^{O} 2^{O} 3^{O} 4^{O}	0,23 1,40 1,12 1,12	1,57 0,77 0,55 0,77
ΔΙΑΜΕΣΟΣ	1^{O} 2^{O} 3^{O} 4^{O}	0,00 2,00 1,00 1,00	2,00 1,00 0,00 1,00
ΔΙΑΚΥΜΑΝΣΗ	1^{O} 2^{O} 3^{O} 4^{O}	0,397 1,269 1,308 0,974	1,59 0,808 1,397 0,974
ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ	1^{O} 2^{O} 3^{O} 4^{O}	0,63 1,127 0,987 0,987	1,261 0,899 1,144 0,987
ΕΛΑΧΙΣΤΟ	1^{O} 2^{O} 3^{O} 4^{O}	0 0 0 0	0 0 0 0
	1^{O}	2	4

ΜΕΓΙΣΤΟ	2^{O} 3^{O} 4^{O}	4 3 3	3 4 3
ΕΥΡΟΣ	1^{O} 2^{O} 3^{O} 4^{O}	2 4 3 3	4 3 4 3
ΑΣΥΜΜΕΤΡΙΑ	1^{O} 2^{O} 3^{O} 4^{O}	3,33 0,85 0,72 1,42	0,46 1,88 3,47 1,57
ΚΥΡΤΩΣΗ	1^{O} 2^{O} 3^{O} 4^{O}	3,12 0,72 -1,02 0,25	-0,52 1,29 4,08 0,13

Τα συμπεράσματα που προκύπτουν βάσει των τριμηνιαίων αποτελεσμάτων που εξαγάγαμε, μας βοηθούν να συγκρίνουμε σε πιο συνολικό επίπεδο τις δύο τράπεζες, ώστε να καταλήξουμε σε ποια εκ των δύο παρουσιάστηκε η μεγαλύτερη ζήτηση δανείων ανά κατηγορία.
Εκ νέου, λοιπόν και πιο συνοπτικά, θα μπορούσαμε να πούμε ότι η ΕΘΝΙΚΗ ΤΡΑΠΕΖΑ κατέχει ένα υψηλότερο ποσοστό αιτήσεων από την ΑΓΡΟΤΙΚΗ ΤΡΑΠΕΖΑ ειδικότερα στα καταναλωτικά δάνεια και πιο συγκεκριμένα το 3^{o} τρίμηνο με 10,31 δάνεια την εβδομάδα, καθώς επίσης και τα κεφάλαια κίνησης, αφού και στα τέσσερα τρίμηνα του έτους συγκέντρωσε περισσότερες αιτήσεις.
Αντίθετα, η Α.Τ.Ε. προπορεύθηκε σε μεγάλο ποσοστό στα στεγαστικά συγκεντρώνοντας το 3^{o} τρίμηνο 5,23 δάνεια την εβδομάδα σε αντίθεση με τα 2,85 δάνεια της Ε.Τ.Ε. και το 4^{o} 2,08 σε σχέση με τα 1,62 της αντίστοιχης. Επιπλέον, η Α.Τ.Ε. υπερίσχυσε και στα επαγγελματικά δάνεια έχοντας πάνω από ένα αίτημα την εβδομάδα στα τρία τρίμηνα του έτους, εν αντιθέσει με την Ε.Τ.Ε. η οποία στις περισσότερες εβδομάδες τους έτους παρουσίασε σχεδόν καμία αίτηση.
Αναλύοντας τη διάμεσο, δίνουμε έμφαση στο 3^{o} τρίμηνο για ό,τι αφορά τα καταναλωτικά δάνεια όπου το 50% των αιτημάτων της Α.Τ.Ε. ήταν πάνω από 9 την εβδομάδα και αντίστοιχα για την ΕΘΝΙΚΗ πάνω από 11. Στα στεγαστικά ιδιαίτερα στο 3^{o} τρίμηνο η ΑΓΡΟΤΙΚΗ είχε το 50% πάνω από 5, ενώ η Ε.Τ.Ε. κινήθηκε στα υψηλότερα επίπεδα μεταξύ $1^{ου}$ και $2^{ου}$ τριμήνου έχοντας το 50% των αιτήσεων της πάνω από 4. Αναφορικά, αξιόλογη παρατήρηση σε ότι αφορά τα κεφάλαια κίνησης, μπορούμε να τονίσουμε ειδικότερα το 2^{o} τρίμηνο στην ΕΘΝΙΚΗ όπου το ήμισυ των δανείων που αιτήθηκαν ήταν πάνω από 3.
Αναφερόμενοι, τέλος, στις κατανομές των παρατηρήσεων και χωριστά για κάθε προϊόν καταλήγουμε:
<u>Με βάση το συντελεστή ασυμμετρίας:</u>
τα **καταναλωτικά δάνεια** και στις δύο τράπεζες τα δεδομένα παρουσίασαν συμμετρική κατανομή, εκτός από το 1^{o} και 4^{o} τρίμηνο όπου στην ΕΘΝΙΚΗ είχαμε θετική συμμετρία,
τα **στεγαστικά δάνεια** επίσης συμμετρική κατανομή,

τα **κεφάλαια κίνησης** συμμετρική κατανομή, εκτός του 1ου τριμήνου όπου στην ΑΓΡΟΤΙΚΗ παρουσίασαν θετική συμμετρία και τέλος
τα **επαγγελματικά δάνεια** συμμετρική κατανομή, εκτός από το 1ο τρίμηνο όπου στην Α.Τ.Ε. έχουμε θετική συμμετρία, καθώς επίσης και στο 3ο τρίμηνο για τα δάνεια της ΕΘΝΙΚΗΣ.

<u>Με βάση το συντελεστή κύρτωσης:</u>
τα **καταναλωτικά δάνεια** και στις δύο τράπεζες τα δεδομένα παρουσίασαν μεσόκυρτη κατανομή, εκτός από το 1ο τρίμηνο όπου στην ΕΘΝΙΚΗ είχαμε λεπτόκυρτη κατανομή,
τα **στεγαστικά δάνεια** επίσης μεσόκυρτη κατανομή,
τα **κεφάλαια κίνησης** μεσόκυρτη κατανομή και τέλος
τα **επαγγελματικά δάνεια** κανονική κατανομή, εκτός από το 1ο τρίμηνο όπου στην Α.Τ.Ε. έχουμε λεπτόκυρτη κατανομή, καθώς επίσης και στο 3ο τρίμηνο για τα δάνεια της ΕΘΝΙΚΗΣ.

Θέλοντας να κάνουμε και μία αναφορά στα εκατοστημόρια των παρατηρήσεων, οι πίνακες των οποίων παρατίθενται στο κεφάλαιο των παραρτημάτων της εργασίας,

– Για τα καταναλωτικά δάνεια:

διαπιστώνουμε πως στο 3ο τρίμηνο παρουσιάζονται οι υψηλότερες αιτήσεις, όπου το 90% των παρατηρήσεων ήταν πάνω από 16,8 αιτήσεις στην Α.Τ.Ε., ενώ το υπόλοιπο 10% κάτω από αυτό το όριο, ενώ στην ΕΘΝΙΚΗ το 90% ήταν 14,6 αιτήσεις, ενώ το υπόλοιπο 10% κάτω από αυτές.

– Για τα στεγαστικά δάνεια:

στο 3ο τρίμηνο για την ΑΓΡΟΤΙΚΗ το 90% των αιτήσεων ήταν πάνω από 10,20, ενώ στην ΕΘΝΙΚΗ το 90% των αιτήσεων ήταν πάνω από 6.

– Για τα κεφάλαια κίνησης:

στο 4ο τρίμηνο το 75% των αιτήσεων ήταν μία στην ΑΓΡΟΤΙΚΗ, ενώ το 25% καμία αίτηση, ενώ στην ΕΘΝΙΚΗ το 75% ήταν πάνω από 2,50 και το υπόλοιπο 25% κάτω από το όριο αυτό.

– Για τα επαγγελματικά δάνεια:

στο 2ο τρίμηνο το 50% των αιτήσεων της Α.Τ.Ε. ήταν 2 αιτήσεις, ενώ στην Ε.Τ.Ε. το 50% ήταν μία αίτηση και το υπόλοιπο 50% κάτω από το όριο δηλαδή καμία αίτηση.

ΠΑΡΑΡΤΗΜΑ

Statistics

ΕΠΑΓΓΕΛΜΑΤΙΚΑ1

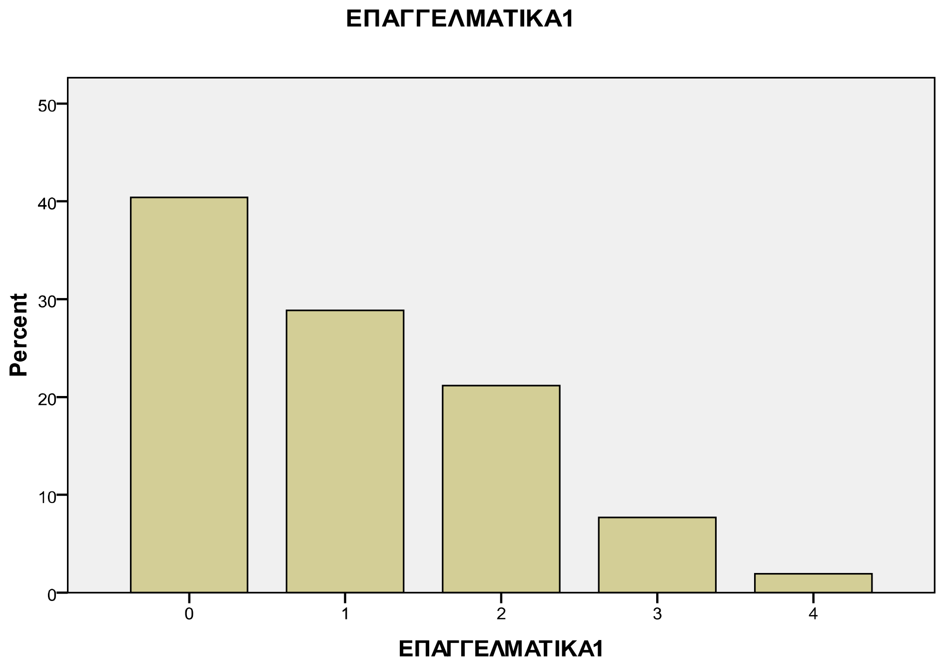
N	Valid	52
	Missing	0

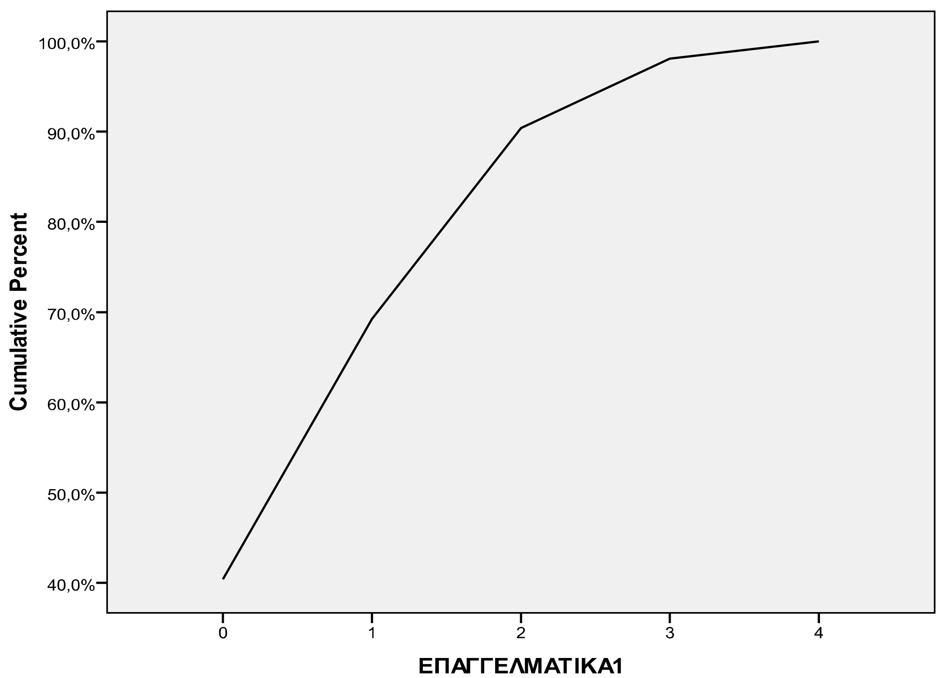
	Range	4
	Minimum	0
	Maximum	4

ΕΠΑΓΓΕΛΜΑΤΙΚΑ1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	21	40,4	40,4	40,4
	1	15	28,8	28,8	69,2
	2	11	21,2	21,2	90,4
	3	4	7,7	7,7	98,1
	4	1	1,9	1,9	100,0
	Total	52	100,0	100,0	

ΠΙΝΑΚΑΣ ΚΑΤΑΝΟΜΗΣ ΣΥΧΝΟΤΗΤΩΝ ΕΠΑΓΓΕΛΜΑΤΙΚΩΝ 1 ΔΑΝΕΙΩΝ.





ΔΙΑΓΡΑΜΜΑΤΑ ΣΥΧΝΟΤΗΤΩΝ ΕΠΑΓΓΕΛΜΑΤΙΚΩΝ 1 ΔΑΝΕΙΩΝ.

Statistics

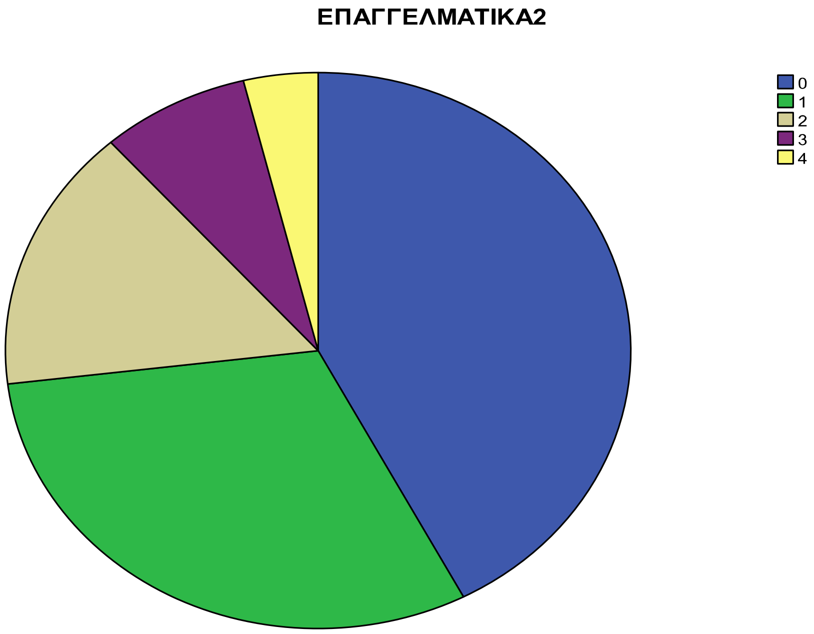
ΕΠΑΓΓΕΛΜΑΤΙΚΑ2

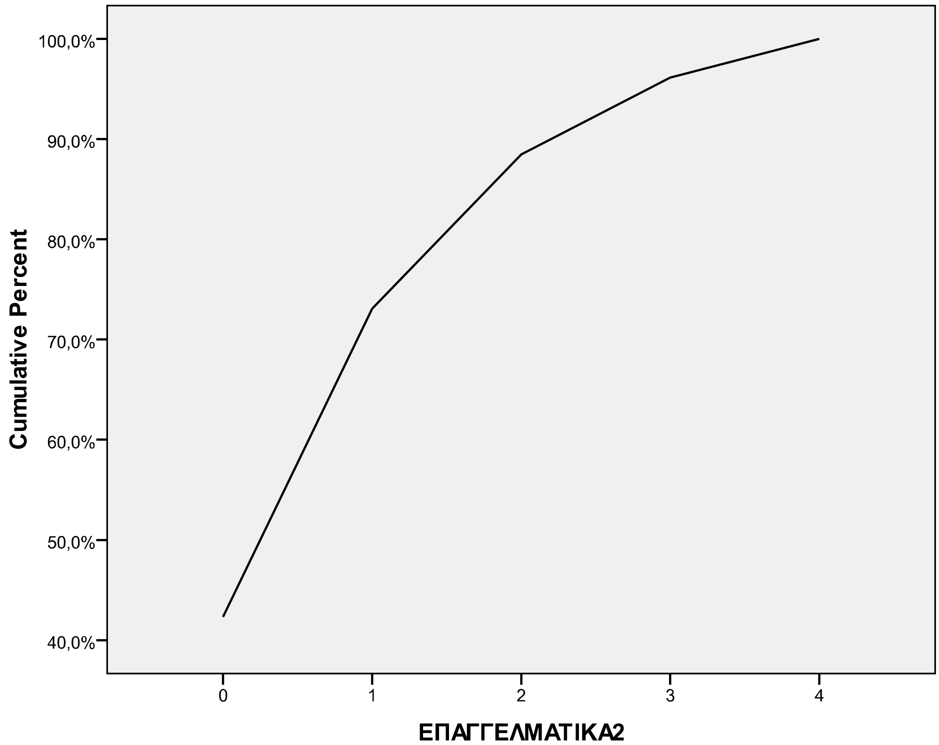
N	Valid	52
	Missing	0
	Range	4
	Minimum	0
	Maximum	4

ΕΠΑΓΓΕΛΜΑΤΙΚΑ2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	22	42,3	42,3	42,3
	1	16	30,8	30,8	73,1
	2	8	15,4	15,4	88,5
	3	4	7,7	7,7	96,2
	4	2	3,8	3,8	100,0
	Total	52	100,0	100,0	

ΠΙΝΑΚΑΣ ΚΑΤΑΝΟΜΗΣ ΣΥΧΝΟΤΗΤΩΝ ΕΠΑΓΓΕΛΜΑΤΙΚΩΝ 2 ΔΑΝΕΙΩΝ.





ΔΙΑΓΡΑΜΜΑΤΑ ΣΥΧΝΟΤΗΤΩΝ ΕΠΑΓΓΕΛΜΑΤΙΚΩΝ 2 ΔΑΝΕΙΩΝ.

Statistics

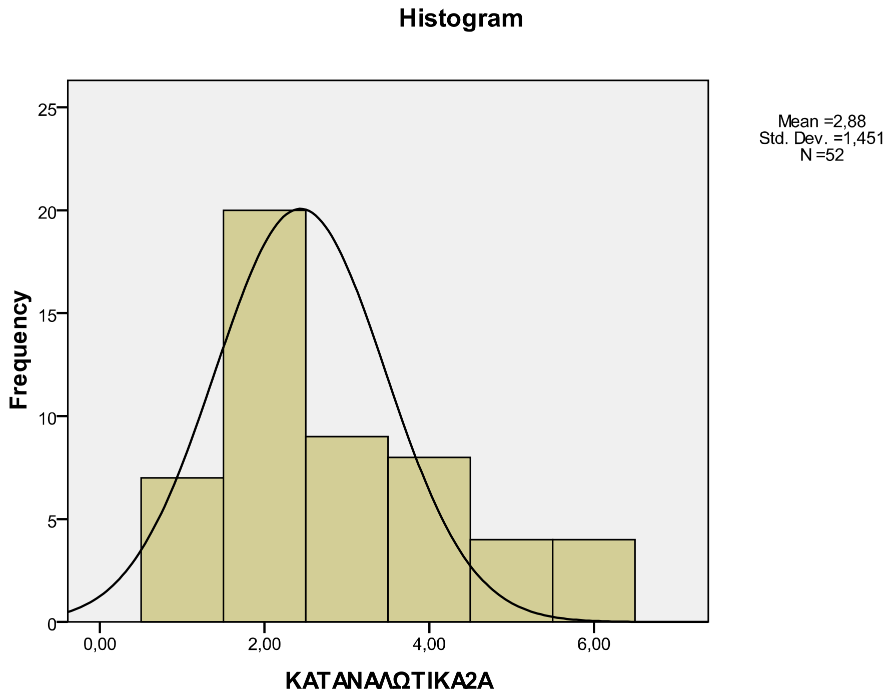
ΚΑΤΑΝΑΛΩΤΙΚΑ2

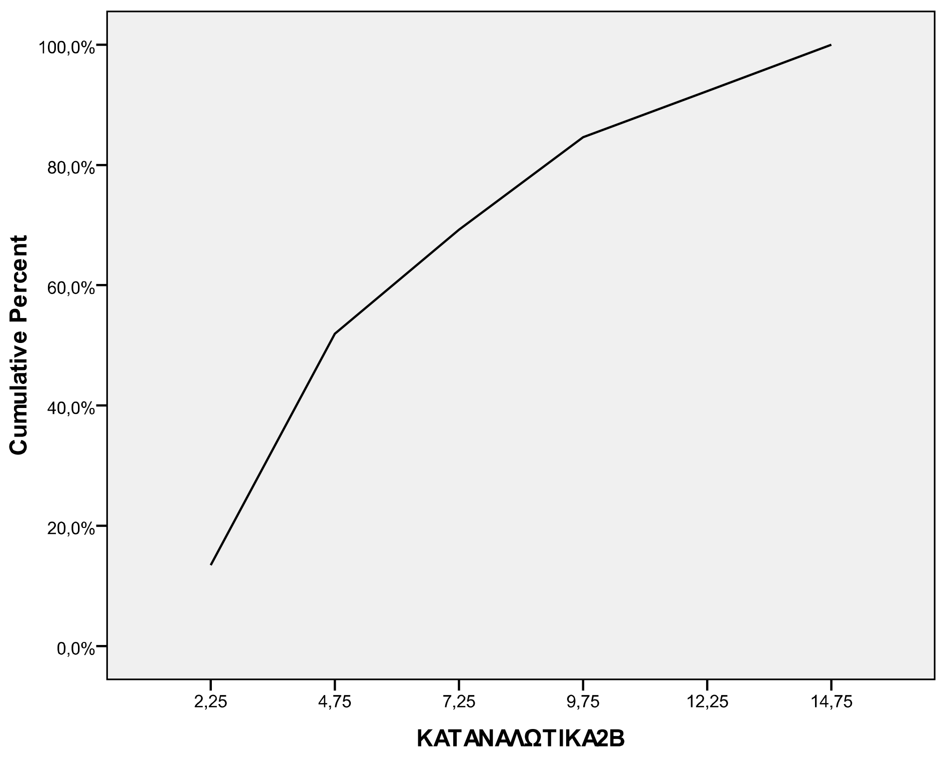
N	Valid	52
	Missing	0
	Range	15
	Minimum	1
	Maximum	16

ΚΑΤΑΝΑΛΩΤΙΚΑ2Α

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	7	13,5	13,5	13,5
	2,00	20	38,5	38,5	51,9
	3,00	9	17,3	17,3	69,2
	4,00	8	15,4	15,4	84,6
	5,00	4	7,7	7,7	92,3
	6,00	4	7,7	7,7	100,0
	Total	52	100,0	100,0	

ΠΙΝΑΚΑΣ ΚΑΤΑΝΟΜΗΣ ΣΥΧΝΟΤΗΤΩΝ ΚΑΤΑΝΑΛΩΤΙΚΩΝ 1 ΔΑΝΕΙΩΝ.





ΔΙΑΓΡΑΜΜΑΤΑ ΣΥΧΝΟΤΗΤΩΝ ΚΑΤΑΝΑΛΩΤΙΚΩΝ 2 ΔΑΝΕΙΩΝ.

Statistics

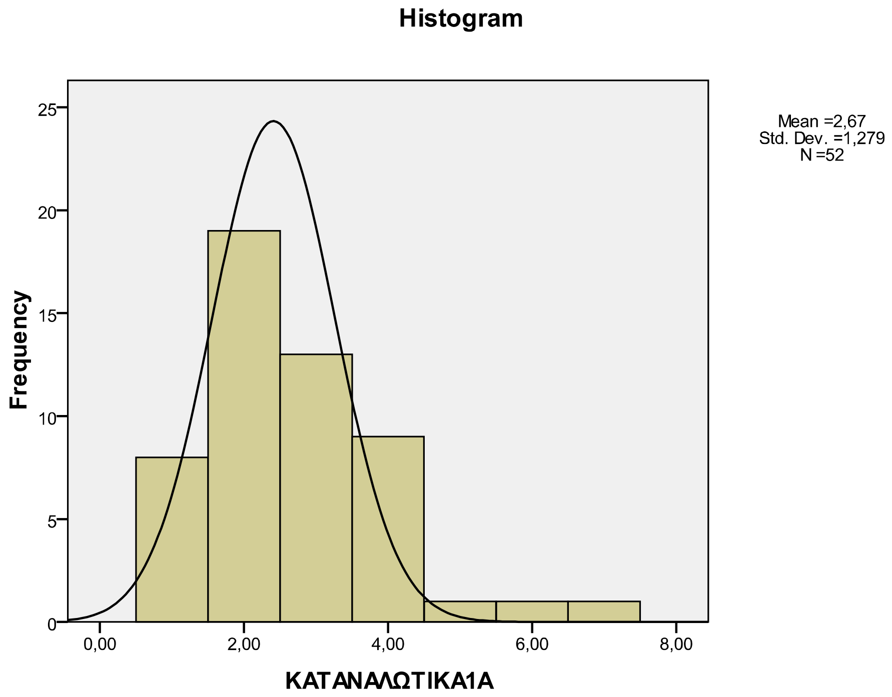
ΚΑΤΑΝΑΛΩΤΙΚΑ1

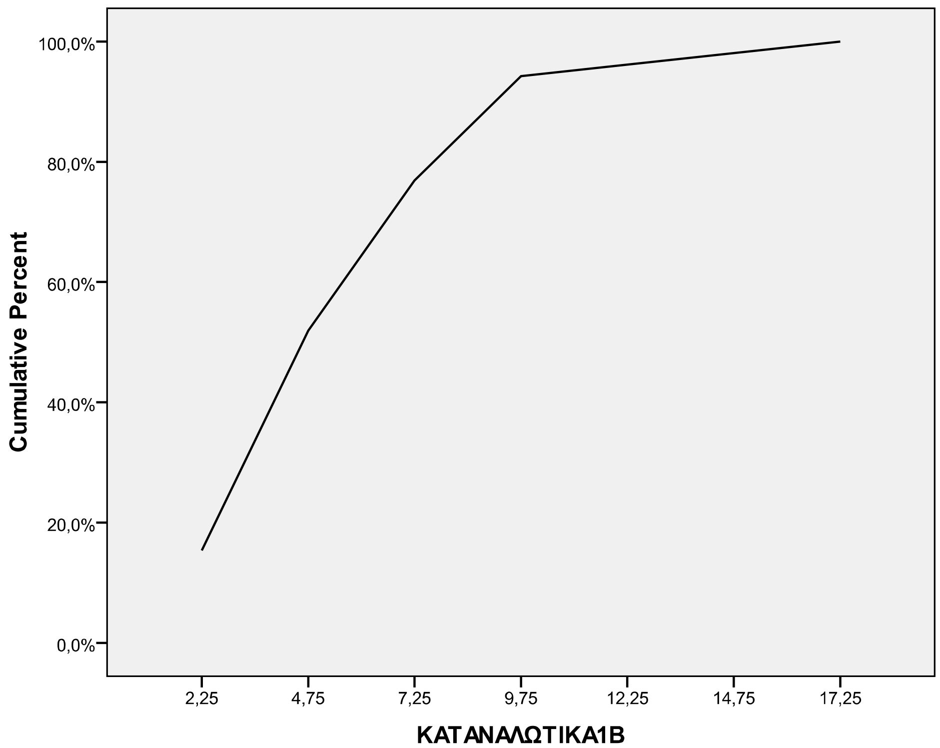
N	Valid	52
	Missing	0
	Range	17
	Minimum	1
	Maximum	18

ΚΑΤΑΝΑΛΩΤΙΚΑ1Α

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	8	15,4	15,4	15,4
	2,00	19	36,5	36,5	51,9
	3,00	13	25,0	25,0	76,9
	4,00	9	17,3	17,3	94,2
	5,00	1	1,9	1,9	96,2
	6,00	1	1,9	1,9	98,1
	7,00	1	1,9	1,9	100,0
	Total	52	100,0	100,0	

ΠΙΝΑΚΑΣ ΚΑΤΑΝΟΜΗΣ ΣΥΧΝΟΤΗΤΩΝ ΚΑΤΑΝΑΛΩΤΙΚΩΝ 1 ΔΑΝΕΙΩΝ.





ΔΙΑΓΡΑΜΜΑΤΑ ΣΥΧΝΟΤΗΤΩΝ ΚΑΤΑΝΑΛΩΤΙΚΩΝ 1 ΔΑΝΕΙΩΝ.

Statistics

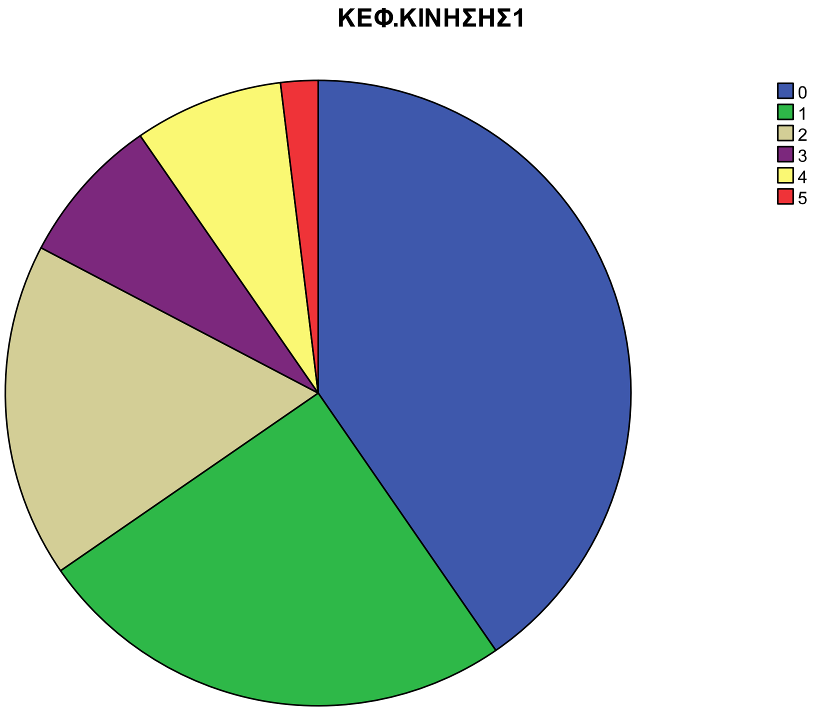
ΚΕΦ.ΚΙΝΗΣΗΣ1

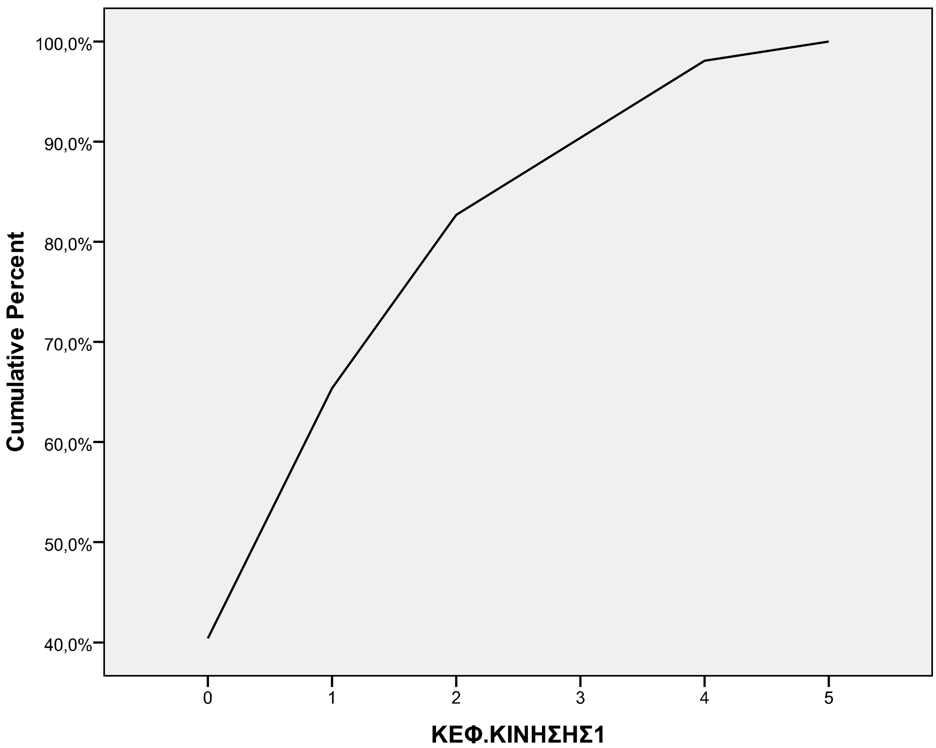
N	Valid	52
	Missing	0
	Range	5
	Minimum	0
	Maximum	5

ΚΕΦ.ΚΙΝΗΣΗΣ1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	21	40,4	40,4	40,4
	1	13	25,0	25,0	65,4
	2	9	17,3	17,3	82,7
	3	4	7,7	7,7	90,4
	4	4	7,7	7,7	98,1
	5	1	1,9	1,9	100,0
	Total	52	100,0	100,0	

ΠΙΝΑΚΕΣ ΚΑΤΑΝΟΜΗΣ ΣΥΧΝΟΤΗΤΩΝ ΚΕΦΑΛΑΙΩΝ ΚΙΝΗΣΗΣ 1.





ΔΙΑΓΡΑΜΜΑΤΑ ΣΥΧΝΟΤΗΤΩΝ ΚΕΦΑΛΑΙΩΝ ΚΙΝΗΣΗΣ 1.

Statistics

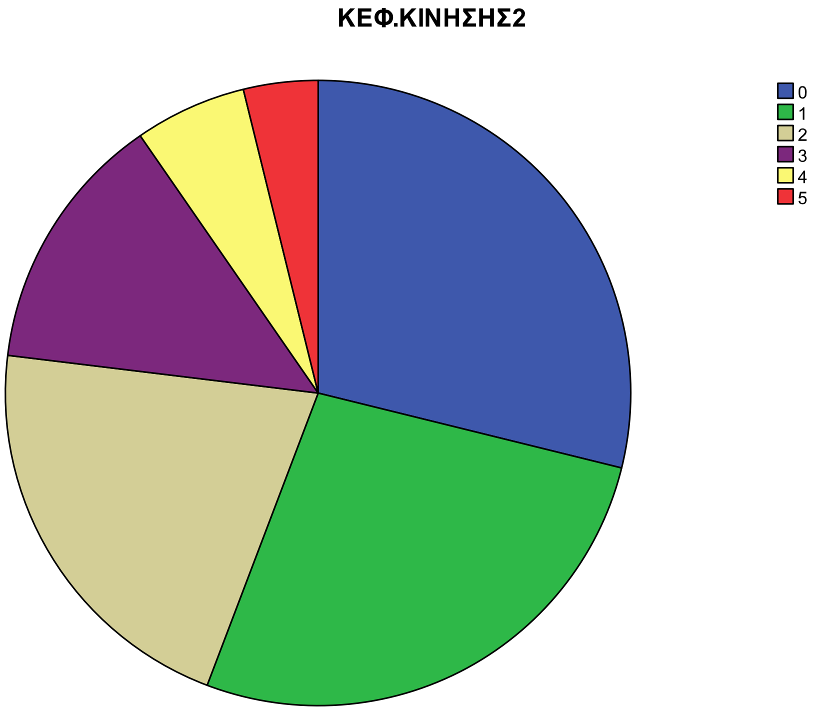
ΚΕΦ.ΚΙΝΗΣΗΣ2

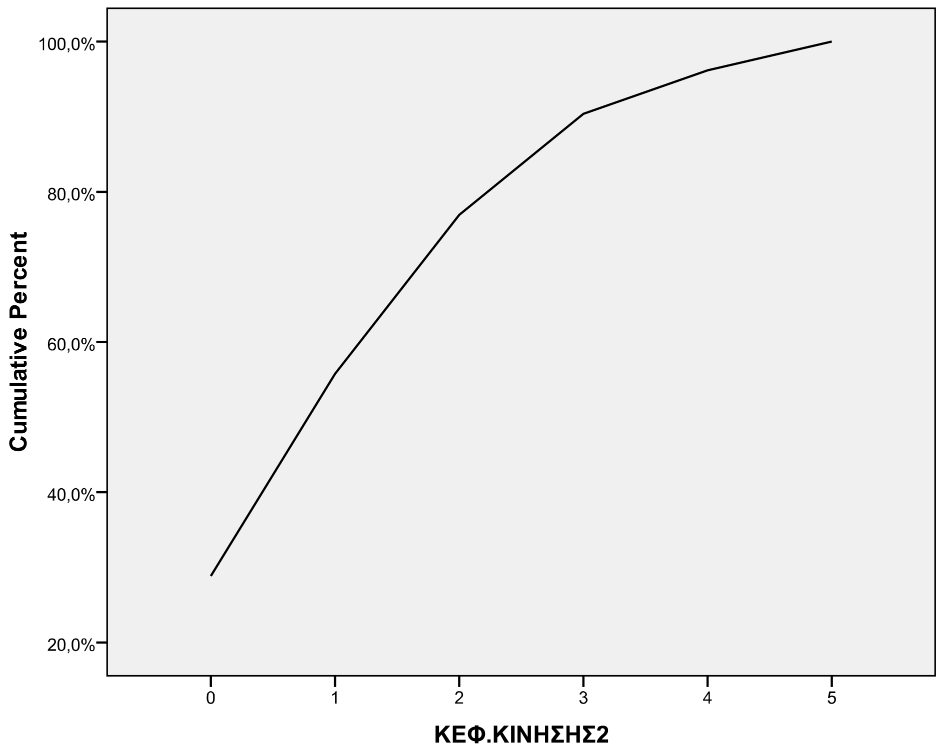
N	Valid	52
	Missing	0
	Range	5
	Minimum	0
	Maximum	5

ΚΕΦ.ΚΙΝΗΣΗΣ2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	15	28,8	28,8	28,8
	1	14	26,9	26,9	55,8
	2	11	21,2	21,2	76,9
	3	7	13,5	13,5	90,4
	4	3	5,8	5,8	96,2
	5	2	3,8	3,8	100,0
	Total	52	100,0	100,0	

ΠΙΝΑΚΕΣ ΚΑΤΑΝΟΜΗΣ ΣΥΧΝΟΤΗΤΩΝ ΚΕΦΑΛΑΙΩΝ ΚΙΝΗΣΗΣ 2.





ΔΙΑΓΡΑΜΜΑΤΑ ΣΥΧΝΟΤΗΤΩΝ ΚΕΦΑΛΑΙΩΝ ΚΙΝΗΣΗΣ 2.

Statistics

ΣΤΕΓΑΣΤΙΚΑ1

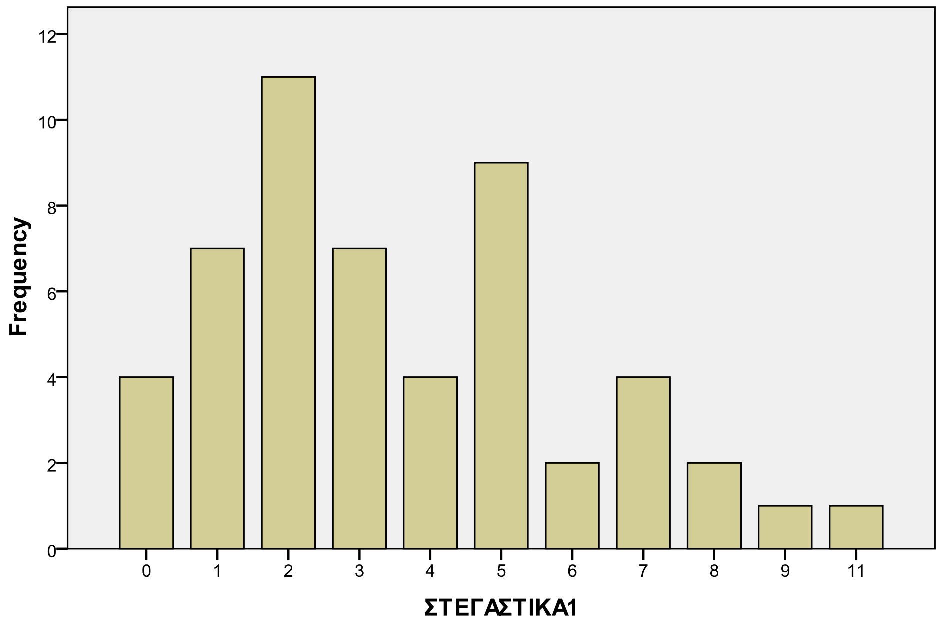
N	Valid	52
	Missing	0
	Range	11
	Minimum	0
	Maximum	11

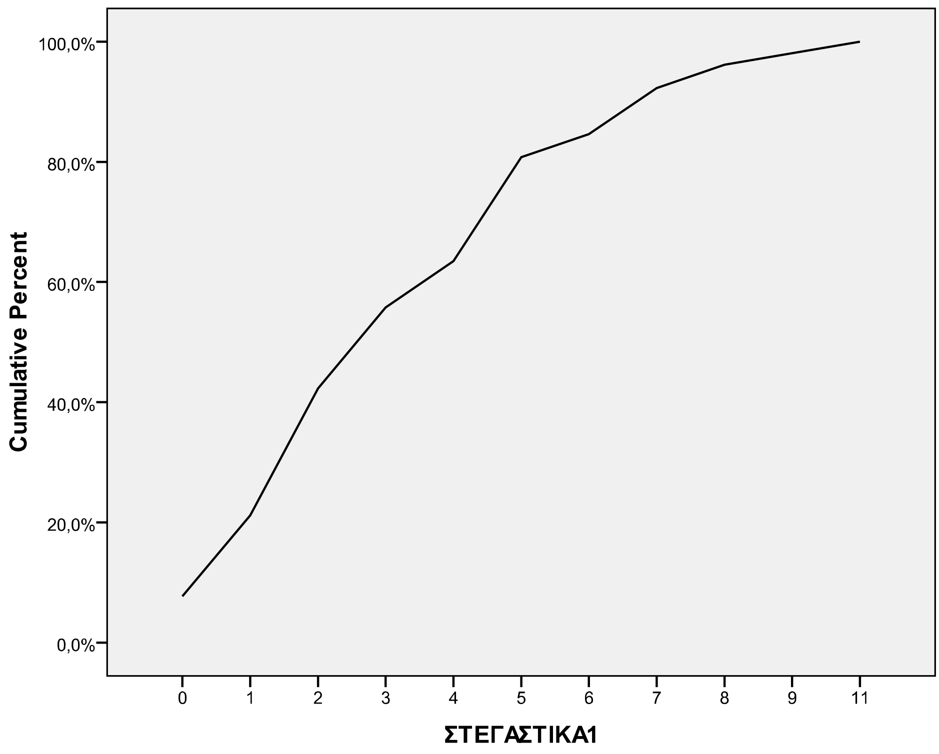
ΣΤΕΓΑΣΤΙΚΑ1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	4	7,7	7,7	7,7
	1	7	13,5	13,5	21,2
	2	11	21,2	21,2	42,3
	3	7	13,5	13,5	55,8
	4	4	7,7	7,7	63,5
	5	9	17,3	17,3	80,8
	6	2	3,8	3,8	84,6
	7	4	7,7	7,7	92,3
	8	2	3,8	3,8	96,2
	9	1	1,9	1,9	98,1
	11	1	1,9	1,9	100,0
	Total	52	100,0	100,0	

ΠΙΝΑΚΕΣ ΚΑΤΑΝΟΜΗΣ ΣΥΧΝΟΤΗΤΩΝ ΣΤΑΓΑΣΤΙΚΩΝ 1 ΔΑΝΕΙΩΝ.

ΣΤΕΓΑΣΤΙΚΑ1





ΔΙΑΓΡΑΜΜΑΤΑ ΣΥΧΝΟΤΗΤΩΝ ΣΤΕΓΑΣΤΙΚΩΝ 1 ΔΑΝΕΙΩΝ.

Statistics

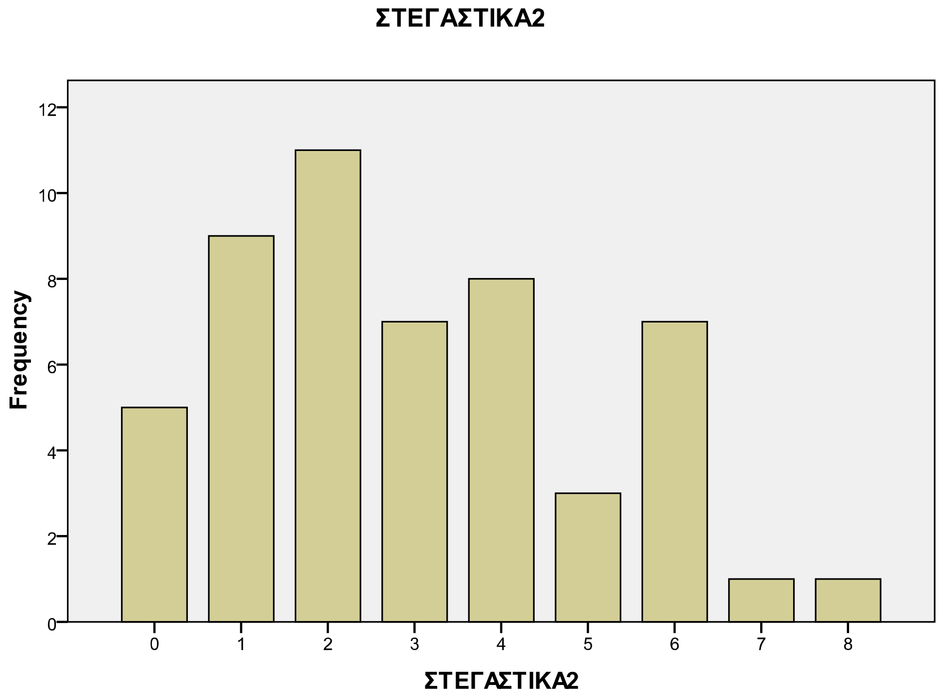
ΣΤΕΓΑΣΤΙΚΑ2

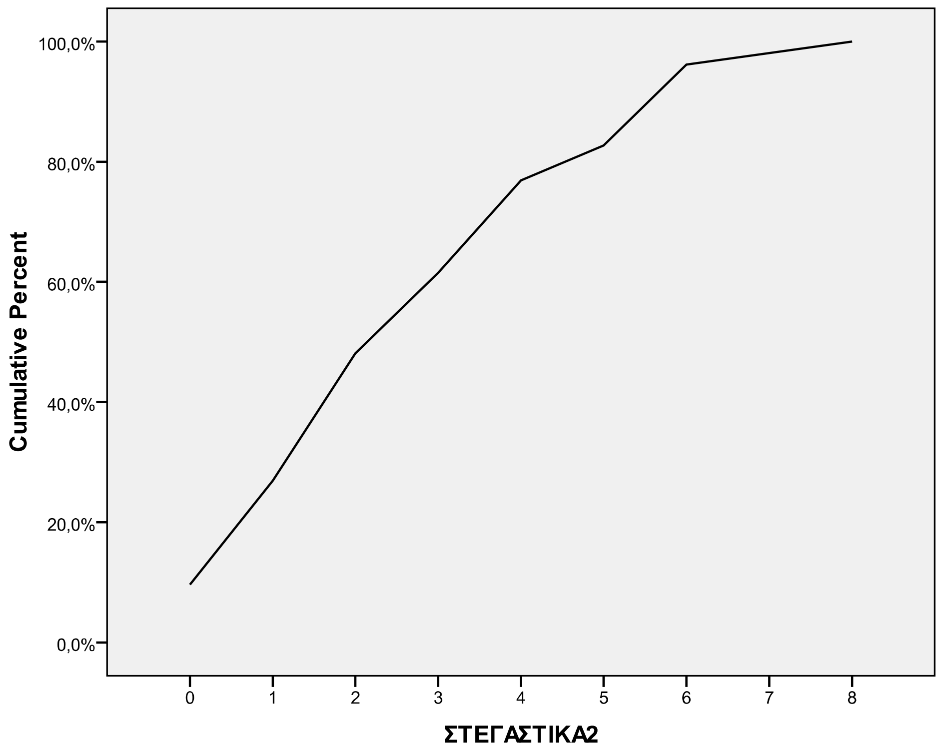
N	Valid	52
	Missing	0
	Range	8
	Minimum	0
	Maximum	8

ΣΤΕΓΑΣΤΙΚΑ2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	5	9,6	9,6	9,6
	1	9	17,3	17,3	26,9
	2	11	21,2	21,2	48,1
	3	7	13,5	13,5	61,5
	4	8	15,4	15,4	76,9
	5	3	5,8	5,8	82,7
	6	7	13,5	13,5	96,2
	7	1	1,9	1,9	98,1
	8	1	1,9	1,9	100,0
	Total	52	100,0	100,0	

ΠΙΝΑΚΕΣ ΚΑΤΑΝΟΜΗΣ ΣΥΧΝΟΤΗΤΩΝ ΣΤΕΓΑΣΤΙΚΩΝ 2 ΔΑΝΕΙΩΝ.





ΔΙΑΓΡΑΜΜΑΤΑ ΣΥΧΝΟΤΗΤΩΝ ΣΤΕΓΑΣΤΙΚΩΝ 2 ΔΑΝΕΙΩΝ.

Descriptives

			Statistic	Std. Error
ΚΑΤΑΝΑΛΩΤΙΚΑ1	Mean		6,73	,430
	95% Confidence Interval for Mean	Lower Bound	5,87	
		Upper Bound	7,59	
	5% Trimmed Mean		6,52	
	Median		6,00	
	Variance		9,612	
	Std. Deviation		3,100	
	Minimum		1	
	Maximum		18	
	Range		17	
	Interquartile Range		3	
	Skewness		1,137	,330
	Kurtosis		2,858	,650
ΚΑΤΑΝΑΛΩΤΙΚΑ2	Mean		7,23	,504
	95% Confidence Interval for Mean	Lower Bound	6,22	
		Upper Bound	8,24	
	5% Trimmed Mean		7,10	
	Median		6,00	
	Variance		13,201	
	Std. Deviation		3,633	
	Minimum		1	
	Maximum		16	
	Range		15	
	Interquartile Range		5	
	Skewness		,603	,330
	Kurtosis		-,367	,650
ΣΤΕΓΑΣΤΙΚΑ1	Mean		3,60	,352
	95% Confidence Interval for Mean	Lower Bound	2,89	
		Upper Bound	4,30	

	5% Trimmed Mean		3,47	
	Median		3,00	
	Variance		6,442	
	Std. Deviation		2,538	
	Minimum		0	
	Maximum		11	
	Range		11	
	Interquartile Range		3	
	Skewness		,736	,330
	Kurtosis		,168	,650
ΣΤΕΓΑΣΤΙΚΑ2	Mean		3,00	,285
	95% Confidence Interval for Mean	Lower Bound	2,43	
		Upper Bound	3,57	
	5% Trimmed Mean		2,94	
	Median		3,00	
	Variance		4,235	
	Std. Deviation		2,058	
	Minimum		0	
	Maximum		8	
	Range		8	
	Interquartile Range		3	
	Skewness		,449	,330
	Kurtosis		-,647	,650
ΚΕΦ.ΚΙΝΗΣΗΣ1	Mean		1,23	,189
	95% Confidence Interval for Mean	Lower Bound	,85	
		Upper Bound	1,61	
	5% Trimmed Mean		1,12	
	Median		1,00	
	Variance		1,867	
	Std. Deviation		1,366	
	Minimum		0	
	Maximum		5	

	Range		5	
	Interquartile Range		2	
	Skewness		1,002	,330
	Kurtosis		,153	,650
ΚΕΦ.ΚΙΝΗΣΗΣ2	Mean		1,52	,193
	95% Confidence Interval for Mean	Lower Bound	1,13	
		Upper Bound	1,91	
	5% Trimmed Mean		1,42	
	Median		1,00	
	Variance		1,941	
	Std. Deviation		1,393	
	Minimum		0	
	Maximum		5	
	Range		5	
	Interquartile Range		2	
	Skewness		,751	,330
	Kurtosis		-,126	,650
ΕΠΑΓΓΕΛΜΑΤΙΚΑ1	Mean		1,02	,147
	95% Confidence Interval for Mean	Lower Bound	,72	
		Upper Bound	1,31	
	5% Trimmed Mean		,94	
	Median		1,00	
	Variance		1,117	
	Std. Deviation		1,057	
	Minimum		0	
	Maximum		4	
	Range		4	
	Interquartile Range		2	
	Skewness		,789	,330
	Kurtosis		-,122	,650
ΕΠΑΓΓΕΛΜΑΤΙΚΑ2	Mean		1,00	,155
	95% Confidence Interval for Mean	Lower Bound	,69	

		Upper Bound	1,31	
	5% Trimmed Mean		,90	
	Median		1,00	
	Variance		1,255	
	Std. Deviation		1,120	
	Minimum		0	
	Maximum		4	
	Range		4	
	Interquartile Range		2	
	Skewness		1,044	,330
	Kurtosis		,401	,650

ΠΙΝΑΚΑΣ ΜΕΤΡΩΝ ΘΕΣΗΣ ΜΕΤΑΒΛΗΤΩΝ.

Percentiles

		Percentiles						
		5	10	25	50	75	90	95
Weighted Average(Definition 1)	ΚΑΤΑΝΑΛΩΤΙΚΑ 1	2,65	3,00	5,00	6,00	8,00	10,00	13,05
	ΚΑΤΑΝΑΛΩΤΙΚΑ 2	2,00	3,00	5,00	6,00	10,00	12,70	14,35
	ΣΤΕΓΑΣΤΙΚΑ1	,00	1,00	2,00	3,00	5,00	7,00	8,35
	ΣΤΕΓΑΣΤΙΚΑ2	,00	,30	1,00	3,00	4,00	6,00	6,35
	ΚΕΦ.ΚΙΝΗΣΗΣ1	,00	,00	,00	1,00	2,00	3,70	4,00
	ΚΕΦ.ΚΙΝΗΣΗΣ2	,00	,00	,00	1,00	2,00	3,70	4,35
	ΕΠΑΓΓΕΛΜΑΤΙΚΑ1	,00	,00	,00	1,00	2,00	2,70	3,00
	ΕΠΑΓΓΕΛΜΑΤΙΚΑ2	,00	,00	,00	1,00	2,00	3,00	3,35
Tukey's Hinges	ΚΑΤΑΝΑΛΩΤΙΚΑ 1			5,00	6,00	8,00		
	ΚΑΤΑΝΑΛΩΤΙΚΑ 2			5,00	6,00	10,00		

	ΣΤΕΓΑΣΤΙΚΑ1			2,00	3,00	5,00		
	ΣΤΕΓΑΣΤΙΚΑ2			1,00	3,00	4,00		
	ΚΕΦ.ΚΙΝΗΣΗΣ1			,00	1,00	2,00		
	ΚΕΦ.ΚΙΝΗΣΗΣ2			,00	1,00	2,00		
	ΕΠΑΓΓΕΛΜΑΤΙΚΑ1			,00	1,00	2,00		
	ΕΠΑΓΓΕΛΜΑΤΙΚΑ2			,00	1,00	2,00		

ΠΙΝΑΚΑΣ ΕΚΑΤΟΣΤΗΜΟΡΙΩΝ ΜΕΤΑΒΛΗΤΩΝ.

Descriptives

			Statistic	Std. Error
ΚΑΤΑΝΑΛΩΤΙΚΑ1	Mean		4,85	,576
	95% Confidence Interval for Mean	Lower Bound	3,59	
		Upper Bound	6,10	
	5% Trimmed Mean		4,77	
	Median		5,00	
	Variance		4,308	
	Std. Deviation		2,075	
	Minimum		2	
	Maximum		9	
	Range		7	
	Interquartile Range		4	
	Skewness		,506	,616
	Kurtosis		-,532	1,191
ΚΑΤΑΝΑΛΩΤΙΚΑ2	Mean		5,31	,711
	95% Confidence Interval for Mean	Lower Bound	3,76	
		Upper Bound	6,86	
	5% Trimmed Mean		5,12	
	Median		5,00	
	Variance		6,564	

	Std. Deviation		2,562	
	Minimum		2	
	Maximum		12	
	Range		10	
	Interquartile Range		3	
	Skewness		1,493	,616
	Kurtosis		3,245	1,191
ΣΤΕΓΑΣΤΙΚΑ1	Mean		3,62	,583
	95% Confidence Interval for Mean	Lower Bound	2,34	
		Upper Bound	4,89	
	5% Trimmed Mean		3,52	
	Median		3,00	
	Variance		4,423	
	Std. Deviation		2,103	
	Minimum		1	
	Maximum		8	
	Range		7	
	Interquartile Range		3	
	Skewness		,603	,616
	Kurtosis		-,211	1,191
ΣΤΕΓΑΣΤΙΚΑ2	Mean		3,62	,703
	95% Confidence Interval for Mean	Lower Bound	2,08	
		Upper Bound	5,15	
	5% Trimmed Mean		3,57	
	Median		4,00	
	Variance		6,423	
	Std. Deviation		2,534	
	Minimum		0	
	Maximum		8	
	Range		8	
	Interquartile Range		5	
	Skewness		,040	,616

	Kurtosis		-1,011	1,191
ΚΕΦ.ΚΙΝΗΣΗΣ1	Mean		,92	,383
	95% Confidence Interval for Mean	Lower Bound	,09	
		Upper Bound	1,76	
	5% Trimmed Mean		,80	
	Median		,00	
	Variance		1,910	
	Std. Deviation		1,382	
	Minimum		0	
	Maximum		4	
	Range		4	
	Interquartile Range		2	
	Skewness		1,279	,616
	Kurtosis		,484	1,191
ΚΕΦ.ΚΙΝΗΣΗΣ2	Mean		,92	,265
	95% Confidence Interval for Mean	Lower Bound	,35	
		Upper Bound	1,50	
	5% Trimmed Mean		,86	
	Median		1,00	
	Variance		,910	
	Std. Deviation		,954	
	Minimum		0	
	Maximum		3	
	Range		3	
	Interquartile Range		2	
	Skewness		,854	,616
	Kurtosis		,221	1,191
ΕΠΑΓΓΕΛΜΑΤΙΚΑ1	Mean		,31	,175
	95% Confidence Interval for Mean	Lower Bound	-,07	
		Upper Bound	,69	
	5% Trimmed Mean		,23	
	Median		,00	

	Variance		,397	
	Std. Deviation		,630	
	Minimum		0	
	Maximum		2	
	Range		2	
	Interquartile Range		0	
	Skewness		2,051	,616
	Kurtosis		3,711	1,191
ΕΠΑΓΓΕΛΜΑΤΙΚΑ2	Mean		1,62	,350
	95% Confidence Interval for Mean	Lower Bound	,85	
		Upper Bound	2,38	
	5% Trimmed Mean		1,57	
	Median		2,00	
	Variance		1,590	
	Std. Deviation		1,261	
	Minimum		0	
	Maximum		4	
	Range		4	
	Interquartile Range		2	
	Skewness		,283	,616
	Kurtosis		-,619	1,191

Percentiles

		Percentiles						
		5	10	25	50	75	90	95
Weighted Average(Definition 1)	ΚΑΤΑΝΑΛΩΤΙΚΑ1	2,00	2,40	3,00	5,00	6,50	8,20	.
	ΚΑΤΑΝΑΛΩΤΙΚΑ2	2,00	2,40	3,50	5,00	6,00	10,40	.
	ΣΤΕΓΑΣΤΙΚΑ1	1,00	1,00	2,00	3,00	5,00	7,20	.
	ΣΤΕΓΑΣΤΙΚΑ2	,00	,00	1,50	4,00	6,00	7,20	.
	ΚΕΦ.ΚΙΝΗΣΗΣ1	,00	,00	,00	,00	2,00	3,60	.
	ΚΕΦ.ΚΙΝΗΣΗΣ2	,00	,00	,00	1,00	1,50	2,60	.
	ΕΠΑΓΓΕΛΜΑΤΙΚΑ1	,00	,00	,00	,00	,50	1,60	.

	ΕΠΑΓΓΕΛΜΑΤΙΚΑ2	,00	,00	,50	2,00	2,50	3,60	.
Tukey's Hinges	ΚΑΤΑΝΑΛΩΤΙΚΑ1			3,00	5,00	6,00		
	ΚΑΤΑΝΑΛΩΤΙΚΑ2			4,00	5,00	6,00		
	ΣΤΕΓΑΣΤΙΚΑ1			2,00	3,00	5,00		
	ΣΤΕΓΑΣΤΙΚΑ2			2,00	4,00	6,00		
	ΚΕΦ.ΚΙΝΗΣΗΣ1			,00	,00	2,00		
	ΚΕΦ.ΚΙΝΗΣΗΣ2			,00	1,00	1,00		
	ΕΠΑΓΓΕΛΜΑΤΙΚΑ1			,00	,00	,00		
	ΕΠΑΓΓΕΛΜΑΤΙΚΑ2			1,00	2,00	2,00		

ΠΙΝΑΚΕΣ ΜΕΤΡΩΝ ΘΕΣΗΣ ΚΑΙ ΕΚΑΤΟΣΤΗΜΟΡΙΩΝ 1^{ΟΥ} ΤΡΙΜΗΝΟΥ ΕΤΟΥΣ 2008.

Descriptives

			Statistic	Std. Error
ΚΑΤΑΝΑΛΩΤΙΚΑ1	Mean		6,00	,689
	95% Confidence Interval for Mean	Lower Bound	4,50	
		Upper Bound	7,50	
	5% Trimmed Mean		6,06	
	Median		6,00	
	Variance		6,167	
	Std. Deviation		2,483	
	Minimum		1	
	Maximum		10	
	Range		9	
	Interquartile Range		4	
	Skewness		-,424	,616
	Kurtosis		,039	1,191
ΚΑΤΑΝΑΛΩΤΙΚΑ2	Mean		6,08	,964
	95% Confidence Interval for Mean	Lower Bound	3,98	
		Upper Bound	8,18	
	5% Trimmed Mean		5,97	

	Median		5,00	
	Variance		12,077	
	Std. Deviation		3,475	
	Minimum		1	
	Maximum		13	
	Range		12	
	Interquartile Range		5	
	Skewness		,557	,616
	Kurtosis		-,336	1,191
ΣΤΕΓΑΣΤΙΚΑ1	Mean		3,46	,514
	95% Confidence Interval for Mean	Lower Bound	2,34	
		Upper Bound	4,58	
	5% Trimmed Mean		3,40	
	Median		3,00	
	Variance		3,436	
	Std. Deviation		1,854	
	Minimum		1	
	Maximum		7	
	Range		6	
	Interquartile Range		3	
	Skewness		,489	,616
	Kurtosis		-,513	1,191
ΣΤΕΓΑΣΤΙΚΑ2	Mean		3,92	,525
	95% Confidence Interval for Mean	Lower Bound	2,78	
		Upper Bound	5,07	
	5% Trimmed Mean		3,91	
	Median		4,00	
	Variance		3,577	
	Std. Deviation		1,891	
	Minimum		1	
	Maximum		7	
	Range		6	

	Interquartile Range		3	
	Skewness		-,132	,616
	Kurtosis		-,795	1,191
ΚΕΦ.ΚΙΝΗΣΗΣ1	Mean		2,38	,446
	95% Confidence Interval for Mean	Lower Bound	1,41	
		Upper Bound	3,36	
	5% Trimmed Mean		2,37	
	Median		2,00	
	Variance		2,590	
	Std. Deviation		1,609	
	Minimum		0	
	Maximum		5	
	Range		5	
	Interquartile Range		3	
	Skewness		-,038	,616
	Kurtosis		-1,058	1,191
ΚΕΦ.ΚΙΝΗΣΗΣ2	Mean		2,62	,446
	95% Confidence Interval for Mean	Lower Bound	1,64	
		Upper Bound	3,59	
	5% Trimmed Mean		2,63	
	Median		3,00	
	Variance		2,590	
	Std. Deviation		1,609	
	Minimum		0	
	Maximum		5	
	Range		5	
	Interquartile Range		3	
	Skewness		,038	,616
	Kurtosis		-1,058	1,191
ΕΠΑΓΓΕΛΜΑΤΙΚΑ1	Mean		1,46	,312
	95% Confidence Interval for Mean	Lower Bound	,78	
		Upper Bound	2,14	

	5% Trimmed Mean		1,40	
	Median		2,00	
	Variance		1,269	
	Std. Deviation		1,127	
	Minimum		0	
	Maximum		4	
	Range		4	
	Interquartile Range		2	
	Skewness		,526	,616
	Kurtosis		,853	1,191
ΕΠΑΓΓΕΛΜΑΤΙΚΑ2	Mean		,85	,249
	95% Confidence Interval for Mean	Lower Bound	,30	
		Upper Bound	1,39	
	5% Trimmed Mean		,77	
	Median		1,00	
	Variance		,808	
	Std. Deviation		,899	
	Minimum		0	
	Maximum		3	
	Range		3	
	Interquartile Range		1	
	Skewness		1,156	,616
	Kurtosis		1,538	1,191

Percentiles

		Percentiles						
		5	10	25	50	75	90	
Weighted Average(Definition 1)	ΚΑΤΑΝΑΛΩΤΙΚΑ1	1,00	1,80	4,00	6,00	7,50	9,60	.
	ΚΑΤΑΝΑΛΩΤΙΚΑ2	1,00	1,40	4,00	5,00	9,00	11,80	.
	ΣΤΕΓΑΣΤΙΚΑ1	1,00	1,00	2,00	3,00	5,00	6,60	.
	ΣΤΕΓΑΣΤΙΚΑ2	1,00	1,00	2,50	4,00	5,50	6,60	.
	ΚΕΦ.ΚΙΝΗΣΗΣ1	,00	,00	1,00	2,00	4,00	4,60	.

	ΚΕΦ.ΚΙΝΗΣΗΣ2	,00	,40	1,00	3,00	4,00	5,00	.
	ΕΠΑΓΓΕΛΜΑΤΙΚΑ1	,00	,00	,50	2,00	2,00	3,20	.
	ΕΠΑΓΓΕΛΜΑΤΙΚΑ2	,00	,00	,00	1,00	1,00	2,60	.
Tukey's Hinges	ΚΑΤΑΝΑΛΩΤΙΚΑ1			4,00	6,00	7,00		
	ΚΑΤΑΝΑΛΩΤΙΚΑ2			4,00	5,00	8,00		
	ΣΤΕΓΑΣΤΙΚΑ1			2,00	3,00	5,00		
	ΣΤΕΓΑΣΤΙΚΑ2			3,00	4,00	5,00		
	ΚΕΦ.ΚΙΝΗΣΗΣ1			1,00	2,00	4,00		
	ΚΕΦ.ΚΙΝΗΣΗΣ2			1,00	3,00	4,00		
	ΕΠΑΓΓΕΛΜΑΤΙΚΑ1			1,00	2,00	2,00		
	ΕΠΑΓΓΕΛΜΑΤΙΚΑ2			,00	1,00	1,00		

ΠΙΝΑΚΕΣ ΜΕΤΡΩΝ ΘΕΣΗΣ ΚΑΙ ΕΚΑΤΟΣΤΗΜΟΡΙΩΝ 2^{ΟΥ} ΤΡΙΜΗΝΟΥ ΕΤΟΥΣ 2008.

Descriptives

			Statistic	Std. Error
ΚΑΤΑΝΑΛΩΤΙΚΑ1	Mean		9,46	1,035
	95% Confidence Interval for Mean	Lower Bound	7,21	
		Upper Bound	11,72	
	5% Trimmed Mean		9,24	
	Median		9,00	
	Variance		13,936	
	Std. Deviation		3,733	
	Minimum		5	
	Maximum		18	
	Range		13	
	Interquartile Range		5	
	Skewness		1,172	,616
	Kurtosis		1,066	1,191
ΚΑΤΑΝΑΛΩΤΙΚΑ2	Mean		10,31	,850
	95% Confidence Interval for Mean	Lower Bound	8,46	
		Upper Bound	12,16	

	5% Trimmed Mean		10,29	
	Median		11,00	
	Variance		9,397	
	Std. Deviation		3,066	
	Minimum		6	
	Maximum		15	
	Range		9	
	Interquartile Range		6	
	Skewness		,005	,616
	Kurtosis		-1,248	1,191
ΣΤΕΓΑΣΤΙΚΑ1	Mean		5,23	,921
	95% Confidence Interval for Mean	Lower Bound	3,22	
		Upper Bound	7,24	
	5% Trimmed Mean		5,20	
	Median		5,00	
	Variance		11,026	
	Std. Deviation		3,320	
	Minimum		0	
	Maximum		11	
	Range		11	
	Interquartile Range		6	
	Skewness		-,013	,616
	Kurtosis		-,880	1,191
ΣΤΕΓΑΣΤΙΚΑ2	Mean		2,85	,492
	95% Confidence Interval for Mean	Lower Bound	1,78	
		Upper Bound	3,92	
	5% Trimmed Mean		2,83	
	Median		2,00	
	Variance		3,141	
	Std. Deviation		1,772	
	Minimum		0	
	Maximum		6	

	Range	6	
	Interquartile Range	2	
	Skewness	,592	,616
	Kurtosis	,003	1,191
ΝΗΣΗΣ1	Mean	,92	,265
	95% Confidence Interval for Mean Lower Bound	,35	
	Upper Bound	1,50	
	5% Trimmed Mean	,86	
	Median	1,00	
	Variance	,910	
	Std. Deviation	,954	
	Minimum	0	
	Maximum	3	
	Range	3	
	Interquartile Range	2	
	Skewness	,854	,616
	Kurtosis	,221	1,191
ΚΕΦ.ΚΙΝΗΣΗΣ2	Mean	1,15	,274
	95% Confidence Interval for Mean Lower Bound	,56	
	Upper Bound	1,75	
	5% Trimmed Mean	1,12	
	Median	1,00	
	Variance	,974	
	Std. Deviation	,987	
	Minimum	0	
	Maximum	3	
	Range	3	
	Interquartile Range	2	
	Skewness	,262	,616
	Kurtosis	-,912	1,191
ΕΠΑΓΓΕΛΜΑΤΙΚΑ1	Mean	1,15	,317
	95% Confidence Interval for Mean Lower Bound	,46	

		Upper Bound	1,84	
	5% Trimmed Mean		1,12	
	Median		1,00	
	Variance		1,308	
	Std. Deviation		1,144	
	Minimum		0	
	Maximum		3	
	Range		3	
	Interquartile Range		2	
	Skewness		,442	,616
	Kurtosis		-1,215	1,191
ΕΠΑΓΓΕΛΜΑΤΙΚΑ2	Mean		,69	,328
	95% Confidence Interval for Mean	Lower Bound	-,02	
		Upper Bound	1,41	
	5% Trimmed Mean		,55	
	Median		,00	
	Variance		1,397	
	Std. Deviation		1,182	
	Minimum		0	
	Maximum		4	
	Range		4	
	Interquartile Range		1	
	Skewness		2,138	,616
	Kurtosis		4,862	1,191

Percentiles

		Percentiles						
		5	10	25	50	75	90	95
Weighted Average(Definition 1)	ΚΑΤΑΝΑΛΩΤΙΚΑ 1	5,00	5,40	6,50	9,00	11,50	16,80	.

	ΚΑΤΑΝΑΛΩΤΙΚΑ2	6,00	6,00	7,50	11,00	13,00	14,60	.
	ΣΤΕΓΑΣΤΙΚΑ1	,00	,40	2,00	5,00	7,50	10,20	.
	ΣΤΕΓΑΣΤΙΚΑ2	,00	,40	2,00	2,00	4,00	6,00	.
	ΚΕΦ.ΚΙΝΗΣΗΣ1	,00	,00	,00	1,00	1,50	2,60	.
	ΚΕΦ.ΚΙΝΗΣΗΣ2	,00	,00	,00	1,00	2,00	2,60	.
	ΕΠΑΓΓΕΛΜΑΤΙΚΑ1	,00	,00	,00	1,00	2,00	3,00	.
	ΕΠΑΓΓΕΛΜΑΤΙΚΑ2	,00	,00	,00	,00	1,00	3,20	.
Tukey's Hinges	ΚΑΤΑΝΑΛΩΤΙΚΑ1			7,00	9,00	11,00		
	ΚΑΤΑΝΑΛΩΤΙΚΑ2			8,00	11,00	12,00		
	ΣΤΕΓΑΣΤΙΚΑ1			2,00	5,00	7,00		
	ΣΤΕΓΑΣΤΙΚΑ2			2,00	2,00	4,00		
	ΚΕΦ.ΚΙΝΗΣΗΣ1			,00	1,00	1,00		
	ΚΕΦ.ΚΙΝΗΣΗΣ2			,00	1,00	2,00		
	ΕΠΑΓΓΕΛΜΑΤΙΚΑ1			,00	1,00	2,00		
	ΕΠΑΓΓΕΛΜΑΤΙΚΑ2			,00	,00	1,00		

ΠΙΝΑΚΕΣ ΜΕΤΡΩΝ ΘΕΣΗΣ ΚΑΙ ΕΚΑΤΟΣΤΗΜΟΡΙΩΝ 3ΟΥ ΤΡΙΜΗΝΟΥ ΕΤΟΥΣ 2008.

Descriptives

			Statistic	Std. Error
ΚΑΤΑΝΑΛΩΤΙΚΑ1	Mean		6,62	,549
	95% Confidence Interval for Mean	Lower Bound	5,42	
		Upper Bound	7,81	
	5% Trimmed Mean		6,63	
	Median		6,00	
	Variance		3,923	

	Std. Deviation		1,981	
	Minimum		3	
	Maximum		10	
	Range		7	
	Interquartile Range		3	
	Skewness		-,126	,616
	Kurtosis		-,347	1,191
ΚΑΤΑΝΑΛΩΤΙΚΑ2	Mean		7,23	,975
	95% Confidence Interval for Mean	Lower Bound	5,11	
		Upper Bound	9,36	
	5% Trimmed Mean		6,98	
	Median		7,00	
	Variance		12,359	
	Std. Deviation		3,516	
	Minimum		3	
	Maximum		16	
	Range		13	
	Interquartile Range		4	
	Skewness		1,297	,616
	Kurtosis		2,274	1,191
ΣΤΕΓΑΣΤΙΚΑ1	Mean		2,08	,487
	95% Confidence Interval for Mean	Lower Bound	1,02	
		Upper Bound	3,14	
	5% Trimmed Mean		2,03	
	Median		2,00	
	Variance		3,077	
	Std. Deviation		1,754	
	Minimum		0	
	Maximum		5	
	Range		5	
	Interquartile Range		3	
	Skewness		,520	,616

	Kurtosis		-,739	1,191
ΣΤΕΓΑΣΤΙΚΑ2	Mean		1,62	,331
	95% Confidence Interval for Mean	Lower Bound	,89	
		Upper Bound	2,34	
	5% Trimmed Mean		1,57	
	Median		1,00	
	Variance		1,423	
	Std. Deviation		1,193	
	Minimum		0	
	Maximum		4	
	Range		4	
	Interquartile Range		2	
	Skewness		,548	,616
	Kurtosis		-,245	1,191
ΚΕΦ.ΚΙΝΗΣΗΣ1	Mean		,69	,208
	95% Confidence Interval for Mean	Lower Bound	,24	
		Upper Bound	1,15	
	5% Trimmed Mean		,66	
	Median		1,00	
	Variance		,564	
	Std. Deviation		,751	
	Minimum		0	
	Maximum		2	
	Range		2	
	Interquartile Range		1	
	Skewness		,611	,616
	Kurtosis		-,776	1,191
ΚΕΦ.ΚΙΝΗΣΗΣ2	Mean		1,38	,385
	95% Confidence Interval for Mean	Lower Bound	,55	
		Upper Bound	2,22	
	5% Trimmed Mean		1,32	
	Median		1,00	

	Variance		1,923	
	Std. Deviation		1,387	
	Minimum		0	
	Maximum		4	
	Range		4	
	Interquartile Range		3	
	Skewness		,503	,616
	Kurtosis		-,972	1,191
ΕΠΑΓΓΕΛΜΑΤΙΚΑ1	Mean		1,15	,274
	95% Confidence Interval for Mean	Lower Bound	,56	
		Upper Bound	1,75	
	5% Trimmed Mean		1,12	
	Median		1,00	
	Variance		,974	
	Std. Deviation		,987	
	Minimum		0	
	Maximum		3	
	Range		3	
	Interquartile Range		1	
	Skewness		,876	,616
	Kurtosis		,294	1,191
ΕΠΑΓΓΕΛΜΑΤΙΚΑ2	Mean		,85	,274
	95% Confidence Interval for Mean	Lower Bound	,25	
		Upper Bound	1,44	
	5% Trimmed Mean		,77	
	Median		1,00	
	Variance		,974	
	Std. Deviation		,987	
	Minimum		0	
	Maximum		3	
	Range		3	
	Interquartile Range		2	

	Skewness	,967	,616
	Kurtosis	,160	1,191

Percentiles

		Percentiles						
		5	10	25	50	75	90	
Weighted Average(Definition 1)	ΚΑΤΑΝΑΛΩΤΙΚΑ1	3,00	3,40	5,50	6,00	8,00	9,60	.
	ΚΑΤΑΝΑΛΩΤΙΚΑ2	3,00	3,00	5,00	7,00	9,00	14,00	.
	ΣΤΕΓΑΣΤΙΚΑ1	,00	,00	,50	2,00	3,50	5,00	.
	ΣΤΕΓΑΣΤΙΚΑ2	,00	,00	1,00	1,00	2,50	3,60	.
	ΚΕΦ.ΚΙΝΗΣΗΣ1	,00	,00	,00	1,00	1,00	2,00	.
	ΚΕΦ.ΚΙΝΗΣΗΣ2	,00	,00	,00	1,00	2,50	3,60	.
	ΕΠΑΓΓΕΛΜΑΤΙΚΑ1	,00	,00	,50	1,00	1,50	3,00	.
	ΕΠΑΓΓΕΛΜΑΤΙΚΑ2	,00	,00	,00	1,00	1,50	2,60	.
Tukey's Hinges	ΚΑΤΑΝΑΛΩΤΙΚΑ1			6,00	6,00	8,00		
	ΚΑΤΑΝΑΛΩΤΙΚΑ2			5,00	7,00	8,00		
	ΣΤΕΓΑΣΤΙΚΑ1			1,00	2,00	3,00		
	ΣΤΕΓΑΣΤΙΚΑ2			1,00	1,00	2,00		
	ΚΕΦ.ΚΙΝΗΣΗΣ1			,00	1,00	1,00		
	ΚΕΦ.ΚΙΝΗΣΗΣ2			,00	1,00	2,00		
	ΕΠΑΓΓΕΛΜΑΤΙΚΑ1			1,00	1,00	1,00		
	ΕΠΑΓΓΕΛΜΑΤΙΚΑ2			,00	1,00	1,00		

ΠΙΝΑΚΕΣ ΜΕΤΡΩΝ ΘΕΣΗΣ ΚΑΙ ΕΚΑΤΟΣΤΗΜΟΡΙΩΝ 4ΟΥ ΤΡΙΜΗΝΟΥ ΕΤΟΥΣ 2008.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Α. ΕΛΛΗΝΙΚΗ

1. Κιόχος Α. Π.: «Στατιστική» , εκδόσεις Interbooks, Αθήνα 1993
2. Ιωαννίδης Α. Δ.: «Στατιστικές Μέθοδοι», εκδόσεις Ζήτη, Θεσσαλονίκη 1999
3. Κιόχος Α. Π.: «Περιγραφική Στατιστική», εκδόσεις Interbooks, Αθήνα 1993
4. Τσέρπες Ν.Α. – Αλεβίζος Φ.: «Εισαγωγή στην Θεωρία της δειγματοληψίας», εκδόσεις Πανεπιστημίου Πατρών, Πάτρα 1999
5. Κουρουκλή Σ.: «Στατιστική Ι», εκδόσεις Πανεπιστημίου Πατρών, Πάτρα 1999
6. Ρόντος Κ. – Παπανής Ε.: «Μέθοδοι συλλογής στατιστικών δεδομένων», εκδόσεις Σιδέρη, Αθήνα 2007
7. Κικιλιάς Π. – Παλαμούρδας Δ. – Πετράκης Α. – Τσουκαλάς Δ.: «Στατιστική - Πιθανότητες», εκδόσεις Δήρος, Αθήνα 2001
8. Γεωργίου Γρ. Ρ.: «Θεωρία Πιθανοτήτων», εκδόσεις Ζήτη, Θεσσαλονίκη 1998

Β. ΗΛΕΚΤΡΟΝΙΚΗ

1. www.statistics.scientist.gr
2. www.nnipi.gr
3. www.uoi.gr
4. www.cc.uoq.gr
5. www.obelix.ee.duth.gr
6. www.math.ntua.gr
7. www.androulakis.bma.upatras.gr
8. www.el.wikipedia.org