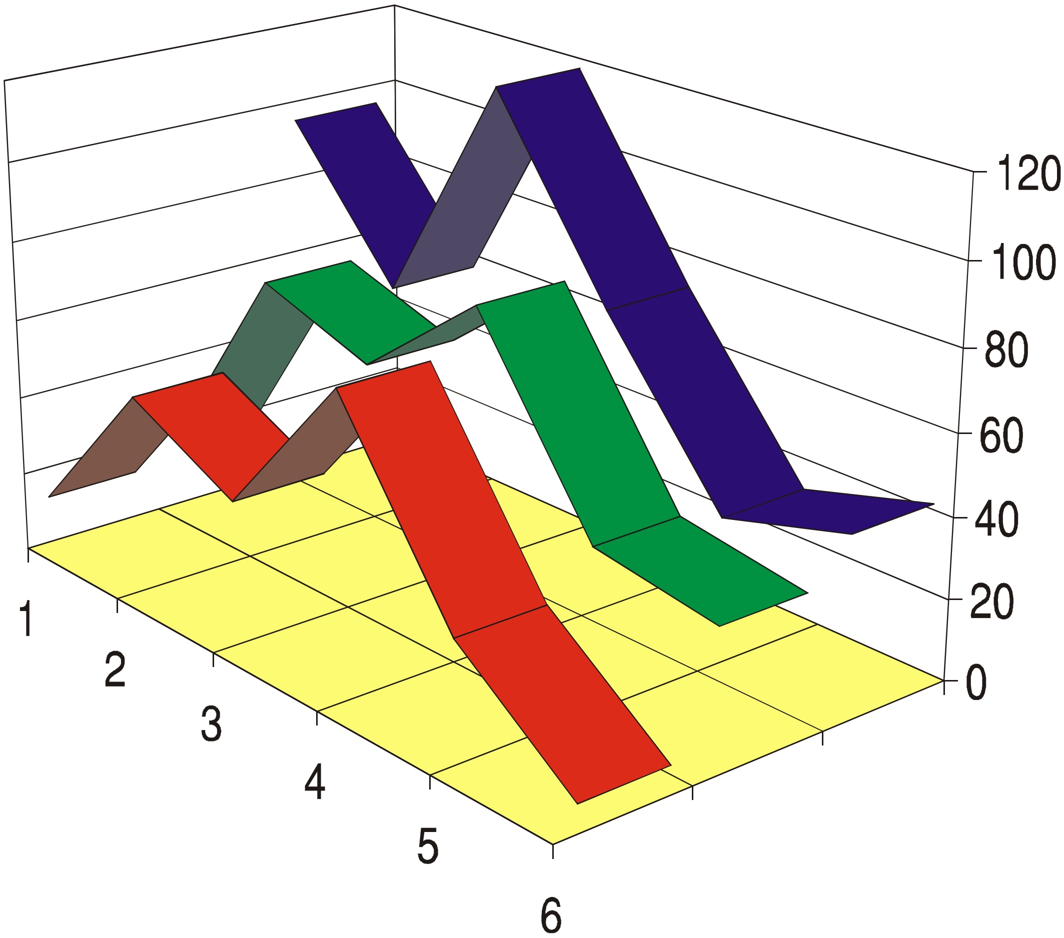
ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ
ΤΜΗΜΑ ΑΓΡΟΤΙΚΗΣ ΑΝΑΠΤΥΞΗΣ
Θεόδωρος Χ. Κουτρουμανίδης
Αναπληρωτής Καθηγητής ΔΠΘ

ΣΤΑΤΙΣΤΙΚΗ



Ορεστιάδα 2007

ΠΕΡΙΕΧΟΜΕΝΑ

Κεφάλαιο 1^o : Στατιστική.

Κεφάλαιο 2^o : Ταξινόμηση, κατάταξη και παρουσίαση των Στατιστικών δεδομένων.

Κεφάλαιο 3^o : Μέτρα θέσης και διασποράς.

Π ρ ό λ ο γ ο ς

Σκοπός του παρόντος βοηθήματος είναι να δώσει στον φοιτητή του τμήματος της Αγροτικής Ανάπτυξης, με όσο το δυνατόν μεγαλύτερη απλότητα και πιο ουσιαστικά, τις βασικές γνώσεις της Στατιστικής και να παρουσιάσει κύριους τομείς εφαρμογής της Στατιστικής στο χώρο της Οικονομικής διαχείρισης και της διαχείρισης των επιχειρήσεων με έμφαση στον τομέα της Αγροτικής Οικονομίας και Ανάπτυξης.

Η εφαρμοσμένη Στατιστική αποτελεί σημαντικό εργαλείο δουλειάς για ένα μελλοντικό επιστήμονα που θα ασχοληθεί με θέματα Αγροτικής Οικονομίας και Ανάπτυξης.

Για τον λόγο αυτό ο φοιτητής ή φοιτήτρια πρέπει να εμπεδώσει τις βασικές έννοιες της Στατιστικής και να εξασκηθεί με πρακτικά προβλήματα που αναδεικνύουν τον σπουδαίο ρόλο της Στατιστικής στο πεδίο της εφαρμογής.

Το ξεκίνημα αυτής της προσπάθειας θέλω να πιστεύω ότι θα έχει συνέχεια και ότι λάθη και παραλήψεις στο παρόν εγχείρημα θα διορθωθούν στην πορεία με τελικό στόχο να προκύψει ένα βοήθημα απλό – ουσιαστικό και κατανοητό για όλους όσους ενδιαφέρονται και ασχολούνται άμεσα με τα θέματα της Αγροτικής Οικονομίας και Ανάπτυξης.

Θ. Χ. Κουτρουμανίδης

Ε ι σ α γ ω γ ή

Η Στατιστική πρωτοεμφανίστηκε στην αρχαιότητα ως πρακτική μεθοδολογία καταμέτρησης δεδομένων και στην πορεία εξελίχθηκε σ΄ ένα κλάδο επιστημονικό με την αντίστοιχη θεωρητική υποδομή κυρίως με την εισαγωγή της θεωρίας των πιθανοτήτων.

Ποτέ όμως δεν έπαψε να αποτελεί εργαλείο δουλειάς και έρευνας για όλους τους επιστημονικούς κλάδους θεωρητικούς και θετικούς.

Σήμερα με την ραγδαία εξέλιξη της πληροφορικής η Στατιστική έχει ενσωματωθεί σε πάρα πολλές επιστήμες και αποτελεί αναπόσπαστο τμήμα τους. Παράλληλα επικρατεί η άποψη ότι μειώνεται ο ρόλος του επιστήμονα – ερευνητή σ΄ ένα περιβάλλον που κατακλύζεται από στατιστικά προγράμματα μέσω Η/Υ.

Εδώ θα πρέπει να είμαστε σαφείς και κατηγορηματικοί ο ερευνητής έχει τον πρώτο και τον τελευταίο λόγο στην επεξεργασία των δεδομένων μέσω στατιστικών πακέτων και θα πρέπει να έχει κατανοήσει καλά τις βασικές έννοιες της Στατιστικής (περιγραφικής και επαγωγικής) για να μπορέσει να κάνει σωστά τη δουλειά του. Άρα προέχει η εκπαίδευση του σε θέματα Στατιστικής και δεν μπορεί κανένα στατιστικό πρόγραμμα να υποκαταστήσει τον ερευνητή.

Το παρόν σύγγραμμα κινείται σ΄ αυτήν την αντίληψη, να δώσει δηλαδή απλά και ουσιαστικά τις βασικές γνώσεις της Στατιστικής και να παρουσιάσει εφαρμογές της Στατιστικής στον χώρο της Αγροτικής Οικονομίας και Ανάπτυξης, ώστε να μπορεί ο μελλοντικός επιστήμονας που θα ασχοληθεί με τα θέματα της Αγροτικής Ανάπτυξης να χρησιμοποιεί στατιστικές μεθόδους και τεχνικές για την δουλειά του.

Η δομή του συγγράμματος περιλαμβάνει βασικά στοιχεία της Περιγραφικής Στατιστικής που αφορούν την συλλογή – Ταξινόμηση – Κατάταξη και παρουσίαση στοιχείων (Κεφάλαια 1,2,3) καθώς επίσης και στοιχεία της Επαγωγικής Στατιστικής, όπως είναι η θεωρία πιθανοτήτων, οι τυχαίες μεταβλητές και οι κατανομές των πιθανοτήτων (Κεφάλαια 4,5,6).

Κεφάλαιο $1^{ο}$

Στατιστική

1.1. Τι είναι στατιστική

1. Ορισμός της Στατιστικής

Στατιστική είναι η επιστήμη που ασχολείται με τη συλλογή, οργάνωση, παρουσίαση ανάλυση και ερμηνεία δεδομένων – πληροφοριών, οι οποίες προκύπτουν από μια διαδικασία μέτρησης ή παρατήρησης ενός ή περισσοτέρων χαρακτηριστικών που αφορούν τον φυσικό κόσμο και όλες τις εκφάνσεις της ζωής του ανθρώπου.

Σήμερα η Στατιστική χρησιμοποιείται ευρύτατα σε όλους τους τομείς της ανθρώπινης δράσης και σε όλες τις Επιστήμες θεωρητικές ή εφαρμοσμένες, (τις Ανθρωπιστικές, Νομικές και τις Κοινωνικές Επιστήμες, τις Φυσικές Επιστήμες, τις Επιστήμες της Υγείας, τις Τεχνολογικές Επιστήμες, τις Οικονομικές Επιστήμες κλπ) αποτελεί οργανικό κομμάτι τους.

Στην οικονομία σήμερα οι μάνατζερ ενδιαφέρονται για τα αριθμητικά δεδομένα που αφορούν την παραγωγή και την πώληση των προιόντων επιχειρήσεων τους, την οικονομική προβολή των επιχειρήσεών τους, την μέτρηση της παραγωγικότητας των εργαζομένων κ.λ.π.

Πολλά θέματα που απασχολούν σήμερα τις επιχειρήσεις και τους διευθύνοντές τους αφορούν την αλλαγή ή διατήρηση του προσωπικού, την αλλαγή ή αντικατάσταση μέρους ενός προϊόντος, τον ποιοτικό έλεγχο του προϊόντος, την παραγωγικότητα της όλης διαδικασίας, την αποτελεσματικότητα της διαφήμισης του προϊόντος, τη γνώση των προτιμήσεων του καταναλωτικού κοινού μέσω έρευνας αγορά κλπ. Όλα τα παραπάνω θέματα εξετάζονται μέσα από στατιστικές έρευνες και στατιστικά μεγέθη τα οποία πρέπει να μπορεί να μελετά το κάθε στέλεχος της επιχείρησης για να παίρνει τις σωστές αποφάσεις προς όφελος της επιχείρησης.

Η Στατιστική πλέον έχει ενσωματωθεί στην οικονομική θεωρία και μάλιστα έχουν δημιουργηθεί νέοι επιστημονικοί κλάδοι όπως είναι η οικονομετρία που προήλθε από την συνύπαρξη της οικονομικής θεωρίας και της Στατιστικής.

Τα αριθμητικά στοιχεία που συστηματικά συγκεντρώνονται καλούνται **δεδομένα** στην Στατιστική. Έτσι εάν ελέγξουμε το όριο ζωής έξι (6) λαμπτήρων και καταλήξουμε στα ακόλουθα αποτελέσματα:

Λαμπτήρας Ι:	713 ώρες
Λαμπτήρας ΙΙ:	697 ώρες
Λαμπτήρας ΙΙΙ:	696 ώρες
Λαμπτήρας IV:	28 ώρες
Λαμπτήρας V:	707 ώρες
Λαμπτήρας VI:	1347 ώρες

Τα στοιχεία αυτά είναι τα δεδομένα της στατιστικής έρευνας που έγινε και μας επιτρέπουν να κάνουμε κάποιες παρατηρήσεις. Συγκεκριμένα παρατηρούμε ότι το όριο ζωής των λαμπτήρων κυμαίνεται από τις 28 ώρες μέχρι 1347 ώρες. Το όριο ζωής είναι περίπου κατά μέσο όρο 700 ώρες και οι τιμές απέχουν λίγο πάνω ή κάτω από το μέσο όρο, αν εξαιρέσουμε τις δύο ακραίες τιμές 28 ωρών και 1347 ωρών οι οποίες άλλωστε δεν είναι και όπως φαίνεται συνήθεις.

Όσον αφορά τον **ποιοτικό έλεγχο** των λαμπτήρων αυτός προκύπτει με την εξαγωγή των συμπερασμάτων.

Αρκετές φορές με τον όρο Στατιστική αναφερόμαστε σε τεχνικές και μεθόδους που χρησιμοποιούνται σε εξεταζόμενα προβλήματα για να αναλυθούν τα στατιστικά δεδομένα που έχουμε στη διάθεσή μας. Η στατιστική μέθοδος είναι συνήθως μαθηματική μέθοδος και επομένως πρέπει να γνωρίζουμε τις μαθηματικές έννοιες όπως είναι μέση τιμή, απόκλιση από την μέση τιμή, κατανομή πιθανότητας κλπ.

Η Στατιστική συνεπώς είναι μια σειρά από μαθηματικές τεχνικές οι οποίες χρησιμεύουν για την ανάλυση προβλημάτων του πραγματικού κόσμου. Τα στατιστικά δεδομένα που χρησιμοποιούνται σε κάθε επιστήμη και ασφαλώς και στην οικονομία μας επιτρέπουν να ελέγχουμε τις υποθέσεις που γίνονται και να αναπτύσσουμε μοντέλα για τον σχεδιασμό και την πρόβλεψη των στόχων-σκοπών που έχουν τεθεί.

2. Κλάδοι της στατιστικής

Η κλάδος της Στατιστικής που ασχολείται με τον σχεδιασμό του τρόπου συλλογής των πληροφοριών καλείται **Σχεδιασμός πειραμάτων,** ενώ ο κλάδος που ασχολείται με την σύντομη και ολοκληρωμένη παρουσίαση των δεδομένων καλείται **Περιγραφική Στατιστική.**

Η περιγραφική Στατιστική μας δίνει τη δυνατότητα να βγάλουμε κάποιες καταρχήν πληροφορίες για τα στοιχεία που μελετάμε.

Ένα παράδειγμα του τρόπου με τον οποίο παρουσιάζονται στην περιγραφική Στατιστική οι πληροφορίες είναι το ακόλουθο. Με μια έρευνα πεδίου σε μια εμπορική επιχείρηση με 30 υπαλλήλους γίνεται καταγραφή των μισθών των υπαλλήλων και τα στοιχεία που προκύπτουν είναι: οι δεκαπέντε πωλητές αμείβονται από 200-350 χιλ., οι πέντε αποθηκάριοι από 150-220 χιλ., οι επτά διοικητικοί από 210-300 χιλ. και τα τρία στελέχη της επιχείρησης από 500-800χιλ. Τα στοιχεία αυτά μπορούν να παρουσιασθούν σε ένα πίνακα της παρακάτω μορφής, που μας δίνει μια συνοπτική εικόνα των αμοιβών των εργαζομένων στην επιχείρηση.

Πίνακας 1.

Κατηγορία εργαζομένων	**Αριθμός**	**Μισθός (χιλ.)**
1. Πωλητές	15	200-350
2. Αποθηκάριοι	5	150-220
3. Διοικητικοί	7	210-300
4. Στελέχη	3	500-800

Ένας άλλος κλάδος της Στατιστικής είναι η **Αναλυτική Στατιστική** η οποία αναλύει τα δεδομένα έτσι ώστε να βοηθά στην λήψη των αποφάσεων, όπως είναι η απόφαση των ιθυνόντων σε μια επιχείρηση να αυξηθεί η παραγωγή ενός προϊόντος στηριζόμενοι στην ανάλυση των στατιστικών στοιχείων τα οποία έδειξαν ότι υπάρχει αυξανόμενη ζήτηση του εν λόγω προϊόντος στην αγορά.

Όμως η λήψη αποφάσεων σε θέματα στα οποία εκτός των μετρήσιμων μεγεθών, όπου η στατιστική έρευνα παίζει καθοριστικό ρόλο, πρέπει να ληφθούν υπόψη και άλλοι μη μετρήσιμοι παράγοντες είναι πολύ δύσκολη. Τέτοιες περιπτώσεις θεμάτων είναι τα έργα στα οποία γίνεται **ανάλυση κόστους - οφέλους** όπου είναι πολύ δύσκολο π.χ να αποτιμηθεί το όφελος από την επέκταση και τον εκσυγχρονισμό μιας επιχείρησης σε κτιριακή και τεχνολογική υποδομή ή από την κατασκευή ενός τεχνικού έργου π.χ αυτοκινητοδρόμου, γέφυρας, λιμένος, αεροδρομίου κλπ.

Προβλήματα που απαιτούν στατιστική διερεύνηση μπορούν να αναφερθούν αρκετά και κυρίως στον οικονομικό τομέα. Συγκεκριμένα ως τέτοια μπορούν να είναι οι απαιτήσεις και οι ανάγκες μιας κοινωνίας για αγαθά και υπηρεσίες και για το ποιους ρόλους πρέπει να παίζουν σε κάθε κοινωνία ο ιδιωτικός και ο δημόσιος τομέας στην παραγωγή αγαθών και υπηρεσιών; Ποιο είναι το επίπεδο ζωής των κατοίκων, ποιες είναι οι προσδοκίες για το μέλλον και ποιο το μέγεθος και το είδος των αγαθών και των υπηρεσιών για επιτευχθούν οι προσδοκίες αυτές; Ποιες οι εισαγόμενες ποσότητες και από ποιες πηγές για κάθε κράτος; Πως διασφαλίζεται μια γενική ποιότητα αγαθών και υπηρεσιών; Ποιοι οι κίνδυνοι και ποιες οι πιθανότητες να συμβεί ένα συγκεκριμένο γεγονός; Ποιες ενέργειες και ποιοι μηχανισμοί ανατροφοδότησης πρέπει να λειτουργήσουν σε μια επιχείρηση για να έχουμε βέλτιστα μεγέθη στην παραγόμενη ποσότητα, στο κόστος παραγωγής, στην ποιότητα του προϊόντος;

Πολλές φορές η περιγραφική στατιστική μελετά, λόγω αδυναμιών στην συγκέντρωση του στατιστικού υλικού, ένα υποσύνολο του συνόλου των δεδομένων που θα έπρεπε να συγκεντρώσει, ένα αντιπροσωπευτικό **δείγμα** του υπό μελέτη συνόλου και προσεγγίζει τα χαρακτηριστικά του συνόλου μέσα από τα χαρακτηριστικά του δείγματος.

Ο κλάδος της Στατιστικής που ασχολείται με τη συστηματική αντιμετώπιση τέτοιων προβλημάτων και χρησιμοποιεί το δείγμα (το υποσύνολο) για να μελετήσει το σύνολο καλείται **Επαγωγική στατιστική.** Είναι ο κλάδος που εφαρμόζει την επαγωγική σκέψη της μαθηματικής θεωρίας για να οδηγηθεί από το μέρος στον όλον.

Εφαρμόζει έρευνες δειγματοληψίας, στηρίζεται στις βασικές αρχές της δειγματοληψίας, χρησιμοποιεί διαδικασίες που μας οδηγούν στην εκτίμηση των χαρακτηριστικών του συνόλου από τα αποτελέσματα των αντιστοίχων χαρακτηριστικών του δείγματος και ελέγχει την αξιοπιστία των συμπερασμάτων στα οποία καταλήγει.

Βασικές μέθοδοι της Επαγωγικής Στατιστικής είναι:

i. Η εκτίμηση παραμέτρων του πληθυσμού.
ii. Ο έλεγχος των Στατιστικών υποθέσεων.
iii. Η θεωρία Λήψης αποφάσεων.

Η Επαγωγική Στατιστική στηρίζεται στη θεωρία των πιθανοτήτων.

3. Τα βήματα μιας στατιστικής έρευνας

1. Πρέπει να καθορίσουμε το πρόβλημα και να μην αναφέρεται αυτό σε γενικούς όρους. Παραδείγματος χάριν αν θέλουμε να ερευνήσουμε τις εργασιακές αμοιβές μιας επιχείρησης θα πρέπει να αναφερόμαστε όχι γενικά σε όλα τα τμήματα της εταιρείας, γιατί υπάρχουν πολλά τμήματα όπου οι εργασιακές αμοιβές δεν παρουσιάζουν διακυμάνσεις, αλλά σε ιδιαίτερα τμήματα. Επίσης πρέπει να αναφερόμαστε σε μια ιδιαίτερη λειτουργία ή προϊόν. Έτσι θα προκύψουν τα θέματα και οι αναφορές με βάση των οποίων θα μπορούμε να αρχίσουμε να συγκεντρώνουμε σχετικά στατιστικά στοιχεία για ανάλυση.
2. Να υπάρξει μια καλή προσέγγιση του θέματος εφόσον το μελετήσουμε μέσα από σχετικό υλικό που θα είναι δημοσιευμένο.
3. Να διερευνήσουμε αν πρέπει και μπορούμε να εξετάσουμε ολόκληρο τον πληθυσμό κάνοντας απογραφή ή αν εξετάσουμε ένα δείγμα αυτού και ποιας μορφής δείγμα.
4. Να εξετάσουμε την αναγκαιότητα να καταφύγουμε σε ερωτηματολόγιο και τότε μπαίνει το ερώτημα πως θα το κατασκευάσουμε.
5. Να συγκεντρώσουμε τα δεδομένα με διάφορες τεχνικές και μεθόδους.
6. Να γράψουμε και να ταξινομήσουμε τα δεδομένα που συγκεντρώσαμε.
7. Τα δεδομένα τα ταξινομούμε σε πίνακες και τα απεικονίζουμε με στατιστικά διαγράμματα.
8. Να κάνουμε την ανάλυση των στοιχείων.
9. Με την ανάλυση των στοιχείων εννοούμε τον υπολογισμό των διαφόρων στατιστικών παραμέτρων που είναι απαραίτητες για να μελετηθεί το στατιστικό υλικό.
10. Να τα παρουσιάσουμε, να τα ερμηνεύσουμε και να βγάλουμε τα συμπεράσματα μας. Τέλος να κάνουμε ενδεχόμενα προτάσεις για την λήψη κάποιων αποφάσεων, για τον προγραμματισμό κλπ.

4. Στατιστική και οικονομική διαχείριση

Πολλά στατιστικά στοιχεία συγκεντρώνονται και παρουσιάζονται από κοινωνιολόγους και στατιστικολόγους, όμως αρκετά από τα στατιστικά στοιχεία και κυρίως τα οικονομικά στοιχεία αφορούν και ενδιαφέρουν τους διευθύνοντες των επιχειρήσεων του ιδιωτικού τομέα αλλά και των δημοσίων οργανισμών που λειτουργούν με ιδιωτικο-οικονομικά κριτήρια όπως και των φορέων της δημόσιας διοίκησης.

Η στατιστική που ασχολείται με την επιχείρηση εξετάζει τα στοιχεία εκείνα που απεικονίζουν τα αποτελέσματα της δραστηριότητας της. Από την μια μεριά εφαρμόζει μεθόδους συγκέντρωσης στατιστικού υλικού από την λειτουργία της επιχείρησης και από την άλλη χρησιμοποιεί μεθόδους ανάλυσης και εξαγωγής συμπερασμάτων που ενδιαφέρουν τους επικεφαλής διευθύνοντες την επιχείρηση.

Όλοι οι τομείς μιας σύγχρονης επιχείρησης χρησιμοποιούν στατιστικές μεθόδους για να λειτουργήσουν αποδοτικά. Έτσι η στατιστική εφαρμόζεται στον έλεγχο της παραγωγικότητας του προσωπικού, στην μεταβολή του προσωπικού, στον έλεγχο του κόστους της εργασίας και των πρώτων υλών, στον έλεγχο της παραγωγής, στον έλεγχο των πωλήσεων, στην παρακολούθηση των πινάκων εισαγωγών-εξαγωγών, τον έλεγχο του επιπέδου των αποθηκευμένων προϊόντων, στην παρακολούθηση των πελατών της επιχείρησης, στον έλεγχο των περιοχών πωλήσεων-του μεγέθους των παραγγελιών-της μεθόδου διανομής των προϊόντων, στον έλεγχο των οχημάτων της επιχείρησης, στον έλεγχο του μηχανολογικού και του τεχνολογικού εξοπλισμού της επιχείρησης, στον ποιοτικό έλεγχο των προϊόντων κλπ.

Οι βασικές στατιστικές δραστηριότητες μιας μεγάλης επιχείρησης είναι:

1. Το τμήμα των Οικονομικών και Στατιστικών μελετών:
   Ασχολείται με την ανάλυση των επιχειρηματικών τάσεων και την πρόβλεψη των επιχειρηματικών δραστηριοτήτων. Εξετάζει όμως και άλλους οικονομικούς παράγοντες της επιχείρησης.
2. Το τμήμα Marketing ασχολείται με την έρευνα των προτιμήσεων του καταναλωτικού κοινού, την έρευνα της αγοράς για την εισαγωγή ενός νέου προϊόντος, της τιμής και της ποιότητας ενός προϊόντος που ήδη κυκλοφορεί κλπ.
3. Το τμήμα Παραγωγής εφαρμόζει στατιστικούς ελέγχους ποιότητας των παραγόμενων προϊόντων.
4. Το τμήμα του Οικονομικού ελέγχου χρησιμοποιεί στατιστικές και λογιστικές μεθόδους για να συντάσσει τους ετήσιους προϋπολογισμούς.
5. Το τμήμα Προσωπικού χρησιμοποιεί στατιστικές επεξεργασίας για να παρακολουθεί την δύναμη του Προσωπικού (σύνδεση-αμοιβές-χρόνος εργασίας κλπ.)

1.2. Περιγραφική Στατιστική

1. Το αντικείμενο της περιγραφικής Στατιστικής

Η περιγραφική Στατιστική ασχολείται με τα ακόλουθα:

α) Την ταξινόμηση των στατιστικών δεδομένων σε πίνακες (συχνοτήτων, σχετικών συχνοτήτων, αθροιστικών συχνοτήτων κλπ.) και την απεικόνιση των στατιστικών δεδομένων σε διαγράμματα.

β) Τον υπολογισμό των βασικών στατιστικών παραμέτρων (μέτρων θέσεως, μέτρων διασποράς, μέτρων ασυμμετρίας, μέτρων συγκέντρωσης).

γ) Την μελέτη στατιστικών πληθυσμών ως προς δύο μεταβλητές ή και περισσότερες. Εξετάζεται δηλαδή η τυχόν επίδραση μιας μεταβλητή επί της άλλης σε ένα δείγμα (παλινδρόμηση, συσχέτιση).

δ) Την διαμόρφωση των διαφόρων Αριθμοδεικτών οι οποίοι χρησιμεύουν για τη διαχρονική ή διατοπική παρακολούθηση της εξέλιξης των διαφόρων οικονομικών μεγεθών.

ε) Την ανάλυση των Χρονολογικών σειρών για την μελέτη της εξέλιξης στο χρόνο των διαφόρων φαινομένων (οικονομικών, κοινωνικών κλπ.).

2. Βασικές έννοιες

Πληθυσμός ή στατιστικός πληθυσμός είναι το σύνολο των μετρήσεων ή των παρατηρήσεων σε μια έρευνα, που αναφέρονται σε ένα χαρακτηριστικό γνώρισμα ενός συνόλου που εξετάζεται. Τα στοιχεία αυτού του πληθυσμού καλούνται **στατιστικές μονάδες.** Τα όρια του συνόλου-πληθυσμού πρέπει να είναι σαφή και καθορισμένα.

Μεταβλητή είναι ένα χαρακτηριστικό γνώρισμα του πληθυσμού, ως προς το οποίο εξετάζουμε τα στοιχεία του και συμβολίζεται με τα κεφαλαία γράμματα **X,Ψ,Z…,** ο όρος **παρατήρηση** ή **παρατηρούμενη τιμή** είναι η αριθμητική τιμή της μεταβλητής και συμβολίζεται με το αντίστοιχο μικρό γράμμα **x, y,z,……**

Οι μεταβλητές διακρίνονται σε:

1. **Ποσοτικές** που εκφράζονται με αριθμούς και τέτοιες είναι π.χ. οι τιμές πώλησης αγροτικών προϊόντων σε μια περίοδο κατά νομό της χώρας, το κατά κεφαλήν ακαθάριστο εθνικό προϊόν (ΑΕΠ) κατά περιφέρεια της χώρας σε ένα συγκεκριμένο έτος, οι μισθοί των δημοσίων υπαλλήλων μια δεδομένη χρονική στιγμή κ.λ.π.

Οι ποσοτικές μεταβλητές διακρίνονται:

α. Στις **συνεχείς** μεταβλητές που μπορούν να πάρουν οποιαδήποτε τιμή σε ένα διάστημα των πραγματικών αριθμών (α,β), όπως το βάρος ενός ατόμου ή ο μισθός ενός εργαζομένου δηλαδή $\alpha<\chi<\beta$.

β. Στις **ασυνεχείς** μεταβλητές που παίρνουν διακριτές τιμές, όπως ο αριθμός των εργαζομένων κατά οικογένεια σε μια πόλη ή ο αριθμός των εργαζομένων στις επιχειρήσεις μεταποίησης αγροτικών προϊόντων κατά νομό της χώρας μια συγκεκριμένη χρονιά.

2. **Ποιοτικές** μεταβλητές που δεν μπορούν να μετρηθούν, όπως το φύλο των εργαζομένων σε μια δημόσια υπηρεσία, το επάγγελμα των κατοίκων σε μια περιοχή κάποιας πόλης. Οι μεταβλητές αυτές μπορούν να χωρισθούν σε **κατηγορίες-κλάσεις** και γι' αυτό καλούνται και **κατηγορικές**. Χωρίζονται:

α. **Ονομαστικές** μεταβλητές, όπως το φύλο, η θρησκεία

β. **Διατάξιμες** μεταβλητές όπως η κατάσταση της υγείας των εργαζομένων στα λατομεία της χώρας, η ποιοτική διαβάθμιση ενός προϊόντος.

3. Το υλικό προς συλλογή για στατιστική ανάλυση

Το υλικό που αφορά τα οικονομικά μεγέθη των επιχειρήσεων, όπως η παραγωγή των αγαθών, το επίπεδο των προσφερομένων υπηρεσιών, η κατανάλωση-ζήτηση των προϊόντων, το κόστος των παραγομένων προϊόντων κλπ., αφού πρώτα συγκεντρωθεί με την έρευνα, γίνεται η παραβολή των στοιχείων που περιλαμβάνει και η ανάλυση αυτών, κατόπιν συντάσσονται οι σχετικές αναφορές και οι προτάσεις που αποτελούν την βάση πάνω στην οποία το κάθε στέλεχος θα στηριχθεί για να πάρει τις αποφάσεις του και να κάνει τις απαιτούμενες κινήσεις.

Έχουμε δύο τύπους στατιστικού υλικού το **πρωτεύον** ή **πρωτογενές υλικό** ή τα **πρωτογενή δεδομένα** και το **δευτερεύον** ή **δευτερογενές υλικό** ή τα **δευτερογενή δεδομένα.** Τα πρωτογενή δεδομένα συγκεντρώνονται από «πρώτο χέρι» είναι **έρευνα πεδίου** και απαιτούν ασφαλώς μια πρώτη παραβολή, ένα πρώτο έλεγχο και ένα ξεκαθάρισμα καταρχήν ώστε να μπορούν να ταξινομηθούν κατόπιν σε πίνακες κ.α. για μια δεύτερη και πιο ουσιαστική επεξεργασία περαιτέρω.

Τα δευτερογενή στοιχεία που λαμβάνονται από διάφορους οργανισμούς και εταιρείες δημόσιου χαρακτήρα ή ιδιωτικού δεν απαιτούν τον πρώτο κύκλο ξεκαθαρίσματος γιατί πολλές φορές λαμβάνονται πίνακες και διαγράμματα αυτούσια που αναφέρονται σε οικονομικά και κοινωνικά μεγέθη όπως εκπαίδευση, παραγωγή, υγεία, μεταφορές, εξαγωγικό εμπόριο, εισαγωγές, ισολογισμούς πληρωμών κλπ.

Πολλές επιχειρήσεις ταξινομούν τα δεδομένα προς στατιστική ανάλυση σε **εσωτερικά** και **εξωτερικά δεδομένα.** Τα εσωτερικά δεδομένα παράγονται μέσα στην επιχείρηση και είναι οι αναφορές ρουτίνας της επιχείρησης όπως οι πωλήσεις, ο ποιοτικός έλεγχος των προιόντων, το επενδυόμενο κεφάλαιο, οι μισθοί κλπ. Τα εξωτερικά δεδομένα προέρχονται από δημόσιες ή ιδιωτικές πηγές, όπως περιοδικά, πιστωτικά ιδρύματα (Τράπεζες), δημόσιοι οργανισμοί και υπηρεσίες, ιδιωτικές εταιρίες κλπ.

Τα στατιστικά δεδομένα ποικίλουν από επιχείρηση σε επιχείρηση, από ίδρυμα σε ίδρυμα και οι εκάστοτε υπεύθυνοι παράγοντες ενδιαφέρονται για εκείνα τα στοιχεία που είναι της αρμοδιότητάς τους.

Συγκεκριμένα ο επιχειρηματίας χρειάζεται ασφαλώς στατιστικά στοιχεία που αφορούν:

α. Το **προϊόν** (τιμές και κόστη πρώτων υλών, παραγωγή, φθορές, κλπ).
β. Τις **πωλήσεις** (ποσότητα του προϊόντος στην αποθήκη, όγκος πωλήσεων, τιμές πώλησης, κατανομή κόστους κλπ).
γ. Τον **κατάλογο πληρωμής ημερομισθίων** (εργασία, ημερομίσθιο, τζίρος, κόστος εκπαίδευσης κλπ).
δ. **έξοδα-δαπάνες** (συντήρηση κτηρίου, συντήρηση εγκαταστάσεων, αποθήκες οχήματα κλπ).
ε. **άλλα θέματα.**

Οι υπεύθυνοι παράγοντες μιας κεντρικής διοίκησης για να αποφασίσουν σε διάφορα θέματα πρέπει να λάβουν υπόψη τους στατιστικά στοιχεία που αφορούν το εθνικό προϊόν, το εισόδημα των πολιτών, τις δαπάνες, την απασχόληση, τις εισαγωγές και τις εξαγωγές κλπ. Οι υπεύθυνοι μιας τοπικής αυτοδιοίκησης πρέπει να λάβουν υπόψη τους στατιστικά στοιχεία που αφορούν την εκπαίδευση, την υγεία, την κοινωνική πρόνοια, την ευημερία των κατοίκων (όπως είναι οι δείκτες κοινωνικής ευημερίας, υγείας, παιδείας, κοινωνικής πρόνοιας), το περιβάλλον, τις μεταφορές, τις κοινωνικές υποδομές της περιοχής τους κλπ.

Οι πηγές της κάθε πληροφόρησης για την επιχείρηση μπορεί να είναι:

α. **Μόνιμες** πηγές (προσωπικό, πελάτες, προμήθειες, αντιπρόσωποι, ενοικιαστές κλπ.).

β. **Μεταβαλλόμενες** πηγές (καταστάσεις πωλήσεων, καταστάσεις αγοραστών, καταστάσεις πληρωμής ημερομισθίων, πράκτορες, τμήματα παραρτήματα κλπ.).

γ. **Συνήθεις** πηγές (φορείς του κράτους, τοπικές εξουσίες για θέματα π.χ. εργατικών ατυχημάτων, επιπέδου λειτουργίας των δημόσιων σχολείων, των νοσοκομείων κλπ.).

Όταν αναφερόμαστε στις πωλήσεις μιας επιχείρησης συνήθως αναγράφονται οι ποσότητες που πουλήθηκαν ανά εβδομάδα (μια συνήθης αναφορά σε επιχειρήσεις) και δεν μας ενδιαφέρουν τα ονόματα και οι διευθύνσεις των πωλητών (πχ.).

Πίνακας 2.

ΕΒΔΟΜΑΔΑ	ΠΟΣΟΤΗΤΑ	ΕΒΔΟΜΑΔΑ	ΠΟΣΟΤΗΤΑ
1	326,43	8	528,87
2	435,45	9	432,98
3	876,34	10	435,97
4	654,32	11	763,26
5	324,76	12	786,65
6	765,87	13	675,76
7	437,98	14	456,87
Σύνολο	**3.821,51**		**4.080,36**
Γενικό Σύνολο	**7.901,87**		

Επίσης ο έλεγχος της δραστηριότητας των πωλητών από αυτούς που ασκούν την διοίκηση σε μια επιχείρηση (μάνατζερς) μπορεί να γίνει μέσα από το στατιστικό υλικό το οποίο συγκεντρώνουν και το οποίο αναφέρεται σε::

- Πωλήσεις ανά μήνα
- Τηλεφωνικές συνδιαλέξεις ανά μήνα
- Έξοδα ανά μήνα
- Καλυπτόμενα χιλιόμετρα ανά μήνα
- Πωλήσεις ανά τηλεφωνική συνδιάλεξη
- Νέοι πελάτες ανά μήνα
- Επανάληψη παραγγελιών από παλιούς πελάτες
- Έξοδα οχήματος ανά μήνα
- Πωλήσεις ανά χιλιόμετρο μετακίνησης

1.3. Μέθοδοι έρευνας – Απογραφή και δείγμα

1. Τρόποι συγκέντρωσης στατιστικών στοιχείων

Τα δεδομένα μπορούν να συγκεντρωθούν με διάφορους τρόπους, όπως με **παρατηρήσεις**, με **φύλλα ποιοτικού ελέγχου,** με **καταγραφές περιοδικές ή τακτικές,** με **ερωτηματολόγιο,** με **συνέντευξη.**

Παρατηρήσεις

Είναι ο τρόπος με τον οποίο γίνονται οι μετρήσεις με βάση την χρονική εξέλιξη σε εργοστάσια, επιχειρήσεις, υπηρεσίες του ιδιωτικού και του δημοσίου τομέα και αφορούν το προσωπικό, τις μετακινήσεις (προϊόντων και ατόμων), τα πειραματικά δεδομένα κλπ. Παραδείγματος χάριν σε ένα αυτοκινητόδρομο οι παρατηρήσεις των διερχομένων βαρέων οχημάτων, του είδους και της έντασης του κυκλοφοριακού φόρτου αυτών, χρησιμοποιούνται για τον προσδιορισμό του κόστους κατασκευής του αυτοκινητοδρόμου.

Φύλλα ποιοτικού ελέγχου.

Χρησιμοποιούνται κυρίως για αντικείμενα, όταν χρειαζόμαστε π.χ να γνωρίζουμε το βάρος, τη σύνθεση, το χρώμα, την αντοχή κλπ. των αντικειμένων που παράγονται στις βιομηχανίες. Συνήθως εξετάζουμε ένα δείγμα των προϊόντων αυτών γιατί δεν μπορεί να ελεγχθεί το σύνολο των παραγομένων προϊόντων (λόγω όγκου, αλυσίδας παραγωγής, κόστους κλπ.).

Ένα συνηθισμένο φύλλο ποιοτικού ελέγχου είναι:

<table>
<tr><td colspan="2">Σειρά προϊόντων Νο:</td><td colspan="2">Επιλεγόμενες μονάδες:
(αριθμός)</td><td>χρώμα:</td></tr>
<tr><td>Διαστάσεις:</td><td>μήκος:</td><td>πλάτος:</td><td>ύψος:</td><td>Διάμετρος</td></tr>
<tr><td colspan="2">Έλεγχος Θραύσης:</td><td colspan="2">επιτυχία:</td><td>αποτυχία</td></tr>
<tr><td colspan="2">Έλεγχος Χρώματος:</td><td colspan="2">επιτυχία:</td><td>αποτυχία:</td></tr>
<tr><td colspan="2" rowspan="2">Ημερομηνία:</td><td colspan="3">Ειδική αναφορά:</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="2">Τελικό αποτέλεσμα:</td><td colspan="2">επιτυχία:</td><td>αποτυχία:</td></tr>
</table>

Καταγραφές

Γίνονται συνήθως στις επιχειρήσεις και αφορούν την καταγραφή των πρώτων υλών ή των προϊόντων που υπάρχουν στις αποθήκες, την καταγραφή των εργαλειομηχανών, των οχημάτων, του τεχνικού και τεχνολογικού εξοπλισμού, του προσωπικού κλπ.

Ερωτηματολόγιο

Καταρχήν πρέπει πρώτα να εξεταστεί πιο θα είναι το περιεχόμενο του ερωτηματολογίου και δεύτερον πως θα πρέπει να διατυπωθεί. Πρέπει το ερωτηματολόγιο να σχεδιάζεται ανάλογα με το εξεταζόμενο θέμα, με το χρόνο που διατίθεται για να απαντήσουν οι ερωτώμενοι, την προθυμία τους να συνεργασθούν, και φυσικά το πόσο γνωρίζουν τα θέματα οι ερωτώμενοι. Συνήθως απευθυνόμαστε σε κατάλληλα άτομα που είναι γνώστες των θεμάτων που ερωτώνται εκτός εάν παραδείγματος χάριν κάνουμε έρευνα αγοράς για ένα νέο προϊόν και ζητάμε τις καταναλωτικές προτιμήσεις του κοινού.

Το περιεχόμενο των ερωτήσεων πρέπει να είναι κατανοητό κυρίως με μορφή πολλαπλών επιλογών και οι απαντήσεις να δίνονται εύκολα και γρήγορα είτε με ένα τσεκάρισμα είτε με ένα κύκλο.

Να δίνεται όμως απαραιτήτως και η δυνατότητα στον ερωτώντα να δίνει και κάποια άλλη απάντηση εκτός αυτών που του παρουσιάζει το ερωτηματολόγιο, εφόσον θα θελήσει να κάνει κάτι τέτοιο.

Παρατίθεται ένα ερωτηματολόγιο το οποίο απευθύνεται σε επιχειρηματίες και αφορά την γνώμη τους για τον ρόλο που παίζουν κάποιοι παράγοντες στην απόφαση ενός επιχειρηματία να επενδύσει σε μια περιοχή.

Συνέντευξη

Η συνέντευξη απευθύνεται σε άτομα τα οποία είναι εξειδικευμένα στα θέματα που διερευνούμε με το ερωτηματολόγιο και κατέχουν θέσεις διευθυντικές. Παρουσιάζει το μειονέκτημα ότι υπάρχει η πίεση του χρόνου με αποτέλεσμα μερικές φορές να μην έχουμε την νηφάλια προσέγγιση ενός θέματος και απαιτεί μεγαλύτερο κόστος από ότι η χρήση του ερωτηματολογίου. Έχει εκτιμηθεί ότι μπορούν να συμπληρωθούν μέχρι και 1000 ερωτηματολόγια την ημέρα που σημαίνει ότι θα έπρεπε να γίνει ο ίδιος αριθμός συνεντεύξεων την ημέρα κάτι το αδύνατο.

ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ ΓΙΑ ΤΗ ΒΑΣΙΚΗ ΕΙΚΟΝΑ ΜΙΑΣ ΠΕΡΙΟΧΗΣ (ΓΙΑ ΕΡΓΟΔΟΤΕΣ – ΕΠΙΧΕΙΡΗΜΑΤΙΕΣ)

ΣΤΟΙΧΕΙΑ ΕΡΩΤΩΜΕΝΟΥ:

ΟΝΟΜΑΤΕΠΩΝΥΜΟ:
ΙΔΙΟΤΗΤΑ:
ΕΠΩΝΥΜΙΑ ΕΠΙΧΕΙΡΗΣΗΣ (αν πρόκειται για επιχειρηματίες):
ΔΙΕΥΘΥΝΣΗ: Οδός: Τ.Κ.:
ΤΗΛΕΦΩΝΟ: FAX:
ΤΟΠΟΣ ΣΥΜΠΛΗΡΩΣΗΣ ΕΡΩΤΗΜΑΤΟΛΟΓΙΟΥ:
ΗΜΕΡΟΜΗΝΙΑ ΣΥΜΠΛΗΡΩΣΗΣ ΕΡΩΤΗΜΑΤΟΛΟΓΙΟΥ:

Εισαγωγή

Αυτή η έρευνα με ερωτηματολόγιο διεξάγεται στο πλαίσιο του ερευνητικού προγράμματος «Χωροθέτηση επιχειρήσεων και περιφερειακή ανάπτυξη των νομών Μακεδονίας, Θράκης και Ηπείρου». Η έρευνα διεξάγεται από το Τμήμα Διοίκησης

Επιχειρήσεων του Πανεπιστημίου του Αιγαίου και χρηματοδοτείται από τη Γενική Γραμματεία Έρευνας και Τεχνολογίας. Σ΄ αυτή τη φάση των εργασιών προσπαθούμε να διαμορφώσουμε το «κοινωνικο-οικονομικό προφίλ» των νομών της Βόρειας Ελλάδας. Σε μια μεταγενέστερη φάση θα γίνει προσπάθεια ανίχνευσης των παραγόντων εκείνων και των μέσων που επηρεάζουν και αλλάζουν αυτό το προφίλ και συνεπώς των ενεργειών που πρέπει να λάβουν χώρα ώστε να γίνει αυτή η αλλαγή. Αυτό το προφίλ, η **βασική εικόνα** μιας περιοχής, είναι ένας δείκτης που εκφράζει την ελκτικότητα της περιοχής στους επενδυτές και στους εργαζομένους. Έτσι, αν η βασική εικόνα μιας περιοχής είναι καλή, τότε αυτή η περιοχή είναι πόλος έλξης επενδύσεων και εργατικού δυναμικού ενώ αντίθετα αν είναι κακή, η προσέλευση πόρων σ΄ αυτήν είναι φτωχή. Η βασική εικόνα είναι η συνισταμένη της οικονομικής και της κοινωνικής εικόνας της περιοχής. Και η οικονομική και η κοινωνική εικόνα είναι αντίστοιχα οι συνισταμένες μιας σειράς παραγόντων όπως αυτοί οι οποίοι παρατίθενται προς στάθμιση στο ερωτηματολόγιο.

Η συμπλήρωση του ερωτηματολογίου αυτού θα αποτελέσει εξαιρετικά πολύτιμη συμβολή στην επιτυχία της έρευνας μας που αφορά τις ευαίσθητες ακριτικές περιοχές της πατρίδας μας.

Παρατήρηση: Αν κρίνετε ότι πρέπει σε κάποια ερώτηση να προσθέσετε κάποιον παράγοντα που θεωρείται σημαντικό μπορείτε να το κάνετε.

ΕΡΩΤΗΣΕΙΣ

1. Στην επιλογή του τόπου εργασίας – εγκατάστασής σας πόσο βαρύνουν οι ακόλουθοι παράγοντες. (μοιράστε 100 μονάδες στους παρακάτω παράγοντες).

α) Οικονομική εικόνα της περιοχής.
β) Κοινωνική εικόνα της περιοχής.

2. Στη διαμόρφωση της <u>οικονομικής εικόνας</u> μιας περιοχής πόσο βαρύνουν οι παρακάτω παράγοντες. (μοιράστε 100 μονάδες στους παρακάτω παράγοντες).

α) Σχετική γεωγραφική θέση του νομού εγκατάστασης ως προς τις κυριότερες αγορές και πηγές εισροών για το νομό.
β) Η διαθεσιμότητα εργατικού δυναμικού και κεφαλαίων.
γ) Οι οικονομικές συνθήκες στην ευρύτερη περιφέρεια.

3. Πόσο βαρύνει η <u>προσέγγιση</u> – ευκολία πρόσβασης <u>στις κυριότερες αγορές</u> και πόσο <u>στις πηγές εισροών</u> για το νομό. (μοιράστε 100 μονάδες στους παρακάτω παράγοντες).

α) προσέγγιση σε αγορές.
β) προσέγγιση σε πηγές εισροών (πρώτων υλών).

4. Σε ότι αφορά τη <u>διαθεσιμότητα πόρων</u> πόσο βαρύνουν οι παρακάτω πόροι (μοιράστε 100 μονάδες στους παρακάτω παράγοντες)

α) Διαθέσιμο εργατικό δυναμικό.
β) Διαθέσιμα κεφάλαια.

5. Στη διαμόρφωση των <u>οικονομικών συνθηκών στην ευρύτερη περιφέρεια</u> τι βαρύτητα δίνετε στους παρακάτω παράγοντες. (μοιράστε 100 μονάδες στους παρακάτω παράγοντες).

α) Επενδύσεις σε έργα υποδομής.
β) Κατά κεφαλήν Α.Ε.Π.

6. Στη διαμόρφωση της κοινωνικής εικόνας μιας περιοχής πόσο βαρύνουν οι παρακάτω παράγοντες. (μοιράστε 100 μονάδες στους παρακάτω παράγοντες).
α) Στεγαστικές συνθήκες περιοχής.
β) Ποιότητα περιβάλλοντος περιοχής.
γ) Συνθήκες παρεχόμενης υγειονομικής περίθαλψης περιοχής.
δ) Συνθήκες παρεχόμενης εκπαίδευσης περιοχής.
ε) Δυνατότητες ψυχαγωγίας περιοχής.
στ. Πολιτιστικές συνθήκες περιοχής.

7. Στη διαμόρφωση των στεγαστικών συνθηκών τι βαρύτητα δίνετε στους παρακάτω παράγοντες. (μοιράστε 100 μονάδες στους παρακάτω παράγοντες.
α) Ποσοτική επάρκεια κατοικιών.
β) Ποιότητα κατοικιών.
γ) Μέσος αριθμός δωματίων ανά κατοικία.

8. Στη διαμόρφωση των συνθηκών της ποιότητας περιβάλλοντος της περιοχής τι βαρύτητα δίνετε στους παρακάτω παράγοντες. (μοιράστε 100 μονάδες στους παρακάτω παράγοντες).
α) Υποβάθμιση δομημένου περιβάλλοντος.
β) Ατμοσφαιρική ρύπανση.
γ) Ρύπανση λιμνών και ποταμών.

9. Στη διαμόρφωση των συνθηκών της παρεχόμενης υγειονομικής περίθαλψης της περιοχής τι βαρύτητα δίνετε στους παρακάτω παράγοντες. (μοιράστε 100 μονάδες στους παρακάτω παράγοντες).
α) Αναλογία κατοίκων ανά γιατρό.
β) Αναλογία κατοίκων ανά διαθέσιμες κλίνες.

10. Στη διαμόρφωση των συνθηκών της παρεχόμενης εκπαίδευσης της περιοχής τι βαρύτητα δίνετε στους παρακάτω παράγοντες. (μοιράστε 100 μονάδες στους παρακάτω παράγοντες).

$$\text{A.}\ \left\{\begin{array}{l}\text{Αναλογία μαθητών προσχολικού επιπέδου ανά διδάσκοντα}\\ \text{Αναλογία μαθητών δημοτικού επιπέδου ανά διδάσκοντα}\\ \text{Αναλογία μαθητών μέσου επιπέδου ανά διδάσκοντα.}\end{array}\right\}\ \text{Μαθ./διδασκοντα.}$$

Β. Αναλογία μαθητών (προσχ.-δημοτ.-μεσ./ανα διδακτήριο Μαθ./διδακτήριο.

11. Στη διαμόρφωση των δυνατοτήτων ψυχαγωγίας της περιοχής τι βαρύτητα δίνετε στους παρακάτω παράγοντες. (μοιράστε 100 μονάδες στους παρακάτω παράγοντες).
α) Αριθμός θεάτρων.
β) Αριθμός κινηματογράφων.
γ) Αριθμός κέντρων ψυχαγωγίας (εστιατόρια, bar, κ.τ.λ.).

12. Στη διαμόρφωση των πολιτιστικών συνθηκών της περιοχής τι βαρύτητα δίνετε στους παρακάτω παράγοντες (μοιράστε 100 μονάδες στους παρακάτω παράγοντες.
α. Αριθμό βιβλιοθηκών.
β. Αριθμό Συνεδρίων.
γ. Αριθμό Μουσείων.

2. Απογραφή και δειγματοληψία

Η συγκέντρωση του στατιστικού υλικού γίνεται με **απογραφή** όλου του πληθυσμού σε ορισμένες μεταβλητές του, όπως η απογραφή των βιομηχανικών επιχειρήσεων της χώρας σε μια συγκεκριμένη χρονιά, η απογραφή των εμπορικών καταστημάτων μιας περιφέρειας μια συγκεκριμένη χρονιά, ο γεωργικός πληθυσμός κατά νομό μια χρονική περίοδο, τα καλλιεργούμενα στρέμματα ανά είδος καλλιέργειας κλπ.

Ο τρόπος αυτός συγκέντρωσης του στατιστικού υλικού απαιτεί χρόνο, αρκετά μέσα και κυρίως άτομα για να φέρουν σε πέρας την μεγάλη σε έκταση αυτή δουλειά, μεγάλο κόστος και είναι χρονοβόρα η επεξεργασία των στοιχείων και η διάθεση των αποτελεσμάτων. Παρόλα αυτά είναι η μόνη έρευνα που μας δίνει καθαρά την σύνθεση και τη δομή του πληθυσμού που ερευνούμε. (ΕΣΥΕ ανά 10 έτη)

Αρκετές φορές λόγω των παραπάνω αναφερόμενων δυσκολιών στην απογραφή του πληθυσμού καταφεύγουμε στην εξέταση ενός τμήματος αυτού του πληθυσμού που καλείται **δείγμα.** Η επιλογή του δείγματος μπορεί να γίνει κατά διάφορους τρόπους. Ο τρόπος δουλειάς στην περίπτωση αυτή είναι να προσδιορίσουμε τα χαρακτηριστικά γνωρίσματα του συνολικού πληθυσμού μέσα από την στατιστική εξέταση των αντιστοίχων χαρακτηριστικών γνωρισμάτων του δείγματος και αυτό γίνεται μέσα από την επαγωγική διαδικασία που ακολουθείται, βέβαια εδώ παίζει καθοριστικό ρόλο το δείγμα που λαμβάνουμε.

Αν το δείγμα είναι **τυχαίο** και έχουμε μια τυχαία δειγματοληψία προφανώς και δεν μπορούμε τα όποια συμπεράσματα που θα εξάγουμε από την στατιστική μελέτη του δείγματος να τα γενικεύσουμε για ολόκληρο τον πληθυσμό, όταν όμως το δείγμα είναι **αντιπροσωπευτικό** τότε και η γενίκευση των συμπερασμάτων επί του συνολικού πληθυσμού μπορεί να γίνει.

Η αντιπροσωπευτικότητα του δείγματος είναι βασική προϋπόθεση για να έχουμε πειστικά και αξιόπιστα συμπεράσματα επί του συνολικού πληθυσμού. Συνήθως δειγματοληψίες με προϋποθέσεις ή **συστηματική** δειγματοληψία (όχι τυχαία) κάνουμε όταν εξετάζουμε αντιπροσωπευτικό δείγμα σε επιχειρήσεις, όπως πχ. του αγροτοβιομηχανικού κλάδου μεταποίησης αγροτικών προϊόντων, τροφίμων, ποτών, μηχανημάτων κλπ. αφού πρώτα τις εντάξουμε σε κατηγορίες ανάλογα με το παραγόμενο προϊόν, την εξειδίκευσή τους, τον τζίρο τους (κύκλο εργασιών), την χωροταξική τους κατανομή κλπ. είτε όταν εξετάζουμε τον πληθυσμό μιας χώρας, μιας ευρύτερης γεωγραφικά πληθυσμιακής περιοχής (π.χ Ευρώπη), για να γνωρίσουμε τις απόψεις του πληθυσμού πάνω σε θέματα οικονομικά, κοινωνικά, πολιτιστικά, πολιτικά κλπ.

Η συστηματική δειγματοληψία αναφέρεται επίσης σε θέματα έρευνας αγοράς (μάρκετινγκ) που αφορούν την γνώμη του κοινού για ένα προϊόν που κυκλοφορεί ή τις επιθυμίες και τις ανάγκες του αγοραστικού κοινού για να μπει ένα νέο προϊόν στην αγορά ή να χρησιμοποιηθεί η διαφήμιση για την προώθηση ενός προϊόντος κλπ.

Τα **είδη** του **δείγματος** μπορούν τελικά να διακριθούν στα εξής:

1. **απλό τυχαίο δείγμα**
Στο απλό τυχαίο δείγμα η πιθανότητα να επιλεγεί μια στατιστική μονάδα του πληθυσμού στο δείγμα είναι η ίδια για όλες τις στατιστικές μονάδες.
2. **τυχαίο συστηματικό, περιοδικό ή ίσου διαστήματος δείγμα**
Στο συστηματικό τυχαίο δείγμα επιλέγουμε ένα ποσοστό επί τοις εκατό του συνολικού πληθυσμού, πχ. επιλέγουμε το 10% σε ένα πλήθος 30 στατιστικών μονάδων οπότε επιλέγουμε τυχαία 3 στατιστικές μονάδες και η επιλογή γίνεται:
Η κάθε τελεία **.** αφορά μια στατιστική μονάδα.
Το * (αστεράκι) αφορά την επιλεγόμενη στατιστική μονάδα.

Είτε **τυχαία**
..*.....*.
.....*....
....

Είτε **συστηματική όχι τυχαία**
.........* (δέκατη θέση)
.........*
.........*

είτε **συστηματική τυχαία**
......*... (έβδομη θέση)
......*...
......*...

είτε **τυχαία για την πρώτη στατιστική μονάδα**
......*...

3. τυχαίο δείγμα σε στρώματα

Χρησιμοποιείται όταν ο πληθυσμός είναι κατανεμημένος σε ομάδες πληθυσμιακές που παρουσιάζουν ομοιογένεια ως προς τα κοινωνικά, τα οικονομικά, τα δημογραφικά, τα πολιτισμικά κλπ. χαρακτηριστικά.

Τότε επιλέγεται τυχαία ένα δείγμα από κάθε ομάδα και το σύνολο των δειγμάτων αποτελεί το δείγμα που θα μελετηθεί. Αν ο συνολικός πληθυσμός παρουσιάζει μεγάλη ανομοιογένεια θα χρειασθούμε μεγάλο δείγμα για να είναι αντιπροσωπευτικό του πληθυσμού.

4. τυχαίο δείγμα κατά περιοχή (δειγματοληψία σε φάσεις)

Μια όμοια με την δειγματοληψία κατά στρώματα είναι η μέθοδος της δειγματοληψίας σε φάσεις. Αυτή εντάσσεται στην δειγματοληψία κατά στρώσεις και αφορά κυρίως δειγματοληψία κατά περιοχές.

Αναφέρεται παρακάτω ένα παράδειγμα αυτής της δειγματοληψίας. Έστω ότι θέλουμε να εξετάσουμε ένα δείγμα νοικοκυριών ως προς κάποια χαρακτηριστικά γνωρίσματά τους από ένα πληθυσμό 400 δήμων, όπως αυτοί κατατάχθηκαν με κάποια κριτήρια.

Ακολουθούμε τα εξής βήματα:

Βήμα 1^{ο} : 400 δήμοι
Επιλέγω ένα τυχαίο δείγμα 2% αυτών των δήμων = 8 δήμοι

Βήμα 2^{ο} : 8 δήμοι = 70 εκλογικά τμήματα
Επιλέγω ένα τυχαίο δείγμα 10% αυτών των τμημάτων = 7 τμήματα

Βήμα 3^{ο} : 7 τμήματα= 25.000 νοικοκυριά
Επιλέγω ένα τυχαίο δείγμα 5% αυτών των νοικοκυριών =1250 νοικοκυριά

5. τυχαίο δείγμα κατά κλάσεις-τμήματα

Αρκετές φορές οι ερωτώμενοι αδιαφορούν να δώσουν απαντήσεις και είναι χαμηλό το ποσοστό αυτών που ανταποκρίνονται και μέσω τηλεφωνικής επικοινωνίας.

Επίσης ακόμα και αν είναι με αμοιβή αυτός που αναλαμβάνει την συγκέντρωση των απαντήσεων μπροστά στην αδιαφορία των ερωτώμενων μπορεί να στραφεί στους γείτονές του, στους συγγενείς του, και να λείπουν από την ομάδα των ερωτώμενων οι νέοι, οι ηλικιωμένοι κλπ.

Έτσι δεν θα υπάρχει μια διαστρωμάτωση του δείγματος κατά φύλο, ηλικία, επάγγελμα, κοινωνική τάξη κλπ., αυτό πρέπει να προσεχθεί ιδιαίτερα. Δύο παραδείγματα τυχαίου δείγματος με διαστρωμάτωση είναι τα ακόλουθα.

Πίνακας 3.

Ηλικία	**Φύλο**	
	Άρρεν	**θήλυ**
Κάτω των 15	0	0
15-29	4	4
30-44	9	9
45-59	7	7
60 και άνω	5	5
Σύνολο	**25**	**25**

Πίνακας 4.

Κοινωνική-εθνική Ομάδα	**Καυκάσια**	**Αφρικανική**	**Ασιατική**
Επάγγελμα			
Διευθυντικό στέλεχος	8	2	2
Διοικητικός	10	3	2
Εργάτης	10	3	1
Συνταξιούχος	6	2	1
ΣΥΝΟΛΟ	**34**	**10**	**6**

6. τυχαίο δείγμα χαρακτηριστικών

Είναι απλή μέθοδο και αναφέρεται σε άτομα που έχουν ένα συγκεκριμένο χαρακτηριστικό γνώρισμα παραδείγματος χάριν αυτοί που γεννήθηκαν τον Απρίλιο μιας συγκεκριμένης χρονιάς. Οι πληροφορίες που μπορούμε να έχουμε σχετικά με αυτήν την ομάδα είναι για το βάρος τους, το ύψος τους, την πρόοδό τους και την επίδοσή τους στα μαθήματα κλπ. και αυτές τις πληροφορίες μπορούμε να τις χρησιμοποιήσουμε για εκπαιδευτικά θέματα, για να αποτιμήσουμε τον δείκτη ευφυΐας τους και αλλού.

Κεφάλαιο 2^{o}

Ταξινόμηση, κατάταξη και παρουσίαση των στατιστικών δεδομένων

2.1. Πίνακες κατανομής συχνοτήτων - σχετικών συχνοτήτων

Θεωρούμε ένα δείγμα ν =20 αγροτικών οικογενειών και μελετάμε το δείγμα αυτό ως προς την μεταβλητή Χ «**ετήσιο εισόδημα σε εκατομμύρια δραχμές**». Τα αποτελέσματα καταγράφονται ως εξής:

2	3	2	3	3	2	4
4	2	3	5	5	4	3
5	6	3	2	4	3	

Οι τιμές της μεταβλητής Χ είναι οι x_i, i = 1, 2, 3, 4, 5

$x_1 = 2$, $x_2 = 3$, $x_3 = 4$, $x_4 = 5$, $x_5 = 6$.

Σε κάθε τιμή x_i αντιστοιχεί ο φυσικός v_i που δείχνει πόσες φορές εμφανίζεται η τιμή x_i της μεταβλητής Χ στο σύνολο των παρατηρήσεων.

Παραδείγματος χάριν η $x_1 = 2$ εμφανίζεται $v_1 = 5$ φορές.

Ο αριθμός v_i καλείται (απόλυτη) **συχνότητα** της τιμής x_i (της παρατήρησης).

Έτσι θα έχουμε: $v_1+v_2+ \ldots +v_5 = v$

Επίσης αν διαιρέσουμε τη συχνότητα v_i με το μέγεθος του δείγματος το ν δημιουργούμε τη **σχετική συχνότητα** fi της τιμής (παρατήρησης) x_i.

Ήτοι: $$fi = \frac{v_i}{v} \quad i = 1,2,...,5$$

Η σχετική συχνότητα εκφράζεται σε δεκαδική μορφή και επί τοις εκατό % ( ).

$$f_\iota \% = \frac{v_i}{v} \cdot 100$$

Ισχύουν για τη σχετική συχνότητα οι σχέσεις:

1. $0 \le f_i \le 1 \quad , \quad i = 1,2,...,5$
2. $f_i + f_2 + f_3 + f_4 + f_5 = 1$

Γενικά αν έχουμε $x_1, x_2, ..., x_\kappa$ τιμές της μεταβλητής Χ με συχνότητες $v_1, v_2, ..., v_\kappa$ $(\kappa \le v)$. Θα έχουμε τις σχετικές συχνότητες $f_i, i = 1,....,\kappa$

Και θα ισχύουν:

$$0 \le fi \le 1 \quad , \quad x_1, x_2, ..., x_\kappa$$

$$\sum_{i=1}^{k} f_i = 1 \qquad (f_1 + f_2 + ... + f_\kappa = 1)$$

Ο Πίνακας Κατανομής συχνοτήτων – σχετικών συχνοτήτων είναι ο ακόλουθος για το εν λόγω παράδειγμα:

Πίνακας 1.

x_i	v_i	f_i	$f_i\%$
2	5	0,25	25,0
3	7	0,35	35,0
4	4	0,20	20,0
5	3	0,15	15,0
6	1	0,05	5,0
	20	1,00	100,0

2.2. Αθροιστική συχνότητα – Σχετική αθροιστική συχνότητα

Αρκετές φορές θέλουμε να μάθουμε το πλήθος ή το ποσοστό των παρατηρήσεων που οι τιμές τους είναι μικρότερες ή ίσες μιας ορισμένης τιμής x_i της μεταβλητής X.

Π.χ. αν θέλουμε να πληροφορηθούμε στο προηγούμενο παράδειγμα τον αριθμό των αγροτικών οικογενειών που έχουν ετήσιο εισόδημα κάτω από 4 εκατ. δρχ. ή το ποσοστό των αγροτικών οικογενειών που έχουν ετήσιο εισόδημα κάτω των 4 εκατ. δραχμών θα χρειασθεί να προσθέσουμε την πρώτη φορά: 5+7 = 12 οικογένειες και την δεύτερη φορά 25+35= 60 % των οικογενειών.

Αυτό που βρίσκουμε καλείται **αθροιστική συχνότητα** την πρώτη φορά και συμβολίζεται με Nj και την δεύτερη φορά **Σχετική Αθροιστική συχνότητα** και συμβολίζεται με Fj.

Δηλαδή έχουμε:

$$Nj = \sum_{i=1}^{j} v_i = v_1 + v_2 + ... + v_j, \qquad 1 \le j \le \kappa$$

$$fj = \sum_{i=1}^{j} f_i = f_1 + f_2 + ... + fj, \qquad 1 \le j \le \kappa$$

Πίνακας 2.

x_i	v_i	N_i	f_i	f_i %	F_j	F_j %
2	5	5	0,25	25	0,25	25
3	7	12	0,35	35	0,60	60
4	4	16	0,20	20	0,80	80
5	3	19	0,15	15	0,95	95
6	1	20	0,05	5	1,00	100
	20	---	1,00	100	---	---

Το F_j εκφράζεται και επί τις εκατό %
F_j % = F_j 100

2.3. Ομαδοποίηση των παρατηρήσεων

Σε περίπτωση όπου έχουμε διακριτή μεταβλητή που παίρνει πολλές τιμές ή όταν έχουμε συνεχή μεταβλητή που παίρνει οποιαδήποτε τιμή στο διάστημα ορισμού της, τότε χωρίζουμε τα δεδομένα σε μικρό αριθμό **κλάσεων**, ώστε κάθε τιμή να ανήκει μόνο σε μια κλάση.

Οι κλάσεις είναι διαστήματα της μορφής [,) δηλαδή μια κλάση περιέχει το κάτω άκρο της αλλά όχι το άνω άκρο της. Κάθε κλάση αντιπροσωπεύεται από την κεντρική της τιμή x_i. Το κέντρο της κλάσης και οι παρατηρήσεις σε κάθε κλάση θεωρούνται όμοιες. Το πλάτος της κλάσης c θεωρείται σταθερό οπότε έχουμε κλάσεις ίσου πλάτους ή μπορεί να έχουμε κλάσεις άνισου πλάτους.

Στην περίπτωση κλάσεων ίσου πλάτους βρίσκουμε το πλάτος της κλάσης διαιρώντας **το εύρος** R του δείγματος με τον αριθμό των κλάσεων $\kappa \cdot \left(c = \frac{R}{\kappa} \right)$

Ο αριθμός κ μπορεί να ληφθεί από τον πίνακα που παρατίθεται:

Πίνακας 3.

Μέγεθος δείγματος ν	Αριθμός κ	Μέγεθος δείγματος ν	Αριθμός κ
<20	5	200- 400	9
20- 50	6	400- 700	10
50-100	7	700-1000	11
100-200	8	≥ 1000	12

ή μπορεί να υπολογισθεί από τον τύπο:
$K = 1 + 3{,}32 \log v$ (Κανόνας Sturges) όπου ν το μέγεθος του δείγματος.

Το **εύρος R** = **Τιμή της μεγαλύτερης παρατήρησης - Τιμή της μικρότερης παρατήρησης**.

Για να φτιάξουμε τις κλάσεις πρέπει να ξεκινήσουμε λίγο πιο κάτω από την μικρότερη παρατήρηση και να προσθέτουμε κάθε φορά το πλάτος c της κλάσης και η μεγαλύτερη τιμή πρέπει να περιέχεται στην ανώτερη κλάση.

Ως **συχνότητα της κλάσης** i θεωρούμε το πλήθος των παρατηρήσεων v_i που βρίσκονται μέσα στην κλάση και καλείται και **συχνότητα της κεντρικής τιμής x_i της κλάσης i.**

Παράδειγμα:

Ένας αγρότης σημείωσε τον αριθμό των αυγών που συγκέντρωσε σε 40 ημέρες, και πήρε τα ακόλουθα αποτελέσματα:

15	16	16	18	18	18	20	20	20	23	23	23	25	25	25
30	30	30	35	35	38	38	38	39	43	43	43	44	44	44
44	49	49	49	49	50	50	50	50	50					

Έχουμε ν = 40 άρα ο αριθμός των κλάσεων $\kappa = 1 + 3{,}3 \log 40 = 6{,}28 \rightarrow \kappa \cong 6$ (σύμφωνα και με τον πίνακα 3).

$R = 50 - 15 = 35$

Συνεπώς $C = \dfrac{R}{\kappa} = \dfrac{35}{6} \approx 6$

Και οι κλάσεις που δημιουργούνται είναι:

$[15,21)$, $[21,27)$, $[27,33)$, $[33,39)$, $[39,45)$, $[45,51)$.

Ο Πίνακας συχνοτήτων – σχετικών συχνοτήτων – αθροιστικών συχνοτήτων και σχετικών αθροιστικών συχνοτήτων είναι ο εξής:

Πίνακας 4.

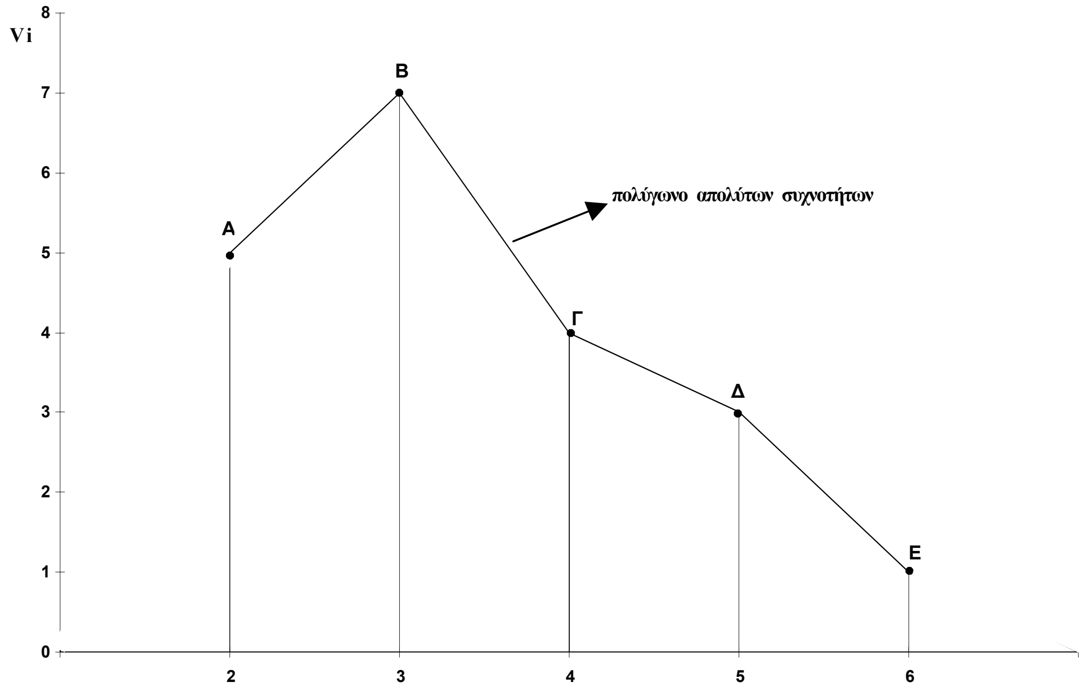
Κλάσεις [,)	**Κεντρικές τιμές x_i**	**Συχν. v_i**	**Σχετ. Συχν. f_i**	**Σχετ. Συχν. f_i %**	**Αθρ. Συχν. N_i**	**Αθρ. Σχ. συχν. F_i**	**Αθρ.Σχετ. Συχ. F_i %**
15-21	18	9	0,225	22,5	9	0,225	22,5
21-27	24	6	0,150	15,0	15	0,375	37,5
27-33	30	3	0,075	7,5	18	0,450	45,0
33-39	36	6	0,150	15,0	24	0,600	60,0
39-45	42	7	0,175	17,5	31	0,775	77,5
45-51	48	9	0,225	22,5	40	1,000	100,0
		40	1,000	100,0			

2.4. Γραφική παράσταση μιας κατανομής συχνοτήτων

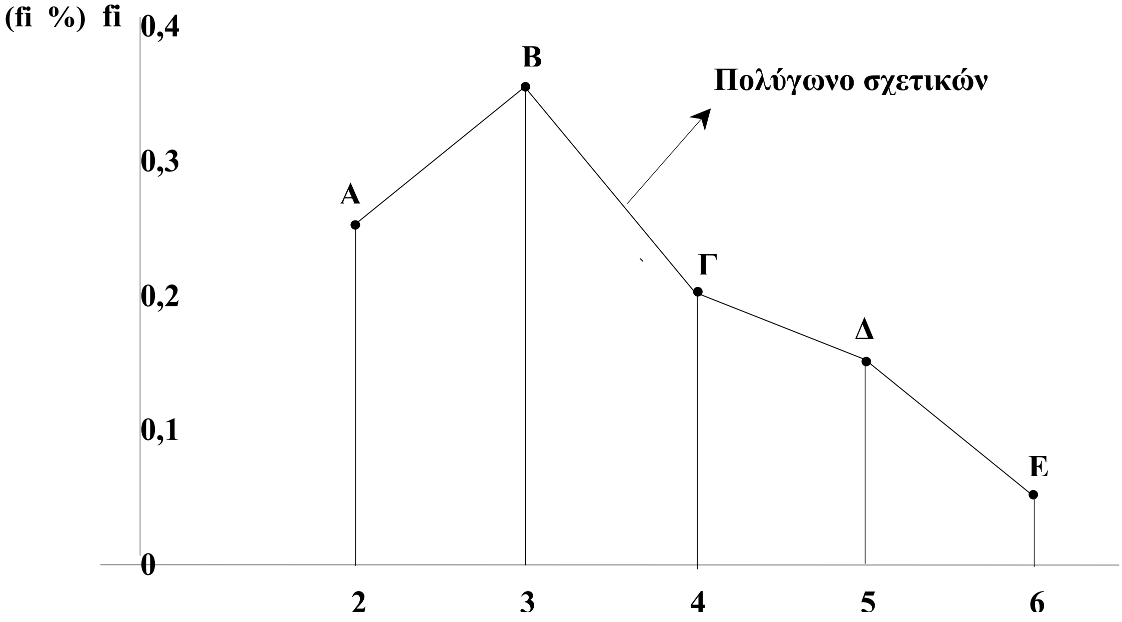
(α) Διαγράμματα – Ιστογράμματα.
ι/ Μη ομαδοποιημένες παρατηρήσεις

Διάγραμμα και πολύγωνο απολύτων – σχετικών συχνοτήτων

Στο διάγραμμα σε κάθε τιμή x_i της μεταβλητής X έχουμε υψώσει κάθετη γραμμή ύψους ίσο με την αντίστοιχη απόλυτη συχνότητα v_i της τιμής x_i.



Διάγραμμα απολύτων συχνοτήτων στο παράδειγμα των 20 αγροτικών οικογενειών με ετήσια εισοδήματα.

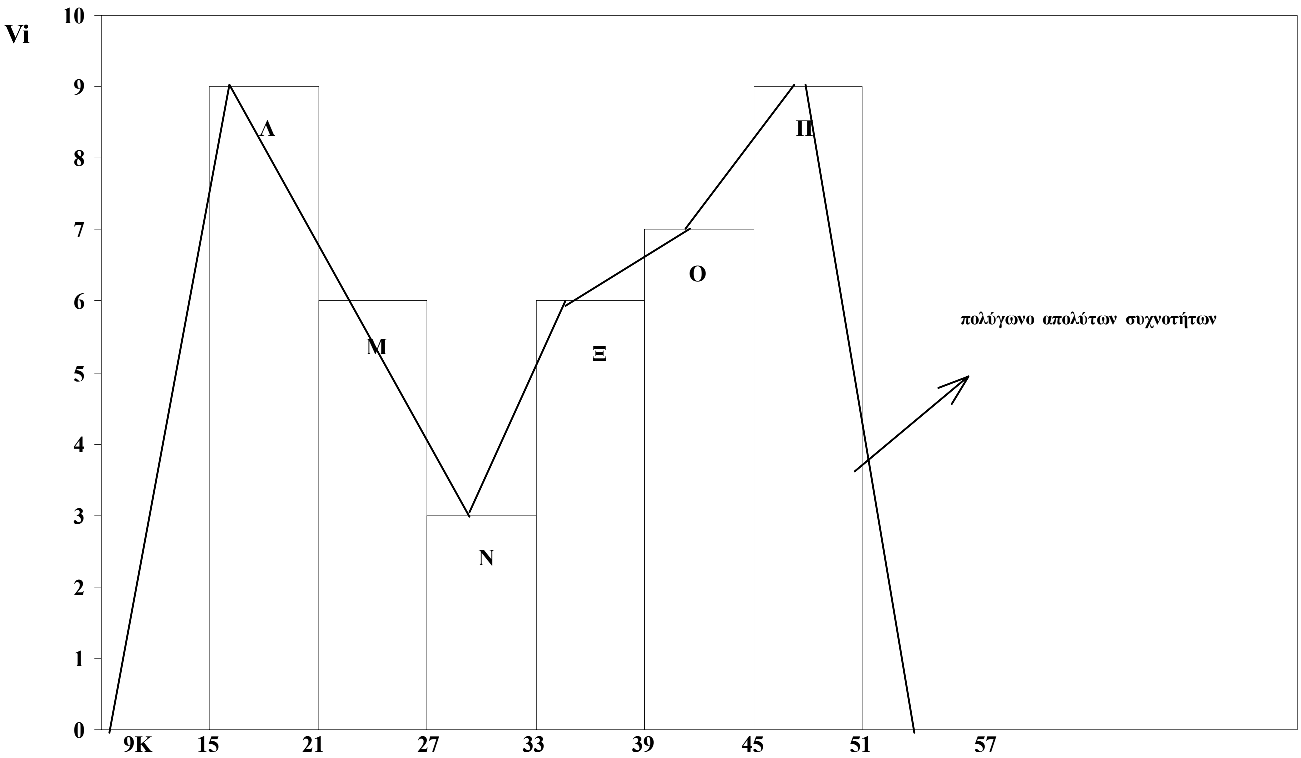


Διάγραμμα σχετικών συχνοτήτων

ii. Ομαδοποιημένες παρατηρήσεις σε κλάσεις ίσου πλάτους.

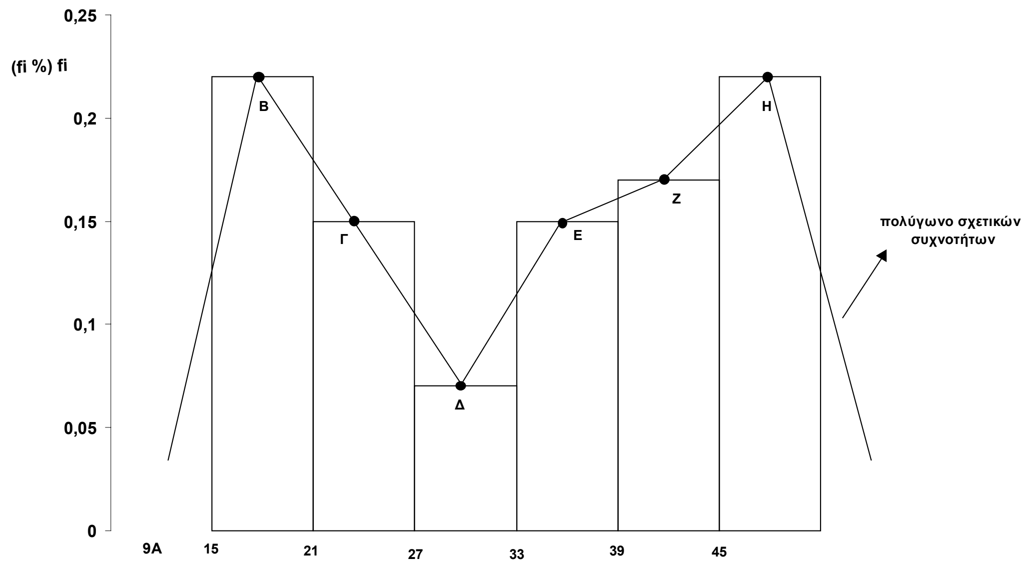
Ιστόγραμμα και πολύγωνο απολύτων και σχετικών συχνοτήτων.

Στο Ιστόγραμμα στον οριζόντιο άξονα σημειώνουμε τα όρια των κλάσεων και στη συνέχεια κατασκευάζουμε ορθογώνια (Ιστούς) καθένα από τα οποία έχει βάση ίση με το πλάτος της κλάσης και ύψος τέτοιο ώστε **το εμβαδόν του Ορθογωνίου να ισούται με τη συχνότητα της κλάσης αυτής.**
Σύμφωνα με το παράδειγμα της καταμέτρησης των αυγών έχουμε:



Ιστόγραμμα απόλυτων συχνοτήτων

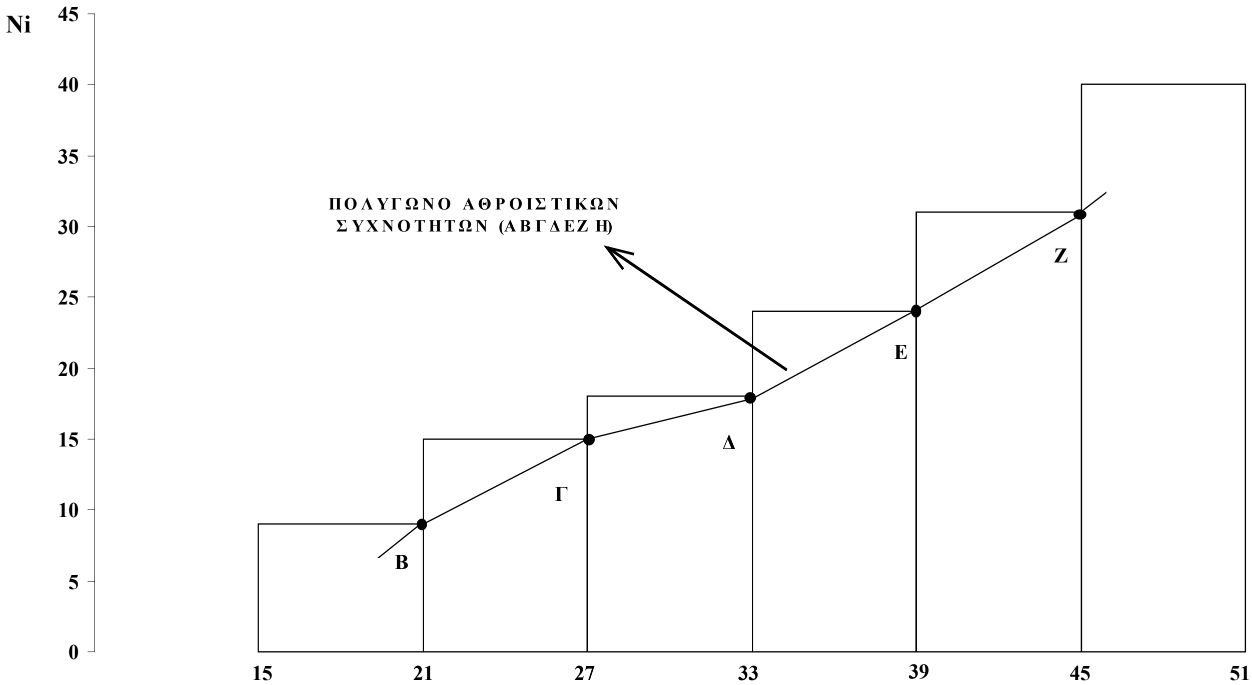
Το πολύγωνο των απολύτων συχνοτήτων κατασκευάζεται αν πάρουμε δύο υποθετικές κλάσεις στην αρχή και στο τέλος [9,15) και [51,57) με συχνότητα μηδέν και ενώσουμε τα μέσα των άνω βάσεων των ορθογωνίων ΛΜΝΞΟΠ με τα μέσα Κ,Ρ των υποθετικών κλάσεων. Το εμβαδόν του χωρίου που ορίζεται από το πολύγωνο ΚΛΜΝΞΟΠΡ και την οριζόντιο είναι ίσο με το άθροισμα των απολύτων συχνοτήτων δηλαδή ίσο με το μέγεθος του δείγματος



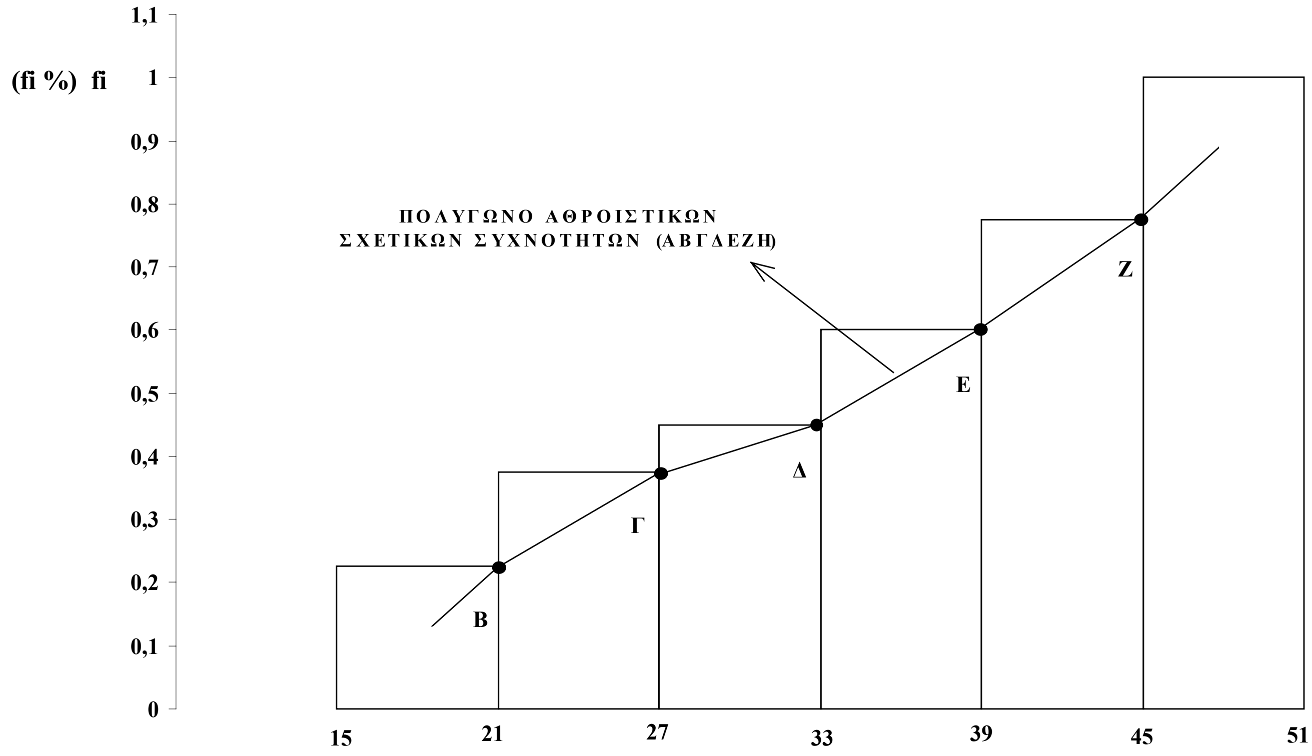
Ιστόγραμμα σχετικών συχνοτήτων.

Όμοια κατασκευάζεται και το πολύγωνο των σχετικών συχνοτήτων ΑΒΓΔΕΖΗΘ. Το εμβαδόν του χωρίου μεταξύ του πολυγώνου ΑΒΓΔΕΖΗΘ και του άξονα Χ ισούται με 1 (άθροισμα των σχετικών συχνοτήτων, όταν εκφράζονται σε δεκαδική μορφή) ή 100 (όταν είναι επί τοις εκατό).

Ιστόγραμμα και πολύγωνο αθροιστικών συχνοτήτων και αθροιστικών σχετικών συχνοτήτων.



Ιστόγραμμα αθροιστικών συχνοτήτων.



Ιστόγραμμα αθροιστικών σχετικών συχνοτήτων.

Το πολύγωνο των αθροιστικών συχνοτήτων προκύπτει αν ενώσουμε τα δεξιά άκρα των άνω βάσεων των ορθογωνίων με ευθύγραμμα τμήματα.

(β) Ραβδογράμματα

Χρησιμοποιούνται κυρίως σε ποιοτικές μεταβλητές.

Το ραβδόγραμμα περιέχει ορθογώνια των οποίων οι βάσεις καθορίζονται αυθαίρετα και τα ύψη αντιστοιχούν στη συχνότητα ή σχετική συχνότητα της παρατήρησης.

Η απόσταση μεταξύ των ορθογωνίων είναι αυθαίρετη.

Παράδειγμα.

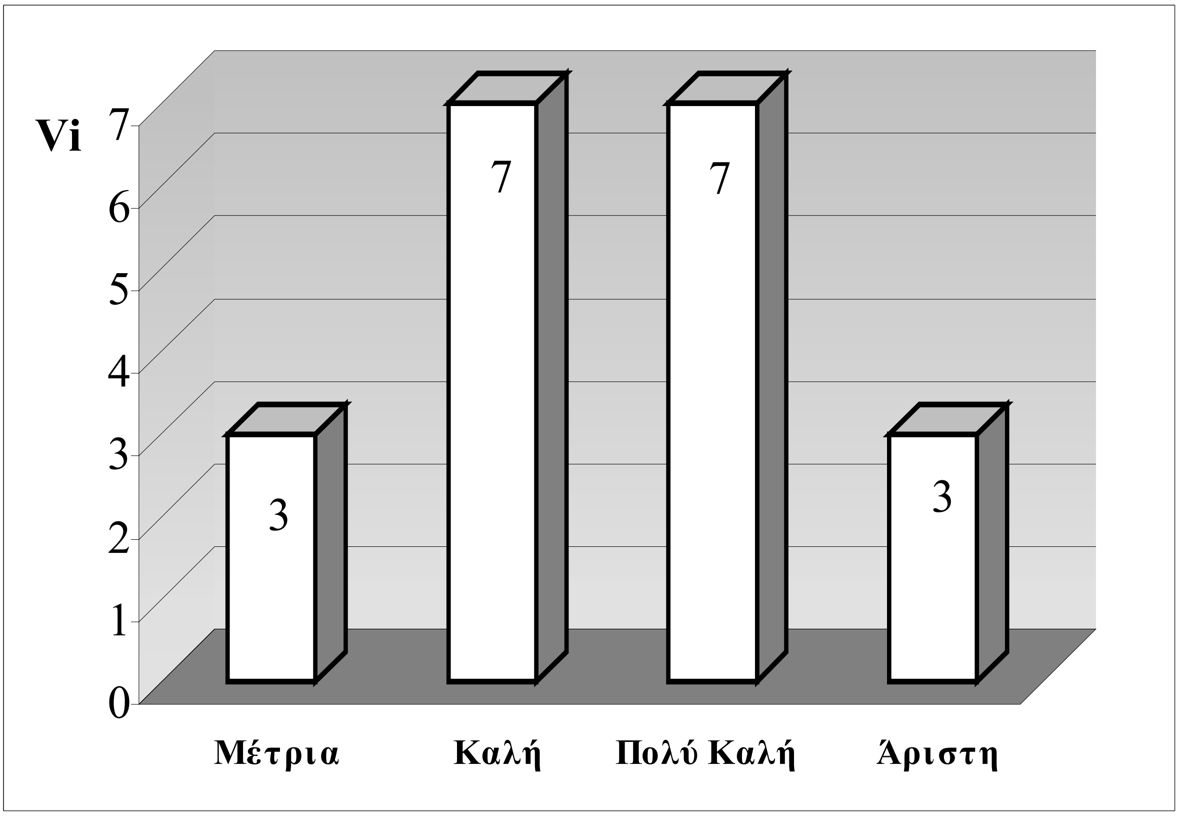
Εξετάσθηκαν 20 καλλιεργημένες εκτάσεις βάμβακος ως προς την ποιότητα του παραγόμενου προϊόντος.

Ως κλίμακα ποιότητας θεωρήθηκε η ακόλουθη: μέτρια-καλή- πολύ καλή, άριστη, (παρατηρήσεις της ποιοτικής μεταβλητής Χ «**ποιότητα βάμβακος**») και προέκυψαν τα αποτελέσματα που δίνεται στον παρακάτω πίνακα.

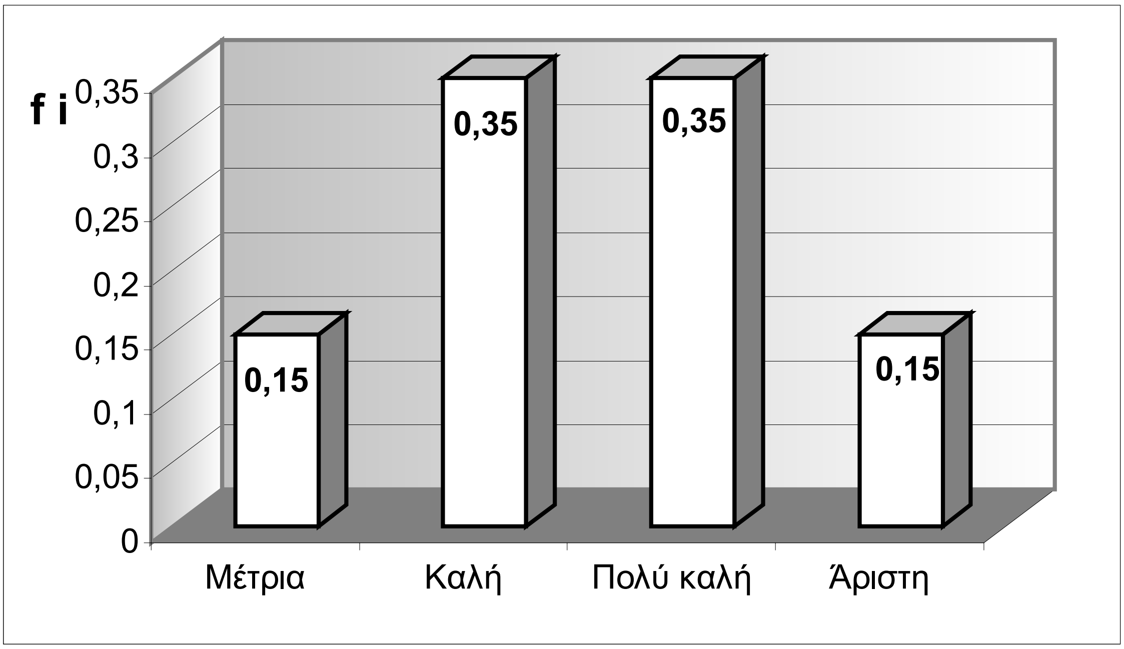
Ο πίνακας συχνότητας – σχετικής συχνότητας αθροιστικής συχνότητας – αθροιστική σχετικής συχνότητας

x_i	v_i	f_i	f_i%	N_i	F_i	F_i %
Μέτρια	3	0,15	15	3	0,15	15
Καλή	7	0,35	35	10	0,50	50
Πολύ καλή	7	0,35	35	17	0,85	85
Άριστη	3	0,15	15	20	1	100
	20	1	100			

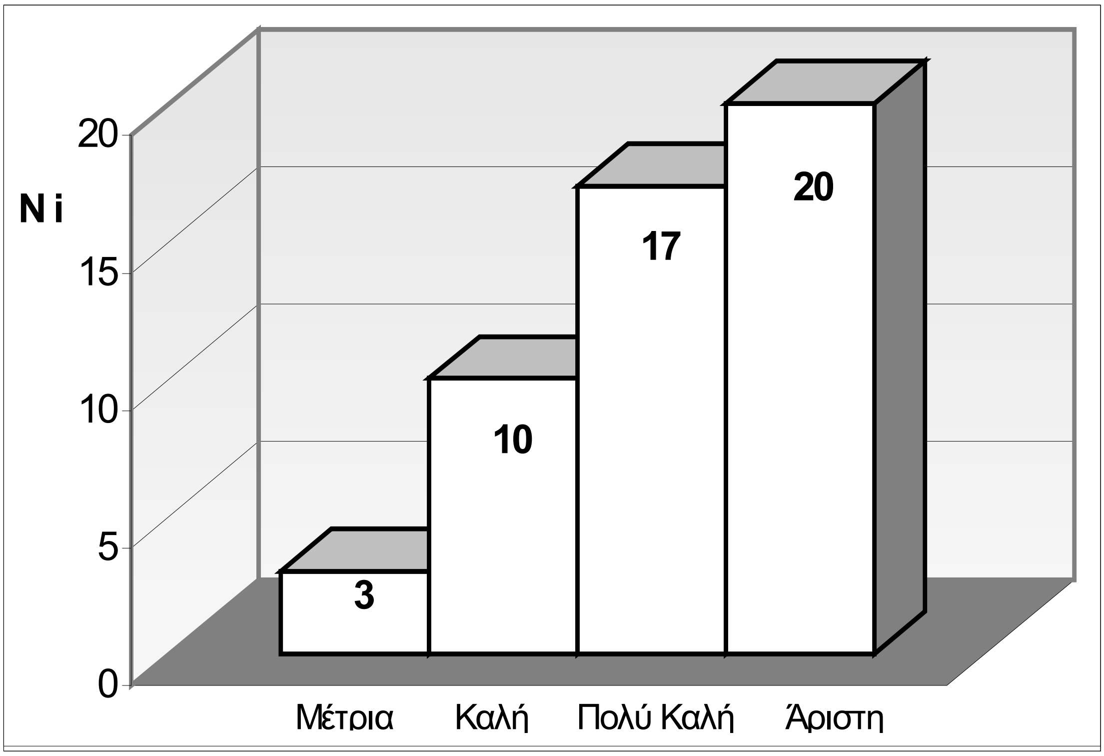
Ραβδόγραμμα συχνοτήτων



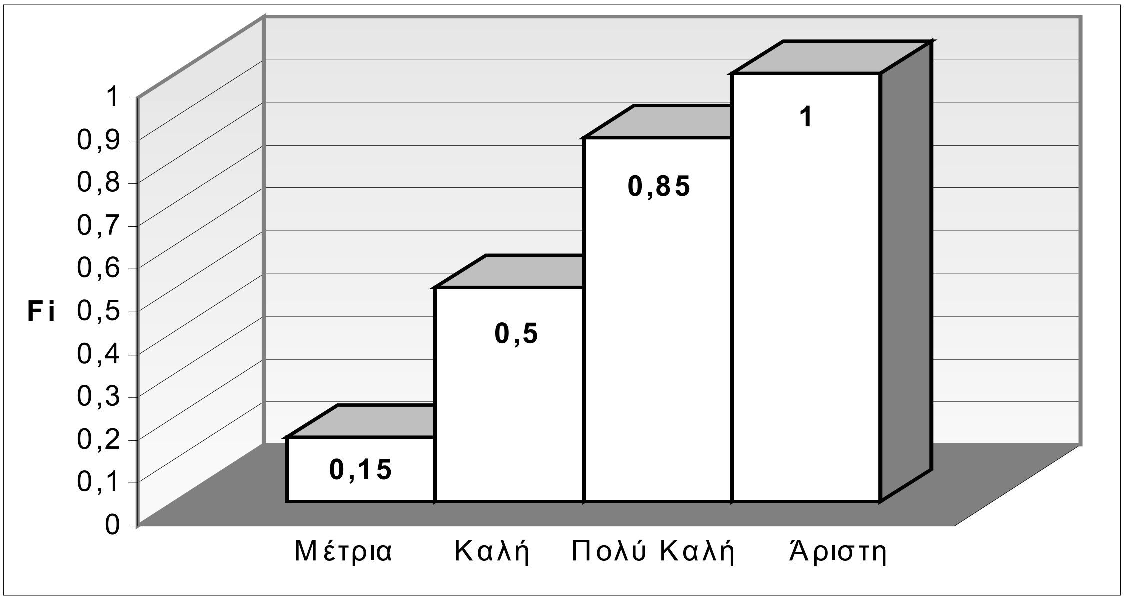
Ραβδογράμμα σχετικών συχνοτήτων



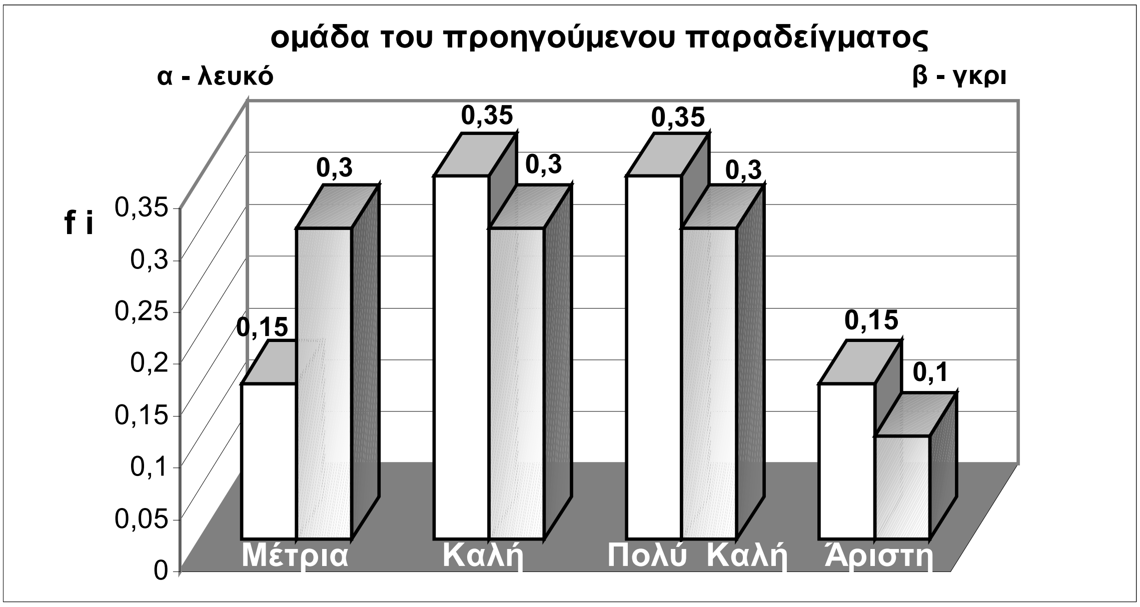
Ραβδόγραμμα αθροιστικών συχνοτήτων



Ραβδόγραμμα αθροιστικών σχετικών συχνοτήτων



Παρατήρηση: Το ραβδόγραμμα προσφέρεται για να συγκρίνουμε δύο ή και περισσότερες ομάδες παρατηρήσεων ως προς την ίδια μεταβλητή (ποιοτική).
Π.χ. Το ραβδόγραμμα που ακολουθεί προέκυψε από δύο ομάδες καλλιεργούμενων εκτάσεων που ελέγχθηκαν ως προς την ποιότητα του βάμβακος που παράγουν (η μια εξ΄ αυτών είναι η εξεταζόμενη στο προαναφερόμενο παράδειγμα).



Ραβδόγραμμα σχετικών συχνοτήτων των δύο ομάδων

Το ραβδόγραμμα φανερώνει ότι η ομάδα β υστερεί της ομάδας α ως προς την ποιότητα του παραγόμενου βάμβακος.

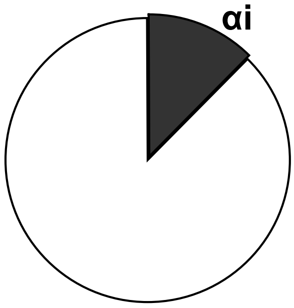
(γ) Κυκλικά διαγράμματα: Αναφέρεται και σε ποιοτικές και σε ποσοτικές μεταβλητές.

Χρησιμοποιείται όταν οι παρατηρήσεις είναι σχετικά λίγες. Το κυκλικό διάγραμμα είναι ένας κυκλικός δίσκος που χωρίζεται σε κυκλικούς τομείς με εμβαδά ή (τόξα) που είναι ανάλογα των συχνοτήτων v_i ή των σχετικών συχνοτήτων f_i των τιμών x_i της μεταβλητής Χ.

Αν θέσουμε x_i το τόξο ενός κυκλικού τομέως τότε έχουμε:

$$\alpha i = v_i \frac{360}{v} = 360 \cdot \frac{v_i}{v} = 360 \cdot fi,$$

όπου v_i, f_i, v η συχνότητα, σχετική συχνότητα και το μέγεθος του δείγματος.



$$\boxed{\alpha i = 360 \frac{v_i}{v} = 360 \quad f_i}$$

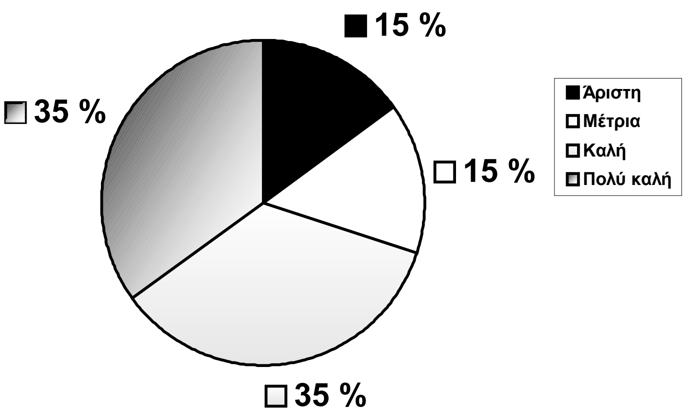
Στο παράδειγμα με την ποιοτική μεταβλητή Χ «**ποιότητα βάμβακος**» έχουμε:

Μέτρια:	$v_1 = 3$,	$f_1 = 0{,}15$
Καλή:	$v_2 = 7$,	$f_2 = 0{,}35$
Πολύ καλή:	$v_3 = 7$,	$f_3 = 0{,}35$
Άριστη:	$v_4 = 3$,	$f_4 = 0{,}15$

Άρα θα έχουμε τα τόξα:

$$\alpha_1 = 360° \cdot f_1 = 360 \cdot 0{,}15 = 54°$$
$$\alpha_2 = 360° \cdot f_2 = 360 \cdot 0{,}35 = 126°$$
$$\alpha_3 = 360° \cdot f_3 = 360 \cdot 0{,}35 = 126°$$
$$\alpha_4 = 360° \cdot f_4 = 360 \cdot 0{,}15 = 54°$$

Κυκλικό διάραμμα της μεταβλητής Χ "Ποιότητα Βάμβακος"



Κεφάλαιο 3ο

Στατιστικά μέτρα θέσης και διασποράς

3.1. Μέτρα θέσης

Τα στατιστικά μέτρα θέσης όπως και τα στατιστικά μέτρα διασποράς μπορούν να αντικαταστήσουν ένα μεγάλο πλήθος δεδομένων και να μας δώσουν το μεγαλύτερο ποσοστό της πληροφορίας που περιέχεται στα δεδομένα.

Τα στατιστικά μέτρα θέσης δίνουν πληροφορία για το μέγεθος των τιμών των δεδομένων και για την κεντρική τάση αυτών. Αυτά χωρίζονται στα εξής μεγέθη:

1. Μέση τιμή ή μέσος αριθμητικός $\bar{x}$

Έστω οι παρατηρήσεις $x_1, x_2, \ldots, x_\nu$ μιας μεταβλητής X.
Τότε ορίζουμε:

$$\bar{\mathbf{x}} = \frac{1}{\nu}\sum_{i=1}^{\nu} x_i$$

Αν θεωρήσουμε τις παρατηρήσεις $x_1, x_2, \ldots, x_\kappa$ ως τιμές της μεταβλητής X με τις αντίστοιχες συχνότητες $\nu_1, \nu_2, \ldots, \nu_\kappa$ $(\kappa \leq \nu)$

Τότε ορίζουμε:

$$\bar{x} = \frac{1}{\nu}\sum_{i=1}^{\nu} x_i \nu_i = \sum_{i=1}^{\kappa} x_i \frac{\nu_i}{\nu} = \sum_{i=1}^{\kappa} x_i\, f_i$$

Τον τύπο αυτό παίρνουμε και όταν έχουμε ομαδοποιημένες παρατηρήσεις με x_i, το κέντρο της κλάσης i και με συχνότητα αυτής ν_i.

Παρατήρηση: Ο Μέσος όρος μειονεκτεί ως στατιστικό μέτρο κεντρικής θέσης διότι επηρεάζεται πολύ από τις ακραίες τιμές της μεταβλητής X.

Παραδείγματα

1) Δέκα γεωργοί της περιοχής Ροδόπης που καλλιέργησαν καλαμπόκι το 1999 είχαν τις επιδόσεις ανά στρέμμα που δίνονται στο παρακάτω πίνακα.

Πίνακας στρεμματικής απόδοσης	
Αριθμός γεωργών	**Στρεμ.Απόδ.(Kgr/στρεμ.)**
1	930
2	1235
3	1100
4	950
5	1230
6	1120
7	1210
8	1150
9	1180
10	1030

Να βρεθεί η μέση απόδοση κατά στρέμμα για τους δέκα εν λόγω γεωργούς.

$$\bar{x} = \frac{930 + 1235 + 1100 + 950 + \dots\dots + 1030}{10} = 1113{,}5 \text{ Kgr / στρέμ.}$$

2) Σε μια εταιρεία οι πωλήσεις που έγιναν από τους πωλητές της εταιρείας είναι οι εξής: (σε εκατομ. δραχμές)

1	2	1	3	5	7	8	9	11	12
5	3	2	10	12	1	4	3	2	1
6	7	6	8	5	8	1	2	5	6

Να βρεθεί ο μέσος όρος πωλήσεων ανά πωλητή.

Ομαδοποιούμε τα δεδομένα σε κλάσεις ίσου πλάτους: και δημιουργούμε τον πίνακα.

Κλάσεις [-)	x_i	v_i	$X_i . v_i$	f_i	$f_I . x_i$
1,3	2	9	18	0,30	0,60
3,5	4	4	16	0,13	0,52
5,7	6	7	42	0,23	1,38
7,9	8	5	40	0,17	1,36
9,11	10	2	20	0,07	0,70
11,13	12	3	36	0,10	1,20
		30	172	1,00	5,76

$v = 30$

$R = 12 - 1 = 11$

$(\kappa = 6)$ (από τον πίνακα)

$$\text{άρα } c = \frac{R}{\kappa} = \frac{11}{6} \simeq 2$$

Ο Μέσος όρος των πωλήσεων είναι:

$$\bar{x} = \frac{1}{v}\sum_{i=1}^{6} x_i \, v_i = \frac{1}{30}\, 172 = 5{,}73$$

$$\text{ή} \quad \bar{x} = \sum_{i=1}^{6} x_i \, f_i = 5{,}76 \cong 5{,}73$$

$$\sum_{i=1}^{6} x_i \, v_i = 172, \qquad \sum_{i=1}^{6} x_i \, f_i = 5{,}76$$

παρατηρούμε απόκλιση λόγω προσεγγίσεων στις τιμές των fi.

2. Σταθμικός μέσος όρος $\bar{x}$

Μερικές φορές οι παρατηρήσεις x_1, x_2,…,x_v μιας μεταβλητής X δεν θεωρούνται ίσως σημασίας (βαρύτητας) για την μεταβλητή X που εξετάζουμε, τότε προσδίδουμε σε κάθε μια παρατήρηση - τιμή x_i και ένα **συντελεστή βάρους ή στάθμισης w_i**.

Αν δηλαδή θεωρήσουμε στις προαναφερόμενες παρατηρήσεις τους συντελεστές βάρους w_1, w_2, ..., w_v, τότε μπορούμε να ορίσουμε τον **σταθμικό μέσο όρο $\bar{x}$** των παρατηρήσεων ως εξής:

$$\bar{x} = \frac{w_1 \, x_1 + w_2 x_2 + \ldots + w_v x_v}{w_1 + w_2 + \ldots + w_v} = \frac{\sum w_i x_i}{\sum w_i}$$

Παράδειγμα:

Υποθέτουμε ότι σε μια αγροτική περιοχή κατ' αρχήν οι γεωργοί ασχολούνται ***όλοι*** με τις διαφορετικές καλλιέργειες που γίνονται στην περιοχή κατ' έτος. Ο πίνακας που ακολουθεί δίνει τον αριθμό των γεωργών και τις διαφορετικές καλλιέργειες κατ' έτος της περιοχής:

Πίνακας

Έτος	**Αριθμ. Γεωργών**	**Αριθ ιαφορ ν κ ιεργε**	**x_i w_i**
	x_i	w_i	
1991	900	3	2700
1992	1000	2	2000
1993	1100	3	3300
1994	1200	2	2400
Σύνολο	4200	10	10400

Ο Σταθμισμένος μέσος αριθμός των γεωργών ανά καλλιέργεια κατ' έτος, χρησιμοποιείται για να πάρουμε τον αριθμητικό μέσο των γεωργών ανά καλλιέργεια.

$\bar{x} = \dfrac{\Sigma x_i w_i}{\Sigma w_i} = \dfrac{10400}{10} = 1040$ γεωργοί σε κάθε καλλιέργεια κατ' έτος.

Αν πάρουμε τώρα ότι η περιοχή έχει 30 Γεωπόνους για να παρακολουθούν τις καλλιέργειες τότε η αναλογία γεωπόνων προς γεωργούς είναι η εξής:
Σε κάθε γεωπόνο αντιστοιχούν ετησίως 10/30 = 0,3 καλλιέργειες κατά μέσο όρο άρα <u>και</u> 0,30 ·1040 = 312 γεωργοί.

3. Γεωμετρικός μέσος όρος $\bar{x}_Γ$

i. Έστω $x_1, x_2, \dots, x_ν$ παρατηρήσεις της μεταβλητής X.
Ορίζουμε ως γεωμετρικό μέσο όρο $\bar{x}_Γ = \sqrt[ν]{x_1 \cdot x_2 \cdot \ldots \cdot x_ν}$

Για τον υπολογισμό του $\bar{x}_Γ$ είναι προτιμότερο να λογαριθμίζουμε και τα δύο μέλη (με δεκαδικό ή νεπέρειο λογάριθμο).

$$\log\bar{x}_Γ = \log\sqrt[ν]{x_1 \cdot x_2 \cdot \ldots \cdot x_ν} = \frac{1}{ν}\log(x_1 \cdot x_2 \cdot \ldots \cdot x_ν) = \frac{\log x_1 + \log x_2 + \ldots + \log x_ν}{ν} \Rightarrow \log\bar{x}_Γ = \frac{1}{ν}\sum_{i=1}^{ν}\log x_1$$

ii. Αν έχουμε παρατηρήσεις με συχνότητα ή ομαδοποιημένες παρατηρήσεις $x_1, x_2, \ldots, x_κ$ της μεταβλητής X, όπου $κ \le ν$ με συχνότητες $ν_1, ν_2, \ldots, ν_κ$ αντίστοιχα τότε

$\bar{x}_Γ = \sqrt[ν]{x_1^{ν_1} \cdot x_2^{ν_2} \cdot \ldots \cdot x_κ^{ν_κ}}$ όπου $ν = ν_1 + ν_2 + \ldots + ν_κ$.

$$\Rightarrow \log\bar{x}_Γ = \frac{1}{ν}\log\left(x_1^{ν_1} x_2^{ν_2} \ldots x_κ^{ν_κ}\right) \Rightarrow$$

$$\log\bar{x}_Γ = \frac{1}{ν}\left(ν_1 \log x_1 + ν_2 \log x_2 + \ldots + ν_κ \log x_κ\right)$$

Παραδείγματα

1. Έστω 10 εργάτες με ημερομίσθια (σε χιλ. δρχ.).

5 4 4 5 6 5 4 5 4 5

$\bar{x}_Γ = \sqrt[10]{5 \cdot 4 \cdot 4 \cdot 5 \cdot 6 \cdot 5 \cdot 4 \cdot 5 \cdot 4 \cdot 5}$

$$\log \overline{x}_\Gamma = \frac{1}{10}\sum_{i=1}^{10}\log x_1 = \frac{1}{10}(\log 5 + \log 4 + ... + \log 5) = \frac{1}{10}\cdot 6{,}63 = 0{,}663 \rightarrow \overline{x}_\Gamma \cong 4{,}6$$

2. Ας θεωρήσουμε 100 εργάτες με τα δεδομένα του πίνακα που ακολουθεί:

Ημερομ. σε χιλ.	x_i	v_i	log x_i	v_i log x_i
4-5	4,5	58	0,65	37,7
5-6	5,5	37	0,74	27,38
6-7	6,5	3	0,81	2,43
7-8	7,5	2	0,87	1,74
		100		

$$\sum v_i \cdot \log x_i = 69{,}25$$

Τότε έχουμε: $\log \overline{x}_\Gamma = \frac{1}{100}\cdot 69{,}25 = 0{,}69 \rightarrow \overline{x}_\Gamma \cong 4{,}9 \rightarrow \overline{x}_\Gamma \cong 4.900$ δρχ.

4. Αρμονικός μέσος όρος h

i. Έστω ότι η τιμή ενός προϊόντος για μια σειρά χρονικών περιόδων είναι x_1, x_2,….,x_v I=1, 2,…., v χρονικές περίοδοι.

Έστω επίσης ότι δαπανούμε ένα ποσό κ χρημάτων για να αγοράσουμε το προϊόν κάθε χρονική περίοδο.

Τότε θα παίρνουμε: $\frac{\kappa}{x_1}$ μονάδες του προϊόντος την 1 περίοδο

$\frac{\kappa}{x_2}$ μονάδες του προϊόντος την 2 περίοδο

…………………………………………...

$\frac{\kappa}{x_v}$ μονάδες του προϊόντος την v περίοδο

Άρα τελικά θα πάρουμε στις v – περιόδους:

$$\frac{\kappa}{x_1} + \frac{\kappa}{x_2} + ... + \frac{\kappa}{x_v} = \kappa\sum_{i=1}^{v}\frac{1}{x_i}$$ μονάδες του προϊόντος.

Θα έχουμε ωστόσο δαπανήσει v·κ δραχμές:

Άρα η μέση δαπάνη h ανά μονάδα του προϊόντος είναι:

$$h = \frac{v\cdot\kappa}{\kappa\cdot\sum_{i=1}^{v}\frac{1}{x_i}} = \frac{v}{\sum_{i=1}^{v}\frac{1}{x_i}}$$ που καλείται **αρμονικός μέσος των τιμών x_1, x_2,….,x_v.**

ii. Αν έχουμε τις τιμές x_1, x_2,…., $x_κ$ με τις αντίστοιχες συχνότητες v_1, v_2,…., $v_κ$ για την τιμή ενός προϊόντος στις ν-περιόδους $v = v_1 + v_2 + ... + v_κ$.

Τότε: $h = \dfrac{v_1 + v_2 + ... + v_κ}{\sum_{i=1}^{κ} \frac{1}{x_i} v_i}$ ο **αρμονικός μέσος** των x_1, x_2,….., $x_κ$.

Παραδείγματα

1. Έστω οι τιμές ενός αγροτικού προϊόντος στα χρόνια 1991-1995 είναι:

Πίνακας

Έτη	**x_i**	**$\frac{1}{x_1}$**
1991	320	0,0031
1992	330	0,0030
1993	350	0,0028
1994	400	0,0025
1995	450	0,0114

Και διαθέτουμε κ= 10.000 δρχ. κατ' έτος για να αγοράζουμε το προϊόν.
Θα 'χουμε ως μέση δαπάνη h ανά μονάδα του προϊόντος τα έτη 1991-1995:

$h = \dfrac{5}{\sum_{i=1}^{5} \frac{1}{x_i}} = \dfrac{5}{0,0114} = 438,59$ δρχ./ ανά μονάδα του προϊόντος.

2. Αν στο προηγούμενο παράδειγμα είχαμε τα δεδομένα στα έτη 1980-1999

x_i	v_i	$\frac{1}{x_1}$	$v_i \frac{1}{x_1}$
320	5	0,0031	0,0155
330	7	0,0030	0,021
350	6	0,0028	0,0168
400	2	0,0025	0,005
	20		

$\sum v_i \frac{1}{x_i} = 0,058$

Τότε αν διαθέταμε επίσης 10.000 δρχ. κατ' έτος για την αγορά του προϊόντος, θα είχαμε μέση δαπάνη ανά μονάδα προϊόντος:

$$h = \frac{20}{\sum_{i=1}^{\kappa} \frac{1}{x_i} \cdot v_i} = \frac{20}{0{,}058} = 344{,}83 \text{ δρχ. /ανά μονάδα προϊόντος}$$

5. Διάμεσος Μ

Θεωρείται το σπουδαιότερο στατιστικό μέτρο θέσης.

Είναι η τιμή που βρίσκεται στο μέσο των παρατηρήσεων όταν αυτές τοποθετηθούν σε αύξουσα ή φθίνουσα σειρά.

i. Η θέση της διαμέσου καθορίζεται από το πλήθος των παρατηρήσεων.
Αν το πλήθος είναι περιττό, δηλαδή αν ν = περιττός, τότε ως διάμεσο λαμβάνουμε τον **μεσαίο όρο** π.χ.

2 3 5 7 8 10 11 13 15

Μ = 8 εφόσον ν = 9.
Αν το πλήθος είναι άρτιο, δηλαδή αν είναι ν = άρτιος, τότε ως διάμεσο λαμβάνουμε το μέσο όρο των δύο μεσαίων όρων π.χ.

2 3 5 8 9 10 11 13

$$M = \frac{8+9}{2} = 8{,}5 \text{ εφόσον } \nu = 8.$$

Παρατήρηση: Γενικά η θέση της διαμέσου δίνεται από τον τύπο:
$\frac{\nu+1}{2}$, όπου ν το πλήθος των παρατηρήσεων.

Παρατήρηση: Η θέση της διαμέσου στο σύνολο των παρατηρήσεων είναι τέτοια ώστε να χωρίζει το πλήθος των παρατηρήσεων σε δύο ισοπληθείς ομάδες.
Έτσι αφήνει το πολύ το 50% των παρατηρήσεων κάτω από αυτήν και το πολύ το 50% των παρατηρήσεων πάνω από αυτήν.

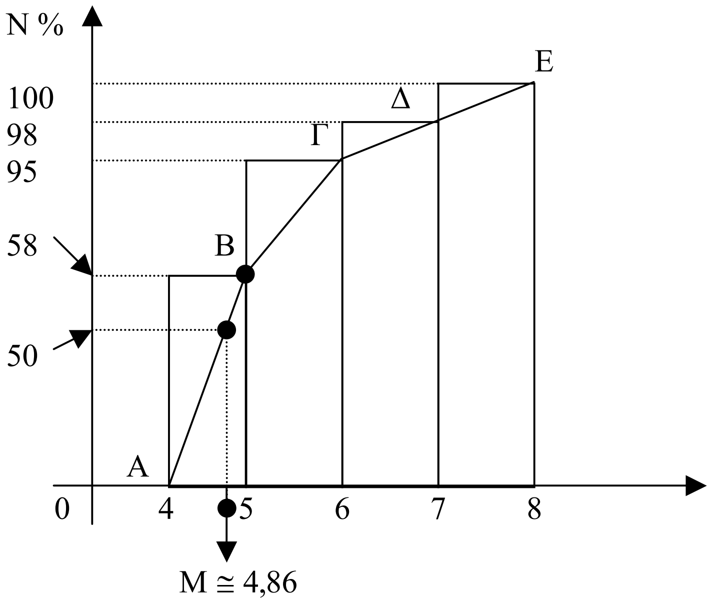
ii. Αν οι παρατηρήσεις συνοδεύονται από συχνότητες ή αν έχουμε ομαδοποιημένες παρατηρήσεις η διάμεσος μπορεί να προσδιορισθεί γραφικά ως εξής:

Παράδειγμα

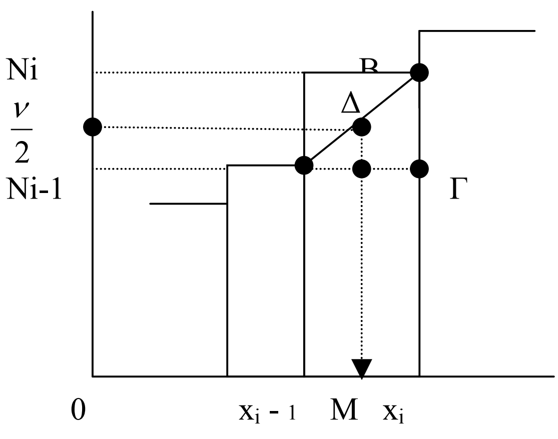
1. Σύμφωνα με το παράδειγμα των ημερομισθίων των 100 εργατών βρίσκουμε την αθροιστική συχνότητα, σχηματίζουμε το πολύγωνο των αθροιστικών σχετικών συχνοτήτων και αντιστοιχούμε στο 50% των παρατηρήσεων, δηλαδή Ν=50, μέσω της πολυγωνικής γραμμής το Μ ως εξής: (Γραφικός εντοπισμός της διαμέσου Μ).

Ημερομ. Κλάσεις	x_i	v_i	f_i	f_I %	F_i	Fi %	N_i
4-5	4,5	58	0,58	58	0,58	58	58
5-6	5,5	37	0,37	37	0,95	95	95

6-7 7-8	6,5 7,5	3 2	0,03 0,02	3 2	0,98 1,00	98 100	98 100
		100					



Παρατήρηση: Ο προσδιορισμός της διαμέσου Μ γίνεται και ***υπολογιστικά***:



Για παρατηρήσεις με αύξουσα σειρά:

$$\boxed{\frac{\mathrm{AE}}{\mathrm{A\Gamma}} = \frac{\Delta\mathrm{E}}{\mathrm{B\Gamma}}}$$

$$\mathrm{AE} = \mathrm{M} - x_{i-1}$$
$$\mathrm{A\Gamma} = x_i - x_{i-1}$$
$$\Delta\mathrm{E} = \frac{\nu}{2} - N_{i-1}$$
$$\mathrm{B\Gamma} = N_i - N_{i-1}$$

Τύπος της διαμέσου Μ

$$\frac{M - x_{i-1}}{x_i - x_{i-1}} = \frac{\frac{\nu}{2} - N_{i-1}}{N_i - N_{i-1}}$$

$$\frac{\mathrm{M}-4}{5-4} = \frac{50-0}{58-0} \to \mathrm{M} - 4 = \frac{50}{58} \to \mathrm{M} = 4 + \frac{50}{58} = 4{,}86$$

$$\begin{matrix} x_i - 1 = 4 \\ x_i = 5 \end{matrix}, \frac{\nu}{2} = 50,\ N_{i-1} = 0,\ N_i = 58$$

6. Επικρατούσα τιμή Τ

Είναι η τιμή που αντιστοιχεί στη μεγαλύτερη συχνότητα.

Ο υπολογισμός είναι εύκολος για παρατηρήσεις που δεν συνοδεύονται από αντίστοιχες συχνότητες.

Π.χ.

2 3 7 5 6 7 2 3 2

Τ=2 εφόσον έχει συχνότητα την μεγαλύτερη 3.

Υπάρχει περίπτωση να μην έχουμε επικρατούσα τιμή

Π.χ.

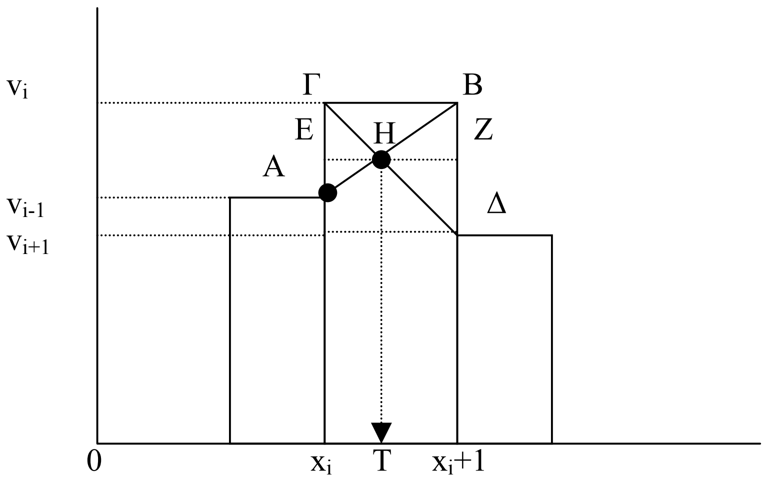
2 3 7 9 11 13

ή να έχουμε περισσότερες από μια επικρατούσες τιμές π.χ.

2 3 5 2 3 4 7

$\begin{matrix} \mathrm{T}_1 = 2 \\ \mathrm{T}_2 = 3 \end{matrix}$ με συχνότητα το 2.

Γραφικά: Ισχύει για δείγμα με μια επικρατούσα τιμή.



Ενώνω τα σημεία Α,Β και Γ,Δ. Το σημείο τομής Η των ΑΒ, ΓΔ δίνει μέσω της τετμημένης του την επικρατούσα τιμή Τ.

Υπολογιστικά: Τα τρίγωνα $\overset{\Delta}{\mathrm{AH\Gamma}}$ *και* $\overset{\Delta}{\mathrm{HB\Delta}}$ είναι όμοια άρα:

$$\frac{\mathrm{A\Gamma}}{\mathrm{B\Delta}} = \frac{\mathrm{EH}}{\mathrm{HZ}}, \quad \acute{o}\pi o\upsilon \quad \begin{matrix} \mathrm{A\Gamma} = v_i - v_{i-1} & \mathrm{EH} = \mathrm{T} - x_i \\ \mathrm{B\Delta} = v_i - v_{i+1} & \mathrm{HZ} = x_{i+1} - T \end{matrix}$$

$$\frac{T - x_i}{x_{i+1} - T} = \frac{v_i - v_{i-1}}{v_i - v_{i+1}}$$

Παράδειγμα:

Στο παράδειγμα των 100 εργατών μέσω του πίνακα μπορούμε να επαληθεύσουμε την επικρατούσα τιμή (Τ= 4,5).

Έχουμε:

$$\begin{matrix} x_i = 4 & v_i = 58 & \\ x_{i+1} = 5 & v_{i-1} = 0 & v_{i+1} = 37 \end{matrix}$$

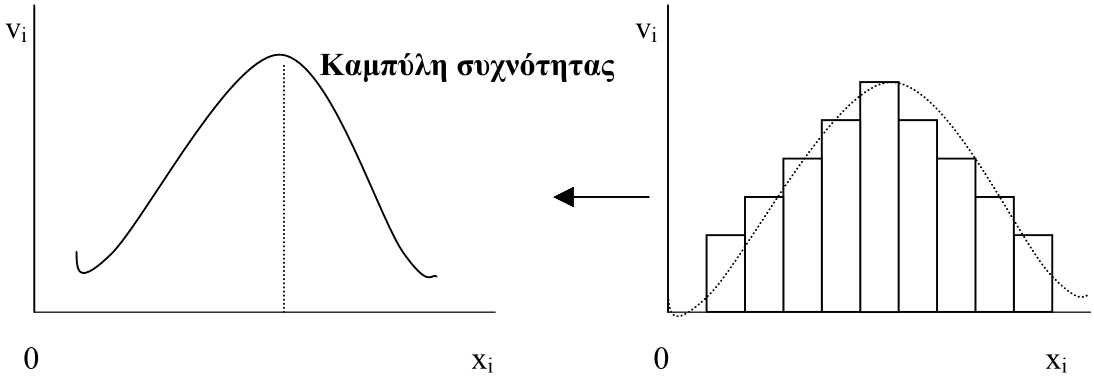
$$\frac{\mathrm{T} - 4}{5 - \mathrm{T}} = \frac{58 - 0}{58 - 37} \rightarrow \frac{\mathrm{T} - 4}{5 - \mathrm{T}} = \frac{58}{21} \simeq 2{,}76$$

$\Rightarrow \mathrm{T} - 4 = 2{,}76(5 - \mathrm{T}) \rightarrow \mathrm{T} = 4{,}73$ προσεγγίζει την τιμή 4,50

7. Σχέσεις Αριθμητικού μέσου, διαμέσου, επικρατούσας τιμής.

Τα τρία αυτά μέτρα θέσης όταν η κατανομή της συχνότητας είναι συμμετρική τότε συμπίπτουν.

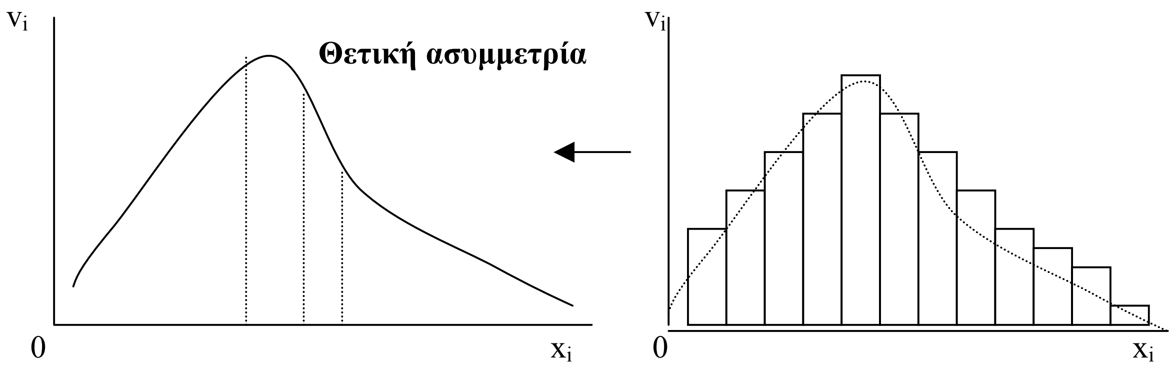
Π.χ.



$\bar{x} = M = T$

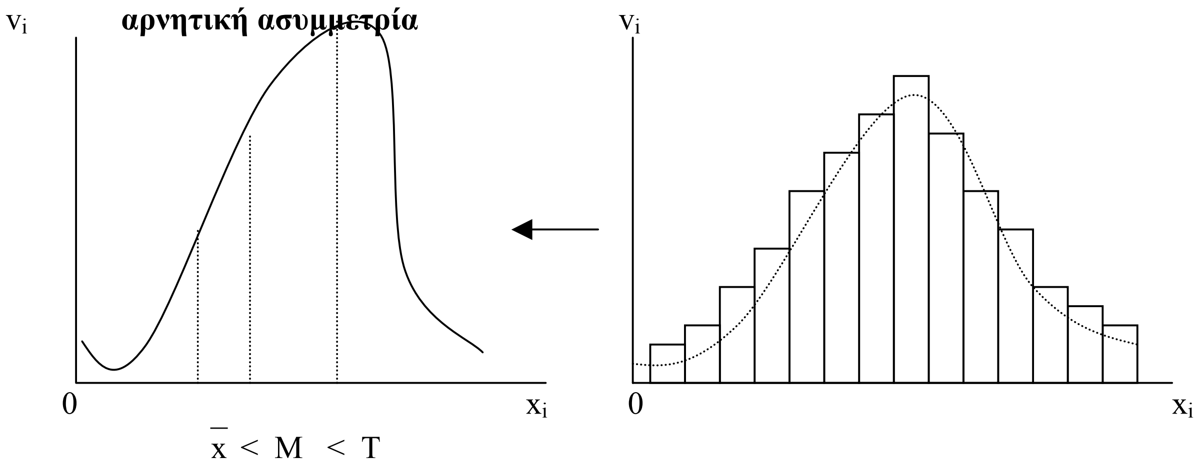
Η καμπύλη συχνότητας είναι οριακή κατάληξη του πολυγώνου συχνότητας όταν ο αριθμός των κλάσεων τείνει απεριόριστα να αυξηθεί και το πλάτος των κλάσεων τείνει στο μηδέν.

Όταν έχουμε ασυμμετρία στην κατανομή της συχνότητας ήτοι θετική ασυμμετρία (κοίλα προς τα δεξιά) έχουμε μακριά ουρά στην περιοχή των μεγάλων τιμών.



$T < M < \bar{x}$

Όταν έχουμε ασυμμετρία στην κατανομή της συχνότητας, αρνητική ασυμμετρία (κοίλα προς τα αριστερά) έχουμε μακριά ουρά στην περιοχή των μικρών τιμών.



3.2. Μέτρα διασποράς.

Πολλές φορές τα μέτρα θέσης δεν μας δίνουν την πραγματική εικόνα των παρατηρήσεων και κυρίως όταν θέλουμε να συγκρίνουμε δείγματα μεταξύ τους.

Συμβαίνει σε δύο δείγματα να έχουμε την ίδια μέση τιμή αλλά να μην έχουμε την ίδια κατανομή συχνότητας.

Όταν υπάρχει μια διασπορά των παρατηρήσεων από την κεντρική θέση (τάση) τότε τα μέτρα θέσης αδυνατούν να μας δώσουν την πραγματική κατάσταση του δείγματος και χρειαζόμαστε άλλα μέτρα τα μέτρα διασποράς που προσδιορίζουν την διασπορά των παρατηρήσεων σε σχέση με τον μέσο όρο του δείγματος.
Τα **μέτρα διασποράς** είναι το εύρος, το ενδοτεταρτημοριακό εύρος, η μέση απόκλιση, η διακύμανση ή διασπορά, η τυπική απόκλιση και ο συντελεστής μεταβολής ή μεταβλητότητας που είναι μέτρο **σχετικής διασποράς.**

1. **Εύρος R**.
Προσδιορίζεται ως η διαφορά της μικρότερης τιμής από την μεγαλύτερη τιμή των παρατηρήσεων.

$\mathbf{R= x_{\text{μεγαλύτερη τιμή}} - x_{\text{μικρότερη τιμή}}}$

Μειονεκτεί ως μέτρο διασποράς γιατί επηρεάζεται απόλυτα από τις ακραίες τιμές. Χρησιμοποιείται κυρίως σε ορισμένα μεγέθη όπου χρειάζεται να βρούμε ανώτερες και κατώτερες τιμές maximum-minimum.

Μετεωρολογία (θερμοκρασίες)- Χρηματιστήριο (μετοχές),

Έλεγχο ποιότητας της μαζικής παραγωγής προϊόντων κ.α.

Χρησιμοποιείται κυρίως σε ολιγομελή δείγματα.

Παράδειγμα:

Έχουμε 10 υπαλλήλους εταιρείας με αμοιβές σε χιλ. δραχμές.

210	280	310	250	400
450	320	280	270	320

Το εύρος R=450-210=240 χιλ. δραχμές.

2. Ενδοτεταρτημοριακό Εύρος Q.

Σε ένα δείγμα παρατηρήσεων x_1, x_2,….,x_ν ορίσαμε την διάμεσο M. Μπορούμε επίσης να ορίσουμε την διάμεσο Q_1 την αριστερά της διαμέσου M παρατηρήσεων όπως και την διάμεσο Q_3 των δεξιά της διαμέσου M παρατηρήσεων. Π.χ.

3 4 7 10 13 15

$Q_1=4$ $Q_3=13$

M=8,5
M= Q_2

Q_1, Q_2, Q_3: Τεταρτημόρια.
Το Q_1 αφήνει κάτω από την τιμή του το πολύ το 25% των παρατηρήσεων, ενώ το Q3 αφήνει πάνω από την τιμή του το πολύ το 25%.
Έχουμε ότι το $Q = Q_3 - Q_1$

Παράδειγμα

Σύμφωνα με το προηγούμενο παράδειγμα θα ‘χουμε:

210 250 270 280 280 310 320 320 400 450

Q_1=270 Q_2=M=295 Q_3=320

$Q = 320 - 270 = 50$ χιλ. δραχμές.

3. Μέση απόκλιση Μ.Α.

Ορίζεται για το δείγμα των παρατηρήσεων x_1, x_2,…, x_ν με μέσο αριθμητικό $\bar{x}$ ως εξής:

$$MA = \frac{\sum_{i=1}^{v} \left| x_i - \overline{x} \right|}{v}$$

Παράδειγμα:

Σε μια περιοχή η στρεμματική απόδοση μιας καλλιέργειας για 10 αγρότες της περιοχής είναι (Kgr/στρεμ.).

1280	1100	950	1000	960
1010	1500	1900	1200	1530

$\overline{x} = 1243 \text{ Kgr/στρ.}$

$$MA = \frac{1}{10}\left(\left|1280 - 1243\right| + \left|1100 - 1243\right| + ... + \left|1530 - 1243\right|\right) =$$

$$= \frac{1}{10}\left(37 + 143 + 293 + 243 + 283 + 233 + 257 + 657 + 43 + 287\right)$$

$\Rightarrow MA = 247{,}6 \text{ Kgr/στρ.}$

4. Διασπορά ή διακύμανση S^2 (ή μέση τετραγωνική απόκλιση) και τυπική απόκλιση S.

Η Διασπορά των τιμών $x_1, x_2, \ldots, x_v$ ενός δείγματος μιας μεταβλητής X με μέσο όρο $\overline{x}$ δίνεται από τον τύπο:

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

Αν ονομάσουμε σ^2 την διασπορά – διακύμανση του πληθυσμού από τον οποίο παίρνουμε το δείγμα, τότε η διακύμανση S^2 αποτελεί τον καλύτερο εκτιμητή της σ^2 του πληθυσμού:

$S^2 = \frac{1}{v-1}\sum_{i=1}^{v}\left(x_i - \overline{x}\right)^2$ (αμερόληπτη εκτίμηση του σ^2). Όταν έχουμε παρατηρήσεις x_1, $x_2, \ldots, x_\kappa$ με συχνότητες $v_1, v_2, \ldots, v_\kappa$ $(\kappa \le v)$ ή ομαδοποιημένες παρατηρήσεις τότε

$$S^2 = \frac{1}{v-1}\sum_{i=1}^{\kappa}\left(x_i - \overline{x}\right)^2 \cdot v_i$$

Συνήθως αποφεύγεται ο τύπος αυτός όπως και ο προηγούμενος και χρησιμοποιούνται τύποι χωρίς την $\overline{x}$.

Αυτό προκύπτει ως εξής:

$$S^2 = \frac{1}{v-1}\sum_{i=1}^{v}\left(x_i^{\,2} - 2x_i\overline{x} + \overline{x}^2\right) = \frac{1}{v-1}\left(\sum_{i=1}^{v} x_i^{\,2} - 2\overline{x}\sum_{i=1}^{v} x_i + v\overline{x}^2\right)$$

$$= \frac{1}{\nu-1}\left(\sum_{i=1}^{\nu} x_i^2 - 2\bar{x}\cdot\left(\nu\cdot\bar{x}\right)^* + \nu\bar{x}^2\right) = \frac{1}{\nu-1}\left(\sum_{i=1}^{\nu} x_i^2 - \nu\bar{x}^2\right) = \frac{1}{\nu-1}\left[\sum_{i=1}^{\nu} x_i^2 - \frac{\left(\sum_{i=1}^{\nu} x_i\right)^2}{\nu}\right]$$

$$\left[* \bar{x} = \frac{1}{\nu}\sum x_i \rightarrow \sum_{i=1}^{\nu} x_i = \nu\bar{x}\cdot\right]$$

Όμοια έχουμε:

$$S^2 = \frac{1}{\nu-1}\left[\sum_{i=1}^{\kappa} \nu_i x_i^2 - \frac{\left(\sum_{i=1}^{\kappa} x_i \nu_i\right)^2}{\nu}\right]$$, για τις ομαδοποιημένες τιμές ή για τις τιμές με συχνότητες

Παραδείγματα

1. Η Διασπορά των 10 τιμών που εκφράζουν την στρεμματική απόδοση σύμφωνα με το προηγούμενο παράδειγμα είναι:

$$S^2 = \frac{1}{10-1}\cdot\left(37^2 + 143^2 + 293^2 + ... + 287^2\right)$$
$$S^2 = \frac{1}{9}\cdot 883\cdot 010 = 98.112\ (\text{Kgr}/\text{στρ.})^2$$

2. Έχουμε 200 επιχειρήσεις που το ύψος των μηνιαίων πωλήσεων τους δίνεται από τον πίνακα.
Σε εκατομμύρια δραχμές.

Αξία πωλήσεων	x_i	ν_i	$x_i\nu_i$	x_i^2	$\nu_i x_i^2$
4 - 6	5	10	50	25	250
6 - 8	7	20	140	49	980
8 -10	9	30	270	81	2430
10-12	11	80	880	121	9680
12-14	13	30	390	169	5070
14-16	15	20	300	225	4500
16-18	17	10	170	289	2890
		200			

$$\sum_{i=1}^{7} x_i v_i = 2200 \qquad \sum_{i=1}^{7} v_i x_i^{\,2} = 25800$$

Η διασπορά των ομαδοποιημένων τιμών είναι:

$$S^2 = \frac{1}{200-1}\left[25800 - \frac{(2200)^2}{200}\right] = \frac{1}{199}(25800 - 24200)$$

$$S^2 = 8{,}04 \cdot (\text{Εκατ})^2$$

Τυπική Απόκλιση S

Η τυπική απόκλιση $S = \sqrt{S^2}$ είναι η θετική τετραγωνική ρίζα της διασποράς. Χρησιμοποιείται πιο πολύ εφόσον παρακάμπτει το μειονέκτημα της διασποράς που μετριέται σε τετραγωνικές μονάδες, σε σχέση με τις απλές μονάδες που εκφράζουν τις παρατηρούμενες τιμές x_i μιας μεταβλητής X.

Έτσι σύμφωνα με τα δύο παραδείγματα που διατυπώθηκαν έχουμε:

$S^2 = 98.112\ (\text{Kgr}/\text{στρ.})^2 \Rightarrow S = \sqrt{98112}\ \text{Kgr}/\text{στρεμ}$

$\Rightarrow S = 313{,}23\ \text{Kgr}/\text{στρεμ}.$

Ενώ από την σχέση $S^2 = 8{,}04(\text{εκατ})^2 \Rightarrow S = \sqrt{8{,}04}\ \text{εκ} \Rightarrow S = 2{,}84\ \text{εκατ}.$

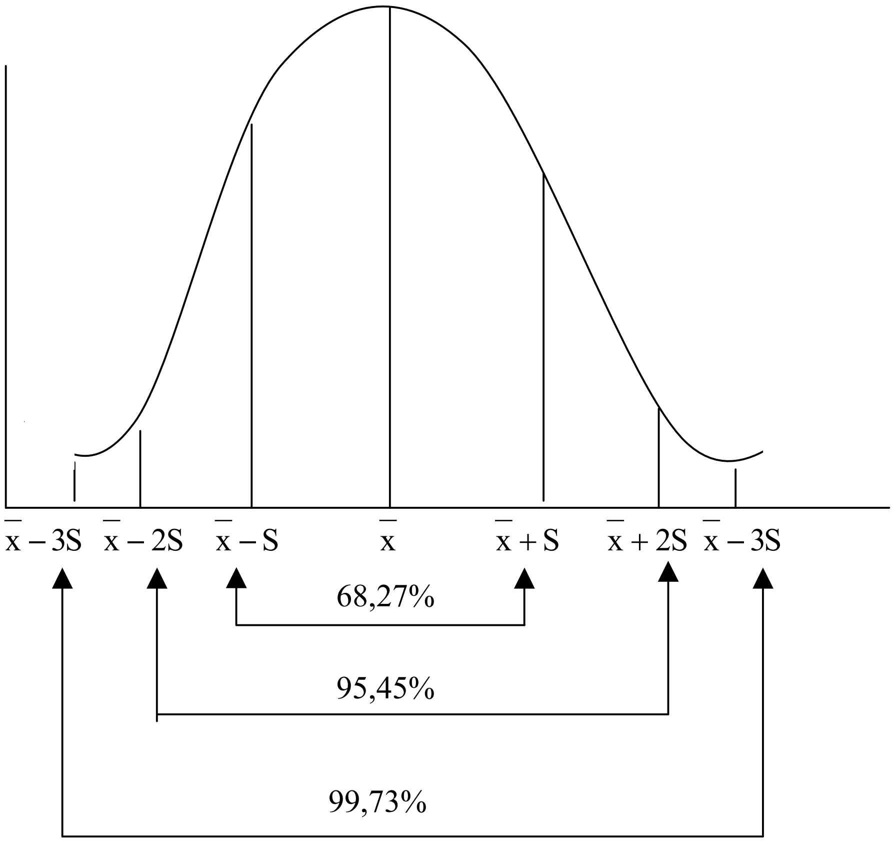
Η τυπική απόκλιση S αποτελεί το κυριότερο μέτρο διασποράς. Όταν αναφέρουμε την τυπική απόκλιση S παράλληλα πρέπει να αναφερθεί και η μέση τιμή των παρατηρήσεων έτσι ώστε να γίνει αντιληπτό ότι οι παρατηρήσεις κυμαίνονται σε ένα διάστημα π.χ. $(\bar{x} - S, \quad \bar{x} + S)$ ή $(\bar{x} - 2S, \quad \bar{x} + 2S)$ κ.ο.κ.

Έτσι στο παράδειγμα των μηνιαίων πωλήσεων των 200 επιχειρήσεων έχουμε S=2,84 εκατ. με μέσο όρο $\bar{x} = \frac{1}{v}\sum_{i=1}^{7} x_i v_i = \frac{2200}{200} = 11$ εκατ.

Αυτό σημαίνει ότι ένα ποσοστό από τις παρατηρήσεις των 200 επιχειρήσεων κυμαίνονται π.χ. στο διάστημα (11-2,84, 11+2,84) ήτοι (8,16, 13,84), ένα μεγαλύτερο ποσοστό από τις παρατηρήσεις βρίσκεται στο διάστημα (11-2·2,84, 11+2·2,84) ήτοι (5,32, 16,68) κ.ο.κ.

Σύμφωνα με το θεώρημα του Chebyshev το ποσοστό των παρατηρήσεων – τιμών ενός δείγματος που βρίσκονται εντός του διαστήματος $\bar{x} \pm \kappa S$ είναι μεγαλύτερο ή ίσο του $1 - \frac{1}{\kappa^2}$ $\mu\varepsilon \quad \kappa > 0$

Όταν η κατανομή της συχνότητας των παρατηρήσεων είναι συμμετρική (έχει κωδωνοειδή μορφή) τότε έχει εκτιμηθεί ότι:



Εντός του διαστήματος $\overline{x}-S, \quad \overline{x}+S$ βρίσκεται το 68,27% των παρατηρήσεων, εντός του διαστήματος $\left(\overline{x}-2S, \quad \overline{x}+2S\right)$.Το 95,45% και εντός του διαστήματος $\left(\overline{x}-3S, \quad \overline{x}+3S\right)$ το 99,73% των παρατηρήσεων.

Έχει εκτιμηθεί επίσης ότι εφόσον στην κωδωνοειδή κατανομή συχνοτήτων το 99,73% περίπου βρίσκεται εντός του διαστήματος $\left(\overline{x}-3S, \quad \overline{x}+3S\right)$που είναι μήκους 6S, μπορούμε να θεωρήσουμε ότι το εύρος των τιμών των παρατηρήσεων $R \cong 6S$ και έτσι η τυπική απόκλιση $S \cong \frac{R}{6}$.

Αυτό μας επιτρέπει να προσδιορίσουμε την τυπική απόκλιση μέσω του εύρους, όταν έχουμε συμμετρική κατανομή της συχνότητας των παρατηρήσεων.

5. Συντελεστή διασποράς (συντελεστής Pearson) v.

Η διασπορά δεν μπορεί να χρησιμοποιηθεί για να συγκριθούν δύο διαφορετικές κατανομές συχνοτήτων όταν έχουν την ίδια τιμή διασποράς αλλά διαφορετική μέση τιμή π.χ. όταν έχουμε δύο κατανομές συχνοτήτων που αναφέρονται σε μηνιαίους μισθούς υπαλλήλων δύο επιχειρήσεων Α και Β με διασπορές

S_A και $S_B = 40.000$ δρχ. αλλά $\overline{x}_A = 300.000$ δρχ. και $\overline{x}_B = 200.000$ τότε δεν μπορούμε να πούμε ότι έχουμε το ίδιο βαθμό διασποράς των παρατηρήσεων στις δύο κατανομές γύρω από τους αντίστοιχους μέσους όρους.

Αυτό διότι διαφορετικό είναι στην επιχείρηση Α να κυμαινόμαστε $\pm$ 40.000 δρχ. γύρω από το $\overline{x}_A = 300.000$ και διαφορετικό είναι να κυμαινόμαστε ± 40.000 δρχ. γύρω από το $\overline{x}_B = 200.000$.

Το ποσοστό του 40.000 δρχ. στις 300.000 δρχ. είναι μικρότερο από το ποσοστό του 40.000 δρχ. στις 200.000 δρχ.

Επίσης η διασπορά δεν μπορεί να χρησιμοποιηθεί και όταν επιχειρούμε να συγκρίνουμε δύο κατανομές συχνότητας με διαφορετικές μονάδες μέτρησης των επιμέρους παρατηρήσεων.

Π.χ. Δεν εξάγεται άμεσα συμπέρασμα ως προς τον βαθμό διασποράς σε δύο κατανομές συχνοτήτων που αναφέρονται σε εισοδήματα αγροτικών νοικοκυριών, όταν η μια ομάδα αναφέρεται σε ελληνικά νοικοκυριά άρα μετριέται σε δραχμές το εισόδημα τους, ενώ η άλλη ομάδα αναφέρεται σε γαλλικά αγροτικά νοικοκυριά και τα εισοδήματά τους μετρώνται σε φράγκα π.χ. ομάδα αγροτικών νοικοκυριών στην Ελλάδα με $\overline{x}_E = 2,5$ και $S_E = 0,5$ (σε εκατ. δραχμές) και ομάδα αγροτικών νοικοκυριών στη Γαλλία με $\overline{x}_\Gamma = 100.000$ φράγκα και $\overline{S}_\Gamma = 14.000$ φράγκα.

Ένα μέτρο που χρησιμοποιείται στις περιπτώσεις που αναφέρθηκαν είναι **ο συντελεστής διασποράς** v που ονομάζεται και **συντελεστής μεταβλητότητας**. Είναι μέτρο σχετικής διασποράς γιατί γίνεται αναφορά της απόκλισης στον μέσο όρο και εκφράζεται επί τοις εκατό: $v = \frac{S}{x}100$. Ένα δείγμα θεωρείται ομοιογενές όταν η τιμή του v είναι το πολύ 10%.

Παραδείγματα

1. Στις προαναφερόμενες περιπτώσεις έχουμε:

i) $$\left.\begin{aligned} v_A &= \frac{40.000}{300.000}\cdot 100 = \frac{40}{3} = 13,3\ \% \\ v_B &= \frac{40.000}{200.000}\cdot 100 = \frac{40}{2} = 20\ \% \end{aligned}\right\} \Rightarrow v_B > v_A$$

Άρα η ομάδα των μισθών των υπαλλήλων της Β επιχείρησης παρουσιάζει μεγαλύτερη διασπορά απ΄ ότι η ομάδα των μισθών των υπαλλήλων της Α επιχείρησης.

ii) $$\left.\begin{aligned} v_E &= \frac{0,5}{2,5}\cdot 100 = 20\ \% \\ v_\Gamma &= \frac{14.000}{100.000}\cdot 100 = 14\ \% \end{aligned}\right\} \Rightarrow v_E > v_\Gamma$$

Τα αγροτικά εισοδήματα των Ελλήνων παρουσιάζουν μεγαλύτερη διασπορά απ΄ ότι τα αγροτικά εισοδήματα των γάλλων.

2. Έχουμε τις στρεμματικές αποδόσεις του βαμβακιού στην Θράκη, Μακεδονία και του καλαμποκιού στη Θεσσαλία που είναι οι ακόλουθες:

$$\bar{x}_{\theta\rho} = 300kgr \quad \kappa\alpha\iota \quad S_{\theta\rho} = 58kgr$$

$$\bar{x}_{M} = 350kgr \quad \kappa\alpha\iota \quad S_{M} = 65kgr$$

$$\bar{x}_{\theta\varepsilon\sigma} = 1150kgr \quad \kappa\alpha\iota \quad S_{\theta\varepsilon\sigma} = 210kgr$$

Οι τυπικές αποκλίσεις που αναφέρονται δεν μας βοηθούν να βρούμε πιο δείγμα παρουσιάζει μεγαλύτερη διασπορά αποδόσεων γύρω από τη μέση απόδοση. Αν πάρουμε όμως τους συντελεστές διασποράς θα έχουμε:

$$v_{\theta\rho} = \frac{58}{300} \cdot 100 = 19{,}33\ \%$$

$$v_{M} = \frac{65}{350} \cdot 100 = 18{,}57\ \%$$

$$v_{\theta\varepsilon\sigma} = \frac{210}{1150} \cdot 100 = 18{,}26\ \%$$

Αυτό μας επιτρέπει να πούμε ότι:

$$v_{\theta\rho} > v_{M} > v_{\theta\varepsilon\sigma}$$

Δηλαδή η στρεμματική απόδοση στη Θράκη έχει μεγαλύτερη διασπορά απ΄ ότι στη Μακεδονία ως προς την καλλιέργεια του βαμβακιού.

Συνεπώς η απόδοση κατά στρέμμα στη Μακεδονία είναι πιο σταθερή.

Επίσης συγκρίνοντας τους συντελεστές διασποράς στη Μακεδονία και στη Θεσσαλία για δύο διαφορετικές καλλιέργειες έχουμε να παρατηρήσουμε ότι:

Η στρεμματική απόδοση του βαμβακιού στη Μακεδονία παρουσιάζει μεγαλύτερες διακυμάνσεις απ΄ ότι η στρεμματική απόδοση του καλαμποκιού στη Θεσσαλία.

6. Σχετική θέση τιμών δύο διαφορετικών δειγμάτων.

Αρκετές φορές γνωρίζουμε το μέσο όρο και την τυπική απόκλιση δύο διαφορετικών δειγμάτων αλλά θέλουμε να προσδιορίσουμε τη σχετική θέση δύο τιμών στα δύο διαφορετικά δείγματα.

Δηλαδή σε ποιο δείγμα η προαναφερόμενη τιμή είναι σε καλύτερη θέση σε σχέση με την θέση της άλλης τιμής στο δεύτερο δείγμα.

Το καλύτερο έχει να κάνει με το ποσό σε μεγέθη τυπικής απόκλισης η τιμή είναι πάνω ή κάτω του αντίστοιχου μέσου όρου του δείγματος στο οποίο ανήκει η τιμή.

Μια τέτοια ποσότητα στη στατιστική είναι το μέγεθος Z:

$$Z = \frac{x - \bar{x}}{S}$$ η οποία καλείται **ανηγμένη μεταβλητή**,

όπου x η τιμή, $\bar{x}$ η μέση τιμή και S η τυπική απόκλιση. (θα γίνει εκτενέστερος λόγος στις κατανομές τυχαίων μεταβλητών παρακάτω για την ανηγμένη μεταβλητή Z).

Παράδειγμα

Έχουμε 6 γεωργούς που καλλιέργησαν σιτάρι και καλαμπόκι ένα έτος σε μια περιοχή της Θράκης και είχαν τις ακόλουθες στρεμματικές αποδόσεις.

Γεωργοί	Αποδ. Σιτάρι	Αποδ. Καλαμπόκι
1	150	880
2	170	960
3	180	1100
4	200	1200
5	210	950
6	190	1110

Να βρεθεί η σχετική θέση του τρίτου γεωργού ως προς τους υπόλοιπους γεωργούς στις δύο αναφερόμενες καλλιέργειες.

Έστω $\overline{x}_6$ η μέση τιμή και S_6 η τυπική απόκλιση στο σιτάρι και

$\overline{x}_\kappa$ η μέση τιμή και S_κ η τυπική απόκλιση στο καλαμπόκι

Έχουμε: $\overline{x}_6 = \dfrac{1100}{6} = 183{,}3 \text{ Kgr/στρ.}$

$$\overline{x}_\kappa = \frac{6200}{6} = 1.033{,}3 \text{ Kgr/στρ.}$$

Επίσης: $S_6^2 = \dfrac{(150-183{,}3)^2 + \ldots\ldots + (190-183{,}3)^2}{6} =$

$$= \frac{1.108{,}89 + 176{,}89 + 10{,}897 + 278{,}89 + 712{,}89 + 44{,}89}{6} = 388{,}89$$

και $S_6 = 19{,}72 \quad \text{Kgr/στρ.}$

$$S_\kappa^2 = \frac{(180-1.033{,}3)^2 + (960-1.033{,}3)^2 + \ldots. + (1.110-1.033{,}3)^2}{6} =$$

$$= \frac{23.500{,}89 + 5.372{,}89 + 4.448{,}89 + 27.788{,}89 + 6.938{,}89 + 5.882{,}89}{6} \cdot S_\kappa^2 = 12.322{,}22 \rightarrow$$

$$\rightarrow S_\kappa = 111{,}00 \text{ Kgr/στρ.}$$

Παίρνω την ανηγμένη μεταβλητή Z:

$$Z_6 = \frac{x_6 - \overline{x}_6}{S_6} = \frac{180-183{,}3}{19{,}72} = -0{,}167$$

$$Z_\kappa = \frac{x_\kappa - \overline{x}_\kappa}{S_\kappa} = \frac{1.100-1.033{,}3}{111} = 0{,}601$$

Παρατηρούμε ότι ο τρίτος γεωργός παρουσίασε καλύτερη απόδοση, σε σχέση με τους υπόλοιπους γεωργούς, στην καλλιέργεια του καλαμποκιού απ΄ ότι στην καλλιέργεια του σιταριού.

Κεφάλαιο 4^{ο}

Βασικές αναφορές στη θεωρία των πιθανοτήτων

4.1. Πείραμα τύχης – ενδεχόμενα – δειγματικός χώρος.

Πείραμα τύχης: Κάθε πείραμα το οποίο μπορεί να επαναληφθεί κάτω από τις ίδιες συνθήκες χωρίς να είναι προβλέψιμο το αποτέλεσμά του.

Ενδεχόμενα ή γεγονότα: Τα δυνατά αποτελέσματα ενός πειράματος τύχης.
Διακρίνονται σε απλά ενδεχόμενα ή σύνθετα ενδεχόμενα ανάλογα αν έχουν ένα ή περισσότερα στοιχεία.

Δειγματικός χώρος: Το σύνολο των απλών ενδεχομένων ενός πειράματος τύχης.

Παραδείγματα

1. Αν σε ένα πείραμα μετράμε τη ζωή ηλεκτρικών αντιστάσεων μιας κατασκευάστριας εταιρείας το αποτέλεσμα θα είναι t σε ώρες $0 \le t \le 5.000$ (δεχόμαστε ότι καμιά ηλεκτρική αντίσταση δεν ξεπερνά σε ζωή τις 5.000 ώρες). Το πείραμα αυτό είναι πείραμα τύχης.

2. Σε ένα πείραμα επιλέγουμε τυχαία ένα μηχάνημα που κατασκευάζεται από μια εταιρεία. Το μηχάνημα μπορεί να είναι καλό ή ελαττωματικό. Στην ερώτηση αν το μηχάνημα είναι καλό ή ελαττωματικό δεν μπορούμε να απαντήσουμε διότι δεν έχουμε πάντα το ίδιο αποτέλεσμα. Αυτό θεωρείται ένα πείραμα τύχης.

3. Αν ρίψουμε ένα νόμισμα δύο φορές τα αποτελέσματα θα είναι ΚΚ, ΚΓ, ΓΚ, ΓΓ (Κ: κορώνα, Γ: γράμμα). Εφόσον δεν υπάρχει το ίδιο πάντα αποτέλεσμα θεωρείται πείραμα τύχης.

4. Ο δειγματικός χώρος του 1^{ου} παραδείγματος είναι: $0 \le \mathrm{t} \le 5.000$. Αυτός ονομάζεται άπειρος, μη αριθμήσιμος, συνεχής.

5. Ο δειγματικός χώρος του 2^{ου} παραδείγματος είναι: {Κ,Ε}: Κ: καλό, Ε: ελαττωματικό. Ο δειγματικός αυτός χώρος είναι πεπερασμένος.

6. Αν ο δειγματικός χώρος ενός πειράματος τύχης είναι το σύνολο IN*={1,2,3,…..} τότε ονομάζεται άπειρος αριθμήσιμος δειγματικός χώρος (δ.χ.).

7. Αν ορισθούν τα δύο γεγονότα-ενδεχόμενα ενός πειράματος τύχης ως σύνολα Α και Β τότε:

- Η πραγματοποίηση του Α ή του Β συμβολίζεται στη γλώσσα των συνόλων ως $\mathrm{A} \cup \mathrm{B}$.

- Η πραγματοποίηση του Α και του Β: $A \cap B$.
- Η μη πραγματοποίηση του Α: $\overline{A}$ (συμπληρωματικό του Α ως προς το σύνολο αναφοράς Ω).
- Η πραγματοποίηση του Α και όχι του Β: Α-Β.

Βέβαιο ενδεχόμενο: Το σύνολο των δυνατών αποτελεσμάτων Ω, εφόσον πάντοτε πραγματοποιείται.

Αδύνατο ενδεχόμενο: Το ενδεχόμενο που δεν μπορεί να πραγματοποιηθεί.

Ασυμβίβαστα ή αμοιβαίως αποκλειόμενα ενδεχόμενα: Δύο ενδεχόμενα Α, Β που $A \cap B = \phi$ (η τομή τους είναι το κενό σύνολο).

4.2 Ορισμός της πιθανότητας

1. Εμπειρικός Ορισμός της πιθανότητας.

Έστω ένα πείραμα τύχης και ένα ενδεχόμενο Ε αυτού.

Σε ν φορές επανάληψη του πειράματος τύχης πραγματοποιείται μ φορές το ενδεχόμενο Ε.

Θεωρούμε την σπάνια περίπτωση που το πείραμα τύχης επαναλαμβάνεται άπειρες φορές.

Τότε έχουμε ως πιθανότητα του ενδεχομένου Ε το :

$$P(E) = \lim_{\nu \to +\omega} \frac{\mu}{\nu}.$$

2. Κλασσικός Ορισμός της Πιθανότητας

Θεωρούμε N(E) το πλήθος των ευνοϊκών περιπτώσεων πραγματοποίησης του ενδεχομένου Ε και N(Ω) το πλήθος των δυνατών περιπτώσεων του ενδεχομένου Ε (το πλήθος των δειγματικών σημείων του δειγματικού χώρου Ω). Ορίζουμε ως πιθανότητα του Ε:

$\mathrm{P(E)} = \dfrac{\mathrm{N(E)}}{\mathrm{N}(\Omega)}$. Θεωρούμε τα στοιχειώδη ενδεχόμενα ισοπίθανα.

4.3. Αξιώματα της πιθανότητας.

Σε κάθε γεγονός – ενδεχόμενο Α ενός πειράματος τύχης με δειγματικό χώρο Ω αντιστοιχεί ένας πραγματικός αριθμός P(A) μέσω της συνάρτησης που ορίσθηκε πιο πάνω.

Αξίωμα 1^{o} : Για κάθε γεγονός Α: $P(A) \geq 0$.

Αξίωμα 2^{o} : Για το Ω ισχύει: $P(\Omega) = 1$.

Αξίωμα 3^{o} : Αν έχουμε τα ενδεχόμενα A_1, A_2,…, του ίδιου πειράματος τύχης που είναι ανά **δύο ασυμβίβαστα,** δηλαδή $A_i \cap A_j = \phi$, $\iota \neq j$ Τότε:

$$P(A_1 \cup A_2 \cup A_3 \cup ....) = P(A_1) + P(A_2) + P(A_3) + ...$$

Ειδικότερα: Αν $A_1 \cap A_2 = \phi$: $P(A_1 \cup A_2 = P(A_1) + P(A_2)$.

Ακόμη ισχύει:

1. $P(\overline{A}) = 1 - P(A)$ εφόσον

$A \cap \overline{A} = \phi και$ $A \cup \overline{A} = \Omega$ $P(A \cup \overline{A}) = P(\Omega) = 1 = P(A) + P(\overline{A}) \Rightarrow$.

2. $P(\phi) = 0$

3. Αν $A = A_1 \cup A_2 \cup .... \cup A_V$ με $A_{i,} i = 1,2,3,...., v$ ανά δύο ξένα μεταξύ τους: $P(A) = P(A_1) + P(A_2) + ... + P(A_V)$.

4. $P(A \cup B) = P(A) + P(B) - P(A \cap B), A, B$ δύο τυχαία ενδεχόμενα.

5. Αν $A \subseteq B \Rightarrow P(A) \leq P(B)$

Παραδείγματα:

1. Ένας γεωργός θέλει να διαλέξει από 4 είδη φυτοφαρμάκων Α,Β,Γ,Δ, δύο είδη για να τα χρησιμοποιήσει στις καλλιέργειές του.

i. Με πόσους τρόπους μπορεί να κάνει αυτή την επιλογή;

ii. Θεωρούμε ότι οι συνδυασμοί που θα κάνει έχουν την ίδια πιθανότητα.

Να βρεθούν οι πιθανότητες των γεγονότων:

1. A_1 = «Η επιλογή να περιέχει το φυτοφάρμακο Α».
2. A_2 = «Η επιλογή να περιέχει το φυτοφάρμακο Β».
3. A_3 = «Η επιλογή να περιέχει το φυτοφάρμακο Α ή Β».
4. A_4 = «Η επιλογή να περιέχει τα φυτοφάρμακα Α και Β».
5. A_5 = «Η επιλογή να μην περιέχει το φυτοφάρμακο Γ ή Δ».
6. A_6 = «Η επιλογή να μην περιέχει το φυτοφάρμακο Α».
7. A_7 = «Η επιλογή να μην περιέχει το φυτοφάρμακο Β».

i. Ζητάμε το δειγματικό χώρο που περιέχει όλους τους συνδυασμούς $S = \{AB, A\Gamma, A\Delta, B\Gamma, B\Delta, \Gamma\Delta\}$.

ii. Η πιθανότητα κάθε συνδυασμού είναι $\frac{1}{6}$.

Οι πιθανότητες των γεγονότων $A_{1,} A_2, ..., A_7$ είναι:

$$P(A_1) = \frac{3}{6} = \frac{1}{2} \qquad P(A_2) = \frac{3}{6} = \frac{1}{2}$$

$$P(A_3) = P(A_1 \cup A_2) = \frac{5}{6} P(A_{4)} = \frac{1}{6} = P(A_1 \cap A_{2)}$$

$$P(A_5) = \frac{1}{6}, \qquad P(A_6) = P(\overline{A}_1) = 1 - P(A_1) = 1 - \frac{1}{2} = \frac{1}{2}$$

$$P(A_7) = P(\overline{A}_2) = 1 - P(A_2) = 1 - \frac{1}{2} = \frac{1}{2}$$

Παρατηρούμε ότι:

$$P(A_3) = P(A_1 \cup A_2) = \frac{5}{6}$$

$$P(A_1) + P(A_2) - P(A_1 \cap A_2) = \frac{1}{2} + \frac{1}{2} - \frac{1}{6} = \frac{5}{6}$$

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

4.4. Δεσμευμένη πιθανότητα.

Ορίζουμε ως δεσμευμένη πιθανότητα του A ως προς B και συμβολίζουμε: P(A/B) την πιθανότητα:

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

Παράδειγμα:

Σε μια αποθήκη υπάρχουν 100 γεωργικά μηχανήματα δύο τύπων I και II. Από αυτά τα 30 είναι του τύπου I και τα 70 του τύπου II. Από τα 30 μηχανήματα του τύπου I τα 5 είναι χαλασμένα, ενώ από τα 70 μηχανήματα του τύπου II τα 8 είναι χαλασμένα.

Ένας γεωργός μπαίνει στην αποθήκη και παίρνει τυχαία ένα μηχάνημα και διαπιστώνει ότι είναι χαλασμένο.

Ποια η πιθανότητα αυτό να είναι τύπου I.

Να υπολογισθούν ακόμη οι πιθανότητες:

1. Να έχει επιλέξει τύπου II, όταν διαπιστώσει ότι είναι χαλασμένο.
2. Να έχει επιλέξει καλό μηχάνημα, όταν διαπιστώσει ότι είναι τύπου I.
3. Να έχει επιλέξει καλό μηχάνημα, όταν διαπιστώσει ότι είναι τύπου II.
4. Να έχει επιλέξει χαλασμένο μηχάνημα, όταν διαπιστώσει ότι είναι τύπου I.
5. Να έχει επιλέξει χαλασμένο μηχάνημα, όταν διαπιστώσει ότι είναι τύπου II.
6. Να έχει επιλέξει μηχάνημα τύπου I, όταν διαπιστώσει ότι είναι καλό.
7. Να έχει επιλέξει μηχάνημα τύπου II, όταν διαπιστώσει ότι είναι καλό.

	Χαλασμένο μηχάνημα	Καλό μηχάνημα	Σύνολο
Τύπος Ι	5	25	30
Τύπος ΙΙ	8	62	70
	13	87	100

X: Χαλασμένο μηχάνημα
K: Καλό μηχάνημα

Θέλουμε να υπολογίσουμε την δεσμευμένη πιθανότητα:

$$P(I/X) = \frac{P(I \cap X)}{P(X)}.$$

Έχουμε : $P(I) = \frac{30}{100} = 0,3 \quad , \quad P(II) = \frac{70}{100}$

$$P(X) = \frac{13}{100} = 0,13 \quad , \quad P(I \cap X) = \frac{5}{100} = 0,05$$

$$P(II \cap X) = \frac{8}{100} = 0,08 \quad , \quad P(I \cap k) = \frac{25}{100} = 0,25$$

$$P(II \cap K) = \frac{62}{100} = 0,62$$

Άρα:

$$P(I/X) = \frac{P(I \cap X)}{P(X)} = \frac{0,05}{0,13} = 0,38$$

$$P(II/X) = \frac{0,08}{0,13} \simeq 0,61 \quad P(K/I) = \frac{0,25}{0,30} = 0,83$$

$$P(K/II) = \frac{0,62}{0,70} = 0,89 \qquad P(X/II) = \frac{0,08}{0,7} = 0,11$$

$$P(X/I) = \frac{0,05}{0,30} = 0,17 \qquad P(I/K) = \frac{0,25}{0,87} = 0,29$$

$$P(II/K) = \frac{0,62}{0,87} = 0,71$$

4.5. Ανεξάρτητα ενδεχόμενα

Δύο ενδεχόμενα A, B ονομάζονται ανεξάρτητα όταν η εμφάνιση του A δεν αποκλείει την εμφάνιση του B και αντίστροφα.
Η πιθανότητα να εμφανισθούν **ταυτόχρονα** και τα δύο συμβάντα A και B ισούται με το γινόμενο των πιθανοτήτων τους.

$P(A \cap B) = P(A) \cdot P(B)$ **«Κανόνας του πολλαπλασιασμού των πιθανοτήτων»**

Παράδειγμα

Εξετάζουμε τους 300 εργαζόμενους σε μια επιχείρηση ως προς το φύλλο και το είδος της εργασίας τους.

Από την εξέταση προέκυψαν τα ακόλουθα στοιχεία.

Φύλο	Εργάτης Εργάτρ	Διοικ Υπά +Στέλεχος	
Άνδρες	160	40	200
Γυναίκες	20	80	100
Σύνολο	180	120	300

Επιλέγουμε στην τύχη ένα εργαζόμενο, να υπολογισθούν οι πιθανότητες:

i) Να είναι εργάτης και άνδρας.

ii) Να είναι Δ. υπάλληλος και άνδρας.

iii) Να είναι εργάτρια και γυναίκα

iv) Να είναι Δ. υπάλληλος και γυναίκα.

V) Να εξετασθούν αν τα χαρακτηριστικά ***«είδος εργασίας : εργάτης»*** και ***«φύλλο: άνδρας»*** είναι ανεξάρτητα μεταξύ τους.

Έχουμε: $\mathrm{P}(\text{εργάτης} - \alpha) = \dfrac{180}{300} = 0{,}6$ $\quad \mathrm{P}(\text{άνδρας}) = \dfrac{200}{300} = 0{,}67$

$\mathrm{P}(\Delta.\text{υπάλ.} + \text{Στέλ.}) = \dfrac{120}{300} = 0{,}4$ $\quad \mathrm{P}(\text{γυναίκα}) = \dfrac{100}{300} = 0{,}33$

i) $P(\varepsilon\rho\gamma\acute{\alpha}\tau\eta\varsigma\,\&\,\acute{\alpha}\nu\delta\rho\alpha\varsigma) = 0{,}53$

ii) $P(\Delta.\upsilon\pi\alpha\lambda.\,\&\,\acute{\alpha}\nu\delta\rho\alpha\varsigma) = 0{,}13$

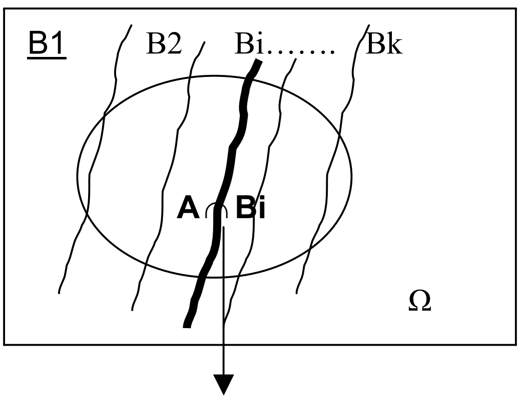
iii) $P(\varepsilon\rho\gamma\acute{\alpha}\tau\rho\iota\alpha\,\&\,\gamma\upsilon\nu\alpha\acute{\iota}\kappa\alpha) = 0{,}07$

iv) $P(\Delta.\upsilon\pi\acute{\alpha}\lambda.\,\&\,\gamma\upsilon\nu\alpha\acute{\iota}\kappa\alpha) = 0{,}27$

Έχουμε : $\mathrm{P}\,(\text{εργάτης}) \cdot \mathrm{P}(\text{άνδρας}) = 0{,}6 \cdot 0{,}67 = 0{,}402 \neq 0{,}53 = \mathrm{P}(\text{εργάτης και άνδρας})$:
Τα χαρακτηριστικά "**φύλλο: άνδρας**" και "**είδος εργασίας: εργάτης**" δεν είναι ανεξάρτητα.

4.6. Νόμος του Bayes

Έστω Β1, Β2,….,Βκ ο διαμελισμός του δειγματικού χώρου S και ένα ενδεχόμενο Α του δειγματικού χώρου S.



Τότε:

$$P(Bi/A) = \frac{P(A \cap Bi}{P(A)} \quad (\textbf{Νόμος Bayes})$$

Έχουμε όμως: $P(A) = P(A \cap B_1) + P(A \cap B_2) + ... + P(A \cap B_k)$

Και τελικά

$$P(Bi/A) = \frac{P(A \cap Bi)}{P(A \cap B_1) + ... + P(A \cap B_K)}$$

$$\Rightarrow P(Bi/A) = \frac{P(A/Bi) \cdot P(Bi)}{P(A/B_1)P(B_1) + P(A/B_2) \cdot P(B_2) + ... + P(A/B_K) \cdot P(B_{K)}}$$

Παράδειγμα

Ένα μεταποιημένο γεωργικό προϊόν μπορεί να παραχθεί σε τρία διαφορετικά εργοστάσια. Το πρώτο εργοστάσιο παράγει τριπλάσια ποσότητα από το δεύτερο την ίδια χρονική περίοδο, ενώ το δεύτερο παράγει διπλάσια ποσότητα από το τρίτο εργοστάσιο την ίδια χρονική περίοδο. Είναι γνωστό ότι το 3% των προϊόντων που παράγονται στα δύο πρώτα εργοστάσια είναι ελαττωματικά, ενώ στο τρίτο εργοστάσιο τα ελαττωματικά είναι το 5% των παραγόμενων προϊόντων στην ίδια πάντα χρονική περίοδο.
Όλα τα προϊόντα συγκεντρώνονται σε ένα αποθηκευτικό χώρο και γίνεται ποιοτικός έλεγχος. Ο ελεγκτής παίρνει τυχαία ένα προϊόν και το βρίσκει καλό. Ποια η πιθανότητα να έχει κατασκευασθεί στο δεύτερο εργοστάσιο;

Έστω Α το ενδεχόμενο **«το προϊόν καλό»** και Bi τα ενδεχόμενα «**Το προϊόν κατασκευάζεται στο i - Εργοστάσιο» με i =1,2,3.**

Ζητάμε την πιθανότητα:

$$P(B_2 / A) = \frac{P(A / B_2) \cdot P(B_2)}{P(A / B_1) P(B_1) + P(A / B_2) \cdot P(B_2) + P(A / B_3) \cdot P(B_3)}$$

Έχουμε:

$$P(B_1) = \frac{6}{9} \qquad P(B_2) = \frac{2}{9} \qquad P(B_3) = \frac{1}{9}$$

Εξήγηση: Αν x η ποσότητα η παραχθείσα στο τρίτο εργοστάσιο τότε θα έχει παραχθεί 2 x ποσότητα στο δεύτερο και 3·2 x = 6x στο πρώτο εργοστάσιο.

Άρα στην ποσότητα 9X:

το X αντιστοιχεί στο 3^o εργοστάσιο,
το 2X αντιστοιχεί στο 2^o εργοστάσιο
και το 6X αντιστοιχεί στο 1^o εργοστάσιο

Συνεπώς : $P(B_1) = \frac{6x}{9x} = \frac{6}{9} \qquad P(B_2) = \frac{2x}{9x} = \frac{2}{9} \qquad P(B_3) = \frac{x}{9x} = \frac{1}{9}$

Έχουμε επίσης:

$\mathrm{P(A / B_1)} = \frac{97}{100} = 0{,}97$ (εφόσον το 3% είναι ελαττωματικά)

$\mathrm{P(A / B_2)} = \frac{97}{100} = 0{,}97$ (όμοια)

$P(A / B_3) = \frac{95}{100} = 0{,}95$ (εφόσον το 5% είναι ελαττωματικά)

Άρα $\mathrm{P(B_2 / A)} = \dfrac{0{,}97 \cdot \frac{2}{9}}{0{,}97 \cdot \frac{6}{9} + 0{,}97 \cdot \frac{2}{9} + 0{,}95 \cdot \frac{1}{9}} = \dfrac{0{,}22}{0{,}65 + 0{,}22 + 0{,}11}$

$\mathrm{P(B_2 / A)} = \frac{0{,}22}{0{,}98} = 0{,}22.$

4.7. Στοιχεία Συνδυαστικής.

4.7.1. Βασική (πολλαπλασιαστική) αρχή της απαρίθμησης - Δενδροδιάγραμμα.

Αν ένα πείραμα τύχης χωρίζεται σε n φάσεις και έχουμε:
n $1^{η}$ φάση να πραγματώνεται με K_1 τρόπους
n $2^{η}$ φάση να πραγματώνεται με K_2 τρόπους
…………
$n^{η}$ φάση να πραγματώνεται με K_n τρόπους
τότε το πείραμα πραγματώνεται με K_1 K_2 ….K_n διαφορετικούς τρόπους.

Δενδροδιάγραμμα

Η αρχή της απαρίθμησης παρουσιάζεται αναλυτικά με ένα διάγραμμα μορφής δένδρου το δενδροδιάγραμμα.

Παράδειγμα

Θέλουμε να σχηματίσουμε μια τριμελή επιτροπή τεχνικών η οποία να έχει ένα αρχιτέκτονα, ένα πολιτικό μηχανικό και ένα μηχανικό περιβάλλοντος. Έχουμε στην διάθεσή μας δύο αρχιτέκτονες, τρεις πολιτικούς μηχανικούς και τρεις μηχανικούς περιβάλλοντος.

Με πόσους τρόπους μπορούμε να σχηματίσουμε την τεχνική αυτή επιτροπή; Να γίνει το αντίστοιχο δενδροδιάγραμμα.

Εφόσον η επιλογή μπορεί να γίνει σε 3 φάσεις και έχουμε
n $1^{η}$ φάση να πραγματώνεται με 2 τρόπους,
n $2^{η}$ φάση να πραγματώνεται με 3 τρόπους και
n $3^{η}$ φάση να πραγματώνεται με 3 τρόπους
τότε υπάρχουν σύμφωνα με την βασική αρχή της απαρίθμησης 2·3·3=18 διαφορετικοί τρόποι για να σχηματίσουμε την τεχνική επιτροπή.

4.7.2. Πιθανότητες στη δειγματοληψία.

Ορισμοί: Συνδυασμοί ν αντικειμένων ανά $\mu : \begin{pmatrix} \nu \\ \mu \end{pmatrix} = \frac{\nu!}{\mu!(\nu-\mu)!} = {}_{\nu}C_{\mu}$

Διατάξεις ν αντικειμένων ανά $\mu : \Delta_{\mu}^{\nu} = \frac{\nu!}{(\nu-\mu)!} = {}_{\nu}P_{\mu}$

Μεταθέσεις ν αντικειμένων: $M_{\nu} = \nu!$

${}_{\nu}P_{\mu} = \Delta_{\mu}^{\nu} = {}_{\nu}C_{\mu} \cdot M_{\mu}$ (Διατάξεις ίσον συνδυασμοί επί μεταθέσεις).

Οι αριθμοί $\begin{pmatrix} \nu \\ \mu \end{pmatrix} = \frac{V!}{\mu!(\nu-\mu)!}$ ονομάζονται διωνυμικοί συντελεστές επειδή υπάρχουν στο ανάπτυγμα του διώνυμου $(a+\beta)^{\nu}$:

$$(a+\beta)^{v}=\binom{v}{o}a^{v}\beta^{o}+\binom{v}{1}a^{v-1}\beta+\binom{v}{2}a^{v-2}\beta^{2}+...+\binom{v}{v-1}a\beta^{v-1}+\binom{v}{v}a^{o}\beta^{v}$$

Παράδειγμα

Υποθέτουμε ότι από τις 20 ημέρες που μετράμε την συγκέντρωση στα νερά ποταμού ενός τοξικού ρυπαντή, ο οποίος χύνεται από παρακείμενη χημική βιομηχανία, στις επτά (7) παρουσιάζεται υπέρβαση του ορίου συγκέντρωσής του. Ποια η πιθανότητα αν πάρουμε τυχαία 3 ημέρες να έχουμε υπέρβαση του ορίου συγκέντρωσης του ρυπαντή.

Δυνατές περιπτώσεις $\binom{20}{3}=N(\Omega)$

Ευνοϊκές περιπτώσεις $\binom{7}{3}=N(E)$

Έστω Ε το ενδεχόμενο να έχουμε σε 3 τυχαίες ημέρες υπέρβαση του ορίου ρύπανσης.

Ζητάμε: $P(E)=\dfrac{N(E)}{N(\Omega)}$

$$P(E)=\frac{\binom{7}{3}}{\binom{20}{3}}=\frac{\frac{7!}{3!4!}}{\frac{20!}{3!17!}}=\frac{7!17!}{4!20!}=\frac{5\cdot 6\cdot 7}{18\cdot 19\cdot 20}=0{,}031\rightarrow 3{,}1\%$$

4.7.3. Παραδείγματα.

1. Ένα δοχείο περιέχει τέσσερις κίτρινους κύβους και τρεις πράσινους. Ένα άλλο δοχείο περιέχει τρεις κίτρινους κύβους και δύο πράσινους. Ρίχνουμε ένα ζάρι και αν έρθει άρτιος αριθμός βγάζουμε ένα κύβο από το πρώτο δοχείο ενώ αν έρθει περιττός αριθμός βγάζουμε ένα κύβο από το δεύτερο δοχείο.

Ποια η πιθανότητα να βγει πράσινος κύβος;

Δοχείο πρώτο Δ_1 Πράσινος κύβος Π.

Δοχείο δεύτερο Δ_2 Κίτρινος κύβος Κ.

Ζητάμε $P(\Pi)=P(\Pi/\Delta_1)\cdot P(\Delta_1)+P(\Pi/\Delta_2)\cdot P(\Delta_2)=P(\Pi\bigcap\Delta_1)+P(\Pi\cap\Delta_2)$

$$P(\Delta_1)=\frac{1}{2},\qquad P(\Delta_2)=\frac{1}{2},\qquad P(\Pi/\Delta_1)=\frac{3}{4+3}=\frac{3}{7},\qquad P(\Pi/\Delta_2)=\frac{2}{3+2}=\frac{2}{5}$$

Άρα $P(\Pi)=\dfrac{3}{7}\cdot\dfrac{1}{2}+\dfrac{2}{5}\cdot\dfrac{1}{2}=\dfrac{3}{14}+\dfrac{1}{5}=\dfrac{29}{70}$

2. Στο προηγούμενο παράδειγμα υποθέτουμε ότι δεν γνωρίζουμε αν ήρθε άρτιος ή περιττός αριθμός από την ρίψη του ζαριού, άρα δεν γνωρίζουμε από ποιο δοχείο βγήκε ο πράσινος κύβος, μόνο ότι βγήκε.
Ποια η πιθανότητα να πήραμε τον πράσινο κύβο από το δεύτερο δοχείο;

$$P(\Delta_2/\Pi)=\frac{P(\Pi/\Delta_2)\cdot P(\Delta_2)}{P(\Pi)}=\frac{P(\Pi/\Delta_2)\cdot P(\Delta_2)}{\sum_{i=1}^{2}P(\Pi/\Delta i)P(\Delta i)}=\frac{\frac{2}{5}\cdot\frac{1}{2}}{\frac{29}{70}}=\frac{70}{529}=\frac{14}{29}$$

(Νόμος του Bayes).

3. Επιλέγουμε στην τύχη τρεις λάμπες από ένα κουτί που περιέχει 5 κόκκινες, 3 άσπρες και 2 πράσινες (i) με επανατοποθέτηση και (ii) χωρίς επανατοποθέτηση. Να βρεθεί η πιθανότητα να είναι οι λάμπες στη σειρά άσπρη, πράσινη, κόκκινη.

Έστω A_1: η άσπρη στην πρώτη επιλογή.
Π_2: η πράσινη στην δεύτερη επιλογή.
K_3: η κόκκινη στην τρίτη επιλογή.

(i) Τα τρία ενδεχόμενα A_1, Π_2, K_3 εφόσον έχουμε επανατοποθέτηση είναι ανεξάρτητα.

$$P(A_1\cap\Pi_2\cap K_3)=P(A_1)\cdot P(\Pi_2/A_1)\cdot P(K_3/A_1\cap\Pi_2)=P(A_1)\cdot P(\Pi_2)\cdot P(K_3)$$

$$=\frac{3}{5+3+2}\cdot\frac{2}{5+3+2}\cdot\frac{5}{5+3+2}=\frac{3\cdot2\cdot5}{10\cdot10\cdot10}=\frac{3}{100}=0{,}03.$$

(ii) Τα ενδεχόμενα A_1, Π_2, K_3 δεν είναι ανεξάρτητα (δεν έχουμε επανατοποθέτηση)

Άρα $P(A_1\cap\Pi_2\cap K_3)=P(A_1)\cdot P(\Pi_2/A_1)\cdot P(K_3/A_1\cap\Pi_2)$

$$=\frac{3}{5+3+2}\cdot\frac{2}{5+2+2}\cdot\frac{5}{5+2+1}=\frac{3\cdot2\cdot5}{10\cdot9\cdot8}=\frac{1}{24}.$$ Αρχή της απαρίθμησης.

4. Ένα εργοστάσιο παράγει 1800 μηχανές την ημέρα από τις οποίες το 4% είναι ελαττωματικές. Να υπολογισθεί η πιθανότητα να είναι ελαττωματικές 20 από ένα δείγμα 800 μηχανών.

Έχουμε $1800\cdot\frac{4}{100}=72$ ελαττωματικές μηχανές άρα 1628 καλές.

Τα δυνατά δείγματα είναι ασφαλώς $\binom{1800}{800}$ σε αριθμό.

Τις ευνοϊκές περιπτώσεις θα τις έχουμε παίρνοντας τις 20 από τις 72 ελαττωματικές και τις υπόλοιπες 780 καλές από τις 1628 καλές του πλήθους.

Άρα $\binom{72}{20}\cdot\binom{1628}{780}$

Επομένως η ζητούμενη πιθανότητα είναι: $\dfrac{\binom{72}{20}\cdot\binom{1628}{780}}{\binom{1800}{800}}$.

5. Οι πιθανότητες να είναι σε λειτουργία μετά από 10 χρόνια δύο μηχανές είναι 0,6 και 0,3 αντίστοιχα. Να βρεθεί η πιθανότητα να λειτουργούν μετά από 10 χρόνια:

(i) και οι δύο
(ii) καμιά και
(iii) τουλάχιστον μια.

Έστω M_1 η πρώτη μηχανή και M_2 η δεύτερη.
Έστω $ΛM_1$ λειτουργεί η M_1 μετά από 10 χρόνια.
Έστω $ΛM_2$ λειτουργεί η M_2 μετά από 10 χρόνια.
Έχουμε: $P(ΛM_1) = 0{,}6$, $P(ΛM_2) = 0{,}3$, $P(\overline{ΛM_1}) = 0{,}4$, $P(\overline{ΛM_2}) = 0{,}7$.
Τότε έχουμε:

(i) $P(ΛM_1 \cap ΛM_2) = P(ΛM_1)\cdot P(ΛM_2) = 0{,}6\cdot 0{,}3 = 0{,}18$

(ii) $P(\overline{ΛM_1} \cap \overline{ΛM_2}) = P(\overline{ΛM_1})\cdot P(\overline{ΛM_2}) = 0{,}4\cdot 0{,}7 = 0{,}28$

(iii) P (τουλάχιστον μια μηχανή να λειτουργεί) = 1-P (να μην λειτουργεί καμία) = 1-0,28=0,72.

Κεφάλαιο 5ο

Τυχαίες Μεταβλητές

5.1. Ορισμός τυχαίας μεταβλητής

Τυχαία μεταβλητή λέγεται η συνάρτηση η οποία μας επιτρέπει να απεικονίσουμε όλα τα δυνατά αποτελέσματα ενός πειράματος τύχης στο σύνολο των πραγματικών αριθμών.

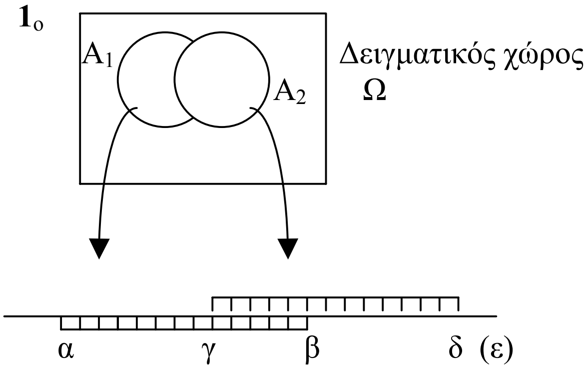
Τις τυχαίες μεταβλητές τις συμβολίζουμε με κεφαλαία γράμματα X, Y, Z και με μικρά γράμματα τις τιμές που παίρνουν αυτές $x_1, x_2, \ldots, y_1, y_2, \ldots z_1, z_2, \ldots$

Η τιμή της $X = x$ εκφράζει ένα συγκεκριμένο ενδεχόμενο.

Αν ο δειγματικός χώρος απεικονίζεται σε συνεχή διάστημα τιμών τότε η τυχαία μεταβλητή (τ.μ.) X καλείται συνεχής τ.μ., ενώ αν απεικονίζεται σε διακριτές τιμές η τ.μ. X καλείται διακριτή τ.μ.

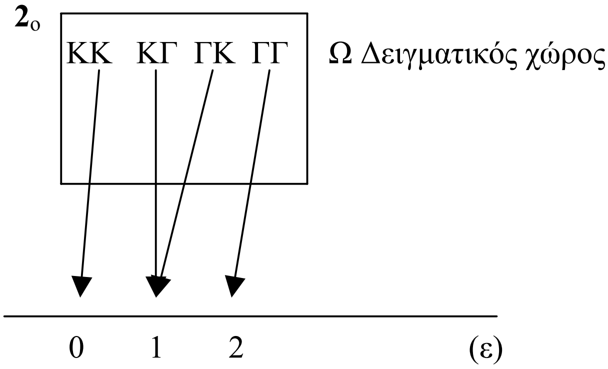
Παραδείγματα:

1ο



X τ.μ. συνεχής

2ο



Πείραμα τύχης: Ρίψη νομίσματος δύο φορές ένδειξη «Γράμμα».
Ω= {ΚΚ, ΚΓ, ΓΚ, ΓΓ}

3^o Παράδειγμα:

Ένας αγρότης προσδοκά στα επόμενα τρία χρόνια να έχει ετήσια δαπάνη σε σταθερές τιμές το πολύ 300.000 δρχ. για την αγορά λιπασμάτων.

Αν συμβολίσουμε με:

Α: το ενδεχόμενο να έχει ετήσια δαπάνη $\leq$ 300.000 δρχ.

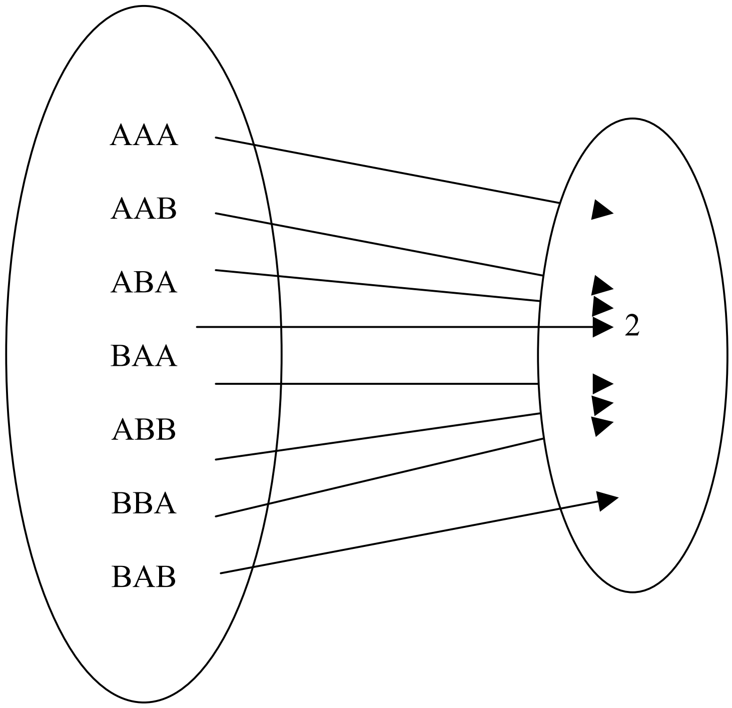
Β: το ενδεχόμενο να έχει ετήσια δαπάνη > 300.000 δρχ.

Τότε ο δειγματικός χώρος S για τις τρεις επόμενες χρονιές θα είναι:
S = {ΑΑΑ, ΑΑΒ, ΑΒΑ, ΒΑΑ, ΑΒΒ, ΒΒΑ, ΒΑΒ, ΒΒΒ).

Μπορούμε να ορίσουμε μια συνάρτηση Χ που να αντιστοιχεί σε κάθε στοιχείο του δειγματικού χώρου S τον αριθμό των ετών που η ετήσια δαπάνη είναι μικρότερη ή ίση από τις 300.000 δρχ.

Έτσι έχουμε ότι η τυχαία μεταβλητή Χ παίρνει τις τιμές

$x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = 3$.



5.2. Κατανομή πιθανότητας μιας τυχαίας Μεταβλητής

Η πιθανότητα η τ.μ. X να λάβει την τιμή x (X=x) είναι ίση με την πιθανότητα του ενδεχόμενου το x να γίνει γεγονός.

Ο κανόνας που δίνει το μέτρο πιθανότητας κάθε δυνατής τιμής μιας τ.μ. X είναι η κατανομή πιθανότητας ή «Νόμος πιθανότητας».

Περιγραφή της κατανομής πιθανότητας.

5.2.1. Διακριτές κατανομές πιθανότητας (για διακριτές τιμές X)

5.2.1.1. Συνάρτηση πιθανότητας ή κατανομής πιθανότητας.

Θεωρούμε μια τ.μ. X διακριτή να παίρνει τιμές αύξουσες $x_1, x_2, \ldots\ldots$

Ας θεωρήσουμε ότι οι πιθανότητες να πάρει η τ.μ. X τις τιμές αυτές είναι:

$P(xi) = P(X = xi) = f(xi) \quad$ με $\quad i = 1,2,.....$

Ορίζουμε την συνάρτηση πιθανότητας (ΣΠ) ή κατανομή πιθανότητας ως

$P(x) = P(X = x) = f(x)$

έτσι ώστε: για X = $x_1, x_2, \ldots\ldots$ $P(X = xi) = f(xi)$

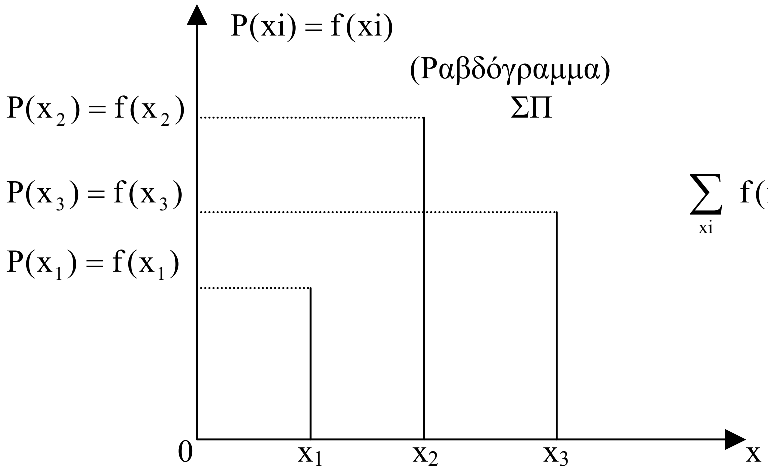
και για τις υπόλοιπες τιμές του $x : f(x) = 0$

Η συνάρτηση αυτή ορίζεται μόνο στις διακριτές τ.μ.

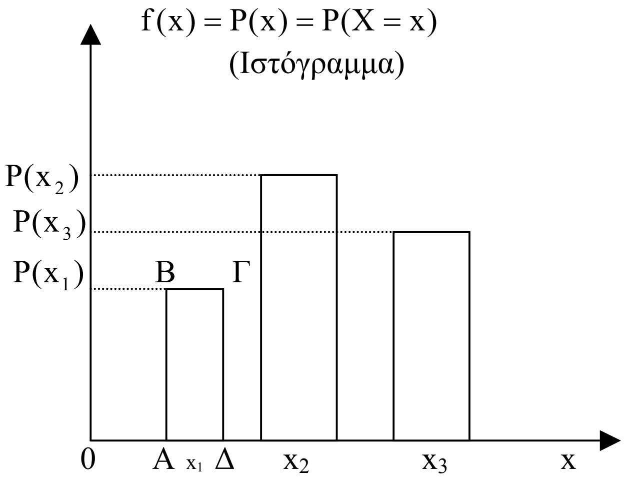
Η $f(x)$ για να είναι συνάρτηση πιθανότητας ικανοποιεί δύο βασικές προϋποθέσεις:

1. $f(x) \geq 0$
2. $\sum_{xi} f(xi) = 1$, για τις δυνατές τιμές xi της X.

Γραφική παράσταση της f(x) (υπό μορφή Ραβδογραμμάτων ή Ιστογραμμάτων).



$$\sum_{xi} f(xi) = \sum_{xi} P(xi) = 1$$



Παρατήρηση:
Θεωρούμε την τ.μ. X «συνεχή» με πλάτος γύρω από το xi ίσο με την μονάδα. Το εμβαδόν (ΑΒΓΔ) =(AB)=f(xi).
Το άθροισμα των εμβαδών των καθέτων παραλληλο - γράμμων ισούται με την μονάδα.

5.2.1.2. Αθροιστική συνάρτηση κατανομής (ΑΣΚ).

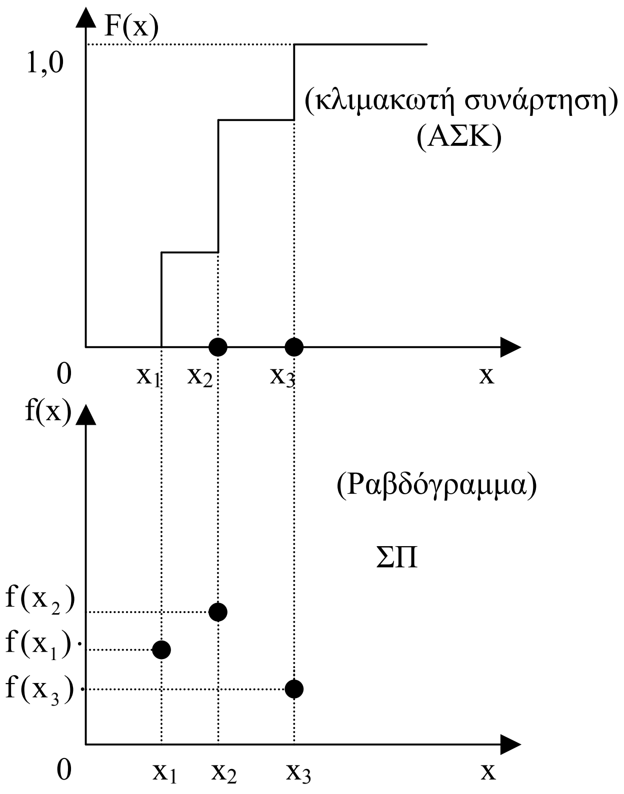
Ως αθροιστική συνάρτηση κατανομής για τ.μ. X διακριτή ορίζουμε την συνάρτηση:

$$F(x) = P(X \le x) \quad \text{με} \quad -\infty < x < +\infty \; \text{(ορίζεται και για συνεχείς τ.μ.)}$$
$$F(x) = P(X \le x) = \sum_{xi \le x} P(xi) = \sum_{xi \le x} P(X = xi)$$

Η $F(x)$ ορίζεται ως εξής:

$$F(x) = \begin{cases} 0, & -\infty < x < x_1 \\ f(x_1), & x_1 \le x < x_2 \\ f(x_1) + f(x_2) & x_2 \le x < x_3 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots & \\ f(x_1) + \dots + f(x_n), & x_n \le x < +\infty \end{cases}$$

Γραφική παράσταση της ΑΣΚ και της ΣΠ.



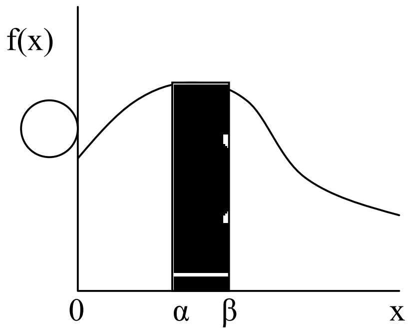
$$\sum_{i=1}^{3} f(xi) = 1$$

5.2.2. Συνεχείς κατανομές πιθανότητας (για συνεχείς τ.μ. X).

5.2.2.1. Συνάρτηση πυκνότητας πιθανότητας (ΣΠΠ) - Αθροιστική συνάρτηση κατανομής (ΑΣΚ).

Η f (x) ονομάζεται συνάρτηση πυκνότητας πιθανότητας για την οποία ισχύουν αξιωματικά ότι:

1. $f(x) \geq 0 \qquad \forall x \varepsilon R$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$



Και ορίζεται εκ της σχέσεως: *

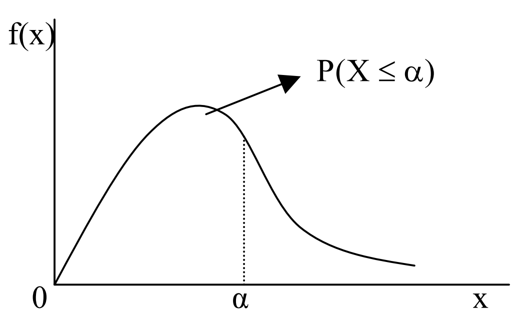
$$P(\alpha \leq X < \beta) = P(\alpha \leq X \leq \beta) = P(\alpha < X < \beta) = P(\alpha < X \leq \beta) = \int_{\alpha}^{\beta} f(x)dx.$$

που δίνει την πιθανότητα η τ.μ. X να παίρνει τιμές μεταξύ των α, β (ορίζεται μόνο για συνεχείς τ.μ. X).

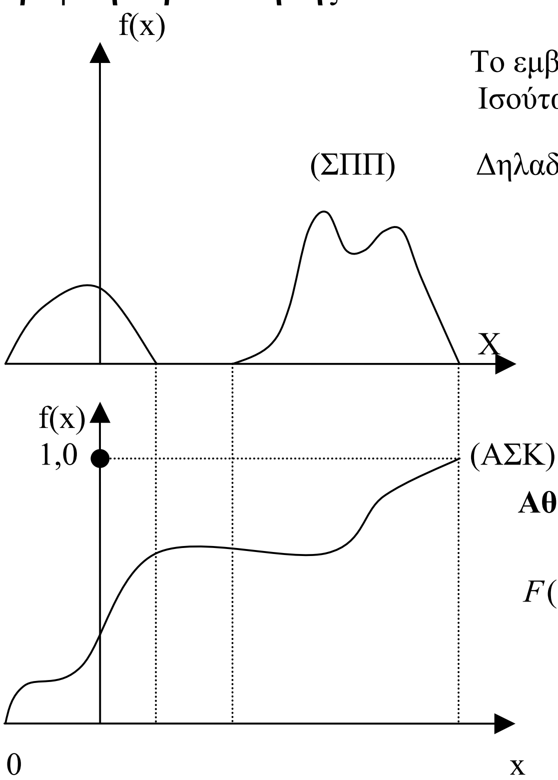
Η ισότητα των πιθανοτήτων $P(\alpha \leq X < \beta) = ...$προκύπτει από το γεγονός ότι η πιθανότητα σημείων στις συνεχείς τ.μ. X είναι μηδέν.

Ισχύουν οι σχέσεις:

1. $P(\alpha \leq X \leq \beta) = P(X \leq \beta) - P(X \leq \alpha)$
2. $P(X \leq \alpha) = \int_{-\infty}^{\alpha} f(x)dx.$
3. $P(X \geq \alpha) = 1 - P(X \leq \alpha)$



Γραφική παράσταση της ΣΠΠ.



Το εμβαδόν κάτω από την ΣΠΠ Ισούται με 1.0

Δηλαδή $P(-\infty \le X \le +\infty) = \int_{-\infty}^{\infty} f(x)dx = 1$

Αθροιστική συνάρτηση κατανομής (ΑΣΚ).

$$F(x) = P(-\infty < X \le x) = \int_{-\infty}^{x} f(\xi)d\xi$$

Εφόσον υπάρχει:

$$\frac{dF(x)}{dx} = f(x)$$

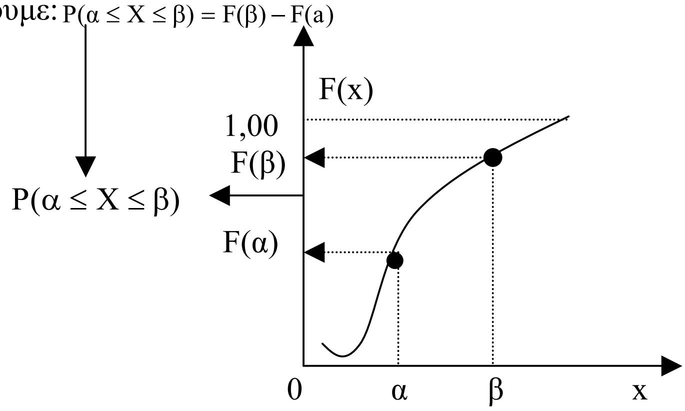
Προσοχή: Η f (x) δεν εκφράζει πιθανότητα. Το γινόμενο **f(x)dx** εκφράζει την πιθανότητα $P(x \le X \le x + dx)$.

Παρατήρηση:

Για την F(x) ως συνάρτηση κατανομής της X ισχύουν οι ιδιότητες:

1. $F(-\infty) = 0$ και $F(+\infty) = 1$	2. $F(x) \ge 0$ $\forall \chi \varepsilon R$	3. Η F (x) είναι μια συνεχής μη φθίνουσα συνάρτηση του X. $F(x_1) \le F(x_2)$ $\mu\varepsilon$ $x_1 \le x_2$

4. Έχουμε: $P(\alpha \le X \le \beta) = F(\beta) - F(a)$



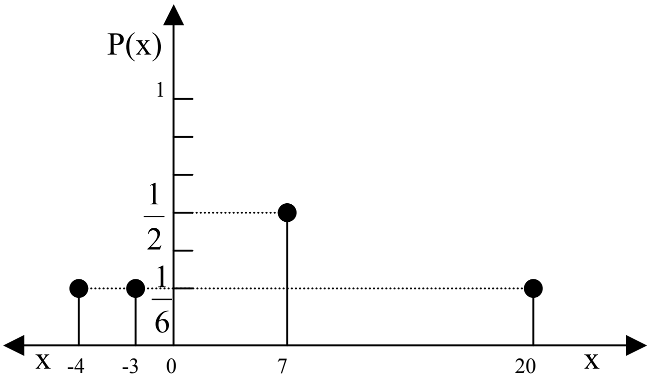
Παραδείγματα

1. Σε ένα τυχερό παιχνίδι ο παίκτης ρίχνει ένα ζάρι και κερδίζει 20 δραχμές αν έρθει η όψη «5», 7 δραχμές αν έρθει μια από τις όψεις «1», «2», «6» και χάνει 3 και 4 δραχμές αν έρθει αντίστοιχα η όψη «4» ή «3». Έστω Χ η διακριτή τυχαία μεταβλητή που εκφράζει το χρηματικό αποτέλεσμα του παιχνιδιού δηλ. οι τιμές της είναι:

$x_1=20$, $x_2= 7$, $x_3=-3$, $x_4= -4$.

Ζητείται η κατανομή πιθανοτήτων της διακριτής τυχαίας μεταβλητής Χ όπως και η γραφική της παράσταση.

x_i	P(xi)
20	$\frac{1}{6}$
7	$\frac{1}{2}$
-3	$\frac{1}{6}$
-4	$\frac{1}{6}$

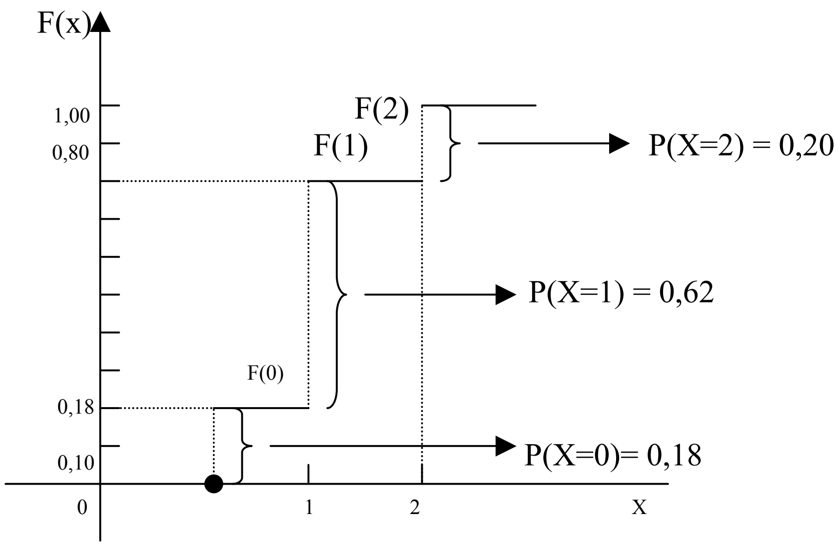


2. Σε ένα ποιοτικό έλεγχο ενός προϊόντος, από ένα σύνολο ομοειδών προϊόντων που παίρνουμε με απλή τυχαία δειγματοληψία, έστω Χ η τυχαία διακριτή μεταβλητή που μετρά την ποιότητα του προϊόντος.

Η τυχαία μεταβλητή Χ ανάλογα με το εάν το προϊόν είναι ελαττωματικό, καλό ή άριστο παίρνει τις τιμές $x_1 = 0$, $x_2 = 1$, $x_3 = 2$ αντίστοιχα με πιθανότητες $P(x_1) = 0,18$, $P(x_2) = 0,62$, $P(x_3) = 0,20$.

Να βρεθεί η αθροιστική συνάρτηση πιθανότητας της Χ και να γίνει η γραφική της παράσταση.

$$F(x) = \begin{cases} 0 & , x < 0 \\ 0,18 & ,0 \le x < 1 \\ 0,80 & ,1 \le x < 2 \\ 1,00 & ,2 \le x \end{cases}$$



3. Ο χρόνος σε ώρες που μπορεί ένα ευπαθές αγροτικό προϊόν να μένει εκτός ψυγείου το καλοκαίρι θεωρείται μια συνεχής τυχαία μεταβλητή Χ με συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \begin{cases} 80/x^2 & \gamma\iota\alpha \quad x > 80 \\ 0 & \gamma\iota\alpha \quad x \le 80 \end{cases}$$

Ποια η πιθανότητα το ευπαθές αυτό αγροτικό προϊόν να διατηρείται εκτός ψυγείου χρόνο που κυμαίνεται μεταξύ 130 και 160 ώρες. Να βρεθεί η αθροιστική συνάρτηση πιθανότητας.

Έχουμε: $P(130 < X < 160) =$

$$P(130 < X < 160) = \int_{130}^{160} \frac{80}{x^2} dx = 80\int_{130}^{160} x^{-2} dx =$$

$$= 80\left[\frac{x^{-2+1}}{-2+1}\right]_{130}^{160} = 80\left(-\frac{1}{x}\right)_{130}^{160} = -80\left(\frac{1}{160} - \frac{1}{130}\right)$$

$$P(130 < X < 160) = -8\left(\frac{13-16}{13\cdot 16}\right) = +\frac{3\cdot 8}{13\cdot 16} = \frac{3}{26} = 0{,}12.$$

Επίσης μπορεί να βρεθεί η P(130<X<160) μέσω της σχέσης:

P (130<X<160)=F(160)-F(130).

Γι' αυτό χρειάζεται να υπολογίσουμε την αθροιστική συνάρτηση πιθανότητας.

$$F(X) = \int_{-\infty}^{x} f(t)dt = \int_{80}^{x} \frac{80}{t^2}dt = 80\int_{80}^{x} t^{-2}dt$$

$$F(X) = 80\left[\frac{t^{-1}}{-1}\right]_{80}^{x} = 80\left(-\frac{1}{x}+\frac{1}{80}\right) = 1-\frac{80}{x}$$

Συνεπώς: P (130<X<160)=F(160)-F(130)=

$$=\left(1-\frac{80}{160}\right)-\left(1-\frac{80}{130}\right) = \frac{80}{130}-\frac{80}{160} = \frac{8}{13}-\frac{8}{16} = \frac{24}{208} = 0{,}12.$$

5.3. Μέση τιμή και διακύμανση τυχαίας μεταβλητής X.

5.3.1. Μέση τιμή ή Αναμενόμενη τιμή ή μαθηματική ελπίδα ή ροπή πρώτης τάξης της X.

i/ **Διακριτή τυχαία μεταβλητή X.**
Συμβολίζεται με E(X) = μ και ισούται:

$\mu = E(X) = \sum_{i=1} xi \cdot P(xi),$ όπου P(xi) η συνάρτηση πιθανότητας.

ii/ **Συνεχής τυχαία Μεταβλητή X.**
Συμβολίζεται επίσης με E(X) = μ και ισούται:

$\mu = E(X) = \int_{-\infty}^{+\infty} xf(x)dx,$ όπου f(x) η συνάρτηση πυκνότητας πιθανότητας.

Παραδείγματα

1. Έστω σε ένα τυχερό παιχνίδι «ρίψη νομίσματος» όταν πέσει η όψη «κεφαλή» ο παίκτης κερδίζει 30 δραχμές ενώ αν πέσει «γράμματα» χάνει 8 δραχμές.

Αν θεωρήσουμε X την τυχαία διακριτή μεταβλητή που παίρνει ως τιμές τα χρηματικά αποτελέσματα του παιχνιδιού, δηλαδή x_1 = 30, x_2 = -8 με αντίστοιχες πιθανότητες $P(x_1) = P(30) = \frac{1}{2}$ και $P(x_2) = P(-8) = \frac{1}{2}$

τότε η μέση τιμή της X είναι:

$$\mu = \mathrm{E(X)} = \sum_{i=1}^{2} x_i \cdot P(x_i) = x_1 P(x_1) + x_2 P(x_2) = 30 \cdot \frac{1}{2} - 8 \cdot \frac{1}{2} = 11$$

2. Έστω η συνεχής τυχαία μεταβλητή X με συνάρτηση πυκνότητας πιθανότητας

$$\mathrm{f(x)} = \begin{cases} \dfrac{2}{17} & \text{όταν} \quad 0 \le x \le 17 \\ 0 & \text{για τις υπόλοιπες τιμές του X.} \end{cases}$$

Έχουμε ως μέση τιμή E(X):

$$\mu = \mathrm{E(X)} = \int_{-\infty}^{\infty} \mathrm{xf(x)dx} = \int_{0}^{17} \mathrm{x} \cdot \frac{2}{17}\mathrm{dx} = \frac{2}{17}\left[\frac{\mathrm{x}^2}{2}\right]_0^{17} = \frac{2}{17} \cdot \frac{17^2}{2} = 17$$

3. Στο παράδειγμα του αγρότη με την ετήσια δαπάνη για την αγορά λιπασμάτων το πολύ 300.000 δραχμές ετησίως για τρεις συνεχόμενες χρονιές, έχουμε την μέση τιμή της διακριτής μεταβλητής X με τις τιμές $x_1=0$, $x_2= 1$, $x_3=2$, $x_4=3$ και τις αντίστοιχες πιθανότητες

$$\mathrm{P(x_1)} = \frac{1}{8}, \qquad \mathrm{P(x_2)} = \frac{3}{8}, \qquad \mathrm{P(x_3)} = \frac{3}{8}, \qquad \mathrm{P(x_4)} = \frac{1}{8}.$$

$$\mu = \mathrm{E(X)} = \sum_{\mathrm{i}=1}^{4} \mathrm{x_1 \cdot P(x_1) = x_1P(x_1) + x_2P(x_2) + x_3P(x_3) + x_4P(x_4)} =$$

$$= 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = 1{,}5$$

5.3.2. Διακύμανση τυχαίας μεταβλητής X - Τυπική απόκλιση μεταβλητής X.

i) **Διακριτή τυχαία μεταβλητή X:** Η Διακύμανση της διακριτής τυχαίας μεταβλητής X συμβολίζεται με Var (X) και ισούται με:

ii)

$\mathrm{Var(X)} = \mathrm{E(X^2)} - [\mathrm{E(X)}]^2$. (Υπάρχει και ο συμβολισμός σ^2).

Η ποσότητα $\mathrm{E(X^2)}$ λέγεται ροπή δεύτερης τάξης και είναι:

$$E(X^2) = \sum_{i=1}^{\nu} x_i^{\,2} P(x_i)$$

Η τυπική απόκλιση $\sigma = \sqrt{\mathrm{Var(X)}}$ είναι η θετική τετραγωνική ρίζα της διακύμανσης.

ii) Συνεχής τυχαία μεταβλητή Χ:

Η διακύμανση της συνεχούς τυχαίας μεταβλητής Χ συμβολίζεται με Var (X) και ισούται επίσης με:

$$\mathrm{Var(X)} = \mathrm{E(X^2)} - [\mathrm{E(X)}]^2. = \int_{-\infty}^{+\infty} (x - E[x])^2 f(x)dx$$

Η ποσότητα $E(X^2)$ λέγεται επίσης ροπή δεύτερης τάξης και δίνεται από τον τύπο:

$$\mathrm{E(X^2)} = \int_{-\infty}^{\infty} \mathrm{x^2 f(x)dx}.$$

Γενικότερα **οι ροπές ρ - τάξης** μιας τυχαίας μεταβλητής Χ είναι:

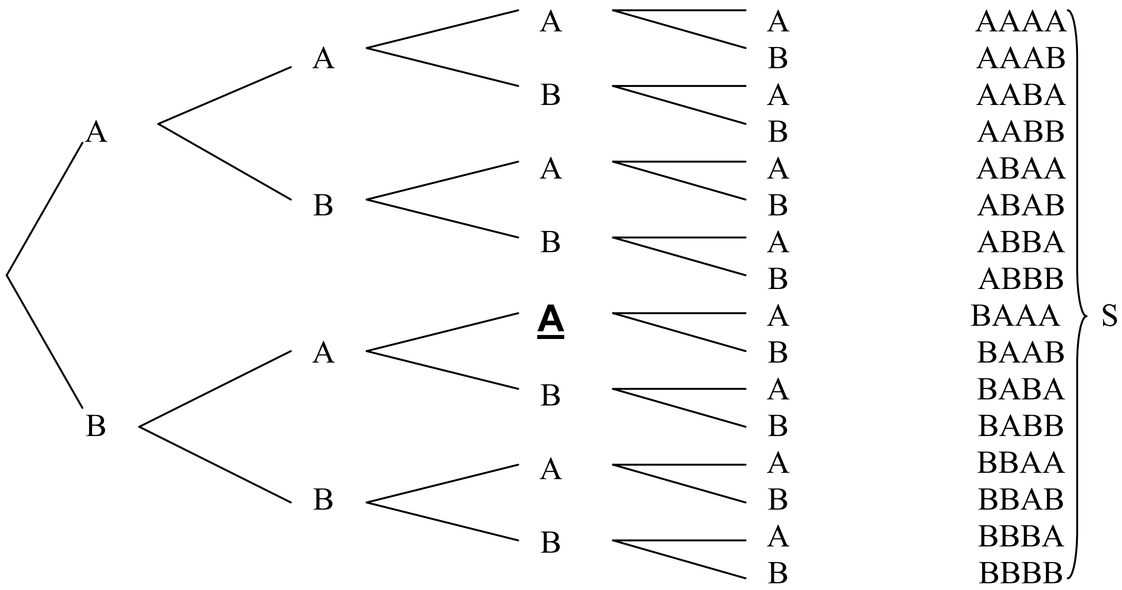
$$\mathrm{E(X^p)} = \sum_{i=1}^{\nu} \mathrm{x_i^{\,2} P(x_i)} \quad \text{(για διακριτή τυχαία μεταβλητή Χ).}$$

$$\mathrm{E(X^p)} = \int_{-\infty}^{\infty} \mathrm{x^p f(x)dx} \quad \text{(για συνεχή τυχαία μεταβλητή Χ).}$$

Παραδείγματα

1. Τα διευθυντικά στελέχη μιας μεγάλης εξαγωγικής βιομηχανικής μονάδας έχουν σχεδιάσει για τα επόμενα τέσσερα συνεχή χρόνια να πραγματοποιήσουν εξαγωγές κατ' έτος τουλάχιστον 1 δισεκατομμυρίου δραχμών.

Αν θεωρήσουμε ως Α το ενδεχόμενο να έχουν ετήσιες εξαγωγές ≥ 1 δις δραχμές και Β το ενδεχόμενο να έχουν ετήσιες εξαγωγές < 1 δις – τότε θα έχουμε για τα επόμενα τέσσερα χρόνια τον δειγματικό χώρο:



Σχηματίζουμε Δενδρόγραμμα για τον προσδιορισμό του δειγματικού χώρου s.

Μπορούμε να ορίσουμε μια συνάρτηση X που να αντιστοιχεί σε κάθε στοιχείο του δειγματικού χώρου S τον αριθμό των ετών όπου οι ετήσιες εξαγωγές είναι τουλάχιστον 1 δις δραχμές.

Κατ' αυτόν τον τρόπο ορίζουμε την τυχαία διακριτή μεταβλητή X που παίρνει τις τιμές: $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = 3$, $x_5 = 4$, με τις αντίστοιχες πιθανότητες:

$$P(x_1) = \frac{1}{16}, \qquad P(x_2) = \frac{4}{16}, \qquad P(x_3) = \frac{6}{16}, \qquad P(x_4) = \frac{4}{16}, \qquad P(x_5) = \frac{1}{16}.$$

Διακύμανση της διακριτής αυτής τυχαίας μεταβλητής X :

$$Var(X) = E(X^2) - [E(X)]^2.$$

$$\mu = E(X) = \sum_{i=1}^{5} x_i \cdot P(x_i) = x_1 \cdot P(x_1) + x_2 P(x_2) + x_3 P(x_3) + x_4 P(x_4) + x_5 P(x_5) =$$

$$= 0 \cdot \frac{1}{16} + 1 \cdot \frac{4}{16} + 2 \cdot \frac{6}{16} + 3 \cdot \frac{4}{16} + 4 \cdot \frac{1}{16} = \frac{32}{16} = 2.$$

$$E(X^2) = \sum_{i=1}^{5} x_i^{\,2} P(x_i) = 0 \cdot \frac{1}{16} + 1^2 \cdot \frac{4}{16} + 2^2 \cdot \frac{6}{16} + 3^2 \cdot \frac{4}{16} + 4^2 \cdot \frac{1}{16} = \frac{80}{16} = 5$$

$$E(X^2) = 5$$

Οπότε: Var (X) = 5-2^2 = 5-4 =1,

Και η τυπική απόκλιση είναι: $\sigma = \sqrt{Var(X)} = 1.$

2. Να υπολογισθεί η διακύμανση και η τυπική απόκλιση της συνεχούς τυχαίας μεταβλητής X με συνάρτηση πυκνότητας πιθανότητας $f(x) = 0.2x, \quad x \in [1,3]$.

$$Var(X) = E(X^2) - [E(X)]^2.$$

$$E(X) = \int_1^3 x(0,2x)dx = 0,2\int_1^3 x^2 dx = 0,2\frac{x^3}{3}\int_1^3 \; = 0,2\left(\frac{3^3}{3} - \frac{1}{3}\right)$$

$$E(X) = 0,2\frac{26}{3} = 1,73$$

$$E(X^2) = \int_1^3 x^2 \cdot 0,2 \cdot x dx = 0,2\int_1^3 x^3 dx = 0,2\frac{x^4}{4}\int_1^3 \; = \frac{0,2}{4}\left(3^4 - 1\right)$$

$$E(X^2) = 4$$

$$Var(X) = 4 - 1,73^2 = 4 - 2,99 = 1,01.$$

$$\sigma = \sqrt{Var(X)} \approx 1.$$

3. Δίνεται η συνεχής τυχαία μεταβλητή X που παίρνει τιμές στο διάστημα [0,2] και έχει συνάρτηση πυκνότητας πιθανότητας $f(x) = 1 - kx, k\varepsilon R$.

Να βρεθεί η σταθερά κ και να υπολογισθούν:

i. Η μέση τιμή

ii. Η τυπική απόκλιση

iii. Η πιθανότητα P(1<X<2).

iv. Η συνάρτηση κατανομής και να γίνει η γραφική της παράσταση.

Θα πρέπει $\int_0^2 f(x)dx = 1$ (εφόσον το εμβαδόν του χωρίου μεταξύ των f(x), του άξονα των x και των ευθειών x = 0 και x = 2 ισούται με μονάδα).

$$\int_0^2 (1-\kappa x)dx = 1 \Leftrightarrow \int_0^2 dx - \int_0^2 \kappa dx = 1 \Leftrightarrow x\Big|_0^2 - \kappa\frac{x^2}{2}\Big|_0^2 = 1$$

$$\Leftrightarrow 2 - \kappa\frac{2^2}{2} = 1 \Leftrightarrow 2\kappa = 1 \Leftrightarrow \kappa = \frac{1}{2}.$$

Άρα $f(x) = 1 - kx = 1 - \frac{1}{2}x$.

i: $\mu = E(X) = \int_0^2 x \cdot f(x)dx = \int_0^2 x \cdot \left(1 - \frac{1}{2}x\right)dx =$

$$= \frac{x^2}{2}\Big|_0^2 - \frac{1}{2}\frac{x^3}{3}\Big|_0^2 = \frac{2^2}{2} - \frac{1}{2}\frac{2^3}{3} = 2 - \frac{4}{3} = \frac{2}{3} = 0{,}67.$$

$$E(X^2) = \int_0^2 x^2 f(x)dx = \int_0^2 x^2\left(1 - \frac{1}{2}x\right)dx = \frac{x^3}{3}\Big|_0^2 - \frac{1}{2}\frac{x^4}{4}\Big|_0^2$$

$$= \frac{2^3}{3} - \frac{1}{2}\frac{2^4}{4} = \frac{8}{3} - 2 = \frac{2}{3} = 0{,}67.$$

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = 0{,}67 - 0{,}67^2 = 0{,}67 - 0{,}45 = 0{,}22.$$

ii. $\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{0{,}22} = 0{,}47.$ Τυπική απόκλιση.

iii. $P(1 < X < 2) = P(X < 2) - P(X < 1)$

$$P(X < 2) = \int_0^2 f(x)dx = \int_0^2 \left(1 - \frac{1}{2}x\right)dx =$$

$$= x\Big|_0^2 - \frac{1}{2} - \frac{x^2}{2}\Big|_0^2 = 2 - \frac{2^2}{4} = 2 - 1 = 1.$$

$$P(X < 1) = \int_0^1 f(x)dx = \int_0^1 \left(1 - \frac{1}{2}x\right)dx = x\left(\Big|_0^1 - \frac{1}{2}\frac{x^2}{2}\right)\Big|_0^1 =$$

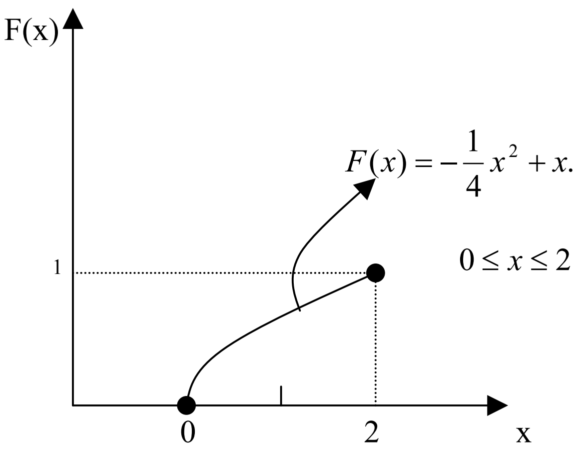
$$= 1 - \frac{1}{2}\cdot\frac{1^2}{2} = 1 - \frac{1}{4} = \frac{3}{4} = 0{,}75.$$

Άρα $P(1 < X < 2) = 1 - \frac{3}{4} = \frac{1}{4} = 0{,}25.$

iv. $F(X) = \int_{-\infty}^{x} f(\xi)d\xi$

$$F(X) = \int_0^x f(\xi)d\xi = \int_0^x \left(1 - \frac{1}{2}\xi\right)d\xi = \xi\Big|_0^x - \frac{1}{2}\frac{\xi^2}{2}\Big|_0^x$$

$$F(X) = x - \frac{1}{2}\frac{x^2}{2} = -\frac{1}{4}x^2 + x. \qquad x \in [0,2].$$



Κεφάλαιο 6ο

Βασικές θεωρητικές κατανομές πιθανότητας

6.1. Εισαγωγή.

Στην θεωρία των πιθανοτήτων έχουν αναπτυχθεί θεωρητικές κατανομές πιθανότητας οι οποίες προήλθαν μέσα από υποθέσεις που ελήφθησαν εξ' αρχής.

Οι θεωρητικές αυτές κατανομές αποτελούν μαθηματικά προτεινόμενα μοντέλα για να περιγράψουν φαινόμενα τα οποία παρουσιάζουν μεταβλητότητα και τα οποία κατ' αρχήν επιχειρούνται να περιγραφούν με νόμους μεταβλητότητας.

Στην πράξη όταν θέλουμε να κρίνουμε αν μια θεωρητική κατανομή αποτελεί κατάλληλο μαθηματικό μοντέλο πιθανότητας για μια τυχαία μεταβλητή X, που ορίσθηκε σ' ένα πείραμα, πρέπει να γνωρίζουμε ότι ισχύουν και για το πείραμα οι προϋποθέσεις που έχουν τεθεί για την θεωρητική κατανομή.

Στην περίπτωση που δεν λάβουμε υπόψη τις προϋποθέσεις που τίθενται για την θεωρητική κατανομή μπορούμε να συγκρίνουμε αν οι πιθανότητες από την θεωρητική κατανομή που χρησιμοποιήσαμε και σε τι βαθμό προσαρμόζονται στις σχετικές συχνότητες των παρατηρήσεων. Ο στατιστικός έλεγχος της προσαρμογής μιας θεωρητικής κατανομής σε μια κατανομή που προέκυψε από παρατηρήσεις αποτελεί βασικό τμήμα της Στατιστικής και περιλαμβάνει την προσαρμογή κατανομών δειγματοληψίας, την κατασκευή των διαστημάτων εμπιστοσύνης, τους ελέγχους υποθέσεων κ.α.

Οι βασικότερες θεωρητικές κατανομές που παρουσιάζονται στη συνέχεια είναι για μεν τις διακριτές κατανομές (που αφορούν τις διακριτές τυχαίες μεταβλητές) η κατανομή Bernoulli, η Διωνυμική κατανομή, η κατανομή Poisson, η Γεωμετρική κατανομή, η Αρνητική Διωνυμική Κατανομή και για δε τις συνεχείς κατανομές (που αφορούν τις συνεχείς τυχαίες μεταβλητές) η κανονική κατανομή, η τυπική κανονική κατανομή, η Αρνητική Εκθετική κατανομή, η κατανομή X^2, η κατανομή t-student, η κατανομή F, η ορθογώνια κατανομή.

6.2. Ειδικές κατανομές διακριτής τυχαίας μεταβλητής X.

6.2.1. Κατανομή Bernoulli

Αν σε ένα πείραμα τύχης υπάρχουν δύο μόνο δυνατά αποτελέσματα τότε το πείραμα καλείται **δοκιμή Bernoulli.**

Ακολουθίες Bernoulli ή Επαναληπτικές δοκιμές καλούνται ν-ανεξάρτητες επαναλήψεις του πειράματος με δύο δυνατά αποτελέσματα τα οποία ονομάζονται.

Επιτυχία (Ε): Όταν έχουμε πραγματοποίηση του ενδεχομένου.

Αποτυχία (Α): Όταν δεν πραγματοποιείται το ενδεχόμενο.

Η πιθανότητα της επιτυχίας Ρ(Ε) = ρ είναι σταθερή για όλες τις δοκιμές.
Σε κάθε δοκιμή Bernoulli ορίζουμε μια τυχαία μεταβλητή Χ που παίρνει τις τιμές $x_1 = 0$ και $x_2 = 1$ με πιθανότητες αντίστοιχα ρ επιτυχίας και q αποτυχίας.

Έχουμε: $q + \rho = 1$

Δηλαδή έχουμε: $P(x) = \begin{cases} \rho, & \text{όταν} \quad x=1 \\ q, & \text{όταν} \quad x=0 \end{cases}$

ή $P[X = x] = p^x q^{1-x}$ με $x = 0,1.$

Για $x = 0 \rightarrow P[X = 0] = p^0 q^1 = q.$

$x = 1 \rightarrow P[X = 1] = p^1 q^0 = p.$

Έχουμε ότι η μέση τιμή και η διακύμανση στην κατανομή Bernoulli της τυχαίας μεταβλητής Χ είναι:

$$E(X^2) = \sum_{I=1}^{2} x_i^2 P(x_i) = 0P(0) + 1^2 \cdot p(1) = p.$$

$$E(X) = \sum_{I=1}^{2} x_i P(x_i) = 0P(0) + 1 \cdot p(1) = p.$$

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = p - p^2 = p(1-p) = p \cdot q.$$

Παρατήρηση: Κατανομή Bernoulli έχουμε όταν μας ενδιαφέρει αν ένα προϊόν είναι ελαττωματικό ή όχι, η τιμή ενός προϊόντος ανεβαίνει ή όχι, η τιμή μιας μετοχής ανεβαίνει ή όχι κ.α.

Παράδειγμα:

Μια τυχαία μεταβλητή Χ παίρνει τις τιμές x = 0 με πιθανότητα q = 0,4 και x =1 με πιθανότητα p = 0,6.

Η πιθανότητα $P[X = x] = p^x q^{1-x} = 0{,}6^x \cdot 0{,}4^{1-x}$

$E(X) = p = 0{,}6$ και $\mathrm{Var}(X) = p \cdot q = 0{,}4 \cdot 0{,}6 = 0{,}24.$

6.2.2. Διωνυμική κατανομή B(ν, p)

Έστω μια ακολουθία ν-δοκιμών (ακολουθία Bernoulli), όπως διετυπώθη. Έστω επίσης η τυχαία μεταβλητή X που εκφράζει (μετρά) **τον αριθμό των επιτυχιών στις ν-δοκιμές.**

Παίρνουμε X = x **τον αριθμό των επιτυχιών** που θέλουμε, δηλαδή την πραγματοποίηση ενός ενδεχομένου που την θεωρούμε «επιτυχία», **x φορές στις ν-δοκιμές.**

Έστω ένας συνδυασμός των x-επιτυχιών στις ν-δοκιμές είναι ο εξής:

$$\cdots * \cdots\cdots * \cdot * \cdots\cdots * \cdots$$

x = επιτυχίες

θεωρώ με * την επιτυχημένη δοκιμή με πιθανότητα p (δηλαδή πραγματοποίηση του ενδεχομένου) και με . την αποτυχημένη δοκιμή με πιθανότητα q (μη πραγματοποίηση του ενδεχομένου).

Η πιθανότητα για τον συνδυασμό αυτό θα είναι:

$$p^x \cdot q^{\nu-x} = p^x (1-p)^{\nu-x}.$$

Τέτοιους συνδυασμούς έχουμε $\binom{\nu}{x}$, άρα η τελική πιθανότητα που θα πάρουμε είναι:

$$p[X = x] = \binom{\nu}{x} p^x \cdot q^{\nu-x} = \binom{\nu}{x} p^x (1-p)^{\nu-x}$$

Η μέση τιμή της τυχαίας μεταβλητής X στην διωνυμική κατανομή είναι

E(X) = ν·p και η διακύμανση Var(X) = νpq.

Παράδειγμα

Το 10% των δοκιμών για τον ποιοτικό έλεγχο ενός προϊόντος, έχει την πιθανότητα να μην περάσει το όριο που ετέθη κατά τον ποιοτικό έλεγχο. Παίρνουμε έξη προϊόντα τυχαίως.

Να βρεθεί η πιθανότητα να είναι ελαττωματικά (κάτω του ορίου ποιότητας όπως ετέθη).

i. Ένα εξ’ αυτών
ii. Δύο εξ’ αυτών
iii. Τουλάχιστον δύο προϊόντα
iv. Το πολύ δύο προϊόντα
v. Να βρεθεί η μέση τιμή, η διακύμανση και η τυπική απόκλιση της τυχαίας μεταβλητής Χ.

Θεωρούμε ως «επιτυχία» στο πείραμα του Bernoulli που έχουμε το να είναι το προϊόν ελαττωματικό.

Δηλαδή $p = 0{,}10$ και $q = 0{,}90$.

Η τυχαία μεταβλητή Χ παίρνει τιμές $x = 0,1,2,\ldots,6$ και μετρά τον αριθμό των ελαττωματικών προϊόντων.

i. $p[X = 1] = \binom{6}{1} 0{,}10^1 \cdot 0{,}90^5$

ii. $p[X = 2] = \binom{6}{2} 0{,}10^2 \cdot 0{,}90^4$

iii. $p[X \geq 2] = p[X = 2] + p[X = 3] + p[X = 4] + p[X = 5] + p[X = 6] =$

$$= \binom{6}{2} 0{,}10^2 \cdot 0{,}90^4 + \binom{6}{3} 0{,}10^3 \cdot 0{,}90^3 + \binom{6}{4} 0{,}10^4 \cdot 0{,}90^2 + \binom{6}{5} 0{,}10^5 \cdot 0{,}90 +$$

$$+ \binom{6}{6} 0{,}10^6 \cdot 0{,}90^0. \qquad \text{όπου } \binom{6}{6} = 1.$$

iv. $p[X \leq 2] = p[X = 0] + p[X = 1] + p[X = 2] =$

$$= \binom{6}{0} 0{,}10^0 \cdot 0{,}90^6 + \binom{6}{1} 0{,}10^1 \cdot 0{,}90^5 + \binom{6}{2} 0{,}10^2 \cdot 0{,}90^4.$$

vi. $E(X) = 6 \cdot 0{,}10 = 0{,}60.$

$\mathrm{Var}(X) = 6 \cdot 0{,}10 \cdot 0{,}90 = 0{,}54.$

$\sigma = \sqrt{\mathrm{Var}(X)} = 0{,}73.$

6.2.3. Κατανομή Poisson.

Όταν κατά την Διωνυμική κατανομή έχουμε την πιθανότητα να συμβεί ένα ενδεχόμενο «επιτυχία» να είναι πολύ μικρή $p \leq 0{,}10$ και το πλήθος των δοκιμών ν μεγάλο ν>50, έτσι ώστε η μέση τιμή Ε(Χ) = μ = ν·p να είναι $0 \leq \nu p \leq 10$, τότε χρησιμοποιούμε τον τύπο:

$$P[X = x] = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad \text{όπου} \qquad \lambda = \nu \cdot p \qquad \text{και} \qquad x = 0,1,2,3,.....$$

και η κατανομή ονομάζεται Poisson με Ε[Χ] = Var[X]= λ.

Ο τύπος αυτός δίνει **τον αριθμό των επιτυχιών x** μέσα σε μια διαδικασία δοκιμών Bernoulli, όπου **έχουμε πολλά συμβάντα (ν μεγάλο)** και **μικρή πιθανότητα της επιτυχίας (p μικρό).**

Η κατανομή Poisson χρησιμοποιείται όταν η πραγματοποίηση ενός γεγονότος είναι **τυχαία σε κάποια χρονική στιγμή** και δεν έχει έννοια η μη πραγματοποίηση του ενδεχομένου.

Παραδείγματος χάριν όταν μελετάμε τα τροχαία ατυχήματα που έγιναν σ' ένα δρόμο σε μια χρονική περίοδο. Η κατανομή Poisson χρησιμοποιείται και σε θέματα αφίξεων. (π.χ. αεροπλάνων, πελατών σε σούπερ μάρκετ, γεννήσεων παιδιών, ασθενών στα Νοσοκομεία, κλήσεων σε τηλεφωνικά κέντρα κ.α.).

Με την συνάρτηση πιθανότητας Poisson μελετάμε την πιθανότητα να **πραγματοποιηθεί το ενδεχόμενο (επιτυχία) x φορές μέσα σ' ένα χρονικό διάστημα,** με την προϋπόθεση ότι ο μέσος αριθμός πραγματοποιήσεων του ενδεχομένου στο ίδιο χρονικό διάστημα είναι λ = ν·p (παράδειγμα να προσδιορίσουμε την πιθανότητα να έλθουν x πελάτες σε μια ώρα σ' ένα κατάστημα).

Παράδειγμα

Ο αριθμός των παραγωγών που φθάνουν με το προϊόν τους για επεξεργασία σ' ένα εργοστάσιο σακχάρεως ακολουθεί την κατανομή Poisson. Αν ο μέσος αριθμός των παραγωγών που φθάνει στο εργοστάσιο μέσα σε μια ώρα είναι 6 παραγωγοί, να βρεθεί η πιθανότητα να φθάσουν μέσα σε μια ώρα 5 παραγωγοί.

Έχουμε: $\mathrm{P(x)} = \dfrac{e^{-\lambda}\lambda^{x}}{x!}, \qquad \lambda = 6 \qquad , x = 5$

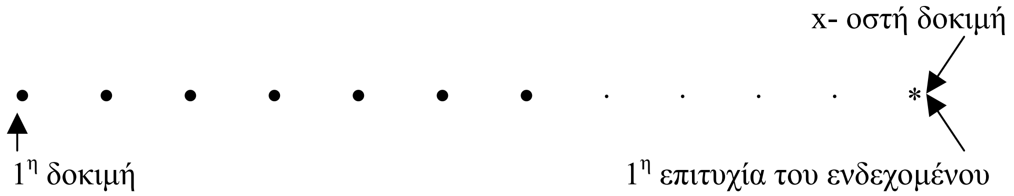
$$\mathrm{P(5)} = \frac{e^{-6}6^{5}}{5!} = \frac{7776}{403{,}43 \cdot 120} = \frac{7776}{484116} = 0{,}016.$$

6.2.4. Γεωμετρική Κατανομή.

Στην ακολουθία Bernoulli ο αριθμός των δοκιμών έως ότου συμβεί για **πρώτη φορά** ένα ορισμένο ενδεχόμενο (η επιτυχία) είναι τυχαία μεταβλητή X με **γεωμετρική κατανομή.**

Αν η πρώτη «επιτυχία» (δηλαδή η πραγματοποίηση του ενδεχομένου) παρουσιαστεί στη ν-οστή δοκιμή Bernoulli θα πρέπει προηγουμένως να έχουμε ν-1 συνεχείς «αποτυχίες».

Αν λοιπόν **x είναι ο αριθμός των δοκιμών μέχρι την 1η επιτυχία** θα έχουμε:



Έχουμε p η πιθανότητα της επιτυχίας μιας δοκιμής και q η πιθανότητα της αποτυχίας μιας δοκιμής.

$$\text{Άρα } P[X = x] = \mathrm{p} \cdot q^{x-1}, \qquad \acute{o}\pi o\upsilon \qquad x = 1,2,3,.....$$

Επίσης έχουμε:

$$E(X) = \frac{1}{p} \qquad \kappa\alpha\iota \qquad Var(X) = \frac{1-p}{p^{2}}$$

Παράδειγμα

Η πιθανότητα να πετύχουμε ένα ελαττωματικό προϊόν είναι ένα σε κάθε 100 προϊόντα. Ποια η πιθανότητα να πετύχουμε ελαττωματικό προϊόν, όταν λαμβάνουμε τυχαίως ένα, την πέμπτη φορά (δοκιμή).

Να βρεθεί η πιθανότητα το ελαττωματικό προϊόν να είναι μέσα στα πέντε πρώτα προϊόντα που ανασύρθηκαν.

Έχουμε x = 5, $p = \dfrac{1}{100} = 0{,}01 \qquad q = 0{,}99$

Άρα $P[X = 5] = p \cdot q^{5-1} = 0{,}01 \cdot 0{,}99^4 = 0{,}096.$

$$P[X \le 5] = \sum_{x=1}^{5} 0{,}01 \cdot 0{,}99^{x-1} = 0{,}01 \cdot 0{,}99^0 + 0{,}01 \cdot 0{,}99 + 0{,}01 \cdot 0{,}99^2 +$$
$$+ 0{,}01 \cdot 0{,}99^3 + 0{,}01.0{,}99^4$$

6.2.5. Αρνητική Διωνυμική Κατανομή.

Στις δοκιμές Bernoulli η πιθανότητα να έχουμε **κ-επιτυχίες στις x-δοκιμές** είναι:

$$\mathrm{P[X = x]} = \binom{\mathrm{x}-1}{\kappa-1} \mathrm{p}^{\kappa} \cdot \mathrm{q}^{\mathrm{x}-\kappa}, \qquad \mathrm{x} = \kappa, \kappa+1.....$$

Η τυχαία μεταβλητή **X μετράει τις δοκιμές για να έχουμε κ επιτυχίες** (x ≥ κ).

Έχουμε ότι: $E(X) = \dfrac{\kappa}{p} \kappa\alpha\iota \qquad Var(X) = \sigma^2 = \dfrac{\kappa(1-p)}{p^2} = \dfrac{\kappa q}{p^2}$

p: η πιθανότητα της επιτυχίας.

q: η πιθανότητα της αποτυχίας, p + q =1.

Παράδειγμα.

Παίρνουμε τυχαία ένα δείγμα 6 αγροτών και τους ρωτάμε αν τα χωράφια τους είναι ποτιστικά. Από τις μελέτες της γεωργικής διεύθυνσης της περιοχής απ΄ όπου κατάγονται οι 12 αγρότες γνωρίζουμε ότι το 60% είναι ποτιστικά τα χωράφια και το 40% δεν είναι.

Ποια η πιθανότητα να έχουμε πάρει 4 θετικές απαντήσεις (δηλαδή να έχουν ποτιστικά χωράφια) όταν ρωτήσουμε και τον 6^o αγρότη.

Έχουμε X τυχαία μεταβλητή που ακολουθεί αρνητική διωνυμική κατανομή.

$$P[X = x] = \binom{x-1}{\kappa-1} p^{\kappa} \cdot q^{x-\kappa}, \acute{o}\pi o \upsilon \qquad x = 6, \kappa = 4.$$

Η τ.μ **Χ μετρά τους αγρότες μέχρι την κ - επιτυχία**.

Δηλαδή έχουμε 6 δοκιμές και 4 «επιτυχίες» = θετικές απαντήσεις. p= 0,60 και q=0,40.

$$\mathrm{P}[\mathrm{X} = 6] = \binom{6-1}{4-1} 0{,}60^4 \cdot 0{,}40^{6-4} = \frac{5!}{3!2!} 0{,}60^4 \cdot 0{,}40^2 = 0{,}21.$$

6.3. Ειδικές κατανομές συνεχούς τυχαίας μεταβλητής Χ.

6.3.1. Κανονική Κατανομή

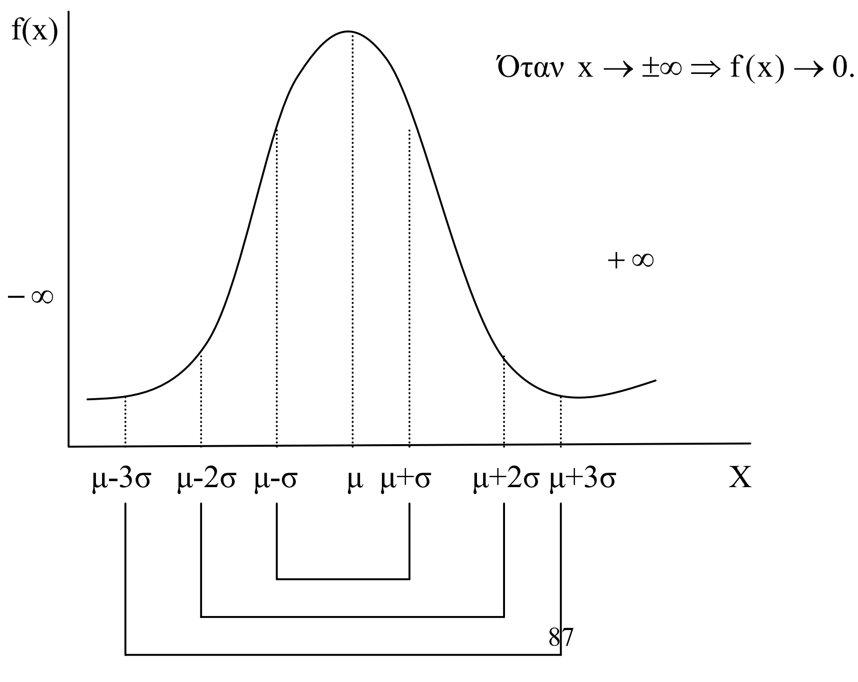
Είναι η κατανομή κατά την οποία μια συνεχής τυχαία μεταβλητή Χ έχει συνάρτηση πυκνότητας πιθανότητας.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < +\infty$$

$\sigma > 0$, $\mu = E(X)$, τυπική απόκλιση και μέση τιμή της Χ.

Ο συμβολισμός της κανονικής κατανομής είναι $N(\mu, \sigma^2)$.

Η γραφική παράσταση της f(x) δίνει μια κωδωνοειδής καμπύλη συμμετρική ως προς την x = μ.



Όπως αναφέρθηκε στην κανονική κατανομή έχουμε:

i) Το 68,27% των τιμών της Χ να βρίσκεται στο διάστημα [μ-σ, μ+σ]

ii) Το 95,45% των τιμών της Χ να βρίσκεται στο διάστημα [μ-2σ, μ+2σ] και

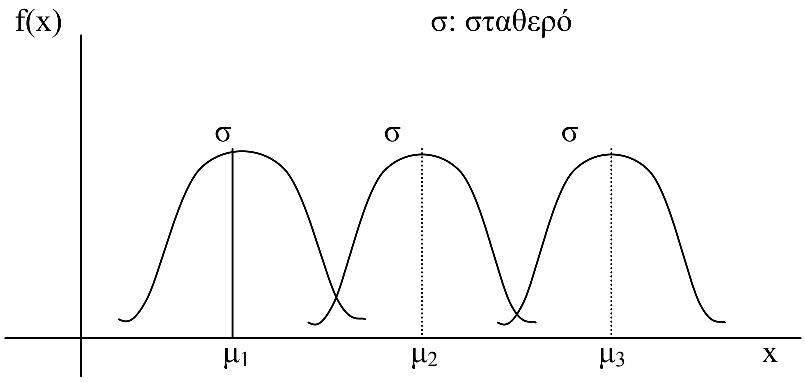
iii) Το 99,73% των τιμών της Χ να βρίσκεται στο διάστημα [μ-3σ, μ+3σ].

Η κανονική κατανομή χρησιμοποιείται πολύ διότι πολλά φαινόμενα έχουν κατανομή που προσεγγίζει την κανονική.

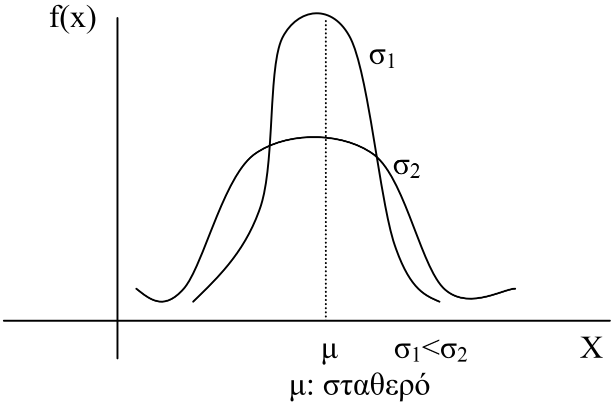
Μεταβλητές που η κατανομή τους προσεγγίζει την κανονική είναι το ύψος, το βάρος και οι διαστάσεις διαφόρων οργάνων του ανθρώπινου σώματος, η ετήσια γαλακτοπαραγωγή των αγελάδων, τα σφάλματα μέτρησης διαφόρων οικονομικών και κοινωνικών μεγεθών. Η ανακάλυψη της κανονικής κατανομής προήλθε από την μελέτη των σφαλμάτων παρατήρησης στη Φυσική και στην Αστρονομία Γι’ αυτό η **κανονική καμπύλη ονομάζεται και καμπύλη των σφαλμάτων.**

Το εμβαδόν κάτω από την κανονική καμπύλη από το $-\infty$ έως το $+\infty$ ισούται με την μονάδα.

Σε διαφορετικές κανονικές κατανομές (καμπύλες) αν έχουμε την ίδια τυπική απόκλιση σ και διαφορετική μέση τιμή θα έχουμε τις μορφές:



Επίσης αν σε διαφορετικές κανονικές κατανομές (καμπύλες) έχουμε την ίδια μέση τιμή μ και διαφορετική τυπική απόκλιση σ θα πάρουμε τις μορφές:



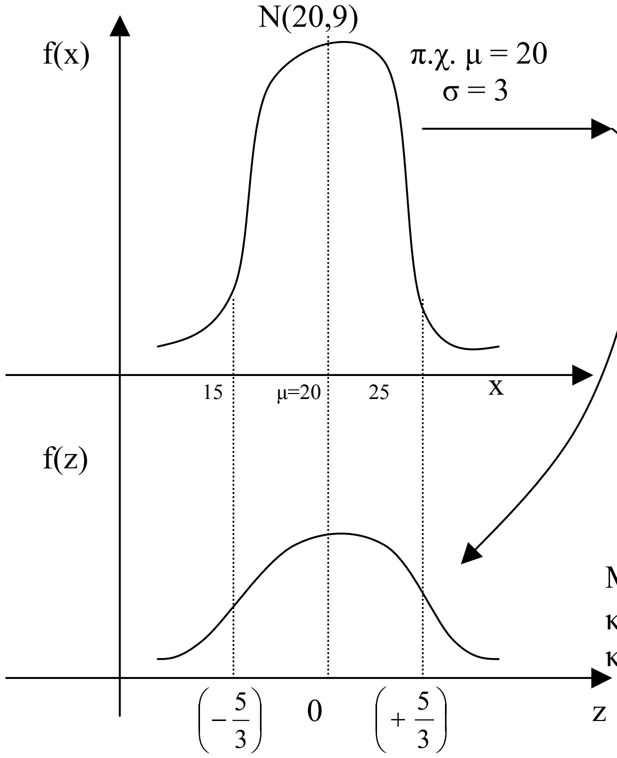
6.3.2. Τυπική Κανονική Κατανομή.

Αν έχουμε μια τυχαία μεταβλητή X με μέση τιμή μ και τυπική απόκλιση σ τότε η τυχαία μεταβλητή $Z = \frac{X - \mu}{\sigma}$ ονομάζεται **τυπική τυχαία μεταβλητή** της X και έχει μέση τιμή μηδέν και τυπική απόκλιση 1.

Όταν η τυχαία μεταβλητή X ακολουθεί κανονική κατανομή τότε η τυπική τυχαία μεταβλητή Z ακολουθεί επίσης κανονική κατανομή.

Η κατανομή που ακολουθεί η μεταβλητή Z καλείται **Τυπική Κανονική Κατανομή** και συμβολίζεται με N (0,1), έχει δε συνάρτηση πυκνότητας:

$$f(z) = \frac{1}{\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, -\infty < z < +\infty$$

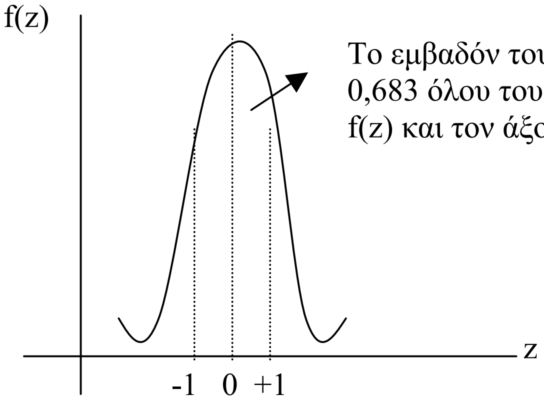


$$z = \frac{x - \mu}{\sigma}$$

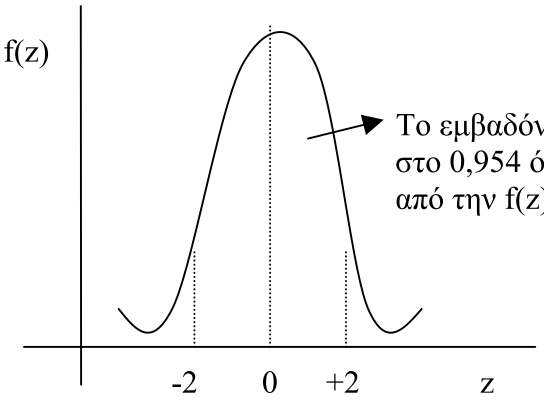
Μετασχηματισμός της κανονικής κατανομής Ν (μ, σ^2) σε τυπική κατανομή Ν (0,1).

$$\frac{15 - 20}{3} = -\frac{5}{3} \qquad \frac{25 - 20}{3} = \frac{5}{3}$$

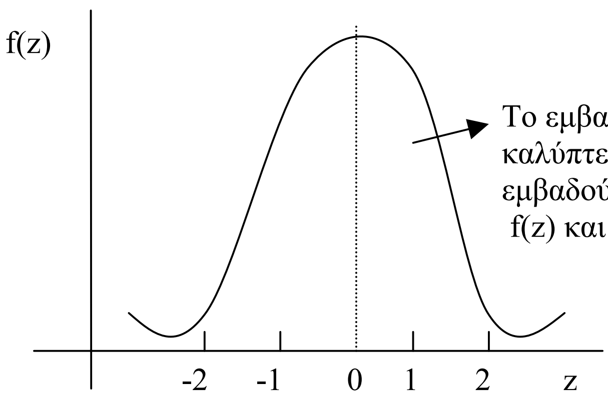
Στην τυπική κανονική κατανομή Ν (0,1) έχουμε:



Το εμβαδόν του χωρίου αυτού αντιστοιχεί στο 0,683 όλου του εμβαδού που ορίζεται από την f(z) και τον άξονα z.



Το εμβαδόν του χωρίου αυτού αντιστοιχεί στο 0,954 όλου του εμβαδού που ορίζεται από την f(z) και του άξονα z.



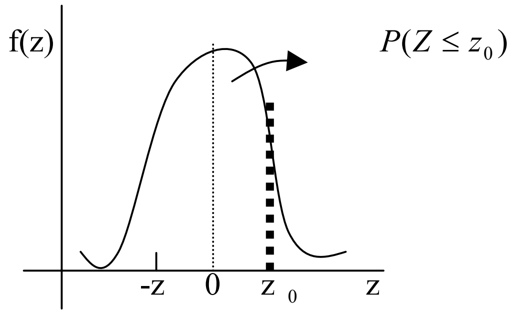
Το εμβαδόν του χωρίου αυτού καλύπτει το 0,997 όλου του εμβαδού που ορίζεται από την f(z) και τον άξονα z.

Η αθροιστική συνάρτηση πυκνότητας πιθανότητας που δίνει τις πιθανότητες $P(Z \le z)$ ορίζεται ως:

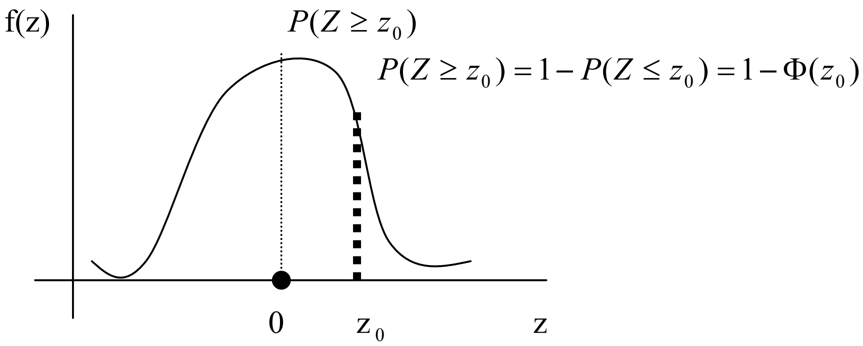
$$\Phi(z) = P(Z \le z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{u^2}{2}} du$$

Η τιμή της Φ(z) δίνεται από πίνακες για τις τιμές z της μεταβλητής Z. Η $\Phi(z) = P(Z \le z)$, δηλαδή εκφράζει την πιθανότητα η Z να παίρνει τιμές μέχρι το z.

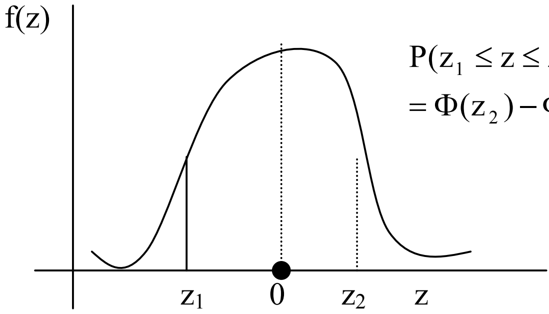
Η πιθανότητα $\Phi(z_0) = P(Z \le z_0)$
εκπροσωπεί το εμβαδόν
αριστερά του z_0 και κάτω της f(z)



Η πιθανότητα $P(Z \ge z_0)$ εκπροσωπεί το εμβαδόν δεξιά του z_0 και κάτω της f(z).



Η πιθανότητα $P(z_1 \le z \le z_2)$ εκπροσωπεί το εμβαδόν μεταξύ των z_1, z_2 και κάτω της f(z).

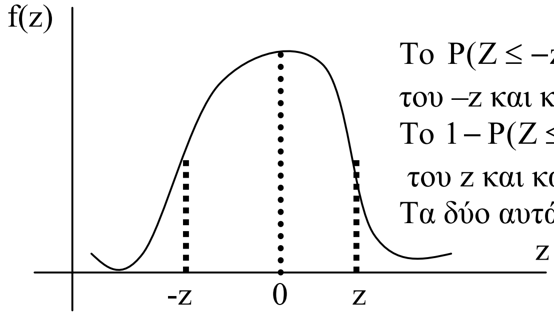


$$P(z_1 \le z \le z_2) = P(z \le z_2) - P(z \le z_1)$$
$$= \Phi(z_2) - \Phi(z_1)$$

Επειδή η τυπική κανονική κατανομή είναι συμμετρική γύρω από το z = 0 είναι

$$\Phi(-z)=1-\Phi(z).$$

Πράγματι έχουμε $P(Z \le -z) = 1 - P(Z \le z)$.



Το $P(Z \le -z)$ εκπροσωπεί το εμβαδόν αριστερά του –z και κάτω της f(z).
Το $1 - P(Z \le z)$ εκπροσωπεί το εμβαδόν δεξιά του z και κάτω της f(z)
Τα δύο αυτά εμβαδά είναι ίσα λόγω συμμετρίας.

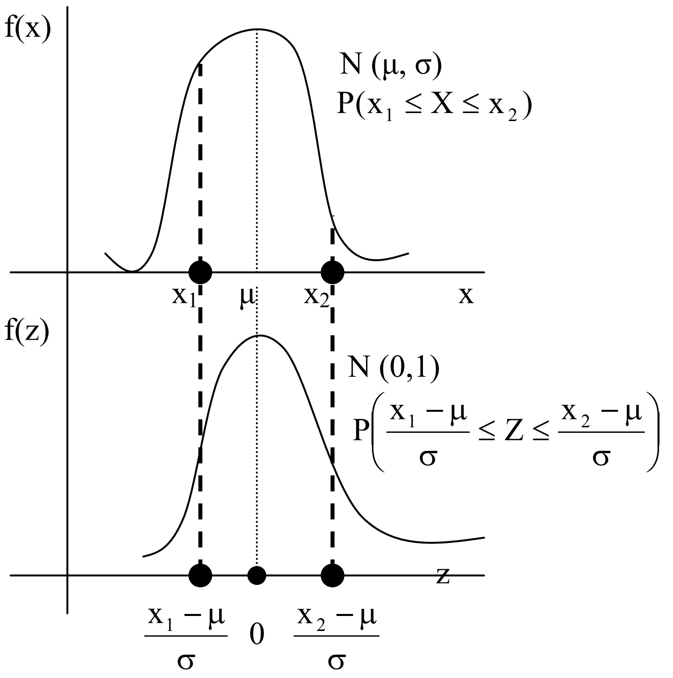
Η σχέση Φ(-z) =1-Φ(z) μας επιτρέπει να υπολογίσουμε πιθανότητες $P(Z \le -z)$ με αρνητικές τιμές του z.

Εκ του ορισμού της $\Phi(z) = P(Z \le z)$ συνεπάγεται ότι:

$$\Phi(+\infty) = 0 \;\text{ και }\; \Phi(-\infty) = 0$$

Αν θέλουμε να υπολογίσουμε την πιθανότητα $P(x_1 \le X \le x_2)$ σε μια κανονική κατανομή Ν (μ, σ) τότε μπορούμε να την ορίσουμε ως εξής:

$$P(x_1 \le X \le x_2) = P\left(\frac{x_1-\mu}{\sigma} \le Z \le \frac{x_2-\mu}{\sigma}\right) = \Phi\left(\frac{x_2-\mu}{\sigma}\right) - \Phi\left(\frac{x_1-\mu}{\sigma}\right).$$



Παραδείγματα

1. Η κατανομή των μισθών 1000 εργαζομένων σε μια επιχείρηση ακολουθεί την κανονική κατανομή με μέσο μισθό 200.000 δρχ. και τυπική απόκλιση 30.000 δρχ. Να βρεθεί ο αριθμός των μισθωτών που παίρνει μισθό

i) μικρότερο ή ίσο των 180.000 δρχ.

ii) Ανάμεσα στις 170.000 δρχ. και 220.000 δρχ.

iii) Μεγαλύτερο ή ίσο των 210.000 δρχ.

θεωρώ την τυπική μεταβλητή Ζ με τιμή $z = \frac{x-\mu}{\sigma}$ και έχω:

$$z_1 = \frac{180.000 - 200.000}{30.000} = -\frac{20.000}{30.000} = -0{,}66.$$

Άρα η πιθανότητα να έχουμε μισθό ≤ 180.000 δραχμών είναι:

$$P(Z \le -0{,}66) = \Phi(-0{,}66) = 1 - \Phi(0{,}66) = 1 - 0{,}7454 = 0{,}2546.\,(25{,}46\%)$$

Οπότε ο αριθμός των μισθωτών που παίρνουν μισθό ≤ 180.000 θα είναι: $N \cdot P(Z \le z_1) = N \cdot \Phi(z_1) = 1000 \cdot 0{,}2546 \simeq 255$ μισθωτοί., όπου N =1000 εργαζόμενοι.

ii)Για να βρω τον αριθμό των μισθωτών με μισθό που κυμαίνεται μεταξύ 170.000 δρχ. και 220.000 δρχ. θα υπολογίσω:

$$z_1 = \frac{170.000 - 200.000}{30.000} = -1, \qquad z_2 = \frac{220.000 - 200.000}{30.000} = \frac{20.000}{30.000} = 0{,}66.$$

και $P(Z \le z_2) = \Phi(z_2) = \Phi(+0{,}66) = 0{,}7454.$

$P(Z \le z_1) = \Phi(z_1) = \Phi(-1) = 1 - \Phi(1) = 1 - 0{,}8413 = 0{,}1587.$

$\Phi(z_2) - \Phi(z_1) = 0{,}7454 - 0{,}1587 = 0{,}5867.$

Συνεπώς ο αριθμός των μισθωτών που ζητάμε ισούται με:

$$1000 \cdot 0{,}5867 \simeq 587 \text{ μισθωτούς.}$$

iii). Για να βρούμε τέλος τον αριθμό των μισθωτών με μισθούς ≥ 210.000 βρίσκουμε πρώτα την

$$z_1 = \frac{210.000 - 200.000}{30.000} = \frac{10.000}{30.000} \approx 0{,}3333.$$

$P(Z \le 0{,}3333) = \Phi(0{,}33) = 0{,}6293$

Άρα $P(Z \ge 0{,}3333) = 1 - P(Z \le 0{,}3333)1 - \Phi(0{,}33) = 1 - 0{,}6293 = 0{,}3707$

Οπότε ο αριθμός των μισθωτών είναι:$1000 \cdot 0{,}3707 \simeq 371$ μισθωτοί.

2. Σε μια φάρμα η ετήσια παραγωγή γάλακτος (ανά αγελάδα) είναι μια μεταβλητή που ακολουθεί κανονική κατανομή με μέση τιμή 5.000 kgr και τυπική απόκλιση 600 kgr.

i) Να βρεθεί ο αριθμός των αγελάδων στη φάρμα όταν 100 αγελάδες παράγουν γάλα μεταξύ 5.500 – 6000 kgr.

ii) Πόσες αγελάδες παράγουν γάλα ετησίως μέχρι και 5.800 kgr.

iii) Πόσες αγελάδες παράγουν γάλα ετησίως πάνω από 6.100 kgr και πάνω.

iv) Κάτω από ποια κιλά (kgr) βρίσκεται το 20% της ετήσιας παραγωγής.

v) Πάνω από ποια κιλά (kgr) βρίσκεται το 10% της ετήσια παραγωγής.

i) $$z_1 = \frac{5500-5000}{600} = \frac{500}{600} = 0{,}8333.$$

$$z_2 = \frac{6000-5000}{600} = \frac{1000}{600} = 1{,}6666.$$

$$P(5500 \le X \le 6000) = P(0{,}8333 \le Z \le 1{,}6666) = \Phi(1{,}6666) - \Phi(0{,}8333)$$
$$= 0{,}9515 - 0{,}7967 = 0{,}1548.$$

Οπότε αν θέσουμε N τον αριθμό των αγελάδων στη Φάρμα θα χουμε:

$N[\Phi(1{,}66) - \Phi(0{,}83)] = 100 \to N \cdot 0{,}1548 = 100 \Rightarrow N \approx 646$ αγελάδες.

Ποσοστό αγελάδων που παράγουν μεταξύ 5500-6000 kg.

$$\frac{100}{646} \approx 15{,}47\%.$$

ii) $P(X \le 5800) = P(Z \le 1{,}33) = \Phi(1{,}33) = 0{,}9082$

$$z = \frac{5800-5000}{600} = \frac{800}{600} = 1{,}33.$$

Και οι αγελάδες που παράγουν γάλα κατά μέσο όρο <u>μέχρι</u> και 5800 kgr είναι:

$N \cdot \Phi(1{,}33) = 646 \cdot 0{,}9082 = 587$ αγελάδες $\rightarrow$ ποσοστό $\frac{587}{646} \approx 90{,}8\%.$

iii) $P(X \geq 6100) = P(z \geq 1{,}83) = 1 - P(Z \leq 1{,}83) = 1 - \Phi(1{,}83) = 1 - 0{,}9664 = 0{,}0336.$

$$z = \frac{6100 - 5000}{600} = \frac{1100}{600} = 1{,}83.$$

Οι αγελάδες που παράγουν γάλα κατά μέσο όρο πάνω από 6100 kgr είναι:

$N \cdot 0{,}0336 = 646 \cdot 0{,}0336 \approx 22$ αγελάδες $\rightarrow$ ποσοστό : $\frac{22}{646} \approx 3{,}4\%.$

iv)Εδώ μας δίνεται ότι Φ(z) = 0,20. [Εφόσον το εμβαδόν του χωρίου κάτω της f(x) και τον άξονα x είναι ίσο με την μονάδα και το εμβαδόν από το $-\infty$ έως μιας τιμής της τυπικής μεταβλητής Z δίνεται από την Φ(z)].

Από την σχέση Φ(z)= 0,20 → από τους πίνακες έχουμε:

z = - 0,84: (μπορεί να υπολογισθεί από τον πίνακα με z>0 και από τον πίνακα με z<0).

Τώρα αν θέσω στη σχέση $z = \frac{x - \mu}{\sigma}$, z = -0,84 θα έχουμε:

$$-0{,}84 = \frac{x - 5000}{600} \Rightarrow x - 5000 = -504 \Rightarrow x = 5000 - 504 = 4.496\text{kgr}.$$

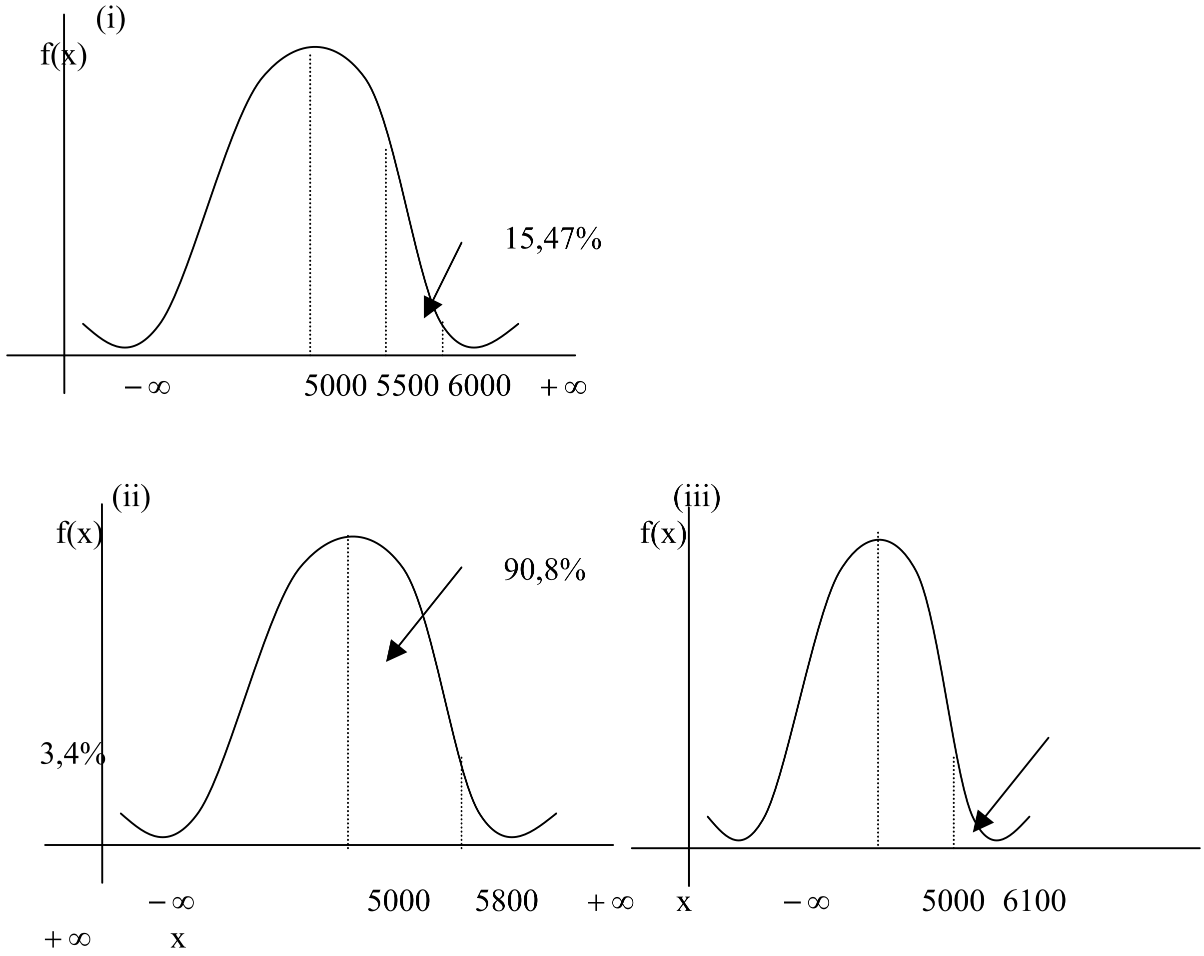
Άρα κάτω από τα 4.496Kgr βρίσκεται το 20% της ετήσιας παραγωγής

v. Γνωρίζουμε ότι το εμβαδόν από μια τιμή της z μέχρι το $+\infty$ είναι 10% = 0,10.

Δηλαδή: $1 - \Phi(z) = 0{,}10 \rightarrow \Phi(z) = 0{,}90 \rightarrow z = 1{,}28.$

Και $z = \frac{x - \mu}{\sigma} \Rightarrow 1{,}28 = \frac{x - 5000}{600} \Rightarrow x - 5000 = 768 \rightarrow x = 5.768\text{kgr}.$

vi) Άρα πάνω από τα 5.768Kgr βρίσκεται το 10% της ετήσια παραγωγής.



3. Ένα εργοστάσιο κατασκευάζει κυλινδρικά έμβολα με διάμετρο X η οποία ακολουθεί μια κανονική κατανομή με μέση τιμή 43,0 cm και τυπική απόκλιση 0,6 cm. Αν επιτρέπεται μια ανοχή στην διάμετρο από 42,4 cm μέχρι 43,6 cm να βρεθεί το ποσοστό των ελαττωματικών εμβόλων.

Κατ' αρχήν βρίσκουμε το ποσοστό των καλών εμβόλων:

$$P(42{,}4 \le X \le 43{,}6) = P(-1 \le Z \le 1) = \Phi(1) - \Phi(-1) = 0{,}8413 - 0{,}1587 = 0{,}6826.$$

$$z_1 = \frac{42{,}4 - 43}{0{,}6} = \frac{-0{,}6}{0{,}6} = -1, \qquad z_2 = \frac{43{,}6 - 43}{0{,}6} = 1$$

$\Phi(z_1) = \Phi(-1) = 0{,}1587, \qquad \Phi(1) = 0{,}8413$

Άρα το ποσοστό των ελαττωματικών είναι: 1-0,6826= 0,3174 $\Rightarrow$31,74%.

6.3.3. Προσέγγιση Διωνυμικής κατανομής από την τυπική κανονική κατανομή.

Γνωρίζουμε ότι η τυχαία μεταβλητή X που παρουσιάζει Διωνυμική κατανομή έχει συνάρτηση πιθανότητας.

$$\mathrm{P}[X = x] = \binom{v}{x} p^x \cdot q^{v-x} = \binom{v}{x} p^x (1-p)^{v-x}$$

με μέση τιμή μ = E(X) = ν p , διακύμανση

$\sigma^2 = Var(\mathrm{X}) = vpq = vp(1-p)$ και τυπική απόκλιση $\sigma = \sqrt{vp(1-p)}.$

Δημιουργούμε την τυπική μεταβλητή $Z = \dfrac{x-\mu}{\sigma}$, δηλαδή την

$$\mathrm{z} = \frac{\mathrm{x} - \mathrm{vp}}{\sqrt{\mathrm{vp}(1-\mathrm{p})}}$$

Έχει αποδειχθεί ότι αυτή ακολουθεί τυπική κανονική κατανομή, όταν $\mathrm{v} \to +\infty$ (στην πράξη όταν $\mathrm{v} \geq 30$ και $\mathrm{vp} \geq 5$, Θεώρημα Laplace Κ.Ο.Θ).

Έχουμε συνεπώς Z ~ N (0,1).

Έτσι αν θελήσουμε μπορούμε, όταν ισχύουν οι προϋποθέσεις που ετέθησαν ($v \geq 30 \quad vp \geq 5$ και $\mathrm{vq} \geq 5$, να υπολογίσουμε την πιθανότητα $\mathrm{P}[\mathrm{X} \leq \mathrm{x}]$ ή την πιθανότητα $P[\alpha \leq X \leq \beta]$ όχι μέσω διωνυμικής κατανομής, αλλά μέσω της τυπικής κανονικής κατανομής της τυπικής μεταβλητής Z.

Παίρνουμε δηλαδή:

$$P[X \leq x] = \Phi\left(\frac{x + 0{,}5 - \mu}{\sigma}\right) = \Phi\left(\frac{x + 0{,}5 - v\mathrm{p}}{\sqrt{vp(1-p)}}\right)$$

$$\text{και } P[\alpha \le X \le \beta] = \Phi\left(\frac{\beta + 0{,}5 - \nu p}{\sqrt{\nu p(1-p)}}\right) - \Phi\left(\frac{\alpha + 0{,}5 - \nu p}{\sqrt{\nu p(1-p)}}\right)$$

Παρατήρηση:

Επειδή η διωνυμική κατανομή είναι διακριτής μεταβλητής κατανομή έχει έννοια η πιθανότητα P[X=x] σε μια τιμή της X.

Η τυπική κανονική κατανομή όμως δεν μπορεί να δώσει την P[X=x], διότι

$$\mathrm{P}[\mathrm{X} = \mathrm{x}] = \int_{\mathrm{x}}^{\mathrm{x}} \mathrm{f}(\mathrm{u})\mathrm{du} = 0.$$

Έτσι όταν θέλουμε να αντικαταστήσουμε (να προσεγγίζουμε) την διωνυμική κατανομή σε τιμή x της X από την κανονική κατανομή, δηλαδή → P[X=x], παίρνουμε την σχέση:

$$\mathrm{P}[\mathrm{X} = \mathrm{x}] =$$

$$P(x - 0{,}5 < X < x + 0{,}5) = P\left(\frac{x - 0{,}5 - \nu \mathrm{p}}{\sqrt{\nu p(1-p)}} < z < \frac{x + 0{,}5 - \nu p}{\sqrt{\nu p(1-p)}}\right) =$$

$$= \phi\left(\frac{x + 0{,}5 - \nu p}{\sqrt{\nu p(1-p)}}\right) - \phi\left(\frac{x - 0{,}5 - \nu p)}{\sqrt{\nu p(1-p)}}\right).$$

Παραδείγματα

1. Να βρεθεί η πιθανότητα σε ένα δείγμα 30 γεωργικών προϊόντων να υπάρχουν 5 χαλασμένα αν η πιθανότητα ρ των χαλασμένων προϊόντων στο σύνολο είναι p = 0,20.

Αν χρησιμοποιήσουμε την διωνυμική κατανομή θα έχουμε την πιθανότητα:

$$\mathrm{P}[\mathrm{X} = 5] = \binom{30}{5} 0{,}20^{5} \cdot 0{,}80^{25} = 0{,}1624$$

Επειδή νp = 30·0,20 = 6>5 και νq = 30.0,8 = 24>5 μπορούμε να προσεγγίσουμε την πιθανότητα αυτήν μέσω κανονικής κατανομής.

Έχουμε $\sqrt{\nu p(1-p)} = \sqrt{6 \cdot 0{,}80} = 2{,}19$.

Παίρνουμε την πιθανότητα:

$$P(4{,}5 < X < 5{,}5) = \Phi\left(\frac{5{,}5-6}{2{,}19}\right) - \Phi\left(\frac{4{,}5-6}{2{,}19}\right) =$$
$$\Phi(-0{,}23) - \Phi(-0{,}68) = 1 - \Phi(0{,}23) - (1 - \Phi(0{,}68)) =$$
$$= -\Phi(0{,}23) + \Phi(0{,}68) = 0{,}7517 - 0{,}5910 = 0{,}1607$$

Παρατηρείται ένα **σφάλμα προσέγγισης** 0,1624-0,1607 = 0,0017.

2. Στο προηγούμενο παράδειγμα η πιθανότητα να είναι χαλασμένα είναι p = 0,80. Ζητάμε την πιθανότητα να έχουμε τουλάχιστον 28 γεωργικά προϊόντα χαλασμένα.

Σύμφωνα με την διωνυμική κατανομή έχουμε:

$$P[X \geq 28] = P[X = 28] + P[X = 29] + P[X = 30].$$

$$P[X = 28] = \binom{30}{28} 0{,}80^{28} \cdot 0{,}2^{2} = 435 \cdot 0{,}0019 \cdot 0{,}04 = 0{,}033$$

$$P[X = 29] = \binom{30}{29} 0{,}80^{29} \cdot 0{,}2^{1} = 30 \cdot 0{,}0015 \cdot 0{,}2 = 0{,}009$$

$$P[X = 30] = \binom{30}{30} 0{,}80^{30} \cdot 0{,}2^{0} = 0{,}001$$

$$P[X \geq 28] = 0{,}043$$

Αν χρησιμοποιήσουμε την Τυπική κανονική κατανομή για να προσεγγίσουμε την πιθανότητα αυτή θα έχουμε:

$$P[X \geq 28] = 1 - P[X \leq 28] = 1 - \Phi\left(\frac{28 + 0{,}5 - vp}{\sqrt{vp(1-p)}}\right) = 1 - \Phi\left(\frac{28{,}5 - 24}{2{,}19}\right) =$$

$$v \cdot p = 30 \cdot 0{,}8 = 24$$

$$\sqrt{vp(1-p)} = \sqrt{24 \cdot 0{,}2} = \sqrt{4{,}8} = 2{,}19$$

$$= 1 - \Phi(2{,}05) = 1 - 0{,}9798 = 0{,}0302.$$

Παρατηρείται **σφάλμα προσέγγισης** 0,043-0,0302=0,0128.

3. Η πιθανότητα να είναι χαλασμένα γεωργικά προϊόντα σε μια παραγωγή γεωργικών προϊόντων είναι p = 0,50. Παίρνουμε τυχαία 30 προϊόντα ποια ή πιθανότητα να έχουμε το πολύ επτά χαλασμένο προϊόντα.

Σύμφωνα με την διωνυμική κατανομή έχουμε:

$P[X \leq 7] = P[X = 0] + P[X = 1] +$ P[X=2]+…+ P[X=7].

$$P[X = 0] = \binom{30}{0} 0{,}50^0 \cdot 0{,}50^{30}$$

$$P[X = 1] = \binom{30}{1} 0{,}50^1 \cdot 0{,}50^{29}$$

………………..

$P[X \leq 7]$= …… υπολογίζεται

Αν προσεγγίσουμε την πιθανότητα αυτή με τυπική κανονική κατανομή:

v p= 30·0,5 = 15 > 5

$\mu = vp = 15, \qquad \sqrt{vp(1-p)} = \sqrt{15.0{,}50} = 2{,}74$

$$P[X \leq 7] = \Phi\left(\frac{7 - 15}{2{,}79}\right) = \Phi(-2{,}92) = 1 - \Phi(2{,}92) = 0{,}0018$$

6.3.4. Εκθετική Κατανομή

Η τυχαία μεταβλητή Χ έχει εκθετική κατανομή όταν εφαρμόζεται ένα πείραμα Bernoulli και ο **Χ μετρά το χρόνο που μεσολαβεί από ένα αρχικό χρονικό σημείο**

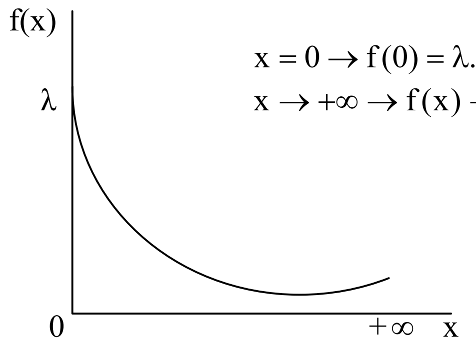
μέχρις ότου πραγματοποιηθεί ένα συμβάν (φαινόμενο). Ο αριθμός των συμβάντων σε ένα χρονικό διάστημα ακολουθεί κατανομή Poisson.

Η συνάρτηση πυκνότητας πιθανότητας της εκθετικής κατανομής είναι:

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & x \geq 0 \\ o, & x < 0 \end{cases}$$

Η παράμετρος λ>0 εκφράζει το **μέσο αριθμό πραγματοποίησης του ενδεχομένου (φαινομένου).**

Αποδεικνύει ότι $\mu = E(X) = \dfrac{1}{\lambda}$ και $\sigma^2 = Var(X) = \dfrac{1}{\lambda^2}$.



Η εκθετική κατανομή μπορεί να χρησιμοποιηθεί όταν **η τυχαία μεταβλητή εκφράζει χρόνο εξυπηρέτησης, χρόνο διάρκειας ενός τηλεφωνήματος, χρόνο μεταξύ δύο αφίξεων αεροπλάνων σ' ένα αεροδρόμιο, διάρκεια συνεχούς λειτουργίας ενός λαμπτήρα μέχρι να καεί. Η εκθετική κατανομή χρησιμοποιείται σε συνδυασμό με την Poisson.**

Παράδειγμα

Ο αριθμός των παραγωγών που φθάνουν στο εργοστάσιο Σακχάρεως με την παραγωγή τους ακολουθεί την κατανομή Poisson.

Αν κατά μέσο όρο χρειάζονται 20 λεπτά για να εξυπηρετηθεί ένας παραγωγός, ποια η πιθανότητα ένας παραγωγός να εξυπηρετηθεί σε λιγότερο από 15 λεπτά;

Έχουμε:

$$P(X<15)=\int_0^{15}\lambda e^{-\lambda x}dx \qquad \text{όπου}: \lambda=\frac{1}{20}$$

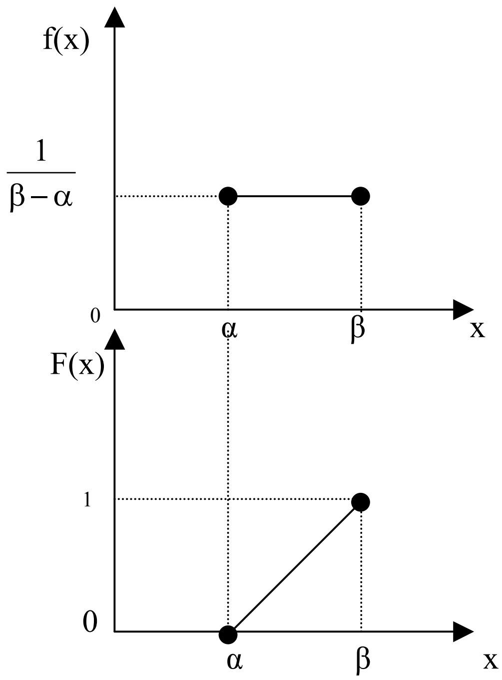
$$P(X<15)=\frac{1}{20}\int_0^{15} e^{-\frac{1}{20}x}dx$$
$$=\frac{1}{20}(-20)\int_0^{15} de^{-\frac{1}{20}x}=-e^{-\frac{x}{20}}\Big|_0^{15}=-e^{-\frac{15}{20}}-(-e^0)$$
$$=-e^{-\frac{3}{4}}+1=1-\frac{1}{e^{\frac{3}{4}}}=1-\frac{1}{2{,}18}=0{,}54.$$

6.3.5. Ορθογώνια ή Ομοιόμορφη Κατανομή.

Η τυχαία μεταβλητή X λέμε ότι έχει **ορθογώνια κατανομή ή ομοιόμορφη κατανομή στο διάστημα [α, β]** εάν η συνάρτηση πυκνότητας είναι:

$$f(x)=\begin{cases}\dfrac{1}{\beta-\alpha} & \chi\varepsilon[\alpha,\beta]\\[2ex] 0 & \chi\notin[\alpha,\beta]\end{cases}$$

$$\text{Έχουμε}: \ E(X)=\frac{1}{2}(\alpha+\beta), \qquad Var(X)=\frac{1}{12}(\beta-\alpha)^2.$$



$$F(X) = \frac{x-\alpha}{\beta-\alpha} \quad \text{για} \quad \alpha \leq x \leq \beta \quad \text{και} \quad F(X) = 1 \quad \text{για} \quad x \geq \beta$$

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Αποστολόπουλος Θ. : «Στατιστική Επιχειρήσεων», Αθήνα 1991.
2. Αποστολόπουλος Θ. : «Περιγραφική Στατιστική», Αθήνα 1996
3. Δρακάτος Κ. : «Στατιστική», εκδόσεις Παπαζήση, Αθήνα 1984.
4. Δρακάτος Κ. : «Περιγραφική Οικονομική Στατιστική», εκδόσεις Παπαζήση, Αθήνα 1993.
5. Δρακάτος Κ. «Ασκήσεις Στατιστικής», εκδόσεις Παπαζήση, Αθήνα 1984.
6. Ζαχαροπούλου Χ. : «Στατιστική, μέθοδοι-εφαρμογές», Θεσ/νίκη 1996.
7. Λαμπράκη Δ. : «Στατιστική», Αθήνα 1980.
8. Μπένος Β.Κ. : «Στατιστική – Περιγραφική Στατιστική», εκδόσεις Σταμούλη, Πειραιάς 1973.
9. Μπένος Β.Κ. : «Οικονομική Στατιστική», εκδόσεις Σταμούλη, Πειραιάς 1973.
10. Μπένος Β.Κ. : «Στατιστική Επιχειρήσεων», εκδόσεις Σταμούλη, Πειραιάς 1996.
11. Μπόρα Ε. – Κολύβα Φ. : «Στατιστική – Θεωρία – Εφαρμογές», εκδόσεις Ζήτη, Θεσσαλονίκη 1999.
12. Μπόρα Ε. – Μωυσιάδης Χ. : «Εφαρμοσμένη Στατιστική», εκδόσεις Ζήτη, Θεσσαλονίκη 1997.
13. Μάνος Β. : «Γεωργική Οικονομική Στατιστική», εκδόσεις Ζήτη, Θεσ/νίκη 1997.
14. Παναγιωτακόπουλος Δ. : «Εφαρμογές πιθανοτήτων & Στατιστική», Ξάνθη 1997.
15. Παπαδημητρίου Γ. : «Στατιστική» περιγραφική Στατιστική Τεύχος 1, εκδόσεις Παρατηρητής, Θεσ/νίκη 1986.
16. Περσίδης Σ. : «Πιθανότητες και Στατιστική» Schaurns Outline Series. Mc Graun-Hill, N. York. ΕΣΠΙ, Αθήνα 1977.
17. Τζωρτζόπουλος Θ. : «Αριθμοδείκτες» , Αθήνα 1978.
18. Τζωρτζόπουλος Θ. : «Ανάλυση χρονολογικών σειρών» , Αθήνα 1979.
19. GRAIS B. : “Methodes Statistiques” Paris 1978.
20. WAYNE R. OTT: “Environmental Statistics and Data Analysis”.
21. P. WHITEHEAD-G. WHITEHEAD: “Statistics for Business” 1992 PITMAN PUBLISHING LONDON.