ΚΩΝ/ΝΟΣ Α. ΖΩΓΡΑΦΟΣ
Επίκουρος Καθηγητής, Τμήματος Μαθηματικών
Πανεπιστημίου Ιωαννίνων

ΜΑΘΗΜΑΤΑ ΠΙΘΑΝΟΤΗΤΩΝ ΚΑΙ ΣΤΑΤΙΣΤΙΚΗΣ

Ιωάννινα 1994

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΒΙΒΛΙΟΘΗΚΗ
ΠΑΝΕΠΙΣΤΗΜΙΟΥ ΙΩΑΝΝΙΝΩΝ
026000136275
519.2
2ΟΓ
ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΚΩΝ/ΝΟΣ Α. ΖΩΓΡΑΦΟΣ
Επίκουρος Καθηγητής, Τμήματος Μαθηματικών
Πανεπιστημίου Ιωαννίνων

Εισαγωγή στην υπολογιστική στατιστική

ΜΑΘΗΜΑΤΑ ΠΙΘΑΝΟΤΗΤΩΝ ΚΑΙ ΣΤΑΤΙΣΤΙΚΗΣ

Ιωάννινα 1994

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Π Ρ Ο Λ Ο Γ Ο Σ

Οι Πιθανότητες και η Στατιστική αποτελούν ένα κλάδο της Μαθηματικής Επιστήμης που ασχολείται με την περιγραφή και τη μελέτη τυχαίων φαινομένων. Οι αρχές και οι μέθοδοι των Πιθανοτήτων και της Στατιστικής αποτελούν απαραίτητο εργαλείο για κάθε ερευνητή, γεγονός που αποδεικνύεται από τον αυξανόμενο ρόλο τους σε όλες τις εκφράσεις της επιστήμης, της τεχνολογίας και της κοινωνίας.

Το βιβλίο αυτό γράφτηκε για να βοηθήσει τους φοιτητές του Παιδαγωγικού Τμήματος Νηπιαγωγών του Πανεπιστημίου Ιωαννίνων στα πλαίσια του Μαθήματος "Στατιστική Ι" του Β' εξαμήνου σπουδών τους. Επιχειρεί μια εισαγωγή στα πιο βασικά θέματα των πιθανοτήτων και της στατιστικής, χωρίς να απαιτείται προηγούμενη γνώση των θεμάτων αυτών και εξειδικευμένη μαθηματική γνώση. Έτσι πρωταρχικό ρόλο στην εισαγωγή των εννοιών και την ανάπτυξη της μεθοδολογίας παίζει η διαίσθηση. Η μαθηματική αυστηρότητα διατηρείται όπου αυτό είναι δυνατόν.

Στο πλαίσιο αυτό, μετά από ένα εισαγωγικό κεφάλαιο στο οποίο εισάγονται οι έννοιες του πληθυσμού, του δείγματος και της μεταβλητής, στα Κεφάλαια 2 και 3 αναπτύσσονται οι αρχές και οι μέθοδοι που επιτελούν στην οργάνωση και παρουσίαση των δεδομένων και δίνουν περιεχόμενο στον όρο περιγραφική στατιστική.

Στα Κεφάλαια 4 και 5 αναπτύσσονται οι πιο βασικές έννοιες της θεωρίας πιθανοτήτων. Η πιθανότητα προσεγγίζεται ποσοτικά μέσω του κλασσικού και εμπειρικού της ορισμού και ακολουθεί η αξιωματική της

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

θεμελίωση και η εξέταση των βασικών της ιδιοτήτων. Οι έννοιες της δεσμευμένης πιθανότητας και της ανεξαρτησίας ολοκληρώνουν το κεφάλαιο αυτό. Στο Κεφάλαιο 5 παρουσιάζονται οι θεμελιώδεις έννοιες της τυχαίας μεταβλητής και της συνάρτησης κατανομής. Εισάγονται επίσης η διωνυμική κατανομή, η κατανομή Poisson και η κανονική κατανομή που αποτελούν πιθανοθεωρητικά μοντέλα στη βάση των οποίων μπορεί να προσαρμοστεί και επιλυθεί μεγάλη ποικιλία προβλημάτων της καθημερινότητας.

Τα Κεφάλαια 6 και 7 αποτελούν μια εισαγωγή στη Στατιστική Συμπερασματολογία. Ειδικότερα στο Κεφάλαιο 6 παρουσιάζονται οι αρχές και οι μέθοδοι για την εκτίμηση των παραμέτρων ενός πιθανοθεωρητικού μοντέλου ενώ στο Κεφάλαιο 7 αναπτύσσεται η φιλοσοφία και η μεθοδολογία για τον έλεγχο στατιστικών υποθέσεων.

Θεωρώ υποχρέωσή μου να ευχαριστήσω την κ. Κατερίνα Χρηστίδη για την προσεκτική δακτυλογράφηση του κειμένου και το Τυπογραφείο του Πανεπιστημίου Ιωαννίνων για την επιμέλεια της έκδοσης αυτής.

Ιωάννινα Νοέμβριος 1993 Κων/νος Α. Ζωγράφος

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1 : ΕΙΣΑΓΩΓΗ

1.1 Εισαγωγικά

1.2 Πληθυσμός, Δείγμα-Μεταβλητές

ΚΕΦΑΛΑΙΟ 2 : ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

2.1 Εισαγωγικά

2.2 Πίνακες Συχνοτήτων

2.3 Γραφικές Παραστάσεις

ΚΕΦΑΛΑΙΟ 3 : ΑΡΙΘΜΗΤΙΚΑ ΜΕΓΕΘΗ

3.1 Εισαγωγικά

3.2 Μέτρα Θέσεως

3.3 Μέτρα Διασποράς

Ασκήσεις 2ου και 3ου Κεφαλαίου

ΚΕΦΑΛΑΙΟ 4 : ΣΤΟΙΧΕΙΑ ΑΠΟ ΤΗ ΘΕΩΡΙΑ ΠΙΘΑΝΟΤΗΤΩΝ

4.1 Εισαγωγικά

4.2 Κλασική, εμπειρική πιθανότητα

4.3 Εμπειρικός ορισμός πιθανότητας

4.4 Τεχνάσματα Δειγματοληψίας

Ενδεχόμενα-Πράξεις μεταξύ τους

4.5 Αξιωματικός ορισμός πιθανότητας

4.6 Δεσμευμένη ή υπό συνθήκη πιθανότητα

4.7 Ανεξάρτητα Ενδεχόμενα

Ασκήσεις 4ου Κεφαλαίου

Στην κόρη μου

Ελένη

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1 : ΕΙΣΑΓΩΓΗ

1.1	Εισαγωγικά	1
1.2	Πληθυσμός-Δείγμα-Μεταβλητή	5

ΚΕΦΑΛΑΙΟ 2 : ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

2.1	Εισαγωγικά	17
2.2	Πίνακες Συχνοτήτων	18
2.3	Γραφικές Παραστάσεις	31

ΚΕΦΑΛΑΙΟ 3 : ΑΡΙΘΜΗΤΙΚΑ ΜΕΓΕΘΗ

3.1	Εισαγωγικά	45
3.2	Μέτρα Θέσεως	46
3.3	Μέτρα Διασποράς	53
Ασκήσεις 2ου και 3ου Κεφαλαίου		72

ΚΕΦΑΛΑΙΟ 4 : ΣΤΟΙΧΕΙΑ ΑΠΟ ΤΗ ΘΕΩΡΙΑ ΠΙΘΑΝΟΤΗΤΩΝ

4.1	Εισαγωγικά	75
4.2	Κλασσικός ορισμός πιθανότητας	77
4.3	Εμπειρικός ορισμός πιθανότητας	81
4.4	Τυχαίο Πείραμα-Δειγματικός χώρος -Ενδεχόμενο-Πράξεις ενδεχομένων	84
4.5	Αξιωματικός ορισμός πιθανότητας	89
4.6	Δεσμευμένη ή υπό συνθήκη πιθανότητα	91
4.7	Ανεξαρτησία Ενδεχομένων	95
Ασκήσεις 4ου Κεφαλαίου		98

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΚΕΦΑΛΑΙΟ 5 : ΤΥΧΑΙΑ ΜΕΤΑΒΛΗΤΗ-
-ΚΑΤΑΝΟΜΗ ΠΙΘΑΝΟΤΗΤΑΣ
-ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΥΧΑΙΑΣ
ΜΕΤΑΒΛΗΤΗΣ & ΚΑΤΑΝΟΜΗΣ

5.1	Εισαγωγικά	101
5.2	Τυχαία Μεταβλητή	102
5.3	Διακριτές τυχαίες μεταβλητές -Συνάρτηση Πιθανότητας	106
5.3.1	Διωνυμικό Πείραμα-Διωνυμική Κατανομή	110
5.3.2	Κατανομή Poisson	115
5.4	Συνεχείς Τυχαίες Μεταβλητές -Συνάρτηση Πυκν. Πιθανότητας	119
5.4.1	Κανονική Καταν. ή Κατανομή Gauss	122
5.4.2	Τυπική Κανονική Κατανομή	125
5.5	Χαρακτηριστικά Τυχαίων Μεταβλητών και Κατανομών: Μέση Τιμή και Διακύμανση	131
5.6	Ανακεφαλαίωση	137
Ασκήσεις 5ου Κεφαλαίου		141

ΚΕΦΑΛΑΙΟ 6 : ΣΤΑΤΙΣΤΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ
ΕΚΤΙΜΗΣΗ ΠΑΡΑΜΕΤΡΩΝ

6.1	Εισαγωγικά	143
6.2	Εκτίμηση σε σημείο	146
6.3	Εκτίμηση σε διάστημα-Διαστήματα Εμπιστοσύνης	150
6.3.1	Διάστημα Εμπιστοσύνης για τη μέση τιμή κανονικού πληθυσμού	151
6.3.2	Διάστημα Εμπιστοσύνης για τη διαφορά των μέσων τιμών δύο ανεξάρτητων κανο-	

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

νικών πληθυσμών 162
6.3.3 Διάστημα Εμπιστοσύνης για τη μέση τιμή οποιουδήποτε πληθυσμού 170
6.3.4 Γενικές Παρατηρήσεις 171

ΚΕΦΑΛΑΙΟ 7 : ΕΛΕΓΧΟΣ ΣΤΑΤΙΣΤΙΚΩΝ ΥΠΟΘΕΣΕΩΝ

7.1 Εισαγωγικά 173
7.2 Το στατιστικό τεστ - Οι πρώτες έννοιες 174
7.3 Αξιολόγηση στατιστικών τεστ - Σφάλματα τύπου I και II 179
7.4 Στατιστικά τεστ για τη μέση τιμή κανονικής κατανομής 182
7.5 Στατιστικά τεστ για τη διαφορά των μέσων τιμών δύο ανεξάρτητων κανονικών πληθυσμών 189
7.6 Τέστ συγκρίσεως ζευγών 197
7.7 Γενικές Παρατηρήσεις 202
Ασκήσεις 6ου και 7ου Κεφαλαίου 204

ΠΑΡΑΡΤΗΜΑ : Στατιστικοί πίνακες 207
ΑΠΑΝΤΗΣΕΙΣ ΣΤΙΣ ΑΣΚΗΣΕΙΣ 219
ΒΙΒΛΙΟΓΡΑΦΙΑ 223

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

1.1. Εισαγωγικά

Σε καθημερινή βάση, στις σελίδες του ημερήσιου και περιοδικού τύπου, σε κυβερνητικά έντυπα, σε έντυπα και αφίσες σχετικές με την αναπτυξιακή πορεία επιχειρήσεων και οργανισμών, σε νοσοκομεία, στην τηλεοπτική μετάδοση εκλογικών αποτελεσμάτων, βρισκόμαστε αντιμέτωποι με πλήθος εννοιών και στοιχείων, με καλοφτιαγμένες γραφικές παραστάσεις, με πίνακες αριθμητικών δεδομένων, με χρωματιστούς χάρτες κλπ. Τα στοιχεία και οι παραστάσεις αυτές, βασισμένες στο γεγονός ότι μια εικόνα "μιλάει" ίσως καλύτερα από χίλιες λέξεις, σκοπό έχουν να περιγράψουν, για παράδειγμα, την αύξηση ή μείωση της παραγωγής μιας βιομηχανίας την καλοκαιρινή περίοδο σε σχέση με την αντίστοιχη περσινή, το πλήθος των τροχαίων ατυχημάτων σε σχέση με τις χώρες της ΕΟΚ, να μας πείσουν για την ανωτερότητα ενός προϊόντος ή μιας μεθόδου, για την επιτυχία της κυβερνητικής πολιτικής σε ένα θέμα. Σε εθνικό ή και διεθνές επίπεδο το κάθε κράτος ή διεθνής οργανισμός, μέσω στατιστικών υπηρεσιών απ' τις οποίες στελεχώνεται, διενεργεί σε τακτά χρονικά διαστήματα δειγματοληπτικές έρευνες για οικονομικά και κοινωνικά προβλήματα όπως ο πληθωρισμός, η ανεργία των νέων, η εξάπλωση των ναρκωτικών και του AIDS κλπ. Οι μοντέρνες στατιστικές μεθοδολογίες βοηθούν στη μελέτη των παραπάνω

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

προβλημάτων, και γενικότερα αποτελούν χρήσιμα εργαλεία για τη μελέτη και διερεύνηση κάθε επιστημονικής, τεχνολογικής και κοινωνικής δραστηριότητας. Το γεγονός αυτό τεκμηριώνεται από τον εκθετικά αυξανόμενο ρόλο της στατιστικής σε ευρύ φάσμα επιστημών όπως η βιολογία, δημογραφία, οικονομία, εκπαίδευση, υγεία, ψυχολογία και ψυχομετρία, κοινωνικές επιστήμες.

Η στατιστική είναι μια επιστήμη με επαγωγικό χαρακτήρα και μέθοδο, μια επιστήμη που επεκτείνει τα συμπεράσματά της στο όλο (πληθυσμό) από τη μελέτη και ανάλυση του μέρους (δείγμα). Στις δραστηριότητές της εντάσσεται:

i) Η διατύπωση και ο σχεδιασμός του προς επίλυση προβλήματος.

ii) Η συλλογή, οργάνωση και παρουσίαση των δεδομένων που προκύπτουν είτε από διαθέσιμες πηγές είτε από την εκτέλεση πειραμάτων.

iii) Η ανάλυση των δεδομένων.

iv) Η ερμηνεία των αποτελεσμάτων της ανάλυσης, η εξαγωγή συμπερασμάτων και η λήψη αποφάσεων.

v) Η ανάπτυξη νέων στατιστικών μεθόδων ή η βελτίωση των ήδη υπαρχόντων.

Οι κατευθύνσεις αυτές δίνουν "σάρκα και οστά" στον όρο στατιστική. Οι τέσσερις πρώτες αφορούν τη χρήση αρχών και μεθόδων που έχουν αναπτυχθεί για την επεξεργασία δεδομένων, την εξαγωγή συμπερασμάτων και τη λήψη αποφάσεων. Η πέμπτη δραστηριότητα αφορά την ανάπτυξη της ίδιας της στατιστικής που συνίσταται στη δημιουργία στατιστικών μεθοδολογιών για την προσέγγιση και ερμηνεία σύνθετων πρακτικών προβλημάτων.

Οι αρχές και οι μέθοδοι που επιτελούν στην συλλογή, οργάνωση και παρουσίαση των δεδομένων, δίνουν περιεχόμενο στον όρο **περιγραφική στατιστική** (descriptive statistics) . Η ανάλυση των

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

δεδομένων, η ερμηνεία της ανάλυσης και η εξαγωγή συμπερασμάτων για το όλο από τη μελέτη του μέρους ανήκει στο χώρο της **στατιστικής συμπερασματολογίας** (statistical inference) ή **επαγωγικής στατιστικής**. Επισημαίνεται ότι οποιαδήποτε σύγχρονη ερευνητική ή λειτουργική δραστηριότητα στο χώρο της στατιστικής αποτελεί αντικείμενο της στατιστικής συμπερασματολογίας.

Το θεωρητικό υπόβαθρο της στατιστικής αποτελεί η θεωρία πιθανοτήτων, μια μαθηματικά δομημένη θεωρία με σπουδαίες εφαρμογές σε διαφορετικές περιοχές της ανθρώπινης δραστηριότητας. Η θεωρία πιθανοτήτων και η στατιστική εισάγουν μια επανάσταση στο χώρο της ανθρώπινης σκέψης που έγκειται στην αμφισβήτηση της ντετερμινιστικής αντίληψης της φύσης. Έτσι η φύση που τον $18^{ο}$ αιώνα θεωρούνταν αυστηρά προκαθορισμένη και οχεδιασμένη σύμφωνα με απαράλλακτους μαθηματικούς νόμους, παρουσιάστηκε σαν χαώδης, άτακτη και απρόβλεπτη. Αυτή η άποψη απέναντι στη φύση και τους νόμους που τη διέπουν είναι γνωστή σαν στατιστική αντίληψη της φύσης.

Σε κάθε επιστημονική μελέτη ενός φαινομένου επιδίωξη αποτελεί η δημιουργία ενός μαθηματικού μοντέλου, μιας μαθηματικής κατασκευής που δίνει τη δυνατότητα περιγραφής, πρόβλεψης των χαρακτηριστικών του φαινομένου. Τα μαθηματικά μοντέλα για την προσέγγιση και ερμηνεία της φύσης διακρίνονται σε ντετερμινιστικά (deterministic), και σε πιθανοθεωρητικά (probabilistic) ή στοχαστικά (stochastic) ανάλογα με τη φύση του φαινομένου και την αντίληψή μας, αναφορικά με την προσέγγιση και ερμηνεία της φύσης και των φαινομένων της. Ένα πιθανοθεωρητικό λοιπόν μοντέλο είναι μια μαθηματική κατασκευή που σχεδιάζεται για να περιγράψει φαινόμενα που διέπονται από τον παράγοντα τύχη. Το αντικείμενο της στατιστικής είναι η διάγνωση της ορθότητας του πιθανοθεωρητικού μοντέλου στη βάση δεδομένων που συλλέγονται μετά από πειράματα πάνω στην κατάσταση που περιγράφει το μοντέλο. Έτσι η

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

θεωρία πιθανοτήτων και η στατιστική είναι κλάδοι των μαθηματικών που γενικά καταπιάνονται με τους νόμους τυχαίων φαινομένων.

Η θεωρία πιθανοτήτων οφείλει μάλλον τη γέννησή της στην ενασχόληση του ανθρώπου με τα τυχερά παιχνίδια. Από τα μέσα του $17^{ου}$ αιώνα πολλοί διάσημοι μαθηματικοί όπως οι Fermat, Pascal, Bernoulli, De Moivre, ασχολήθηκαν με τη λύση προβλημάτων που αναφέρονται στα τυχερά παιγνίδια. Η γρήγορη εξέλιξη των φυσικών επιστημών εκείνη την εποχή έθετε επιτακτικά την παραπέρα εξέλιξη της θεωρίας πιθανοτήτων με τη δημιουργία πιο εξειδικευμένων αναλυτικών εργαλείων - μοντέλων. Ιδιαίτερα σημαντικό ρόλο στην ανάπτυξη αναλυτικών μεθόδων έπαιξαν οι De Moivre , Laplace, Gauss και Poisson. Ως κλάδος των μαθηματικών καθιερώνεται το 1812 με την εργασία του Laplace: "*Theorie Analytique de Probabilite*". Στη συνέχεια αρχίζει μια προσπάθεια αξιωματικής θεμελίωσης της θεωρίας πιθανοτήτων με τους Bertrand, Poincare κ.ά. η οποία ολοκληρώνεται από το Σοβιετικό μαθηματικό Kolmogorov περί το 1930. Σήμερα η θεωρία πιθανοτήτων έχει εκτείνει την επιρροή και τις πρακτικές της εφαρμογές σε ευρύ φάσμα επιστημών. Ερευνητές, σ' όλο τον κόσμο, πλουτίζουν τη θεωρία με σπουδαία αποτελέσματα.

Παράλληλα με τη θεωρία πιθανοτήτων, αλλά ανεξάρτητα, αναπτύχθηκε η στατιστική. Σκοπός της στατιστικής από παλιά ήταν η περιγραφή των πιο σημαντικών πράξεων του κράτους (γεννήσεις, θάνατοι, φόροι κλπ.). Στατιστικά δεδομένα που αναφέρονται σ' αυτή τη δραστηριότητα έχουν βρεθεί, για παράδειγμα, στην Ιουδαία και αφορούν απογραφή του 2030 π.Χ. Η στατιστική αρχίζει να συνδέεται με τα μαθηματικά τον $17^{ο}$ αιώνα. Σημαντικό ρόλο σ' αυτό έπαιξαν οι Fermat και Pascal. Η ανάπτυξη της θεωρίας πιθανοτήτων, συντέλεσε στην αποδέσμευση της στατιστικής από τον περιγραφικό της ρόλο και θεμελίωσε τη σημερινή της φυσιογνωμία, το συμπερασματολογικό της ρόλο. Παρά το γεγονός ότι προσπάθειες ανάδειξης του σημερινού

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

χαρακτήρα της στατιστικής άρχισαν το 18^{ο} αιώνα, ο χαρακτήρας αυτός δημιουργείται στις αρχές του αιώνα μας με τις πρωτοποριακές εργασίες των K. Pearson και A. Fisher. Τα ονόματα επίσης των W. Gosset, J. Neyman, A. Wald και οι εργασίες τους στο σχεδιαμό πειραμάτων, τον έλεγχο υποθέσεων, τη θεωρία αποφάσεων κλπ. κινούν τη στατιστική σε νέους επαναστατικούς δρόμους σκέψης και σηματοδοτούν την τεράστια σημερινή της ανάπτυξη τόσο στο επίπεδο της θεωρίας όσο και στο επίπεδο των εφαρμογών.

1.2 Πληθυσμός - Δείγμα - Μεταβλητή

Οι έννοιες του πληθυσμού και του δείγματος είναι κεντρικής σημασίας στη στατιστική. Με την καθημερινή χρήση ο όρος πληθυσμός χρησιμοποιείται για να αναφερθούμε σε ομάδες ή σύνολα ανθρώπων. Αναφερόμαστε για παράδειγμα στον πληθυσμό της Ελλάδας ή της Ηπείρου ή της πόλης των Ιωαννίνων και εννοούμε το σύνολο των ανθρώπων που κατέχουν μια ορισμένη γεωγραφική περιοχή. Η θεώρηση αυτή του όρου πληθυσμός είναι περιορισμένη για τη στατιστική. Ο στατιστικός τον εμπλουτίζει ώστε να περιλαμβάνει την ευρύτερη δυνατή συλλογή οντοτήτων ή συμβάντων. Έτσι ένας πληθυσμός μπορεί να είναι ένα σύνολο από αντικείμενα, έντομα, μετρήσεις, ύλη, μικροοργανισμούς, βιομηχανική παραγωγή, τροχαία ατυχήματα κλπ. Επιγραμματικά, **πληθυσμός** (population) είναι το σύνολο των οντοτήτων ή συμβάντων των οποίων ένα ή περισσότερα χαρακτηριστικά ενδιαφερόμαστε να μελετήσουμε.

Ένας πληθυσμός μπορεί να είναι **πεπερασμένος ή άπειρος**. Οι μαθητές της Α' τάξης των Δημοτικών Σχολείων των Ιωαννίνων, οι τόμοι των λογοτεχνικών βιβλίων της Δημοτικής βιβλιοθήκης, τα τροχαία ατυχήματα στο Νομό Ιωαννίνων το 1989 αποτελούν παραδείγματα

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

πεπερασμένων πληθυσμών. Κοινό τους χαρακτηριστικό είναι το γεγονός ότι το μέγεθός τους, το πλήθος των στοιχείων τους είναι πεπερασμένο.

Ένας άπειρος πληθυσμός απαρτίζεται από ένα ατέλειωτο πλήθος στοιχείων, όπως για παράδειγμα ο απεριόριστος αριθμός ρίψεων ενός νομίσματος ή οποιαδήποτε πειραματική διαδικασία αφού θεωρητικά τουλάχιστον μπορεί να επαναλαμβάνεται αδιάκοπα. Σε πολλές περιπτώσεις πληθυσμοί τους οποίους ο στατιστικός προτίθεται να μελετήσει είναι πεπερασμένοι, αλλά τόσο μεγάλοι ώστε για πρακτικούς σκοπούς μπορούν να θεωρηθούν ως άπειροι. Καταφεύγουμε στην υπόθεση του άπειρου πληθυσμού, οποτεδήποτε το μέγεθος του πεπερασμένου πληθυσμού είναι αρκετά μεγάλο λόγω της αξιοπιστίας των στατιστικών μοντέλων όταν αυτά εφαρμόζονται σε άπειρους πληθυσμούς.

Σε πολλά πρακτικά προβλήματα το μέγεθος του υπό εξέταση πληθυσμού είναι τόσο μεγάλο, ώστε είναι δύσκολο ή και αδύνατο να μελετηθεί. Σε μια δημοσκόπιση για την κατάργηση της θανατικής ποινής στην Ελλάδα και αντιοικονομικό αλλά και πιθανόν αδύνατο θα ήταν να ρωτηθούν όλοι οι Έλληνες. Έτσι εκλέγουμε ένα μέρος του πληθυσμού και από την μελέτη και ανάλυσή του με κατάλληλες στατιστικές μεθόδους καταλήγουμε σε συμπεράσματα τα οποία επεκτείνουμε στον πληθυσμό. Το μέρος ή το υποσύνολο του υπό μελέτη πληθυσμού ονομάζεται **δείγμα** (sample).

Τα συμπεράσματα που προκύπτουν από την ανάλυση του δείγματος είναι αξιόπιστα όταν το δείγμα είναι αντιπροσωπευτικό του πληθυσμού. Στο παράδειγμά μας, αν ως δείγμα θεωρηθούν μόνο οι συγγενείς θυμάτων εγκληματικών ενεργειών, είναι φανερή η αναξιοπιστία των συμπερασμάτων. Ένα στοιχείο της αντιπροσωπευτικότητας του δείγματος είναι η τυχαία εκλογή του από τον πληθυσμό.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Τυχαίο δείγμα (random sample) είναι το δείγμα που εκλέγεται κατά τέτοιο τρόπο ώστε όλα τα μέλη του πληθυσμού έχουν την ίδια δυνατότητα να συμπεριληφθούν σ' αυτό.

Ανάλογη σπουδαιότητα με τις έννοιες του πληθυσμού και του δείγματος έχει για τη στατιστική και γενικότερα για τα μαθηματικά, η έννοια της μεταβλητής. Ο όρος **μεταβλητή** (variable) χρησιμοποιείται για να περιγράψει μια ιδιότητα ή ένα χαρακτηριστικό γνώρισμα, των μελών ενός πληθυσμού, το οποίο διαφοροποιείται ή μεταβάλεται από μέλος σε μέλος του πληθυσμού.

Παραδείγματα μεταβλητών είναι το ύψος, το βάρος, ο δείκτης νοημοσύνης, το χρώμα των ματιών, το επάγγελμα, η κοινωνικοοικονομική κατάσταση, το φύλο, η εθνικότητα, το θρήσκευμα. Τα προηγούμενα είναι μεγέθη ή χαρακτηριστικά ενός πληθυσμού τα οποία μεταβάλονται από μέλος σε μέλος του. Χαρακτηρισμοί ή αριθμητικές τιμές μπορούν να χρησιμοποιηθούν για να περιγράψουν τον τρόπο με τον οποίο ένα μέλος του πληθυσμού είναι το ίδιο ή διαφορετικό από ένα άλλο ως προς την υπό θεώρηση ιδιότητα ή χαρακτηριστικό γνώρισμα. Έτσι ως προς το φύλο ένα παιδί μπορεί να χαρακτηρισθεί ως αρσενικό ή θηλυκό, ως προς το θρήσκευμα ως χριστιανός ορθόδοξος, καθολικός, μωαμεθανός κλπ. Για τις μεταβλητές ύψος, βάρος, δείκτης νοημοσύνης, αριθμητικές τιμές περιγράφουν το μέγεθός τους. Όλοι οι δυνατοί χαρακτηρισμοί ή οι αριθμητικές τιμές που μπορεί να πάρει μια μεταβλητή ονομάζονται **τιμές της μεταβλητής**.

Οι μεταβλητές ανάλογα με τη φύση, τον τύπο της ιδιότητας των μελών του πληθυσμού που περιγράφουν, διακρίνονται σε ποσοτικές και ποιοτικές.

Ποσοτική μεταβλητή (quantitative variable) είναι εκείνη που μπορεί να μετρηθεί. Παραδείγματα ποσοτικών μεταβλητών είναι η ηλικία, το βάρος, το ύψος, το εισόδημα, το μέγεθος της οικογένειας, ο δείκτης

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

νοημοσύνης κλπ. Κοινό χαρακτηριστικό τους είναι ότι μπορούν να μετρηθούν και κάθε μέλος του πληθυσμού να περιγραφεί από μια αριθμητική τιμή που αναφέρεται στο υπό εξέταση μεταβαλόμενο χαρακτηριστικό του. Ένα σύνολο από αριθμητικές τιμές που περιγράφουν για παράδειγμα το ύψος των μαθητών του Α' Νηπιαγωγείου Ιωαννίνων ή το αντίστοιχο μέγεθος των οικογενειών τους μπορούν να διαταχθούν σε τάξη μεγέθους από τις μεγαλύτερες στις μικρότερες ή και αντίστροφα. Αυτή ακριβώς η δυνατότητα διάταξης, σε τάξη μεγέθους, των τιμών των ποσοτικών μεταβλητών, αποτελεί και μια χαρακτηριστική ένδειξη διάκρισής τους από τις αντίστοιχες ποιοτικές.

Ποιοτική μεταβλητή (qualitative variable) είναι εκείνη που περιγράφει χαρακτηριστικά του πληθυσμού που μεταβάλονται, από μέλος σε μέλος του πληθυσμού, όχι κατά μέγεθος αλλά κατά ποιότητα ή είδος. Έτσι το φύλο μπορεί να είναι αρσενικό ή θηλυκό, η οικογενειακή κατάσταση ανύπαντρος, παντρεμένος ή διαζευγμένος, η εθνικότητα, Έλληνας, Τσεχοσλοβάκος, Τούρκος ή Ιταλός. Οι χαρακτηρισμοί ή οι τιμές μιας ποιοτικής μεταβλητής, δεν διατάσσονται κατά κανόνα, σε τάξη μεγέθους. Μια εθνικότητα ή μια οικογενειακή κατάσταση δεν μπορεί να θεωρηθεί μεγαλύτερη ή μικρότερη από κάποια αντίστοιχη.

Οι ποσοτικές μεταβλητές διακρίνονται σε διακριτές και συνεχείς.

Διακριτή μεταβλητή (discrete variable) είναι εκείνη που παίρνει μόνο μεμονωμένες τιμές που διαφέρουν η μια από την άλλη κατά σταθερή, συνήθως, ποσότητα. Μια οικογένεια μπορεί να αποτελείται από 0, 1, 2, 3 ή περισσότερα παιδιά, όχι όμως από 2.45 παιδιά ή οποιαδήποτε άλλη ενδιάμεση τιμή. Συνήθως οι τιμές των διακριτών μεταβλητών είναι φυσικοί αριθμοί. Ο αριθμός των ασθενών σ' ένα νοσοκομείο, των γεννήσεων το 1989, των κατοίκων μιας πόλης, των τροχαίων ατυχημάτων τον Ιανουάριο αποτελούν παραδείγματα διακριτών μεταβλητών.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Σε αντίθεση με τη διακριτή, μια **συνεχής μεταβλητή** (continuous variable) είναι δυνατόν να πάρει οποιαδήποτε τιμή ενός καθορισμένου πεδίου τιμών της που είναι συνήθως ένα διάστημα της πραγματικής ευθείας. Μεταξύ δύο οποιονδήποτε τιμών μιας συνεχούς μεταβλητής, υπάρχει απεριόριστος αριθμός ενδιάμεσων τιμών. Το βάρος, το ύψος, ο χρόνος αποτελούν παραδείγματα συνεχών μεταβλητών. Η διάκριση των μεταβλητών σε διακριτές και συνεχείς έχει ουσιαστικά θεωρητική αξία. Ενώ το βάρος ενός μαθητή είναι συνεχής μεταβλητή το μεταχειριζόμαστε ως διακριτή αφού το στρογγυλεύουμε συνήθως στον πλησιέστερο φυσικό αριθμό.

Εκτός από τη διάκριση των μεταβλητών που προηγούμενα δόθηκε, υπάρχει η δυνατότητα περαιτέρω ταξινόμησής τους. Η ταξινόμηση αυτή βασίζεται στον διαφορετικό τύπο πληροφορίας που αποκτάται από τη χρησιμοποίηση διαφορετικών τρόπων μέτρησης.

Υποθέτουμε ότι ενδιαφερόμαστε για το ύψος των μαθητών του Α' Νηπιαγωγείου Ιωαννίνων. Υπάρχουν δύο τρόποι για να αποκτήσουμε μια άποψη, να πάρουμε κάποιες πληροφορίες σχετικά με το ύψος των μαθητών. Ο πρώτος είναι ο κατ' ευθείαν έλεγχος του ύψους με τη διάταξη κατά αύξουσα τάξη ύψους των μαθητών. Ο άλλος είναι η μέτρηση και η καταγραφή του ύψους του κάθε μαθητή. Οι δύο αυτές διαδικασίες είναι διαφορετικές όπως διαφορετικός είναι και ο τύπος, η φύση της πληροφορίας που προκύπτει από τις δύο διαδικασίες μέτρησης. Έτσι στην πρώτη περίπτωση είμαστε σε θέση να αποφανθούμε αν κάποιος μαθητής είναι ψηλότερος ή κοντότερος από κάποιο άλλο. Στη δεύτερη περίπτωση επιπρόσθετα δίνεται η δυνατότητα σύγκρισης κατά πόσο ψηλότερος ή κοντότερος είναι ένας μαθητής από κάποιον άλλο. Ανάλογα λοιπόν με τον τύπο της πληροφορίας, που μας παρέχουν διαφορετικοί τρόποι μέτρησης, κατατάσσουμε τις μεταβλητές στις ακόλουθες κατηγορίες: ονοματικές, διατάξιμες, διαστηματικές και αναλογικές. Ο

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

τρόπος μέτρησης ή η διαδικασία μέτρησης αναφέρεται σε ένα σύνολο κανόνων, στην εκτέλεση μαθηματικών και εμπειρικών υπολογισμών για τον προσδιορισμό των τιμών μιας μεταβλητής. Διαφορετικοί τρόποι ή κλίμακες μέτρησης προκύπτουν όταν χρησιμοποιηθούν διαφορετικοί κανόνες για τη μέτρηση ενός χαρακτηριστικού.

Στη βιβλιογραφία, συχνά συναντάμε, τις ονοματικές, διατάξιμες κλπ. μεταβλητές, με τον όρο ονοματική, διατάξιμη, διαστηματική, αναλογική κλίμακα.

Η **ονοματική μεταβλητή** (nominal variable) αναφέρεται και περιγράφει μια ιδιότητα, ως προς την οποία τα μέλη του πληθυσμού κατατάσσονται σε δύο ή περισσότερες, ξένες (διαφορετικές) μεταξύ τους, κατηγορίες. Η μοναδική σχέση σύγκρισης μεταξύ των μελών του πληθυσμού είναι της ισότητας ή της διαφορετικότητας ως προς το υπό μελέτη χαρακτηριστικό. Ενδείξεις για "μικρότερο" ή "μεγαλύτερο" μεταξύ των μελών του πληθυσμού δεν υπάρχουν. Παραδείγματα ονοματικών μεταβλητών είναι: το θρήσκευμα, η εθνικότητα, το φύλο, το χρώμα των μαλλιών.

Συχνά οι τιμές μιας ονοματικής μεταβλητής είναι αριθμοί. Στην περίπτωση αυτή οι αριθμοί χρησιμοποιούνται σαν "ετικέτες" για να χαρακτηρίσουν, να κωδικοποιήσουν τις κατηγορίες στις οποίες τα μέλη του πληθυσμού ταξινομούνται και σε καμιά περίπτωση δεν παριστάνουν το αποτέλεσμα της μέτρησης κάποιου μεγέθους.

Η **διατάξιμη μεταβλητή** (ordinal variable) διαθέτει τα χαρακτηριστικά της ονοματικής μεταβλητής και επιπλέον παρέχει τη δυνατότητα διατύπωσης προτάσεων διάταξης των μελών του υπό μελέτη πληθυσμού. Έτσι τα μέλη του πληθυσμού μπορεί να είναι ίσα, διαφορετικά, το ένα μεγαλύτερο ή μικρότερο από κάποιο άλλο ως προς την ιδιότητα που περιγράφεται από μια διατάξιμη μεταβλητή. Δεν υπάρχουν όμως ενδείξεις για την εξαγωγή συμπερασμάτων της μορφής πόσες φορές ένα

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

μέλος του πληθυσμού είναι μεγαλύτερο ή μικρότερο από κάποιο άλλο. Παράδειγμα διατάξιμης μεταβλητής αποτελεί η έννοια της "διαγωγής". Οι μαθητές μιας τάξης ταξινομούνται στους έχοντες διαγωγή κοσμία και διαγωγή κοσμιώτατη. Οι μαθητές μπορούν να είναι "ίσοι" ή "διαφορετικοί" αν έχουν τον ίδιο τύπο διαγωγής ή όχι. Επιπλέον όμως μπορούν να διαταχθούν, αφού ισχύει η σχέση: κοσμία < κοσμιώτατη , μια σχέση, μια διάκριση που ορίζεται από τους κανόνες συμπεριφοράς που διέπουν τη σχολική ζωή.

Πολλές από τις μεταβλητές που χρησιμοποιούνται στην ψυχολογία και την εκπαίδευση είναι διατάξιμες. Στις επιστήμες αυτές όροι όπως επιθετικότητα, πνεύμα συνεργασίας, συμπεριφορά δικαιολογούν τον ορισμό σχέσης διάταξης μεταξύ των μελών ενός πληθυσμού.

Η **διαστηματική μεταβλητή** (interval variable) αποτελεί μια πολυπλοκότερη και πιο επιτηδευμένη έννοια μεταβλητής. Αναφέρεται σε ιδιότητες, σε χαρακτηριστικά των μελών του πληθυσμού για τα οποία είναι δυνατόν να οριστεί μια σχέση διάταξης μεταξύ των μελών. Επιπλέον υπάρχει η δυνατότητα γνώσης της απόστασης μεταξύ δύο μετρήσεων, μεταξύ δύο τιμών της μεταβλητής. Η δυνατότητα αυτή παρέχεται από τη χρήση μονάδων μέτρησης και ενός μηδενικού σημείου. Και τα δύο ορίζονται αυθαίρετα. Το επιλεγμένο μηδενικό σημείο δεν είναι το απόλυτο μηδέν που απεικονίζει την ολική απουσία, την ανυπαρξία της ποσότητας η οποία μετριέται. Το καλύτερο ίσως παράδειγμα διαστηματικής μεταβλητής αποτελεί η θερμοκρασία. Η μονάδα μέτρησης είναι ο βαθμός και το σημείο σύγκρισης, το αυθαίρετα επιλεγμένο μηδενικό σημείο που είναι οι "μηδέν βαθμοί". Αν θεωρήσουμε τρία αντικείμενα τα Α, Β και Γ με αντίστοιχες θερμοκρασίες 10° C, 20° C και 30° C τότε η διαφορά θερμοκρασίας μεταξύ των Α και Β είναι ίση με τη διαφορά θερμοκρασίας μεταξύ των Β και Γ. Επίσης η διαφορά θερμοκρασίας των Α και Γ είναι διπλάσια της αντίστοιχης διαφοράς μεταξύ των Α και Β ή των Β και Γ. Δεν είναι όμως

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

σωστή η πρόταση ότι το Β έχει διπλάσια θερμοκρασία από το Α ή το Γ τριπλάσια θερμοκρασία από το Α. Με άλλα λόγια αν η χθεσινή θερμοκρασία ήταν 26° C και η σημερινή 13° C δεν σημαίνει ότι η χθεσινή είναι διπλάσια της σημερινής. Από τα παραδείγματα αυτά προκύπτει ότι για μια διαστηματική μεταβλητή αναλογίες έχουν νόημα όταν σχηματίζονται από τις διαφορές μεταξύ των τιμών της και όχι από τις διαφορές μεταξύ των τιμών και του "αυθαίρετου μηδέν".

Ο πιο εξελιγμένος τύπος μεταβλητής είναι η **αναλογική μεταβλητή** (ratio variable). Απαραίτητη προϋπόθεση για τον ορισμό της είναι η ύπαρξη του "αληθινού μηδέν". Οι τιμές της μεταβλητής αποτελούν αποστάσεις από το σημείο αυτό. Έτσι προσδιορίζεται η ισότητα των διαστημάτων και των αναλογιών. Πολλές από τις μεταβλητές των φυσικών επιστημών όπως μήκος, βάρος, επιφάνεια, όγκος, είναι παραδείγματα αναλογικών μεταβλητών. Γι' αυτές τις μεταβλητές έχουν νόημα προτάσεις της μορφής: τρεις φορές βαρύτερο, δύο φορές ψηλότερο κλπ.

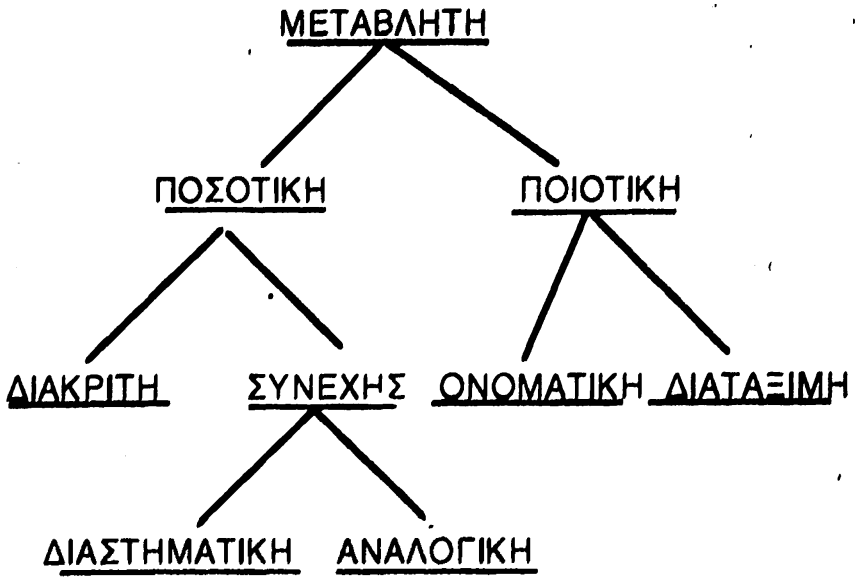
Η ουσιώδης διαφορά μεταξύ αναλογικών και διαστηματικών μεταβλητών έγκειται στο γεγονός ότι οι τιμές της πρώτης είναι αποστάσεις από το αληθινό μηδέν ενώ της δεύτερης από το αυθαίρετα επιλεγμένο μηδέν. Εξ' αιτίας αυτού του γεγονότος για μια αναλογική μεταβλητή, οι ίδιες οι τιμές της δίνουν τη δυνατότητα σχηματισμού αναλογιών. Για μια διαστηματική μεταβλητή αναλογίες έχουν νόημα όταν σχηματίζονται από τις διαφορές μεταξύ των τιμών της. Οι διαφορές αυτές συνιστούν μια αναλογική μεταβλητή γιατί η διαδικάσία που τις δημιουργεί εξαλείφει ή ακυρώνει τη σημασία του αυθαίρετου μηδενικού σημείου.

Από όσα πιο πάνω εκτέθηκαν για τις ονοματικές, διατάξιμες, διαστηματικές και αναλογικές μεταβλητές - που ονομάζονται και κλίμακες μέτρησης - προκύπτει ότι η ονοματική μεταβλητή είναι η ατελέστερη. Κάθε μια από τις υπόλοιπες και με τη σειρά που παρουσιάστηκαν, διατηρεί

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

τα χαρακτηριστικά της προηγούμενης και διαθέτει και κάτι επιπρόσθετο. Τέλος, από τον τύπο των ιδιοτήτων του πληθυσμού που περιγράφουν, οι ονοματικές και διατάξιμες μεταβλητές ανήκουν στην κατηγορία των ποιοτικών μεταβλητών, ενώ οι διαστηματικές και οι αναλογικές στην κατηγορία των ποσοτικών μεταβλητών.

Οι διάφορες κατηγορίες μεταβλητών παραστατικά δίνονται στο παρακάτω διάγραμμα.



Παράδειγμα: Ας υποθέσουμε ότι πρόκειται να εφαρμοστεί ένα πρόγραμμα ανίχνευσης και επιλογής παιδιών προσχολικής ηλικίας με "αυξημένες αθλητικές ικανότητες". Εκτιμάται ότι οι παράγοντες που επηρεάζουν τις αθλητικές ικανότητες είναι ο δείκτης νοημοσύνης, το ύψος, ο χρόνος που απαιτείται για να διανυθεί τρέχοντας απόσταση 100 μέτρων, η συμπεριφορά, η οικονομική κατάσταση της οικογένειας.

Σε μια προκαταρτική μελέτη, από τον πληθυσμό των παιδιών προσχολικής ηλικίας των Ιωαννίνων, εκλέγεται ένα τυχαίο δείγμα 35 παιδιών και για κάθε παιδί εξετάζεται ο δείκτης νοημοσύνης του, το ύψος, ο χρόνος σε δευτερόλεπτα, που διανύει τα 100 μέτρα η συμπεριφορά και η οικονομική κατάσταση της οικογένειας. Η συμπεριφορά του αξιολογείται

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

με βάση τη διαγωγή του και χαρακτηρίζεται ως Α αν πρόκειται για κοσμιώτατη και ως Β αν πρόκειται για κόσμια. Η οικονομική κατάσταση της οικογένειας χαρακτηρίζεται ως Α, Β, Γ και Δ αν το μηνιαίο εισόδημα της οικογένειας σε δραχμές είναι από 0-150.000 δρχ., από 150.000-300.000 δρχ. από 300.000-400.000 δρχ. και από 400.000 δρχ. και πανω αντίστοιχα.

Τα αποτελέσματα της εξέτασης των παιδιών προσχολικής ηλικίας ως προς τα χαρακτηριστικά που προαναφέρθηκαν δίνονται στον επόμενο πίνακα. Επίσης καταγράφεται το φύλο κάθε παιδιού, με την ένδειξη "Α" αν πρόκειται για αγόρι και την ένδειξη "Θ" αν πρόκειται για κορίτσι.

Α/Α	ΦΥΛΟ	ΔΙΑΓΩΓΗ	ΟΙΚΟΝΟΜΙΚΗ ΚΑΤΑΣΤΑΣΗ ΟΙΚΟΓΕΝ.	ΔΕΙΚΤΗΣ ΝΟΗΜΟΣΥΝ.	ΥΨΟΣ ΣΕ ΕΚΑΤΟΣΤΑ (cm)	ΧΡΟΝΟΣ ΣΕ ΔΕΥΤΕΡΟΛΕΠΤΑ
1	Α	Β	Β	111	95	22
2	Θ	Α	Γ	90	98	25
3	Θ	Α	Γ	90	92	18
4	Θ	Α	Γ	90	104	19
5	Α	Α	Α	104	85	21
6	Α	Α	Β	72	96	20
7	Θ	Β	Β	105	89	21
8	Α	Α	Δ	93	103	22
9	Α	Α	Γ	99	110	18
10	Α	Α	Β	93	85	27
11	Α	Α	Β	84	94	30
12	Θ	Α	Α	95	98	21
13	Θ	Α	Γ	93	96	24
14	Α	Α	Δ	78	99	26
15	Θ	Α	Β	108	83	19
16	Θ	Β	Β	100	87	27
17	Α	Α	Α	81	85	25
18	Α	Α	Γ	77	97	24
19	Α	Α	Γ	67	96	23
20	Α	Α	Α	100	107	28
21	Α	Α	Β	104	102	29
22	Θ	Α	Δ	111	106	19
23	Θ	Α	Δ	122	95	20
24	Α	Α	Β	99	82	28
25	Θ	Α	Β	108	94	31
26	Θ	Α	Α	126	90	19
27	Α	Β	Α	90	90	23
28	Α	Α	Γ	110	96	32
29	Α	Α	Γ	117	87	27
30	Θ	Α	Α	119	97	24

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

31	Θ	Α	Δ	105	90	22
32	Α	Β	Β	100	107	18
33	Θ	Α	Α	75	92	25
34	Α	Α	Γ	96	98	30
35	Θ	Α	Δ	81	95	23

Πίνακας 1.1: Δεδομένα παιδιών προσχολικής ηλικίας

Οι μεταβλητές που χρησιμοποιούνται στο προηγούμενο παράδειγμα αναφέρονται στο φύλο, τη διαγωγή, την οικονομική κατάσταση, το δείκτη νοημοσύνης, το ύψος και το χρόνο στον οποίο ένα παιδί τρέχοντας καλύπτει την απόσταση των 100 μέτρων. Είναι προφανές ότι οι τρεις πρώτες μεταβλητές είναι ποιοτικές και οι υπόλοιπες ποσοτικές. Ειδικότερα, η μεταβλητή που παριστά το φύλο είναι ονοματική, ενώ οι μεταβλητές που περιγράφουν τη διαγωγή και την οικονομική κατάσταση, διατάξιμες. Επίσης οι μεταβλητές που παριστούν το ύψος και το χρόνο είναι αναλογικές ενώ η μεταβλητή που περιγράφει το δείκτη νοημοσύνης είναι διαστηματική λόγω μη ύπαρξης του απόλυτου μηδέν. Στο πίνακα που ακολουθεί κωδικοποιούνται οι μεταβλητές του Πίνακα 1.1.

Α/Α	ΟΝΟΜΑΣΙΑ ΜΕΤΑΒΛΗΤΗΣ	ΚΩΔΙΚΟΣ	ΤΥΠΟΣ ΜΕΤΑΒΛΗΤΗΣ
1	ΦΥΛΟ	1. ΑΓΟΡΙ 2. ΚΟΡΙΤΣΙ	ΟΝΟΜΑΤΙΚΗ
2	ΔΙΑΓΩΓΗ	1. Α (Κοσμιώτατη) 2. Β (Κοσμία)	ΔΙΑΤΑΞΙΜΗ
3	ΟΙΚΟΝΟΜΙΚΗ ΚΑΤΑΣΤΑΣΗ	1. Α 2. Β 3. Γ 4. Δ	ΔΙΑΤΑΞΙΜΗ
4	ΔΕΙΚΤΗΣ ΝΟΗΜΟΣΥΝΗΣ		ΔΙΑΣΤΗΜΑΤΙΚΗ
5	ΥΨΟΣ		ΑΝΑΛΟΓΙΚΗ
6	ΧΡΟΝΟΣ		ΑΝΑΛΟΓΙΚΗ

Πίνακας 1.2: Κωδικοποίηση μεταβλητών Πίνακα 1.1.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

31	Θ	A	Δ	105	30	22
32	A	B	B	100	10	18
33	Θ	A	A	75	32	25
34	A	A	A	98	35	20
35	Θ	A	Δ	81	35	23

Πίνακας 1.1: Δεδομένα παίδων προσχολικής ηλικίας

Οι μεταβλητές του χαρακτηριστούνται στα προηγούμενο παράδειγμα αναφέρονται στο φύλο, τη διαγωγή, την οικονομική κατάσταση, το δείκτη νοημοσύνης, το ύψος και το χρόνο στον οποίο ένα παιδί τρέχοντας καλύπτει την απόσταση των 100 μέτρων. Είναι προφανές ότι οι τρεις πρώτες μεταβλητές είναι ποιοτικές και οι υπόλοιπες ποσοτικές. Εξάπτερα, η μεταβλητή που παριστά το φύλο είναι ανομαστική, ενώ οι μεταβλητές που περιγράφουν τη διαγωγή και την οικονομική κατάσταση, διατάξεις. Επίσης οι μεταβλητές που παριστούν το ύψος και το χρόνο είναι αναλογικές ενώ η μεταβλητή που περιγράφει το δείκτη νοημοσύνης είναι διαστηματική λόγω μη ύπαρξης του απόλυτου μηδέν. Στο πίνακα που ακολουθεί κωδικοποιούνται οι μεταβλητές του Πίνακα 1.1.

ΑΑ	ΟΝΟΜΑΣΙΑ ΜΕΤΑΒΛΗΤΗΣ	ΚΩΔΙΚΟΣ	ΤΥΠΟΣ ΜΕΤΑΒΛΗΤΗΣ
1	ΦΥΛΟ	1. ΑΓΟΡΙ 2. ΚΟΡΙΤΣΙ	ΟΝΟΜΑΤΙΚΗ
2	ΔΙΑΓΩΓΗ	1. Α (Καλοσύνη) 2. Β (Κοπιάζει)	ΔΙΑΤΑΞΙΜΗ
3	ΟΙΚΟΝΟΜΙΚΗ ΚΑΤΑΣΤΑΣΗ	1. Α 2. Β 3. Γ 4. Δ	ΔΙΑΤΑΞΙΜΗ
4	ΔΕΙΚΤΗΣ ΝΟΗΜΟΣΥΝΗΣ		ΔΙΑΣΤΗΜΑΤΙΚΗ
5	ΥΦΟΣ		ΑΝΑΛΟΓΙΚΗ
6	ΧΡΟΝΟΣ		ΑΝΑΛΟΓΙΚΗ

Πίνακας 1.2: Κωδικοποίηση μεταβλητών Πίνακα 1.1

ΚΕΦΑΛΑΙΟ 2

ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

2.1 Εισαγωγικά

Σε οποιαδήποτε στατιστική μελέτη ενός προβλήματος βρισκόμαστε αντιμέτωποι με ένα πλήθος δεδομένων, κατά κανόνα αριθμητικών, που είναι το αποτέλεσμα κάποιας πειραματικής διαδικασίας ή προκύπτουν από διαθέσιμες πηγές. Σκοπός μας είναι η ανάλυση των δεδομένων αυτών και η εξαγωγή συμπερασμάτων για τον πληθυσμό απ' τον οποίο προέρχονται. Πριν όμως από οποιαδήποτε ανάλυση καθίσταται προφανής η ανάγκη περιληπτικής παρουσίασης των δεδομένων. Οι αρχές και οι μέθοδοι που χρησιμοποιούνται για την οργάνωση και την περιληπτική παρουσίαση των δεδομένων αποτελούν αντικείμενο της **περιγραφικής στατιστικής**. Η συνοπτική παρουσίαση των δεδομένων επιτυγχάνεται: i) με τους πίνακες συχνοτήτων και ii) με τις γραφικές τους παραστάσεις. Κάθε μια από αυτές τις μεθόδους βοηθά, με το δικό της τρόπο, στην ανίχνευση της φύσης σπουδαίων χαρακτηριστικών των δεδομένων και κατ' επέκταση του πληθυσμού από τον οποίο προέρχονται. Μια ακόμη δραστικότερη

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

σύμπτυξη των δεδομένων επιτυγχάνεται με την χρήση αριθμητικών μεγεθών τα οποία θα μας απασχολήσουν στο επόμενο κεφάλαιο.

2.2 Πίνακες συχνοτήτων

9/03/2001.

Αποτελούν μια ταξινόμηση, μια κωδικοποίηση των δεδομένων στην οποία καταγράφεται η συχνότητα εμφάνισης των τιμών σε αυθαίρετα ορισμένα διαστήματα. Στα εδάφια που ακολουθούν δίνονται οι αρχές και μέθοδοι κατασκευής πινάκων συχνοτήτων για ποσοτικές και ποιοτικές μεταβλητές.

2.2.1 Πίνακες συχνοτήτων ποσοτικών μεταβλητών

2.2.1.1 Διατεταγμενος πίνακας συχνοτήτων

Περιγράφεται η κατασκευή του για τα δεδομένα του δείκτη νοημοσύνης των 35 παιδιών προσχολικής ηλικίας του Πίνακα 1.1. Τα δεδομένα αυτά είναι:

111,	90,	90,	90,	104,	72,	105,
93,	99,	93,	84,	95,	93,	78,
108,	100,	81,	77,	67,	100,	104,
111,	122,	99,	108,	126,	90,	110,
117,	119,	105,	100,	75,	96,	81.

Από τις μετρήσεις αυτές παρατηρούμε ότι σε περισσότερα από ένα παιδιά αντιστοιχεί ο ίδιος δείκτης νοημοσύνης ή μια μέτρηση

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

εμφανίζεται περισσότερο από μία φορά στο δείγμα. Στο παράδειγμα, η τιμή 90 εμφανίζεται 4 φορές, η τιμή 93 εμφανίζεται 3 φορές, η τιμή 108 εμφανίζεται 2 φορές κ.ο.κ. Αυτονόητο είναι σε μια προσπάθεια σύμπτυξης των δεδομένων, οι επαναλαμβανόμενες τιμές να μην λαμβάνονται υπ' όψη και να αντιστοιχίζονται με τον αριθμό των φορών που αυτές εμφανίζονται στο δείγμα. Έτσι σαν ένα πρώτο βήμα περιληπτικής παρουσίασης των δεδομένων, διατάσσονται οι τιμές του δείγματος σε αύξουσα (ή φθίνουσα) τάξη, για τον εντοπισμό των τιμών που εμφανίζονται περισσότερο από μία φορά. Για το παράδειγμά μας τα διατεταγμένα δεδομένα είναι:

67,	72,	75,	77,	78,	81,	81,
84,	90,	90,	90,	90,	93,	93,
93,	95,	96,	99,	99,	100,	100,
100,	104,	104,	105,	105,	108,	108,
110,	111,	111,	117,	119,	122,	126.

Αγνοώντας τις τιμές που εμφανίζονται στο δείγμα περισσότερο από μια φορά, και χρησιμοποιώντας την έννοια της **συχνότητας** (frequency) - τη συμβολίζουμε με f_i και παριστά τον αριθμό των φορών που η τιμή x_i εμφανίζεται στο δείγμα - κατασκευάζουμε τον διατεταγμένο πίνακα συχνοτήτων της επόμενης σελίδας.

Ο διατεταγμένος πίνακας συχνοτήτων μπορεί να εμπλουτιστεί ώστε να περιλαμβάνει τη σχετική συχνότητα, την αθροιστική συχνότητα και τη σχετική αθροιστική συχνότητα, έννοιες που κάθε μιά με το δικό της τρόπο συμβάλλει στην περιγραφή ενός συνόλου δεδομένων.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

αριθμός αυξοντα

Α/Α	ΜΕΤΡΗΣΗ (x_i)	ΣΥΧΝΟΤΗΤΑ (f_i)	Α/Α	ΜΕΤΡΗΣΗ (x_i)	ΣΥΧΝΟΤΗΤΑ (f_i)
1	67	1	12	99	2
2	72	1	13	100	3
3	75	1	14	104	2
4	77	1	15	105	2
5	78	1	16	108	2
6	81	2	17	110	1
7	84	1	18	111	2
8	90	4	19	117	1
9	93	3	20	119	1
10	95	1	21	122	1
11	96	1	22	126	1

→ είναι ταυτόσημη με την έννοια του πλήθους.

Η σχετική συχνότητα της τιμής x_i του δείγματος είναι το πηλίκο f_i / N, με N το πλήθος των μετρήσεων δηλ. το μέγεθος του δείγματος, και παριστά το ποσοστό ή το επί τοις εκατό (%) των φορών που η τιμή x_i εμφανίζεται στο δείγμα για $i=1,...,v$, με v πλέον τον αριθμό των διαφορετικών τιμών x_i του δείγματος. **Η αθροιστική συχνότητα** της τιμής x_i του δείγματος, συμβολίζεται με F_i, και παριστά το πλήθος των τιμών του δείγματος που είναι μικρότερες ή το πολύ ίσες από την x_i, ενώ η **σχετική αθροιστική συχνότητα** της τιμής x_i είναι το πηλίκο F_i / N και παριστά το ποσοστό ή το επί τοις εκατό (%) των τιμών του δείγματος που είναι μικρότερες ή ίσες από την τιμή x_i, $i=1,...,v$. Οι αθροιστικές συχνότητες και οι σχετικές αθροιστικές συχνότητες της τιμής x_i, $i=1,...,v$, από τον τρόπο που πιο πάνω ορίστηκαν, προκύπτουν αντίστοιχα από την πρόσθεση των συχνοτήτων

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

και των σχετικών συχνοτήτων που αντιστοιχούν στις τιμές x_j του δείγματος για τις οποίες $x_j \le x_i$, $i,j = 1,\dots,v$.

Έτσι ο πλήρης διατεταγμένος πίνακας συχνοτήτων για τα δεδομένα του δείκτη νοημοσύνης του Πίνακα 1.1 είναι:

κατασκευή

διαίρεση → τη συχνότητα της καθε μετρησης με το πληθος των μετρήσεων

Α/Α	ΜΕΤΡΗΣΗ (x_j)	ΣΥΧΝΟΤΗΤΑ (f_j)	ΣΧΕΤ. ΣΥΧΝ. (f_j/ N)	ΑΘΡ. ΣΥΧΝΟΤ. (F_j)	ΣΧ.ΑΘΡ.ΣΥΧ. (F_j/ N)
1	67	1	0.0286	1	0.0286
2	72	1	0.0286	2	0.0571
3	75	1	0.0286	3	0.0857
4	77	1	0.0286	4	0.1142
5	78	1	0.0286	5	0.1428
6	81	2	0.0571	7	0.2
7	84	1	0.0286	8	0.2286 %
8	90	4	0.1142	12	0.3428
9	93	3	0.0857 %	15	0.4286
10	95	1	0.0286	16	0.4571
11	96	1	0.0286	17	0.4857
12	99	2	0.0571	19	0.5428
13	100	3	0.0857	22	0.6286
14	104	2	0.0571	24	0.6857
15	105	2	0.0571	26	0.7428
16	108	2	0.0571	28	0.8
17	110	1	0.0286	29	0.8286
18	111	2	0.0571	31	0.8857
19	117	1	0.0286	32	0.9143

εχουν δεικτη νοημοσ. ίσο ή μικροτερο απο το ιος.

Ερμηνεία → ειδραση

(διαιρώ τη καθε αθρ. συχνοτητα με το πληθος των ~~[illegible]~~ μετρήσεων.

12 → προσθετουμε τη συχνότητα f_j απο το [illegible]

20	119	1	0.0286	33	0.9428
21	122	1	0.0286	34	0.9714
22	126	1	0.0286	35	1.0000
ΣΥΝΟΛΟ		35	1		

Πίνακας 2.1: Διατεταγμένος πίνακας συχνοτήτων των δεδομένων του δείκτη νοημοσύνης του Πίνακα 1.1.

Η πρώτη στήλη του διατεταγμένου πίνακα συχνοτήτων δημιουργείται από τους αύξοντες αριθμούς των διαφορετικών τιμών του δείγματος. Η τελευταία τιμή της στήλης αυτής είναι το πλήθος των διαφορετικών τιμών του δείγματος που έχουμε συμβολίσει με ν. Στη δεύτερη στήλη καταγράφονται σε αύξουσα τάξη μεγέθους οι διαφορετικές τιμές του δείγματος, ενώ στην τρίτη στήλη οι αντίστοιχες συχνότητες. Η στήλη των σχετικών συχνοτήτων δημιουργείται από το πηλίκο της αντίστοιχης συχνότητας δια του μεγέθους Ν του δείγματος. Για το παράδειγμά μας το Ν=35. Κάθε στοιχείο της στήλης των αθροιστικών συχνοτήτων είναι ίσο με το άθροισμα των προηγουμένων και της ιδίας θέσης συχνοτήτων, ενώ η στήλη των σχετικών αθροιστικών συχνοτήτων δημιουργείται από το πηλίκο της αντίστοιχης αθροιστικής συχνότητας δια του μεγέθους Ν του δείγματος. Επίσης κάθε στοιχείο της στήλης των σχετικών αθροιστικών συχνοτήτων είναι ίσο με το άθροισμα των προηγουμένων και της ίδιας θέσης σχετικών συχνοτήτων.

Για τον έλεγχο της ορθότητας του διατεταγμένου πίνακα συχνοτήτων θα πρέπει το άθροισμα των συχνοτήτων να είναι ίσο με το

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

μεγέθος N του δείγματος και το άθροισμα των σχετικών συχνοτήτων ίσο με 1, όπως φαίνεται στην τελευταία γραμμή του Πίνακα 2.1. Επίσης η τελευταία τιμή της στήλης των αθροιστικών συχνοτήτων να είναι ίση με το μέγεθος του δείγματος ενώ η τελευταία τιμή της στήλης των σχετικών αθροιστικών συχνοτήτων ίση με 1.

9/03/2001.

2.2.1.2 Ομαδοποιημένος πίνακας συχνοτήτων

Η προηγούμενη διαδικασία σύμπτυξης των δεδομένων παρά το γεγονός ότι δεν οδηγεί σε απώλειες μετρήσεων, λίγο συμβάλλει στην περιληπτική παρουσίαση των δεδομένων. Μία περαιτέρω σύμπτυξη των δεδομένων μπορεί να επιτευχθεί με την ομαδοποίησή τους. Για την ομαδοποίηση ενός συνόλου μετρήσεων επιλέγουμε ένα σύνολο από συγγενή, μη επικαλυπτόμενα διαστήματα τέτοια ώστε κάθε τιμή του δείγματος να ανήκει σε ένα και μόνο ένα από αυτά. Τα διαστήματα αυτά ονομάζονται **κλάσεις** ή **ομάδες**.

Το πρώτο πρόβλημα που εμφανίζεται κατά την ομαδοποίηση των δεδομένων αφορά το πλήθος των ομάδων που θα χρησιμοποιηθούν. Αν χρησιμοποιηθούν λίγες ομάδες υπάρχει ο κίνδυνος σημαντικής απώλειας πληροφορίας λόγω απώλειας μετρήσεων, ενώ η χρησιμοποίηση πολλών ομάδων λειτουργεί σε βάρος της δραστικής σύμπτυξης των δεδομένων. Ο καλύτερος ίσως οδηγός δίνεται από τον εμπειρικό κανόνα του Sturges σύμφωνα με τον οποίο προτείνεται η χρησιμοποίηση

ομάδες. > πλήθος δεδομένων.

$$k = 1 + 3.322(\log_{10} N)$$

ομάδων, με N το πλήθος των μετρήσεων ή το μέγεθος του δείγματος.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Ένα άλλο πρόβλημα, κατά την ομαδοποίηση των δεδομένων, αφορά το μήκος των ομάδων. Συνήθως κατασκευάζονται ομάδες ίσου μήκους. Το μήκος προκύπτει από το πηλίκο του εύρους των μετρήσεων δια του αριθμού των ομάδων.

Το **εύρος** (range) των μετρήσεων συμβολίζεται με R και ορίζεται να είναι η διαφορά της ελάχιστης τιμής του δείγματος ($\min x_i$) από τη μέγιστη τιμή του δείγματος ($\max x_i$) δηλαδή

μεγαλύτερη από τις μετρήσεις αν αφαιρέσω τη μικρότερη.

$$R = \max x_i - \min x_i .$$

Τότε το **μήκος των ομάδων** συμβολίζεται με d και είναι:

πλήθος ομάδων

$$d = \frac{R}{k} = \frac{\max x_i - \min x_i}{k} .$$

Για τα δεδομένα του δείκτη νοημοσύνης του Πίνακα 1.1 ο αριθμός των ομάδων είναι:

$$k = 1 + 3.322 * \log_{10} 35 = 1 + 3.322 * 1.54$$

$$= 6.1159 \approx 6 \text{ ομάδες.}$$

Το εύρος R των μετρήσεων

$$R = \max x_i - \min x_i = 126 - 67 = 59$$

και το μήκος της κάθε ομάδας

$$d = \frac{R}{k} = \frac{59}{6} = 9.833 \approx 10 .$$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Αφού προσδιοριστεί το πλήθος των ομάδων που θα χρησιμοποιηθούν και το αντίστοιχο μήκος τους η κατασκευή πλέον των ομάδων επιτυγχάνεται ξεκινώντας από τη μικρότερη τιμή του δείγματος και καταλήγοντας στη μεγαλύτερη τιμή του δείγματος, προσθέτοντας κάθε φορά ποσότητα ίση με το μήκος της ομάδας.

Έτσι για τα δεδομένα του δείκτη νοημοσύνης του Πίνακα 1.1 ορίζονται οι ακόλουθες ομάδες:

67-77, 77- 87, 87- 97, 97-107, 107-117, 117-127.

Η έννοια της συχνότητας μιας δειγματικής τιμής, που πιο πάνω ορίστηκε, κατ' αναλογία μεταφέρεται και στην περίπτωση των ομάδων. Έτσι η **συχνότητα της i-ομάδας**, συμβολίζεται με f_i , i = 1,2,...,k, και ορίζεται να είναι το πλήθος των τιμών του δείγματος που ανήκουν σ' αυτή την ομάδα. Ανάλογα η σχετική συχνότητα μιας ομάδας, f_i / N, i = 1,...,k , παριστά το ποσοστό των τιμών του δείγματος που ανήκουν στην ομάδα, η αθροιστική συχνότητα F_i το πλήθος των δειγματικών τιμών που ανήκουν στις προηγούμενες αλλά και στην εν λόγω ομάδα, και η σχετική αθροιστική συχνότητα ομάδας, F_i / N, i = 1,...,k, το αντίστοιχο ποσοστό των τιμών του δείγματος.

Με τη χρήση των εννοιών αυτών η κατασκευή του ομαδοποιημένου πίνακα συχνοτήτων, που θα δοθεί στη συνέχεια, είναι εντελώς ανάλογη με εκείνη του διατεταγμένου πίνακα συχνοτήτων.

Ένα πρόβλημα που συχνά εμφανίζεται στην πράξη είναι η ύπαρξη τιμών του δείγματος που συμπίπτουν με το άνω όριο (άκρο) μιας ομάδας και με το κάτω όριο της επόμενης. Για το παράδειγμα μας οι δειγματικές τιμές 77 και 117 συμπίπτουν με το άνω όριο της πρώτης ομάδας και το κάτω όριο της δεύτερης, και με το άνω όριο της πέμπτης

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

όμάδας και το κάτω όριο της έκτης ομάδας αντίστοιχα. Το πρόβλημα αυτό μπορεί να ξεπεραστεί με δυο τρόπους ανάλογα αν η ποσοτική μεταβλητή είναι διακριτή ή συνεχής. Για διακριτές μεταβλητές συνίσταται η χρησιμοποίηση "κλειστών - ανοιχτών" διαστημάτων, δηλαδή διαστημάτων στα οποία περιλαμβάνονται το κάτω όριο όχι όμως και το άνω.

Για το παράδειγμά μας οι ομάδες θα αποτελούνταν από τα διαστήματα:

[67,77), [77,87), [87,97), [97,107) , [107,117), [117,127).

Επίσης συχνά οι ομάδες είναι "κλειστά" διαστήματα της μορφής:

[67,77], [78,88], [89,99], [100,110], [111,121], [122,132].

Όταν η υπό θεώρηση μεταβλητή είναι συνεχής και οι δειγματικές τιμές έχουν παραχθεί με κάποια ακρίβεια μέτρησης (με τη χρήση συγκεκριμένου αριθμού δεκαδικών ψηφίων), συνίσταται η χρησιμοποίηση ομάδων των οποίων τα όρια δημιουργούνται με μεγαλύτερη ακρίβεια από τις δειγματικές τιμές κατά ένα δεκαδικό ψηφίο. Τα όρια των ομάδων που προκύπτουν με τον τρόπο αυτό, ονομάζονται **πραγματικά όρια** (exact limits) και δημιουργούνται:

α. Από τη μείωση κατά μισή μονάδα των ορίων των "κλειστών - ανοιχτών " ομάδων ή

β. Από τη μείωση κατά μισή μονάδα του κάτω ορίου και την αύξηση κατά μισή μονάδα του άνω ορίου αν πρόκειται για κλειστές ομάδες, έτσι ώστε οι ομάδες με μεγαλύτερη ακρίβεια ορίων που προκύπτουν να μην επικαλύπτονται. Έτσι για το παράδειγμά μας, αν ο δείκτης νοημοσύνης θεωρηθεί συνεχής μεταβλητή, τα πραγματικά όρια θα είναι:

[66.5,76.5], [76.5,86.5], [86.5,96.5], [96.5,106.5], [106.5,116.5] και

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

[116.5,126.5] για "κλειστές - ανοικτές" ομάδες και

[66.5,77.5], [77.5,88.5], [88.5,99.5], [99.5,110.5], [110.5,121.5],

και [121.5,132.5], για "κλειστές" ομάδες.

Η δημιουργία των πραγματικών ορίων, που πιο πάνω αναφέρθηκε, δεν είναι αυθαίρετη όσο από πρώτη άποψη φαίνεται. Πράγματι αν ο δείκτης νοημοσύνης είναι συνεχής μεταβλητή τότε ένας μαθητής με δείκτη νοημοσύνης 77 μπορεί στην πραγματικότητα να έχει δείκτη νοημοσύνης οποιαδήποτε τιμή μεταξύ του 76.5 και του 77.5, ή γενικότερα οποιαδήποτε τιμή σ' ένα οσοδήποτε μικρό διάστημα που περιέχει το 77. Η αντιστοίχιση της νοημοσύνης του μαθητή με το 77 και όχι, για παράδειγμα, με το 76.993 .... είναι επιλογή του ατόμου που εκτέλεσε το τεστ νοημοσύνης, μια επιλογή που απορρέει από τη φύση του προβλήματος που ενδιαφέρεται να μελετήσει. Για να γίνει πιο κατανοητή η έκφραση από τη φύση του προβλήματος, αναφέρουμε ότι η στρογγύλευση του βάρους ενός ανθρώπου στον πλησιέστερο ακέραιο δεν θα είχε πιθανώς κανένα κόστος, ενώ η αντίστοιχη στρογγύλευση του βάρους μιας ποσότητας χρυσού θα είχε τεράστιο κόστος.

Η έννοια της **μέσης τιμής ή αντιπροσωπευτικής τιμής** μιας ομάδας επιβάλλεται όχι μόνο από την ανάγκη ευκολότερου υπολογισμού αριθμητικών μεγεθών και κατασκευής γραφικών παραστάσεων, όπως θα φανεί σε επόμενα εδάφια, αλλά από την ανάγκη ύπαρξης μιας τιμής που θα αναπαριστάνει, θα αντιπροσωπεύει όλες τις υπόλοιπες που βρίσκονται στα ίδια όρια. Ορίζουμε σαν μέση τιμή m_i μιας ομάδας, εκείνο το σημείο που απέχει εξίσου από τα πραγματικά της όρια. Ο προσδιορισμός της μπορεί να γίνει με δυο τρόπους:

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

α) Διαιρούμε το πραγματικό εύρος της ομάδας με το 2 και προσθέτουμε το αποτέλεσμα στο κάτω πραγματικό όριο.

β) Προσθέτουμε τα πραγματικά όρια της ομάδας και διαιρούμε με το 2.

Έτσι η μέση τιμή της ομάδας [86.5,96.5] είναι:

$$\frac{96.5-\ 86.5}{2} = 5, \quad m_i = 86.5+ 5 = 91.5$$

ή $$m_i = \frac{86.5+96.5}{2} = 91.5 .$$

Ο ομαδοποιημένος πίνακας συχνοτήτων των δεδομένων του δείκτη νοημοσύνης του Πίνακα 1.1 είναι:

ΟΜΑΔΑ	ΟΡΙΑ ΟΜΑΔΑΣ	ΠΡΑΓΜΑΤΙΚΑ ΟΡΙΑ	Μ.ΤΙΜ. ΟΜΑΔΑΣ (m_i)	ΣΥΧΝ. (f_i)	ΣΧΕΤ. ΣΥΧΝΟΤΗΤΑ (f_i / N)	ΑΘΡΟΙΣΤΙΚΗ ΣΥΧΝΟΤΗΤΑ (F_i)	ΣΧ. ΑΘΡ. ΣΥΧΝΟΤ. (F_i / N)
1	[67,77)	[66.5,76.5]	71.5	3	0.0857	3	0.0857
2	[77,87)	[76.5,86.5]	81.5	5	0.1428	8	0.2286
3	[87,97)	[86.5,96.5]	91.5	9	0.2572	17	0.4857
4	[97,107)	[96.5,106.5]	101.5	9	0.2572	26	0.7428
5	[107,117)	[106.5,116.5]	111.5	5	0.1428	31	0.8857
6	[117,127)	[116.5,126.5]	121.5	4	0.1141	35	1.000
ΣΥΝΟΛΟ				35	1.0000		

Πίνακας 2.2: Ομαδοποιημένος πίνακας συχνοτήτων των δεδομένων του δείκτη νοημοσύνης του Πίνακα 1.1.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Παρατηρήσεις: 1. Ο Πίνακας 2.2 κατασκευάστηκε θεωρώντας το δείκτη νοημοσύνης συνεχή μεταβλητή. Αν θεωρούνταν διακριτή μεταβλητή η στήλη των πραγματικών ορίων θα ήταν η ίδια με την προηγούμενη.

2. Η στήλη των μέσων τιμών των ομάδων προκύπτει αν στην προηγούμενη μέση τιμή προστεθεί το μήκος της ομάδας.

3. Η στήλη των συχνοτήτων δημιουργείται από την απαρίθμηση και καταγραφή των δειγματικών τιμών που βρίσκονται μεταξύ των πραγματικών ορίων.

4. Οι στήλες των σχετικών, αθροιστικών και σχετικών αθροιστικών συχνοτήτων καθώς και ο έλεγχος ορθότητας του ομαδοποιημένου πίνακα συχνοτήτων γίνεται όπως και στον διατεταγμένο πίνακα συχνοτήτων.

5. Συχνά τα δεδομένα περιέχουν τιμές που είναι πολύ μικρότερες ή πολύ μεγαλύτερες από τις άλλες. Στην περίπτωση αυτή αν γίνει ομαδοποίηση σύμφωνα με όσα έχουν αναφερθεί, θα προκύψουν ομάδες με μηδενική συχνότητα. Το πρόβλημα αντιμετωπίζεται με τη χρήση διαστημάτων απείρου μήκους για την πρώτη και την τελευταία ομάδα. Έτσι έχουμε καλύτερη εικόνα της μεταβλητής, προκύπτουν όμως δυσκολίες στον υπολογισμό των αριθμητικών μεγεθών του δείγματος. Στο παράδειγμά μας αν υπήρχαν οι μετρήσεις 12 και 138 η πρώτη ομάδα θα μπορούσε να γίνει $(-\infty , 76.5)$ και η τελευταία $[116.5,\infty)$.

6. Η ομαδοποίηση των μετρήσεων έχει σαν συνέπεια την απώλεια πληροφορίας. Η απώλεια αυτή είναι το κόστος της πρακτικής σημασίας αυτών των πινάκων των οποίων η ύπαρξη αποτελεί προϋπόθεση για την κατασκευή γραφικών παραστάσεων.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

2.2.2 Πίνακες συχνοτήτων ποιοτικών μεταβλητών

Η περιληπτική, συνοπτική παρουσίαση ποιοτικών δεδομένων είναι σχετικά απλούστερη. Από τη φύση των ποιοτικών μεταβλητών δεν υπάρχει διάκριση σε διατεταγμένο και ομαδοποιημένο πίνακα συχνοτήτων, δεν εμφανίζεται η έννοια της ομάδας, των πραγματικών ορίων, των μέσων τιμών των ομάδων. Η ποιοτική μεταβλητή περιγράφει χαρακτηριστικά των μελών του πληθυσμού που μεταβάλλονται κατά ποιότητα ή είδος και όχι κατά μέγεθος. Κάθε άτομο του πληθυσμού και επομένως και του δείγματος ανήκει σε μια και μόνο μια από τις κατηγορίες που χαρακτηρίζουν την ποιοτική μεταβλητή. Έτσι ο πίνακας συχνοτήτων μιας ποιοτικής μεταβλητής προκύπτει από την απαρίθμηση και καταγραφή των δειγματικών τιμών στην αντίστοιχη κατηγορία.

Οι πίνακες συχνοτήτων που αφορούν το φύλο και την οικονομική κατάσταση της οικογένειας των δεδομένων του Πίνακα 1.1 είναι αντίστοιχα:

ΚΑΤΗΓΟΡΙΑ	ΣΥΧΝΟΤΗΤΑ (f_i)	ΣΧΕΤΙΚΗ ΣΥΧΝΟΤΗΤΑ (f_i / N)
ΑΓΟΡΙ Α	19	0.5428
ΚΟΡΙΤΣΙ Θ	16	0.4572
ΣΥΝΟΛΟ	35	1.0000

Πίνακας 2.3: Πίνακας συχνοτήτων των δεδομένων του φύλου του Πίνακα 1.1.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΚΑΤΗΓΟΡΙΑ	ΣΥΧΝΟ - ΤΗΤΑ (f_j)	ΣΧΕΤΙΚΗ ΣΥΧΝΟΤΗΤΑ (f_j / N)	ΑΘΡ. ΣΥ- ΧΝΟΤΗΤΑ (F_j)	ΣΧΕΤ.ΑΘΡ. ΣΥΧΝΟΤΗΤΑ (F_j / N)
"μέχρι 150000" Α	8	0.2286	8	0.2286
"150000-300000" Β	11	0.3143	19	0.5429
"300000-400000" Γ	10	0.2857	29	0.8286
"400000 & πάνω" Δ	6	0.1714	35	1.0000
ΣΥΝΟΛΟ	35	1.0000		

Πίνακας 2.4: Πίνακας συχνοτήτων των δεδομένων της οικονομικής κατάστασης του Πίνακα 1.1.

Η στήλη των αθροιστικών και σχετικών αθροιστικών συχνοτήτων δεν έχει έννοια και δεν περιλαμβάνεται στον πίνακα συχνοτήτων ονοματικών ποιοτικών μεταβλητών, μπορεί όμως να συμπεριληφθεί στον πίνακα συχνοτήτων διατάξιμων μεταβλητών.

2.3 Γραφικές Παραστάσεις

Γραφική παράσταση είναι η γεωμετρική εικόνα ενός συνόλου δεδομένων, η μαθηματική εικόνα ενός πίνακα συχνοτήτων. Δίνουν τη

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

δυνατότητα άμεσης κατανόησης των ουσιωδών χαρακτηριστικών της κατανομής συχνοτήτων και επιτρέπουν την άμεση σύγκριση μιας κατανομής συχνοτήτων με μια άλλη. Οι γραφικές παραστάσεις αποτελούν μέρος της καθημερινής μας δραστηριότητας. Εφημερίδες, περιοδικά, επιστημονικά και διαφημιστικά έντυπα τις χρησιμοποιούν σε μεγάλη έκταση. Υπάρχουν πολλοί τύποι γραφικών παραστάσεων που κάθε ένας, με τον τρόπο του, αποκαλύπτει χρήσιμα χαρακτηριστικά του πληθυσμού. Οι πιο βασικοί που θα συζητηθούν στη συνέχεια είναι: ιστόγραμμα, πολύγωνο συχνοτήτων, αθροιστικό ιστόγραμμα, αθροιστικό πολύγωνο συχνοτήτων, ραβδόγραμμα, κυκλικό διάγραμμα και στατιστικός χάρτης. Οι τέσσερις πρώτοι εφαρμόζονται σε ποσοτικές μεταβλητές ενώ οι υπόλοιποι σε ποιοτικές.

Παρασκευή 16/03/2001.

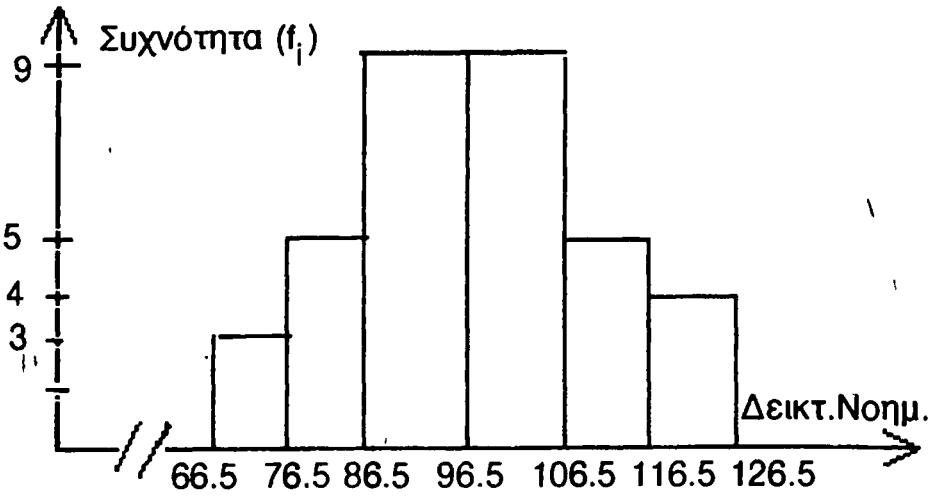
2.3.1 Γραφικές παραστάσεις ποσοτικών μεταβλητών.

2.3.1.1 Ιστόγραμμα συχνοτήτων

Συνίσταται από ένα σύνολο από συγγενή ορθογώνια παραλληλόγραμμα των οποίων το ύψος είναι ανάλογο με τη συχνότητα της ομάδας και το μήκος τους ανάλογο με το μήκος της ομάδας. Το ιστόγραμμα που αντιστοιχεί στον πίνακα συχνοτήτων 2.2 των δεδομένων του δείκτη νοημοσύνης δίνεται στην επόμενη σελίδα.

Για την κατασκευή του ιστογράμματος οι τιμές της μεταβλητής (στην περίπτωσή μας ο δείκτης νοημοσύνης) τοποθετούνται στον οριζόντιο άξονα και οι συχνότητες στον κατακόρυφο άξονα. Κατά μήκος του οριζόντιου άξονα δημιουργούνται τμήματα ίσα με τα μήκη των

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ



Σχήμα 2.1: Ιστόγραμμα συχνοτήτων των δεδομένων του δείκτη νοημοσύνης του Πίνακα 1.1.

ομάδων και στα άκρα των τμημάτων αναγράφονται τα πραγματικά όρια των ομάδων. Κατά μήκος του κατακόρυφου άξονα δημιουργούνται τμήματα ίσου μήκους που παριστούν τις συχνότητες της κατανομής που θα παρασταθεί γραφικά. Έχοντας σαν βάση το μήκος των πραγματικών ορίων των ομάδων κατασκευάζονται διαδοχικά ορθογώνια παραλληλόγραμμα των οποίων το ύψος είναι ίσο με τη συχνότητα της κάθε ομάδας. Συνήθως στον κατακόρυφο άξονα αντί των συχνοτήτων τοποθετούνται οι τιμές $(f_i / N) \cdot 100\%$ ώστε οι συγκρίσεις και τα συμπεράσματα να εκφράζονται επί τοις εκατό. Το ιστόγραμμα συχνοτήτων δίνει τη δυνατότητα να αποφανθούμε με αρκετή ακρίβεια για το ποσοστό των μαθητών με δείκτη νοημοσύνης μεγαλύτερο ή μικρότερο από κάποια τιμή. Η κατασκευή του ιστογράμματος συχνοτήτων βασίζεται στην υπόθεση ότι όλες οι δειγματικές τιμές που

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ανήκουν σε μια ομάδα κατανέμονται ομοιόμορφα μεταξύ των ορίων της. Η υπόθεση αυτή δικαιολογεί και τον χαρακτηρισμό της μέσης τιμής της ομάδας σαν την αντιπροσωπευτική τιμή των δειγματικών τιμών που ανήκουν σ' αυτή.

16/03/2001.

2.3.2 Πολύγωνο συχνοτήτων

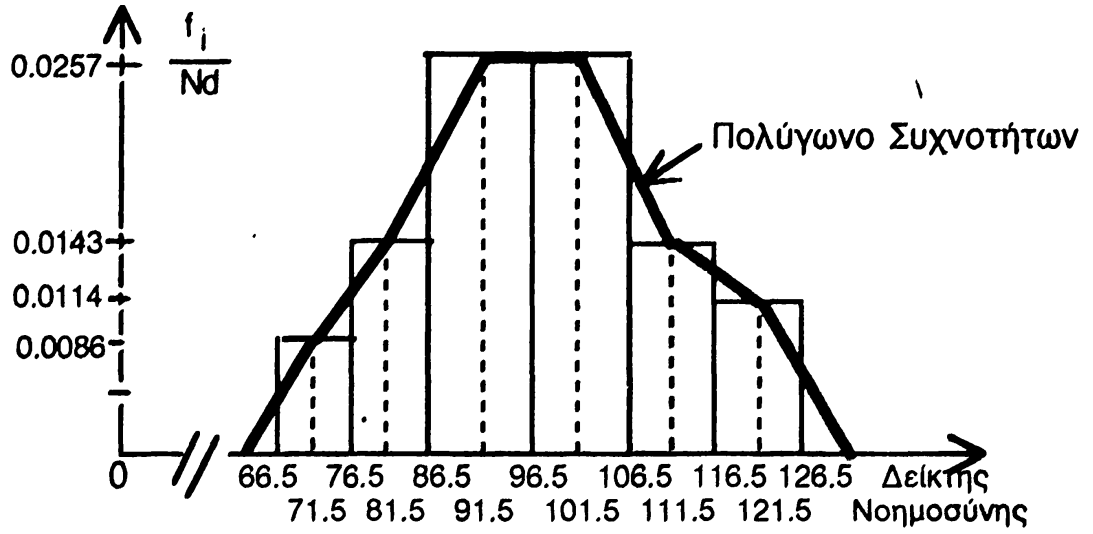
Αν τα μέσα των άνω πλευρών των ορθογωνίων παραλληλογράμμων του ιστογράμματος ενωθούν διαδοχικά με ευθύγραμμα τμήματα το πολύγωνο που θα προκύψει ονομάζεται πολύγωνο συχνοτήτων.

Το εμβαδόν του ιστογράμματος θα ισούται με το εμβαδό του πολύγωνου συχνοτήτων αν ενώσουμε και τα δύο άκρα της πολυγωνικής γραμμής με τα μέσα των δύο διαστημάτων εκ των οποίων το ένα προηγείται και το άλλο έπεται του ιστογράμματος. Επειδή το πολύγωνο συχνοτήτων δίνει μια πρώτη εικόνα της κατανομής της μεταβλητής την οποία αφορούν τα δεδομένα, στον κατακόρυφο άξονα τοποθετούνται οι τιμές f_i / Nd, $i=1,...,k$, ώστε το συνολικό εμβαδό του πολυγώνου να είναι ίσο με τη μονάδα. Στο επόμενο σχήμα δίνεται το πολύγωνο συχνοτήτων που αντιστοιχεί στον πίνακα συχνοτήτων 2.2 των δεδομένων του δείκτη νοημοσύνης των παιδιών προσχολικής ηλικίας.

Το πολύγωνο συχνοτήτων μπορεί να κατασκευαστεί και ανεξάρτητα από το ιστόγραμμα. Συνδέουμε τη μέση τιμή κάθε ομάδας με μια κατακόρυφη με το αντίστοιχο σημείο f_i / Nd, $i=1,...,k$. Αν ενώσουμε τα άνω άκρα των f_i / Nd διαδοχικά, προκύπτει το πολύγωνο

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

συχνοτήτων το οποίο είναι συνήθως ακανόνιστης μορφής. Αν αυξηθεί όμως το μήκος των ομάδων προκύπτει ομαλότερη κατασκευή.



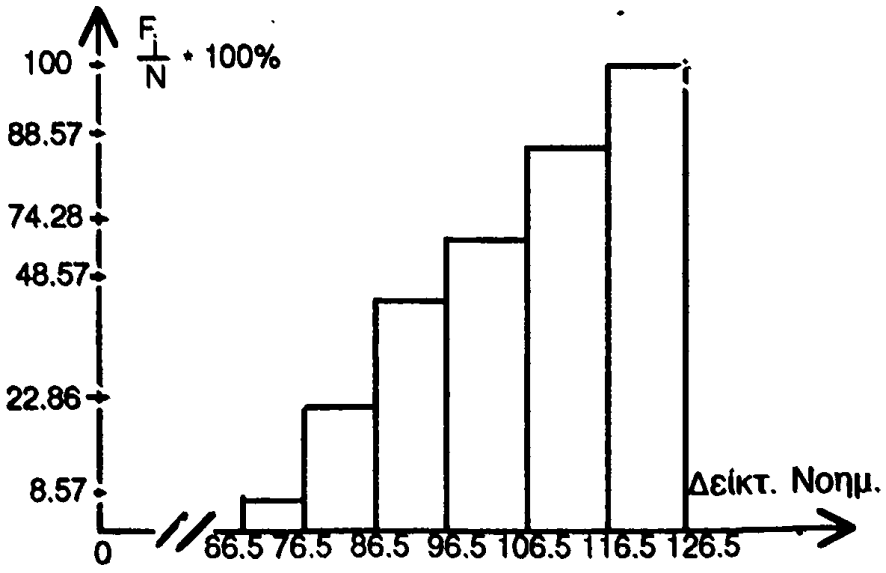
Σχήμα 2.2: Πολύγωνο συχνοτήτων των δεδομένων του δείκτη νοημοσύνης του Πίνακα 1.1.

Το ιστόγραμμα και το πολύγωνο συχνοτήτων είναι η γεωμετρική εικόνα ενός πίνακα συχνοτήτων. Ανάλογα με την περίπτωση ενδέχεται το ιστόγραμμα να παρουσιάζει περισσότερα πλεονεκτήματα από το πολύγωνο συχνοτήτων και αντίστροφα. Για παράδειγμα το πολύγωνο συχνοτήτων είναι προτιμότερο από το ιστόγραμμα όταν πρόκειται να συγκριθούν οι κατανομές δύο ή περισσότερων συνόλων δεδομένων. Η ουσιώδης όμως διαφορά μεταξύ του πολυγώνου συχνοτήτων και του ιστογράμματος εντοπίζεται στο γεγονός ότι για την κατασκευή του πολυγώνου συχνοτήτων υποθέτουμε ότι όλες οι δειγματικές τιμές που ανήκουν σε μια ομάδα συγκεντρώνονται γύρω από τη μέση της τιμή και δεν κατανέμονται ομοιόμορφα μεταξύ των ορίων της, υπόθεση στην οποία βασίζεται η κατασκευή του ιστογράμματος.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

2.3.1.3 Αθροιστικό ιστόγραμμα συχνοτήτων

Το αθροιστικό ιστόγραμμα συχνοτήτων αποτελεί τη γεωμετρική εικόνα της κατανομής των αθροιστικών συχνοτήτων και δίνει τη δυνατότητα εύρεσης του ποσοστού των περιπτώσεων, του ποσοστού των τιμών του δείγματος, που ειναι μεγαλύτερες ή μικρότερες απο μια δοθείσα. Όπως και για την κατασκευή του ιστογράμματος συχνοτήτων στον οριζόντιο άξονα σημειώνονται τα πραγματικά όρια των ομάδων ενώ στον κατακόρυφο άξονα συνίσταται η τοποθέτηση των τιμών $(F_i / N) \cdot 100\%$, $i = 1,2,...,k$ ώστε οι συγκρίσεις να εκφράζονται επί τοις εκατό. Στο σχήμα που ακολουθεί δίνεται το αθροιστικό ιστόγραμμα συχνοτήτων που αντιστοιχεί στον πίνακα συχνοτήτων 2.2 των δεδομένων του δείκτη νοημοσύνης των παιδιών προσχολικής ηλικίας.



Σχήμα 2.3: Αθροιστικό Ιστόγραμμα Συχνοτήτων των δεδομένων του δείκτη νοημοσύνης του Πίνακα 1.1.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

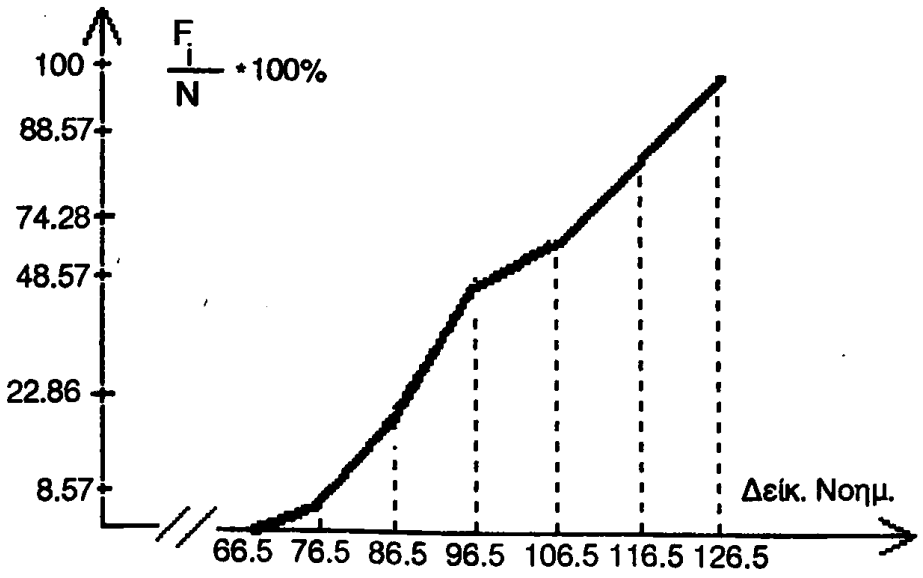
Η προηγούμενη γραφική παράσταση επιτρέπει την εξαγωγή συμπερασμάτων της μορφής: Το 22.8% των μαθητών έχουν δείκτη νοημοσύνης μικρότερο από 86.5, το 11.43% των μαθητών έχει δείκτη νοημοσύνης μεγαλύτερο από 116.5 ή το (88.57-22.86) % = 65.71% έχει δείκτη νοημοσύνης μεταξύ 86.5 και 116.5.

2.3.1.4 Αθροιστικό πολύγωνο συχνοτήτων

Κατασκευάζεται ανάλογα με το πολύγωνο συχνοτήτων με μια και μόνο διαφορά. Για την εύρεση του πολυγώνου συχνοτήτων κατασκευάζονταν κάθετος στην μέση τιμή κάθε ομάδας ύψους ίσου προς την αντίστοιχη τιμή f_i / Nd, $i = 1,...,k$. Για το αθροιστικό πολύγωνο συχνοτήτων κατασκευάζεται κάθετη, ύψους ίσου προς την τιμή $(F_i / N) * 100\%$, $i = 1,...,k$, στο πραγματικό άνω όριο της ομάδας και όχι στη μέση της τιμής. Ο λόγος γίνεται περισσότερο κατανοητός από τον τρόπο δημιουργίας της στήλης των αθροιστικών συχνοτήτων ενός πίνακα συχνοτήτων. Το αθροιστικό πολύγωνο συχνοτήτων προκύπτει από την ένωση με διαδοχικά ευθύγραμμα τμήματα των διατεταγμένων ζευγών των τιμών του οριζόντιου άξονα και των αντίστοιχων τιμών του κατακόρυφου.

Η επόμενη γραφική παράσταση δίνει και πάλι τη δυνατότητα να αποφανθούμε με αρκετή ακρίβεια για το ποσοστό των μαθητών με δείκτη νοημοσύνης μεγαλύτερο ή μικρότερο από κάποια τιμή. Έτσι το 8.57% των μαθητών έχουν δείκτη νοημοσύνης μικρότερο από 76.5, το 25.72% των μαθητών έχει δείκτη νοημοσύνης μεγαλύτερο από 106.5 και το 51.42% έχει δείκτη νοημοσύνης μεταξύ 86.5 και 106.5 μονάδων.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ



Σχήμα 2.4: Αθροιστικό πολύγωνο συχνοτήτων των δεδομένων του δείκτη νοημοσύνης του Πίνακα 1.1.

Παράδειγμα 2.1: Οι βαθμοί (διατεταγμένοι σε αύξουσα τάξη μεγέθους) 70 φοιτητών στις εξετάσεις του μαθήματος αυτού ήταν :

30, 32, 34, 37, 38, 40, 40, 40, 41, 41, 42, 45, 47, 50, 50, 50, 50, 50, 51, 52, 53, 57, 57, 58, 59, 60, 60, 60, 60, 62, 62, 63, 64, 65, 65, 65, 68, 69, 69, 70, 71, 71, 71, 73, 73, 73, 75, 75, 75, 77, 78, 79, 80, 80, 80, 80, 83, 84, 88, 88, 88, 89, 89, 90, 90, 90, 90, 95, 96, 99.

α) Να κατασκευαστεί ο ομαδοποιημένος πίνακας συχνοτήτων.

β) Το ιστόγραμμα συχνοτήτων.

γ) Να δοθούν κατά προσέγγιση τα ποσοστά των φοιτητών με βαθμό κάτω του πέντε, μεταξύ έξι και εννέα και ανω του οκτώ.

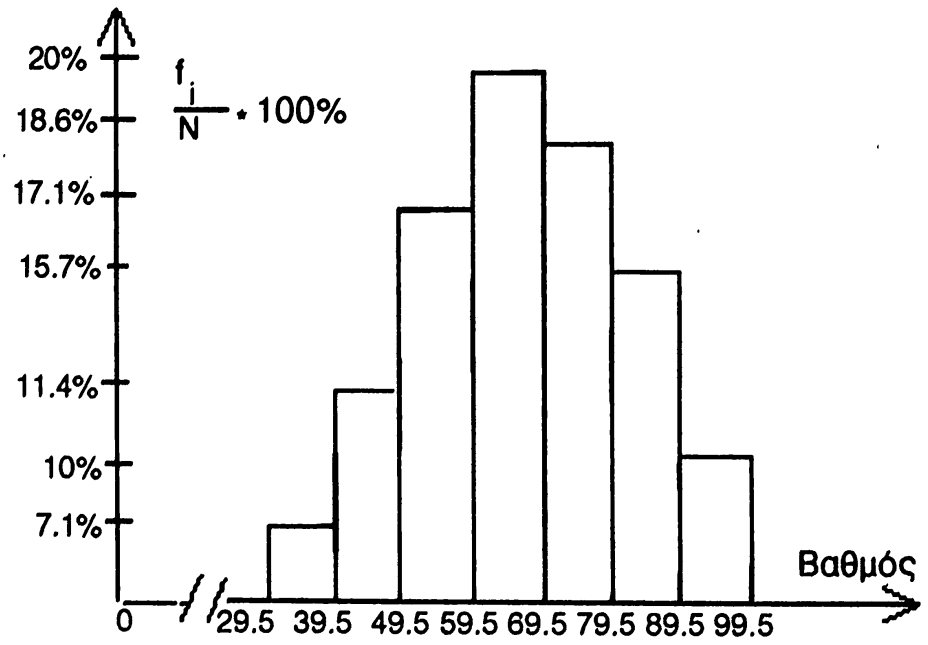
ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

α) Ο Ομαδοποιημένος πίνακας συχνοτήτων είναι:

Α\Α	ΠΡΑΓΜ. ΟΡΙΑ	m_i	f_i	f_i / N	F_i	F_i / N
1	[29.5 , 39.5]	34.5	5	0.071	5	0.071
2	[39.5 , 49.5]	44.5	8	0.114	13	0.186
3	[49.5 , 59.5]	54.5	12	0.171	25	0.357
4	[59.5 , 69.5]	64.5	14	0.2	39	0.557
5	[69.5 , 79.5]	74.5	13	0.186	52	0.743
6	[79.5 , 89.5]	84.5	11	0.157	63	0.9
7	[89.5 , 99.5]	94.5	7	0.1	70	1.000

Έχουν χρησιμοποιηθεί k=7 ομάδες με μήκος κάθε ομάδας d = 10.

β) Το ιστόγραμμα συχνοτήτων είναι:



ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

γ) Ποσοστό με βαθμό ≤ 5 περίπου: 19%.

Ποσοστό με βαθμό μεταξύ 6 και 9 περίπου: 54%

Ποσοστό με βαθμό ≥ 8 περίπου: 26%.

Παράδειγμα 2.2: Αν οι βαθμοί 70 φοιτητριών στις εξετάσεις του παρόντος μαθήματος είναι:

30, 40, 40, 40, 42, 42, 43, 48, 48, 53, 53, 53, 54, 54, 54, 54, 55, 55, 56, 58, 58, 58, 60, 60, 60, 63, 63, 63, 63, 65, 65, 65, 65, 66, 66, 66, 66, 68, 68, 68, 69, 69, 69, 70, 70, 70, 70, 70, 72, 73, 73, 73, 76, 76, 77, 77, 77, 79, 79, 79, 80, 80, 80, 80, 85, 85, 88, 88, 95, 99.

α) Να κατασκευαστεί ο ομαδοποιημένος πίνακας συχνοτήτων.

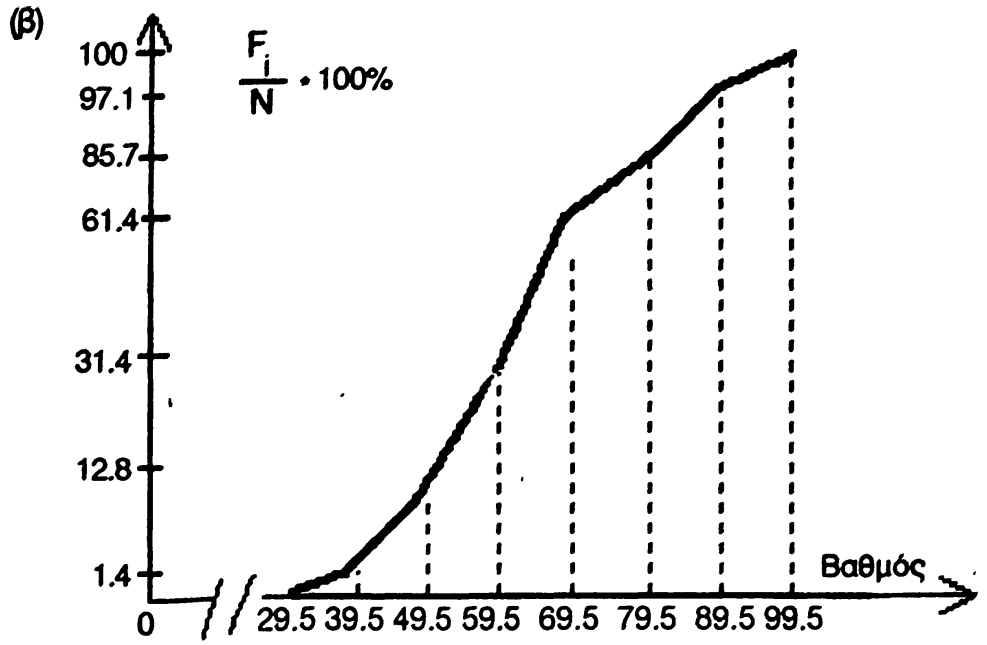
β) Το αθροιστικό πολύγωνο συχνοτήτων

γ) Τα ποσοστά των φοιτητριών με βαθμό κάτω του πέντε, μεταξύ έξι και εννιά και άνω του οκτώ.

α) Ο ομαδοποιημένος πίνακας συχνοτήτων είναι

Α/Α	ΠΡΑΓΜ.ΟΡΙΑ	m_i	f_i	f_i / N	F_i	F_i / N
1	[29.5 , 39.5]	34.5	1	0.014	1	0.014
2	[39.5 , 49.5]	44.5	8	0.114	9	0.128
3	[49.5 , 59.5]	54.5	13	0.186	22	0.314
4	[59.5 , 69.5]	64.5	21	0.3	43	0.614
5	[69.5 , 79.5]	74.5	17	0.243	60	0.857
6	[79.5 , 89.5]	84.5	8	0.114	68	0.971
7	[89.5 , 99.5]	94.5	2	0.028	70	1.000

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ



γ) Ποσοστό με βαθμό ≤ 5 περίπου: 13%

Ποσοστό με βαθμό μεταξύ έξι και εννέα περίπου: (97-31.5)% = 65.5%.

Ποσοστό με βαθμό ≥ 8 περίπου (100-86)% = 14%.

2.3.2 Γραφικές παραστάσεις ποιοτικών μεταβλητών

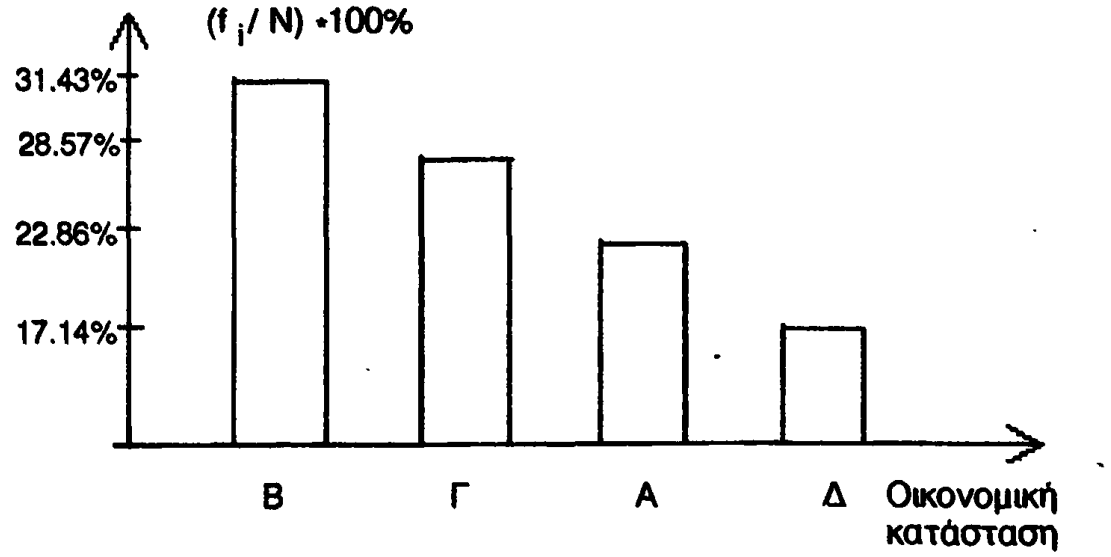
2.3.2.1 Ραβδόγραμμα

16/03/2001

Για την κατασκευή του χρησιμοποιείται και πάλι ένα σύστημα ορθογωνίων αξόνων. Στον οριζόντιο άξονα σημειώνονται οι κατηγορίες στις οποίες τα μέλη του πληθυσμού κατατάσσονται ενώ στον κατακόρυφο άξονα οι αντίστοιχες συχνότητες, ή οι αντίστοιχες σχετικές συχνότητες πολλαπλασιασμένες επί εκατό ώστε τα αποτελέσματα και οι συγκρίσεις να εκφράζονται επί τοις εκατό (%).

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Το ραβδόγραμμα κατασκευάζεται με τη βοήθεια ορθογωνίων στηλών με ίσες βάσεις, που αναφέρονται σ' όλες τις δυνατές κατηγορίες, και ύψη ίσα με την αντίστοιχη συχνότητα. Για καλύτερη παρουσίαση αλλά και για μια συγκριτική αναπαράσταση οι ορθογώνιες στήλες τοποθετούνται σε ίση απόσταση η μία από την άλλη και σε φθίνουσα τάξη μεγέθους. Το ραβδόγραμμα των δεδομένων της οικονομικής κατάστασης του Πίνακα 1.1 που αντιστοιχεί στον πίνακα συχνοτήτων 2.4 είναι:



Σχήμα 2.5: Ραβδόγραμμα των δεδομένων της οικονομικής κατάστασης του Πίνακα 1.1.

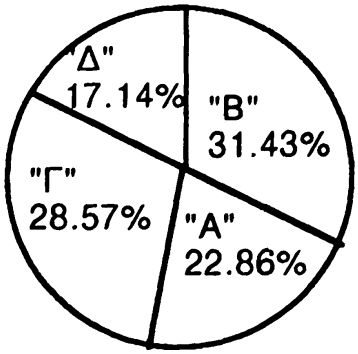
16/03/2001

2.3.2.2 Κυκλικό διάγραμμα

Όπως ο όρος υποδηλώνει, το κυκλικό διάγραμμα είναι ένας κυκλικός δίσκος χωρισμένος σε τομείς, όσες και οι κατηγορίες στις οποίες τα μέλη του πληθυσμού κατατάσσονται. Το εμβαδόν κάθε τομέα απεικονίζει το ποσοστό των ατόμων του πληθυσμού που ανήκουν στην αντίστοιχη κατηγορία. Το κυκλικό διάγραμμα των δεδομένων της οικονομικής

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

κατάστασης του Πίνακα 1.1 που αντιστοιχεί στον πίνακα συχνοτήτων 2.4 είναι:



Σχήμα 2.6: Κυκλικό διάγραμμα των δεδομένων της οικονομικής κατάστασης του Πίνακα 1.1.

Παρά το γεγονός ότι το κυκλικό διάγραμμα είναι η πιο διαδεδομένη γραφική παράσταση ποιοτικών δεδομένων και χρησιμοποιείται ευρύτατα από τα μέσα μαζικής ενημέρωσης παρουσιάζει δύο αδυναμίες σε σχέση με το ραβδόγραμμα. Πρώτον γίνεται πάρα πολύ σύνθετο όταν χωριστεί σε πολλούς τομείς με αποτέλεσμα να είναι δυσδιάκριτα και δυσκολοσύγκριτα τα ποσοστά των κατηγοριών και δεύτερον οι συγκρίσεις των εμβαδών του κυκλικού διαγράμματος είναι πολύ πιο δύσκολες από τις γραμμικές συγκρίσεις του ραβδογράμματος.

2.3.2.3 Στατιστικός χάρτης

Χρησιμοποιείται για να παραστήσει τη γεωγραφική κατανομή διαφόρων φαινομένων. Παράδειγμα στατιστικού χάρτη αποτελεί η τηλεοπτική μετάδοση εκλογικών αποτελεσμάτων, όπου παρουσιάζεται στην οθόνη ο χάρτης της Ελλάδος με τους διάφορους νομούς χρωματισμένους με το χρώμα του πρώτου κόμματος στο νομό.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΚΕΦΑΛΑΙΟ 3

16/03/2001

ΑΡΙΘΜΗΤΙΚΑ ΜΕΓΕΘΗ

3.1 Εισαγωγικά

Οι πίνακες συχνοτήτων και οι γραφικές τους παραστάσεις, που εκτέθηκαν στο προηγούμενο κεφάλαιο, σκοπό έχουν τη συνοπτική παρουσίαση των δεδομένων και την εξαγωγή κάποιων συμπερασμάτων για τον πληθυσμό απ' τον οποίο τα δεδομένα προέρχονται. Η κατασκευή τους συνίσταται στην υποκατάσταση των μετρήσεων από τις συχνότητες εμφάνισής τους σε κατάλληλα ορισμένα διαδοχικά και ξένα μεταξύ τους διαστήματα.

Μια πιο δραστική σύμπτυξη των δεδομένων επιτυγχάνεται με τα αριθμητικά μεγέθη, ποσότητες οι οποίες υπολογίζονται μέσω των δεδομένων και φιλοδοξούν να τα περιγράψουν στο σύνολό τους. Ορίζονται για ποσοτικά δεδομένα και η δημιουργία τους συνίσταται στην υποκατάσταση των μετρήσεων από δύο αριθμητικές τιμές. Η πρώτη ενδεικτική της θέσης, του σημείου γύρω από το οποίο το σύνολο των δεδομένων συγκεντρώνεται και η δεύτερη ενδεικτική του τρόπου με τον οποίο τα δεδομένα κατανέμονται γύρω από το σημείο αυτό. Ανάλογα με το σκοπό τον οποίο επιτελούν, τα αριθμητικά μεγέθη διακρίνονται στα μέτρα θέσεως και τα μέτρα διασποράς.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

3.2 Μέτρα θέσεως

Ένα μέτρο θέσεως είναι μια αριθμητική τιμή συνήθως πολύ κοντά στο σημείο με τη μεγαλύτερη συγκέντρωση μετρήσεων και έτσι κατά κάποιο τρόπο θεωρείται ότι αντιπροσωπεύει ή αναπαριστάνει το σύνολο των μετρήσεων. Τα μέτρα θέσεως είναι η μέση τιμή, η διάμεσος και η κορυφή.

3.2.1 Μέση τιμή ή μέσος όρος

Υποθέτουμε ότι τα δεδομένα αποτελούνται από N μετρήσεις της ίδιας μεταβλητής x και συμβολίζονται με $x_1, x_2, \ldots, x_N$. Ο υπολογισμός της μέσης τιμής (mean) των δεδομένων, που συμβολίζεται με $\bar{x}$, επιτυγχάνεται με τους ακόλουθους τρόπους:

α. **Απ' τα δεδομένα**

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{x_1+x_2+ \ldots +x_N}{N}.$$

β. **Από το διατεταγμένο πίνακα συχνοτήτων**

$$\bar{x} = \frac{\sum_{i=1}^{v} x_i f_i}{N} = \frac{x_1f_1+x_2f_2+ \ldots + x_v f_v}{N},$$

με v το πλήθος των διακεκριμένων x_i και f_i τις αντίστοιχες συχνότητες, i = 1, ... ,v.

γ. **Απ' τον ομαδοποιημένο πίνακα συχνοτήτων**

$$\bar{x} = \frac{\sum_{i=1}^{k} m_i f_i}{N} = \frac{m_1f_1+m_2f_2+ \ldots + m_kf_k}{N},$$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

με k το πλήθος των ομάδων, m_i τις μέσες τιμές των ομάδων και f_i τις αντίστοιχες συχνότητες, i = 1,..., k.

Οι δύο πρώτοι τρόποι μπορούν να χρησιμοποιηθούν όταν το πλήθος N των δεδομένων είναι μικρό. Συνήθως χρησιμοποιείται ο τρίτος τρόπος γιατί σε κάθε πρακτικό πρόβλημα το πλήθος N των μετρήσεων είναι πολύ μεγάλο και έτσι ο ομαδοποιημένος πίνακας συχνοτήτων είναι ένα αναγκαίο πρώτο βήμα οποιασδήποτε στατιστικής ανάλυσης.

Παράδειγμα 3.2.1: Να υπολογιστεί και με τους τρεις τρόπους η μέση τιμή των δεδομένων του δείκτη νοημοσύνης των N= 35 παιδιών προσχολικής ηλικίας του Πίνακα 1.1.

Από τους προηγούμενους τύπους έχουμε

$$\text{α)}\quad \bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{111+90+90+90+104+\ \dots\ +75+96+81}{35} = 96.94.$$

β) Από τον Πίνακα 2.1

$$\bar{x} = \frac{\sum_{i=1}^{v} x_i\, f_i}{N} = \frac{67*1+72*1+75*1+\ \dots\ +122*1+126*1}{35} = 96.94.$$

γ) Από τον Πίνακα 2.2:

$$\bar{x} = \frac{\sum_{i=1}^{k} m_i\, f_i}{N} = \frac{71.5*3+81.5*5+\ \dots\ +121.5*4}{35} = 97.21.$$

Η μέση τιμή που προκύπτει από τον α και β τρόπο είναι διαφορετική από εκείνη του γ τρόπου. Το γεγονός αυτό οφείλεται στην ομαδοποίηση των μετρήσεων, στην υποκατάσταση των μετρήσεων

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

από τις μέσες τιμές των ομάδων και είναι το κόστος της δραστικής σύμπτυξης μεγάλου αριθμού δεδομένων. Η μέση τιμή λοιπόν των δεδομένων του δείκτη νοημοσύνης των παιδιών προσχολικής ηλικίας είναι περίπου 97 γεγονός που σημαίνει ότι οι περισσότερες μετρήσεις συγκεντρώνονται γύρω από την τιμή 97 ή ο δείκτης νοημοσύνης των περισσοτέρων παιδιών είναι κοντά στο 97. Στο ίδιο συμπέρασμα καταλήγουμε και από την προσεκτική παρατήρηση του Πίνακα 2.1 ή του Σχήματος 2.2.

Παράδειγμα 3.2.2: Να υπολογιστεί απ' ευθείας από τα δεδομένα και από τον ομαδοποιημένο πίνακα συχνοτήτων ή μέση τιμή α) των δεδομένων του Παραδείγματος 2.1 και β) των δεδομένων του Παραδείγματος 2.2.

$$\text{α)}\quad \bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{30+32+34+\ \dots\ +96+99}{70} = \frac{4576}{70} = 65.37,$$

ή

$$\bar{x} = \frac{\sum_{i=1}^{k} m_i f_i}{N} = \frac{34.5\cdot 5+\ \dots\ +94.5\cdot 7}{70} = \frac{4645}{70} = 66.36.$$

$$\text{β)}\quad \bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{30+40+40+\ \dots\ +95+99}{70} = \frac{4576}{70} = 65.37,$$

ή

$$\bar{x} = \frac{\sum_{i=1}^{k} m_i f_i}{N} = \frac{34.5\cdot 1+\ \dots\ +94.5\cdot 2}{70} = \frac{4585}{70} = 65.5.$$

3.2.2 Διάμεσος

Ένα άλλο χρήσιμο μέτρο θέσεως είναι η διάμεσος (median). Είναι το σημείο που χωρίζει τα δεδομένα σε δύο ίσα μέρη ώστε το πλήθος

των μετρήσεων που βρίσκονται αριστερά της διαμέσου να είναι ίσο με το πλήθος των μετρήσεων που βρίσκονται δεξιά από τη διάμεσο. Συμβολίζεται συνήθως με M και ο υπολογισμός της επιτυγχάνεται με τους ακόλουθους τρόπους.

α. Απ' τα δεδομένα

Απαραίτητη προϋπόθεση είναι οι μετρήσεις να είναι διατεταγμένες σε αύξουσα (ή φθίνουσα) τάξη μεγέθους και να μην υπάρχουν μετρήσεις οι οποίες εμφανίζονται περισσότερο από μία φορά.

i) Αν το πλήθος N των δεδομένων είναι περιττός αριθμός (π.χ. 1,3,5, κ.ο.κ.) η διάμεσος M είναι η τιμή της $\frac{N+1}{2}$ θέσης.

π.χ. για τα δεδομένα: 2,5,7,8,11,12,14 το N=7 είναι περιττός αριθμός. Η διάμεσος είναι η τιμή που βρίσκεται στην $\frac{N+1}{2} = \frac{7+1}{2} = 4$ θέση δηλ. M= 8.

ii) Αν το πλήθος N των δεδομένων είναι άρτιος αριθμός (π.χ. 2,4,6 κ.ο.κ.) η διάμεσος είναι το ημιάθροισμα των τιμών της $\frac{N}{2}$ και $\frac{N}{2} + 1$ θέσης.

π.χ. για τα δεδομένα: 2,4,5,7,9,10,11,13 το N= 8 είναι άρτιος. Οι τιμές της $\frac{N}{2} = \frac{8}{2} = 4$ και $\frac{N}{2} + 1 = \frac{8}{2} + 1 = 5$ θέσης είναι αντίστοιχα οι 7 και 9. Έτσι η διάμεσος M είναι το ημιάθροισμά τους δηλ. $M = \frac{7+9}{2} = \frac{16}{2} = 8$. Όπως φαίνεται από το παράδειγμα αυτό η διάμεσος δεν είναι πάντοτε ένα σημείο του συνόλου των δεδομένων.

β. Απ' τον ομαδοποιημένο πίνακα συχνοτήτων

Το πρώτο βήμα για τον υπολογισμό της διαμέσου απ' τον ομαδοποιημένο πίνακα συχνοτήτων είναι ο εντοπισμός της ομάδας στην

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

οποία ανήκει. Επειδή η διάμεσος είναι το σημείο που χωρίζει τα δεδομένα σε δύο ίσα μέρη και για την κατασκευή του ομαδοποιημένου πίνακα συχνοτήτων τα δεδομένα έχουν ήδη διαταχθεί σε αύξουσα (ή φθίνουσα) τάξη μεγέθους, προσδοκούμε η διάμεσος να είναι μια τιμή κοντά σ' εκείνη της $\frac{N}{2}$ θέσης. Έτσι η διάμεσος θα βρίσκεται στην ομάδα με αθροιστική συχνότητα αμέσως μεγαλύτερη από $\frac{N}{2}$ ενώ η αθροιστική συχνότητα της προηγούμενης ομάδας θα πρέπει να είναι μικρότερη από $\frac{N}{2}$. Δηλαδή αν

$$F_{i-1} \leq \frac{N}{2} \leq F_i ,$$

για κάποιο $i = 1,2, \ldots ,k$ με $F_0 = 0$ η διάμεσος βρίσκεται στην i ομάδα. Αφού εντοπιστεί η ομάδα η διάμεσος υπολογίζεται από τον ακόλουθο τύπο:

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

όπου L_i το κάτω όριο της ομάδας που περιέχει τη διάμεσο, d το μήκος των ομάδων, F_{i-1} η αθροιστική συχνότητα της προηγούμενης ομάδας και f_i η συχνότητα της ομάδας που περιέχει τη διάμεσο.

Παράδειγμα 3.2.3: Να υπολογιστεί από τον ομαδοποιημένο πίνακα συχνοτήτων, η διάμεσος των δεδομένων: α) του δείκτη νοημοσύνης του Πίνακα 2.2 β) του Παραδείγματος 2.1. γ) του Παραδείγματος 2.2.

α) Από τον Πίνακα 2.2 έχουμε :

$$F_3 = 17 < \frac{N}{2} = 17.5 < 26 = F_4 .$$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Έτσι η διάμεσος βρίσκεται στην $i = 4$ ομάδα. Είναι $L_4 = 96.5$, $d = 10$, $F_3 = 17$ και $f_4 = 9$.

Επομένως

$$M = 96.5+10 \left(\frac{17.5-17}{9} \right) = 96.5+10 \cdot 0.0556 = 97.056.$$

β) Απ' τον πίνακα του Παραδείγματος 2.1 προκύπτει

$$F_3 = 25 < \frac{N}{2} = 35 < 39 = F_4 .$$

Η διάμεσος βρίσκεται στην $i = 4$ ομάδα και είναι

$$M = L_4 + d \left(\frac{N/2-F_3}{f_4} \right) = 59.5+10 \left(\frac{35-25}{14} \right) = 66.64.$$

γ) Ανάλογα απ' τον πίνακα του Παραδείγματος 2.2 η διάμεσος βρίσκεται στην $i = 4$ ομάδα και είναι:

$$M = L_4 + d \left(\frac{N/2-F_3}{f_4} \right) = 59.5+10 \left(\frac{35-22}{21} \right) = 65.69.$$

3.2.3 Κορυφή ή επικρατούσα τιμή 16/03/2002

Ως κορυφή ή επικρατούσα τιμή (mode) των δεδομένων ορίζεται η τιμή με τη μεγαλύτερη συχνότητα και συμβολίζεται συνήθως με K. Για τα δεδομένα του δείκτη νοημοσύνης του Πίνακα 1.1 η $K = 90$, ενώ για τα δεδομένα των Παραδειγμάτων 2.1 και 2.2 οι κορυφές τους αντίστοιχα είναι $K = 50$ και $K = 70$.

Αν όλες οι μετρήσεις εμφανίζονται με την ίδια συχνότητα δεν υπάρχει κορυφή ενώ αν δύο τιμές έχουν την ίδια συχνότητα, μεγαλύτερη από τις υπόλοιπες, τότε υπάρχουν δύο κορυφές.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Χρησιμοποιώντας τον ομαδοποιημένο πίνακα συχνοτήτων, ως κορυφή λαμβάνεται η μέση τιμή της ομαδας με τη μεγαλύτερη συχνότητα. Για τα δεδομένα του δείκτη νοημοσύνης από τον Πίνακα 2.1 οι κορυφές είναι Κ = 91.5 και 101.5, ενώ από τους πίνακες των Παραδειγμάτων 2.1 και 2.2 οι κορυφές είναι 64.5 και 64.5 αντίστοιχα.

Η κορυφή είναι το μόνο μέτρο θέσεως που εφαρμόζεται και σε ποιοτικά δεδομένα. Η κορυφή ενός συνόλου ποιοτικών δεδομένων είναι η κατηγορία, ο χαρακτηρισμός με τη μεγαλύτερη συχνότητα. Για τα δεδομένα της οικονομικής κατάστασης του Πίνακα 1.1 η κατηγορία Β, των οικογενειών με εισόδημα μεταξύ 150.000 και 300.000 δρχ., έχει την μεγαλύτερη συχνότητα ίση με 11 και αποτελεί την κορυφή των δεδομένων.

3.2.4 Σύγκριση των μέτρων θέσεως

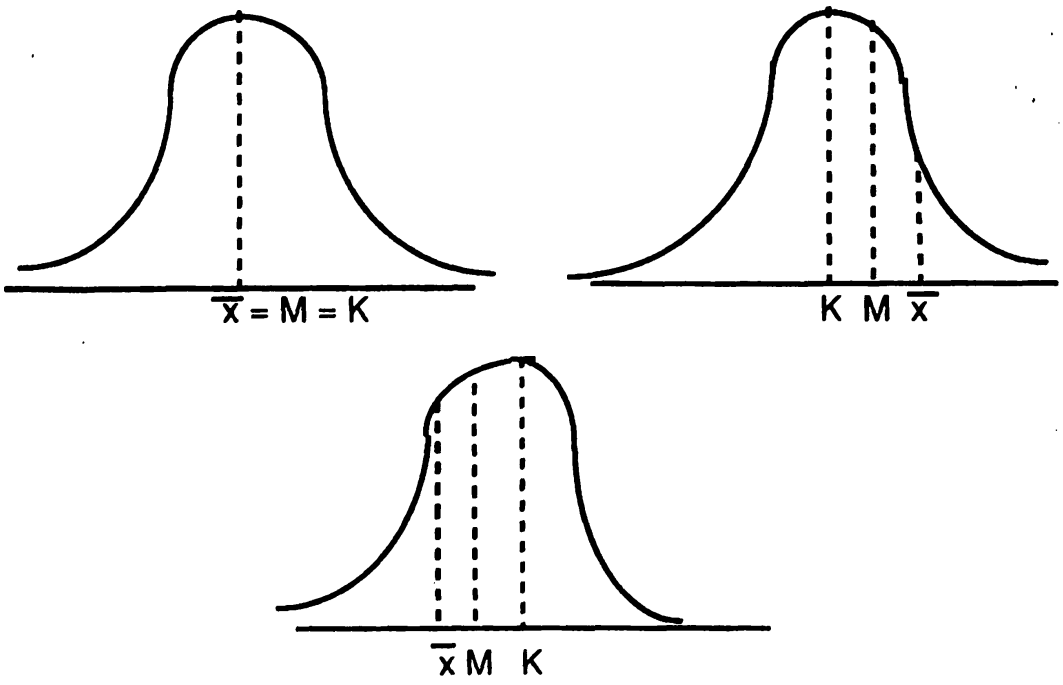
Τα μέτρα θέσης χρησιμοποιούνται για την περιγραφή της θέσης, ή τη σύγκριση των θέσεων δύο ή περισσότερων συνόλων αριθμητικών δεδομένων. Η σύγκριση γίνεται με τη χρησιμοποίηση κάθε φορά του ίδιου μέτρου θέσεως. Το ερώτημα που γεννιέται αφορά την καταλληλότερη επιλογή ενός μέτρου θέσεως. Παρά το γεγονός ότι δεν υπάρχουν σαφείς κανόνες για τη χρησιμοποίηση του ενός ή του άλλου μέτρου, η μέση τιμή προτιμάται συνήθως από τη διάμεσο και την κορυφή.

Ο εύκολος υπολογισμός της, που στηρίζεται μάλιστα στο σύνολο των δεδομένων, και οι καλές στατιστικές ιδιότητες που ικανοποιεί συνηγορούν σ' αυτή την προτίμηση. Η διάμεσος χρησιμοποιείται όταν στο σύνολο των δεδομένων υπάρχουν ακραίες τιμές, δηλαδή τιμές που είναι κατά πολύ μεγαλύτερες ή μικρότερες από τις υπόλοιπες. Τέτοιες μετρήσεις επηρεάζουν κατά πολύ τη μέση τιμή. Θεωρούμε για παράδειγμα τα δεδομένα: 2,4,5,6,8,9,50. Η τιμή 50

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

απέχει πολύ από τις προηγούμενες και επηρεάζει τη μέση τιμή. Πράγματι η μέση τιμή των μετρήσεων αυτών, που είναι ίση με 12, είναι μεγαλύτερη από τις 6 μετρήσεις του παραπάνω συνόλου. Η διάμεσος του είναι ίση με 6 και στην περίπτωση αυτή χαρακτηρίζει καλύτερα το σύνολο των μετρήσεων, περιγράφει καλύτερα την τάση των δεδομένων.

Μια σύγκριση των μέτρων θέσεως με την κατανομή συχνοτήτων (προσέγγισή της αποτελεί το πολύγωνο συχνοτήτων) δίνεται στο σχήμα που ακολουθεί. Αν η κατανομή είναι συμμετρική τότε η μέση τιμή, η διάμεσος και η κορυφή συμπίπτουν. Αν η κατανομή είναι λοξή προς τα δεξιά ή θετικώς λοξή τότε $K < M < \bar{x}$. Αν η κατανομή είναι λοξή προς τα αριστερά ή αρνητικώς λοξή τότε $\bar{x} < M < K$.



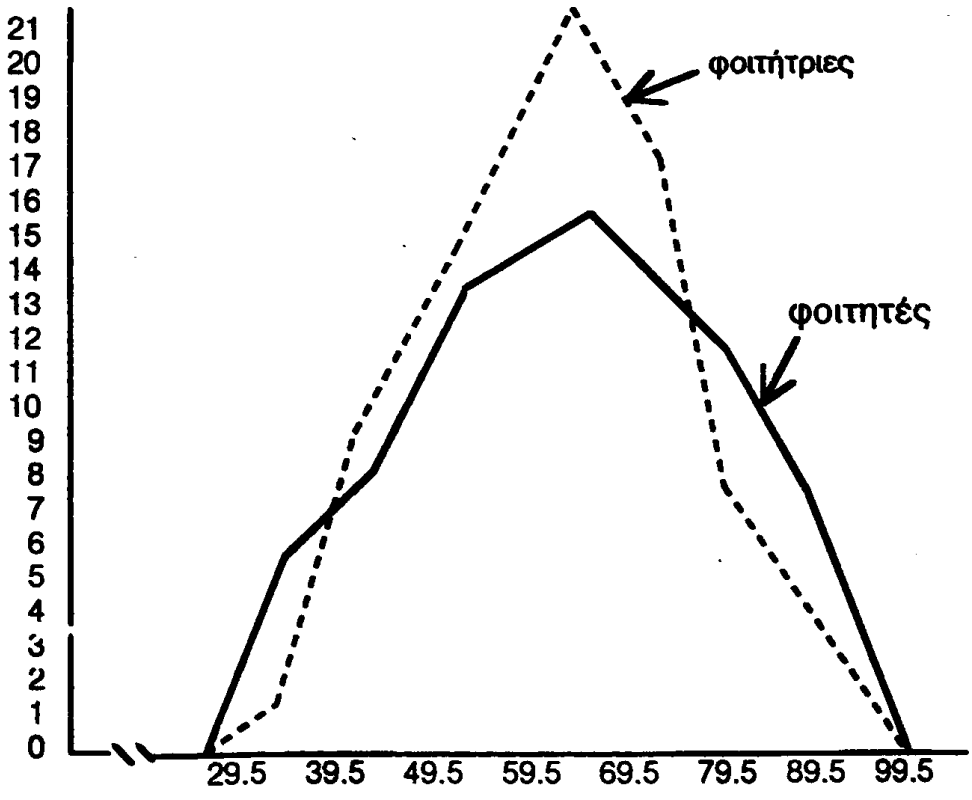
Σχήμα 3.1: Σύγκριση μέτρων θέσης και κατανομής συχνοτήτων.

3.3 Μέτρα διασποράς

Τα πολύγωνα συχνοτήτων των βαθμών των 70 φοιτητών και φοιτητριών των Παραδειγμάτων 2.1 και 2.2 είναι στο Σχήμα 3.2.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Η μέση τιμή, η διάμεσος και η κορυφή για τους βαθμούς των φοιτητών είναι 66.36, 66.64 και 64.5 αντίστοιχα ενώ για τους βαθμούς των φοιτητριών 65.5, 65.69 και 64.5. Από το παράδειγμα αυτό προκύπτει ότι για δύο σύνολα δεδομένων, παρά το γεγονός ότι τα μέτρα θέσης είναι σχεδόν τα ίδια, οι γραφικές τους παραστάσεις ή οι κατανομές συχνοτήτων τους είναι διαφορετικές. Η διαφορά έγκειται στο γεγονός ότι η μια κατανομή είναι περισσότερο σπαρμένη,απλωμένη γύρω από τη μέση τιμή απ' ότι η άλλη. Έτσι για να περιγραφεί πλήρως ένα σύνολο μετρήσεων, δεν αρκεί μόνο η γνώση των μέτρων θέσης αλλά και η γνώση του τρόπου με τον οποίο οι μετρήσεις αραδειάζονται, κατανέμονται γύρω από τη μέση τιμή.



Σχήμα 3.2: Πολύγωνα συχνοτήτων βαθμών

Ακριβώς αυτόν τον τρόπο με τον οποίο τα δεδομένα κατανέμονται γύρω

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

από τη μέση τιμή εκφράζουν τα μέτρα διασποράς που είναι: το εύρος, η διακύμανση, η τυπική απόκλιση, ο συντελεστής μεταβλητότητας και τα εκατοστιαία σημεία.

3.3.1 Εύρος

Συμβολίζεται με R και ορίζεται:

$$R = \max x_i - \min x_i$$

όπου $\max x_i$ και $\min x_i$ παριστούν αντίστοιχα την μέγιστη και ελάχιστη τιμή του συνόλου των δεδομένων. Για τα δεδομένα του δείκτη νοημοσύνης των 35 παιδιών προσχολικής ηλικίας του Πίνακα 1.1 το εύρος εύκολα υπολογίζεται: R = 126-67= 59 ενώ για τα δεδομένα των παραδειγμάτων 2.1. και 2.2 είναι: R= 99-30= 69.

Το εύρος είναι ατελές μέτρο διασποράς αφού βασίζεται μόνο στις ακραίες τιμές των δεδομένων.

3.3.2 Διακύμανση

Το σημαντικότερο ίσως μέτρο που εκφράζει τη διασπορά ή τη μεταβλητότητα ενός συνόλου αριθμητικών δεδομένων από τη μέση τους τιμή, είναι η διασπορά ή διακύμανση των μετρήσεων, και συμβολίζεται με S^2. Ο υπολογισμός της βασίζεται στις μετρήσεις: x_1, x_2, ... , x_N και επιτυγχάνεται, όπως και ο υπολογισμός της μέσης τιμής, α) άμεσα από τα δεδομένα, β) απ' το διατεταγμένο πίνακα συχνοτήτων, γ) απ' τον ομαδοποιημένο πίνακα συχνοτήτων. Οι τρεις αυτοί τρόποι υπολογισμού της διακύμανσης S^2 των δεδομένων: 1, 2, 2, 2, 3, 4, 4, 4, 5 περιγράφονται στη συνέχεια.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

3.3.2.1 Υπολογισμός S^2 απ' τα δεδομένα

Για τον υπολογισμό της διακύμανσης S^2 άμεσα από τα δεδομένα χρησιμοποιείται ο τύπος:

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 \qquad (3.3.1)$$

ή η ισοδύναμη μορφή του, που είναι γνωστή και ως τύπος μηχανής και δίνεται από τη σχέση:

$$S^2 = \frac{1}{N-1} \left[\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N} \right] \qquad (3.3.2)$$

$$= \frac{1}{N-1} \left[\sum_{i=1}^{N} x_i^2 - N\bar{x}^2 \right] . \qquad (3.3.3)$$

Για τον εύκολο υπολογισμό της S^2 των δεδομένων: (1) 2, 2, 2, 3, 4, 4, 4, 5, με τη χρήση των προηγούμενων τύπων, κατασκευάζεται συνήθως βοηθητικός πίνακας, όπως αυτός της επόμενης σελίδας, στις στήλες του οποίου καταγράφονται τα κύρια συστατικά των τύπων (3.3.1) και (3.3.2).

Έτσι $\bar{x} = \frac{1}{9} \sum_{i=1}^{9} x_i = \frac{27}{9} = 3.$

Από τον τύπο (3.3.1)

$$S^2 = \frac{1}{9-1} \sum_{i=1}^{9} (x_i - \bar{x})^2 = \frac{1}{9-1}\ 14 = \frac{14}{8} = 1.75.$$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	x_i^2
	1	-2	4	1
	2	-1	1	4
	2	-1	1	4
	2	-1	1	4
	3	0	0	9
	4	1	1	16
	4	1	1	16
	4	1	1	16
	5	2	4	25
ΣΥΝΟΛΟ	27		14	95

Από τον τύπο (3.3.2)

$$S^2 = \frac{1}{9-1}\left[\sum_{i=1}^{9} x_i^2 - \frac{(\sum_{i=1}^{9} x_i)^2}{9}\right] = \frac{1}{8}\left[95 - \frac{(27)^2}{9}\right] = \frac{1}{8}(95-81) = 1.75$$

και από τον τύπο (3.3.3)

$$S^2 = \frac{1}{9-1}\left(\sum_{i=1}^{9} x_i^2 - N\bar{x}^2\right) = \frac{1}{8}(95-9(3)^2) = \frac{1}{8}(95-81) = 1.75.$$

3.3.2.2 Υπολογισμός S^2 απ' το διατεταγμένο πίνακα συχνοτήτων

Αν ν είναι το πλήθος των διακεκριμένων μετρήσεων x_i και f_i οι αντίστοιχες συχνότητες εμφάνισής τους τότε:

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{v}(x_i - \bar{x})^2 f_i \qquad (3.3.4)$$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ή η ισοδύναμη έκφραση, που είναι γνωστή ως τύπος μηχανής και δίνεται από τη σχέση:

$$S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{v} x_i^2 f_i - \frac{(\sum_{i=1}^{v} x_i f_i)^2}{N}\right] \qquad (3.3.5)$$

$$= \frac{1}{N-1}\left(\sum_{i=1}^{v} x_i^2 f_i - N\bar{x}^2\right). \qquad (3.3.6)$$

Για τον υπολογισμό της διακύμανσης των δεδομένων: 1, 2, 2, 2, 3, 4, 4, 4, 5 απ' τον διατεταγμένο πίνακα συχνοτήτων κατασκευάζεται βοηθητικός πίνακας που περιλαμβάνει τα κύρια συστατικά των προηγούμενων τύπων:

x_i	f_i	$(x_i-\bar{x})^2$	$(x_i-\bar{x})^2 f_i$	x_i^2	$x_i f_i$	$x_i^2 f_i$
1	1	4	4	1	1	1
2	3	1	3	4	6	12
3	1	0	0	9	3	9
4	3	1	3	16	12	48
5	1	4	4	25	5	25
			14		27	95

με $\bar{x} = \frac{1}{9}\sum_{i=1}^{v} x_i f_i = \frac{1}{9}\,27 = 3.$

Από την χρησιμοποίηση των τύπων (3.3.4), (3.3.5) και (3.3.6) προκύπτει αντίστοιχα:

$$S^2 = \frac{1}{9-1}\sum_{i=1}^{5}(x_i - \bar{x})^2 f_i = \frac{1}{8}\,14 = 1.75,.$$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

$$S^2 = \frac{1}{9-1}\left[\sum_{i=1}^{5} x_i^2 f_i - \frac{(\sum_{i=1}^{5} x_i f_i)^2}{9}\right] = \frac{1}{8}(95 - \frac{(27)^2}{9}) = \frac{1}{8}(95-81) = 1.75,$$

$$S^2 = \frac{1}{9-1}(\sum_{i=1}^{5} x_i^2 f_i - 9\,\bar{x}^2) = \frac{1}{8}(95-9\,(3^2)) = \frac{1}{8}(95-81) = 1.75.$$

3.3.2.3 Υπολογισμός S^2 απ' τον ομαδοποιημένο πίνακα συχνοτήτων

Αν k είναι το πλήθος των ομάδων, m_i οι μέσες τους τιμές και f_i οι αντίστοιχες συχνότητες των ομάδων, τότε η διακύμανση ορίζεται:

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{k}(m_i - \bar{x})^2 f_i \qquad (3.3.7)$$

ή η ισοδύναμη έκφραση που είναι γνωστή ως τύπος μηχανής και δίνεται απ' τη σχέση:

$$S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{k} m_i^2 f_i - \frac{(\sum_{i=1}^{k} m_i f_i)^2}{N}\right] \qquad (3.3.8)$$

$$= \frac{1}{N-1}\left(\sum_{i=1}^{k} m_i^2 f_i - N\bar{x}^2\right). \qquad (3.3.9)$$

Ο ομαδοποιημένος πίνακας συχνοτήτων, με k=5 ομάδες, των δεδομένων: 1, 2, 2, 2, 3, 4, 4, 4, 5 είναι:

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΟΜΑΔΑ	ΟΡΙΑ	m_i	f_i	f_i/N	F_i	F_i/N
1	[0.5 , 1.5]	1	1	0.11	1	0.11
2	[1.5 , 2.5]	2	3	0.33	4	0.44
3	[2.5 , 3.5]	3	1	0.11	5	0.55
4	[3.5 , 4.5]	4	3	0.33	8	0.88
5	[4.5 , 5.5]	5	1	0.11	9	0.99

Για τον υπολογισμό της διακύμανσης από τον ομαδοποιημένο πίνακα συχνοτήτων κατασκευάζεται ο ακόλουθος βοηθητικός πίνακας που περιέχει τα κύρια συστατικά των τύπων (3.3.7), (3.3.8) και (3.3.9).

m_i	f_i	$(m_i - \bar{x})^2$	$(m_i - \bar{x})^2 f_i$	m_i^2	$m_i f_i$	$m_i^2 f_i$
1	1	4	4	1	1	1
2	3	1	3	4	6	12
3	1	0	0	9	3	9
4	3	1	3	16	12	48
5	1	4	4	25	5	25
			14		27	95

Η χρησιμοποίηση των τύπων (3.3.7), (3.3.8) και (3.3.9) δίνει:

$$S^2 = \frac{1}{9-1} \sum_{i=1}^{5} (m_i - \bar{x})^2 f_i = \frac{1}{8} 14 = 1.75,$$

$$S^2 = \frac{1}{9-1} \left[\sum_{i=1}^{5} m_i^2 f_i - \frac{(\sum_{i=1}^{5} m_i f_i)^2}{9} \right] = \frac{1}{8} \left(95 - \frac{(27)^2}{9}\right) = 1.75,$$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

και $S^2 = \frac{1}{9-1}\left(\sum_{i=1}^{5} m_i^2 f_i - 9\bar{x}^2\right) = \frac{1}{8}(95 - 9(3^2)) = 1.75.$

Παράδειγμα 3.3.1: Να υπολογιστεί με τους τρεις τρόπους η διακύμανση S^2 των δεδομένων του δείκτη νοημοσύνης των 35 παιδιών του Πίνακα 1.1

α) Το S^2 απ' τα δεδομένα

Κατασκευάζουμε τον βοηθητικό πίνακα.

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	x_i^2
67	-29.94	896.4	4489
72	-24.94	622	5184
75	-21.94	481.4	5625
.	.	.	.
.	.	.	.
.	.	.	.
119	22.06	486.64	14161
122	25.06	628	14884
126	29.06	844.48	15876
3393		7113.9	336041

όπου $\bar{x} = \frac{1}{35}\sum_{i=1}^{35} x_i = \frac{1}{35}\,3393 = 96.94.$

Από τους τύπους (3.3.1), (3.3.2), (3.3.3) προκύπτει αντίστοιχα:

$$S^2 = \frac{1}{34}\,7113.9 = 209.23, \quad S^2 = \frac{1}{34}\left(336041 - \frac{(3393)^2}{35}\right) = 209.23$$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

και $S^2 = \frac{1}{34}\ (336041\text{-}35\ (96.94)^2) = 209.8$.

β) <u>Το S^2 απ΄ τον διατεταγμένο πίνακα συχνοτήτων</u>

Με τη βοήθεια του Πίνακα 2.1, για τον υπολογισμό της S^2, κατασκευάζεται ο ακόλουθος βοηθητικός πίνακας.

x_i	f_i	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f_i$	x_i^2	$x_i f_i$	$x_i^2 f_i$
67	1	896.4	896.4	4489	67	4489
72	1	622	622	5184	72	5184
75	1	481.4	481.4	5625	75	5625
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
111	2	197.68	395.36	12321	222	24642
117	1	402.4	402.4	13689	117	13689
119	1	486.64	486.64	14161	119	14161
122	1	628	628	14884	122	14884
126	1	844.48	844.48	15876	126	15876
			7113.9		3393	336041

Οι τύποι (3.3.4) , (3.3.5) και (3.3.6) δίνουν αντίστοιχα.

$$S^2 = \frac{1}{34}\ 7113.9 = 209.23\ , \quad S^2 = \frac{1}{34}\ \left(336041 - \frac{(3393)^2}{35}\right) = 209.23$$

και $S^2 = \frac{1}{34}\ (336041 - 35\ (96.94)^2) = 209.8.$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

γ) Το S^2 απ' τον ομαδοποιημένο πίνακα συχνοτήτων

Με τη βοήθεια του ομαδοποιημένου Πίνακα 2.2 κατασκευάζεται ο ακόλουθος βοηθητικός πίνακας:

m_i	f_i	$(m_i - \bar{x})^2$	$(m_i - \bar{x})^2 f_i$	m_i^2	$m_i f_i$	$m_i^2 f_i$
71.5	3	661	1983.01	5112.2	214.5	15336.6
81.5	5	246.8	1234.02	6642.2	407.5	33211
91.5	9	32.6	293.44	8372.2	823.5	75349.8
101.5	9	18.4	165.64	10302.2	913.5	92719.8
111.5	5	204.2	1021.02	12432.2	557.5	62161
121.5	4	590	2360	14762.2	486	59048.8
			7057.13		3402.5	337827

Από τον πίνακα αυτό το $\bar{x} = \frac{1}{35}\, 3402.5 = 97.21$ και από τους τύπους (3.3.7), (3.3.8) και (3.3.9) παίρνουμε:

$$S^2 = \frac{1}{34}\, 7057.13 = 207.56, \quad S^2 = \frac{1}{34}\left(337827 - \frac{(3402.5)^2}{35}\right) = 207.51$$

και $S^2 = \frac{1}{34}\left(337827 - 35\,(97.21)^2\right) = 208.37.$

Παρατήρηση 3.3.1: Η τιμή της διακύμανσης που προκύπτει από τον α και β τρόπο είναι διαφορετική από την τιμή που προκύπτει με το γ τρόπο. Η διαφορά αυτή είναι αναμενόμενη και οφείλεται στην απώλεια της πληροφορίας λόγω ομαδοποίησης των μετρήσεων, στην υποκατάσταση δηλ. των μετρήσεων από τις μέσες τιμές των ομάδων στις οποίες ανήκουν.

Η διαφορά που υπάρχει στη διακύμανση S^2 από τη χρησιμοποίηση των τριών διαφορετικών τύπων, για κάθε μια από τις

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

περιπτώσεις α, β, και γ, οφείλεται σε σφάλματα στρογγυλεύσεως που γίνονται κατά τη διάρκεια των υπολογισμών.

Παράδειγμα 3.3.2: Να υπολογιστεί με τη χρήση του ομαδοποιημένου πίνακα συχνοτήτων: α) Η διακύμανση των δεδομένων του Παραδείγματος 2.1. β) Η διακύμανση των δεδομένων του Παραδείγματος 2.2.

α) Στο παράδειγμα 3.2.2 α) έχει υπολογιστεί ότι: $\bar{x} = 66.36$. Επίσης με τη βοήθεια του ομαδοποιημένου πίνακα του Παραδείγματος 2.1 κατασκευάζεται ο επόμενος βοηθητικός πίνακας:

m_i	f_i	$(m_i - \bar{x})^2$	$(m_i - \bar{x})^2 f_i$	m_i^2	$m_i f_i$	$m_i^2 f_i$
34.5	5	1015.06	5075.3	1190.25	172.5	5951.25
44.5	8	477.86	3822.88	1980.25	356	15842
54.5	12	140.66	1687.92	2970.25	654	35643
64.5	14	3.46	48.43	4160.25	903	58243.5
74.5	13	66.26	861.37	5550.25	968.5	72153.25
84.5	11	329.06	3619.65	7140.25	929.5	78542.75
94.5	7	791.86	5543.02	8930.25	661.5	62511.75
			20658.57		4645	328887.5

Έτσι από τους τύπους (3.3.7), (3.3.8), (3.3.9) προκύπτει αντίστοιχα $S^2 = \frac{1}{69}\, 20658.57 = 299.4$, $S^2 = \frac{1}{69}\,(328887.5 - \frac{(4645)^2}{70}) = 299.4$ και $S^2 = \frac{1}{69}\,(328887.5 - 70(66.36)^2) = 299.01$.

β) Όπως και στην περίπτωση α) απ' το παράδειγμα 3.2.2 β) προκύπτει

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ότι $\bar{x} = 65.5$ ενώ με τη βοήθεια του πίνακα του Παραδείγματος 2.2 κατασκευάζεται ο πίνακας:

m_i	f_i	$(m_i - \bar{x})^2$	$(m_i - \bar{x})^2 f_i$	m_i^2	$m_i f_i$	$m_i^2 f_i$
34.5	1	961	961	1190.25	34.5	1190.25
44.5	8	441	3528	1980.25	356	15842
54.5	13	121	1573	2970.25	708.5	38613.25
64.5	21	1	21	4160.25	1354.5	87365.25
74.5	17	81	1377	5550.25	1266.5	94354.25
84.5	8	361	2888	7140.25	676	57122
94.5	2	841	1682	8930.25	189	17860.5
			12030		4585	312347.5

Απ' τους τύπους (3.3.7), (3.3.8) και (3.3.9) προκύπτει ότι $S^2 = 174.35$, $S^2 = 174.35$ και $S^2 = 174.35$ αντίστοιχα.

Παρατήρηση 3.3.2: α) Από τη σύγκριση των διακυμάνσεων των βαθμών των φοιτητών και φοιτητριών, που υπολογίστηκαν στο προηγούμενο παράδειγμα, προκύπτει ότι: $S^2_{\text{φοιτητριών}} < S^2_{\text{φοιτητών}}$.

Το γεγονός αυτό ήταν αναμενόμενο και από το Σχήμα 3.2.

β) Συχνά στη βιβλιογραφία η διακύμανση S^2 ενός συνόλου αριθμητικών δεδομένων ορίζεται με έναν από τους τύπους (3.3.1) κ.ο.κ. (3.3.9) με τον παράγοντα $\frac{1}{N-1}$ να αντικαθίσταται από τον παράγοντα $\frac{1}{N}$. Έτσι οι τύποι (3.3.1) κ.ο.κ. (3.3.9) γίνονται:

$$S^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2 \quad \text{κ.ο.κ.} \quad S^2 = \frac{1}{N} \left(\sum_{i=1}^{k} m_i^2 f_i - N\bar{x}^2 \right).$$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

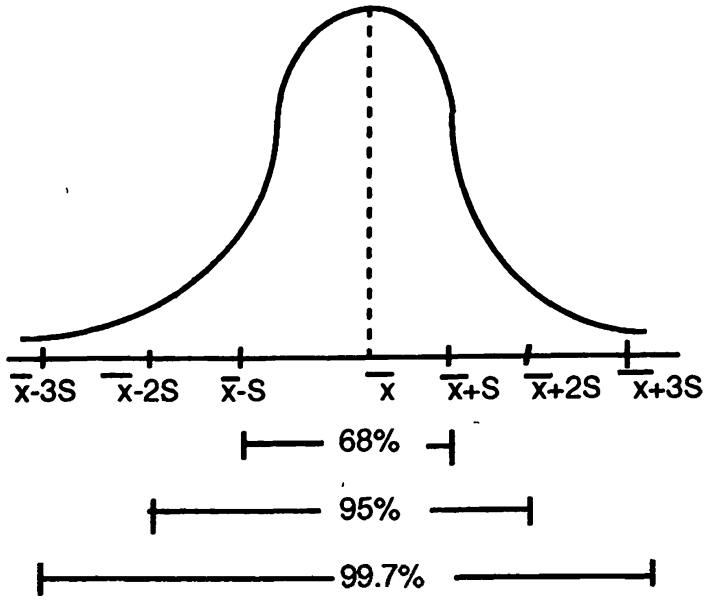
Η διακύμανση S^2 δόθηκε με τους τύπους (3.3.1) - (3.3.9) λόγω των καλών στατιστικών ιδιοτήτων που ικανοποιεί η S^2 όταν ορίζεται με αυτό τον τρόπο.

3.3.3 Τυπική Απόκλιση

Συμβολίζεται με S, ορίζεται ως η θετική τετραγωνική ρίζα της διακύμανσης, δηλ. $S = + \sqrt{S^2}$ και εκφράζεται σε μονάδες μέτρησης ίδιες με εκείνες των δεδομένων. Από το παράδειγμα 3.3.1 γ) προκύπτει ότι η τυπική απόκλιση του δείκτη νοημοσύνης των 35 παιδιών προσχολικής ηλικίας είναι: $S = + \sqrt{207.51} = 14.41$ ενώ από το Παράδειγμα 3.3.2 η τυπική απόκλιση των βαθμών των φοιτητών και φοιτητριών είναι αντίστοιχα $S = \sqrt{299.4} = 17.30$ και $S = \sqrt{174.35} = 13.20$.

Μια χαρακτηριστή ιδιότητα της τυπικής απόκλισης είναι η εξής: Αν η κατανομή συχνοτήτων (ή το πολύγωνο συχνοτήτων που την προσεγγίζει) ενός συνόλου αριθμητικών δεδομένων είναι συμμετρική γύρω από τη μέση τους τιμή $\bar{x}$, τότε το 68% των μετρήσεων βρίσκονται σε διάστημα μιας τυπικής απόκλισης από το $\bar{x}$, δηλ. στο διάστημα $(\bar{x} - S, \bar{x} + S)$, το 95% των μετρήσεων στο διάστημα $(\bar{x} - 2S, \bar{x} + 2S)$ και τέλος το 99.7% των μετρήσεων στο διάστημα $(\bar{x} - 3S, \bar{x} + 3S)$. Η ιδιότητα αυτή συχνά συναντάται στη βιβλιογραφία εκφραζόμενη με τη βοήθεια σχήματος ανάλογου με το σχήμα που ακολουθεί.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ



Η προηγούμενη ιδιότητα επαληθεύεται για τα δεδομένα του Παραδείγματος 3.3.1 με μέση τιμή: $\bar{x}$ = 97.21 και S = 14.41. Από τον ομαδοποιημένο πίνακα συχνοτήτων των μετρήσεων αυτών (Πίνακας 2.2) προκύπτει ότι η κατανομή συχνοτήτων τους είναι σχεδόν συμμετρική γύρω από το $\bar{x}$ = 97.21. Επίσης στο διάστημα ($\bar{x}$ -S, $\bar{x}$ +S) = (82.8, 111.62) βρίσκονται 24 μετρήσεις ή ποσοστό 68.57%, και στο διάστημα ($\bar{x}$ -2S, $\bar{x}$ + 2S) = (68.39, 126.03) βρίσκονται 34 μετρήσεις ή ποσοστό 97.14%.

3.3.4 Συντελεστής μεταβλητότητας

Η τυπική απόκλιση είναι χρήσιμο μέτρο διασποράς όταν εφαρμόζεται σε ένα σύνολο αριθμητικών δεδομένων ή σε δύο σύνολα δεδομένων με ίσες μέσες τιμές. Όταν όμως μας ενδιαφέρει η σύγκριση της μεταβλητότητας δύο συνόλων δεδομένων με διαφορετικές μέσες τιμές, η εξαγωγή συμπερασμάτων από τη σύγκριση των διακυμάνσεων ή των τυπικών αποκλίσεων μπορεί να οδηγήσει σε ολέθρια αποτελέσματα.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Για παράδειγμα, μεταβλητότητα 5000 δρχ. σε μέσο μηνιαίο εισόδημα 50.000 δρχ. έχει διαφορετική έννοια από ότι σε μέσο μηνιαίο εισόδημα 200.000 δρχ. Επίσης μεταβλητότητα 2cm στο ύψος παιδιών προσχολικής ηλικίας είναι ίσως άνευ σημασίας αλλά μεταβλητότητα 2cm στο μέσο μήκος της μύτης των ίδιων παιδιών μπορεί να είναι καταστροφική. Γίνεται έτσι φανερή η ανάγκη εισαγωγής ενός μέτρου που θα εκφράζει την μεταβλητότητα των μετρήσεων απαλλαγμένη από τη μέση τους τιμής. Το μέτρο αυτό ονομάζεται **συντελεστής μεταβλητότητας** (coefficient of variation), είναι καθαρός αριθμός (αναξάρτητος από μονάδες μέτρησης), συμβολίζεται με C.V. και ορίζεται

$$C.V = S/\bar{x} \quad \text{ή} \quad C.V = (S/\bar{x})\ 100\%.$$

Από το Παράδειγμα 3.3.1 γ) προκύπτει ότι ο συντελεστής μεταβλητότητας του δείκτη νοημοσύνης των 35 παιδιών είναι: $C.V = \frac{14.41}{97.21} = 0.148$ ή 14.8% ενώ, από το Παράδειγμα 3.3.2, ο συντελεστής μεταβλητότητας των βαθμών των φοιτητών και φοιτητριών είναι αντίστοιχα: C.V = 0.261 ή 26.1% και C.V = 0.202 ή 20.2% γεγονός που σημαίνει ότι ο βαθμός διασποράς της βαθμολογίας των φοιτητών είναι μεγαλύτερος από το βαθμό διασποράς της βαθμολογίας των φοιτητριών. Γενικά θα δεχόμαστε ότι ένα σύνολο μετρήσεων είναι ομοιογενές αν ο συντελεστής μεταβλητότητας δεν ξεπερνά το 10%.

3.3.5 Εκατοστιαία σημεία

Η έννοια της συχνότητας, της αθροιστικής συχνότητας, των αντιστοίχων σχετικών συχνοτήτων καθώς επίσης και οι γραφικές τους παραστάσεις δίνουν τη δυνατότητα προσδιορισμού των περιπτώσεων, του ποσοστού των μετρήσεων που είναι μεγαλύτερες, μικρότερες ή

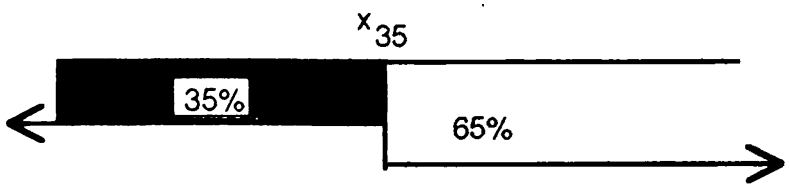
ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

βρίσκονται μεταξύ δύο άλλων τιμών. Η εφαρμογή τους παρουσιάστηκε στα Παραδείγματα 2.1γ και 2.2γ.

Στο αντίστροφο πρόβλημα απαντούν τα εκατοστιαία σημεία (quantiles). Είναι εκείνες οι τιμές - του συνόλου τιμών των δεδομένων - για τις οποίες είναι γνωστό εκ των προτέρων το ποσοστό των μετρήσεων που είναι μικρότερες ή ίσες από τις τιμές αυτές.

Πιο συγκεκριμένα το **p-οστό εκατοστιαίο σημείο** των μετρήσεων, συμβολίζεται με x_p και είναι το σημείο με την εξής χαρακτηριστική ιδιότητα: Το p % των μετρήσεων είναι μικρότερες ή ίσες από το x_p και συνεπώς το (100-p)% των μετρήσεων μεγαλύτερες του x_p . Το p είναι ένας αριθμός μεταξύ του 0 και του 100. Έτσι το x_{35} είναι το σημείο για το οποίο το 35% των μετρήσεων είναι μικρότερες ή ίσες του x_{35} και το 65% των μετρήσεων μεγαλύτερες του x_{35}.

Σχηματικά:



Ιδιαίτερο ενδιαφέρον παρουσιάζουν τα εκατοστιαία σημεία x_{25}, x_{50} , x_{75} τα οποία χωρίζουν το σύνολο των μετρήσεων σε τέσσερα μέρη και για το λόγο αυτό ονομάζονται **τεταρτημόρια.** Ειδικότερα το x_{25} είναι το πρώτο τεταρτημόριο, το x_{50} το δεύτερο τεταρτημόριο και κατ' αναλογία το x_{75} το τρίτο τεταρτημόριο. Τα τεταρτημόρια είναι τα σημεία αριστερά των οποίων βρίσκεται το 25%, το 50% και το 75% του συνόλου των μετρήσεων αντίστοιχα. Το δεύτερο τερταρτημόριο, x_{50}, χωρίζει το σύνολο των δεδομένων σε δυο ίσα μέρη και επομένως συμπίπτει με τη διάμεσό τους Μ. Στη συνέχεια περιγράφεται η διαδικασία υπολογισμού εκατοστιαίων σημείων από ομαδοποιημένο

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

πίνακα συχνοτήτων. Η διαδικασία είναι ανάλογη με εκείνη της διαμέσου.

3.3.5.1 Υπολογισμός του εκατοστιαίου σημείου x_p

Το πρώτο βήμα για τον υπολογισμό του x_p από τον ομαδοποιημένο πίνακα συχνοτήτων, είναι ο εντοπισμός της ομάδας η οποία το περιέχει. Με συλλογισμό ανάλογο με εκείνο της διαμέσου, το x_p βρίσκεται στην ομάδα i για την οποία συμβαίνει

$$F_{i-1} < \frac{pN}{100} \leq F_i \text{ ή } \frac{F_{i-1}}{N} < \frac{p}{100} \leq \frac{F_i}{N} ,$$

με $F_0 = 0$ και $i = 1, 2, \ldots, k$.

Αφού εντοπιστεί η ομάδα i η οποία το περιλαμβάνει, το εκατοστιαίο σημείο x_p υπολογίζεται από τον τύπο:

$$x_p = L_i + d\left(\frac{\frac{pN}{100} - F_{i-1}}{f_i}\right),$$

με L_i το κάτω όριο της ομάδας που περιέχει το x_p, d το μήκος των ομάδων, F_{i-1} η αθροιστική συχνότητα της προηγούμενης ομάδας και f_i η συχνότητα της ομάδας που περιέχει το εκατοστιαίο σημείο.

Παράδειγμα 3.3.3: Για τα δεδομένα του δείκτη νοημοσύνης των 35 παιδιών του Πίνακα 1.1 να υπολογιστούν τα εκατοστιαία σημεία x_{25}, x_{50} και x_{75}.

Χρησιμοποιείται ο ομαδοποιημένος Πίνακας 2.2.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

i) Για $p = 25$ είναι $\frac{pN}{100} = 8.75$. Επειδή $F_2 = F_{3-1} = 8 < 8.75 = \frac{pN}{100} \leq 17 = F_3$ το x_{25} βρίσκεται στην $3^{η}$ ομάδα και είναι:

$$x_{25} = L_3 + d\ (\frac{8.75-F_2}{f_3}) = 86.5 + 10\ (\frac{8.75-8}{9}) = 87.3\ .$$

Έτσι το 25% των μετρήσεων είναι μικρότερες ή ίσες του 87.3.

ii) Για $p = 50$ είναι $\frac{pN}{100} = 17.5.$ Επειδή

$$F_3 = F_{4-1} = 17 < 17.5 = \frac{pN}{100} \leq 26 = F_4$$ το x_{50} βρίσκεται στην $4^{η}$ ομάδα και είναι:

$$x_{50} = L_4 + d\ (\frac{17.5-F_3}{f_4}) = 97.06\ .$$

Έτσι το 50% των μετρήσεων είναι μικρότερες του 97.06. Παρατηρείστε επίσης ότι το x_{50} είναι περίπου ίσο με τη διάμεσο που υπολογίστηκε στο Παράδειγμα 3.2.3 α .

iii) Για $p = 75$ το $\frac{pN}{100} = 26.25.$ Είναι

$$F_4 = 26 < 26.25 = \frac{pN}{100} \leq 31 = F_5$$ και το x_{75} βρίσκεται στην $5^{η}$ ομάδα και είναι

$$x_{75} = L_5 + d\ (\frac{26.25-F_4}{f_5}) = 107\ .$$

Δηλαδή το 75% των μετρήσεων είναι μικρότερες του 107.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Ασκήσεις $2^{ου}$ και $3^{ου}$ Κεφαλαίου

2-3.1: Οι παρακάτω μετρήσεις αποτελούν τα ορθογραφικά λάθη σε κάποια έκθεση 30 μαθητών της Γ' τάξης ενός Δημοτικού Σχολείου και δίνονται διατεταγμένα σε αύξουσα τάξη μεγέθους:

8, 9, 9, 9, 10, 11, 11, 11, 11, 12, 12, 12, 12, 12, 12,
13, 13, 14, 14, 14, 14, 14, 15, 16, 16, 17, 17, 18, 18, 18.

i) Να κατασκευαστεί ο ομαδοποιημένος πίνακας συχνοτήτων (να δοθούν τα πραγματικά όρια των ομάδων, μέση τιμή ομάδας, συχνότητα, σχετική συχνότητα, αθροιστική συχνότητα, σχετική αθροιστική συχνότητα) θεωρώντας τα δεδομένα συνεχή.

ii) Να κατασκευαστεί το ιστόγραμμα και το πολύγωνο συχνοτήτων, και να υπολογιστεί με τη βοήθειά τους το ποσοστό των μετρήσεων (λαθών στην προκειμένη περίπτωση) που είναι μικρότερες, μεγαλύτερες ή μεταξύ δυο τιμών της επιλογής σας.

iii) Να υπολογιστεί η μέση τιμή $\bar{x}$, η διακύμανση S^2, η διάμεσος M, η κορυφή K, ο συντελεστής μεταβλητότητας C.V. και το πρώτο, δεύτερο και τρίτο τεταρτημόριο.

iv) Σχολιάστε τις τιμές που βρήκατε.

2-3.2: Να απαντηθούν τα ερωτήματα της προηγούμενης άσκησης για τα ακόλουθα δεδομένα που αφορούν τα μηνιαία έξοδα σε χιλιάδες δραχμές 20 φοιτητών.

20, 30, 35, 40, 40, 45, 50, 50, 50, 55
58, 60, 60, 63, 64, 67, 68, 70, 73, 89.

2-3.3: Για τις μετρήσεις

1, 0, 2, 1, 0, 2, 3, 2, 4, 3, 1, 2, 5, 3, 2, 2,

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

2, 1, 3, 2, 4, 1, 3, 0, 4, 1, 4, 2, 5, 0, 3, 2,

που αφορούν τον αριθμό των παιδιών 32 οικογενειών να γίνει ο πίνακας συχνοτήτων και η γραφική του παράσταση.

2-3.4: Για τη βαθμολογία στα Αγγλικά 40 φοιτητών που είναι:

42, 88, 37, 75, 96, 93, 73, 62,
96, 80, 52, 76, 66, 54, 73, 69,
83, 62, 53, 79, 69, 56, 81, 75,
52, 65, 49, 80, 67, 59, 88, 80,
44, 71, 72, 87, 91, 82, 89, 79.

να απαντηθούν τα ερωτήματα της Άσκησης 2-3.1.

2-3.5: Τα 28 πρώτα νεογέννητα του 1990 σε μαιευτήριο της Αθήνας κατανέμονται ως προς το φύλο τους ως εξής:

Α Κ Α Κ Κ Κ Κ Κ Α Κ Α Κ Α Κ
Κ Α Α Κ Α Κ Κ Α Κ Α Α Κ Κ Α

Να γίνει περιγραφή των παραπάνω παρατηρήσεων με πίνακα και οι γραφικές παραστάσεις του πίνακα αυτού.

2-3.6: Για να εξεταστεί πόσο πλούσιο είναι το λεξιλόγιο 40 παιδιών μιας ορισμένης ηλικίας, τους ζητήθηκε να περιγράψουν πως περνά μια συνηθισμένη μέρα τους. Ο αριθμός των διαφορετικών λέξεων που χρησιμοποίησαν είναι:

138,164,150,132,144,125,149,157,146,158,140,147, 136,
148,152,144,168,126,138,176,163,119,154,165,146,173,
142,147,135,153,140,135,161,145,135,142,150,156,145,128.

Να απαντηθούν τα ερωτήματα της άσκησης 2-3.1.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΚΕΦΑΛΑΙΟ 4

ΣΤΟΙΧΕΙΑ ΑΠΟ ΤΗ ΘΕΩΡΙΑ ΠΙΘΑΝΟΤΗΤΩΝ

Παρασκευή 30/03/2001.

4.1 Εισαγωγικά

Η έννοια της πιθανότητας (probability) αποτελεί έκφραση του καθημερινού λεξιλόγιου. Στις προτάσεις: "Δεν νομίζω ότι αύριο θα βρέξει", "οι πιθανότητες να περάσω το μάθημα είναι 50%", "κατά πάσα πιθανότητα αύριο θα πάμε σινεμά" διατυπώνεται η άποψη, η πίστη, η προσδοκία, για την πραγματοποίηση ενός φαινομένου, μιας κατάστασης, ενός γεγονότος. Κοινό χαρακτηριστικό και των τριών προτάσεων είναι το γεγονός ότι αναφέρονται σε καταστάσεις, η πραγματοποίηση των οποίων είναι αδύνατο να προβλεφθεί. Επίσης φραστικά ή και αριθμητικά εκφράζεται ο βαθμός της βεβαιότητας, ή της πίστης ή της προσδοκίας σχετικά με την πραγματοποίηση της μελλοντικής κατάστασης. Αυτός ακριβώς ο βαθμός βεβαιότητας που με υποκειμενικά κριτήρια διατυπώνεται αποτελεί την υποκειμενική θεώρηση της πιθανότητας.

Το ενδιαφέρον εντοπίζεται στην αντικειμενική θεώρηση, στην αξιωματική θεμελίωση και προσέγγιση της πιθανότητας που επιτυγχάνεται με τη διατύπωση διαισθητικά αποδεκτών νόμων και

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

αρχών στις οποίες πρέπει να υπακούει. Η θεωρία πιθανοτήτων είναι μια μαθηματική θεωρία, αντικείμενο της οποίας είναι η δημιουργία μοντέλων - πιθανοθεωρητικών μοντέλων - για την περιγραφή φαινομένων η έκβαση των οποίων δεν είναι δυνατό να προβλεφθεί. Αποτελεί το θεωρητικό υπόβαθρο της στατιστικής και η αναγκαιότητα γνώσης ορισμένων στοιχείων της φαίνεται στο ακόλουθο παράδειγμα. Υποθέτουμε ότι σε δυο διαφορετικές ομάδες των 50 μαθητών η κάθε μία, με ανάλογα πνευματικά προσόντα, εφαρμόζονται δύο διαφορετικές μέθοδοι διδασκαλίας των μαθηματικών. Μετά την ολοκλήρωση της διδασκαλίας διαπιστώνεται ότι η ομάδα στην οποία εφαρμόστηκε η Α μέθοδος διδασκαλίας είχε καλύτερη επίδοση από την ομάδα στην οποία εφαρμόστηκε η Β μέθοδος διδασκαλίας των μαθηματικών. Είναι η μέθοδος Α πραγματικά καλύτερη από την Β; Πρέπει στο μέλλον να υιοθετηθεί η Α ως μέθοδος διδασκαλίας των μαθηματικών;

Ερωτήματα όπως αυτά διατυπώνονται σε κάθε στατιστική διερεύνηση ενός προβλήματος. Από τα δεδομένα φαίνεται ότι η μέθοδος Α είναι αποτελεσματικότερη από τη Β. Μπορούμε όμως να είμαστε βέβαιοι γι' αυτό και πολύ περισσότερο να υιοθετήσουμε την Α ως μέθοδο διδασκαλίας; Μήπως η διαφορετικότητα στις επιδόσεις των δύο ομάδων οφείλεται στο γεγονός ότι η ομάδα που παρακολούθησε τη Β μέθοδο διδασκαλίας ήταν περισσότερο αδιάφορη; Είναι φανερό πλέον ότι οποιαδήποτε απάντηση δοθεί σε ερωτήσεις που αφορούν στατιστικά δεδομένα πρέπει να διατυπωθεί με κάποιο βαθμό αβεβαιότητας. Αν η αβεβαιότητα είναι μικρή (ή η βεβαιότητα μεγάλη) τότε το συμπέρασμα θα είναι ακλόνητο και η απόφαση ασφαλής. Προκύπτει έτσι επιτακτικά η ανάγκη μέτρησης της αβεβαιότητας και τα κατάλληλα εργαλεία για το σκοπό αυτό δίδονται από τη θεωρία πιθανοτήτων. Υπάρχουν τρεις ορισμοί της πιθανότητας: Ο κλασσικός, ο

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

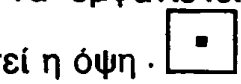
εμπειρικός και ο αξιωματικός. Οι ορισμοί αυτοί παρουσιάζονται στις επόμενες παραγράφους.

4.2 Κλασσικός ορισμός της πιθανότητας

Υπάρχουν δυό απόψεις σχετικά με τη δημιουργία του κλασσικού ορισμού της πιθανότητας και τα προβλήματα που έδωσαν το έναυσμα για τη γέννησή του. Σύμφωνα με τη μια άποψη ο κλασσικός ορισμός της πιθανότητας ξεπήδησε από την ενασχόληση του ανθρώπου με τα τυχερά παιχνίδια. Η άλλη άποψη ισχυρίζεται ότι τα προβλήματα των ασφαλιστικών εταιρειών, που δημιουργούνται από την αλματώδη ανάπτυξη του εμπορίου στα χρόνια της Αναγέννησης, δίνουν το έναυσμα για τη γέννηση του κλασσικού ορισμού και της θεωρίας πιθανοτήτων γενικότερα. Ανεξάρτητα όμως από αυτά και παρά το γεγονός ότι η έννοια της πιθανότητας βρίσκεται από την εποχή του Αρχιμήδη στον ανθρώπινο πολιτισμό, περί τα μέσα του 17ου αιώνα εμφανίζεται ο κλασσικός της ορισμός στις εργασίες των μαθηματικών Β. Pascal και P. Fermat με αφορμή προβλήματα που προέρχονται από το χώρο των τυχερών παιχνιδιών. Η θεωρία πιθανοτήτων καθιερώνεται όμως ως κλάδος των μαθηματικών με την εργασία του Laplace: "*Theorie analytique des Probabilities*" που δημοσιεύθηκε το 1812 και άνοιξε νέους ορίζοντες γι' αυτήν.

Τις ιδέες που περιέχονται στον κλασσικό ορισμό της πιθανότητας διευκρινίζουν με τον καλύτερο τρόπο παραδείγματα από το χώρο των τυχερών παιχνιδιών όπως ρίψεις ζαριών, νομισμάτων, εκλογή καρτών από μια τράπουλα. Στο πλαίσιο αυτό θεωρούμε ένα ζάρι το οποίο ρίχνουμε μια φορά. Το αποτέλεσμα της ρίψης είναι η εμφάνιση μιας από τις έξι όψεις του ζαριού που είναι:

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

και βέβαια δεν υπάρχει κανείς λόγος να πιστεύουμε ότι κάποια όψη είναι πιο δυνατό, πιο εύκολο, πιο "πιθανό" να εμφανιστεί από ότι κάποια άλλη. Έτσι η πιθανότητα να εμφανιστεί η όψη  είναι 1/6 και το ίδιο ισχύει για την εμφάνιση οποιασδήποτε όψης. Αν από μια τράπουλα των 52 χαρτιών εκλεγεί ένα στην τύχη, η πιθανότητα να εκλεγεί σπαθί είναι 13/52, αφού στην τράπουλα υπάρχουν 13 σπαθιά, και το ίδιο ισχύει για καρρώ, κούπα ή μπαστούνι. Με αφετηρία τα παραδείγματα αυτά ο κλασσικός ορισμός της πιθανότητας διατυπώνεται ως εξής:

Ορισμός: Αν N, πεπερασμένος αριθμός, είναι το πλήθος των δυνατών, το ίδιο πιθανών αποτελεσμάτων μιας διαδικασίας και m από αυτά ευνοούν την πραγματοποίηση ενός χαρακτηριστικού, μιας κατάστασης E, η πιθανότητα πραγματοποίησης του E ορίζεται να είναι ίση με $\frac{m}{N}$. Αν με P(E) συμβολιστεί η πιθανότητα πραγματοποίησης του E, ο κλασσικός ορισμός της πιθανότητας συνοψίζεται στον τύπο:

$$P(E)=\frac{m}{N}=\frac{\text{αριθμός αποτελεσμάτων που ευνοούν την πραγματοποίηση του E}}{\text{ολικός αριθμός των ισοπίθανων αποτελεσμάτων}}$$

Ο κλασσικός ορισμός της πιθανότητας στηρίζεται στην ιδέα, στην υπόθεση του "ισοπίθανου" των αποτελεσμάτων π.χ. της ρίψης του ζαριού, της ρίψης ενός νομίσματος. Η υπόθεση αυτή παρά το γεγονός ότι είναι διαισθητικά αποδεκτή για παραδείγματα από το χώρο των τυχερών παιχνιδιών, αποτελεί και την αδυναμία του κλασσικού ορισμού. Αδυναμία αφ' ενός μεν γιατί ο ορισμός διατυπώνεται χρησιμοποιώντας την έννοια που φιλοδοξεί να ορίσει και αφ' ετέρου

ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ ΒΙΒΛΙΟΘΗΚΗ

Για 'σένα έχω κανει κρατηρίο!
ΠΑΟΚ
1926

γιατί δεν μπορεί να χρησιμοποιηθεί σαν μοντέλο για τον υπολογισμό πιθανοτήτων σε καταστάσεις, σε διαδικασίες τα δυνατά αποτελέσματα των οποίων δεν είναι "ισοπίθανα". Επίσης ο κλασσικός ορισμός εφαρμόζεται μόνο για διαδικασίες που το πλήθος N των δυνατών αποτελεσμάτων τους είναι πεπερασμένος αριθμός.

Παράδειγμα 4.2.1: Σε μια αίθουσα των 50 μαθητών οι 3 εμφανίζουν χαρακτηριστικά δυσλεξίας, οι 5 γράφουν με το αριστερό χέρι και οι υπόλοιποι δεν εμφανίζουν κανένα από τα δύο αυτά χαρακτηριστικά. Αν ένας μαθητής εκλεγεί στην τύχη, να υπολογιστούν οι πιθανότητες: α) Να παρουσιάζει δυσλεξία. β) Να γράφει με το αριστερό χέρι. γ) Να μη παρουσιάζει κανένα από τα δύο αυτά χαρακτηριστικά. δ) Να παρουσιάζει δυσλεξία ή να γράφει με το αριστερό χέρι.

Αν συμβολίσουμε με Δ το γεγονός ένας μαθητής να είναι δυσλεκτικός, με A να γράφει με το αριστερό χέρι και με Φ το γεγονός να μη παρουσιάζει κανένα από τα δυο αυτά χαρακτηριστικά, τότε από το κλασσικό ορισμό οι ζητούμενες πιθανότητες είναι:

α) $P(\Delta) = \frac{3}{50} = 0.06.$ β) $P(A) = \frac{5}{50} = 0.1.$ γ) $P(\Phi) = \frac{42}{50} = 0.84$

και δ) $P(\Delta \text{ ή } A) = \frac{3+5}{50} = \frac{8}{50} = 0.16$.

Παράδειγμα 4.2.2: Σε ένα διαγώνισμα υπάρχουν ερωτήσεις με τέσσερις απαντήσεις στην κάθε ερώτηση. Αν μια απάντηση είναι η σωστή, ποιά η πιθανότητα ένας φοιτητής που απαντά στην τύχη να διαλέξει τη σωστή απάντηση; Αν δύο από τις απαντήσεις θεωρηθούν σωστές ποιά η πιθανότητα να δώσει τη σωστή απάντηση;

Και στις δύο περιπτώσεις τέσσερις είναι οι δυνατοί τρόποι με τους οποίους ο φοιτητής μπορεί να διαλέξει στην τύχη μία απάντηση.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Αν μία είναι η σωστή απάντηση η πιθανότητα να την διαλέξει είναι 1/4 ενώ αν οι σωστές απαντήσεις είναι δυο η πιθανότητα να δώσει τη σωστή απάντηση είναι 2/4.

Παρατήρηση 4.2.1: Μια προσεκτική ματιά στα αποτελέσματα του Παραδείγματος 4.2.1 οδηγεί στις ακόλουθες διαπιστώσεις για την πιθανότητα όπως αυτή προσεγγίζεται μέσω του κλασσικού της ορισμού.

i) Η πιθανότητα είναι ένας αριθμός μεγαλύτερος ή ίσος από το 0 και μικρότερος ή ίσος από το 1.

ii) Για το παράδειγμα αυτό ένας μαθητής ανήκει σε μια και μόνο από τις κατηγορίες: Δυσλεκτικός, Αριστερόχειρας ή Φυσιολογικός. Παρατηρούμε ότι το άθροισμα των πιθανοτήτων των διαφορετικών κατηγοριών στις οποίες μπορεί να ανήκει ένας μαθητής είναι ίσο με 1. Δηλαδή $P(\Delta) + P(A) + P(\Phi) = 1$.

iii) Η πιθανότητα ένας μαθητής να έχει τουλάχιστον ένα από δύο διαφορετικά χαρακτηριστικά είναι ίση με το άθροισμα των πιθανοτήτων των χαρακτηριστικών αυτών. Έτσι $P(\Delta \text{ ή } A) = P(\Delta) + P(A)$.

Ο κλασσικός ορισμός της πιθανότητας παρέχει τη δυνατότητα υπολογισμού πιθανοτήτων στηριζόμενος στην υπόθεση του "ισοπίθανου" των αποτελεσμάτων. Η υπόθεση αυτή ικανοποιείται σε πολλές περιπτώσεις, γενικά όμως δεν μπορεί να υιοθετηθεί και αποτελεί την αδυναμία του κλασσικού ορισμού στο επίπεδο των εφαρμογών. Ο εμπειρικός ορισμός της πιθανότητας, που θα παρουσιαστεί στη συνέχεια, ξεπερνάει την αδυναμία αυτή του κλασσικού ορισμού προσεγγίζοντας την έννοια της πιθανότητας μέσω της έννοιας της σχετικής συχνότητας.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Παραουσιή 30/05/2009

4.3 Εμπειρικός ορισμός της πιθανότητας

Ο εμπειρικός ορισμός της πιθανότητας εισάγεται στις αρχές του 20ου αιώνα από τους R. von Mises και A. Fisher και αποτελεί προσέγγιση της πιθανότητας ενός γεγονότος από τη σχετική συχνότητα εμφάνισής του σε μια ακολουθία επαναλήψεων μιας διαδικασίας. Προυποθέτει τη δυνατότητα επανάληψης μιας διαδικασίας, κάτω από τις ίδιες ακριβώς συνθήκες, την απαρίθμηση των επαναλήψεων καθώς και των φορών εμφάνισης, πραγματοποίησης του υπό μελέτη γεγονότος ή χαρακτηριστικού.

Σε πολλά φαινόμενα, δημογραφικής φύσης, έχει παρατηρηθεί ότι το πηλίκο n(E)/ N, του αριθμού n(E) των εμφανίσεων ενός γεγονότος E στις N επαναλήψεις μιας διαδικασίας, παραμένει σχεδόν αμετάβλητο καθώς ο αριθμός N των επαναλήψεων της διαδικασίας μεγαλώνει. Από την αρχαιότητα, για παράδειγμα, ήταν ήδη γνωστό ότι για μεγάλες πόλεις - κράτη το πηλίκο του αριθμού των νεογέννητων αγοριών δια του ολικού αριθμού των γεννήσεων παραμένει σταθερό από χρόνο σε χρόνο και περίπου ίσο με 1/2 . Ανάλογη συμπεριφορά της σχετικής συχνότητας εμφάνισης ενός γεγονότος παρατηρήθηκε και σε μεγάλο αριθμό ρίψεων ενός συμμετρικού, τέλεια κατασκευασμένου, νομίσματος. Ειδικότερα (B. Gnedenko (1976), *The Theory of Probability*, MIR Publishers, Moscow) ο αριθμός των "κεφαλών" σε μεγάλο αριθμό ρίψεων ενός συμμετρικού νομίσματος, που έγινε από τους Buffon και Karl Pearson φαίνεται στον πίνακα:

	ΑΡΙΘ. ΡΙΨΕΩΝ N	ΑΡ. ΚΕΦΑΛΩΝ n(K)	ΣΧΕΤ. ΣΥΧΝΟΤ. n(K)/ N
Buffon	4040	2048	0.5080
Karl Pearson	12000	6019	0.5016
"	24000	12012	0.5005

2048/4040

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Καθώς ο αριθμός των ρίψεων μεγαλώνει η σχετική συχνότητα n(K)/ N τείνει να σταθεροποιηθεί στην τιμή 0.5 που είναι η πιθανότητα εμφάνισης "κεφαλής" σε μια ρίψη του νομίσματος (σύμφωνα με τον κλασσικό ορισμό της πιθανότητας που για την περίπτωση του νομίσματος μπορεί να εφαρμοστεί). Η αναλογία που διαπιστώνεται, μέσω του παραδείγματος αυτού, μεταξύ της κλασσικής πιθανότητας και της σχετικής συχνότητας ενός γεγονότος, υποδεικνύει τη χρησιμοποίηση της σχετικής συχνότητας για την προσέγγιση της πιθανότητας ενός γεγονότος, και σε περιπτώσεις που ο κλασσικός ορισμός δεν εφαρμόζεται. Στο πλαίσιο αυτό ο εμπειρικός ορισμός της πιθανότητας είναι ο ακόλουθος:

Ορισμός: Αν μια διαδικασία επαναληφθεί κάτω από τις ίδιες ακριβώς συνθήκες N φορές, και n(E) φορές πραγματοποιηθεί το γεγονός E, η πιθανότητα P(E) ορίζεται:

$$P(E) = \lim_{N \to \infty} \frac{n(E)}{N} .$$

Η σημασία της προηγούμενης σχέσης είναι η εξής: Καθώς ο αριθμός N των επαναλήψεων μιας διαδικασίας μεγαλώνει το πηλίκο n(E)/N τείνει να σταθεροποιηθεί σε έναν αριθμό τον οποίο θεωρούμε σαν την πιθανότητα του γεγονότος, της κατάστασης E.

Ο εμπειρικός ορισμός της πιθανότητας αποφεύγει την υπόθεση του ισοπίθανου των αποτελεσμάτων μιας διαδικασίας, που αποτελεί την αδυναμία του κλασσικού ορισμού, προϋποθέτει όμως επανάληψη της διαδικασίας μεγάλο αριθμό φορών και μάλιστα κάτω από τις ίδιες ακριβώς συνθήκες.

Παράδειγμα 4.3.1: Ο δείκτης νοημοσύνης(δ.ν) κατατάσσεται στις εξής κατηγορίες: Κατώτερος, Μέτριος, Φυσιολογικός, Καλός,

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Εξαιρετικός και Ανώτερος. Η εμπειρία αποκλείει την παραδοχή του ισοπίθανου των κατηγοριών αυτών. Για την εύρεση της πιθανότητας ένας μαθητής να έχει δ.ν σε μια απ' τις κατηγορίες αυτές ή του ποσοστού των μαθητών με δ.ν σε κάθε μια απ' τις κατηγορίες αυτές, εκτελείται το εξής πείραμα: Ένα τυποποιημένο τεστ νοημοσύνης δίνεται σε 2000 μαθητές. Τα αποτελέσματα του τεστ καταγράφονται στον πίνακα:

Δείκτης Νοημοσ.	Κατώτερ. (Κ)	Μέτριος (Μ)	Φυσιολ. (Φ)	Καλός (ΚΛ)	Εξαιρ. (Ε)	Ανώτερ. (Α)
Μαθητές	80	480	840	520	60	20

Να υπολογιστεί η πιθανότητα ένας μαθητής (ή το ποσοστό των μαθητών) i) να έχει δ.ν. Κατώτερο, Μέτριο, Φυσιολογικό, Καλό, Εξαιρετικό, Ανώτερο. ii) Καλό ή εξαιρετικό iii) Τουλάχιστο μέτριο iv) Το πολύ φυσιολογικό.

Χρησιμοποιώντας τον εμπειρικό ορισμό:

i) $P(K) = \frac{80}{2000} = 0.04$, $P(M) = \frac{480}{2000} = 0.24$, $P(\Phi) = \frac{840}{2000} = 0.42$,

$P(K\Lambda) = \frac{520}{2000} = 0.26$, $P(E) = \frac{60}{2000} = 0.03$, $P(A) = \frac{20}{2000} = 0.01$.

ii) $P(K\Lambda \text{ ή } E) = \frac{520+60}{2000} = 0.29$.

iii) P(τουλάχιστον μέτριο) = P(Μ ή Φ ή ΚΛ ή Ε ή Α) =

$$= \frac{480+840+520+60+20}{2000} = 0.96 .$$

iv) P(Το πολύ Φυσιολογικό) = $P(K \text{ ή } M \text{ ή } \Phi) = \frac{80+480+840}{2000} = 0.7$.

Τα αντίστοιχα ποσοστά μαθητών είναι: 4%, 24% κ.ο.κ. 70% .

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Παρατήρηση 4.3.1: i) Η προσέγγιση της έννοιας της πιθανότητας μέσω του εμπειρικού ορισμού προϋποθέτει τον ορισμό ενός πληθυσμού από γεγονότα, από καταστάσεις όπως οι κατηγορίες του παραπάνω παραδείγματος στις οποίες ταξινομείται ο δ.ν. Η πιθανότητα είναι η σχετική συχνότητα ενός χαρακτηριστικού του πληθυσμού. Είναι επομένως μια παράμετρος του πληθυσμού η οποία προσεγγίζεται από τη σχετική συχνότητα ενός συνόλου μετρήσεων από τον πληθυσμό αυτό.

ii) Μέσω του Παραδείγματος 4.3.1 διαπιστώνεται ότι για την πιθανότητα, όπως ορίζεται από τον εμπειρικό ορισμό, ισχύουν όσα αναφέρθηκαν στην Παρατήρηση 4.2.1.

Οι προηγούμενοι ορισμοί της πιθανότητας παρά το γεγονός ότι είναι εντελώς διαφορετικοί έχουν κοινές ιδιότητες. Επί πλέον οι ιδιότητες αυτές συμφωνούν με την καθημερινή, διαισθητική ερμηνεία του όρου πιθανότητα. Η διαπίστωση αυτή υπαινίσεται τη δυνατότητα ορισμού της πιθανότητας χρησιμοποιώντας τις αποδεκτές αυτές ιδιότητες σαν αξιώματα. Με τον τρόπο αυτό αποφεύγονται και οι αδυναμίες του κλασσικού και εμπειρικού ορισμού.

Πριν δοθεί ο αξιωματικός ορισμός της πιθανότητας θα αναφερθούν τα στοιχεία με την βοήθεια των οποίων οικοδομείται.

4.4. Τυχαίο Πείραμα - Δειγματικός Χώρος - Ενδεχόμενο - Πράξεις Ενδεχομένων.

Έχει ήδη αναφερθεί ότι αντικείμενο της θεωρίας πιθανοτήτων είναι η δημιουργία μαθηματικών μοντέλων που σχεδιάζονται για να περιγράψουν ή να προβλέψουν τα χαρακτηριστικά ενός φαινομένου ή μιας διαδικασίας το αποτέλεσμα ή τα αποτελέσματα της οποίας δεν

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

είναι δυνατό να προβλεφτούν με ακρίβεια, αλλά επηρεάζονται, διέπονται από τον παράγοντα τύχη. Τέτοια φαινόμενα ονομάζονται **τυχαία φαινόμενα**, ενώ η διαδικασία, ο μηχανισμός που μας επιτρέπει να τα παρατηρήσουμε ονομάζεται **τυχαίο πείραμα** ή **πείραμα τύχης**. Τα δυνατά αποτελέσματα ενός τυχαίου πειράματος, είναι κατά κανόνα γνωστά πριν την εκτέλεση του πειράματος. Το σύνολο των αποτελεσμάτων αυτών λέγεται **δειγματικός χώρος** του πειράματος και συμβολίζεται συνήθως με S. Έτσι

i) Ο δειγματικός χώρος του τυχαίου πειράματος που συνίσταται από τη ρίψη ενός ζαριού μια φορά είναι

$S =$.

ii) Η ρίψη μια φορά ενός νομίσματος συνιστά τυχαίο πείραμα ο δειγματικός χώρος του οποίου είναι: S = {Κ(εφαλή), Γ(ράμματα)}.

iii) Μια μελέτη για το δείκτη νοημοσύνης παιδιών προσχολικής ηλικίας στηρίζεται στη μέτρηση του δείκτη των παιδιών αυτών. Η μέτρηση αυτή συνιστά τυχαίο πείραμα. Ο δειγματικός χώρος είναι το σύνολο:

S= {Κατώτερος, Μέτριος, Φυσιολογικός, Καλός, Εξαιρετικός, Ανώτερος}.

iv) Στη μελέτη που αφορά τα ορθογραφικά λάθη, σ' ένα πολυσέλιδο κείμενο, τελειοφοίτων μαθητών Λυκείου ο δειγματικος χώρος είναι: S = {0, 1, 2, ...}.

v) Μια μελέτη που αφορά το ύψος παιδιών προσχολικής ηλικίας στηρίζεται στη μέτρηση του ύψους των παιδιών αυτών. Η μέτρηση του ύψους συνιστά τυχαίο πείραμα με δειγματικό χώρο το σύνολο:

S = {κάθε πραγματικός αριθμός μεγαλύτερος του μηδέν}

$= \{s: s \in \mathbf{R} \text{ και } s > 0\}$,

ή πιο ρεαλιστικά

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

$S = \{s : s \in \mathbf{R} \text{ και } 88cm \leq s \leq 105cm\}$

αφού έχει παρατηρηθεί ότι το ύψος των παιδιών προσχολικής ηλικίας βρίσκεται μεταξύ αυτών των ορίων.

Από τα προηγούμενα παραδείγματα προκύπτει ότι ο δειγματικός χώρος ενός πειράματος μπορεί να είναι πεπερασμένο σύνολο (Παραδ. i, ii, iii) ή αριθμήσιμο σύνολο (Παραδ. iv) ή μη αριθμήσιμο σύνολο (Παράδ. v).

Κάθε δυνατό αποτέλεσμα ενός τυχαίου πειράματος, δηλαδή κάθε στοιχείο του S ονομάζεται **απλό ή στοιχειώδες ενδεχόμενο** ενώ κάθε υποσύνολο του S ονομάζεται **ενδεχόμενο**. Για το Παράδειγμα iii) κάθε ένας από τους χαρακτηρισμούς: Κατώτερος κ.ο.κ. Ανώτερος είναι στοιχειώδες ενδεχόμενο ενώ τα υποσύνολα:
A={ Κατώτερος, Μέτριος, Καλός} , B = { Μέτριος, Εξαιρετικός} του S είναι ενδεχόμενα. Είναι φανερό ότι τα ενδεχόμενα συνίστανται από δύο ή περισσότερα απλά ενδεχόμενα.

Ένα ενδεχόμενο A λέγεται ότι έχει **συμβεί ή πραγματοποιηθεί** αν το αποτέλεσμα του τυχαίου πειράματος είναι στοιχείο του ενδεχόμενου A. Το αποτέλεσμα αυτό λέγεται **ευνοϊκό αποτέλεσμα** για το ενδεχόμενο A. Για το παράδειγμα v) θεωρούμε το ενδεχόμενο $A = \{ s : 95cm \leq S \leq 105cm \}$ που εκφράζει το γεγονός, το ύψος ενός παιδιού προσχολικής ηλικίας να είναι μεταξύ 95 και 105 εκατοστών. Αν το ύψος ενός παιδιού μετρηθεί και βρεθεί να είναι 102cm τότε το ενδεχόμενο A θα έχει πραγματοποιηθεί και το αποτέλεσμα 102cm είναι ευνοϊκό αποτέλεσμα για το ενδεχόμενο A.

Πράξεις μεταξύ ενδεχομένων και ερμηνεία αυτών

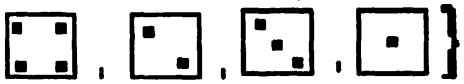
Ο ορισμός των ενδεχομένων σαν υποσυνόλων του δειγματικού χώρου επιτρέπει τον ορισμό πράξεων μεταξύ τους, ανάλογων με

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

εκείνες της θεωρίας συνόλων. Έστω λοιπόν Α και Β είναι δυο ενδεχόμενα ενός πειράματος τύχης με δειγματικό χώρο το σύνολο S.

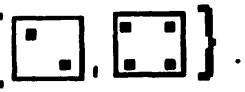
Το ενδεχόμενο **Α συνεπάγεται το Β**, συμβολικά $A \subseteq B$, αν κάθε στοιχείο του ενδεχομένου Α είναι και στοιχείο του ενδεχομένου Β ή ισοδύναμα αν η πραγματοποίηση του ενδεχόμενου Α έχει σαν συνέπεια την πραγματοποίηση του Β. Για το δειγματικό χώρο του δείκτη νοημοσύνης αν: Α = { Κατώτερος, Καλός}, και Β = {Κατώτερος, Μέτριος, Καλός, Ανώτερος} προφανώς $A \subseteq B$.

Η ένωση ή το άθροισμα των ενδεχομένων Α και Β , συμβολίζεται με $A \cup B$, και είναι το ενδεχόμενο που απαρτίζεται από τα κοινά και μη κοινά στοιχειώδη ενδεχόμενα των Α και Β. Για το δειγματικό χώρο της ρίψης ενός ζαριού μια φορά αν τα ενδεχόμενα Α και Β είναι :

$A =$, $B =$ 

τότε το $A \cup B =$.

Η πραγματοποίηση του ενδεχομένου $A \cup B$ είναι ισοδύναμη με την πραγματοποίηση τουλάχιστον ενός από τα ενδεχόμενα Α και Β (του Α ή του Β ή και των δύο).

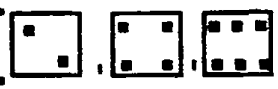
Η τομή ή το γινόμενο των ενδεχομένων Α και Β, συμβολίζεται με $A \cap B$, και είναι το ενδεχόμενο που απαρτίζεται από τα κοινά στοιχειώδη ενδεχόμενα των Α και Β. Για τα προηγούμενα ενδεχόμενα Α και Β της ρίψης του ζαριού είναι $A \cap B =$  .

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Η πραγματοποίηση του ενδεχομένου $A \cap B$ είναι ισοδύναμη με την ταυτόχρονη πραγματοποίηση των Α και Β.

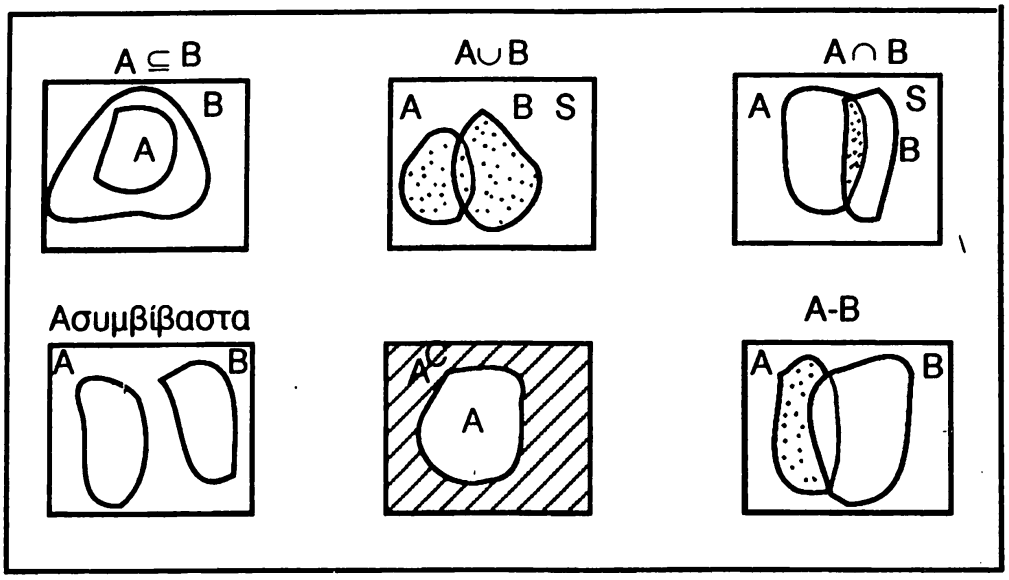
Τα ενδεχόμενα Α και Β λέγονται **ξένα ή ασυμβίβαστα** αν η τομή τους είναι το κενό σύνολο. Δηλαδή η πραγματοποίηση του ενός αποκλείει την πραγματοποίηση του άλλου. Για το δειγματικό χώρο του δείκτη νοημοσύνης αν Α= {Μέτριος, Καλός} και Β={Κατώτερος, Εξαιρετικός} προφανώς τα Α και Β είναι ξένα.

Το **συμπλήρωμα** του Α (ως προς το δειγματικό χώρο S), συμβολίζεται με A^c και ορίζεται να είναι το ενδεχόμενο που απαρτίζεται από όλα τα στοιχειώδη ενδεχόμενα που δεν είναι στοιχεία του Α. Αν Α={Μέτριος, Καλός} τότε A^c = {Κατώτερος, Φυσιολογικός, Εξαιρετικός, Ανώτερος}. Η πραγματοποίηση του ενδεχομένου A^c είναι ισοδύναμη με τη μη πραγματοποίηση του ενδεχομένου Α.

Η **διαφορά των ενδεχομένων Α και Β**, συμβολίζεται με Α-Β, και ορίζεται να είναι το ενδεχόμενο που αποτελείται από τα στοιχεία του Α που δεν ανήκουν στο Β. Για το δειγματικό χώρο της ρίψης του ζαριού αν Α=  και Β= τότε Α-Β = . Η πραγματοποίηση του Α-Β συνίσταται στην πραγματοποίηση του Α και τη μη πραγματοποίηση του ενδεχομένου Β.

Η γραφική παράσταση των προηγούμενων πράξεων μεταξύ ενδεχομένων δίνεται στο σχήμα που ακολουθεί:

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ



Σχήμα 4.1: Γραφικές παραστάσεις πράξεων ενδεχομένων

4.5 Αξιωματικός ορισμός πιθανότητας

Η αξιωματική θεμελίωση της πιθανότητας στηρίζεται σε τρια αξιώματα που συμβιβάζονται απόλυτα με την αντίληψη του κλασσικού και εμπειρικού ορισμού. Ο αξιωματικός ορισμός δόθηκε από τον A. N. Kolmogorov το 1933 και σύμφωνα με αυτόν η πιθανότητα είναι μια συνάρτηση που ορίζεται στο σύνολο των ενδεχομένων ενός τυχαίου πειράματος και ικανοποιεί τα ακόλουθα αξιώματα:

Αξίωμα 1: Η πιθανότητα οποιουδήποτε ενδεχομένου E είναι μεγαλύτερη ή ίση από το μηδέν και μικρότερη ή ίση από τη μονάδα:

$$0 \leq P(E) \leq 1.$$

Αξίωμα 2: Αν $E_1, E_2, \ldots, E_k$ είναι τα στοιχειώδη ενδεχόμενα ενός τυχαίου πειράματος τότε το άθροισμα των πιθανοτήτων τους είναι ίσο με τη μονάδα:

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

$$P(E_1) + P(E_2) + \ldots + P(E_k) = 1 .$$

Αξίωμα 3: Αν $E_1 , E_2, \ldots , E_k$ είναι k ενδεχόμενα ανά δύο ασυμβίβαστα (δηλ. $E_i \cap E_j = \varnothing , i \neq j , i,j = 1,\ldots,k$) τότε

$$P(E_1 \cup E_2 \cup \ldots \cup E_k) = P(E_1) + P(E_2) + \ldots + P(E_k) .$$

Από τα αξιώματα αυτά άμεσα προκύπτουν οι ακόλουθες **ιδιότητες των πιθανοτήτων:**

1. Από τον συνδυασμό των Αξιωμάτων 2 και 3 προκύπτει:

$P(S)=1$.

Λόγω της σχέσης αυτής ο δειγματικός χώρος S ονομάζεται και βέβαιο ενδεχόμενο.

2. Για κάθε ενδεχόμενο A ισχύει: $P(A^c) = 1-P(A)$.

3. Προσθετικός Νόμος: Για οποιαδήποτε ενδεχόμενα A και B ισχύει:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) .$$

Ο αξιωματικός ορισμός της πιθανότητας αποφεύγει τις αδυναμίες του κλασσικού και του εμπειρικού ορισμού και απαντά με σαφήνεια στο ερώτημα ποιούς νόμους, σε ποιούς κανόνες πρέπει να υπακούει η πιθανότητα. Δεν υποδεικνύει όμως και τον τρόπο υπολογισμού της. Έτσι στην αντιμετώπιση κάθε πρακτικού προβλήματος οι τρεις ορισμοί παρά το γεγονός ότι είναι εντελώς διαφορετικοί, παρά τις αδυναμίες τους πρέπει να συνυπάρχουν. Οι δύο πρώτοι παρέχοντας τα εφόδια για τον υπολογισμό πιθανοτήτων ενδεχομένων και ο τρίτος χορηγώντας μαθηματικούς τύπους και κανόνες για τον υπολογισμό πιθανοτήτων σύνθετων ενδεχομένων.

Παράδειγμα 4.5.1: Έχει παρατηρηθεί ότι το 5% του πληθυσμού των παιδιών προσχολικής ηλικίας παρουσιάζουν

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

χαρακτηριστικά δυσλεξίας, το 3% γράφουν με το αριστερό χέρι και το 1% παρουσιάζει και τα δύο χαρακτηριστικά. Να υπολογιστεί η πιθανότητα:

α) Ένα παιδί προσχολικής ηλικίας παρουσιάζει τουλάχιστον ένα από τα χαρακτηριστικά αυτά.

β) Δεν γράφει με το αριστερό χέρι.

Ορίζουμε τα ενδεχόμενα: Δ = { ένα παιδί παρουσιάζει χαρακτηριστικά δυσλεξίας} και Α = { ένα παιδί γράφει με το αριστερό χέρι}.

α) Ζητείται:

$$P(A \cup \Delta) = P(A) + P(\Delta) - P(A \cap \Delta)$$

$$= 0.03 + 0.05 - 0.01$$

$$= 0.07$$

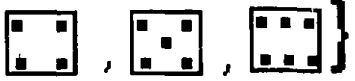
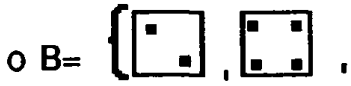
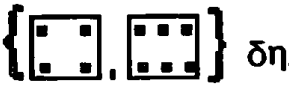
β) Ζητείται:

$$P(A^c) = 1 - P(A) = 1 - 0.03 = 0.97 .$$

4.6 Δεσμευμένη ή υπό συνθήκη πιθανότητα

Η έννοια της δεσμευμένης πιθανότητας είναι σημαντική στη θεωρία πιθανοτήτων για δύο λόγους. Αφ' ενός δίνει τη δυνατότητα υπολογισμού πιθανοτήτων για την περίπτωση που κάποια "μερική πληροφορία" για τα αποτελέσματα του πειράματος είναι διαθέσιμη και αφ' ετέρου δίνει τη δυνατότητα έκφρασης και υπολογισμού της πιθανότητας ενός σύνθετου ενδεχομένου, σε όρους απλούστερων δεσμευμένων πιθανοτήτων.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Η δεσμευμένη πιθανότητα περιγράφεται με τη βοήθεια του παρακάτω παραδείγματος. Ένα ζάρι ρίχνεται μια φορά. Ο δειγματικός χώρος είναι: S = και έστω το ενδεχόμενο: A = { το αποτέλεσμα είναι μεγαλύτερο ή ίσο του τέσσερα } =  . Απ' τον κλασσικό ορισμό της πιθανότητας ισχύει P(A) = 3/6 = 1/2 . Υποθέτουμε ότι αμέσως μετά τη ρίψη του ζαριού δίνεται η πληροφορία ότι το αποτέλεσμα της ρίψης είναι άρτιος αριθμός. Μετά τη λήψη της πληροφορίας αυτής μήπως μεταβάλεται η πιθανότητα του ενδεχομένου A; Πράγματι, επειδή είναι γνωστό ότι το αποτέλεσμα είναι άρτιος αριθμός, τα δυνατά αποτελέσματα της ρίψης του ζαριού είναι τώρα τα: .Έτσι είναι λογικό να δεχτούμε ότι η πιθανότητα του A, λαμβάνοντας υπ' όψη την πληροφορία που δίνεται, είναι τώρα ίση με 2/3. Ο νέος δειγματικός χώρος είναι ο B=  και τα νέα ευνοϊκά αποτελέσματα για το A είναι τα  δηλαδή τα στοιχεία της τομής A∩B. Η πιθανότητα 2/3 λέγεται δεσμευμένη ή υπό συνθήκη πιθανότητα του ενδεχομένου A δοθέντος ότι έχει συμβεί το ενδεχόμενο B και συμβολίζεται με P(A/B). Για το παράδειγμα έχουμε:

$$P(A/B) = \frac{2}{3} = \frac{\text{αριθ. ευνοϊκών αποτελ. για } A\cap B}{\text{αριθ.ευνοϊκών αποτελ. για } B}$$

$$= \frac{\dfrac{\text{αριθ. ευνοϊκών αποτελ. για } A\cap B}{\text{αριθ. δυνατών αποτελ. πειράματος}}}{\dfrac{\text{αριθ. ευνοϊκών αποτελ. για } B}{\text{αριθμ. δυνατών αποτελ. πειράματος}}}$$

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

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Το παράδειγμα αυτό υποδεικνύει τον ακόλουθο ορισμό της δεσμευμένης πιθανότητας.

Ορισμός: Έστω τα ενδεχόμενα A και B με $P(B) \neq 0$. Η δεσμευμένη πιθανότητα του A δοθέντος του B συμβολίζεται με P(A/B) και ορίζεται από τη σχέση

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

Ο ορισμός αυτός είναι αφηρημένος. Η ερμηνεία του και η διαδικασία απ' την οποία προκύπτει δόθηκε στο παραπάνω παράδειγμα. Το παράδειγμα δείχνει ότι η δεσμευμένη πιθανότητα, παριστά πιθανότητα δοθείσης κάποιας πληροφορίας για το αποτέλεσμα του τυχαίου πειράματος. Η μερική αυτή πληροφορία εκφράζεται από το ενδεχόμενο B, μεταβάλοντας το δειγματικό χώρο από S σε B και τα ευνοϊκά αποτελέσματα από στοιχεία του A σε στοιχεία του $A \cap B$. Η διαπίστωση αυτή οδηγεί συχνά στον υπολογισμό δεσμευμένων πιθανοτήτων χωρίς τη χρήση του προηγούμενου τύπου.

Το ενδεχόμενο A/B λέγεται δεσμευμένο ενδεχόμενο, ενώ για τη διάκριση της P(A) από την P(A/B) η πρώτη ονομάζεται και απόλυτη ή αδέσμευτη πιθανότητα του ενδεχομένου A. Ο τύπος που ορίζει τη δεσμευμένη πιθανότητα ισοδύναμα γράφεται με τη μορφή:

$$P(A \cap B) = P(A/B)\, P(B),$$

που είναι γνωστή σαν **πολλαπλασιαστικός νόμος**. Ο τύπος αυτός δίνει τη δυνατότητα υπολογισμού της πιθανότητας της τομής δύο ενδεχομένων με τη βοήθεια δεσμευμένης πιθανότητας.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Παράδειγμα 4.6.1: Σε μία τάξη των 100 μαθητών 50 είναι τα αγόρια και 50 τα κορίτσια. 30 από τα αγόρια έχουν αχρωματοψία ενώ 10 είναι ο αντίστοιχος αριθμός για τα κορίτσια. Ένας μαθητής από τους 100 εκλέγεται στην τύχη. Να υπολογιστούν οι πιθανότητες:

α) Ο μαθητής είναι αγόρι δεδομένου ότι έχει αχρωματοψία

β) Έχει αχρωματοψία δεδομένου ότι είναι κορίτσι.

Τα δεδομένα συνοψίζονται στον ακόλουθο πίνακα

	Αχρωματοψ. (Αχρ.)	όχι Αχρωματοψ.	Σύνολο
Αγόρι (Α)	30	20	50
Κορίτσι (Κ)	10	40	50
Σύνολο	40	60	100

$$\alpha)\ P(A/Αχρ.) = \frac{P(A\cap Αχρ)}{P(Αχρ.)} = \frac{\frac{30}{100}}{\frac{40}{100}} = \frac{30}{40} = \frac{3}{4} = 0.75.$$

$$\beta)\ P(Αχρ/K) = \frac{P(Αχρ\cap K)}{P(K)} = \frac{\frac{10}{100}}{\frac{50}{100}} = \frac{10}{50} = \frac{1}{5} = 0.2\ .$$

Οι πιθανότητες αυτές υπολογίζονται και χωρίς την εφαρμογή του τύπου της δεσμευμένης πιθανότητας. Έτσι για τον υπολογισμό της P(Αχρ/Κ) αφού υπάρχει η γνώση ότι εκλέχτηκε κορίτσι ο δειγματικός χώρος είναι πλέον το σύνολο των 50 κοριτσιών και το γεγονός Αχρωματοψία πραγματοποιείται με 10 τρόπους, αφού 10 είναι τα κορίτσια που έχουν αχρωματοψία. Η ζητούμενη πιθανότητα είναι λοιπόν ίση με 10/50 = 1/5.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Παρατήρηση 4.6.1: Απ' τον παραπάνω πίνακα εύκολα υπολογίζεται ότι: P(A) = 50/100 = 0.5 και P(Αχρ) = 40/100 = 0.4. Έτσι P(A/Αχρ) = 0.75 > 0.5 = P(A) ενώ P(Αχρ/K) = 0.2 < 0.4 = P(Αχρ) . Οι συγκρίσεις αυτές οδηγούν στο συμπέρασμα ότι η δεσμευμένη πιθανότητα ενός ενδεχομένου είναι άλλες φορές μεγαλύτερη και άλλες φορές μικρότερη από την αντίστοιχη αδέσμευτη πιθανότητα.

4.7 Ανεξαρτησία Ενδεχομένων

Από το προηγούμενο παράδειγμα και την παρατήρηση που το ακολουθεί προκύπτει ότι η δεσμευμένη πιθανότητα ενός ενδεχομένου είναι άλλοτε μεγαλύτερη και άλλοτε μικρότερη από την αντίστοιχη αδέσμευτη. Υπάρχουν όμως και περιπτώσεις που οι πιθανότητες αυτές είναι ίσες, όπως φαίνεται στο παράδειγμα που ακολουθεί.

Παράδειγμα 4.7.1: Μια τάξη αποτελείται από 40 αγόρια και 60 κορίτσια. Παρατηρούμε ότι 16 αγόρια και 24 κορίτσια φορούν γυαλιά. Ένας μαθητής εκλέγεται στην τύχη. Ποιά η πιθανότητα ο μαθητής αυτός να φορά γυαλιά; Ποιά η πιθανότητα να φορά γυαλιά δεδομένου ότι είναι αγόρι;

Θεωρούμε τα ενδεχόμενα: A = { φορά γυαλιά} και B = {αγόρι}. Τότε

$$P(A/B) = \frac{P(A\cap B)}{P(B)} = \frac{16/100}{40/100} = \frac{16}{40} = 0.4$$

και $$P(A) = \frac{16+24}{100} = \frac{40}{100} = 0.4 \ .$$

Από το παράδειγμα προκύπτει ότι υπάρχουν και περιπτώσεις στις οποίες η δεσμευμένη πιθανότητα ενός ενδεχομένου συμπίπτει με

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

την αντίστοιχη αδέσμευτη. Διαισθητικά αυτό σημαίνει ότι η μερική πληροφορία για το αποτέλεσμα του πειράματος, που εκφράζεται από το ενδεχόμενο B, δεν μεταβάλει στο παραμικρό την πραγματοποίηση του A. Στην περίπτωση αυτή, που η πραγματοποίηση ενός ενδεχομένου B δεν επηρεάζει την πραγματοποίηση ενός άλλου ενδεχομένου A, τα A και B λέγονται ανεξάρτητα. Ο μαθηματικός ορισμός της ανεξαρτησίας ενδεχομένων είναι ο ακόλουθος.

Ορισμός: Δύο ενδεχόμενα A και B λέγονται ανεξάρτητα αν ισχύει η σχέση:

$$P(A\cap B) = P(A)\,P(B)\,.$$

Διαφορετικά τα A και B λέγονται εξαρτημένα.

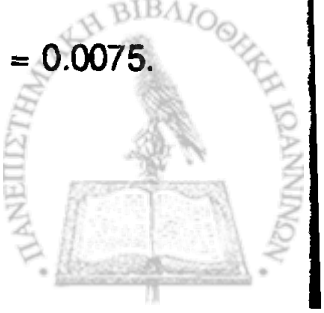
Η έννοια της ανεξαρτησίας επεκτείνεται και για περισσότερα από δύο ενδεχόμενα, ξεφεύγει όμως του σκοπού των σημειώσεων αυτών.

Παράδειγμα 4.7.2: Έχει παρατηρηθεί ότι το ποσοστό των παιδιών προσχολικής ηλικίας που γράφουν με το αριστερό χέρι είναι 15%, 5% είναι δυσλεκτικά και 19.25% παρουσιάζουν τουλάχιστον ένα απ' τα χαρακτηριστικά αυτά. Είναι η δυσλεξία και η γραφή με το αριστερό χέρι ανεξάρτητα χαρακτηριστικά του πληθυσμού των παιδιών προσχολικής ηλικίας;

Θεωρούμε τα ενδεχόμενα: A = {γραφή με αριστερό χέρι}, Δ = {δυσλεξία}. Δίνεται ότι: $P(A) = 0.15$, $P(\Delta) = 0.05$ και $P(A\cup\Delta) = 0.1925$. Για να είναι τα A και Δ ανεξάρτητα πρέπει: $P(A\cap\Delta) = P(A)\,P(\Delta)$.

Είναι $P(A\cup\Delta) = P(A) + P(\Delta) - P(A\cap\Delta)$

ή $P(A\cap\Delta) = P(A) + P(\Delta) - P(A\cup\Delta) = 0.15+0.05-0.1925 = 0.0075.$



Επίσης $P(A)\,P(\Delta) = (0.15)\ (0.05) = 0.0075$.

Παρατηρούμε ότι $P(A\cap\Delta) = 0.0075 = P(A)\,P(\Delta)$ και έτσι η γραφή με το αριστερό χέρι και η δυσλεξία είναι ανεξάρτητα χαρακτηριστικά.

Παράδειγμα 4.7.3: Στον πίνακα που ακολουθεί καταγράφεται η επίδοση 200 φοιτητών στο μάθημα της στατιστικής και η δέσμη την οποία παρακολούθησαν στο Λύκειο.

		ΑΠΟΤΥΧΙΑ (ΑΠ)	ΚΑΛΩΣ (Κ)	ΛΙΑΝ ΚΑΛ. (ΛΚ)	ΑΡΙΣΤΑ (Α)	ΣΥΝΟΛΟ
ΔΕΣΜΗ	Α' (Δ1)	5	3	7	10	25
	Γ' (Δ2)	25	45	40	15	125
	Δ' (Δ3)	10	22	13	5	50
ΣΥΝΟΛΟ		40	70	60	30	200

Ένας φοιτητής εκλέγεται στην τύχη. Να υπολογιστεί η πιθανότητα:

α) Να έχει επίδοση λιάν καλώς. β) Καλώς ή λίαν καλώς

γ) Επίδοση άριστα και να προέρχεται από τη Γ' Δέσμη

δ) Επίδοση λίαν καλώς δεδομένου ότι προέρχεται από την Α' Δέσμη

ε) Είναι η αποτυχία και η προέλευση από τη Γ' Δέσμη ανεξάρτητα ενδοχόμενα;

α) $P(\Lambda K) = \frac{60}{200} = 0.3$, β) $P(K\cup\Lambda K) = \frac{70+60}{200} = \frac{130}{200} = 0.65$,

γ) $P(A\cap\Delta 2) = \frac{15}{200} = 0.075$, δ) $P(\Lambda K/\Delta 1) = \frac{P(\Lambda K\cap\Delta 1)}{P(\Delta 1)} = \frac{7/200}{25/200} = 0.28$,

ε) Είναι $P(A\Pi\cap\Delta 2) = \frac{25}{200} = 0.125 = \frac{40}{200}\,\frac{125}{200} = P(A\Pi)\,P(\Delta 2)$

δηλαδή ανεξάρτητα.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Ασκήσεις 4ου Κεφαλαίου

4.1: Από τα 16 παιδιά που φοιτούν σ' ένα Νηπιαγωγείο τα 7 προέρχονται από το συνοικισμό Α, τα 6 από τον Β και τα 3 από τον Γ. Επιλέγουμε ένα παιδί στην τύχη. Ποιά η πιθανότητα: i) Να είναι από το συνοικισμό Α; ii) Να είναι απ' το συνοικισμό Β; iii) Να είναι απ' τον Γ; iv) Να είναι απ' τον συνοικισμό Α ή τον Β;

4.2: Οι προτιμήσεις 75 μαθητών προσχολικής ηλικίας ως προς τα ομαδικά αθλήματα Ποδόσφαιρο (Π), Μπάσκετ (Μ) και τα ατομικά αθλήματα Κολύμβηση (Κ) και Δρόμος (Δ) καταγράφονται στον πίνακα:

	Άθλημα			
	Π	Μ	Κ	Δ
Αριθμός Μαθητών	30	24	12	9

Ένας μαθητής εκλέγεται στην τύχη. Να υπολογιστούν:
i) Η πιθανότητα να προτιμά το Ποδόσφαιρο. ii) Η πιθανότητα να προτιμά ατομικά αθλήματα iii) Η πιθανότητα να προτιμά το Μπάσκετ δεδομένου ότι προτιμά ομαδικό άθλημα.

4.3: Στον παρακάτω πίνακα δίνεται ο αριθμός των φοιτητών που παρακολουθούσαν τις παραδόσεις του μαθήματος "Στατιστική Ι" το 1990 και η επίδοσή τους στις εξετάσεις του μαθήματος αυτού

	Βαθμός < 5	Βαθμός ≥ 5
παρακολουθούσαν	20	40
Φοιτητές που δεν παρακολουθούσαν	30	10

α) Με βάση τον πίνακα αυτό να υπολογιστεί η πιθανότητα:

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

i) Ένας φοιτητής πέτυχε στις εξετάσεις.

ii) Ένας φοιτητής παρακολουθούσε και πέτυχε.

iii) Ένας φοιτητής πέτυχε δεδομένου ότι παρακολουθούσε

iv) Ένας φοιτητής δεν παρακολουθούσε δεδομένου ότι απέτυχε.

β) Να εξεταστεί κατά πόσο η παρακολούθηση στα μαθήματα και η επιτυχία στις εξετάσεις είναι ή όχι ανεξάρτητα ενδεχόμενα.

4.4: Στον πίνακα που ακολουθεί 502 παιδιά ταξινομούνται ως προς το φύλο και την ομάδα αίματος:

ΦΥΛΟ ΟΜΑΔΑ ΑΙΜΑΤΟΣ	ΑΓΟΡΙ	ΚΟΡΙΤΣΙ	ΣΥΝΟΛΟ
0	113	113	226
A	103	103	206
B	25	25	50
AB	10	10	20
ΣΥΝΟΛΟ	251	251	502

Με βάση τα δεδομένα αυτά είναι η ομάδα αίματος ανεξάρτητη από το φύλο ενός παιδιού; (Η απάντηση να δοθεί μέσω υπολογισμού κατάλληλων πιθανοτήτων).

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΚΕΦΑΛΑΙΟ 5

ΤΥΧΑΙΑ ΜΕΤΑΒΛΗΤΗ -ΚΑΤΑΝΟΜΗ ΠΙΘΑΝΟΤΗΤΑΣ -ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΥΧΑΙΑΣ ΜΕΤΑΒΛΗΤΗΣ ΚΑΙ ΚΑΤΑΝΟΜΗΣ

5.1 Εισαγωγικά

Η θεωρία πιθανοτήτων σκοπό έχει τη δημιουργία μαθηματικών μοντέλων για την περιγραφή φαινομένων, διαδικασιών η έκβαση των οποίων διέπεται από τον παράγοντα τύχη. Στην κατεύθυνση αυτή θεμελιακή είναι η έννοια του τυχαίου πειράματος, του δειγματικού χώρου, του ενδεχομένου και κατακλείδα η έννοια της πιθανότητας, μιας συνολοσυνάρτησης που σε κάθε ενδεχόμενο αντιστοιχίζει ένα αριθμό του διαστήματος [0, 1].

Παρά το θεμελιακό τους χαρακτήρα οι έννοιες που προηγούμενα παρουσιάστηκαν, είναι περιορισμένης έκτασης δεν εκμεταλλεύονται πλήρως το μαθηματικό υπόβαθρο και έτσι είναι αδύνατο να μοντελοποιήσουν και περιγράψουν τα σύνθετα

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

προβλήματα της καθημερινής πραγματικότητας. Η έννοια της τυχαίας μεταβλητής και της κατανομής πιθανότητας, πλουτίζουν την θεωρία πιθανοτήτων με χρήσιμα εργαλεία για την περιγραφή και επίλυση σύνθετων προβλημάτων, και αποτελούν το αντικείμενο αυτού του Κεφαλαίου. Η μέση τιμή και η διακύμανση μιας τυχαίας μεταβλητής, που αποτελούν χαρακτηριστικά της, και η σύνδεσή τους με τις αντίστοιχες έννοιες ενός συνόλου αριθμητικών δεδομένων, θα παρουσιαστούν στο τέλος του κεφαλαίου αυτού.

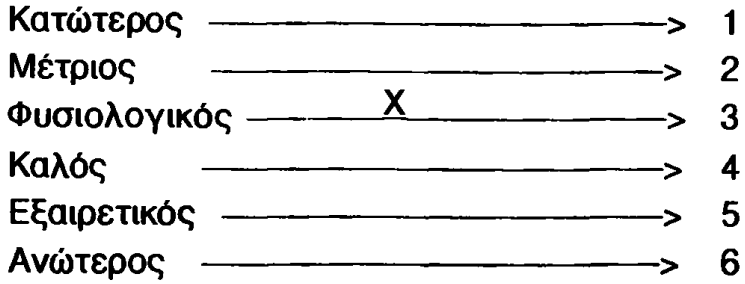
5.2 Τυχαία μεταβλητή

Η μέχρι τώρα περιπλάνηση στον χώρο της θεωρίας πιθανοτήτων διδάσκει ότι ο δειγματικός χώρος ενός τυχαίου πειράματος και τα ενδεχόμενα είναι συνήθως κάποια άμορφα σύνολα με στοιχεία λέξεις ή και ολόκληρες εκφράσεις που χρησιμοποιούνται για να περιγράψουν τα αποτελέσματα ενός τυχαίου πειράματος. Η εξέταση του δείκτη νοημοσύνης ενός παιδιού συνιστά, για παράδειγμα, ένα τυχαίο πείραμα με δειγματικό χώρο το σύνολο: S = {Κατώτερος, Μέτριος, Φυσιολογικός, Καλός, Εξαιρετικός, Ανώτερος}. Τα ενδεχόμενα συνίστανται από μια ή περισσότερες από τις λέξεις αυτές και οι πιθανότητες ενδεχομένων εκφράζονται μέσω των λέξεων αυτών. Η χρησιμοποίηση λέξεων ή και φράσεων για την περιγραφή των αποτελεσμάτων ενός πειράματος έχει το πλεονέκτημα ότι οποιοσδήποτε και ανά πάσα στιγμή έχει τη δυνατότητα εύκολης ενημέρωσης και κατανόησής τους. Αν όμως χρησιμοποιηθεί ο αριθμός 1 για να περιγράψει - κωδικοποιήσει το δείκτη νοημοσύνης "Κατώτερος", ο αριθμός 2 για το δείκτη νοημοσύνης "Μέτριος" κ.ο.κ., ο αριθμός 6 για το δείκτη "Ανώτερος" τότε τα αποτελέσματα του

τυχαίου πειράματος περιγράφονται από τους αριθμούς 1, 2, ... , 6 και ο δειγματικός χώρος είναι πλέον το σύνολο: {1, 2, 3, 4, 5, 6}.

Το μειονέκτημα από τη νέα περιγραφή των αποτελεσμάτων του πειράματος είναι η απώλεια της άμεσης επαφής με τη φύση του υπό μελέτη προβλήματος. Το κέρδος όμως είναι πολλαπλάσιο όχι μόνο γιατί ο δειγματικός χώρος και τα ενδεχόμενα είναι πλέον κομψά σύνολα αριθμών, η πιθανότητα εκφράζεται σύντομα μέσω αριθμών και όχι λέξεων ή φράσεων αλλά κυρίως γιατί ο δειγματικός χώρος γίνεται ένα υποσύνολο των πραγματικών αριθμών. Το σύνολο των πραγματικών αριθμών έχει άριστες μαθηματικές ιδιότητες και μ' αυτό τον τρόπο - τη σύνδεση δηλαδή των θεμελιωδών εννοιών της θεωρίας πιθανοτήτων με το σύνολο των πραγματικών αριθμών - οι μαθηματικές μέθοδοι τίθενται στη διάθεση της θεωρίας πιθανοτήτων. Η σύνδεση αυτή του δειγματικού χώρου ενός πειράματος με ένα υποσύνολο των πραγματικών αριθμών επιτυγχάνεται μέσω της τυχαίας μεταβλητής. Για το προηγούμενο παράδειγμα αν το υπό μελέτη χαρακτηριστικό των παιδιών προσχολικής ηλικίας, ο δείκτης νοημοσύνης, συμβολιστεί με X τότε ο χαρακτηρισμός του ως Κατώτερος, Μέτριος, κ.ο.κ., Ανώτερος μπορεί να γίνει αν χρησιμοποιηθεί η μεταβλητή X με (κωδικές) τιμές 1 όταν ο δείκτης νοημοσύνης είναι Κατώτερος, 2 όταν είναι Μέτριος, κ.ο.κ. 6 όταν ο δείκτης νοημοσύνης είναι Ανώτερος. Έτσι τα ενδεχόμενα: {Κατώτερος}, {Φυσιολογικός}, {Μέτριος, Ανώτερος} γράφονται ισοδύναμα {X=1} ή {1}, {X=3}, ή {3} και {2,6} αντίστοιχα. Από τους περιγραφικούς όρους Κατώτερος κ.ο.κ. Ανώτερος μεταφερόμαστε ισοδύναμα στους αριθμούς 1 κ.ο.κ. 6 αντίστοιχα. Επιτυγχάνται μ' αυτό τον τρόπο μια απεικόνιση του δειγματικού χώρου S του πειράματος στο σύνολο {1, 2, 3, 4, 5, 6} . Η απεικόνιση αυτή γίνεται μέσω της μεταβλητή X, που κωδικοποιεί το υπό μελέτη φαινόμενο ή χαρακτηριστικό, και παριστάνεται με τη μορφή διαγράμματος ως εξής:

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ



Η έννοια της τυχαίας μεταβλητής συνοψίζεται στον ορισμό που ακολουθεί:

Ορισμός: Τυχαία μεταβλητή (random variable) (τ.μ) είναι μια μονοσήμαντη συνάρτηση με πεδίο ορισμού τον δειγματικό χώρο ενός τυχαίου πειράματος και τιμές ένα υποσύνολο των πραγματικών αριθμών.

Η τ.μ. συμβολίζεται συνήθως με κεφαλαία γράμματα X, Y, Z, X_1, Y_1 κλπ. ενώ οι τιμές της με τα αντίστοιχα μικρά x, y, z, x_1, y_1.

Παράδειγμα 5.2.1: Ένας μαθητής απαντά στην τύχη σε τρεις ερωτήσεις που του δίνονται καταφατικά ή αρνητικά. Αν συμβολιστεί με Σ το γεγονός να απαντήσει σωστά σε μια ερώτηση και με Λ το γεγονός να απαντήσει λανθασμένα, τότε οι δυνατές απαντήσεις του μαθητή στα τρία ερωτήματα είναι:

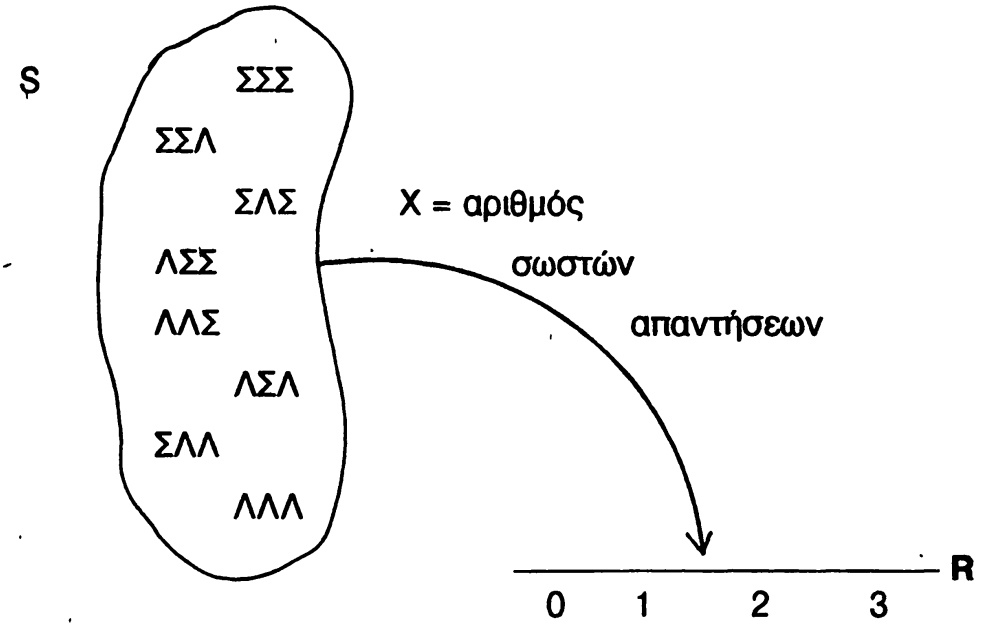
$$S = \{ΣΣΣ, ΣΣΛ, ΣΛΣ, ΛΣΣ, ΛΛΣ, ΛΣΛ, ΣΛΛ, ΛΛΛ\},$$

και συνιστούν το δειγματικό χώρο της διαδικασίας αυτής.

Έστω ότι το ενδιαφέρον επικεντρώνεται στον αριθμό, ας συμβολιστεί με X, των σωστών απαντήσεων. Οι δυνατές τιμές της X είναι: 3 αν εμφανιστεί ΣΣΣ, 2 αν εμφανιστεί ΣΣΛ ή ΣΛΣ ή ΛΣΣ, 1 αν εμφανιστεί ΛΛΣ ή ΛΣΛ ή ΣΛΛ και 0 αν εμφανιστεί το ΛΛΛ. Η X που παριστά τον αριθμό των σωστών απαντήσεων, αντιστοιχεί σε κάθε

ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ ΒΙΒΛΙΟΘΗΚΗ

στοιχείο του S ένα και μόνο ένα στοιχείο του υποσυνόλου {0, 1, 2, 3} των πραγματικών αριθμών, όπως φαίνεται και στο διάγραμμα.



Έτσι η X είναι μια τυχαία μεταβλητή.

Παράδειγμα 5.2.2: Για τα δεδομένα του προηγούμενου παραδείγματος, έστω ότι ο μαθητής κερδίζει μια μονάδα για κάθε σωστή απάντηση και χάνει μια μονάδα για κάθε λανθασμένη απάντηση που δίνει. Το ενδιαφέρον επικεντρώνεται στο κέρδος του μαθητή, ας συμβολιστεί με Y. Οι τιμές του Y είναι: -3 αν απαντήσει ΛΛΛ, -1 αν απαντήσει ΛΛΣ ή ΛΣΛ ή ΣΛΛ, 1 αν απαντήσει ΣΣΛ ή ΣΛΣ ή ΛΣΣ και 3 μονάδες αν οι απαντήσεις είναι ΣΣΣ.

Η Y απεικονίζει το δειγματικό χώρο S στο υποσύνολο {-3, -1, 1, 3} του συνόλου των πραγματικών αριθμών και είναι μια τυχαία μεταβλητή.

Από τα Παραδείγματα 5.2.1 και 5.2.2 προκύπτει ότι στον ίδιο δειγματικό χώρο μπορούν να οριστούν περισσότερες από μια τυχαίες μεταβλητές.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Παράδειγμα 5.2.3: Η διαδικασία μέτρησης του βάρους ενός παιδιού τριών ετών συνιστά τυχαίο πείραμα με δειγματικό χώρο $S = \{s \in \mathbf{R} : 11Kg \leq s \leq 17Kg\}$ αφού έχει παρατηρηθεί ότι το βάρος ενός παιδιού τριών ετών είναι μεταξύ 11 και 17 κιλών (Kg). Αν με X συμβολιστεί το βάρος ενός παιδιού τότε η X είναι μια τυχαία μεταβλητή με τιμές στο διάστημα [11, 17].

Από τα προηγούμενα παραδείγματα προκύπτει ότι το σύνολο τιμών μιας τυχαίας μεταβλητής είναι αριθμήσιμο ή μη αριθμήσιμο. Η παρατήρηση αυτή οδηγεί στη διάκριση των τυχαίων μεταβλητών σε διακριτές και συνεχείς.

5.3 Διακριτές τυχαίες μεταβλητές-Συνάρτηση πιθανότητας

Ορισμός: Μία τυχαία μεταβλητή λέγεται **διακριτή** εάν το σύνολο των τιμών της είναι πεπερασμένο ή αριθμήσιμο.

Οι μεταβλητές των παραδειγμάτων 5.2.1 και 5.2.2 είναι διακριτές . Τα καθημερινά παραδείγματα εφαρμογής των διακριτών τυχαίων μεταβλητών είναι πολλά: ο αριθμός των κλήσεων σ' ένα τηλεφωνικό κέντρο στο διάστημα μιας ημέρας, ο αριθμός των τροχαίων ατυχημάτων στην Ήπειρο σ' ένα μήνα, ο αριθμός των γεννήσεων - θανάτων στην πόλη των Ιωαννίνων σε ένα έτος, ο αριθμός των πελατών ενός super market σε διάστημα δύο ωρών, ο αριθμός των ορθογραφικών λαθών ενός μαθητή σ' ένα δισέλιδο κείμενο.

Γενικά μια διακριτή τυχαία μεταβλητή παίρνει μεμονωμένες τιμές που απέχουν κατά κανόνα το ίδιο η μια από την άλλη.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Παράδειγμα 5.3.1: Θεωρούμε και πάλι την τυχαία μεταβλητή Υ του παραδείγματος 5.2.2, που παριστά το κέρδος του μαθητή. Η Υ είναι η συνάρτηση του

S = {ΛΛΛ, ΛΛΣ, ΛΣΛ, ΣΛΛ, ΣΣΛ, ΣΛΣ, ΛΣΣ, ΣΣΣ}

στο

{ -3 , -1 , 1, 3} .

Οι πιθανότητες με τις οποίες ο μαθητής έχει το αντίστοιχο κέρδος είναι:

$$P(\text{Κέρδος - 3 μονάδες}) = P(Y= -3) = P(\Lambda\Lambda\Lambda)= \frac{1}{8} .$$

$$P(\text{Κέρδος-1 μονάδα}) = P(Y= -1) = P(\Lambda\Lambda\Sigma \text{ ή } \Lambda\Sigma\Lambda \text{ ή } \Sigma\Lambda\Lambda)$$

$$= P(\Lambda\Lambda\Sigma) + P(\Lambda\Sigma\Lambda) + P(\Sigma\Lambda\Lambda)$$

$$= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8} .$$

$$P(\text{Κέρδος 1 μονάδα}) = P(Y= 1) = P(\Sigma\Sigma\Lambda \text{ ή } \Sigma\Lambda\Sigma \text{ ή } \Lambda\Sigma\Sigma)$$

$$= P(\Sigma\Sigma\Lambda) + P(\Sigma\Lambda\Sigma) + P(\Lambda\Sigma\Sigma)$$

$$= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8} .$$

$$P(\text{Κέρδος 3 μονάδες}) = P(Y=3) = P(\Sigma\Sigma\Sigma) = \frac{1}{8} .$$

Το παράδειγμα αυτό υπαινίσσεται το γεγονός ότι η έννοια της τυχαίας μεταβλητής δεν είναι μόνο ένα εργαλείο απλοποιημένης περιγραφής δύσμορφων ενδεχομένων αλλά και ένα εργαλείο υπολογισμού πιθανοτήτων. Από το παράδειγμα φαίνεται ότι σε κάθε τιμή της διακριτής τυχαίας μεταβλητής αντιστοιχεί και μια πιθανότητα, δηλαδή ένας αριθμός του διαστήματος [0,1]. Η αντιστοίχιση αυτή υποδεικνύει τον ορισμό μιας νέας συνάρτησης, που συνοδεύει μια διακριτή τυχαία μεταβλητή, και σε κάθε τιμή της τυχαίας μεταβλητής αντιστοιχεί την πιθανότητα της τιμής αυτής. Η συνάρτηση αυτή

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

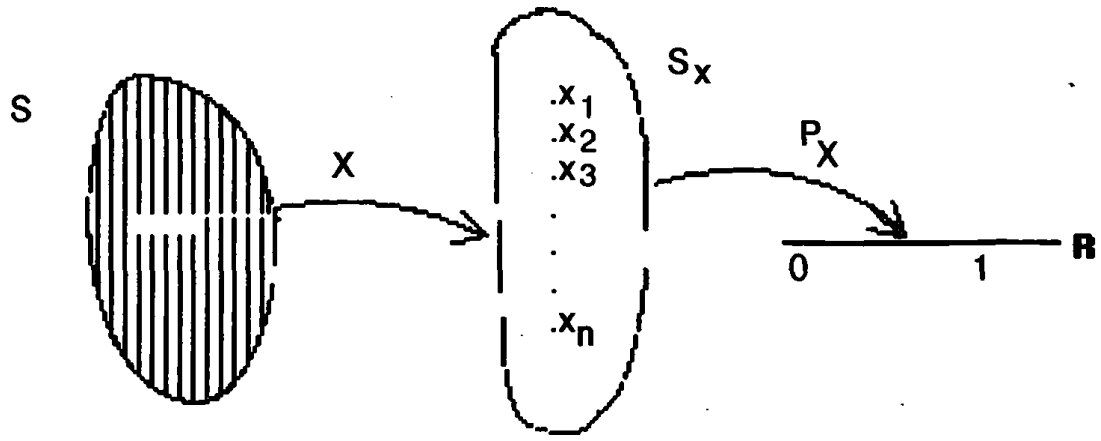
ονομάζεται συνάρτηση πιθανότητας και ο αυστηρός ορισμός της είναι ο ακόλουθος.

Ορισμός: Έστω X μια διακριτή τυχαία μεταβλητή με σύνολο δυνατών τιμών το $S_X = \{x_1, x_2, \ldots, x_n, \ldots, \}$. Η συνάρτηση $p_X : S_X \to [0,1]$ που ορίζεται από τον τύπο

$$p_X(x_i) = P(X=x_i), \quad i = 1, 2, \ldots, n, \ldots.$$

λέγεται **συνάρτηση πιθανότητας** (probability function) (σ.π.) της τυχαίας μεταβλητής X.

Σχηματικά η σχέση που υπάρχει μεταξύ μιας διακριτής τυχαίας μεταβλητής και της αντίστοιχης συνάρτησης πιθανότητάς της δίνεται το διάγραμμα.



Παράδειγμα 5.3.1(Συνέχεια) : Πιο πάνω υπολογίστηκε ουσιαστικά η συνάρτηση πιθανότητας της τυχαίας μεταβλητής Y, του Παραδείγματος 5.2.2, που παριστά το κέρδος του μαθητή. Η συνάρτηση πιθανότητας p_Y συνοπτικά δίνεται ως εξής:

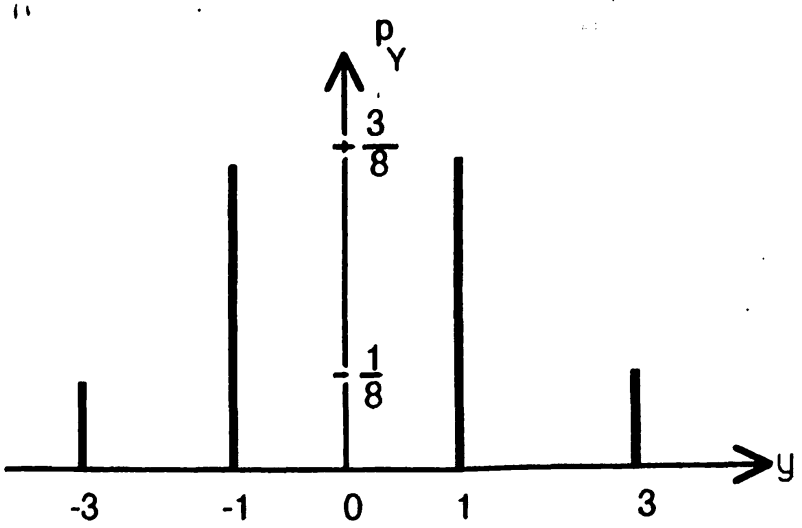
ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

$$p_Y(y)=\begin{cases}\frac{1}{8}\ , & y = -3,3\\ \frac{3}{8}\ , & y = -1,1\end{cases}$$

ή

y	-3	-1	1	3
$p_Y(y)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

και η γραφική της παράσταση είναι:



Από το Παράδειγμα 5.3.1 άμεσα προκύπτουν τα εξής:

i) Η τιμή της συνάρτηση πιθανότητας είναι πάντοτε ένας αριθμός του διαστήματος [0,1]. Δηλαδή

$$0 \le p_Y(y) \le 1\ ,\quad \text{για κάθε τιμή } y \text{ της } Y.$$

ii) Το άθροισμα των τιμών της συνάρτησης πιθανότητας για όλες τις τιμές της τυχαίας μεταβλητής είναι ίσο με το 1. Δηλαδή

$$\sum_y p_Y(y) = 1\ .$$

Οι i) και ii) αποτελούν ιδιότητες της συνάρτησης πιθανότητας.

Η έννοια της τυχαίας μεταβλητής που προηγούμενα παρουσιάστηκε, χρησιμοποιείται για να περιγράψει τα αποτελέσματα και επομένως τα ενδεχόμενα ενός τυχαίου πειράματος, ενώ η έννοια

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

της συνάρτησης πιθανότητας που την ακολουθεί, χρησιμοποιείται για τον υπολογισμό πιθανοτήτων ενδεχομένων που εκφράζονται με τη βοήθεια μιας διακριτής τυχαίας μεταβλητής. Και οι δύο έννοιες αποτελούν το υπόβαθρο για τη δημιουργία πιθανοθεωρητικών μοντέλων που αναφέρονται σε συγκεκριμένους τύπους τυχαίων πειραμάτων. Τα πειράματα αυτά είναι εξιδανικεύσεις φαινομένων και διαδικασιών της καθημερινής πραγματικότητας ενώ τα πιθανοθεωρητικά μοντέλα που αναφέρονται σ' αυτά και ονομάζονται ειδικές κατανομές, βοηθούν στον υπολογισμό πιθανοτήτων χωρίς να ανατρέχουμε κάθε φορά στις βασικές αρχές της θεωρίας πιθανοτήτων. Το διωνυμικό πείραμα και η διωνυμική κατανομή, όπως επίσης η κατανομή Poisson είναι από τις πιο σπουδαίες ειδικές κατανομές διακριτών τυχαίων μεταβλητών.

5.3.1. Διωνυμικό πείραμα - Διωνυμική κατανομή

Πολλά φυσικά φαινόμενα και πολλές καταστάσεις - διαδικασίες της καθημερινότητας έχουν το χαρακτηριστικό γνώρισμα να αποτελούνται από μια σειρά δοκιμών (προσπαθειών) ή επαναλήψεων ενός αρχικού πειράματος του οποίου τα αποτελέσματα είναι μόνο δύο: ένα γεγονός και το αντίθετό του. Για ευκολία ονομάζουμε τα γεγονότα αυτά επιτυχία και αποτυχία και τα συμβολίζουμε με Ε και Α αντίστοιχα. Σε μια τέτοια διαδικασία το ενδιαφέρον επικεντρώνεται στον αριθμό των επιτυχιών που εμφανίζονται στο σύνολο των επαναλήψεων του πειράματος και όχι στο ακριβές αποτέλεσμα της κάθε επανάληψης. Για τη μελέτη της αχρωματοψίας παιδιών προσχολικής ηλικίας τριακόσια παιδιά εκλέγονται και εξετάζονται. Οι δοκιμές του πειράματος είναι η εξέταση των παιδιών. Επιτυχία είναι το γεγονός ένα παιδί να μη πάσχει από αχρωματοψία. Το ενδιαφέρον

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

εντοπίζεται στην κατανομή του αριθμού των παιδιών που δεν παρουσιάζουν αχρωματοψία.

Σε μια μελέτη για το κάπνισμα των μαθητών Λυκείου, 50 μαθητές εκλέγονται και κάθε ένας εξετάζεται για το αν καπνίζει ή όχι. Η εξέταση ενός μαθητή αποτελεί μια εκτέλεση του πειράματος και επιτυχία θεωρείται το γεγονός ο μαθητής να μη καπνίζει. Το ενδιαφέρον επικεντρώνεται και πάλι στην κατανομή του αριθμού των επιτυχιών, στην κατανομή του αριθμού των παιδιών που δεν καπνίζουν.

Το **διωνυμικό τυχαίο πείραμα** αποτελεί εξιδανίκευση διαδικασιών, όπως αυτές που περιγράφονται στα προηγούμενα παραδείγματα, και πληρεί τις ακόλουθες προϋποθέσεις:

i) Αποτελείται από ένα προκαθορισμένο αριθμό n επαναλήψεων, δοκιμών ενός αρχικού πειράματος.

ii) Κάθε δοκιμή, του αρχικού πειράματος, έχει δύο δυνατά αποτελέσματα. Το ένα από αυτά ονομάζεται επιτυχία (Ε) και το άλλο αποτυχία (Α).

iii) Η πιθανότητα επιτυχίας, συμβολίζεται με p, και παραμένει σταθερή σε κάθε επανάληψη του πειράματος. Η πιθανότητα αποτυχίας είναι 1- p και συμβολίζεται με q.

iv) Οι n επαναλήψεις του πειράματος είναι ανεξάρτητες, με την έννοια ότι το αποτέλεσμα μιας δοκιμής δεν έχει καμιά επίδραση στο αποτέλεσμα οποιασδήποτε άλλης.

Ένα τυχαίο πείραμα που πληρεί τις προϋποθέσεις i) - iv) λέγεται **διωνυμικό τυχαίο πείραμα** ενώ κάθε επιμέρους δοκιμή λέγεται **δοκιμή Bernoulli** και p είναι η πιθανότητα επιτυχίας σε κάθε δοκιμή.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Αν X είναι η τυχαία μεταβλητή που παριστά τον αριθμό των επιτυχιών στην ακολουθία των n δοκιμών Bernoulli, η X είναι διακριτή με τιμές 0, 1, 2, ... , n και η συνάρτηση πιθανότητάς της είναι:

$$p_X(x) = \binom{n}{x} p^x q^{n-x}, \quad x=0, 1, ..., n, \quad q=1-p,$$

όπου $\binom{n}{x}$ είναι οι συνδιασμοί των n στοιχείων ανά x και ορίζονται:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \quad \text{με } n! = 1.2.3. ....(n-1)\, n \text{ και } 0! = 1.$$

Η τυχαία μεταβλητή X λέγεται **διωνυμική τυχαία μεταβλητή** και συμβολίζεται X~ B(n,p) ενώ η συνάρτηση πιθανότητάς της λέγεται **διωνυμική κατανομή** (binomial distribution). Τα n και p είναι οι παράμετροι της κατανομής.

Παράδειγμα 5.3.1.1 Είναι γνωστό ότι το 10% των μαθητών ενός δημοτικού σχολείου παρουσιάζουν χαρακτηριστικά δυσλεξίας. 9 μαθητές εκλέγονται στην τύχη και εξετάζονται. Ποιά η πιθανότητα 3 από αυτούς να παρουσιάζουν χαρακτηριστικά δυσλεξίας;

Η εξέταση των 9 μαθητών που εκλέχτηκαν συνιστά επανάληψη μιας διαδικασίας με δύο δυνατά αποτελέσματα: Ε να είναι ένας μαθητής δυσλεκτικός και Α να μην είναι δυσλεκτικός. Αν p είναι η P(E) τότε p = 10% = 0.1 και παραμένει αμετάβλητη για κάθε μαθητή. Το πρόβλημα εντοπίζεται στην εύρεση της πιθανότητας συγκεκριμένου αριθμού επιτυχιών. Αν X είναι η τυχαία μεταβλητή που παριστά τον αριθμό των δυσλεκτικών μαθητών στους 9 που εξετάστηκαν τότε X~B(9, 0.1). Ζητείται η

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

P(3 να είναι δυσλεκτικοί) = P(X=3) = $p_X(3)$.

Η τιμή $p_X(3)$ υπολογίζεται με δύο τρόπους:

α) **Εφαρμογή του τύπου:** $p_X(x) = \binom{n}{x} p^x q^{n-x}$.

Από τα δεδομένα: n = 9, x = 3, p = 0.1, q = 1-p = 0.9. Έτσι

$$p_X(3) = \binom{n}{x} (0.1)^3 (0.9)^{9-3} = \frac{9!}{3!6!} (0.1)^3 (0.9)^6$$

$$= \frac{1.2.3.4.5.6.7.8.9}{1.2.3.1.2.3.4.5.6} (0.1)^3 (0.9)^6$$

$$= \frac{7.8.9}{1.2.3} (0.1)^3 (0.9)^6 = (84)\ (0.001)\ (0.53) = 0.0445.$$

Έτσι η ζητούμενη πιθανότητα είναι 0.0445.

β) **Χρήση πινάκων**: Στο Παράρτημα, δίνονται πίνακες της διωνυμικής κατανομής στους οποίους για δεδομένες τιμές των n,p και x της διωνυμικής κατανομής δίνεται η αντίστοιχη πιθανότητα.Τμήμα των πινάκων αυτών είναι ο ακόλουθος πίνακας.

		p									
n	x	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
9	0	.6302	.3874	.2316	.1342	.0751	.0404	.0207	.0101	.0046	.0020
	1	.2985	.3874	.3679	.3020	.2253	.1556	.1004	.0605	.0339	.0176
	2	.0629	.1722	.2597	.3020	.3003	.2668	.2162	.1612	.1110	.0703
	3	.0077	.0446	.1069	.1762	.2336	.2668	.2716	.2508	.2119	.1641
	4	.0006	.0074	.0283	.0661	.1168	.1715	.2194	.2508	.2600	.2461
	5	.0000	.0008	.0050	.0165	.0389	.0735	.1181	.1672	.2128	.2461
	6	.0000	.0001	.0006	.0028	.0087	.0210	.0424	.0743	.1160	.1641
	7	.0000	.0000	.0000	.0003	.0012	.0039	.0098	.0212	.0407	.0703
	8	.0000	.0000	.0000	.0000	.0001	.0004	.0013	.0035	.0083	.0176
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0008	.0020
10	0	.5987	.3487	.1969	.1074	.0563	.0282	.0135	.0060	.0025	.0010
	1	.3151	.3874	.3474	.2684	.1877	.1211	.0725	.0403	.0207	.0098
	2	.0746	.1937	.2759	.3020	.2816	.2335	.1757	.1209	.0763	.0439
	3	.0105	.0574	.1298	.2013	.2503	.2668	.2522	.2150	.1665	.1172
	4	.0010	.0112	.0401	.0881	.1460	.2001	.2377	.2508	.2384	.2051
	5	.0001	.0015	.0085	.0264	.0584	.1029	.1536	.2007	.2340	.2461
	6	.0000	.0001	.0012	.0055	.0162	.0368	.0689	.1115	.1596	.2051
	7	.0000	.0000	.0001	.0008	.0031	.0090	.0212	.0425	.0746	.1172
	8	.0000	.0000	.0000	.0001	.0004	.0014	.0043	.0106	.0229	.0439
	9	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0016	.0042	.0098
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010
11	0	.5688	.3138	.1673	.0859	.0422	.0198	.0088	.0036	.0014	.0005
	1	.3293	.3835	.3248	.2362	.1549	.0932	.0518	.0266	.0125	.0054

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Για το παράδειγμά μας, εντοπίζεται στην πρώτη στήλη του πίνακα η τιμή του n, ίση με το 9. Γι' αυτή την τιμή του n αναζητείται στη δεύτερη στήλη η τιμή του x, στην περίπτωσή μας ίση με 3. Εντοπίζεται στην πρώτη γραμμή η τιμή του p που για το παράδειγμά μας είναι το 0.1. Η ζητούμενη πιθανότητα βρίσκεται στη διασταύρωση της γραμμής που περνά από το x και της στήλης που περνά από το p και είναι η 0.0446.

Παράδειγμα 5.3.1.2: Το 20% ενός πληθυσμού πάσχει από αχρωματοψία. Από 12 άτομα που εκλέγονται από τον πληθυσμό αυτό ποιά η πιθανότητα: α) Τουλάχιστον 9 έχουν αχρωματοψία. β) Το πολύ 3 έχουν αχρωματοψία. γ) Μεταξύ 5 και 8 έχουν αχρωματοψία.

Αν X ο αριθμός των ατόμων που έχουν αχρωματοψία στους 12 που εκλέχτηκαν τότε $X \sim B(12, 0.2)$. Ζητείται:

α) $P(\text{τουλ. 9 αχρωμ.}) = P(X=9 \text{ ή } X=10 \text{ ή } X=11 \text{ ή } X=12)$

$= p_X(9) + p_X(10) + p_X(11) + p_X(12)$

$= 0.0001 + 0.0000 + 0.0000 + 0.0000 = 0.0001.$

β) $P(\text{το πολύ 3 αχρωμ.}) = P(X=0 \text{ ή } X=1 \text{ ή } X=2 \text{ ή } X=3)$

$= p_X(0) + p_X(1) + p_X(2) + p_X(3)$

$= 0.0687 + 0.2062 + 0.2835 + 0.2362$

$= 0.7946.$

γ) $P(\text{Μεταξύ 5 και 8 έχουν αχρωμ.}) = P(5 \leq X \leq 8)$

$= P(X=5 \text{ ή } X=6 \text{ ή } X=7 \text{ ή } X=8)$

$= p_X(5) + p_X(6) + p_X(7) + p_X(8)$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

$$= 0.0532 + 0.0155 + 0.0033 + 0.0005$$

$$= 0.0725.$$

5.3.2 Κατανομή Poisson

Η κατανομή Poisson, όπως και η διωνυμική κατανομή που προηγούμενα παρουσιάστηκε, είναι από τις πιο σημαντικές κατανομές της θεωρίας πιθανοτήτων με ευρύτατο πεδίο εφαρμογών. Εισήχθη από το Γάλλο μαθηματικό Poisson το 1837, στον οποίο οφείλει και το όνομά της, και μας εφοδιάζει με ένα ρεαλιστικό μοντέλο για την περιγραφή φαινομένων - διαδικασιών που εξελίσσονται με την πάροδο του χρόνου. Ειδικότερα η κατανομή Poisson εφαρμόζεται σε κάθε περίπτωση που τυχαία ενδεχόμενα - συμβάντα εμφανίζονται σε διάφορες χρονικές στιγμές και το ενδιαφέρον εντοπίζεται στην απαρίθμιση των συμβάντων που πραγματοποιούνται σε δεδομένο χρονικό διάστημα. Το χρονικό αυτό διάστημα μπορεί να είναι οποιουδήποτε μήκους, όπως ενός λεπτού, μιας ημέρας, μιας βδομάδας, ενός μήνα κλπ. Εκτός από χρόνο το διάστημα αυτό μπορεί να παριστάνει μήκος, επιφάνεια, όγκο κ.ο.κ.

Με βάση τα προηγούμενα η κατανομή Poisson μπορεί να χρησιμοποιηθεί για να περιγράψει: τον αριθμό των κλίσεων που φτάνουν σ' ένα τηλεφωνικό κέντρο στο διάστημα μιας ώρας, το πλήθος των τροχαίων ατυχημάτων στο νομό Ιωαννίνων το 1990, τον αριθμό των θανάτων τον προηγούμενο μήνα, τον αριθμό των βακτηριδίων που παρατηρούνται με ένα μικροσκόπιο πάνω σε μιά πλάκα ορισμένου εμβαδού, τον αριθμό των μαθητών που αρρώστησαν από γρίπη τον προηγούμενο χειμώνα, τον αριθμό των αεροπλάνων που προσγειώνονται σ' ένα αεροδόμιο σε διάστημα δύο ωρών, τον αριθμό των πελατών ενός super market σε διάστημα μιας μέρας και γενικότερα

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

τον αριθμό των τυχαίων ενδεχομένων που πραγματοποιούνται σε ένα διάστημα που παριστά χρόνο, μήκος, επιφάνεια, όγκο.

Θεωρούμε διαδικασίες, όπως οι προηγούμενες, που ικανοποιούν τις ακόλουθες προϋποθέσεις:

i) Η πιθανότητα πραγματοποίησης ενός τυχαίου ενδεχομένου σ' ένα μικρό χρονικό διάστημα είναι ανάλογη με το μήκος του χρονικού διαστήματος.

ii) Η πιθανότητα πραγματοποίησης περισσοτέρων του ενός ενδεχομένων σ' ένα μικρό χρονικό διάστημα είναι πολύ μικρή.

iii) Ο αριθμός των τυχαίων ενδεχομένων που πραγματοποιούνται σ' ένα χρονικό διάστημα, δεν έχει καμιά επίδραση στον αριθμό των ενδεχομένων που πραγματοποιούνται σε οποιοδήποτε άλλο, ξένο προς το προηγούμενο χρονικό διάστημα.

Αν για μια διαδικασία που εξελίσσεται στο χρόνο οι προηγούμενες υποθέσεις - εξιδανικεύσεις ικανοποιούνται και X είναι η τυχαία μεταβλητή που παριστά τον αριθμό των τυχαίων ενδεχομένων που πραγματοποιούνται σε ένα διάστημα χρόνου (ή μήκους ή επιφάνειας) η X ακολουθεί **κατανομή Poisson** (Poisson distribution), με μια παράμετρο λ, συμβολίζεται με $X \sim P(\lambda)$ και η συνάρτηση πιθανότητας της είναι:

$$p_X(x) = \frac{e^{-\lambda}\,\lambda^x}{x!}\ ,\quad x = 0, 1, 2, \ldots$$

όπου $e = 2.718282$ η βάση των νεπέριων λογαρίθμων.

Το λ είναι ένας σταθερός αριθμός, ονομάζεται παράμετρος της κατανομής Poisson και παριστά το μέσο όρο των ενδεχομένων που πραγματοποιούνται σ' ένα διάστημα. Αν μια τυχαία μεταβλητή ακολουθεί την κατανομή Poisson, τότε η παράμετρος λ έχει την

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ιδιότητα να ακολουθεί αναλογικά το μήκος του διαστήματος που ορίστηκε αρχικά. Αν για παράδειγμα ο μέσος αριθμός γεννήσεων λ σε ένα μήνα είναι 100 γεννήσεις, τότε σε δύο μήνες θα είναι 200 γεννήσεις και σε μισό μήνα 50 γεννήσεις.

Παράδειγμα 5.3.2.1 Ο μέσος όρος των γονέων που έρχονται να εγγράψουν τα παιδιά τους στο νηπιαγωγείο κατά την πρώτη ημέρα των εγγραφών είναι 30 ανά ώρα. Με την προϋπόθεση ότι η προσέλευση των γονέων ακολουθεί την κατανομή Poisson, να βρεθεί η πιθανότητα να έρθουν δυο τουλάχιστον γονείς σε διάστημα ενός λεπτού.

Αν συμβολίσουμε με X την τυχαία μεταβλητή που παριστά τον αριθμό των γονέων που προσέρχονται για την εγγραφή του παιδιού τους σε διάστημα ενός λεπτού, τότε η $X \sim P(\lambda)$. Η παράμετρος λ είναι ο μέσος όρος των γονέων που έρχονται στη μονάδα του χρόνου. Αν ως μονάδα χρόνου ληφθεί η ώρα τότε το $\lambda = 30$ γονείς ανά ώρα και αναλογικά το $\lambda = 1/2$ γονείς ανά λεπτό αφού τελικά ενδιαφέρει η προσέλευση δύο τουλάχιστον γονέων σε διάστημα ενός λεπτού. Έτσι $X \sim P(1/2)$. Ζητείται η πιθανότητα:

P(τουλάχιστον 2 γονείς σε ένα λεπτό) = $P(X \geq 2)$ =

$= 1-P(X<2) = 1-P(X=0 \text{ ή } X=1) =$

$= 1-P(X=0) - P(X=1) = 1-p_x(0) - p_x(1).$

Ο υπολογισμός των $p_x(0)$ και $p_x(1)$ γίνεται με δύο τρόπους:

α) **Από τον τύπο** $p_x(x) = e^{-\lambda} \lambda^x / x!$, $x = 0,1,...$

Έτσι $p_x(0) = \frac{e^{-1/2}\,(\frac{1}{2})^0}{0!} = e^{-1/2} = 0.6065$ και $p_x(1) = \frac{e^{-1/2}\,(\frac{1}{2})^1}{1!}$ $= \frac{1}{2}\, e^{-1/2} = 0.3032$ και η ζητούμενη πιθανότητα είναι: 1- 0.6065 - 0.3032 = 0.0903.

β) **Χρήση πινάκων:** Στο Παράρτημα, δίνονται πίνακες της Poisson κατανομής, μέρος των οποίων ακολουθεί, στους οποίους για δεδομένες τιμές των λ και x δίνεται η αντίστοιχη πιθανότητα.

						λ				
x	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
0	.9048	.8187	.7408	.6703	.6065	.5488	.4966	.4493	.4066	.3679
1	.0905	.1637	.2222	.2681	.3033	.3293	.3476	.3595	.3659	.3679
2	.0045	.0164	.0333	.0536	.0758	.0988	.1217	.1438	.1647	.1839
3	.0002	.0011	.0033	.0072	.0126	.0198	.0284	.0383	.0494	.0613
4	.0000	.0001	.0002	.0007	.0016	.0030	.0050	.0077	.0111	.0153
5	.0000	.0000	.0000	.0001	.0002	.0004	.0007	.0012	.0020	.0031
6	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0003	.0005
7	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

						λ				
x	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0	.3329	.3012	.2725	.2466	.2231	.2019	.1827	.1653	.1496	.1353
1	.3662	.3614	.3543	.3452	.3347	.3230	.3106	.2975	.2842	.2707
2	.2014	.2169	.2303	.2417	.2510	.2584	.2640	.2678	.2700	.2707
3	.0738	.0867	.0998	.1128	.1255	.1378	.1496	.1607	.1710	.1804
4	.0203	.0260	.0324	.0395	.0471	.0551	.0636	.0723	.0812	.0902
5	.0045	.0062	.0084	.0111	.0141	.0176	.0216	.0260	.0309	.0361
6	.0008	.0012	.0018	.0026	.0035	.0047	.0061	.0078	.0098	.0120
7	.0001	.0002	.0003	.0005	.0008	.0011	.0015	.0020	.0027	.0034
8	.0000	.0000	.0001	.0001	.0001	.0002	.0003	.0005	.0006	.0009
9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0002

						λ				
x	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
0	.1225	.1108	.1003	.0907	.0821	.0743	.0672	.0608	.0550	.0498
1	.2572	.2438	.2306	.2177	.2052	.1931	.1815	.1703	.1596	.1494
2	.2700	.2681	.2652	.2613	.2565	.2510	.2450	.2384	.2314	.2240
3	.1890	.1966	.2033	.2090	.2138	.2176	.2205	.2225	.2237	.2240
4	.0992	.1082	.1169	.1254	.1336	.1414	.1488	.1557	.1622	.1680

Για τον υπολογισμό της $p_x(0)$ (οπότε x = 0 και λ = 1/2 = 0.5) εντοπίζεται στην πρώτη στήλη του πίνακα η τιμή του x ίση με 0. Εντοπίζεται στην πρώτη γραμμή η τιμή του λ ίση με 0.5. Η τιμή $p_x(0)$ βρίσκεται στη διασταύρωση της γραμμής που περνά από το x και της στήλης που περνά από το λ και είναι: $p_x(0) = 0.6065$. Όμοια βρίσκεται ότι $p_x(1) = 0.3033$ και η ζητούμενη πιθανότητα είναι 0.0902.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

5.4 Συνεχείς τυχαίες μεταβλητές - Συνάρτηση πυκνότητας πιθανότητας

Ορισμός: Μια τυχαία μεταβλητή λέγεται συνεχής αν το σύνολο των τιμών της είναι μη αριθμήσιμο.

Παραδείγματα συνεχών τυχαίων μεταβλητών είναι: ο χρόνος ζωής του ιού της γρίππης σ' ένα παιδί προσχολικής ηλικίας, το βάρος ενός νεογέννητου, το ύψος ενός ενήλικα, η χρονική στιγμή κατά την οποία ένα τραίνο φτάνει σ' ένα καθορισμένο σταθμό κ.ο.κ. Οι τυχαίες μεταβλητές που χρησιμοποιούνται για την περιγραφή των μεγεθών των πιο πάνω παραδειγμάτων παίρνουν τιμές σε διαστήματα της ευθείας των πραγματικών αριθμών και επομένως είναι συνεχείς.

Έστω X η τυχαία μεταβλητή που παριστά το βάρος ενός νεογέννητου. Οι τιμές της X μπορεί θεωρητικά να είναι οποιοσδήποτε αριθμός μεγαλύτερος του 0. Έχει όμως παρατηρηθεί ότι το βάρος ενός νεογέννητου είναι μεταξύ 1.7 Kg και 5 Kg. Έτσι η τυχαία μεταβλητή X παίρνει τιμές στο διάστημα [1.7Kg, 5Kg] και είναι συνεχής. Στενά συνδεδεμένη με μια διακριτή τυχαία μεταβλητή είναι η συνάρτηση πιθανότητας της, που σε κάθε τιμή της τυχαίας μεταβλητής αντιστοιχεί και μια πιθανότητα. Ισχύει το ίδιο και για μια συνεχή τυχαία μεταβλητή; Αντιστοιχεί σε κάθε τιμή της και μια πιθανότητα; Πιά είναι η πιθανότητα το βάρος ενός νεογέννητου να είναι 3.5Kg; Στο ερώτημα αυτό με τα μέχρι τώρα γνωστά εργαλεία για τον υπολογισμό πιθανοτήτων είναι αδύνατο να απαντήσουμε. Ακόμη και αν αποφύγουμε την μαθηματική αυστηρότητα και χρησιμοποιήσουμε τον κλασσικό ορισμό της πιθανότητας - ας θυμηθούμε ότι ο κλασσικός ορισμός εφαρμόζεται σε τυχαία πειράματα με πεπερασμένο πλήθος ισοπίθανων αποτελεσμάτων - η ζητούμενη πιθανότητα θα ήταν:

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

$P(X=3.5Kg) = \frac{1}{\infty} = 0$ και γενικότερα $P(X=x) = \frac{1}{\infty} = 0$, για κάθε τιμή $x \in [1.7, 5]$ της τυχαίας μεταβλητής X. Έτσι η συνεχής τυχαία μεταβλητή X μπορεί να πάρει οποιαδήποτε τιμή του διαστήματος [1.7, 5] αλλά η πιθανότητα της τιμής αυτής είναι ίση με μηδέν. Από όσα έχουν αναφερθεί φαίνεται ότι υπάρχει κάποια αντινομία ή ότι έχουμε οδηγηθεί σε αδιέξοδο. Είναι όμως έτσι ή η τυχαία μεταβλητή του παραδείγματος αυτού από τη φύση του μεγέθους που περιγράφει απαιτεί διαφορετικό χειρισμό;

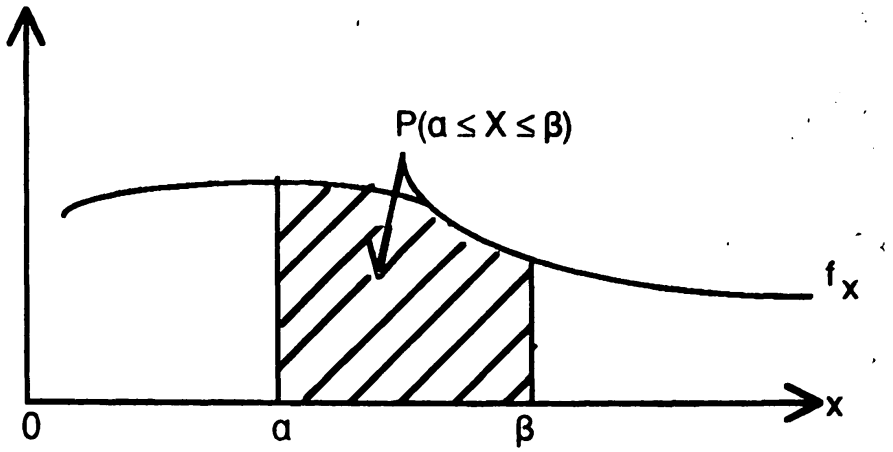
Όταν αναφερόμαστε στο πλήθος των γεννήσεων στην πόλη Ιωαννίνων το 1990 είναι σαφές ότι εννοούμε ένα ακέραιο αριθμό και το γεγονός ο αριθμός αυτός είναι π.χ. 2500 δεν επιδέχεται καμιά άλλη ερμηνεία. Όταν όμως ζητάμε το βάρος ενός νεογέννητου να είναι 3.5Kg και η πλέον τέλεια ζυγαριά και αν χρησιμοποιηθεί δεν είναι δυνατόν να μας πείσει ότι το βάρος είναι ακριβώς 3.5Kg λόγω σφαλμάτων στη μέτρηση αλλά και σφαλμάτων που προκύπτουν από τη στρογγύλευση στο 3.5Kg. Έτσι κατά τη μέτρηση του βάρους του νεογέννητου ενώ πιστεύουμε ότι έχουμε μετρήσει 3.5Kg, στην πραγματικότητα έχουμε μετρήσει ένα αριθμό του διαστήματος (3.5-ε , 3.5+ε) με ε θετικό οσοδήποτε μικρό αριθμό.

Γενικεύοντας τον προηγούμενο συλλογισμό, για συνεχείς τυχαίες μεταβλητές είναι ρεαλιστικό να αναφερόμαστε σε πιθανότητες διαστημάτων, οσοδήποτε μικρών, στα οποία παίρνει τιμές η τυχαία μεταβλητή και όχι σε πιθανότητες συγκεκριμένων της τιμών. Έτσι το πρόβλημα ανάγεται στην εύρεση τρόπου που να επιτρέπει τον ορισμό πιθανοτήτων διαστημάτων στα οποία παίρνει τιμές η τυχαία μεταβλητή. Αυτό μπορεί να επιτευχθεί μέσω μιας μη αρνητικής συνεχούς συνάρτησης, ας τη συμβολίσουμε με f_X, τέτοιας ώστε για κάθε διάστημα [α, β] της πραγματικής ευθείας να ισχύει:

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

$P(\alpha \le X \le \beta)$ = εμβαδό του χωρίου που δημιουργείται από το γράφημα της f_X, τον άξονα των x, και των καθέτων που άγονται στον άξονα αυτό στα σημεία α και β.

Γραφικά το εμβαδό αυτό είναι η γραμμοσκιασμένη περιοχή του παρακάτω σχήματος.



Ο αυστηρός ορισμός της συνάρτησης f_X είναι ο ακόλουθος:

Ορισμός. Έστω X μία συνεχής τυχαία μεταβλητή. Αν υπάρχει μια μη αρνητική συνεχής συνάρτηση $f_X(x)$, με πεδίο ορισμού το πεδίο τιμών της X, τέτοια ώστε η $P(\alpha \le X \le \beta)$ να είναι ίση με το εμβαδό του χωρίου που δημιουργείται από το γράφημα της f_X, τον άξονα των x και των καθέτων που άγονται σ' αυτόν στα σημεία α και β τότε η f_X λέγεται **συνάρτηση πυκνότητας πιθανότητας** (probability density function) της X.

Οι ιδιότητες της συνάρτησης πυκνότητας πιθανότητας μιας συνεχούς τυχαίας μεταβλητής είναι ανάλογες με εκείνες της συνάρτησης πιθανότητας μιας διακριτής τυχαίας μεταβλητής. Έτσι

i) Η τιμή της συνάρτησης πυκνότητας πιθανότητας είναι ένας μη αρνητικός αριθμός.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ii) Το έμβαδόν του χωρίου που σχηματίζεται από το γράφημα της f_X και τον άξονα των x είναι ίσο με τη μονάδα.

Μια συνεχής τυχαία μεταβλητή χρησιμοποιείται για να περιγράψει αποτελέσματα ενός τυχαίου πειράματος. Η συνάρτηση πυκνότητας πιθανότητας που ακολουθεί μια συνεχή τυχαία μεταβλητή είναι το εργαλείο, το μέσο για τον υπολογισμό πιθανοτήτων ενδεχομένων που εκφράζονται με τη βοήθεια της αντίστοιχης συνεχούς τυχαίας μεταβλητής. Ο υπολογισμός πιθανοτήτων με τη χρήση της συνάρτησης πυκνότητας πιθανότητας προϋποθέτει την ύπαρξη και γνώση μεθοδολογιών για τον υπολογισμό εμβαδών χωρίων που δημιουργούνται από το γράφημα της συνάρτησης πυκνότητας πιθανότητας και τον οριζόντιο άξονα. Ο υπολογισμός των εμβαδών αυτών επιτυγχάνεται με την ολοκλήρωση της f_X στο διάστημα [α, β] (συμβολικά $\int_{\alpha}^{\beta} f_X(x)\,dx$) και έτσι

$$P(\alpha \le X \le \beta) = \int_{\alpha}^{\beta} f_X(x)\,dx .$$

Η ολοκλήρωση είναι το συνεχές ανάλογο της άθροισης. Μεθοδολογίες για τον υπολογισμό ολοκληρωμάτων, όπως το προηγούμενο, είναι διαθέσιμες ξεφεύγουν όμως από το σκοπό των σημειώσεων αυτών.

5.4.1. Κανονική κατανομή ή Κατανομή Gauss

Ανάλογη σπουδαιότητα με τη διωνυμική κατανομή και την κατανομή Poisson, της διακριτής περίπτωσης, έχει η κανονική κατανομή για τη συνεχή περίπτωση. Η σπουδαιότητά της οφείλεται: i) στο γεγονός ότι περιγράφει αρκετά καλά μεταβλητές όπως το βάρος, το ύψος, το δείκτη νοημοσύνης, την επίδοση σ' ένα τέστ, την πίεση του

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

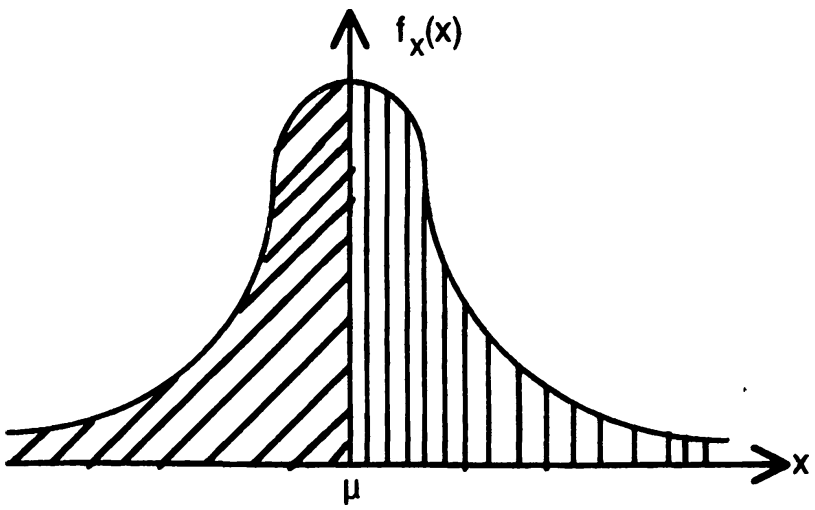
αίματος κλπ. και ii) στο γεγονός ότι κάτω από ορισμένες προϋποθέσεις, προσφέρει ικανοποιητική προσέγγιση της κατανομής του αθροίσματος τυχαίων μεταβλητών. Εισάγεται από το Γάλλο μαθηματικό Abraham de Moivre το 1733, αξιοποιείται όμως στις εργασίες του Gauss το 1809 στον οποίο οφείλει και το όνομά της.

Μια συνεχής τυχαία μεταβλητή X λέγεται ότι ακολουθεί **κανονική κατανομή** (normal distribution) με παραμέτρους μ και σ^2 και συμβολίζεται: $X \sim N(\mu, \sigma^2)$ αν η συνάρτηση πυκνότητας πιθανότητας της X είναι:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

$-\infty < \mu < \infty$, $\sigma^2 > 0$. Τα π και e είναι οι σταθερές 3.14159 και 2.718282 που συχνά συναντώνται στα μαθηματικά.

Η γραφική παράσταση της f_X είναι:



Οι χαρακτηριστικές **ιδιότητες της κανονικής κατανομής** είναι οι ακόλουθες:

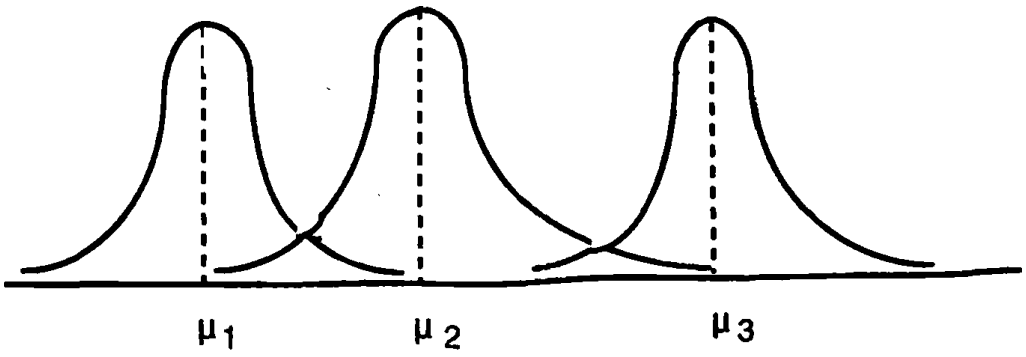
1. Είναι συμμετρική γύρω από την παράμετρο μ.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

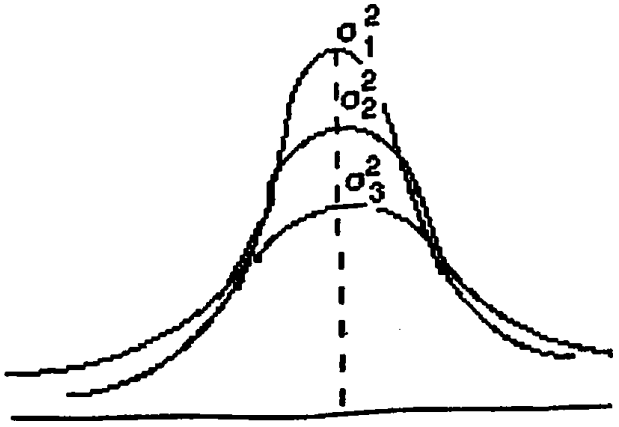
2. Επειδή η f_X είναι συνάρτηση πυκνότητας πιθανότητας το εμβαδό του χωρίου που δημιουργείται από το γράφημα της f_X και του οριζόντιου άξονα είναι ίσο με 1.

3. Λόγω των προηγούμενων ιδιοτήτων, τα εμβαδά των χωρίων που δημιουργούνται αριστερά και δεξιά της καθέτου στον άξονα των x στο σημείο μ είναι ίσα με 0.5.

4. Η κανονική κατανομή προσδιορίζεται πλήρως από τις παραμέτρους μ και σ^2. Διαφορετικές τιμές της μ μετατοπίζουν το γράφημα της f_X στον άξονα των x ενώ διαφορετικές τιμές της σ^2 διαφοροποιούν την επιπεδότητα της f_X .



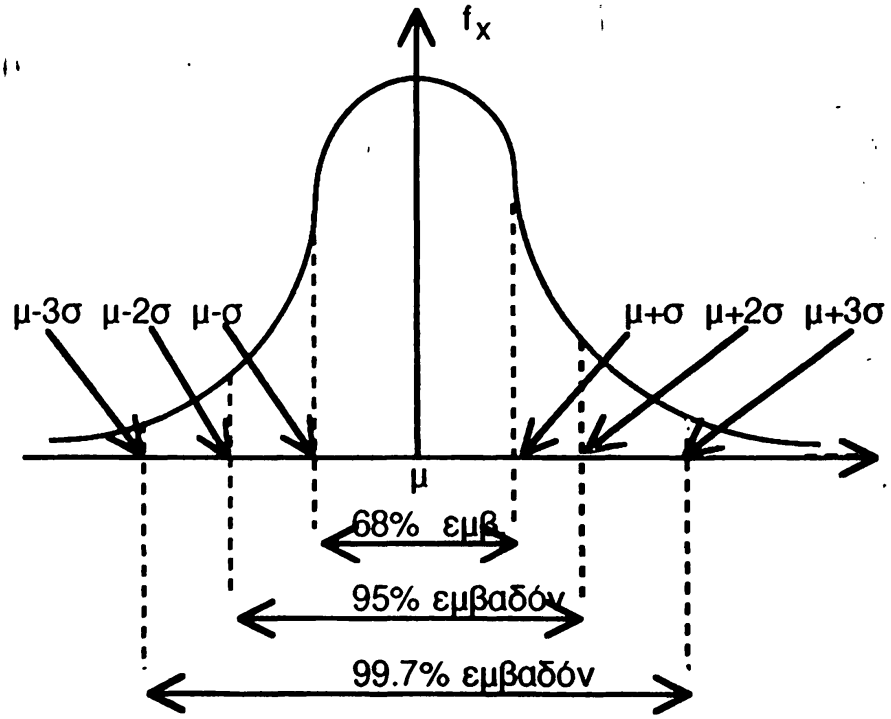
$\mu_1 < \mu_2 < \mu_3$



$\sigma_1^2 < \sigma_2^2 < \sigma_3^2$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

5. Όπως φαίνεται και στο σχήμα που ακολουθεί το 68% του εμβαδού μεταξύ της f_X και του οριζόντιου άξονα, περικλείεται στο διάστημα (μ-σ, μ+σ) , το 95% στο διάστημα (μ-2σ, μ+2σ) και το 99.7% στο διάστημα (μ-3σ, μ+3σ).



5.4.2 Τυπική κανονική κατανομή

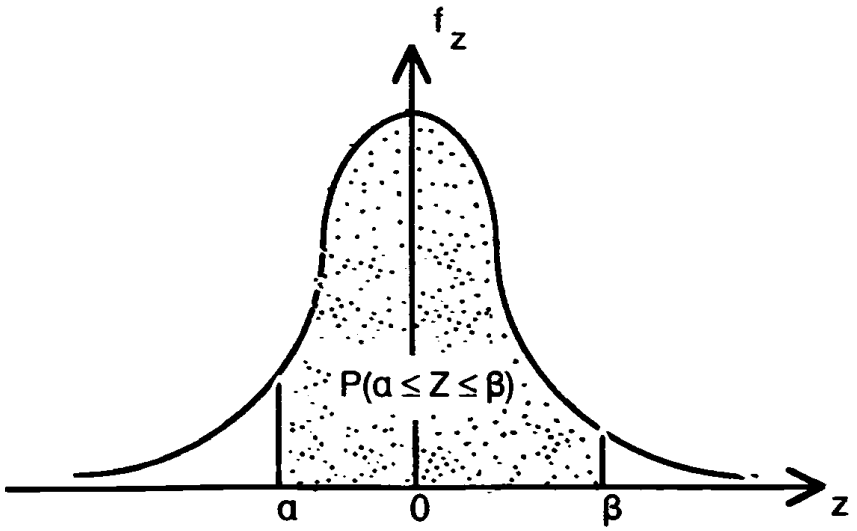
Από την ιδιότητα 4 συνάγεται ότι από την κανονική κατανομή προκύπτουν ειδικότερες κατανομές αν στις παραμέτρους μ και σ^2 δοθούν διάφορες τιμές. Η πιο ενδιαφέρουσα από αυτές είναι η τυπική κανονική κατανομή που προκύπτει από την κανονική κατανομή $N(\mu, \sigma^2)$ αν τεθεί $\mu=0$ και $\sigma^2=1$. Συμβολίζεται με $N(0,1)$, η αντίστοιχη τυχαία μεταβλητή συνήθως με Z και η συνάρτηση πυκνότητας πιθανότητάς της είναι :

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} , \quad -\infty < z < \infty.$$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Ενδιαφέρον, από πρακτική σκοπιά, παρουσιάζει ο υπολογισμός της πιθανότητας ενδεχομένων που εκφράζονται με τη βοήθεια μιας τυχαίας μεταβλητής Z με τυπική κανονική κατανομή $N(0,1)$, δηλαδή πιθανοτήτων της μορφής $P(\alpha \leq Z \leq \beta)$ με α και β δυο πραγματικούς αριθμούς.

Επειδή η τυχαία μεταβλητή Z είναι συνεχής, η πιθανότητα $P(\alpha \leq Z \leq \beta)$



είναι το εμβαδόν του γραμμοσκιασμένου χωρίου του παραπάνω σχήματος, που περικλείεται από το γράφημα της f_Z, τον άξονα των z, και τις κάθετες που άγονται στον άξονα αυτό στα σημεία του α και β. Το εμβαδό αυτό, όπως έχει ήδη τονιστεί, προσδιορίζεται από το ολοκλήρωμα της f_Z στο διάστημα $[\alpha, \beta]$. Στη βιβλιογραφία όμως υπάρχουν διαθέσιμοι πίνακες στους οποίους περιέχονται οι τιμές τέτοιων ολοκληρωμάτων. Ένας από τους πίνακες αυτούς δίνεται στο Παράρτημα και έχει δημιουργηθεί από τον υπολογισμό ολοκληρωμάτων της μορφής $\int_{-\infty}^{z_0} f_Z(z)dz$ ή ισοδύναμα πιθανοτήτων της μορφής $P(-\infty < Z \leq z_0)$ με $Z \sim N(0,1)$ και z_0 ένας γνωστός πραγματικός αριθμός.

Πριν τη χρησιμοποίηση του πίνακα αυτού πρέπει τα ενδεχόμενα, ο υπολογισμός της πιθανότητας των οποίων μας

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ενδιαφέρει, να μετασχηματίζονται ισοδύναμα σε ενδεχόμενα της μορφής ($-\infty < Z \leq z_0$). Ο μετασχηματισμός αυτός επιτυγχάνεται με τη χρήση της συμμετρίας της f_Z γύρω από το 0 και το γεγονός ότι το εμβαδό μεταξύ της f_Z και τον άξονα των z, δηλαδή η πιθανότητα $P(-\infty < Z \leq +\infty)$, είναι ίσο με τη μονάδα.

Η χρήση του παρακάτω πίνακα και ο υπολογισμός πιθανοτήτων

Z_o	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936

ενδεχομένων που εκφράζονται μέσω μιας τυπικής κανονικής τυχαίας μεταβλητής διευκρινίζεται στο παράδειγμα που ακολουθεί:

Παράδειγμα 5.4.2.1: Αν η τυχαία μεταβλητή Z ακολουθεί τυπική κανονική κατανομή N(0,1) να υπολογιστούν οι πιθανότητες: i) $P(Z \leq 0.75)$, ii) $P(-1.24 < Z \leq 0.36)$, iii) $P(0.5 < Z < 1.46)$, iv) $P(Z > 1.6)$.

Οι πιθανότητες αυτές υπολογίζονται μέσω του Πίνακα της τυπικής κανονικής κατανομής του Παραρτήματος.

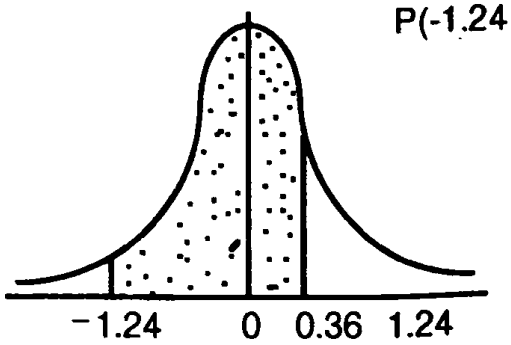
i) Υπολογίζεται άμεσα από τον πίνακα αφού πρόκειται για πιθανότητα ενδεχομένου της μορφής $(-\infty < Z \leq z_0)$, με $z_0 = 0.75$.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Έτσι εντοπίζεται στην πρώτη στήλη του πίνακα η τιμή 0.7 και στην πρώτη γραμμή η τιμή 0.05. Στη διασταύρωση της γραμμής 0.7 και της στήλης 0.05 βρίσκεται η ζητούμενη πιθανότητα. Δηλαδή $P(Z \leq 0.75) = 0.7734$.

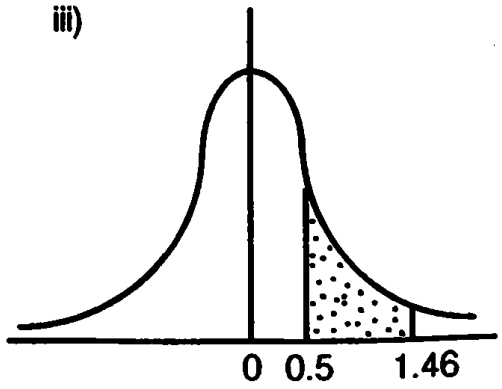
Για τον υπολογισμό των υπολοίπων πιθανοτήτων απαιτείται πρώτα ο μετασχηματισμός των ενδεχομένων τους σε ενδεχόμενα της μορφής: $(-\infty < Z \leq z_0)$ ή $(Z \leq z_0)$ με z_0 γνωστό.

ii)



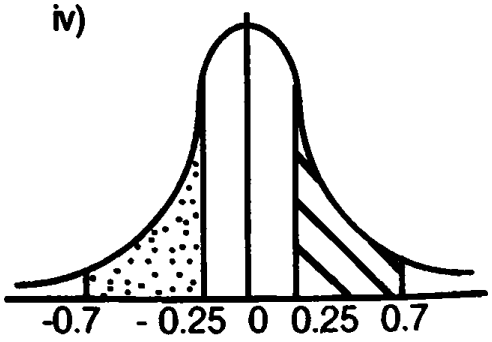
$$P(-1.24 < Z \leq 0.36) = P(-\infty < Z \leq 0.36) - P(Z \leq -1.24)$$
$$= 0.6406 - P(Z \geq 1.24)$$
$$= 0.6406 - [1-P(-\infty < Z \leq 1.24)]$$
$$= 0.6406 - [1- 0.8925] = 0.6406-0.1075$$
$$= 0.5331.$$

iii)



$$P(0.5 < Z < 1.46) =$$
$$= P(-\infty < Z < 1.46) - P(-\infty < Z < 0.5)$$
$$= 0.9279 - 0.6915$$
$$= 0.2364 .$$

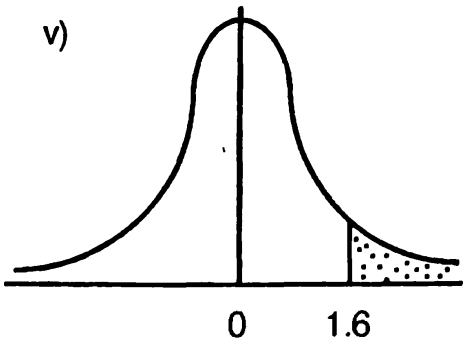
iv)



$$P(-0.7 \leq Z < -0.25) = P(0.25 < Z \leq 0.7)$$
$$= P(-\infty < Z \leq 0.7) - P(-\infty < Z < 0.25)$$
$$= 0.7580 - 0.5987$$
$$= 0.1593.$$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

v)



$P(Z > 1.6) = 1-P(-\infty < Z \leq 1.6)$

$= 1- 0.9452$

$= 0.0548.$

Η σημαντικότερη ίσως ιδιότητα της τυπικής κανονικής κατανομής είναι το γεγονός ότι οποιαδήποτε κανονική κατανομή μπορεί να αναχθεί στην τυπική κανονική αν χρησιμοποιηθεί ο λεγόμενος **τυπικός μετασχηματισμός.** Συγκεκριμένα αποδεικνύεται ότι:

Αν η τ.μ. $X\sim N(\mu,\sigma^2)$ τότε η τ.μ. $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.

Η συνεπαγωγή αυτή είναι πολύ χρήσιμη γιατί επιτρέπει τον υπολογισμό πιθανοτήτων $P(\alpha \leq X \leq \beta)$, $X\sim N(\mu,\sigma^2)$ με τη βοήθεια του πίνακα της τυπικής κανονικής κατανομής, αφού πρώτα εφαρμοστεί κατάλληλα ο πιο πάνω τυπικός μετασχηματισμός. Δηλαδή,

$$P(\alpha \leq X \leq \beta) = P(\frac{\alpha-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{\beta-\mu}{\sigma}) = P(\frac{\alpha-\mu}{\sigma} \leq Z \leq \frac{\beta-\mu}{\sigma}).$$

Παράδειγμα 5.4.2.2: Υποθέτουμε ότι ο δείκτης νοημοσύνης του πληθυσμού των παιδιών προσχολικής ηλικίας έχει κανονική κατανομή με $\mu=100$ και $\sigma=15$. Να υπολογιστούν: i) Η πιθανότητα ένας μαθητής να έχει δείκτη νοημοσύνης μικρότερο των 90 μονάδων. ii) Το ποσοστό των παιδιών με δείκτη μεγαλύτερο του 125 και iii) το ποσοστό των παιδιών με δείκτη νοημοσύνης μεταξύ 90 και 125 μονάδων.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Αν συμβολίσουμε με X την τυχαία μεταβλητή που παριστά το δείκτη νοημοσύνης ενός παιδιού προσχολικής ηλικίας, τότε $X \sim N(100, (15)^2 = 225)$. Χρησιμοποιώντας τον τυπικό μετασχηματισμό της τυχαίας μεταβλητής X, οι ζητούμενες πιθανότητες είναι:

i) $$P(X < 90) = P\left(\frac{X-100}{15} < \frac{90-100}{15}\right) = P(Z < -0.67)$$
$$= P(Z > 0.67)$$
$$= 1-P(Z \leq 0.67)$$
$$= 1- 0.7486$$
$$= 0.2514.$$

ii) $$P(X \geq 125) = P\left(\frac{X-100}{15} \geq \frac{125-100}{15}\right) = P(Z \geq 1.7)$$
$$= 1-P(Z \leq 1.7)$$
$$= 1- 0.9554$$
$$= 0.0446.$$

Έτσι το ποσοστό των παιδιών με δείκτη νοημοσύνης μεγαλύτερο των 125 μονάδων είναι 4.46%.

iii) $$P(90 \leq X \leq 125) = P\left(\frac{90-100}{15} \leq \frac{X-100}{15} \leq \frac{125-100}{15}\right)$$
$$= P(-0.67 \leq Z \leq 1.7)$$
$$= P(-\infty < Z \leq 1.7) - [1-P(-\infty < Z \leq 0.67)]$$
$$= 0.9554 - (1-0.7486)$$
$$= 0.9554 - 0.2514$$
$$= 0.7040.$$

Επομένως το ποσοστό των παιδιών με δείκτη νοημοσύνης μεταξύ 90 και 125 μονάδων είναι 70.40%.

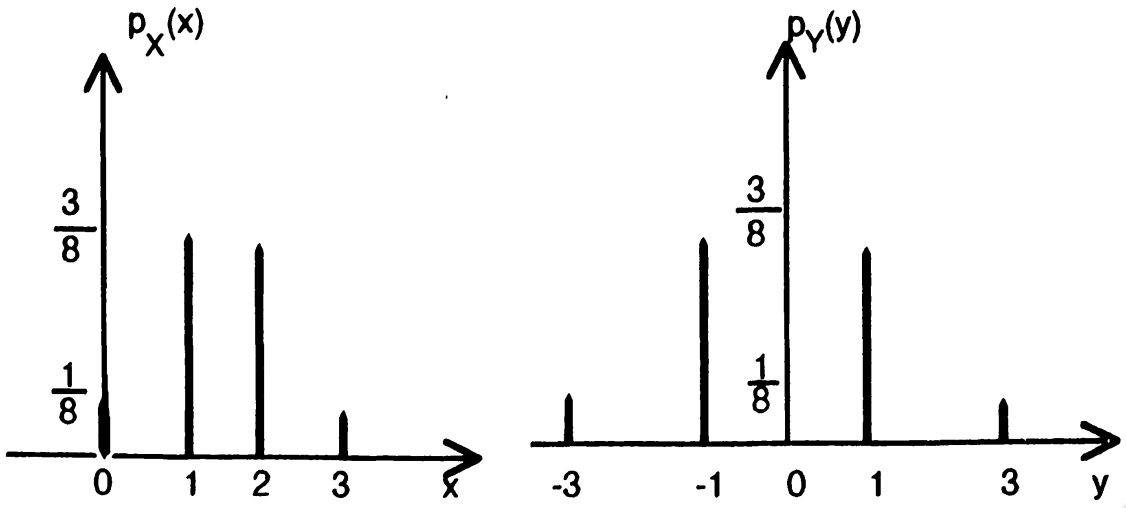
ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

5.5 Χαρακτηριστικά τυχαίων μεταβλητών και κατανομών: Μέση τιμή και διακύμανση

Στην περιγραφή ενός συνόλου αριθμητικών δεδομένων η έννοια της μέσης τιμής και της διακύμανσης των δεδομένων είναι κεντρικής σημασίας αφού με απλό και λακωνικό τρόπο φιλοδοξούν να περιγράψουν το σύνολο των δεδομένων, δίνοντας η μεν μέση τιμή μια ένδειξη της κεντρικής τάσης των μετρήσεων αυτών η δε διακύμανση το βαθμό συγκέντρωσης των μετρήσεων γύρω από τη μέση τους τιμή.

Ανάλογο ενδιαφέρον παρουσιάζουν και οι έννοιες της μέσης τιμής και διακύμανσης μιας τυχαίας μεταβλητής που παρουσιάζονται στη συνέχεια.

Θεωρούμε τις τυχαίες μεταβλητές X και Y των Παραδειγμάτων 5.2.1 και 5.2.2 αντίστοιχα. Η X παριστά το πλήθος των σωστών απαντήσεων στις τρεις ερωτήσεις που απάντησε στην τύχη ο μαθητής ενώ η Y παριστά το συνολικό κέρδος του μαθητή όταν για κάθε σωστή απάντηση κερδίζει μια μονάδα, ενώ για κάθε λάθος απάντηση χάνει μια μονάδα. Οι συναρτήσεις πιθανότητας των τυχαίων μεταβλητών X και Y είναι αντίστοιχα:



Σχήμα 5.1: Συναρτήσεις πιθανότητας των τυχ. μεταβλητών των Παραδειγμάτων 5.2.1.-2.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Εάν ένας πολύ μεγάλος αριθμός N μαθητών απαντούσε στις τρεις ερωτήσεις, τότε η κατανομή συχνοτήτων του αριθμού X των σωστών απαντήσεων θα ήταν:

Αριθμός σωστών απαντήσεων (x)	Συχνότητα (f)
0	$\frac{1}{8}$ N = 0.125N
1	$\frac{3}{8}$ N = 0.375N
2	$\frac{3}{8}$ N = 0.375N
3	$\frac{1}{8}$ N = 0.125N

αφού οι σχετικές συχνότητες των διαφορετικών τιμών της X θα ήταν (μέσω του εμπειρικού ορισμού της πιθανότητας) ίσες με τις αντίστοιχες πιθανότητες των τιμών αυτών.

Η μέση τιμή των N μετρήσεων, που αφορούν τον αριθμό των σωστών απαντήσεων των N μαθητών είναι :

$$\frac{(0\times 0.125N) + (1\times 0.375N) + (2\times 0.375N) + (3\times 0.125N)}{N} =$$

$$= 0\times 0.125 + 1\times 0.375 + 2\times 0.375 + 3\times 0.125$$

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

$$= 0p_X(0) + 1p_X(1) + 2p_X(3) + 3p_X(3) = 1.5$$

Η σχέση αυτή υποδεικνύει τον ορισμό της μέσης τιμής μιας διακριτής τυχαίας μεταβλητής ως εξής:

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Ορισμός: Αν X είναι μια διακριτή τυχαία μεταβλητή, x οι τιμές της και $p_X(x)$ οι πιθανότητες των τιμών αυτών, τότε η **μέση τιμή ή αναμενόμενη τιμή** (mean value, expectation) της τυχαίας μεταβλητής X, συμβολίζεται με μ ή E(X) και ορίζεται:

$$\mu = E(X) = \sum_x x\, p_X(x) \ .$$

Εφαρμόζοντας τον ορισμό αυτό η μέση τιμή της τυχαίας μεταβλητής Y του Παραδείγματος 5.2.2 είναι:

$$E(Y) = \sum_y y\, p_Y(y) = (-3)\frac{1}{8} + (-1)\frac{3}{8} + (1)\frac{3}{8} + (3)\frac{1}{8}$$

$$= -\frac{3}{8} - \frac{3}{8} + \frac{3}{8} + \frac{3}{8} = 0.$$

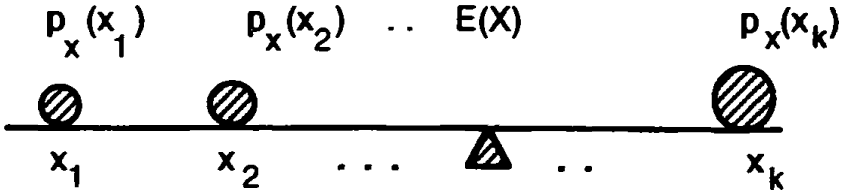
Κέρδος του μαθητή ίσο με 0 ήταν αναμενόμενο αφού κερδίζει ή χάνει τις ίδιες ποσότητες με τις ίδιες πιθανότητες.

Ο ορισμός της μέσης τιμής μιας διακριτής τυχαίας μεταβλητής μέσω της αντίστοιχης έννοιας ενός συνόλου αριθμητικών δεδομένων, έχει σαν συνέπεια και ανάλογη ερμηνεία για τη μέση τιμή της τυχαίας μεταβλητής . Έτσι η μέση τιμή μιας διακριτής τυχαίας μεταβλητής είναι το σημείο γύρω από το οποίο συγκεντρώνονται οι τιμές της μεταβλητής, το σημείο γύρω από το οποίο κατανέμεται μάζα πιθανότητας ίση με τη μονάδα. Παριστά την επικρατέστερη τιμή της τυχαίας μεταβλητής ή την τιμή που μπορεί να αντικαταστήσει, να αντιπροσωπεύσει τις υπόλοιπες και υπό την έννοια αυτή περιγράφει την κεντρική τάση των μελών του πληθυσμού. Επίσης η μέση τιμή χρησιμοποιείται για να περιγράψει και το αναμενόμενο κέρδος που

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

προκύπτει από την επανάληψη ενός τυχερού παιγνιδιού. Η εφαρμογή αυτή δόθηκε στο προηγούμενο παράδειγμα.

Από φυσική άποψη η έννοια της μέσης τιμής είναι ταυτόσημη με την έννοια του κέντρου βάρους.



Πράγματι αν στα σημεία x_i μιας ράβδου τοποθετηθούν βάρη ίσα με $p_X(x_i)$, $i=1,...,k$ τότε η ράβδος πρέπει να στηριχθεί στο σημείο E(X) ώστε να ισσοροπεί.

Ανάλογη ερμηνεία δίνεται και για τη μέση τιμή μιας συνεχούς τυχαίας μεταβλητής. Ο ορισμός της προκύπτει, από την αντικατάσταση του συμβόλου της αθροίσεως με το σύμβολο της ολοκληρώσεως, στον ορισμό της μέσης τιμής διακριτής τυχαίας μεταβλητής και είναι:

$$\mu = E(X) = \int_{-\infty}^{+\infty} x\, f_X(x)\, d\, x\,,$$

με f_X τη συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής X. Μέθοδοι υπολογισμού ολοκληρωμάτων αυτής της μορφής είναι διαθέσιμοι, ξεφεύγουν όμως του σκοπού των σημειώσεων αυτών.

Από τον ορισμό της μέσης τιμής τυχαίας μεταβλητής άμεσα προκύπτουν και οι ακόλουθες ιδιότητές της:

1. $E(aX) = aE(X)$, με a μια σταθερά,
2. $E(aX+b) = aE(X) +b$, με a και b σταθερές,
3. $E(X+Y) = E(X)+E(Y)$ με X, Y τυχαίες μεταβλητές.

Επίσης υπολογίζεται ότι η μέση τιμή μιας τυχαίας μεταβλητής με διωνυμική κατανομή B(n,p) είναι ίση με np, η μέση τιμή μιας

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

κατανομής Poisson $P(\lambda)$ είναι ίση με λ, ενώ η μέση τιμή τυχαίας μεταβλητή με κανονική κατανομή $N(\mu, \sigma^2)$ είναι ίση με την παράμετρο μ.

Η μέση τιμή μιας τυχαίας μεταβλητής δίνει χρήσιμες πληροφορίες για τη θέση της κατανομής, για το σημείο γύρω από το οποίο συγκεντρώνονται, με μεγάλη πιθανότητα οι τιμές της τυχαίας μεταβλητής. Δεν δίνει όμως καμιά πληροφορία για τη μεταβλητότητα ή για τον τρόπο που οι τιμές της τυχαίας μεταβλητής κατανέμονται, αραδειάζονται γύρω από τη μέση τους τιμή. Από την προσεκτική παρατήρηση των κατανομών των τυχαίων μεταβλητών X και Y (Σχήμα 5.1) των Παραδειγμάτων 5.2.1 και 5.2.2, προκύπτει ότι οι αντίστοιχες κατανομές συγκεντρώνονται γύρω από τις μέσες τους τιμές. Η κατανομή όμως της τυχαίας μεταβλητής Y είναι περισσότερο απλωμένη, αραδειασμένη γύρω από τη μέση της τιμή απ' ότι είναι η κατανομή της τυχαίας μεταβλητής X. Γίνεται έτσι φανερή η ανάγκη εισαγωγής ενός μέτρου της μεταβλητότητας, της διασποράς των τιμών της τυχαίας μεταβλητής από τη μέση της τιμή. Το μέτρο αυτό είναι η διακύμανση της τυχαίας μεταβλητής και ορίζεται:

Ορισμός: Η **διακύμανση** (variance) της τυχαίας μεταβλητής X συμβολίζεται με $Var(X)$ ή σ^2 και ορίζεται

$$\sigma^2 = Var(X) = E(X-\mu)^2.$$

Η σχέση αυτή, αν η X είναι διακριτή, γίνεται:

$$\sigma^2 = Var(X) = \sum_x (x-\mu)^2 p_X(x),$$

ενώ αν η τυχαία μεταβλητή X είναι συνεχής, γίνεται

$$\sigma^2 = Var(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)\, d\, x.$$

Η διακύμανση μιας τυχαίας μεταβλητής ορίζεται ως η μέση τιμή του τετραγώνου της διαφοράς της από τη μέση της τιμή. Αποτελεί γενίκευση του ορισμού της διακύμανσης ενός συνόλου μετρήσεων, που δίνεται από την έκφραση $\frac{1}{N-1} \sum_{i=1}^{N} (x_i-\bar{x})^2$, και έχει ανάλογη ερμηνεία με αυτήν. Έτσι η διακύμανση μιας τυχαίας μεταβλητής περιγράφει τον τρόπο κατανομής, αραδειάσματος των τιμών της τυχαίας μεταβλητής γύρω από τη μέση τους τιμή, ή τον τρόπο κατανομής μιας μάζας πιθανότητας ίση με τη μονάδα γύρω από τη μέση τιμή. Επίσης δίνει ένα μέτρο της ομοιογένειας των τιμών της μεταβλητής ή των μελών του πληθυσμού χαρακτηριστικά των οποίων περιγράφει η μεταβλητή.

Οι διακυμάνσεις των τυχαίων μεταβλητών X και Y των Παραδειγμάτων 5.2.1 και 5.2.2 υπολογίζονται ότι είναι:

$$Var(X) = (0-1.5)^2 \frac{1}{8} + (1-1.5)^2 \frac{3}{8} + (2-1.5)^2 \frac{3}{8} + (3-1.5)^2 \frac{1}{8} = 0.75$$

$$Var(Y) = (-3-0)^2 \frac{1}{8} + (-1-0)^2 \frac{3}{8} + (1-0)^2 \frac{3}{8} + (3-0)^2 \frac{1}{8} = 3\ .$$

Ισχύει $Var(X) = 0.75 < 3 = Var(Y)$ γεγονός που ήταν αναμενόμενο.

Ένας εύκολος τρόπος υπολογισμού της διακύμανσης δίνεται από τη σχέση:

$$Var(X) = E(X-\mu)^2 = E(X^2) - \mu^2.$$

Επίσης οι βασικές ιδιότητες της διακύμανσης είναι:

1. $Var(\alpha) = 0$,

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

2. $Var(aX) = a^2\, Var(X)$,

3. $Var(aX+b) = a^2\, Var(X)$,

με a και b σταθερές.

Εύκολα υπολογίζεται ότι η διακύμανση της B(n,p) είναι np(1-p), της P(λ) ίση με την παράμετρο λ και της $N(\mu,\sigma^2)$ ίση με την παράμετρο σ^2.

Η **τυπική απόκλιση** (standard deviation) μιας τυχαίας μεταβλητής, συμβολίζεται με σ, έχει ακριβώς την ίδια ερμηνεία με της διακύμανσης και ορίζεται:

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

Στην πράξη προτιμάται η χρησιμοποίησή της γιατί εκφράζεται με τις ίδιες μονάδες μέτρησης με την τυχαία μεταβλητή.

Εκτός από τη μέση τιμή και τη διακύμανση, χαρακτηριστικά μιας τυχαίας μεταβλητής και της κατανομής της είναι επίσης: οι ροπές, η διάμεσος, η κορυφή, τα εκατοστιαία σημεία, ο συντελεστής λοξότητας, ο συντελεστής κυρτότητας κλπ. για τα οποία ο αναγνώστης παραπέμπεται σε πιο προχωρημένα βιβλία θεωρίας πιθανοτήτων και στατιστικής.

5.6 Ανακεφαλαίωση

Στην παράγραφο αυτή επιχειρείται μια ανακεφαλαίωση με κύρια επιδίωξη τη σύνδεση των εννοιών της θεωρίας πιθανοτήτων και εκείνων που παρουσιάστηκαν στα προηγούμενα κεφάλαια.

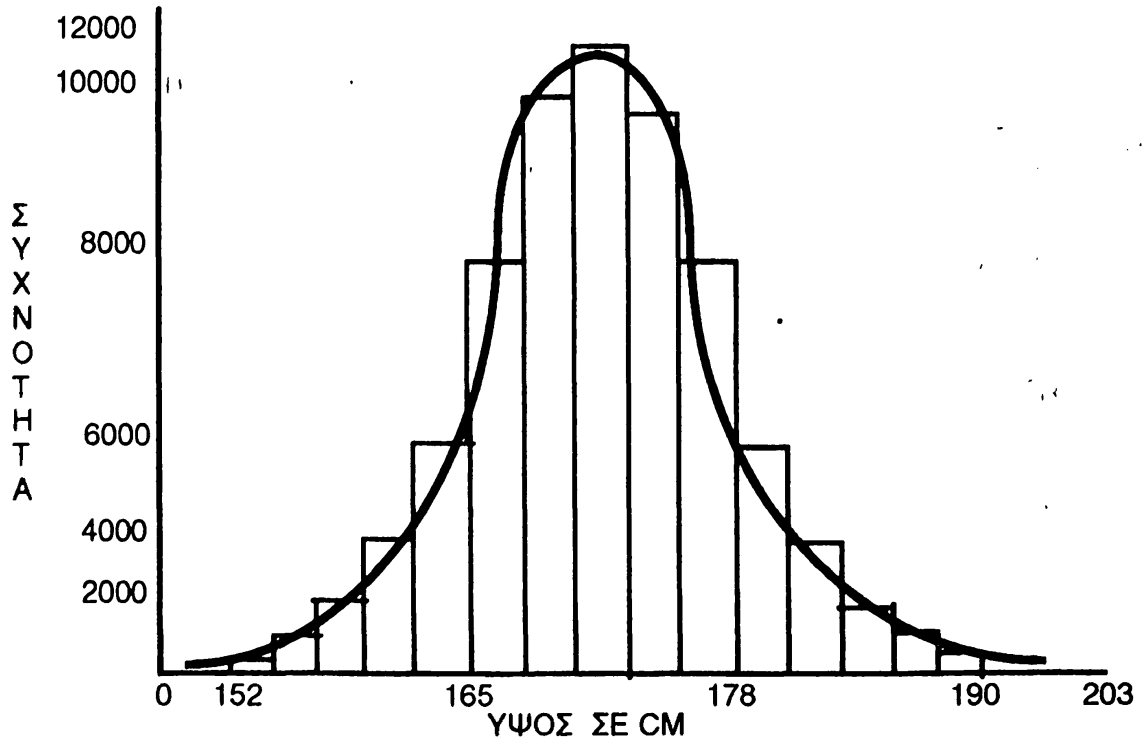
ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Η έννοια της τυχαίας μεταβλητής και της κατανομής πιθανότητάς της - χρησιμοποιείται ο όρος κατανομή πιθανότητας για να ενοποιήσει τις έννοιες της συνάρτησης πιθανότητας και της συνάρτησης πυκνότητας πιθανότητας για τη διακριτή και τη συνεχή περίπτωση αντίστοιχα - είναι κεντρικής σημασίας για τη θεωρία πιθανοτήτων και τη στατιστική. Όπως έχει ήδη τονιστεί, με την πρώτη επιτυγχάνεται η σύνδεση των αποτελεσμάτων ενός τυχαίου πειράματος με την ευθεία των πραγματικών αριθμών, ενώ με τη δεύτερη δίνεται η δυνατότητα υπολογισμού της πιθανότητας ενδεχομένων που εκφράζονται μέσω της αντίστοιχης τυχαίας μεταβλητής.

Σε προηγούμενα Κεφάλαια παρουσιάστηκαν οι αρχές και οι μέθοδοι για την οργάνωση και περιληπτική παρουσίαση ενός συνόλου αριθμητικών δεδομένων. Προκύπτουν συνήθως τα δεδομένα αυτά από τη μέτρηση ενός ή περισσοτέρων χαρακτηριστικών των μελών ενός πληθυσμού και ουσιαστικά είναι τιμές μιας ή περισσότερων τυχαίων μεταβλητών που αναφέρονται στα χαρακτηριστικά αυτά. Η κατανομή των σχετικών συχνοτήτων ενός συνόλου αριθμητικών δεδομένων και η γραφική της παράσταση (ιστόγραμμα, πολύγωνο συχνοτήτων) περιγράφει τον τρόπο με τον οποίο μια μάζα (σχετικών συχνοτήτων) ίση με τη μονάδα, κατανέμεται στο σύνολο των δεδομένων. Η σχέση που υπάρχει, μέσω του εμπειρικού ορισμού της πιθανότητας, μεταξύ της σχετικής συχνότητας και της πιθανότητας ενός ενδεχομένου δικαιολογεί ανάλογη σχέση μεταξύ της κατανομής σχετικών συχνοτήτων ενός συνόλου αριθμητικών δεδομένων και της κατανομής πιθανότητας της αντίστοιχης τυχαίας μεταβλητής. Έτσι η κατανομή σχετικών συχνοτήτων αποτελεί μια προσέγγιση της κατανομής πιθανότητας. Όσο μεγαλύτερο είναι το πλήθος των αριθμητικών δεδομένων τόσο καλύτερη είναι η προσέγγιση αυτή όπως φαίνεται και από το σχήμα 5.2. Στο σχήμα αυτό παρουσιάζεται το ιστόγραμμα

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

συχνοτήτων του ύψους ανδρών, ενώ η συνεχής καμπύλη είναι το γράφημα της συνάρτησης πυκνότητας πιθανότητας μιας κανονικής κατανομής. Η προσέγγιση της κατανομής αυτής από το ιστόγραμμα συχνοτήτων είναι τέλεια.



Σχήμα 5.2: Κατανομή του ύψους ανδρών με κατά προσέγγιση κανονική κατανομή

Η γνώση της κατανομής πιθανότητας μιας τυχαίας μεταβλητής έχει τεράστια σημασία, αφού επιτρέπει τον υπολογισμό πιθανοτήτων, συνδέοντας πολλές φορές τις πιθανότητες αυτές και με τη λήψη μιας απόφασης. Ένας έμπορος παπουτσιών για παράδειγμα, θα ενδιαφερόταν για την κατανομή των μεγεθών των παπουτσιών που εξυπηρετούν την πελατεία του. Έτσι σε κάθε εποχή θα είναι σε θέση να προμηθευτεί τον κατάλληλο (λιγότερο ή περισσότερο) αριθμό ζευγαριών από τα μεγέθη που κυρίως καταναλώνονται. Αν η γνώση αυτής της θεωρητικής κατανομής είναι αδύνατη τότε η εμπειρία του τον οδηγεί στην πιο κατάλληλη παραγγελία. Από το παράδειγμα αυτό

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

διευκρινίζεται πάλι ο ρόλος της θεωρητικής κατανομής πιθανότητας, που κατασκευάζεται στη βάση και της εμπειρίας και κάποιων θεωρητικών υποθέσεων, και της κατανομής συχνοτήτων που την προσεγγίζει και αποτελεί καταγραφή εμπειριών.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Ασκήσεις 5ου Κεφαλαίου

5.1: Είναι γνωστό ότι το 50% των νεογέννητων στο Νοσοκομείο Ιωαννίνων είναι αγόρια. Ποιά η πιθανότητα στις 18 γεννήσεις που έγιναν μια ημέρα στο νοσοκομείο αυτό: i) γεννήθηκαν ακριβώς τρια αγόρια, ii) λιγότερα από τέσσερα αγόρια, iii) περισσότερα από 16 αγόρια, iv) μεταξύ 5 και 10 αγοριών.

5.2: Έχει παρατηρηθεί ότι ο Γιώργος σημαδεύει με επιτυχία ένα στόχο το 25% των φορών. Ποιά η πιθανότητα στις 5 βολές: i) το πολύ 2 να είναι επιτυχείς; ii) 3 βολές να είναι επιτυχείς;

5.3: Έχει παρατηρηθεί ότι ο μέσος αριθμός σοβαρών ατυχημάτων, που συμβαίνουν ανά έτος, στους μαθητές ενός νηπιαγωγείου κατά τη διάρκεια του παιχνιδιού είναι πέντε ατυχήματα ανά έτος. Να υπολογιστεί η πιθανότητα στο έτος που διανύεται να συμβούν: i) επτά σοβαρά ατυχήματα, ii) δέκα ή περισσότερα ατυχήματα, iii) κανένα ατύχημα, iv) λιγότερα από πέντε ατυχήματα.

5.4: Αν η τυχαία μεταβλητή X ακολουθεί κανονική κατανομή με μέση τιμή 75 και διακύμανση 225 να υπολογιστούν οι πιθανότητες: i) $P(50 \leq X \leq 100)$, ii) $P(X > 90)$, iii) $P(X < 60)$, iv) $P(X \geq 85)$, v) $P(30 \leq X < 110)$.

5.5: Από προηγούμενες μελέτες έχει διαπιστωθεί ότι ο δείκτης νοημοσύνης των μαθητών του Λυκείου περιγράφεται από την κανονική κατανομή με μέση τιμή 100 και τυπική απόκλιση 15 μονάδες. Να υπολογιστεί το ποσοστό των μαθητών με δείκτη

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

νοημοσύνης: i) μεταξύ 85 και 115 μονάδες, ii) μεταξύ 115 και 130 μονάδες, iii) μεγαλύτερο από 123.25 μονάδες.

5.6: α) Είναι γνωστό ότι το ύψος του πληθυσμού των παιδιών προσχολικής ηλικίας ακολουθεί κανονική κατανομή με μέση τιμή 100 και τυπική απόκλιση 10 εκατοστά. Να βρεθεί το ποσοστό των παιδιών με ύψος μεταξύ 95 και 105 εκατοστά.

β) Αν το φυσιολογικό ύψος των παιδιών προσχολικής ηλικίας είναι μεταξύ 95 και 105 εκατοστών, να βρεθεί η πιθανότητα από 15 παιδιά που εκλέχτηκαν στην τύχη από τον πληθυσμό αυτό, ακριβώς 2 να έχουν φυσιολογικό ύψος.

5.7: Ένα κουτί περιέχει 3 μαύρες, 2 κόκκινες και 3 πράσινες ολόιδιες μπάλλες. Ένας μαθητής διαλέγει στην τύχη μια μπάλλα απ' το κουτί και κερδίζει 2 μονάδες αν η μπάλλα που διάλεξε είναι μαύρη, χάνει μια μονάδα αν η μπάλλα είναι κόκκινη και δεν κερδίζει τίποτε αν η μπάλλα που διάλεξε είναι πράσινη. α) Να υπολογιστεί και σχεδιαστεί η κατανομή του κέρδους του μαθητή, β) να υπολογιστεί το αναμενόμενο κέρδος και η διακύμανσή του.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΚΕΦΑΛΑΙΟ 6

ΣΤΑΤΙΣΤΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ

ΕΚΤΙΜΗΣΗ ΠΑΡΑΜΕΤΡΩΝ

6.1 Εισαγωγικά

Σε προηγούμενα Κεφάλαια παρουσιάστηκαν οι αρχές και οι μέθοδοι για την παρουσίαση και οργάνωση ενός συνόλου αριθμητικών δεδομένων. Οι αρχές και οι μέθοδοι αυτές δίνουν περιεχόμενο στον όρο περιγραφική στατιστική. Η δεύτερη δραστηριότητα στο χώρο της στατιστικής συνοψίζεται κάτω από τον τίτλο επαγωγική στατιστική ή στατιστική συμπερασματολογία. Αντικείμενο της στατιστικής συμπερασματολογίας είναι η ανάπτυξη μεθοδολογιών που οδηγούν στην εξαγωγή συμπερασμάτων και πιθανώς στη λήψη αποφάσεων για ένα πληθυσμό, από τη μελέτη και ανάλυση ενός τυχαίου δείγματος που εκλέγεται από τον πληθυσμό αυτό. Η στατιστική συμπερασματολογία μας εφοδιάζει με δυναμικές μεθόδους για τη στατιστική διερεύνηση μεγάλης ποικιλίας προβλημάτων.

Σημείο αναφοράς οποιασδήποτε στατιστικής μελέτης ενός φαινομένου, είναι το τυχαίο δείγμα το οποίο συνίσταται από μετρήσεις που γίνονται σε μέλη ενός πληθυσμού και αφορούν ένα ή περισσότερα

ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ ΒΙΒΛΙΟΘΗΚΗ

χαρακτηριστικά τους. Τα χαρακτηριστικά αυτά παριστάνονται - περιγράφονται από μια ή περισσότερες τυχαίες μεταβλητές. Άμεσα συνδεδεμένος με τις μεταβλητές αυτές είναι ο νόμος που τις περιγράφει, η κατανομή πιθανότητάς τους. Η γνώση της κατανομής αυτής έχει τεράστια σημασία αφού παρέχει τη δυνατότητα υπολογισμού πιθανοτήτων, προσδιορισμού των χαρακτηριστικών της αντίστοιχης τυχαίας μεταβλητής (μέση τιμή, διακύμανση κλπ.) και γενικότερα περιγράφει τη συμπεριφορά του πληθυσμού ως προς το υπό μελέτη χαρακτηριστικό του. Γίνεται έτσι φανερό ότι η δραστηριότητα της στατιστικής συμπερασματολογίας στρέφεται στην αξιοποίηση όλων των πληροφοριών που περιέχονται στο τυχαίο δείγμα για την εύρεση της κατανομής της τυχαίας μεταβλητής απ' την οποία προέρχεται αυτό.

Άλλες φορές είναι γνωστή η συναρτησιακή έκφραση της κατανομής, αλλά άγνωστες οι παράμετρες που υπεισέρχονται στην έκφρασή της π.χ. η παράμετρος p της διωνυμικής κατανομής, η λ της Poisson, οι παράμετροι μ και σ^2 της κανονικής κατανομής. Άλλες φορές όμως είναι εντελώς άγνωστη η μορφή της συνάρτησης κατανομής. Μια προσέγγισή της, στην περίπτωση αυτή, προκύπτει από το πολύγωνο σχετικών συχνοτήτων. Οι στατιστικές μέθοδοι που αναπτύσσονται όταν είναι γνωστή η συναρτησιακή έκφραση της κατανομής αποτελούν αντικείμενο της **παραμετρικής στατιστικής** ενώ οι αντίστοιχες μέθοδοι όταν είναι άγνωστη η έκφραση της κατανομής αποτελούν αντικείμενο της **μη παραμετρικής στατιστικής.**

Στη συνέχεια το ενδιαφέρον επικεντρώνεται στο χώρο της παραμετρικής στατιστικής και ειδικότερα στην ανάπτυξη μεθοδολογιών για την προσέγγιση των άγνωστων παραμέτρων που εμφανίζονται στη συναρτησιακή έκφραση της κατανομής απ' την οποία περιγράφεται ο υπό μελέτη πληθυσμός. Η στατιστική συμπερασματολογία στην

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

περίπτωση αυτή χωρίζεται σε δυο κλάδους: **την Εκτιμητική** και τον **Έλεγχο Στατιστικών Υποθέσεων**. Πριν την ανάπτυξη των μεθοδολογιών αυτών, μέσω του Παραδείγματος που ακολουθεί, εισάγεται ο συμβολισμός και η ορολογία που χρησιμοποιείται στη συνέχεια.

Παράδειγμα 6.1.1: Μια μελέτη έχει για στόχο τη διερεύνηση και την εξαγωγή συμπερασμάτων σχετικά με το ύψος των παιδιών προσχολικής ηλικίας της Ηπείρου. Το υπό μελέτη χαρακτηριστικό του πληθυσμού αυτού είναι το ύψος το οποίο περιγράφεται από τη μεταβλητή X. Από τον πληθυσμό αυτό N παιδιά εκλέγονται στην τύχη και μετριέται το ύψος τους. Έστω X_1 παριστά το ύψος του πρώτου παιδιού, X_2 το ύψος του δεύτερου παιδιού, κ.ο.κ. X_N του ύψος του N-ιοστού παιδιού. Η N-ιάδα $X_1, X_2, ..., X_N$ αποτελεί το τυχαίο δείγμα. Κάθε ένα από τα X_i, $i = 1,2,...,N$ έχει διπλή ερμηνεία. Αφ' ενός είναι η τυχαία μεταβλητή που περιγράφει και παριστά το ύψος του i παιδιού και αφ' ετέρου είναι ένας αριθμός που αντιστοιχεί στο ύψος του i παιδιού και αποτελεί έτσι μια τιμή της τυχαίας μεταβλητής X που περιγράφει το ύψος όλων των μελών του πληθυσμού. Η τυχαία μεταβλητή X και επομένως κάθε μία από τις X_i, $i = 1,2,...,N$, υπακούει σε ένα συγκεκριμένο νόμο, την συνάρτηση πυκνότητας πιθανότητάς της. Η κανονική κατανομή περιγράφει αρκετά καλά μεταβλητές όπως το βάρος, το ύψος, το δείκτη νοημοσύνης κλπ. Έτσι μπορούμε να θεωρήσουμε ότι η τυχαία μεταβλητή X ακολουθεί κανονική κατανομή. Ο ισχυρισμός αυτός μπορεί να ελεγθεί και από το πολύγωνο σχετικών συχνοτήτων των αριθμητικών δεδομένων $X_1, X_2, ..., X_N$ το οποίο θα προσεγγίζει την γραφική παράσταση μιας κανονικής κατανομής. Το γεγονός ότι τα $X_1, X_2, ..., X_N$ προέρχονται από την $N(\mu,\sigma^2)$, που είναι τεράστιας σημασίας, παραμένει αναξιοποίητο αν δεν συνοδεύεται και από τη γνώση, ή έστω την προσέγγιση, των άγνωστων κατά κανόνα

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

παραμέτρων μ και σ^2. Εδώ ακριβώς αρχίζει ο ρόλος της στατιστικής συμπερασματολογίας που μας εφοδιάζει με μεθοδολογίες για την προσέγγιση αυτών των παραμέτρων. Οι μεθοδολογίες που αναπτύσσονται βασίζονται στο τυχαίο δείγμα $X_1, X_2, \ldots, X_N$ ή σε συναρτήσεις του τυχαίου δείγματος που ονομάζονται **στατιστικές συναρτήσεις** ή **στατιστικά.** Το $\bar{X}$ και το S^2 που ορίζονται αντίστοιχα από τις σχέσεις: $\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$ και $S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2$ είναι συναρτήσεις μόνο του τυχαίου δείγματος $X_1, X_2, \ldots, X_N$ και αποτελούν έτσι χαρακτηριστικά παραδείγματα στατιστικών.

Υπάρχουν δυο γενικά κλάδοι στατιστικής συμπερασματολογίας: 1) Η εκτιμητική, η οποία επιδιώκει την προσέγγιση των άγνωστων παραμέτρων χωρίς καμιά απολύτως αρχική υπόθεση γι' αυτές και 2) Ο έλεγχος στατιστικών υποθέσεων ο οποίος διατυπώνει μια υπόθεση για την τιμή της άγνωστης παραμέτρου και χρησιμοποιεί το τυχαίο δείγμα για τον έλεγχο της ορθότητας της υπόθεσης αυτής. Ο πρώτος κλάδος, η εκτιμητική, αναπτύσσεται στο κεφάλαιο αυτό, ενώ ο έλεγχος στατιστικών υποθέσεων αποτελεί αντικείμενο του επόμενου κεφαλαίου.

Η εκτιμητική περιλαμβάνει τις μεθοδολογίες για τον προσδιορισμό, από το τυχαίο δείγμα, ενός στατιστικού το οποίο προσφέρεται για την προσέγγιση, την εκτίμηση της άγνωστης παραμέτρου του πληθυσμού από τον οποίο προέρχεται το τυχαίο δείγμα. Η εκτίμηση της άγνωστης παραμέτρου μπορεί να γίνει με δυο τρόπους: την εκτίμηση σε σημείο και την εκτίμηση σε διάστημα.

6.2 Εκτίμηση σε σημείο

Η εκτίμηση σε σημείο μιας παραμέτρου ενός πληθυσμού αναφέρεται σε διαδικασίες εύρεσης μιας στατιστικής συνάρτησης με τη

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

βοήθεια της οποίας προσεγγίζεται η άγνωστη παράμετρος του πληθυσμού. Η στατιστική συνάρτηση - συμβολίζεται συνήθως με $T = T(X_1,X_2,...,X_N)$ - που χρησιμοποιείται για την εκτίμηση μιας άγνωστης παραμέτρου λέγεται **εκτιμήτρια συνάρτηση**, η δε τιμή της - που προκύπτει από την αντικατάσταση στη συναρτησιακή της έκφραση των $X_1, X_2,...,X_N$ από τις τιμές τους - λέγεται **εκτιμητής**. Σαν παράδειγμα θεωρούμε ένα πληθυσμό που περιγράφεται από την κανονική κατανομή $N(\mu,\sigma^2)$ και ένα τυχαίο δείγμα $X_1, X_2,...,X_N$ από τον πληθυσμό αυτό. Η στατιστική συνάρτηση: $T = T(X_1, X_2,...,X_N) = \frac{1}{N} \sum_{i=1}^{N} X_i = \bar{X}$ είναι λογικό να θεωρηθεί ως υποψήφια εκτιμήτρια συνάρτηση της παραμέτρου μ η δε τιμή της, που προκύπτει από την αντικατάσταση των $X_1,...,X_N$ με τις τιμές τους, ως υποψήφιος εκτιμητής της άγνωστης παραμέτρου μ.

Σε πολλές περιπτώσεις για μια παράμετρο είναι δυνατό να βρεθούν περισσότεροι του ενός εκτιμητές. Για παράδειγμα και η διάμεσος του δείγματος $X_1,...,X_N$ μπορεί να χρησιμοποιηθεί ως εκτιμητής της παραμέτρου μ του πληθυσμού $N(\mu,\sigma^2)$. Το ερώτημα που σ' αυτή τη φάση γεννιέται είναι ποιός από τους προηγούμενους εκτιμητές της παραμέτρου μ είναι προτιμότερος, η μέση τιμή $\bar{X}$ του δείγματος ή η διάμεσός του; Η απάντηση βασίζεται σε κριτήρια που αξιολογούν τους υποψήφιους εκτιμητές μιας παραμέτρου. Τα σημαντικότερα από τα κριτήρια αυτά είναι η **αμεροληψία** και η **ελάχιστη διακύμανση**.

Ένας εκτιμητής, έστω ο T, της παραμέτρου θ ενός πληθυσμού λέγεται **αμερόληπτος εκτιμητής** της θ αν: $E(T) = \theta$. Η $E(T)$ είναι η μέση τιμή της στατιστικής συνάρτησης T και ορίζεται όπως και η μέση τιμή μιας τυχαίας μεταβλητής.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Η αμεροληψία ενός εκτιμητή αποτελεί ένα μέτρο της ορθότητας της εκτίμησης αφού επιλέγει ως εκτιμητή εκείνο του οποίου η κατανομή είναι συγκεντρωμένη γύρω από την άγνωστη παράμετρο θ. Έτσι σε πολλές επαναλήψεις της δειγματοληψίας είναι αναμενόμενο, ο μέσος όρος των τιμών του εκτιμητή T να μας δώσει την τιμή της παραμέτρου θ.

Το δεύτερο κριτήριο, της **ελάχιστης διακύμανσης,** αντιμετωπίζει το πρόβλημα της ύπαρξης δύο ή περισσότερων αμερόληπτων εκτιμητών της ίδιας παραμέτρου. Σύμφωνα με το κριτήριο αυτό, μεταξύ των αμερόληπτων εκτιμητών μιας παραμέτρου, καλύτερος είναι εκείνος που έχει τη μικρότερη διακύμανση. Το κριτήριο της ελάχιστης διακύμανσης έχει ακλόνητα διαισθητικά ερίσματα αφού η διακύμανση είναι ένα μέτρο της μεταβλητότητας του εκτιμητή από τη μέση του τιμή.

Εκτός της αμεροληψίας και της ελάχιστης διακύμανσης, η επάρκεια και η συνέπεια είναι άλλα κριτήρια για την αξιολόγηση εκτιμητών, στα οποία όμως δεν θα επεκταθούμε.

Η διατύπωση κριτηρίων για την επιλογή εκτιμητών αποτελεί τη μια κατηγορία προβλημάτων που εμφανίζονται στη σημειοεκτιμητική. Το άλλο πρόβλημα είναι η ανάπτυξη μεθοδολογιών για την εύρεση εκτιμητών που ικανοποιούν τα προηγούμενα κριτήρια. Τέτοιες μέθοδοι υπάρχουν και επιγραμματικά αναφέρονται: η μέθοδος της μέγιστης πιθανοφάνειας, των ροπών, των ελαχίστων τετραγώνων, της ανισότητας των Cramér-Rao. Η παρουσίαση των μεθόδων αυτών ξεφεύγει από το σκοπό των σημειώσεων.

Με τη χρησιμοποίηση των πιο πάνω μεθοδολογιών βρίσκεται ότι: οι καλύτεροι εκτιμητές για τις παραμέτρους μ και σ^2 ενός κανονικού πληθυσμού είναι το $\bar{X}$ και S^2 αντίστοιχα, ο καλύτερος εκτιμητής της παραμέτρου p της διωνυμικής κατανομής B(n,p) είναι το X/ n με X~

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

B(n,p) ενώ ο καλύτερος εκτιμητής της παραμέτρου λ της Poisson P(λ) είναι το $\bar{X}$. Γενικότερα η μέση τιμή $\bar{X}$ και η διακύμανση S^2 του δείγματος είναι αμερόληπτοι εκτιμητές για τη μέση τιμή και τη διακύμανση αντίστοιχα οποιουδήποτε πληθυσμού.

Παράδειγμα 6.2.1 Είναι γνωστό ότι το ύψος παιδιών προσχολικής ηλικίας ακολουθεί κανονική κατανομή. Αν για 35 παιδιά που εκλέχτηκαν τυχαία το ύψος τους (σε εκατοστά) είναι:

85, 86, 88, 88, 90, 90, 90, 91, 92, 92, 92 93, 93, 94, 94, 94, 95, 95, 96, 96, 96, 97, 97, 97, 98, 99, 99, 100,100, 102,102,102, 103, 103, 104.

Να υπολογιστεί το ποσοστό των παιδιών με ύψος μεταξύ 85 και 90 cm.

Αν X είναι η τυχαία μεταβλητή που παριστά το ύψος παιδιών προσχολικής ηλικίας τότε $X \sim N(\mu, \sigma^2)$. Ζητείται η $P(85 \le X \le 90)$. Το $\bar{X} = \frac{1}{35} \sum_{i=1}^{35} X_i = 95.23$ είναι εκτιμητής για την μ και το $S^2 = \frac{1}{34} \sum_{i=1}^{35} (X_i - \bar{X})^2 = 25.53$ είναι εκτιμητής του σ^2. Έτσι, αν και παρακινδυνευμένα όπως θα φανεί στην επόμενη παράγραφο, μπορεί να θεωρηθεί ότι $X \sim N(95.23, 25.53)$ και επομένως

$$P(85 \le X \le 90) = P\left(\frac{85-95.23}{\sqrt{25.53}} \le \frac{X-95.23}{\sqrt{25.53}} \le \frac{90-95.23}{\sqrt{25.53}} \right)$$

$$= P(-2.02 \le Z \le -1.04) = P(1.04 \le Z \le 2.02)$$

$$= P(-\infty < Z \le 2.02) - P(-\infty < Z \le 1.04)$$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

$= 0.9783 - 0.8508 = 0.1275.$

Δηλ. το 12.75% των παιδιών έχει ύψος μεταξύ 85 και 90 cm.

6.3 Εκτίμηση σε διάστημα - Διαστήματα εμπιστοσύνης

Η εκτίμηση σε διάστημα μιας παραμέτρου, αναφέρεται σε διαδικασίες εύρεσης δυο στατιστικών συναρτήσεων τέτοιων ώστε η τιμή της άγνωστης παραμέτρου να βρίσκεται, με προκαθορισμένη πιθανότητα, στο διάστημα που δημιουργείται από τις τιμές αυτών των στατιστικών συναρτήσεων. Το διάστημα αυτό ονομάζεται διάστημα εμπιστοσύνης. Η εκτίμηση σε διάστημα αποτελεί ένα νέο τρόπο προσέγγισης της άγνωστης παραμέτρου του πληθυσμού και ταυτόχρονα αξιοποιεί τον αντίστοιχο σημειακό εκτιμητή αφού ουσιαστικά παρέχει ένα πιθανοθεωρητικό τρόπο αξιολόγησής του. Στη συνέχεια κατασκευάζονται διαστήματα εμπιστοσύνης για την άγνωστη μέση τιμή μ ενός πληθυσμού που περιγράφεται από την κανονική κατανομή $N(\mu, \sigma^2)$. Η εκτίμηση σε διάστημα της μέσης τιμής κανονικής κατανομής είναι η πλέον αντιπροσωπευτική, λόγω της σπουδαιότητας της κατανομής αυτής στην περιγραφή καθημερινών προβλημάτων, αλλά και γιατί διευκρινίζεται με τον καλύτερο τρόπο η αναγκαιότητα της εκτίμησης σε διάστημα μιας παραμέτρου και σκιαγραφείται η μεθοδολογία κατασκευής διαστημάτων εμπιστοσύνης γι' αυτήν. Επίσης διαστήματα εμπιστοσύνης για τη διαφορά των μέσων τιμών δυο ανεξάρτητων κανονικών πληθυσμών θα κατασκευαστούν στη συνέχεια.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

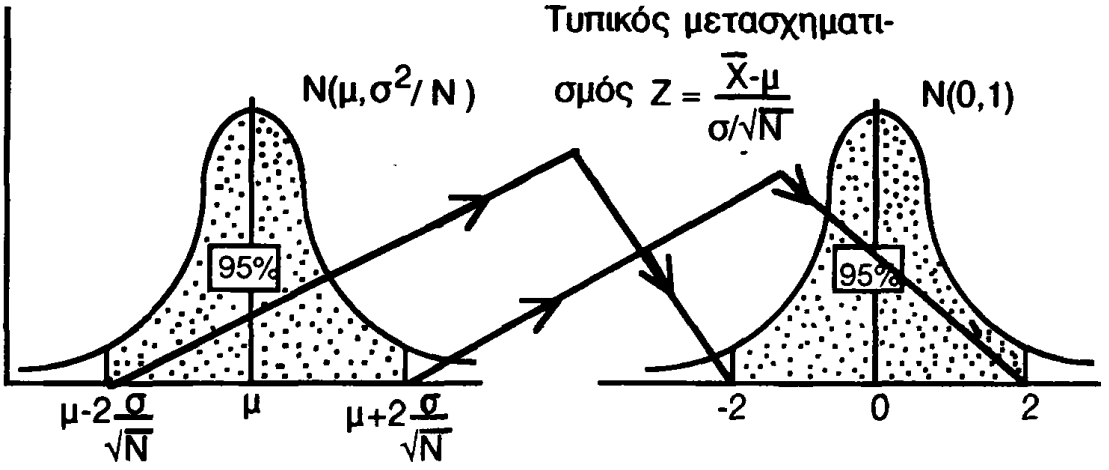
6.3.1 Διάστημα εμπιστοσύνης για τη μέση τιμή κανονικού πληθυσμού

Η αντικατάσταση της άγνωστης παραμέτρου μ ενός κανονικού πληθυσμού $N(\mu, \sigma^2)$, από τον εκτιμητή της $\bar{X}$, όπως έγινε στο Παράδειγμα 6.2.1, είναι παρακινδυνευμένη ακόμη και αν ο χρησιμοποιούμενος εκτιμητής ικανοποιεί τα κριτήρια της αμεροληψίας και της ελάχιστης διακύμανσης. Ο λόγος είναι ότι η εκτιμήτρια συνάρτηση και κάθε στατιστική συνάρτηση γενικότερα, παίρνει κατά κανόνα διαφορετικές τιμές όταν υπολογίζεται σε διαφορετικά δείγματα τα οποία προέρχονται από τον ίδιο πληθυσμό. Έτσι αν εκλεγούν μερικά δείγματα, ίδιου μεγέθους από την πληθυσμό $N(\mu, \sigma^2)$ και υπολογιστεί η μέση τους τιμή $\bar{X}$ είναι σχεδόν αδύνατο να προκύψει η ίδια τιμή για όλα τα δείγματα. Οι τιμές του $\bar{X}$ θα διαφέρουν μεταξύ τους όπως επίσης θα διαφέρουν και από τη μέση τιμή μ του πληθυσμού. Οι αποκλίσεις των τιμών αυτών από τη μέση τιμή του πληθυσμού ονομάζονται σφάλματα δειγματοληψίας και δεν οφείλονται σε λάθη που έγιναν κατά τη διαδικασία δημιουργίας των δειγμάτων. Οφείλονται στις αποκλίσεις, στις διαφορές που υπάρχουν στα μέλη του πληθυσμού και που φυσικά μεταφέρονται στα δείγματα που εξετάζονται και κατ' επέκταση στις στατιστικές συναρτήσεις που δημιουργούνται από τα δείγματα αυτά. Έτσι η αξιοπιστία του εκτιμητή $\bar{X}$ της μέσης τιμής μ του κανονικού πληθυσμού εξαρτάται από τη μεταβλητότητα του πληθυσμού δηλαδή από τη διακύμανσή του σ^2. Σ' αυτήν ακριβώς την εξάρτηση στηρίζεται η κατασκευή διαστημάτων εμπιστοσύνης για την παράμετρο μ, που περιγράφεται πιο κάτω θεωρώντας τη διακύμανση σ^2 πρώτα γνωστή και στη συνέχεια άγνωστη.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

6.3.1.1 Διάστημα εμπιστοσύνης για την μ όταν η σ^2 είναι γνωστή

Ας θεωρήσουμε ένα τυχαίο δείγμα $X_1, X_2,...,X_N$ από κανονικό πληθυσμό $N(\mu, \sigma^2)$ με άγνωστη μέση τιμή μ και γνωστή διακύμανση σ^2. Η μέση τιμή $\bar{X}$ του δείγματος αυτού αποδεικνύεται ότι ακολουθεί κανονική κατανομή $N(\mu, \frac{\sigma^2}{N})$, και επομένως ο τυπικός μετασχηματισμός $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{N}}$ τυπική κανονική κατανομή N(0,1).



Έτσι λόγω των ιδιοτήτων της κανονικής κατανομής το 95% των μέσων τιμών $\bar{X}$ των δειγμάτων που εκλέγονται από τον αρχικό πληθυσμό, βρίσκονται εντός δυο τυπικών αποκλίσεων από το μ δηλαδή στο διάστημα $(\bar{X} - 2\frac{\sigma}{\sqrt{N}}, \bar{X} + 2\frac{\sigma}{\sqrt{N}})$ ή ισοδύναμα το 95% των $\frac{\bar{X}-\mu}{\sigma/\sqrt{N}}$ στο διάστημα (-2, 2) όπως φαίνεται και στο προηγούμενο σχήμα. Επομένως με πιθανότητα 95% έχουμε

$$-2 \le \frac{\bar{X}-\mu}{\sigma/\sqrt{N}} \le 2 \text{ ή ισοδύναμα } \bar{X} - 2\frac{\sigma}{\sqrt{N}} \le \mu \le \bar{X} + 2\frac{\sigma}{\sqrt{N}}.$$

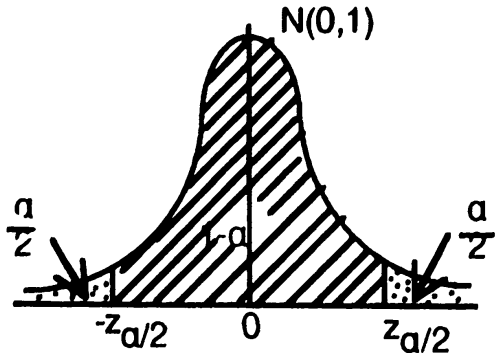
ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Συνεπώς μπορούμε να πούμε ότι με πιθανότητα 95% το διάστημα $(\bar{X} - 2\frac{\sigma}{\sqrt{N}}, \bar{X} + 2\frac{\sigma}{\sqrt{N}})$ ή $\bar{X} \pm 2\frac{\sigma}{\sqrt{N}}$, περιέχει την τιμή της άγνωστης μέσης τιμής μ του πληθυσμού.

Ένα διάστημα, όπως το προηγούμενο, ονομάζεται **διάστημα εμπιστοσύνης** (confidence interval) για την παράμετρο μ με **βαθμό εμπιστοσύνης** (confidence coefficient) 95%. Τα ακραία σημεία $\bar{X} - 2\frac{\sigma}{\sqrt{N}}$ και $\bar{X} + 2\frac{\sigma}{\sqrt{N}}$ λέγονται **όρια εμπιστοσύνης** (confidence limits).

Η κατασκευή του προγούμενου 95% διαστήματος εμπιστοσύνης για την μ υποδεικνύει τον τρόπο κατασκευής διαστήματος εμπιστοσύνης, με βαθμό εμπιστοσύνης γενικά 100(1-α)% με α ένα αριθμό μεταξύ 0 και 1. Και πάλι επειδή το στατιστικό $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{N}}$ ακολουθεί κατανομή N(0,1) , όπως φαίνεται και από το παρακάτω σχήμα υπάρχουν συμμετρικά ως προς το 0, σημεία $z_{\alpha/2}$ και $-z_{\alpha/2}$ τέτοια ώστε με πιθανότητα 1-α να ισχύει:

$$-z_{\alpha/2} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{N}} \le z_{\alpha/2} \text{ ή ισοδύναμα } \bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}} \le \mu \le \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}$$



Επομένως

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Το διάστημα εμπιστοσύνης για τη μέση τιμή μ κανονικού πληθυσμού $N(\mu,\sigma^2)$ με γνωστή διακύμανση σ^2 και βαθμό εμπιστοσύνης $100(1-\alpha)\%$, $0<\alpha<1$ είναι το:

$$(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}},\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}) \text{ ή } \bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{N}}.$$

Στη συνέχεια περιγράφεται η μεθοδολογία προσδιορισμού των σημείων $z_{\alpha/2}$, από τον πίνακα της τυπικής κανονικής κατανομής. Το σημείο $z_{\alpha/2}$ λέγεται άνω αντίστροφο εκατοστιαίο σημείο της τυπικής κανονικής κατανομής και είναι το σημείο για το οποίο:

$P(Z \geq z_{\alpha/2}) = \frac{\alpha}{2}$ και επομένως $P(-\infty < Z \leq z_{\alpha/2}) = 1-\frac{\alpha}{2}$ με $Z \sim N(0,1)$. Έτσι προσδιορίζεται στο κύριο σώμα του πίνακα της τυπικής κανονικής κατανομής πιθανότητα ίση με $1-\frac{\alpha}{2}$. Το σημείο της πρώτης στήλης και της πρώτης γραμμής του πίνακα αυτού που αντιστοιχεί στην τιμή $1-\frac{\alpha}{2}$ είναι το ζητούμενο σημείο $z_{\alpha/2}$.

Με τη βοήθεια του πίνακα της τυπικής κανονικής κατανομής, μέρος του οποίου παρατίθεται στη συνέχεια, θα προσδιορίσουμε για παράδειγμα το σημείο $z_{\alpha/2}$ για διάστημα εμπιστοσύνης με βαθμό 90%. Είναι $100(1-\alpha)\% = 90\%$ ή $1-\alpha = 0.9$ ή $\alpha=0.1$ και επομένως $\frac{\alpha}{2} = 0.05$. Ζητείται δηλαδή το $z_{0.05}$. Είναι $1-\frac{\alpha}{2} = 1-0.05 = 0.95$.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

z_o	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Το 0.95 δεν υπάρχει στο κύριο σώμα του παραπάνω πίνακα. Διαλέγουμε την πλησιέστερη τιμή 0.9495 η οποία αντιστοιχεί στο σημείο 1.64. Επομένως για 90% διάστημα εμπιστοσύνης το $z_{a/2} = 1.64$. Ανάλογα μπορούν να προσδιοριστούν τα $z_{a/2}$ για 95% και 99% διαστήματα εμπιστοσύνης που είναι αντίστοιχα 1.96 και 2.58. Τονίζεται ότι οι βαθμοί εμιστοσύνης 90%, 95% και 99% χρησιμοποιούνται συνήθως στην πράξη.

Παράδειγμα 6.3.1.1. Οι βαθμοί 16 φοιτητών στις εξετάσεις του παρόντος μαθήματος είναι:

3, 5, 4, 7, 8, 9, 9, 8, 5, 5, 6, 8, 6, 9, 10, 10.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Αν είναι γνωστό από προηγούμενες μελέτες ότι τα δεδομένα ακολουθούν κανονική κατανομή με διακύμανση $\sigma^2 = 4$, να εκτιμηθεί η μέση βαθμολογία όλων των φοιτητών και να κατασκευαστεί ένα 95% διάστημα εμπιστοσύνης γι' αυτή.

Αν X είναι η τυχαία μεταβλητή που παριστά το βαθμό ενός φοιτητή τότε $X \sim N(\mu, 4)$ με μ τη μέση βαθμολογία όλων των φοιτητών. Ο καλύτερος εκτιμητής για το μ είναι το $\bar{X}$. Εύκολα υπολογίζεται ότι:

$$\bar{X} = \frac{1}{16} \sum_{i=1}^{16} X_i = 7.$$

Έτσι κατά μέσο όρο η βαθμολογία των φοιτητών είναι ίση με 7. Η αξιοπιστία της εκτίμησης αυτής εκφράζεται από το διάστημα εμπιστοσύνης για την μ που είναι:

$$(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}}, \quad \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}).$$

Για 95% διάστημα εμπιστοσύνης το $z_{\alpha/2} = 1.96$. Επίσης $N = 16$ έτσι $\sqrt{N} = 4$ και $\sigma^2 = 4$ έτσι $\sigma = 2$. Επομένως το διάστημα εμπιστοσύνης για την μ είναι:

$$(7 - 1.96\frac{2}{4}, \ 7+1.96\frac{2}{4}) \text{ ή } (7-0.98, 7+0.98) \text{ ή } (6.02, 7.98)$$

και καθιστά την εκτίμηση της μέσης βαθμολογίας από το 7 αξιόπιστη.

Παράδειγμα 6.3.1.2: Ένας ερευνητής πιστεύει ότι το ύψος των ατόμων ηλικίας 30-35 ετών ακολουθεί την κανονική κατανομή με τυπική απόκλιση 4.5 cm. Ένα τυχαίο δείγμα που πήρε από τον

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

πληθυσμό, μεγέθους 25, έδωσε μέση τιμή 160 cm. Ζητά να κατασκευαστεί το 99% διάστημα εμπιστοσύνης για το μέσο ύψος των ατόμων ηλικίας 30 έως 35 ετών.

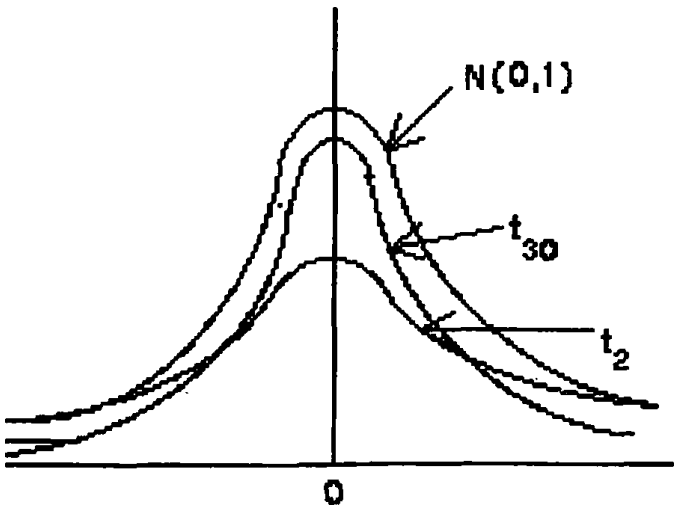
Το διάστημα εμπιστοσύνης για το μέσο ύψος μ είναι $(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}}, \quad \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}})$. Είναι $z_{\alpha/2} = 2.58$, $\sigma = 4.5$, $\bar{X} = 160$ και $\sqrt{N} = 5$. Έτσι το ζητούμενο διάστημα είναι $(160-2.58\ \frac{4.5}{5}$, $160+ 2.58\ \frac{4.5}{5})$ ή (157.7, 162.32).

6.3.1.2 Η κατανομή t .

Στην προηγούμενη παράγραφο κατασκευάστηκε διάστημα εμπιστοσύνης για τη μέση τιμή κανονικού πληθυσμού όταν η διακύμανσή του είναι γνωστή. Η υπόθεση όμως αυτή είναι μάλλον απίθανη. Συνήθως η διακύμανση του κανονικού πληθυσμού είναι και αυτή, όπως και η μέση του τιμή, άγνωστη. Η διαδικασία που ακολουθήθηκε προηγούμενα, δεν μπορεί τώρα να χρησιμοποιηθεί αφού η ποσότητα $\sigma/\sqrt{N}$ δεν μπορεί να υπολογιστεί. Στην περίπτωση αυτή είναι φυσιολογική πλέον η αντικατάσταση της σ, στην έκφραση $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{N}}$, από τον εκτιμητή της S, την δειγματική τυπική απόκλιση. Έτσι από το στατιστικό Z προκύπτει το ανάλογο στατιστικό $t = \frac{\bar{X}-\mu}{S/\sqrt{N}}$. Το στατιστικό αυτό αναμένεται να έχει παρόμοια κατανομή με εκείνη του Z, που είναι η N(0,1), όταν το μέγεθος του δείγματος N είναι μεγάλο γιατί σ' αυτή την περίπτωση η S αποτελεί πολύ καλή προσέγγιση της άγνωστης τυπικής απόκλισης σ του πληθυσμού. Έτσι όταν το μέγεθος του δείγματος N είναι μεγάλο ($N \geq 30$) η κατανομή

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

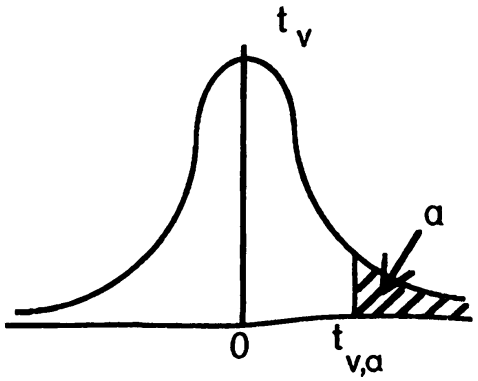
του $t = \frac{\bar{X}-\mu}{S/\sqrt{N}}$ είναι ουσιαστικά και πάλι η τυπική κανονική N(0,1). Όταν όμως το N είναι μικρό (N < 30), τότε η S διαφέρει σημαντικά από την σ. Σ' αυτή την περίπτωση η στατιστική συνάρτηση $t = \frac{\bar{X}-\mu}{S/\sqrt{N}}$ λέγεται ότι ακολουθεί την t-**κατανομή** (ή κατανομή Student) με N-1 **βαθμούς ελευθερίας** και συμβολίζεται t_{N-1}.



Η t - κατανομή έχει σχήμα καμπάνας (όπως και η κανονική κατανομή) και είναι συμμετρική γύρω από τη μέση της τιμή που είναι το 0. Η διακύμανσή της είναι μεγαλύτερη από τη μονάδα γεγονός που την κάνει να είναι χαμηλότερη της τυπικής κανονικής κατανομής. Επίσης για κάθε διαφορετικό βαθμό ελευθερίας προκύπτει και μια διαφορετική t - κατανομή. Όσο οι βαθμοί ελευθερίας μεγαλώνουν, τόσο η t - κατανομή πλησιάζει καλύτερα την N(0,1), όπως φαίνεται και στο παραπάνω σχήμα.

Για την t - κατανομή υπάρχουν διαθέσιμοι πίνακες, όπως αυτός του Παραρτήματος, μέρος του οποίου δίνεται παρακάτω. Ο πίνακας αυτός ουσιαστικά, για διάφορους βαθμούς ελευθερίας ν της t - κατανομής, δίνει το σημείο $t_{v,a}$ για το οποίο ισχύει: $P(t_v \geq t_{v,a}) = a$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ



όπως φαίνεται και στο πιο πάνω σχήμα. Το $t_{v,a}$ λέγεται άνω α-**αντίστροφο εκατοστιαίο σημείο** της κατανομής t_v .

v	a=.10	a=.05	a=.025	a=.01	a=.005	v
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24

Τα σημεία αυτά βρίσκονται στη διασταύρωση της γραμμής που καθορίζεται από τους βαθμούς ελευθερίας v και της στήλης που καθορίζεται από το ποσοστό α. Έτσι για παράδειγμα $t_{21,0.05} = 1.721$ και $t_{14,0.01} = 2.624$.

Επίσης ο παραπάνω πίνακας μπορεί να χρησιμοποιηθεί και για τον υπολογισμό πιθανοτήτων που εκφράζονται μέσω τυχαίων μεταβλητών με κατανομή t_v . Η χρήση του για το σκοπό αυτό περιγράφεται στο ακόλουθο παράδειγμα:

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

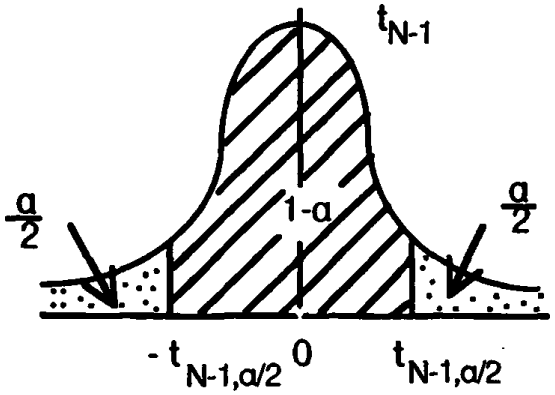
$$P(-1.33 \le t_{18} \le 2.1) = P(-\infty < t_{18} \le 2.1) - P(-\infty < t_{18} \le -1.33)$$

$$= 1-P(t_{18} \ge 2.1) - P(t_{18} \ge 1.33)$$

$$= 1 - 0.025 - 0.1 = 0.875.$$

6.3.1.3. Διάστημα εμπιστοσύνης για την μ όταν η σ^2 είναι άγνωστη

Έστω και πάλι $X_1, X_2,\ldots,X_N$ είναι ένα τυχαίο δείγμα από κανονικό πληθυσμό $N(\mu, \sigma^2)$, με σ άγνωστο. Για την κατασκευή διαστήματος εμπιστοσύνης για την παράμετρο μ, ακολουθείται η διαδικασία της παραγράφου 6.3.1.1 χρησιμοποιώντας την t-κατανομή αντί της τυπικής κανονικής.



Επειδή το στατιστικό $t = \dfrac{\bar{X}-\mu}{S/\sqrt{N}}$ έχει κατανομή την t_{N-1} υπάρχουν συμμετρικά σημεία $-t_{N-1,\,\alpha/2}$ και $t_{N-1,\,\alpha/2}$ τέτοια ώστε:

$$\bar{X} - t_{N-1,\alpha/2}\frac{S}{\sqrt{N}} \le \mu \le \bar{X} + t_{N-1,\alpha/2}\frac{S}{\sqrt{N}},$$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

με πιθανότητα 100(1-α)%, 0<α<1. Έτσι

Το διάστημα εμπιστοσύνης για τη μέση τιμή κανονικού πληθυσμού $N(\mu, \sigma^2)$ με άγνωστη διακύμανση σ^2 και βαθμό εμπιστοσύνης 100(1-α)%, 0<α<1 είναι το

$$\left(\bar{X} - t_{N-1,\alpha/2}\frac{S}{\sqrt{N}},\ \bar{X} + t_{N-1,\alpha/2}\frac{S}{\sqrt{N}}\right) \text{ ή } \bar{X} \pm t_{N-1,\alpha/2}\frac{S}{\sqrt{N}}.$$

Παράδειγμα 6.3.1.3. Αν τα δεδομένα του Παραδείγματος 6.2.1 προέρχονται από κανονική κατανομή να υπολογιστεί ένα 95% διάστημα εμπιστοσύνης για το μέσο ύψος των παιδιών προσχολικής ηλικίας.

Από τα δεδομένα του Παραδείγματος 6.2.1 προκύπτει ότι $\bar{X} = 95.23$ και $\sqrt{S} = \sqrt{25.53} = 5.05$. Επίσης 100(1-α)% = 0.95% ή α = 0.05. Έτσι $\frac{\alpha}{2} = 0.25$ και $t_{N-1,\alpha/2} = t_{34,0.025} = 1.96$. Το 95% διάστημα εμπιστοσύνης για την μ είναι:

$$\left(95.23 - 1.96\frac{5.05}{\sqrt{35}},\ 95.23 + 1.96\frac{5.05}{\sqrt{35}}\right) \text{ ή } (93.56, 96.9).$$

Παράδειγμα 6.3.1.4: Ένα δείγμα από 16 κορίτσια δέκα ετών έδωσε μέσο βάρος 32 κιλά και τυπική απόκλιση 5 κιλά. Υποθέτοντας ότι το βάρος ακολουθεί κανονική κατανομή, να βρεθεί το 90% διάστημα εμπιστοσύνης για το μέσο βάρος του πληθυσμού των κοριτσιών ηλικίας δέκα ετών.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Από τα δεδομένα προκύπτει ότι: $\bar{X} = 32$, $S = 5$, $N = 16$. Επίσης $100(1-\alpha)\% = 90\%$ ή $1-\alpha = 0.9$ ή $\alpha = 0.1$ ή $\frac{\alpha}{2} = 0.05$. Έτσι $t_{15,0.05} = 1.753$. Το 90% διάστημα εμπιστοσύνης για το μέσο βάρος του πληθυσμού των κοριτσιών ηλικίας 10 ετών είναι:

$$(32 - 1.753 \frac{5}{\sqrt{16}},\ 32+1.753 \frac{5}{\sqrt{16}}) \text{ ή } (29.8, 34.2).$$

6.3.2 Διάστημα εμπιστοσύνης για τη διαφορά των μέσων τιμών δύο ανεξάρτητων κανονικών πληθυσμών

Διόρθ. 22/5/2000.

Αρκετά συχνά εμφανίζεται στην πράξη, η ανάγκη διερεύνησης της συμπεριφοράς δυο πληθυσμών ως προς ένα κοινό τους χαρακτηριστικό. Ας θεωρήσουμε για παράδειγμα τον πληθυσμό των παιδιών προσχολικής ηλικίας που προέρχονται από οικογένειες με υψηλό εισόδημα και τον πληθυσμό των παιδιών που προέρχονται από οικογένειες με χαμηλό εισόδημα. Κοινό χαρακτηριστικό και των δυο πληθυσμών είναι ο δείκτης νοημοσύνης τους. Έστω μ_1 είναι ο μέσος δείκτης νοημοσύνης του πληθυσμού των παιδιών από υψηλόμισθες οικογένειες και μ_2 ο μέσος δείκτης νοημοσύνης του πληθυσμού των παιδών από χαμηλόμισθες οικογένειες. Οποιαδήποτε σύγκριση των δεικτών νοημοσύνης των δυο πληθυσμών εκφράζεται με τη βοήθεια των μέσων τιμών μ_1 και μ_2 των δυο πληθυσμών. Ειδικότερα η διαφορά $\mu_1 - \mu_2$ αποτελεί ένα μέτρο της διαφορετικότητας της συμπεριφοράς των δύο πληθυσμών ως προς το κοινό τους χαρακτηριστικό που αντιπροσωπεύεται από τα μ_1 και μ_2. Τιμή της διαφοράς $\mu_1 - \mu_2$ κοντά στο 0 σημαίνει ότι κατά μέσο όρο δεν υπάρχει διαφορά στους δείκτες νοημοσύνης των παιδιών των δυο πληθυσμών. Θετική τιμή για

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

τη διαφορά $\mu_1 - \mu_2$ σημαίνει ότι κατά μέσο όρο τα παιδιά που προέρχονται από υψηλόμισθες οικογένειες έχουν υψηλότερο δείκτη νοημοσύνης από τα παιδιά που προέρχονται από χαμηλόμισθες οικογένειες ενώ στο αντίθετο συμπέρασμα καταλήγουμε όταν η τιμή της διαφοράς $\mu_1 - \mu_2$ είναι αρνητική.

Στην πράξη όμως η τιμή της διαφοράς $\mu_1 - \mu_2$ είναι κατά κανόνα άγνωστη αφού άγνωστες είναι οι μέσες τιμές μ_1 και μ_2. Η προσέγγιση της διαφοράς $\mu_1 - \mu_2$ μπορεί να επιτευχθεί με τη βοήθεια στατιστικών μεθοδολογιών η εφαρμογή των οποίων στηρίζεται στις εξής υποθέσεις - εξιδανικεύσεις:

i) Οι δυο πληθυσμοί περιγράφονται από κανονικές κατανομές με μέσες τιμές μ_1 και μ_2 και διακυμάνσεις σ_1^2 και σ_2^2.

ii) Οι πληθυσμοί είναι ανεξάρτητοι γεγονός που διαισθητικά σημαίνει ότι η δραστηριότητα - συμπεριφορά των μελών του ενός πληθυσμού δεν επηρεάζει την συμπεριφορά των μελών του άλλου πληθυσμού.

Αν ισχύουν οι προηγούμενες υποθέσεις τότε ο καλύτερος εκτιμητής, η καλύτερη προσέγγιση της διαφοράς $\mu_1 - \mu_2$ είναι ο $\bar{X}_1 - \bar{X}_2$, με $\bar{X}_1$ και $\bar{X}_2$ τις μέσες τιμές δύο τυχαίων δειγμάτων που λαμβάνονται από τους δυο ανεξάρτητους κανονικούς πληθυσμούς. Έτσι οι συγκρίσεις του δείκτη νοημοσύνης των παιδιών από υψηλόμισθες και χαμηλόμισθες οικογένειες μπορεί να γίνουν στη βάση της διαφοράς $\bar{X}_1 - \bar{X}_2$. Η αξιοπιστία όμως των συγκρίσεων και κατ' επέκταση των συμπερασμάτων βασίζεται στην αξιοπιστία του εκτιμητή $\bar{X}_1 - \bar{X}_2$. Η αξιολόγηση του εκτιμητή $\bar{X}_1 - \bar{X}_2$ επιτυγχάνεται από το διάστημα εμπιστοσύνης της παραμέτρου $\mu_1 - \mu_2$ που εκτιμά.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Στην συνέχεια κατασκευάζονται τα διαστήματα εμπιστοσύνης της διαφοράς $\mu_1 - \mu_2$ θεωρώντας τις διακυμάνσεις σ_1^2 και σ_2^2 πρώτα γνωστές και μετά άγνωστες αλλά ίσες μεταξύ τους.

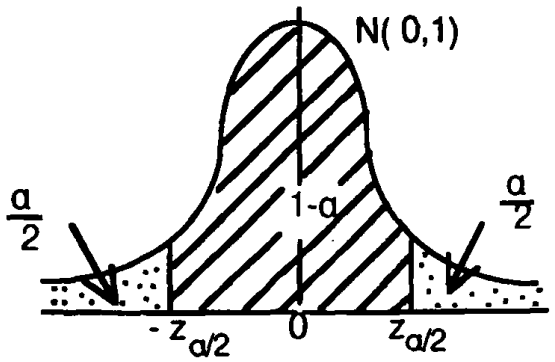
6.3.2.1 Διάστημα εμπιστοσύνης για την $\mu_1 - \mu_2$ όταν οι σ_1^2 και σ_2^2 είναι γνωστές

Θεωρούμε δύο τυχαία δείγματα τα $X_{11}, X_{12}, \ldots, X_{1N_1}$ και $X_{21}, X_{22}, \ldots, X_{2N_2}$ που προέρχονται αντίστοιχα από τους ανεξάρτητους κανονικούς πληθυσμούς $N(\mu_1, \sigma_1^2)$ και $N(\mu_2, \sigma_2^2)$ με διακυμάνσεις σ_1^2 και σ_2^2 γνωστές.

Η κατασκευή διαστήματος εμπιστοσύνης για τη διαφορά $\mu_1-\mu_2$ στηρίζεται σ' ένα πιθανοθεωρητικό αποτέλεσμα σύμφωνα με το οποίο η διαφορά $\bar{X}_1 - \bar{X}_2$ των μέσων τιμών των δυο δειγμάτων ακολουθεί κανονική κατανομή $N(\mu_1-\mu_2, \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2})$ και επομένως το στατιστικό

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

έχει τυπική κανονική κατανομή $N(0,1)$.



ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Έτσι υπάρχουν συμμετρικά σημεία $-z_{\alpha/2}$ και $z_{\alpha/2}$ τέτοια ώστε

$$-z_{\alpha/2} \le \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \le z_{\alpha/2},$$

ή ισοδύναμα με πιθανότητα $100(1-\alpha)\%$, $0 < \alpha < 1$:

$$\bar{X}_1 - \bar{X}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} \le \mu_1 - \mu_2 \le \bar{X}_1 - \bar{X}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}.$$

Το διάστημα εμπιστοσύνης για τη διαφορά $\mu_1 - \mu_2$ των μέσων τιμών δυο ανεξάρτητων κανονικών πληθυσμών $N(\mu_1, \sigma_1^2)$ και $N(\mu_2, \sigma_2^2)$ με γνωστές διακυμάνσεις σ_1^2 και σ_2^2 και βαθμό εμπιστοσύνης $100(1-\alpha)\%$, $0 < \alpha < 1$ είναι το

$$\left((\bar{X}_1 - \bar{X}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} \, , (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} \right)$$

ή

$$(\bar{X}_1 - \bar{X}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}.$$

Παράδειγμα 6.3.2.1: Δυο ανεξάρτητες ομάδες παιδιών 1 και 2 αποτελούνται από 9 παιδιά η κάθε μια. Τα παιδιά της ομάδας 1 προέρχονται από υπερτασικούς γονείς ενώ τα παιδιά της ομάδας 2 προέρχονται από γονείς με κανονική πίεση. Η πίεση του αίματος μετριέται στα παιδιά των δυο ομάδων με τα ακόλουθα αποτελέσματα:

Ομάδα 1: 100, 102 96, 106, 110, 120, 112, 112, 90

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Ομάδα 2: 104, 88, 100, 98, 102, 92, 96, 100, 96.

Υποθέτουμε ότι η πίεση του αίματος ακολουθεί κανονική κατανομή και ότι οι διακυμάνσεις για τους δυο πληθυσμούς είναι αντίστοιχα: $\sigma_1^2 = 70$ και $\sigma_2^2 = 22$. Να υπολογιστεί ο καλύτερος εκτιμητής της διαφοράς των μέσων πιέσεων των δυο ομάδων και να κατασκευαστεί το 95% διάστημα εμπιστοσύνης για τη διαφορά αυτή.

Έστω $N(\mu_1, \sigma_1^2)$ και $N(\mu_2, \sigma_2^2)$, $\sigma_1^2 = 70$, $\sigma_2^2 = 22$, είναι οι πληθυσμοί των παιδιών που προέρχονται από υπερτασικούς και μη γονείς αντίστοιχα. Ο καλύτερος εκτιμητής της διαφοράς $\mu_1-\mu_2$ είναι ο $\bar{X}_1-\bar{X}_2$ με $\bar{X}_1$ και $\bar{X}_2$ τις μέσες τιμές των δειγμάτων από τους δυο πληθυσμούς που είναι:

$$\bar{X}_1 = \frac{1}{9}\sum_{i=1}^{9} X_{1i} = \frac{1}{9}(100+102+\ldots+90) = \frac{1}{9}948 = 105.3,$$

$$\bar{X}_2 = \frac{1}{9}\sum_{i=1}^{9} X_{2i} = \frac{1}{9}(104+88+\ldots+96) = \frac{1}{9}\ 876 = 97.3.$$

Ο καλύτερος εκτιμητής της $\mu_1-\mu_2$ είναι ο $\bar{X}_1-\bar{X}_2 = 105.3 - 97.3 = 8$, απ' τον οποίο συμπεραίνουμε ότι κατά μέσο όρο τα παιδιά που προέρχονται από υπερτασικούς γονείς έχουν υψηλότερη πίεση απ' τα παιδιά που προέρχονται από φυσιολογικούς γονείς, ή η αρτηριακή πίεση έχει σχέση με την κληρονομικότητα. Η αξιοπιστία του εκτιμητή $\bar{X}_1-\bar{X}_2$ και κατ' επέκταση του προηγούμενου συμπεράσματος ελέγχεται με το διάστημα εμπιστοσύνης της διαφοράς $\mu_1-\mu_2$. Το 95% διάστημα εμπιστοσύνης της διάφορας αυτής είναι:

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

$$\left((\bar{X}_1 - \bar{X}_2) - z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} \; , \; (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} \right)$$

Αντικαθιστώντας στην έκφραση αυτή το $\bar{X}_1 = 105.3$, $\bar{X}_2 = 97.3$, $\sigma_1^2 = 70$, $\sigma_2^2 = 22$, $N_1 = 9$, $N_2 = 9$, και $z_{\alpha/2} = 1.96$ προκύπτει το 95% διάστημα εμπιστοσύνης για την $\mu_1 - \mu_2$ που είναι: (1.73, 14.27).

6.3.2.2 Διάστημα Εμπιστοσύνης για την $\mu_1 - \mu_2$ όταν οι σ_1^2 , σ_2^2 είναι άγνωστες αλλά $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Θεωρούμε και πάλι δυο τυχαία δείγματα $X_{11}, X_{12}, \ldots, X_{1N_1}$ και $X_{21}, X_{22}, \ldots, X_{2N_2}$ που προέρχονται από τους ανεξάρτητους κανονικούς πληθυσμούς $N(\mu_1, \sigma^2)$ και $N(\mu_2, \sigma^2)$ με μέσες τιμές μ_1 και μ_2 αντίστοιχα και κοινή αλλά άγνωστη διακύμανση σ^2. Αν $\bar{X}_1$ και $\bar{X}_2$ είναι οι μέσες τιμές των δυο δειγμάτων και S_1^2 , S_2^2 οι αντίστοιχες διακυμάνσεις τότε μπορεί να αποδειχτεί ότι το στατιστικό:

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \quad \text{με} \quad S^2 = \frac{(N_1 - 1)\, S_1^2 + (N_2 - 1)\, S_2^2}{N_1 + N_2 - 2}$$

έχει κατανομή την t με N_1+N_2-2 βαθμούς ελευθερίας. Στηριζόμενοι στο αποτέλεσμα αυτό και ακολουθώντας τη διαδικασία της προηγούμενης παραγράφου:

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Το διάστημα εμπιστοσύνης για τη διαφορά $\mu_1-\mu_2$ των μέσων τιμών δυο ανεξάρτητων κανονικών πληθυσμών $N(\mu_1, \sigma^2)$ και $N(\mu_2, \sigma^2)$ με κοινή αλλά άγνωστη διακύμανση σ^2 και βαθμό εμπιστοσύνης $100(1-\alpha)\%$, $0 < \alpha < 1$ είναι το:

$$\left((\bar{X}_1 - \bar{X}_2) - t_{N_1+N_2-2,\, \alpha/2}\, S \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}\ ,\ (\bar{X}_1 - \bar{X}_2) + t_{N_1+N_2-2,\, \alpha/2}\, S \sqrt{\frac{1}{N_1} + \frac{1}{N_2}} \right),$$

ή

$$(\bar{X}_1 - \bar{X}_2) \pm t_{N_1+N_2-2,\, \alpha/2}\, S \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}.$$

Τα $\bar{X}_1$, $\bar{X}_2$, S_1^2, S_2^2 και S^2 ορίζονται:

$$\bar{X}_1 = \frac{1}{N_1} \sum_{i=1}^{N_1} X_{1i}\,,\quad \bar{X}_2 = \frac{1}{N_2} \sum_{i=1}^{N_2} X_{2i}\ ,\quad S_1^2 = \frac{1}{N_1-1} \sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2,$$

$$S_2^2 = \frac{1}{N_2-1} \sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2 \text{ και } S^2 = \frac{(N_1-1)\, S_1^2 + (N_2-1)\, S_2^2}{N_1+N_2-2}\ .$$

Παράδειγμα 6.3.2.2: Θεωρούμε και πάλι τα δεδομένα του παραδείγματος 6.3.2.1 υποθέτουμε όμως ότι η πίεση του αίματος ακολουθεί κανονική κατανομή και ότι οι δύο πληθυσμοί έχουν κοινή

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

αλλά άγνωστη διακύμανση. Ζητείται το 95% διάστημα εμπιστοσύνης για την διαφορά $\mu_1-\mu_2$.

Στην περίπτωση αυτή το 100(1-α)% διάστημα εμπιστοσύνης είναι:

$$\left((\bar{X}_1 - \bar{X}_2) - t_{N_1+N_2-2,\, \alpha/2}\, S \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}\ ,\right.$$

$$\left. (\bar{X}_1 - \bar{X}_2) + t_{N_1+N_2-2,\, \alpha/2}\, S \sqrt{\frac{1}{N_1} + \frac{1}{N_2}} \right).$$

Από τα δεδομένα υπολογίζεται:

$$\bar{X}_1 = 105.3,\quad \bar{X}_2 = 97.3,\quad t_{N_1+N_2-2,\, \alpha/2} = t_{16,0.025} = 2.12,$$

$$S_1^2 = 86,\quad S_2^2 = 25 \quad \text{και} \quad S^2 = 55.5\ .$$

Έτσι το ζητούμενο διάστημα εμπιστοσύνης είναι:

$$(8 - (2.12)\ (7.45)\ (0.47),\ 8 + (2.12)\ (7.45)\ (0.47))\ ,$$

ή (0.58, 15.42).

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

6.3.3 Διάστημα εμπιστοσύνης για τη μέση τιμή οποιουδήποτε πληθυσμού

Στις προηγούμενες παραγράφους κατασκευάστηκαν διαστήματα εμπιστοσύνης για τη μέση τιμή ενός κανονικού πληθυσμού. Παρά το ευρύ πεδίο εφαρμογών της κανονικής κατανομής δεν είναι πάντοτε σκόπιμο να υποθέτουμε ότι ο υπό μελέτη πληθυσμός περιγράφεται από την κανονική κατανομή. Δημιουργείται έτσι το πρόβλημα της κατασκευής διαστημάτων εμπιστοσύνης για τη μέση τιμή οποιουδήποτε πληθυσμού. Η λύση στο πρόβλημα αυτό δίνεται με τη βοήθεια του **Κεντρικού Οριακού Θεωρήματος**, ενός καθαρά πιθανοθεωρητικού αποτελέσματος, σύμφωνα με το οποίο η κατανομή της μέσης τιμής ενός τυχαίου δείγματος το οποίο λαμβάνεται από οποιοδήποτε πληθυσμό, τείνει να είναι μια κανονική κατανομή. Η αυστηρή διατύπωση του θεωρήματος αυτού είναι η ακόλουθη.

Κεντρικό Οριακό Θεώρημα: Έστω $X_1, X_2,...,X_N$ είναι ένα τυχαίο δείγμα μεγέθους N από ένα πληθυσμό με μέση τιμή μ και διακύμανση σ^2. Αν το N είναι μεγάλο, τότε η μέση τιμή του δείγματος $\bar{X}$ έχει κατά προσέγγιση κανονική κατανομή $N(\mu, \frac{\sigma^2}{N})$ με μέση τιμή μ και διακύμανση σ^2/N. Έτσι το στατιστικό $\frac{\bar{X}-\mu}{\sigma/\sqrt{N}}$ έχει κατά προσέγιση τυπική κανονική κατανομή $N(0,1)$.

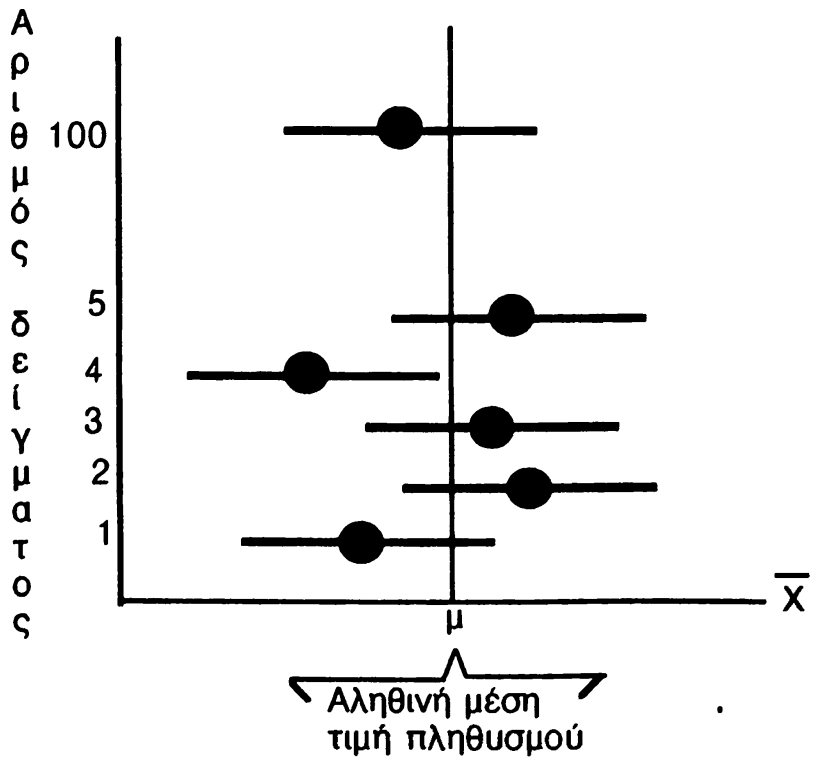
Μέσω το θεωρήματος αυτού είναι πλέον προφανές ότι τα διαστήματα εμπιστοσύνης που κατασκευάστηκαν προηγούμενα για τη μέση τιμή κανονικού πληθυσμού και την διαφορά των μέσων τιμών δυο ανεξάρτητων κανονικών πληθυσμών, είναι και τα (προσεγγιστικά) διαστήματα εμπιστοσύνης για τις μέσες τιμές οποιονδήποτε μη κανονικών πληθυσμών.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

6.3.4 Γενικές Παρατηρήσεις

1) Έχει τονιστεί προηγούμενα ότι τα διαστήματα εμπιστοσύνης αποτελούν ένα αντικειμενικό τρόπο αξιολόγισης του εκτιμητή $\bar{X}$ της μέσης τιμής μ ενός πληθυσμού. Η αλήθεια της πρότασης αυτής φαίνεται από το εξής παράδειγμα: Έστω ότι εκλέγονται δυο τυχαία δείγματα, ίδιου μεγέθους από τον ίδιο πληθυσμό. Αν 50 και 51 είναι οι μέσες τιμές των δυο δειγμάτων και 50 ± 8.5, 51 ± 3.5 είναι τα 95% διαστήματα εμπιστοσύνης για τη μέση τιμή μ του πληθυσμού τότε είναι φανερό ότι η αξιοπιστία του εκτιμητή 51 της μέσης τιμής μ είναι μεγαλύτερη από την αξιοπιστία του άλλου εκτιμητή.

2) Μια ερμηνεία των διαστημάτων εμπιστοσύνης για τη μέση τιμή ενός πληθυσμού είναι η ακόλουθη. Έστω ότι εκλέγεται ένας μεγάλος αριθμός τυχαίων δειγμάτων ίδιου μεγέθους από ένα πληθυσμό



ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

και ότι για κάθε δείγμα υπολογίζεται το 95% διάστημα εμπιστοσύνης για τη μέση τιμή του πληθυσμού. Οι μέσες τιμές και τα διαστήματα εμπιστοσύνης για το κάθε δείγμα φαίνονται στο προηγούμενο σχήμα.

Τότε περίπου 95% από τα διαστήματα αυτά θα περιέχουν τη μέση τιμή του πληθυσμού, ενώ το υπόλοιπο 5% των διαστημάτων δεν θα την περιέχουν. Παρά το γεγονός ότι το 95% των διαστημάτων περιέχει την άγνωστη μέση τιμή μ, δεν γνωρίζουμε αν αυτή η μ περιέχεται σ' ένα συγκεκριμένο διάστημα. Μόλις ένα συγκεκριμένο διάστημα υπολογιστεί η αληθινή τιμή του μ είτε περιέχεται στο διάστημα αυτό είτε όχι, ενώ για κάθε διάστημα που υπολογίζεται η πιθανότητα να περιλαμβάνει το μ είναι 0 ή 1.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΚΕΦΑΛΑΙΟ 7

ΕΛΕΓΧΟΣ ΣΤΑΤΙΣΤΙΚΩΝ ΥΠΟΘΕΣΕΩΝ

7.1. Εισαγωγικά

Στο προηγούμενο κεφάλαιο παρουσιάστηκε ο ένας κλαδος στατιστικής συμπερασματολογίας, η εκτιμητική, που περιλαμβάνει την εκτίμηση σε σημείο και την εκτίμηση σε διάστημα της άγνωστης παραμέτρου ενός πληθυσμού. Αντικείμενο του κεφαλαίου αυτού είναι ο άλλος κλάδος στατιστικής συμπερασματολογίας, ο έλεγχος στατιστικών υποθέσεων. Στις δραστηριότητές του εντάσσεται: η διατύπωση της στατιστικής υπόθεσης για την άγνωστη παράμετρο του πληθυσμού, η ανάπτυξη μεθοδολογιών - που στηρίζονται σ' ένα τυχαίο δείγμα από τον υπό μελέτη πληθυσμό - για τον έλεγχο της υπόθεσης και τέλος η εξαγωγή συμπερασμάτων για την ορθότητα ή μη της διατυπωθείσης υπόθεσης. Η διαδικασία λήψης απόφασης για την αποδοχή ή μη της υπό έλεγχο υπόθεσης λέγεται στατιστικό τεστ. Ειδικότερα στο κεφάλαιο αυτό περιγράφονται οι γενικές έννοιες και αρχές του ελέγχου στατιστικών υποθέσεων και στη συνέχεια δίνονται τα στατιστικά τεστ για τον έλεγχο (i) της μέσης τιμής κανονικού πληθυσμού, (ii) της διαφοράς των μέσων τιμών δύο ανεξάρτητων

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

κανονικών πληθυσμών και (iii) το τεστ συγκρίσεως ζευγών για τη διαφορά των μέσων τιμών δύο κανονικών πληθυσμών για τους οποίους η υπόθεση της ανεξαρτησίας δεν μπορεί να υιοθετηθεί.

7.2 Το στατιστικό τεστ - Οι πρώτες έννοιες

Οι βασικές έννοιες ενός στατιστικού τεστ περιγράφονται μέσω του ακόλουθου παραδείγματος:

Παράδειγμα: Από έρευνες που έχουν γίνει έχει διαπιστωθεί ότι ο μέσος δείκτης νοημοσύνης του πληθυσμού των παιδιών που αποφοιτούν από το Δημοτικό Σχολείο είναι 90 μονάδες. Για να προσδιοριστεί αν η ειδική εκπαίδευση αυξάνει ή όχι το δείκτη νοημοσύνης, ορισμένα Δημοτικά Σχολεία της χώρας εκλέγονται στην τύχη και οι μαθητές της τελευταίας τάξης των σχολείων αυτών παρακολουθούν ένα ειδικό εκπαιδευτικό πρόγραμμα που φιλοδοξεί να βελτιώσει το δείκτη νοημοσύνης. Μετά από μια σχολική χρονιά εφαρμογής του προγράμματος αυτού γεννιέται το ερώτημα αν πραγματικά το πρόγραμμα απέδωσε καρπούς, αν πραγματικά οι απόφοιτοι των Δημοτικών Σχολείων στα οποία εφαρμόστηκε το ειδικό εκπαιδευτικό πρόγραμμα έχουν αυξημένο δείκτη νοημοσύνης.

Είναι ήδη γνωστό ότι ο πληθυσμός των αποφοίτων των Δημοτικών Σχολείων έχει, κατά μέσο όρο, δείκτη νοημοσύνης ίσο με 90 μονάδες. Για τον έλεγχο της αποτελεσματικότητας του νέου εκπαιδευτικού προγράμματος είναι λογικό πλέον το ενδιαφέρον να επικεντρωθεί στον πληθυσμό των αποφοίτων που παρακολούθησαν το πρόγραμμα αυτό και ειδικότερα στο μέσο δείκτη νοημοσύνης, ας συμβολιστεί με μ του πληθυσμού αυτού. Επειδή το νέο πρόγραμμα φιλοδοξεί να έχει θετικές επιπτώσεις στο δείκτη νοημοσύνης των αποφοίτων, η αναποτελεσματικότητά του θα εκφράζεται από την

ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ ΒΙΒΛΙΟΘΗΚΗ

ανισότητα $\mu \leq 90$ που σημαίνει ότι ο μέσος δείκτης νοημοσύνης των αποφοίτων που παρακολούθησαν την ειδική εκπαίδευση δεν αυξήθηκε. Η αποτελεσματικότητα του νέου προγράμματος εκφράζεται από την ανισότητα $\mu > 90$. Έτσι το πρόβλημα ελέγχου της αποτελεσματικότητας του εκπαιδευτικού προγράμματος ανάγεται σε πρόβλημα ελέγχου της ορθότητας των σχέσεων: $\mu \leq 90$ έναντι της $\mu > 90$ ή απλοποιώντας περισσότερο τις σχέσεις, σε πρόβλημα ελέγχου της $\mu = 90$ έναντι της $\mu > 90$.

Σχέσεις όπως οι προηγούμενες, που με τη χρησιμοποίηση των παραμέτρων ενός ή περισσότερων πληθυσμών διατυπώνεται οποιαδήποτε πρόταση, άποψη ή εικασία για τη συμπεριφορά, την κατάσταση των πληθυσμών αυτών λέγονται **στατιστικές υποθέσεις** (statistical hypothesis). Υπόθεση η οποία ορίζεται με την ελπίδα να απορριφθεί, να μηδενιστεί ονομάζεται **μηδενική υπόθεση** (null hypothesis) και συμβολίζεται με H_o. Η αντίθετή της ονομάζεται **εναλλακτική υπόθεση** (alternative hypothesis) και συμβολίζεται με H_a. Έτσι η αποτελεσματικότητα του ειδικού εκπαιδευτικού προγράμματος του προηγούμενου παραδείγματος, ανάγεται σε πρόβλημα ελέγχου της μηδενικής υπόθεσης H_o: $\mu = 90$ έναντι της εναλλακτικής υπόθεσης, H_a: $\mu > 90$. Υποθέτουμε επί πλέον εκ των προτέρων ότι δύο είναι οι δυνατές περιπτώσεις: είτε αληθεύει η H_o είτε αληθεύει η H_a και ότι η εκλογή μας είναι μεταξύ αυτών των υποθέσεων. Ο έλεγχος στατιστικών υποθέσεων δίνει τον τρόπο να αποφασίσουμε για την ορθότητα ή μη της H_o. Η όλη διαδικασία για το σκοπό αυτό λέγεται **στατιστικό τεστ.**

Η αλήθεια μιας στατιστικής υπόθεσης ποτέ δεν είναι γνωστή, εκτός αν ληφθούν και μελετηθούν όλα τα μέλη του πληθυσμού γεγονός που κατά κανόνα είναι αδύνατο ή έχει τεράστιο κόστος. Έτσι εκλέγεται συνήθως ένα τυχαίο δείγμα από τον υπό μελέτη πληθυσμό και με τη χρησιμοποίηση του δείγματος αυτού ελέγχεται αν η

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

εναλλακτική υπόθεση H_a είναι πιο παραδεκτή από την μηδενική υπόθεση H_o. Αν επιστρέψουμε στο παράδειγμα του δείκτη νοημοσύνης για τον έλεγχο της $H_o : \mu = 90$ εκλέγονται στην τύχη N παιδιά από τον πληθυσμό των αποφοίτων που παρακολούθησαν το ειδικό πρόγραμμα, εξετάζονται με ένα βασικό τυποποιημένο τεστ νοημοσύνης και η επίδοση του κάθε παιδιού καταγράφεται. Έστω X_1, X_2, ... , X_N είναι το τυχαίο δείγμα που προκύπτει από την πιο πάνω διαδικασία. Υπενθυμίζεται ότι το X_i, i=1,2,...,N είναι η τυχαία μεταβλητή που παριστά και περιγράφει το δείκτη νοημοσύνης του i-παιδιού και ταυτόχρονα ένας αριθμός που ποσοτικά εκφράζει το δείκτη του παιδιού αυτού. Αν από τη χρησιμοποίηση του δείγματος X_1, X_2, ... ,X_N υπάρξει μαρτυρία ασυμφωνίας προς την υπόθεση $H_o : \mu=90$, οδηγούμαστε στην απόρριψή της. Διαφορετικά αν υπάρξει μαρτυρία που να την ενισχύει οδηγούμαστε στην αποδοχή της. Αποδοχή της μηδενικής υπόθεσης H_o δεν σημαίνει ότι είναι και σωστή. Σημαίνει ότι από το συγκεκριμένο τυχαίο δείγμα δεν προκύπτουν επαρκείς ενδείξεις για την απόρριψή της.

Η μέχρι τώρα επαφή με το περιεχόμενο και τις μεθοδολογίες της στατιστικής συμπερασματολογίας υποδεικνύει τη χρησιμοποίηση κάποιας συνάρτησης π.χ. $\bar{X}$, S^2 του τυχαίου δείγματος και όχι το τυχαίο δείγμα αυτό καθ' εαυτό. Επίσης βασικό ρόλο στην εκτίμηση σε διάστημα και στην κατασκευή των διαστημάτων εμπιστοσύνης, παίζει η γνώση της κατανομής της στατιστικής συνάρτησης που πρόκειται να χρησιμοποιηθεί. Η αναγκαιότητα γνώσης της κατανομής αυτής επιβάλλει την υπόθεση ότι ο υπό μελέτη πληθυσμός περιγράφεται από την κανονική κατανομή ή ισοδύναμα ότι το τυχαίο δείγμα προέρχεται από την κανονική κατανομή. Η υπόθεση της κανονικότητας υιοθετείται και στον έλεγχο στατιστικών υποθέσεων για τη μέση τιμή του

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

πληθυσμού. Όπως θα αναφερθεί στη συνέχεια, στην πράξη η υπόθεση αυτή λόγω του κεντρικού οριακού θεωρήματος δεν είναι περιοριστική.

Επιστρέφουμε και πάλι στο παράδειγμα του δείκτη νοημοσύνης υποθέτοντας ότι το τυχαίο δείγμα $X_1, X_2, \ldots, X_N$ που συνίστατται από το δείκτη νοημοσύνης των N αποφοίτων προέρχεται από την κανονική κατανομή με γνωστή έστω διακύμανση σ^2. Οι υπό έλεγχο υποθέσεις $H_o : \mu = 90$ έναντι της $H_a : \mu > 90$ αφορούν έτσι έλεγχο της μέσης τιμής μ κανονικής κατανομής $N(\mu, \sigma^2)$. Είναι λογικό να χρησιμοποιηθεί για τον έλεγχο αυτό το στατιστικό $\bar{X}$ το οποίο, όπως αναφέρθηκε στην παράγραφο 6.3.1.1, έχει κανονική κατανομή $N(\mu,\sigma^2/ N)$ ή ισοδύναμα χρησιμοποιώντας τον τυπικό μετασχηματισμό το στατιστικό $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{N}}$ με κατανομή την N(0,1). Ειδικότερα επειδή ενδιαφέρει ο έλεγχος της $H_o : \mu = 90$ χρησιμοποιείται η στατιστική συνάρτηση $Z = \frac{\bar{X}-90}{\sigma/\sqrt{N}}$ που η κατανομή της είναι η N(0,1) όταν η μηδενική υπόθεση H_o είναι αληθινή. Αν η τιμή της στατιστικής συνάρτησης Z είναι θετική και μεγάλη τότε το $\bar{X}$ θα είναι μεγαλύτερο του 90. Το $\bar{X}$ όμως αποτελεί έναν εκτιμητή, μια προσέγγιση της μέσης τιμής μ του πληθυσμού των αποφοίτων που παρακολούθησαν το ειδικό εκπαιδευτικό πρόγραμμα. Συνεπώς όταν το $\bar{X}$ είναι αρκετά μεγαλύτερο του 90, η μέση τιμή μ του πληθυσμού θα είναι και αυτή αρκετά μεγαλύτερη του 90 και η απόρριψη της H_o είναι πλέον εύλογη. Ο προηγούμενος συλλογισμός υποδεικνύει τη χρησιμοποίηση των τιμών της στατιστικής συνάρτησης Z σαν ένα κριτήριο για την αποδοχή ή την απόρριψη της H_o.

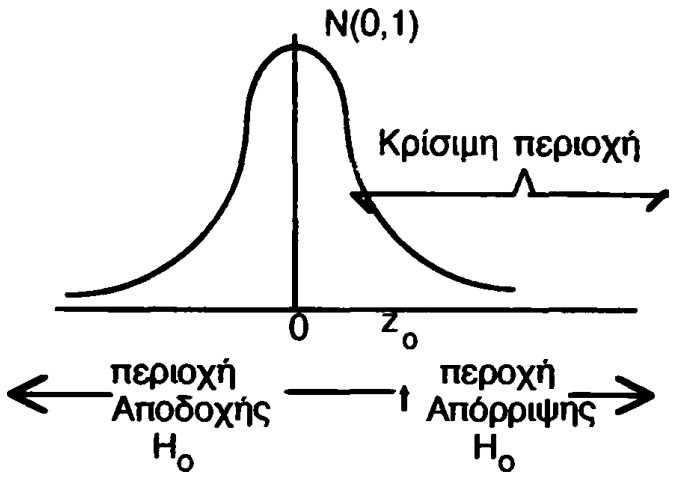
Γενικεύοντας τον προηγούμενο συλλογισμό το σύνολο τιμών της στατιστικής συνάρτησης Z διαιρείται σε δυο περιοχές. Η μια

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

178

ονομάζεται περιοχή απόρριψης ή **κρίσιμη περιοχή** (critical region) και η άλλη περιοχή αποδοχής της H_0. Οι τιμές του στατιστικού Z που περιλαμβάνονται στην κρίσιμη περιοχή είναι σχεδόν αδύνατο να συμβούν αν η H_0 είναι αληθινή, ενώ είναι πάρα πολύ πιθανό να συμβούν αν η H_0 δεν είναι αληθινή. Έτσι αν η τιμή του στατιστικού Z που υπολογίζεται για το συγκεκριμένο τυχαίο δείγμα περιλαμβάνεται στην κρίσιμη περιοχή, η μηδενική υπόθεση απορρίπτεται. Αν η τιμή του Z δεν περιέχεται στην κρίσιμη περιοχή η H_0 γίνεται αποδεκτή.

Στο σχήμα που ακολουθεί δίνονται οι περιοχές αποδοχής και απόρριψης της υπόθεσης $H_0 : \mu = 90$ για το παράδειγμα του δείκτη νοημοσύνης.



$$Z = \frac{\bar{X}-90}{\sigma/\sqrt{N}}$$

Στον οριζόντιο άξονα τοποθετούνται οι τιμές της στατιστικής συνάρτησης Z με κατανομή N(0,1) όταν αληθεύει η H_0 . Η κρίσιμη περιοχή αποτελείται από μεγάλες τιμές του Z, έστω μεγαλύτερες από το σημείο z_0. Η απόρριψη ή αποδοχή της H_0 εξαρτάται από την τιμή της στατιστικής συνάρτησης Z. Αν η τιμή της Z είναι μεγαλύτερη από το z_0 η H_0 απορρίπτεται, διαφορετικά γίνεται αποδεκτή.

Η προηγούμενη παρουσίαση σκιαγραφεί και τη μεθοδολογία κατασκευής στατιστικών τεστ, που επιγραμματικά συνίστανται:

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

i) Στην εύρεση μιας στατιστικής συνάρτησης με γνωστή κατανομή όταν αληθεύει η μηδενική υπόθεση H_o. Η συνάρτηση αυτή λέγεται **στατιστική συνάρτηση του τεστ**.

ii) Στην κατασκευή της κρίσιμης περιοχής, δηλαδή ενός υποσυνόλου τιμών της στατιστικής συνάρτησης του τεστ.

Για το παράδειγμα του δείκτη νοημοσύνης η κρίσιμη περιοχή αποτελείται από μεγάλες τιμές της στατιστικής συνάρτησης $Z = \frac{\bar{X}-90}{\sigma/\sqrt{N}}$ του τεστ και όπως φαίνεται στο προηγούμενο σχήμα είναι ένα διάστημα της μορφής (z_o, ∞). Είναι φανερό με τη μέχρι τώρα ανάλυση ότι το z_o μπορεί να είναι οποιοσδήποτε θετικός αριθμός. Μ' αυτόν τον τρόπο όμως δημιουργούνται πάρα πολλά διαστήματα της μορφής (z_o, ∞) και κατά συνέπεια πάρα πολλές κρίσιμες περιοχές. Το ερώτημα που γεννιέται πλέον είναι πια από αυτές τις κρίσιμες περιοχές θα χρησιμοποιηθεί. Η απάντηση στο ερώτημα αυτό βασίζεται σε κριτήρια αξιολόγησης των τεστ και των αντίστοιχων κρίσιμων περιοχών. Τα κριτήρια αυτά αναπτύσσονται στη συνέχεια.

7.3 Αξιολόγηση των στατιστικών τεστ - Σφάλματα τύπου I και II

Ο έλεγχος στατιστικών υποθέσεων και η στατιστική συμπερασματολογία γενικότερα, μπορεί να θεωρηθεί ως ένα παιχνίδι με παίκτες από την μια μεριά τον στατιστικό και από την άλλη μεριά τον πληθυσμό, την φύση με τα άγνωστα χαρακτηριστικά της. Ο στατιστικός αναπτύσσει τις μεθοδολογίες για να προσεγγίσει τα άγνωστα χαρακτηριστικά του πληθυσμού, την πραγματικότητα. Με βάση τις μεθοδολογίες αυτές καταλήγει σε συμπεράσματα και

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

αποφάσεις χωρίς να είναι ποτέ σίγουρος για το ακλόνητο των συμπερασμάτων και την ορθότητα των αποφάσεων.

Έτσι ο στατιστικός αποφασίζει και για την αποδοχή ή την απόρριψη της H_o χωρίς κατ' ανάγκη η απόφαση να είναι και η σωστή. Οποιαδήποτε και αν είναι η απόφαση για την H_o δύο είναι τα δυνατά σφάλματα στα οποία ο στατιστικός μπορεί να οδηγηθεί:

Σφάλμα Τύπου I: Απόρριψη της H_o όταν η H_o είναι αληθής.
Σφάλμα Τύπου II: Αποδοχή της H_o όταν η H_a είναι αληθής.

Ο κίνδυνος του σφάλματος τύπου I υπάρχει όταν αποφασισθεί απόρριψη της μηδενικής υπόθεσης H_o, ενώ ο κίνδυνος διάπραξης σφάλματος τύπου II υπάρχει όταν αποφασισθεί αποδοχή της υπόθεσης H_o. Επειδή ο κίνδυνος διάπραξης ενός από τα παραπάνω σφάλματα είναι υπαρκτός, το ενδιαφέρον πλέον εντοπίζεται στην ελαχιστοποίηση αυτού του κινδύνου, ή στην ελαχιστοποίηση της πιθανότητας πραγματοποίησης των σφαλμάτων τύπου I και II. Αν συμβολιστούν με α και β οι πιθανότητες αυτές τότε:

$$\alpha = P(\text{Σφάλμα Τύπου I}) = P(\text{Απόρριψη } H_o \mid H_o \text{ αληθής})$$

και

$$\beta = P(\text{Σφάλμα Τύπου II}) = P(\text{Αποδοχή } H_o \mid H_a \text{ αληθής}).$$

Όσο μικρότερες είναι οι πιθανότητες α και β τόσο περισσότερο ακλόνητη είναι η απόφαση για την αποδοχή ή την απόρριψη της H_o. Έτσι γίνεται πλέον φανερό ότι οι πιθανότητες α και β των σφαλμάτων τύπου I και II αντίστοιχα αποτελούν μέτρα αξιολόγησης στατιστικών τεστ και των αντίστοιχων κρίσιμων περιοχών. Η πιθανότητα α του σφάλματος τύπου I λέγεται και μέγεθος της κρίσιμης περιοχής ή **επίπεδο σημαντικότητας** του τεστ ενώ η

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

διαφορά 1-β συμβολίζεται συνήθως με γ και ονομάζεται **ισχύς του τεστ.**

Αν επιστρέψουμε στο παράδειγμα του δείκτη νοημοσύνης, από όλες τις κρίσιμες περιοχές της μορφής (z_o, ∞) προτιμότερη θα είναι εκείνη για την οποία οι πιθανότητες των σφαλμάτων τύπου I και II είναι ελάχιστες. Μπορεί όμως (εύκολα) να αποδεχτεί ότι είναι αδύνατο να επιτευχθεί η ταυτόχρονη ελαχιστοποίηση των α και β. Το πρόβλημα αυτό παρακάμπτεται κρατώντας το επίπεδο σημαντικότητας α σταθερό, καθορισμένο εκ των προτέρων και επιδιώκοντας να βρούμε εκείνη την κρίσιμη περιοχή (z_o, ∞) για την οποία το β γίνεται ελάχιστο ή ισοδύναμα η ισχύς γ γίνεται μέγιστη. Στο προκαθορισμένο επίπεδο σημαντικότητας α δίνονται συνήθως οι τιμές 1%, 5% ή 10% και το αντίστοιχο τέστ λέγεται τεστ επιπέδου ή μεγέθους α.

Για το παράδειγμα του δείκτη νοημοσύνης:

$$\alpha = P(\text{Απόρριψη } H_o \mid H_o \text{ αληθινή}) = P(Z \geq z_o),$$

με $Z = \dfrac{\bar{X}-90}{\sigma/\sqrt{N}}$ την στατιστική συνάρτηση του τεστ και z_o ένα θετικό αριθμό. Από τη σχέση αυτή και τον ορισμό των άνω αντίστροφων εκατοστιαίων σημείων της τυπικής κανονικής κατανομής (δες παράγραφο 6.3.1.1) προκύπτει ότι $z_o = z_\alpha$. Επομένως η καλύτερη κρίσιμη περιοχή για τον έλεγχο του δείκτη νοημοσύνης είναι η (z_α, ∞) . Γι' αυτή την κρίσιμη περιοχή είναι δυνατόν θεωρητικά να αποδειχθεί ότι το β είναι το ελάχιστο δυνατό.

Μετά την παρουσίαση της φιλοσοφίας και των λογικών αρχών των στατιστικών τεστ δίνονται στη συνέχεια στατιστικά τεστ για τον έλεγχο της μέσης τιμής κανονικού πληθυσμού και τον έλεγχο της διαφοράς των μέσων τιμών δύο κανονικών πληθυσμών.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

7.4 Στατιστικά τεστ για τη μέση τιμή κανονικής κατανομής

Ο έλεγχος υποθέσεων γύρω από τη μέση τιμή κανονικού πληθυσμού είναι ο πλέον αντιπροσωπευτικός και χρησιμοποιούμενος στη στατιστική. Στη συνέχεια διατυπώνονται στατιστικά τεστ για τον έλεγχο της μέσης τιμής μ κανονικού πληθυσμού θεωρώντας τη διακύμανση σ^2 του πληθυσμού πρώτα γνωστή και στη συνέχεια άγνωστη.

7.4.1 Στατιστικά τεστ για τη μ όταν η σ^2 είναι γνωστή

Θεωρούμε ένα τυχαίο δείγμα $X_1, X_2, \ldots, X_N$ μεγέθους N από ένα πληθυσμό με κανονική κατανομή $N(\mu, \sigma^2)$ με τη διακύμανση σ^2 γνωστή. Για τον έλεγχο με επίπεδο σημαντικότητας α της μηδενικής υπόθεσης,

$$H_o : \mu = \mu_o$$

με μ_o μια σταθερά, ως προς τις εναλλακτικές υποθέσεις

$$H_a : \mu > \mu_o \quad , \quad H_a : \mu < \mu_o \quad , \quad H_a : \mu \neq \mu_o$$

οι κρίσιμες περιοχές μεγέθους α είναι αντίστοιχα:

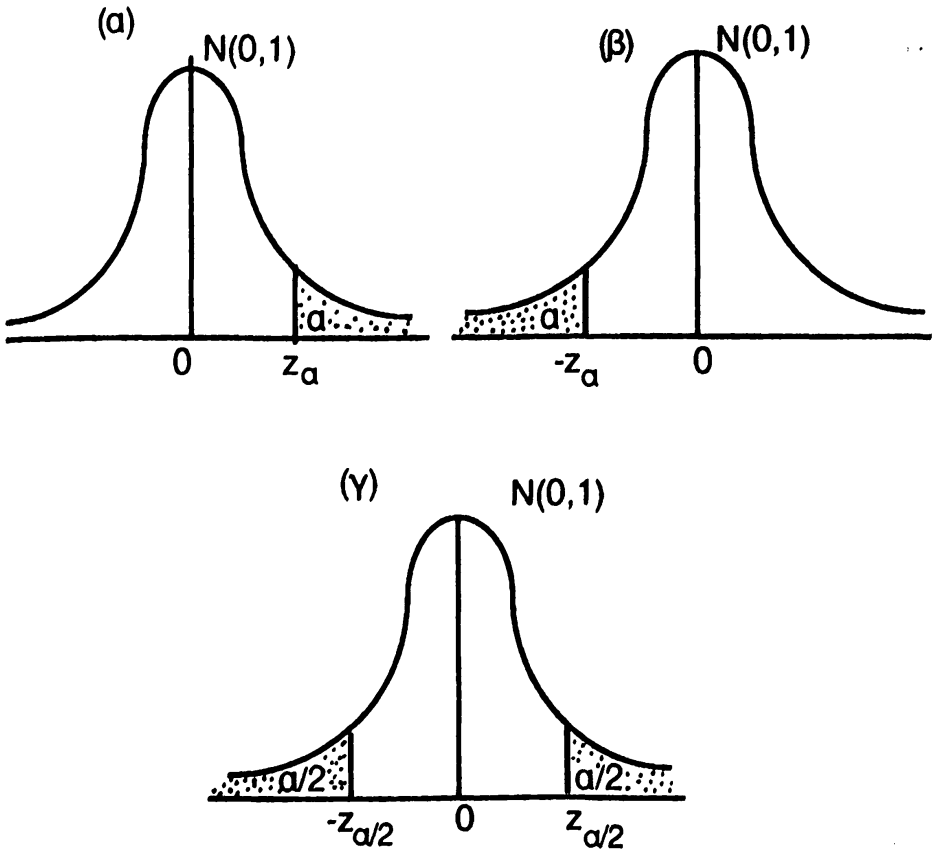
$$Z \geq z_\alpha \quad , \quad Z \leq -z_\alpha \quad , \quad |Z| \geq z_{\alpha/2} \; (Z \geq z_{\alpha/2} \text{ ή } Z \leq -z_{\alpha/2})$$

όπου $Z = \dfrac{\bar{X}-\mu_o}{\sigma/\sqrt{N}}$ είναι η στατιστική συνάρτηση του τεστ με κατανομή την $N(0,1)$ όταν ισχύει η μηδενική υπόθεση. Τα z_α και $z_{\alpha/2}$ είναι αντίστοιχα τα άνω α και $\frac{\alpha}{2}$ αντίστροφα εκατοστιαία σημεία της

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

N(0,1) κατανομής. Οι γραφικές παραστάσεις των παραπάνω κρίσιμων περιοχών δίνονται στο Σχήμα 7.1.

Επίσης η κρίσιμη περιοχή $Z \geq z_\alpha$ χρησιμοποιείται για τον έλεγχο της $H_o: \mu \leq \mu_o$ έναντι της $H_a: \mu > \mu_o$ ενώ η κρίσιμη περιοχή $Z \leq -z_\alpha$ χρησιμοποιείται και για τον έλεγχο της $H_o: \mu \geq \mu_o$ έναντι της $H_a: \mu < \mu_o$. Έτσι για οποιοδήποτε έλεγχο της μέσης τιμής κανονικής κατανομής προσδιορίζεται πρώτα με τη βοήθεια των δεδομένων η τιμή της στατιστικής συνάρτησης του τεστ. Η μορφή της εναλλακτικής υπόθεσης υποδεικνύει την κρίσιμη περιοχή που θα χρησιμοποιηθεί. Αν η τιμή της στατιστικής συνάρτησης του τεστ περιλαμβάνεται στην κρίσιμη περιοχή η μηδενική υπόθεση H_o απορρίπτεται, διαφορετικά γίνεται αποδεκτή.



Σχήμα 7.1: Κρίσιμες περιοχές του Z-τεστ

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Οι προηγούμενες κρίσιμες περιοχές είναι οι καλύτερες. Αυτό αποδεικνύεται με μεθοδολογίες της θεωρίας ελέγχου στατιστικών υποθέσεων, που ξεφεύγουν από το σκοπό των σημειώσεων αυτών. Μπορεί όμως να φανεί και διαισθητικά με συλλογισμούς ανάλογους με εκείνους των παραγράφων 7.2 και 7.3 όπου ουσιαστικά κατασκευάστηκε το στατιστικό τεστ για τον έλεγχο του πρώτου ζεύγους υποθέσεων. Το προηγούμενο τεστ είναι γνωστό στη βιβλιογραφία και ως **Z-τεστ.** Το Z-τεστ λέγεται **μονόπλευρο** αν η κρίσιμη περιοχή του ορίζεται σε ένα από τα δυο άκρα (ουρές) της κατανομής όπως οι περιπτώσεις (α) και (β) του Σχήματος 7.1. Αν η κρίσιμη περιοχή ορίζεται και στα δύο άκρα (ουρές) της κατανομής (Σχήμα 7.1 (γ)) το Z-τεστ λέγεται **δίπλευρο**.

Παράδειγμα 7.4.1. Τα παρακάτω δεδομένα αναφέρονται στο δείκτη νοημοσύνης 10 παιδιών και είναι:

70, 67, 82, 75, 103, 121, 114, 100, 85, 99.

Αν ο δείκτης νοημοσύνης ακολουθεί κανονική κατανομή $N(\mu, \sigma=15)$ να ελεγθεί με επίπεδο σημαντικότητας $\alpha=0.05$ η υπόθεση ότι ο δείκτης νοημοσύνης είναι 89 μονάδες έναντι της υπόθεσης ότι είναι μεγαλύτερος των 89 μονάδων.

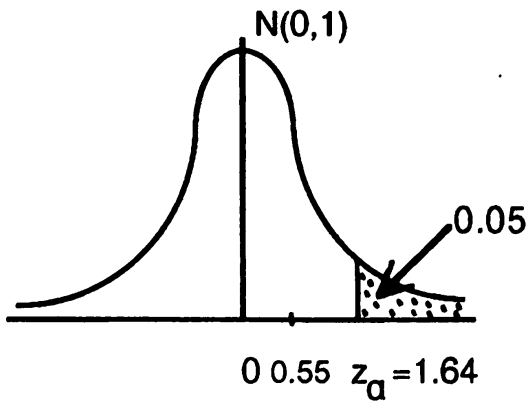
Ζητείται ο έλεγχος της :

$H_o : \mu = 89$ έναντι της $H_a : \mu > 89$.

Για τον έλεγχο αυτό το στατιστικό τεστ είναι το $Z = \dfrac{\bar{X} - 89}{\sigma/\sqrt{N}}$ με κρίσιμη περιοχή την $Z \geq z_\alpha$. Εύκολα υπολογίζεται ότι : $\bar{X} = \dfrac{1}{10} \sum_{i=1}^{10} X_i$ = 91.6.

Έτσι $Z = \frac{91.6-89}{15/\sqrt{10}} = 0.55$ ενώ από τους πίνακες της N(0,1) κατανομής υπολογίζεται ότι : $z_\alpha = z_{0.05} = 1.64$. Με τη βοήθεια και του παρακάτω σχήματος παρατηρούμε ότι : $Z = 0.55 < 1.64 = z_{0.05}$.

Η τιμή της στατιστικής συνάρτησης του τεστ δεν περιλαμβάνεται στη σκιασμένη κρίσιμη περιοχή και έτσι δεν μπορούμε να απορρίψουμε τη μηδενική υπόθεση. Επομένως το μ μπορεί να είναι ίσο με 89 μονάδες, δηλαδή ο μέσος δείκτης νοημοσύνης του πληθυσμού των παιδιών μπορεί να υιοθετηθεί ότι είναι ίσος με 89 μονάδες.



Παράδειγμα 7.4.2. Μια ειδική μέθοδος διδασκαλίας των μαθηματικών θεωρείται επιτυχής αν η βαθμολογία των μαθητών στις εξετάσεις είναι μεγαλύτερη του 5.5. Αν η βαθμολογία 9 μαθητών μιας τάξης ενός Δημοτικού σχολείου που διδάχτηκαν με την ειδική μέθοδο ήταν: 5, 6, 8, 9, 5, 10, 5, 6, 9 μπορούμε να συμπεράνουμε με επίπεδο α= 5% ότι η μέθοδος ήταν επιτυχής; Υποθέτουμε ότι τα δεδομένα προέρχονται από την κανονική κατανομή με μέση τιμή μ και διακύμανση $\sigma^2 = 4$.

Ζητείται ο έλεγχος της $H_o : \mu \leq 5.5$ έναντι της $H_a : \mu > 5.5$. Η στατιστική συνάρτηση του τεστ είναι:

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

$Z = \frac{\bar{X}-\mu_o}{\sigma/\sqrt{N}} = \frac{7-5.5}{2/3} = 2.25$ και η κρίσιμη περιοχή $Z \geq z_\alpha = z_{0.05} = 1.64$. Παρατηρούμε ότι $Z = 2.25 > 1.64 = z_{0.05}$ δηλαδή η τιμή της στατιστικής συνάρτησης του τεστ περιλαμβάνονται στην κρίσιμη περιοχή και επομένως απορρίπτουμε την H_o. Έτσι με πιθανότητα σφάλματος 5% μπορούμε να αποφανθούμε ότι η ειδική μέθοδος διδασκαλίας των μαθηματικών είναι επιτυχής.

7.4.2. Στατιστικά τεστ για το μ όταν η σ^2 είναι άγνωστη

Θεωρούμε και πάλι ένα τυχαίο δείγμα $X_1, X_2, \ldots, X_N$ μεγέθους N από πληθυσμό με κατανομή $N(\mu, \sigma^2)$ με τη διακύμανση σ^2 άγνωστη. Για τον έλεγχο, με επίπεδο σημαντικότητας α, της μηδενικής υπόθεσης:

$$H_o : \mu = \mu_o$$

με μ_o γνωστή σταθερά, ως προς τις εναλλακτικές υποθέσεις

$$H_a : \mu > \mu_o, \quad H_a : \mu < \mu_o, \quad H_a : \mu \neq \mu_o,$$

οι κρίσιμες περιοχές μεγέθους α είναι αντίστοιχα,

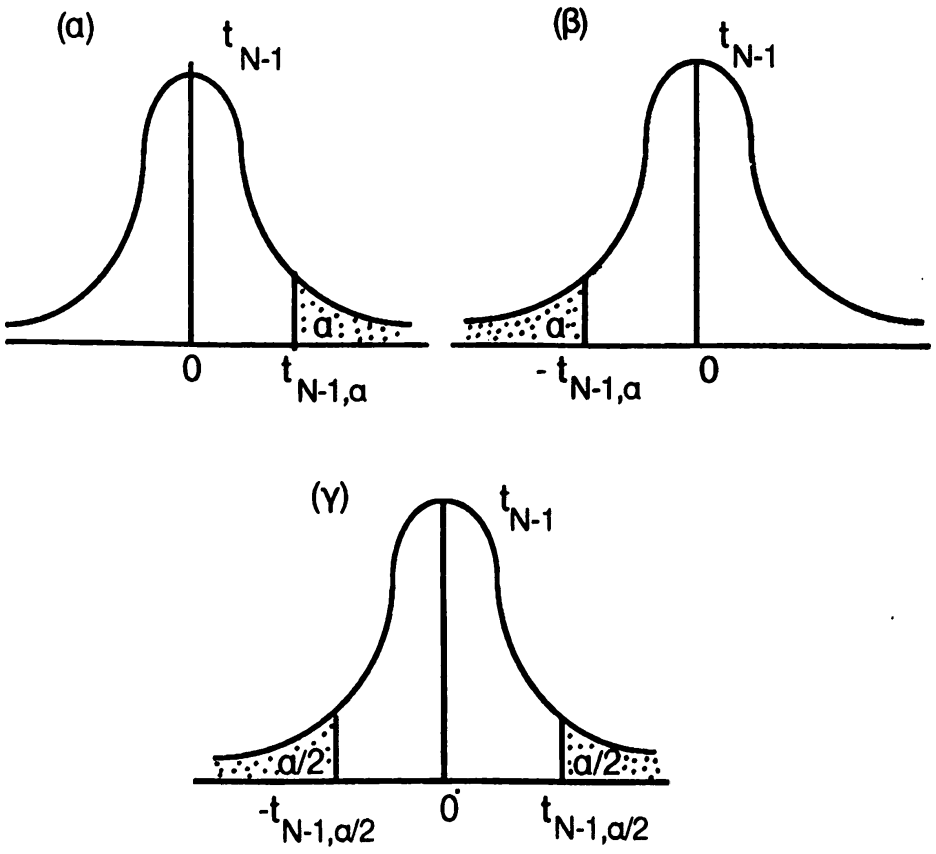
$$t \geq t_{N-1,\alpha}, \; t \leq -t_{N-1,\alpha}, \; |t| \geq t_{N-1,\alpha/2} \; (t \geq t_{N-1,\alpha/2} \text{ ή } t \leq -t_{N-1,\alpha/2}),$$

όπου $t = \frac{\bar{X}-\mu_o}{S/\sqrt{N}}$ η στατιστική συνάρτηση του τεστ με κατανομή την t_{N-1} όταν ισχύει η H_o. S είναι η δειγματική διακύμανση και $t_{N-1,\alpha}$, $t_{N-1,\alpha/2}$ τα α και α/2 αντίστοιχα άνω εκατοστιαία σημεία της

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

t_{N-1} κατανομής. Οι γραφικές παραστάσεις των κρίσιμων περιοχών δίνονται στο Σχήμα 7.2.

Το προηγούμενο τεστ είναι γνωστό ως **t-τεστ** ισχύουν δε και γι' αυτό ανάλογες παρατηρήσεις με εκείνες του Z-τεστ. Τονίζεται ότι για την περίπτωση που το μέγεθος του δείγματος N είναι μεγάλο, συνήθως $N \geq 30$, μπορεί να χρησιμοποιηθεί το Z-τεστ αν στην έκφραση της στατιστικής συνάρτησης του τεστ αυτού ή άγνωστη τυπική απόκλιση σ του πληθυσμού αντικατασταθεί από την αντίστοιχη δειγματική S. Επίσης η κρίσιμη περιοχή $t \geq t_{N-1,\alpha}$ χρησιμοποιείται για τον έλεγχο της $H_o : \mu \leq \mu_o$ έναντι της $H_a : \mu > \mu_o$ ενώ η κρίσιμη περιοχή $t \leq -t_{N-1,\alpha}$ για τον αντίθετο έλεγχο.



Σχήμα 7.2: Γραφικές παραστάσεις κρισίμων περιοχών t-τεστ

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

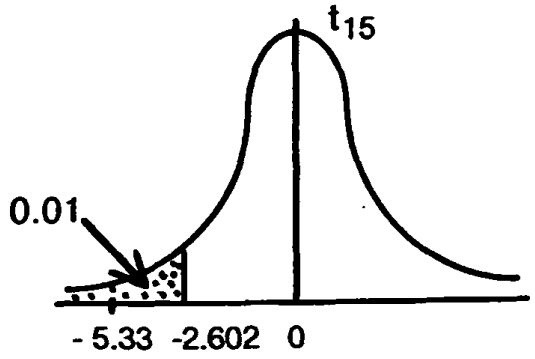
Παράδειγμα 7.4.3: Από προηγούμενες μελέτες είναι γνωστό ότι το μέσο βάρος του πληθυσμού των κοριτσιών ηλικίας δέκα ετών είναι 32 κιλά. Επειδή υπάρχει η υπόνοια ότι σε μια γεωγραφική περιοχή τα κορίτσια ηλικίας δέκα ετών χαρακτηρίζονται από μειωμένο βάρος, ένα δείγμα 16 κοριτσιών έδωσε μέσο βάρος 28 κιλά και τυπική απόκλιση 3 κιλά. Με βάση τα δεδομένα αυτά και επίπεδο σημαντικότητας $\alpha = 1\%$ έχουν τα κορίτσια της περιοχής αυτής μειωμένο βάρος; Υποθέτουμε ότι το βάρος των κοριτσιών ακολουθεί κανονική κατανομή.

Έστω μ είναι το μέσο βάρος του πληθυσμού των κοριτσιών. Έχει παρατηρηθεί ότι το μέσο βάρος είναι $\mu_o=32$. Το γεγονός ότι τα κορίτσια της δεδομένης γεωγραφικής περιοχής έχουν μειωμένο βάρος χαρακτηρίζεται από τη σχέση $\mu \leq 32$. Έτσι το πρόβλημα ανάγεται στον έλεγχο των υποθέσεων: $H_o : \mu \geq 32$ έναντι της $H_a : \mu < 32$. Το βάρος των κοριτσιών ακολουθεί κανονική κατανομή με άγνωστη διακύμανση και ένα τυχαίο δείγμα 16 κοριτσιών έδωσε $\bar{X} = 28$ κιλά και $S = 3$ κιλά. Το στατιστικό τεστ για τον έλεγχο των υποθέσεων είναι το:

$$t = \frac{\bar{X}-\mu_o}{S/\sqrt{N}} \text{ και η κρίσιμη περιοχή } t \leq -t_{N-1,\alpha}.$$

Έτσι $t = \frac{28-32}{3/4} = -5.33$ ενώ από τους πίνακες της t κατανομής το $t_{N-1,\alpha} = t_{15,\,0.01} = 2.602$.

Με τη βοήθεια και του σχήματος:



ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

παρατηρούμε ότι : $t = -5.33 < -2.602 = -t_{15,\ 0.01}$. Η τιμή της στατιστικής συνάρτησης t περιλαμβάνεται στη σκιασμένη κρίσιμη περιοχή και η μηδενική υπόθεση απορρίπτεται. Έτσι με πιθανότητα σφάλματος 1% τα δεδομένα οδηγούν στο συμπέρασμα ότι τα κορίτσια της περιοχής απ' την οποία προέρχεται το δείγμα χαρακτηρίζονται από μειωμένο βάρος.

7.5 Στατιστικά τεστ για τη διαφορά των μέσων τιμών δυο ανεξάρτητων κανονικών πληθυσμών

Ο έλεγχος της διαφοράς των μέσων τιμών προκύπτει συνήθως όταν υπάρχουν δύο διαφορετικοί πληθυσμοί και το ενδιαφέρον επικεντρώνεται στη διερεύνηση της συμπεριφοράς των πληθυσμών αυτών ως προς ένα κοινό τους χαρακτηριστικό. Για παράδειγμα θεωρούμε τους πληθυσμούς των υποψηφίων για τα Α.Ε.Ι. που κατάγονται και αποφοίτησαν από Λύκεια αγροτικών περιοχών και τους υποψήφιους που αποφοίτησαν από Λύκεια αστικών περιοχών. Κοινό χαρακτηριστικό και των δυο πληθυσμών είναι η επίδοση στην έκθεση. Έστω μ_1 είναι η μέση επίδοση στην έκθεση του πληθυσμού των αγροτικών περιοχών και μ_2 η μέση επίδοση στο ίδιο μάθημα του πληθυσμού των αστικών περιοχών. Οποιαδήποτε σύγκριση της επίδοσης των δύο πληθυσμών στην έκθεση εκφράζεται με τη βοήθεια των μέσων τιμών μ_1 και μ_2 των δυο πληθυσμών. Έτσι η σχέση $\mu_1=\mu_2$ σημαίνει ότι η επίδοση των δύο πληθυσμών στην έκθεση είναι η ίδια, η σχέση $\mu_1 > \mu_2$ την ανωτερότητα του αγροτικού πληθυσμού έναντι του αστικού όσον αφορά την επίδοση στην έκθεση. Τα στατιστικά τεστ που δίνονται στη συνέχεια παρέχουν την δυνατότητα ελέγχου υποθέσεων για τις μέσες τιμές δυο ανεξάρτητων πληθυσμών. Ο έλεγχος γίνεται

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

στη βάση τυχαίων δειγμάτων που εκλέγονται από τους πληθυσμούς αυτούς, ενώ η κατασκευή των στατιστικών τεστ στηρίζεται στην υπόθεση ότι τα δείγματα προέρχονται από κανονικές κατανομές με γνωστές διακυμάνσεις ή με άγνωστες αλλά ίσες μεταξύ τους διακυμάνσεις.

7.5.1 Στατιστικά τεστ για τη διαφορά των μέσων τιμών δυο ανεξάρτητων κανονικών πληθυσμών με γνωστές διακυμάνσεις

Θεωρούμε τα τυχαία δείγματα $X_{11}, X_{12}, \ldots, X_{1N_1}$ και $X_{21}, X_{22}, \ldots, X_{2N_2}$ που προέρχονται αντίστοιχα από τους ανεξάρτητους πληθυσμούς με κατανομές $N(\mu_1, \sigma_1^2)$ και $N(\mu_2, \sigma_2^2)$, με γνωστές διακυμάνσεις σ_1^2 και σ_2^2. Η κατασκευή στατιστικών τεστ για τον έλεγχο της διαφοράς των μέσων τιμών μ_1 και μ_2 των δυο πληθυσμών στηρίζεται στο γεγονός ότι η διαφορά $\bar{X}_1 - \bar{X}_2$ των μέσων τιμών των δυο δειγμάτων ακολουθεί κανονική κατανομή με μέση τιμή $\mu_1-\mu_2$ και διακύμανση $(\sigma_1^2/N_1) + (\sigma_2^2/N_2)$. Με βάση το αποτέλεσμα αυτό για τον έλεγχο, με επίπεδο σημαντικότητας α, της μηδενικής υπόθεσης:

$$H_o : \mu_1 = \mu_2$$

ως προς τις εναλλακτικές υποθέσεις.

$$H_a : \mu_1 > \mu_2, \quad H_a : \mu_1 < \mu_2, \quad H_a : \mu_1 \neq \mu_2,$$

οι κρίσιμες περιοχές μεγέθους α είναι αντίστοιχα

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

$$Z \geq z_\alpha \ , \ Z \leq -z_\alpha \ , \quad |Z| \geq z_{\alpha/2} \ (Z \geq z_{\alpha/2} \text{ ή } Z \leq -z_{\alpha/2})$$

όπου $Z = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{\sigma_1^2}{N_1} + \dfrac{\sigma_2^2}{N_2}}}$ είναι η στατιστική συνάρτηση του τεστ, με κατανομή N(0,1) όταν ισχύει η H_o.

Τα $\bar{X}_1$ και $\bar{X}_2$ είναι οι μέσες τιμές των δυο δειγμάτων και ορίζονται: $\bar{X}_1 = \dfrac{1}{N_1} \sum_{i=1}^{N_1} X_{1i}$ και $\bar{X}_2 = \dfrac{1}{N_2} \sum_{i=1}^{N_2} X_{2i}$. Οι γραφικές παραστάσεις των κρίσιμων περιοχών είναι οι ίδιες με εκείνες του Σχήματος 7.1. Οι προηγούμενες κρίσιμες περιοχές προκύπτουν διαισθητικά με συλλογισμούς ανάλογους με εκείνους των παραγράφων 7.2 και 7.3 και της γνώσης της κατανομής του $\bar{X}_1 - \bar{X}_2$, ενώ με μεθοδολογίες του ελέγχου στατιστικών υποθέσεων αποδεικνύεται ότι αυτές οι κρίσιμες περιοχές είναι οι καλύτερες. Επίσης η κρίσιμη περιοχή $Z \geq z_\alpha$ χρησιμοποιείται και για τον έλεγχο της $H_o: \mu_1 \leq \mu_2$ έναντι της $H_\alpha : \mu_1 > \mu_2$ ενώ η κρίσιμη περιοχή $Z \leq -z_\alpha$ χρησιμοποιείται και για τον έλεγχο της $H_o : \mu_1 \geq \mu_2$ έναντι της $H_\alpha : \mu_1 < \mu_2$. Η απόρριψη ή αποδοχή της H_o προκύπτει από τον υπολογισμό από τα δεδομένα της τιμής της στατιστικής συνάρτησης Z και της εξέτασης αν η τιμή αυτή περιλαμβάνεται ή όχι στην αντίστοιχη κρίσιμη περιοχή.

Παράδειγμα 7.5.1: Για να ελεγχθεί αν υπάρχει διαφορά στις επιδόσεις στην έκθεση υποψηφίων για Α.Ε.Ι. που αποφοίτησαν από Λύκεια αστικών και αγροτικών περιοχών εκτελείται το εξής πείραμα. Εκλέγονται στην τύχη 20 απόφοιτοι λυκείων αγροτικών περιοχών και οι βαθμοί τους στην έκθεση καταγράφονται.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Επίσης καταγράφονται και οι βαθμοί 15 αποφοίτων λυκείων αστικών περιοχών. Η μέση βαθμολογία των 20 αποφοίτων των αγροτικών περιοχών είναι 14 ενώ η μέση βαθμολογία των 15 αποφοίτων αστικών περιοχών είναι 16.5. Υποθέτουμε ότι ο βαθμός στην έκθεση των αποφοίτων και των δυο περιοχών ακολουθεί κανονική κατανομή με τυπική απόκλιση για τους απόφοιτους αγροτικών περιοχών ίση με 3.5 και τυπική απόκλιση για τους απόφοιτους αστικών περιοχών ίση με 2. Με βάση τα παραπάνω δεδομένα και με επίπεδο σημαντικότητας α=5% υπάρχει διαφορά στην επίδοση στην έκθεση των αποφοίτων από αγροτικές και αστικές περιοχές;

Τα δεδομένα συνοψίζονται στον πίνακα:

Απόφοιτοι Αγροτικών περιοχών	$N(\mu_1, \sigma_1 = 3.5)\ \ \bar{X}_1 = 14,\ N_1 = 20$
Απόφοιτοι Αστικών περιοχών	$N(\mu_2, \sigma_2 = 2)\ \ \bar{X}_2 = 16.5,\ N_2 = 15$

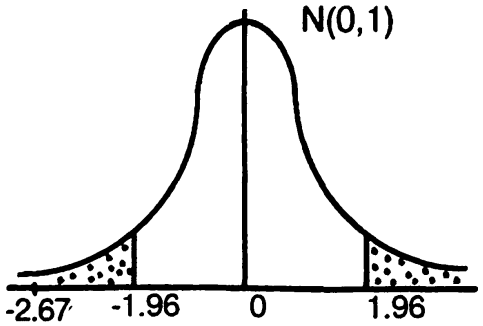
Ο ζητούμενος έλεγχος κωδικοποιείται σε έλεγχο της $H_o : \mu_1 = \mu_2$ έναντι της $H_a : \mu_1 \neq \mu_2$. Η στατιστική συνάρτηση του τεστ είναι:

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} = \frac{14-16.5}{\sqrt{\frac{12.25}{20} + \frac{4}{15}}} = -2.67.$$

Η κρίσιμη περιοχή είναι $|Z| \geq z_{\alpha/2}$. Απ' τους πίνακες της $N(0,1)$ κατανομής υπολογίζεται: $z_{\alpha/2} = z_{0.025} = 1.96$. Επίσης μέσω και του παρακάτω σχήματος παρατηρούμε ότι:

$$|Z| = |-2.67| = 2.67 > 1.96 = z_{0.025}\ .$$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ



Η τιμή της Z περιλαμβάνεται στη σκιασμένη κρίσιμη περιοχή και η H_o απορρίπτεται. Έτσι υπάρχουν ενδείξεις ότι η επίδοση στην έκθεση είναι διαφορετική.

7.5.2 Στατιστικά τεστ για τη διαφορά των μέσων τιμών δυο ανεξάρτητων κανονικών πληθυσμών με άγνωστες και ίσες διακυμάνσεις.

Θεωρούμε και πάλι δυο ανεξάρτητα τυχαία δείγματα $X_{11}, X_{12}, ..., X_{1N_1}$ και $X_{21}, X_{22}, ..., X_{2N_2}$ που προέρχονται από ανεξάρτητους κανονικούς πληθυσμούς $N(\mu_1, \sigma^2)$ και $N(\mu_2, \sigma^2)$ με μέσες τιμές μ_1 και μ_2 αντίστοιχα και κοινή αλλά άγνωστη διακύμανση σ^2. Αν $\bar{X}_1$, $\bar{X}_2$ είναι οι μέσες τιμές των δυο δειγμάτων και S_1^2, S_2^2 οι αντίστοιχες διακυμάνσεις τότε η

$$\frac{\bar{X}_1-\bar{X}_2 - (\mu_1-\mu_2)}{S\sqrt{\frac{1}{N_1}+\frac{1}{N_2}}} \quad \text{με} \quad S^2 = \frac{(N_1-1)S_1^2 + (N_2-1)S_2^2}{N_1+N_2-2},$$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

έχει κατανομή την t με N_1+N_2-2 βαθμούς ελευθερίας. Στηριζόμενοι στο αποτέλεσμα αυτό, για τον έλεγχο με επίπεδο σημαντικότητας α, της μηδενικής υπόθεσης

$$H_o : \mu_1 = \mu_2$$

ως προς τις εναλλακτικές υποθέσεις:

$$H_a : \mu_1 > \mu_2 , \qquad H_a : \mu_1 < \mu_2 , \qquad H_a : \mu_1 \neq \mu_2$$

οι κρίσιμες περιοχές μεγέθους α είναι αντίστοιχα

$$t \geq t_{N_1+N_2-2,\,\alpha} , \qquad t \leq - t_{N_1+N_2-2,\,\alpha} , \qquad |t| \geq t_{N_1+N_2-2,\,(\alpha/2)},$$

όπου $t = \dfrac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\dfrac{1}{N_1} + \dfrac{1}{N_2}}}$ είναι η στατιστική συνάρτηση του τεστ με κατανομή $t_{N_1+N_2-2}$, όταν ισχύει η H_o.

Τα $\bar{X}_1, \bar{X}_2$ και S^2 ορίζονται : $\bar{X}_1 = \dfrac{1}{N_1} \sum_{i=1}^{N_1} X_{1i}, \qquad \bar{X}_2 = \dfrac{1}{N_2} \sum_{i=1}^{N_2} X_{2i}$,

$S^2 = \dfrac{(N_1-1)S_1^2 + (N_2-1)S_2^2}{N_1+N_2-2}$ με $S_1^2 = \dfrac{1}{N_1-1} \sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2$ και

$S_2^2 = \dfrac{1}{N_2-1} \sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2$. Οι γραφικές παραστάσεις των κρίσιμων περιοχών είναι εκείνες του σχήματος 7.2 μόνο που οι βαθμοί ελευθερίας της t κατανομής γίνονται N_1+N_2-2 . Επίσης η κρίσιμη περιοχή $t \geq t_{N_1+N_2-2,\alpha}$ χρησιμοποιείται και για τον έλεγχο της

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

$H_o : \mu_1 \leq \mu_2$ έναντι της $H_a : \mu_1 > \mu_2$, ενώ η κρίσιμη περιοχή $t \leq -t_{N_1+N_2-2,a}$ χρησιμοποιείται και για τον έλεγχο της $H_o : \mu_1 \geq \mu_2$ έναντι της $H_a : \mu_1 < \mu_2$.

Παράδειγμα 7.5.2: Δυο ανεξάρτητες ομάδες παιδιών, η 1 και 2 αποτελούνται από 9 παιδιά η κάθε μια. Το ύψος των παιδιών και των δυο ομάδων μετριέται και βρίσκεται σε εκατοστά ότι είναι:

Ομάδα 1: 100, 102, 96, 106, 110, 120, 112, 112, 90
Ομάδα 2: 104, 98, 100, 90, 110, 116, 120, 115, 88.

Υποθέτουμε ότι το ύψος και των δυο ομάδων παιδιών ακολουθεί κανονική κατανομή με άγνωστες αλλά ίσες διακυμάνσεις. Με επίπεδο σημαντικότητας $\alpha = 10\%$ να ελεγχθεί αν υπάρχει διαφορά στο μέσο ύψος των πληθυσμών απ' τους οποίους προέρχονται οι δύο ομάδες παιδιών.

Αν $N(\mu_1, \sigma^2)$ και $N(\mu_2, \sigma^2)$ είναι οι κατανομές των ανεξάρτητων πληθυσμών απ' τους οποίους προέρχονται οι δυο ομάδες παιδιών ο ζητούμενος έλεγχος είναι ο έλεγχος της $H_o : \mu_1 = \mu_2$ έναντι της $H_a : \mu_1 \neq \mu_2$. Για τον έλεγχο των υποθέσεων αυτών χρησιμοποιείται η στατιστική συνάρτηση $t = \dfrac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\dfrac{1}{N_1} + \dfrac{1}{N_2}}}$ με κρίσιμη περιοχή $|t| \geq t_{N_1+N_2-2,\, \alpha/2}$. Από τα δεδομένα εύκολα υπολογίζεται

Ομάδα 1	$N_1 = 9,\ \bar{X}_1 = 105.3,\ S_1^2 = 86$
Ομάδα 2	$N_2 = 9,\ \bar{X}_2 = 104.6,\ S_2^2 = 132.3$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Έτσι

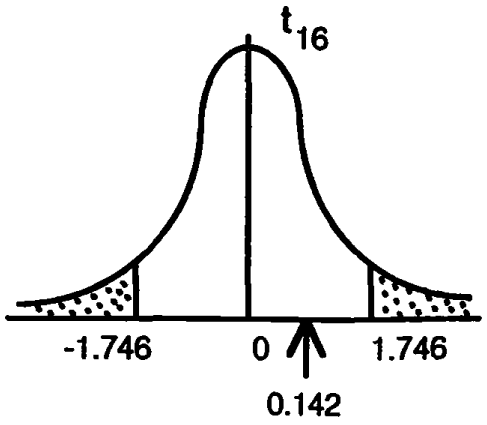
$$S = \sqrt{\frac{(N_1-1)\ S_1^2 + (N_2-1)S_2^2}{N_1+N_2-2}} = \sqrt{\frac{8\times86 + 8\times132.3}{9+9-2}} =$$

10.45.

Επομένως $t = \dfrac{105.3-104.6}{10.45\sqrt{\frac{1}{9} + \frac{1}{9}}} = 0.142$.

Παρατηρούμε ότι : $|t| = |0.142| = 0.142 < 1.746 = t_{16,0.05}$.

Η τιμή της t δεν περιλαμβάνεται στη σκιασμένη κρίσιμη περιοχή και η H_o γίνεται αποδεκτή. Έτσι υπάρχουν ενδείξεις σύμφωνα με τις οποίες δεν υπάρχει διαφορά στο μέσο ύψος των πληθυσμών απ' τους οποίους προέρχονται οι δυο ομάδες παιδιών.



Παρατήρηση 7.5: Οι υποθέσεις των παραγράφων 7.5.1 και 7.5.2 μπορούν να διατυπωθούν σε γενικότερη μορφή ως εξής:

$H_o : \mu_1 - \mu_2 = \delta$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ως προς τις εναλλακτικές υποθέσεις

$$H_a : \mu_1 - \mu_2 > \delta \,, \quad H_a : \mu_1 - \mu_2 < \delta \,, \qquad H_a : \mu_1 - \mu_2 \neq \delta$$

Με τον τρόπο αυτό ελέγχεται αν η διαφορετικότητα των πληθυσμών είναι ίση με μια δεδομένη ποσότητα δ ή όχι. Οι στατιστικές συναρτήσεις για τους ελέγχους αυτούς είναι:

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - \delta}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \quad \text{και} \quad t = \frac{\bar{X}_1 - \bar{X}_2 - \delta}{S\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

ανάλογα με την γνώση ή όχι των διακυμάνσεων των δυο πληθυσμών. Οι κρίσιμες περιοχές είναι οι ίδιες με εκείνες των παραγράφων 7.5.1, και 7.5.2.

7.6 Τεστ συγκρίσεως ζευγών

Στην προηγούμενη παράγραφο δόθηκαν στατιστικά τεστ για τον έλεγχο της διαφοράς των μέσων τιμών δυο ανεξάρτητων κανονικών πληθυσμών. Η εφαρμογή των τεστ αυτών προϋποθέτει την ανεξαρτησία των δυο πληθυσμών που πρακτικά σημαίνει ότι η συμπεριφορά οποιουδήποτε μέλους του ενός πληθυσμού δεν επηρεάζει τη συμπεριφορά άλλου μέλους του δεύτερου πληθυσμού. Παρά το γεγονός ότι η ανεξαρτησία μπορεί να υιοθετηθεί σε ευρύ φάσμα προβλημάτων υπάρχουν και περιπτώσεις που δεν είναι ρεαλιστική η υιοθέτησή της.

Πολλές φορές στην πράξη, ο έλεγχος μιας μεθόδου, μιας θεραπείας γίνεται με την εφαρμογή της σε N τυχαία επιλεγμένα άτομα ή οντότητες, στο εξής θα ονομάζονται πειραματικές μονάδες,

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

και στη συνέχεια συγκρίνονται τα αποτελέσματα προ και μετά την εφαρμογή της μεθόδου, για να ελεγχθεί η αποτελεσματικότητά της. Ας θεωρήσουμε για παράδειγμα μια νέα μέθοδο προπόνησης που υπόσχεται σημαντική βελτίωση του χρόνου στον οποίο μαθητές του Δημοτικού Σχολείου διανύουν τα 100 μέτρα. Για να ελεγχθεί η αποτελεσματικότητα της μεθόδου, Ν μαθητές εκλέγονται στην τύχη και καταγράφεται ο χρόνος στον οποίο διανύουν τα 100 μέτρα. Οι ίδιοι μαθητές ακολουθούν για κάποιο χρονικό διάστημα τη νέα μέθοδο προπόνησης και στη συνέχεια καταγράφεται και πάλι ο χρόνος στον οποίο καλύπτουν τρέχοντας απόσταση 100 μέτρων. Η αποτελεσματικότητα της μεθόδου προκύπτει από τη σύγκριση των αποτελεσμάτων προ και μετά την εφαρμογή της. Οι χρόνοι στους οποίους οι Ν μαθητές διένυσαν τα 100 μέτρα προ και μετά την εφαρμογή της νέας μεθόδου προπόνησής τους δεν συνιστούν τιμές ανεξάρτητων τυχαίων δειγμάτων, γιατί και οι δυο τιμές αναφέρονται στον χρόνο που καλύπτουν τα 100 μέτρα οι ίδιοι μαθητές οι ίδιες πειραματικές μονάδες. Γενικότερα τα τυχαία δείγματα που προκύπτουν από μετρήσεις στις ίδιες πειραματικές μονάδες προ και μετά την εφαρμογή μιας μεθόδου δεν είναι ανεξάρτητα. Στην περίπτωση αυτή τα προηγούμενα στατιστικά τεστ δεν εφαρμόζονται. Εφαρμόζεται όμως το τεστ συγκρίσεως ζευγών το οποίο δίνεται στη συνέχεια.

Έστω $X_{11}, X_{12}, \ldots, X_{1N}$ και $X_{21}, X_{22}, \ldots, X_{2N}$ είναι δυο τυχαία δείγματα τα οποία περιγράφουν τα αποτελέσματα προ και μετά την εφαρμογή ορισμένης μεθόδου σε Ν τυχαία επιλεγμένες πειραματικές μονάδες. Τα δείγματα αυτά επειδή αναφέρονται στις ίδιες πειραματικές μονάδες δεν είναι ανεξάρτητα. Αν συμβολίσουμε με μ_1 και μ_2 αντίστοιχα τις μέσες τιμές των αποτελεσμάτων προ και μετά την εφαρμογή της μεθόδου η αποτελεσματικότητά της ελέγχεται μέσω των υποθέσεων:

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

$$H_o : \mu_1 = \mu_2$$

έναντι των

$$H_a : \mu_1 > \mu_2 \text{ ή } H_a : \mu_1 < \mu_2 \text{ ή } H_a : \mu_1 \neq \mu_2 .$$

Για τον έλεγχο αυτό τα στατιστικά τεστ της παραγράφου 7.5 δεν εφαρμόζονται αφού τα τυχαία δείγματα δεν είναι ανεξάρτητα. Στην περίπτωση αυτή σχηματίζεται το τυχαίο δείγμα :

$$d_1 = X_{11} - X_{21}, \quad d_2 = X_{12} - X_{22}, \ldots, d_N = X_{1N} - X_{2N} ,$$

των διαφορών των δυο εξαρτημένων δειγμάτων. Αν υποτεθεί ότι το τυχαίο δείγμα των διαφορών $d_1, d_2, \ldots, d_N$ προέρχεται από κανονική κατανομή $N(\mu,\sigma^2)$ τότε $\mu = \mu_1 - \mu_2$ και οι προηγούμενες υποθέσεις για τον έλεγχο της αποτελεσματικότητας της μεθόδου, ισοδύναμα πλέον διατυπώνονται ως εξής :

$$H_o : \mu = 0$$

έναντι των εναλλακτικών υποθέσεων

$$H_a : \mu > 0 , \quad H_a : \mu < 0 , \quad H_a : \mu \neq 0.$$

Οι κρίσιμες περιοχές μεγέθους α για τον έλεγχο των παραπάνω υποθέσεων είναι αντίστοιχα

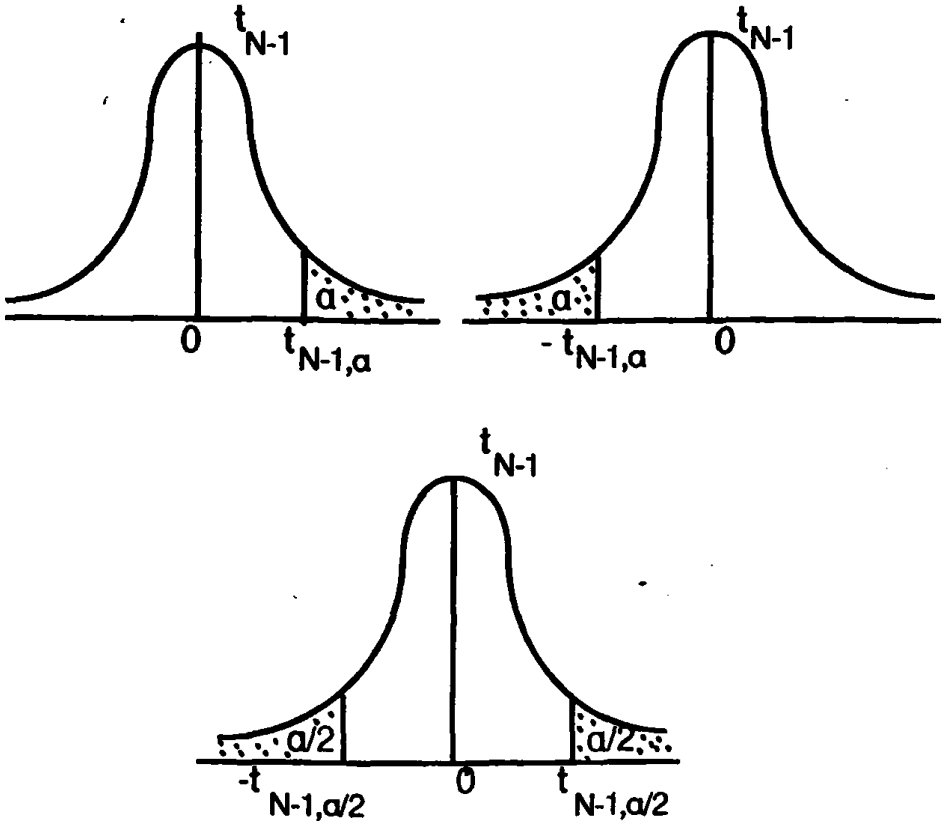
$$t \geq t_{N-1,\alpha} , \quad t \leq - t_{N-1,\alpha} , \quad |t| \geq t_{N-1,\, \alpha/2},$$

όπου $t = \dfrac{\sqrt{N}\,\bar{d}}{S_d}$ με $\bar{d}$ και S_d τη μέση τιμή και την τυπική απόκλιση, του δείγματος $d_1, d_2, \ldots, d_N$ των διαφορών, που ορίζονται:

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

$$\bar{d} = \frac{1}{N} \sum_{i=1}^{N} d_i \,,\; S_d = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (d_i - \bar{d})^2} \,.$$

Η γραφική παράσταση των παραπάνω κρίσιμων περιοχών δίνεται στο Σχήμα 7.3



Σχήμα 7.3: Γραφικές παραστάσεις κρίσιμων περιοχών του τεστ συγκρίσεως ζεύγων .

Παράδειγμα 7.6.1: Σε 9 μαθητές έκτης τάξης Δημοτικού σχολείου εφαρμόστηκε μια νέα μέθοδος προπόνησης με σκοπό να ελλατωθεί ο χρόνος στον οποίο διανύουν τρέχοντας τα 100 μέτρα. Οι ακόλουθες μετρήσεις συνιστούν το χρόνο σε δευτερόλεπτα των μαθητών προ και μετά την εφαρμογή της νέας μεθόδου προπόνησης.

Προ	12	14	17	11	19	20	20	21	16
Μετά	15	11	10	12	13	17	24	15	13

Να ελεγχθεί με επίπεδο α=1% αν η νέα μέθοδος είναι αποτελεσματική.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Επειδή οι μετρήσεις προ και μετά την εφαρμογή της νέας μεθόδου συνιστούν εξαρτημένα τυχαία δείγματα δημιουργείται το δείγμα της διαφοράς:

$$(d_i): -3 \quad 3 \quad 7 \quad -1 \quad 6 \quad 3 \quad -4 \quad 6 \quad 3 .$$

Υποθέτουμε ότι το τυχαίο δείγμα προέρχεται από κατανομή $N(\mu,\sigma^2)$. Για να ελεγχθεί αν η νέα μέθοδος προπόνησης είναι αποτελεσματική αρκεί να ελεγχθεί η μηδενική υπόθεση,

$$H_o : \mu = 0 \text{ έναντι της } H_a : \mu > 0 .$$

Για τον έλεγχο αυτό η κρίσιμη περιοχή μεγέθους $\alpha = 1\%$ είναι: $t \geq t_{8,0.01}$ με $t = \sqrt{N}\,\bar{d} / S_d$.

Είναι $\bar{d} = \frac{1}{9}\sum_{i=1}^{9} d_i = \frac{20}{9} = 2.22$, $S_d = \sqrt{\frac{1}{8}\sum_{i=1}^{9} (d_i-\bar{d})^2} = \sqrt{\frac{1}{8}\ 129.6} =$

4.02 και $t_{8,\,0.01} = 2.896$, ενώ $t = \frac{3\times 2.22}{4.02} = 1.66$.

Επειδή $t = 1.66 < 2.896 = t_{8,\,0.01}$ η H_o γίνεται αποδεκτή και έτσι με βάση τα δεδομένα αυτά η νέα μέθοδος προπόνησης δεν κρίνεται αποτελεσματική.

Παρατήρηση 7.6: Όπως στην περίπτωση των ανεξάρτητων κανονικών πληθυσμών (Παρατήρηση 7.5) έτσι και εδώ, πολλές φορές ενδιαφέρει να ελεγχθεί αν η αποτελεσματικότητα μιας μεθόδου είναι ίση ή διαφορετική με μια προκαθορισμένη ποσότητα δ. Έτσι οι υποθέσεις του τεστ συγκρίσεως ζευγών διατυπώνονται σε γενικότερη μορφή ως εξής:

$$H_o : \mu = \delta$$

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ως προς τις εναλλακτικές υποθέσεις

$$H_a : \mu > \delta , \quad H_a : \mu < \delta , \quad H_a : \mu \neq \delta$$

με μ τη μέση τιμή του πληθυσμού απ' τον οποίο προέρχεται το δείγμα των διαφορών. Για τον έλεγχο των υποθέσεων αυτών η στατιστική συνάρτηση γίνεται $t = \frac{\sqrt{N}(\bar{d}-\delta)}{S_d}$ με $\bar{d}$ και S_d όπως ορίστηκαν προηγούμενα, ενώ οι κρίσιμες περιοχές παραμένουν οι ίδιες.

7.7 Γενικές παρατηρήσεις

1. Τα στατιστικά τεστ που δόθηκαν στις προηγούμενες παραγράφους στηρίζονται στην υπόθεση ότι ο υπό μελέτη πληθυσμός περιγράφεται από την κανονική κατανομή. Για περιπτώσεις που η υπόθεση αυτή δεν μπορεί να υιοθετηθεί χρησιμοποιείται το κεντρικό οριακό θεώρημα σύμφωνα με το οποίο η μέση τιμή μεγάλου αριθμού μετρήσεων ακολουθεί την κανονική κατανομή, οποιαδήποτε και να είναι η κατανομή του πληθυσμού απ' τον οποίο προέρχονται οι μετρήσεις. Έτσι τα στατιστικά τεστ που δόθηκαν στις παραγράφους 7.4, 7.5 και 7.6 αποτελούν (προσεγγιστικά) στατιστικά τεστ για τον έλεγχο της μέσης τιμής οποιουδήποτε (μη κανονικού) πληθυσμού.

2. Ο έλεγχος στατιστικών υποθέσεων και η εκτίμηση σε διάστημα δεν είναι τόσο διαφορετικές μεθοδολογίες στατιστικής συμπερασματολογίας όσο από πρώτη άποψη φαίνεται. Ας θεωρήσουμε για παράδειγμα το 100(1-α)% διάστημα εμπιστοσύνης για τη μέση τιμή

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

κανονικού πληθυσμού με γνωστή διακύμανση σ^2 και το στατιστικό τεστ μεγέθους α για τον έλεγχο της $H_o : \mu = \mu_o$ έναντι της $H_a : \mu \neq \mu_o$. Η H_o απορρίπτεται όταν η τιμή της $Z = \frac{\bar{X}-\mu_o}{\sigma/\sqrt{N}}$ είναι τέτοια ώστε $|Z| \geq z_{\alpha/2}$ ή ισοδύναμα $Z \geq z_{\alpha/2}$ ή $Z \leq -z_{\alpha/2}$ ή ισοδύναμα $\frac{\bar{X}-\mu_o}{\sigma/\sqrt{N}} \geq z_{\alpha/2}$ ή $\frac{\bar{X}-\mu_o}{\sigma/\sqrt{N}} \leq -z_{\alpha/2}$ ή ισοδύναμα $\mu_o \leq \bar{X} - z_{\alpha\backslash 2}\frac{\sigma}{\sqrt{N}}$ ή $\mu_o \geq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}$ ή η τιμή μ_o που καθορίζεται από τη μηδενική υπόθεση, δεν περιέχεται στο 100(1-α)% διάστημα εμπιστοσύνης της άγνωστης παραμέτρου. Συνοψίζοντας, η απόρριψη της $H_o : \mu = \mu_o$ σε επίπεδο α είναι ισοδύναμη με το γεγονός ότι η τιμή μ_o δεν περιέχεται στο 100(1-α)% διάστημα εμπιστοσύνης για τη μέση τιμή μ του πληθυσμού.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Ασκήσεις 6ου & 7ου Κεφαλαίου

6-7.1: Η μέτρηση του ύψους, σε ένα δείγμα 20 ατόμων από ένα πληθυσμό έδωσε τα παρακάτω αποτελέσματα σε εκατοστά:

173, 166, 168, 166, 169, 166, 173, 170, 170, 173,
166, 161, 166, 170, 168, 158, 173, 166, 165, 165.

α) Αν είναι γνωστό ότι τα ύψη ακολουθούν κανονική κατανομή $N(\mu,\sigma^2)$ με $\sigma=5$ εκατοστά:

i) Να κατασκευαστεί το 95% διάστημα εμπιστοσύνης για το μέσο ύψος μ του πληθυσμού.

ii) Να ελεγθεί με επίπεδο σημαντικότητας 5% αν το μέσο ύψος του πληθυσμού είναι μεγαλύτερο των 165 εκατοστών.

β) Αν τα ύψη ακολουθούν κανονική κατανομή με άγνωστη διακύμανση να απαντηθούν τα ερωτήματα i) και ii) της περίπτωσης α).

6-7.2: Για να εκτιμηθεί η μέση βαθμολογία των φοιτητών-τριών στις εξετάσεις κάποιου μαθήματος, παίρνουμε ένα δείγμα 16 φοιτητών - τριών και καταγράφουμε τη βαθμολογία τους:

10, 3, 5, 4, 7, 8, 9, 9, 8, 5, 5, 8, 6, 6, 9, 10.

Να εκτιμηθεί το μ: i) σε σημείο, ii) σε διάστημα με βαθμό εμπιστοσύνης 90%, iii) να ελεγθεί, με επίπεδο σημαντικότητας 5%, ο ισχυρισμός ότι η μέση βαθμολογία είναι διαφορετική του 6.5. Υποθέστε ότι η βαθμολογία στις εξετάσεις ακολουθεί κανονική κατανομή:

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

α) $N(\mu, 4)$, σ^2 γνωστό και β) $N(\mu,\sigma^2)$, σ^2 άγνωστο.

6-7.3: Κάθε χρόνο οι φοιτήτριες που τελειώνουν το Α' έτος σπουδών, υποβάλλονται σ' ένα έλεγχο φυσικής κατάστασης και βαθμολογούνται. Τα προηγούμενα χρόνια ο μέσος όρος βαθμολογίας ήταν 180. Απ' την αρχή της φετεινής χρονιάς εφαρμόστηκε ένα πρόγραμμα βελτίωσης της φυσικής κατάστασης. Στό τέλος του χρόνου επιλέγονται 50 από αυτές βαθμολογούνται μετά από έλεγχο και δίνουν μέσο όρο 190 βαθμών με τυπική απόκλιση 35.2. Να ελεγχθεί με επίπεδο σημαντικότητας $\alpha=0.01$ αν μπορούμε να ισχυριστούμε ότι είχαμε αύξηση της μέσης βαθμολογίας των Αετών φοιτητριών δηλαδή βελτίωση της φυσικής τους κατάστασης που να οφείλεται στο πρόγραμμα.

6-7.4: Ένα τυχαίο δείγμα από 10 μαθητές και ένα τυχαίο δείγμα από 10 μαθήτριες έδωσαν μέσο ύψος αντίστοιχα 152 και 149 εκατοστά.

i) Να εκτιμηθεί η διαφορά των μέσων τιμών των πληθυσμών των μαθητών και μαθητριών.

ii) Να κατασκευαστεί το 95% διάστημα εμπιστοσύνης της διαφοράς των μέσων τιμών των πληθυσμών αυτών.

iii) Να ελεγθεί, με επίπεδο σημαντικότητας 10%, ο ισχυρισμός ότι οι μαθητές είναι ψηλότεροι από τις μαθήτριες.

Υποθέστε: α) ότι το ύψος του πληθυσμού των μαθητών ακολουθεί κανονική κατανομή με διακύμανση 25 ενώ το ύψος των μαθητριών κανονική κατανομή με διακύμανση 49.

β) τα ύψη των μαθητών και μαθητριών προέρχονται από κανονικές κατανομές με άγνωστες αλλά ίσες διακυμάνσεις, ενώ

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

οι δειγματικές διακυμάνσεις των υψών των 10 μαθητών και μαθητριών είναι αντίστοιχα 20 και 30.

6-7.5: Στον πίνακα που ακολουθεί δίνεται το βάρος σε κιλά 9 γυναικών πριν και μετά την εφαρμογή μιας δίαιτας αδυνατίσματος τεσσάρων εβδομάδων.

Βάρος πριν την δίαιτα	67	81	57	68	67	69	66	77	54
Βάρος μετά την δίαιτα	59	69	56	59	63	66	64	71	54

Με βάση τα δεδομένα αυτά και κάνοντας τις κατάλληλες υποθέσεις να ελεγθεί με επίπεδο σημαντικότητας 5% ο ισχυρισμός ότι η δίαιτα είναι αποτελεσματική. Αν είναι αποτελεσματική μπορούμε να ισχυριστούμε με επίπεδο σημαντικότητας 5% ότι η δίαιτα προσφέρει ελάττωση βάρους κατά 3 κιλά σε τέσσερις εβδομάδες;

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Π Α Ρ Α Ρ Τ Η Μ Α

Στατιστικοί Πίνακες

I. Διωνυμική Κατανομή

II. Κατανομή Poisson

III. Τυπική Κανονική Κατανομή

IV. Κατανομή t

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΠΑΡΑΡΤΗΜΑ

Στατιστικοί Πίνακες

I. Διωνυμική Κατανομή

II. Κατανομή Ρούσσας

III. Τυπική Κανονική Κατανομή

IV. Κατανομή 1

ΠΙΝΑΚΑΣ Ι: Διωνυμική Κατανομή

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad , \quad x = 0,1,\ldots,n.$$

Ο πίνακας που ακολουθεί δίνει τις τιμές της $p_X(x)$ για διάφορες τιμές των παραμέτρων n και p.

		p									
n	x	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
1	0	.9500	.9000	.8500	.8000	.7500	.7000	.6500	.6000	.5500	.5000
	1	.0500	.1000	.1500	.2000	.2500	.3000	.3500	.4000	.4500	.5000
2	0	.9025	.8100	.7225	.6400	.5625	.4900	.4225	.3600	.3025	.2500
	1	.0950	.1800	.2550	.3200	.3750	.4200	.4550	.4800	.4950	.5000
	2	.0025	.0100	.0225	.0400	.0625	.0900	.1225	.1600	.2025	.2500
3	0	.8574	.7290	.6141	.5120	.4219	.3430	.2746	.2160	.1664	.1250
	1	.1354	.2430	.3251	.3840	.4219	.4410	.4436	.4320	.4084	.3750
	2	.0071	.0270	.0574	.0960	.1406	.1890	.2389	.2880	.3341	.3750
	3	.0001	.0010	.0034	.0080	.0156	.0270	.0429	.0640	.0911	.1250
4	0	.8145	.6561	.5220	.4096	.3164	.2401	.1785	.1296	.0915	.0625
	1	.1715	.2916	.3685	.4096	.4219	.4116	.3845	.3456	.2995	.2500
	2	.0135	.0486	.0975	.1536	.2109	.2646	.3105	.3456	.3675	.3750
	3	.0005	.0036	.0115	.0256	.0469	.0756	.1115	.1536	.2005	.2500
	4	.0000	.0001	.0005	.0016	.0039	.0081	.0150	.0256	.0410	.0625
5	0	.7738	.5905	.4437	.3277	.2373	.1681	.1160	.0778	.0503	.0312
	1	.2036	.3280	.3915	.4096	.3955	.3602	.3124	.2592	.2059	.1562
	2	.0214	.0729	.1382	.2048	.2637	.3087	.3364	.3456	.3369	.3125
	3	.0011	.0081	.0244	.0512	.0879	.1323	.1811	.2304	.2757	.3125
	4	.0000	.0004	.0022	.0064	.0146	.0284	.0488	.0768	.1128	.1562
	5	.0000	.0000	.0001	.0003	.0010	.0024	.0053	.0102	.0185	.0312
6	0	.7351	.5314	.3771	.2621	.1780	.1176	.0754	.0467	.0277	.0156
	1	.2321	.3543	.3993	.3932	.3560	.3025	.2437	.1866	.1359	.0938
	2	.0305	.0984	.1762	.2458	.2966	.3241	.3280	.3110	.2780	.2344
	3	.0021	.0146	.0415	.0819	.1318	.1852	.2355	.2765	.3032	.3125
	4	.0001	.0012	.0055	.0154	.0330	.0595	.0951	.1382	.1861	.2344
	5	.0000	.0001	.0004	.0015	.0044	.0102	.0205	.0369	.0609	.0938
	6	.0000	.0000	.0000	.0001	.0002	.0007	.0018	.0041	.0083	.0156
7	0	.6983	.4783	.3206	.2097	.1335	.0824	.0490	.0280	.0152	.0078
	1	.2573	.3720	.3960	.3670	.3115	.2471	.1848	.1306	.0872	.0547
	2	.0406	.1240	.2097	.2753	.3115	.3177	.2985	.2613	.2140	.1641
	3	.0036	.0230	.0617	.1147	.1730	.2269	.2679	.2903	.2918	.2734
	4	.0002	.0026	.0109	.0287	.0577	.0972	.1442	.1935	.2388	.2734
	5	.0000	.0002	.0012	.0043	.0115	.0250	.0466	.0774	.1172	.1641
	6	.0000	.0000	.0001	.0004	.0013	.0036	.0084	.0172	.0320	.0547
	7	.0000	.0000	.0000	.0000	.0001	.0002	.0006	.0016	.0037	.0078
8	0	.6634	.4305	.2725	.1678	.1001	.0576	.0319	.0168	.0084	.0039
	1	.2793	.3826	.3847	.3355	.2670	.1977	.1373	.0896	.0548	.0312
	2	.0515	.1488	.2376	.2936	.3115	.2965	.2587	.2090	.1569	.1094
	3	.0054	.0331	.0839	.1468	.2076	.2541	.2786	.2787	.2568	.2188
	4	.0004	.0046	.0185	.0459	.0865	.1361	.1875	.2322	.2627	.2734
	5	.0000	.0004	.0026	.0092	.0231	.0467	.0808	.1239	.1719	.2188
	6	.0000	.0000	.0002	.0011	.0038	.0100	.0217	.0413	.0703	.1094
	7	.0000	.0000	.0000	.0001	.0004	.0012	.0033	.0079	.0164	.0312
	8	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0007	.0017	.0039

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΠΙΝΑΚΑΣ Ι: Διωνυμική Κατανομή (Συνέχεια)

n	x	p = .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
9	0	.6302	.3874	.2316	.1342	.0751	.0404	.0207	.0101	.0046	.0020
	1	.2985	.3874	.3679	.3020	.2253	.1556	.1004	.0605	.0339	.0176
	2	.0629	.1722	.2597	.3020	.3003	.2668	.2162	.1612	.1110	.0703
	3	.0077	.0446	.1069	.1762	.2336	.2668	.2716	.2508	.2119	.1641
	4	.0006	.0074	.0283	.0661	.1168	.1715	.2194	.2508	.2600	.2461
	5	.0000	.0008	.0050	.0165	.0389	.0735	.1181	.1672	.2128	.2461
	6	.0000	.0001	.0006	.0028	.0087	.0210	.0424	.0743	.1160	.1641
	7	.0000	.0000	.0000	.0003	.0012	.0039	.0098	.0212	.0407	.0703
	8	.0000	.0000	.0000	.0000	.0001	.0004	.0013	.0035	.0083	.0176
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0008	.0020
10	0	.5987	.3487	.1969	.1074	.0563	.0282	.0135	.0060	.0025	.0010
	1	.3151	.3874	.3474	.2684	.1877	.1211	.0725	.0403	.0207	.0098
	2	.0746	.1937	.2759	.3020	.2816	.2335	.1757	.1209	.0763	.0439
	3	.0105	.0574	.1298	.2013	.2503	.2668	.2522	.2150	.1665	.1172
	4	.0010	.0112	.0401	.0881	.1460	.2001	.2377	.2508	.2384	.2051
	5	.0001	.0015	.0085	.0264	.0584	.1029	.1536	.2007	.2340	.2461
	6	.0000	.0001	.0012	.0055	.0162	.0368	.0689	.1115	.1596	.2051
	7	.0000	.0000	.0001	.0008	.0031	.0090	.0212	.0425	.0746	.1172
	8	.0000	.0000	.0000	.0001	.0004	.0014	.0043	.0106	.0229	.0439
	9	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0016	.0042	.0098
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010
11	0	.5688	.3138	.1673	.0859	.0422	.0198	.0088	.0036	.0014	.0005
	1	.3293	.3835	.3248	.2362	.1549	.0932	.0518	.0266	.0125	.0054
	2	.0867	.2131	.2866	.2953	.2581	.1998	.1395	.0887	.0513	.0269
	3	.0137	.0710	.1517	.2215	.2581	.2568	.2254	.1774	.1259	.0806
	4	.0014	.0158	.0536	.1107	.1721	.2201	.2428	.2365	.2060	.1611
	5	.0001	.0025	.0132	.0388	.0803	.1321	.1830	.2207	.2360	.2256
	6	.0000	.0003	.0023	.0097	.0268	.0566	.0985	.1471	.1931	.2256
	7	.0000	.0000	.0003	.0017	.0064	.0173	.0379	.0701	.1128	.1611
	8	.0000	.0000	.0000	.0002	.0011	.0037	.0102	.0234	.0462	.0806
	9	.0000	.0000	.0000	.0000	.0001	.0005	.0018	.0052	.0126	.0269
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0007	.0021	.0054
	11	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0005
12	0	.5404	.2824	.1422	.0687	.0317	.0138	.0057	.0022	.0008	.0002
	1	.3413	.3766	.3012	.2062	.1267	.0712	.0368	.0174	.0075	.0029
	2	.0988	.2301	.2924	.2835	.2323	.1678	.1088	.0639	.0339	.0161
	3	.0173	.0852	.1720	.2362	.2581	.2397	.1954	.1419	.0923	.0537
	4	.0021	.0213	.0683	.1329	.1936	.2311	.2367	.2128	.1700	.1208
	5	.0002	.0038	.0193	.0532	.1032	.1585	.2039	.2270	.2225	.1934
	6	.0000	.0005	.0040	.0155	.0401	.0792	.1281	.1766	.2124	.2256
	7	.0000	.0000	.0006	.0033	.0115	.0291	.0591	.1009	.1489	.1934
	8	.0000	.0000	.0001	.0005	.0024	.0078	.0199	.0420	.0762	.1208
	9	.0000	.0000	.0000	.0001	.0004	.0015	.0048	.0125	.0277	.0537
	10	.0000	.0000	.0000	.0000	.0000	.0002	.0008	.0025	.0068	.0161
	11	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010	.0029
	12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002
13	0	.5133	.2542	.1209	.0550	.0238	.0097	.0037	.0013	.0004	.0001
	1	.3512	.3672	.2774	.1787	.1029	.0540	.0259	.0113	.0045	.0016
	2	.1109	.2448	.2937	.2680	.2059	.1388	.0836	.0453	.0220	.0095
	3	.0214	.0997	.1900	.2457	.2517	.2181	.1651	.1107	.0660	.0349
	4	.0028	.0277	.0838	.1535	.2097	.2337	.2222	.1845	.1350	.0873
	5	.0003	.0055	.0266	.0691	.1258	.1803	.2154	.2214	.1989	.1571
	6	.0000	.0008	.0063	.0230	.0559	.1030	.1546	.1968	.2169	.2095
	7	.0000	.0001	.0011	.0058	.0186	.0442	.0833	.1312	.1775	.2095
	8	.0000	.0000	.0001	.0011	.0047	.0142	.0336	.0656	.1089	.1571
	9	.0000	.0000	.0000	.0001	.0009	.0034	.0101	.0243	.0495	.0873
	10	.0000	.0000	.0000	.0000	.0001	.0006	.0022	.0065	.0162	.0349
	11	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0012	.0036	.0095
	12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0016
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΠΙΝΑΚΑΣ Ι: Διωνυμική Κατανομή (Συνέχεια)

		p									
n	x	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
14	0	.4877	.2288	.1028	.0440	.0178	.0068	.0024	.0008	.0002	.0001
	1	.3593	.3559	.2539	.1539	.0832	.0407	.0181	.0073	.0027	.0009
	2	.1229	.2570	.2912	.2501	.1802	.1134	.0634	.0317	.0141	.0056
	3	.0259	.1142	.2056	.2501	.2402	.1943	.1366	.0845	.0462	.0222
	4	.0037	.0349	.0998	.1720	.2202	.2290	.2022	.1549	.1040	.0611
	5	.0004	.0078	.0352	.0860	.1468	.1963	.2178	.2066	.1701	.1222
	6	.0000	.0013	.0093	.0322	.0734	.1262	.1759	.2066	.2088	.1833
	7	.0000	.0002	.0019	.0092	.0280	.0618	.1082	.1574	.1952	.2095
	8	.0000	.0000	.0003	.0020	.0082	.0232	.0510	.0918	.1398	.1833
	9	.0000	.0000	.0000	.0003	.0018	.0066	.0183	.0408	.0762	.1222
	10	.0000	.0000	.0000	.0000	.0003	.0014	.0049	.0136	.0312	.0611
	11	.0000	.0000	.0000	.0000	.0000	.0002	.0010	.0033	.0093	.0222
	12	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0019	.0056
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0009
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
15	0	.4633	.2059	.0874	.0352	.0134	.0047	.0016	.0005	.0001	.0000
	1	.3658	.3432	.2312	.1319	.0668	.0305	.0126	.0047	.0016	.0005
	2	.1348	.2669	.2856	.2309	.1559	.0916	.0476	.0219	.0090	.0032
	3	.0307	.1285	.2184	.2501	.2252	.1700	.1110	.0634	.0318	.0139
	4	.0049	.0428	.1156	.1876	.2252	.2186	.1792	.1268	.0780	.0417
	5	.0006	.0105	.0449	.1032	.1651	.2061	.2123	.1859	.1404	.0916
	6	.0000	.0019	.0132	.0430	.0917	.1472	.1906	.2066	.1914	.1527
	7	.0000	.0003	.0030	.0138	.0393	.0811	.1319	.1771	.2013	.1964
	8	.0000	.0000	.0005	.0035	.0131	.0348	.0710	.1181	.1647	.1964
	9	.0000	.0000	.0001	.0007	.0034	.0116	.0298	.0612	.1048	.1527
	10	.0000	.0000	.0000	.0001	.0007	.0030	.0096	.0245	.0515	.0916
	11	.0000	.0000	.0000	.0000	.0001	.0006	.0024	.0074	.0191	.0417
	12	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0016	.0052	.0139
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010	.0032
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
16	0	.4401	.1853	.0743	.0281	.0100	.0033	.0010	.0003	.0001	.0000
	1	.3706	.3294	.2097	.1126	.0535	.0228	.0087	.0030	.0009	.0002
	2	.1463	.2745	.2775	.2111	.1336	.0732	.0353	.0150	.0056	.0018
	3	.0359	.1423	.2285	.2463	.2079	.1465	.0888	.0468	.0215	.0085
	4	.0061	.0514	.1311	.2001	.2252	.2040	.1553	.1014	.0572	.0278
	5	.0008	.0137	.0555	.1201	.1802	.2099	.2008	.1623	.1123	.0667
	6	.0001	.0028	.0180	.0550	.1101	.1649	.1982	.1983	.1684	.1222
	7	.0000	.0004	.0045	.0197	.0524	.1010	.1524	.1889	.1969	.1746
	8	.0000	.0001	.0009	.0055	.0197	.0487	.0923	.1417	.1812	.1964
	9	.0000	.0000	.0001	.0012	.0058	.0185	.0442	.0840	.1318	.1746
	10	.0000	.0000	.0000	.0002	.0014	.0056	.0167	.0392	.0755	.1222
	11	.0000	.0000	.0000	.0000	.0002	.0013	.0049	.0142	.0337	.0667
	12	.0000	.0000	.0000	.0000	.0000	.0002	.0011	.0040	.0115	.0278
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0008	.0029	.0085
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0018
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
17	0	.4181	.1668	.0631	.0225	.0075	.0023	.0007	.0002	.0000	.0000
	1	.3741	.3150	.1893	.0957	.0426	.0169	.0060	.0019	.0005	.0001
	2	.1575	.2800	.2673	.1914	.1136	.0581	.0260	.0102	.0035	.0010
	3	.0415	.1556	.2359	.2393	.1893	.1245	.0701	.0341	.0144	.0052
	4	.0076	.0605	.1457	.2093	.2209	.1868	.1320	.0796	.0411	.0182
	5	.0010	.0175	.0668	.1361	.1914	.2081	.1849	.1379	.0875	.0472
	6	.0001	.0039	.0236	.0680	.1276	.1784	.1991	.1839	.1432	.0944
	7	.0000	.0007	.0065	.0267	.0668	.1201	.1685	.1927	.1841	.1484
	8	.0000	.0001	.0014	.0084	.0279	.0644	.1134	.1606	.1883	.1855
	9	.0000	.0000	.0003	.0021	.0093	.0276	.0611	.1070	.1540	.1855
	10	.0000	.0000	.0000	.0004	.0025	.0095	.0263	.0571	.1008	.1484
	11	.0000	.0000	.0000	.0001	.0005	.0026	.0090	.0242	.0525	.0944
	12	.0000	.0000	.0000	.0000	.0001	.0006	.0024	.0081	.0215	.0472
	13	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0021	.0068	.0182
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0016	.0052

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΠΙΝΑΚΑΣ Ι: Διωνυμική Κατανομή (Συνέχεια)

n	x	p .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
17	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
18	0	.3972	.1501	.0536	.0180	.0056	.0016	.0004	.0001	.0000	.0000
	1	.3763	.3002	.1704	.0811	.0338	.0126	.0042	.0012	.0003	.0001
	2	.1683	.2835	.2556	.1723	.0958	.0458	.0190	.0069	.0022	.0006
	3	.0473	.1680	.2406	.2297	.1704	.1046	.0547	.0246	.0095	.0031
	4	.0093	.0700	.1592	.2153	.2130	.1681	.1104	.0614	.0291	.0117
	5	.0014	.0218	.0787	.1507	.1988	.2017	.1664	.1146	.0666	.0327
	6	.0002	.0052	.0301	.0816	.1436	.1873	.1941	.1655	.1181	.0708
	7	.0000	.0010	.0091	.0350	.0820	.1376	.1792	.1892	.1657	.1214
	8	.0000	.0002	.0022	.0120	.0376	.0811	.1327	.1734	.1864	.1669
	9	.0000	.0000	.0004	.0033	.0139	.0386	.0794	.1284	.1694	.1855
	10	.0000	.0000	.0001	.0008	.0042	.0149	.0385	.0771	.1248	.1669
	11	.0000	.0000	.0000	.0001	.0010	.0046	.0151	.0374	.0742	.1214
	12	.0000	.0000	.0000	.0000	.0002	.0012	.0047	.0145	.0354	.0708
	13	.0000	.0000	.0000	.0000	.0000	.0002	.0012	.0045	.0134	.0327
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0011	.0039	.0117
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0009	.0031
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0006
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
19	0	.3774	.1351	.0456	.0144	.0042	.0011	.0003	.0001	.0000	.0000
	1	.3774	.2852	.1529	.0685	.0268	.0093	.0029	.0008	.0002	.0000
	2	.1787	.2852	.2428	.1540	.0803	.0358	.0138	.0046	.0013	.0003
	3	.0533	.1796	.2428	.2182	.1517	.0869	.0422	.0175	.0062	.0018
	4	.0112	.0798	.1714	.2182	.2023	.1491	.0909	.0467	.0203	.0074
	5	.0018	.0266	.0907	.1636	.2023	.1916	.1468	.0933	.0497	.0222
	6	.0002	.0069	.0374	.0955	.1574	.1916	.1844	.1451	.0949	.0518
	7	.0000	.0014	.0122	.0443	.0974	.1525	.1844	.1797	.1443	.0961
	8	.0000	.0002	.0032	.0166	.0487	.0981	.1489	.1797	.1771	.1442
	9	.0000	.0000	.0007	.0051	.0198	.0514	.0980	.1464	.1771	.1762
	10	.0000	.0000	.0001	.0013	.0066	.0220	.0528	.0976	.1449	.1762
	11	.0000	.0000	.0000	.0003	.0018	.0077	.0233	.0532	.0970	.1442
	12	.0000	.0000	.0000	.0000	.0004	.0022	.0083	.0237	.0529	.0961
	13	.0000	.0000	.0000	.0000	.0001	.0005	.0024	.0085	.0233	.0518
	14	.0000	.0000	.0000	.0000	.0000	.0001	.0006	.0024	.0082	.0222
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0022	.0074
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0018
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
20	0	.3585	.1216	.0388	.0115	.0032	.0008	.0002	.0000	.0000	.0000
	1	.3774	.2702	.1368	.0576	.0211	.0068	.0020	.0005	.0001	.0000
	2	.1887	.2852	.2293	.1369	.0669	.0278	.0100	.0031	.0008	.0002
	3	.0596	.1901	.2428	.2054	.1339	.0716	.0323	.0123	.0040	.0011
	4	.0133	.0898	.1821	.2182	.1897	.1304	.0738	.0350	.0139	.0046
	5	.0022	.0319	.1028	.1746	.2023	.1789	.1272	.0746	.0365	.0148
	6	.0003	.0089	.0454	.1091	.1686	.1916	.1712	.1244	.0746	.0370
	7	.0000	.0020	.0160	.0545	.1124	.1643	.1844	.1659	.1221	.0739
	8	.0000	.0004	.0046	.0222	.0609	.1144	.1614	.1797	.1623	.1201
	9	.0000	.0001	.0011	.0074	.0271	.0654	.1158	.1597	.1771	.1602
	10	.0000	.0000	.0002	.0020	.0099	.0308	.0686	.1171	.1593	.1762
	11	.0000	.0000	.0000	.0005	.0030	.0120	.0336	.0710	.1185	.1602
	12	.0000	.0000	.0000	.0001	.0008	.0039	.0136	.0355	.0727	.1201
	13	.0000	.0000	.0000	.0000	.0002	.0010	.0045	.0146	.0366	.0739
	14	.0000	.0000	.0000	.0000	.0000	.0002	.0012	.0049	.0150	.0370
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0013	.0049	.0148
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0013	.0046
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0011
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002
	19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	20	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΠΙΝΑΚΑΣ ΙΙ: Κατανομή Poisson

$$p_x(x) = \frac{e^{-\lambda}\ \lambda^x}{x!} \quad , \quad x = 0,1,2,...$$

Ο πίνακας που ακολουθεί δίνει τις τιμές της $p_x(x)$ για διάφορες τιμές της παραμέτρου λ.

						λ				
x	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
0	.9048	.8187	.7408	.6703	.6065	.5488	.4966	.4493	.4066	.3679
1	.0905	.1637	.2222	.2681	.3033	.3293	.3476	.3595	.3659	.3679
2	.0045	.0164	.0333	.0536	.0758	.0988	.1217	.1438	.1647	.1839
3	.0002	.0011	.0033	.0072	.0126	.0198	.0284	.0383	.0494	.0613
4	.0000	.0001	.0002	.0007	.0016	.0030	.0050	.0077	.0111	.0153
5	.0000	.0000	.0000	.0001	.0002	.0004	.0007	.0012	.0020	.0031
6	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0003	.0005
7	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

						λ				
x	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0	.3329	.3012	.2725	.2466	.2231	.2019	.1827	.1653	.1496	.1353
1	.3662	.3614	.3543	.3452	.3347	.3230	.3106	.2975	.2842	.2707
2	.2014	.2169	.2303	.2417	.2510	.2584	.2640	.2678	.2700	.2707
3	.0738	.0867	.0998	.1128	.1255	.1378	.1496	.1607	.1710	.1804
4	.0203	.0260	.0324	.0395	.0471	.0551	.0636	.0723	.0812	.0902
5	.0045	.0062	.0084	.0111	.0141	.0176	.0216	.0260	.0309	.0361
6	.0008	.0012	.0018	.0026	.0035	.0047	.0061	.0078	.0098	.0120
7	.0001	.0002	.0003	.0005	.0008	.0011	.0015	.0020	.0027	.0034
8	.0000	.0000	.0001	.0001	.0001	.0002	.0003	.0005	.0006	.0009
9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0002

						λ				
x	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
0	.1225	.1108	.1003	.0907	.0821	.0743	.0672	.0608	.0550	.0498
1	.2572	.2438	.2306	.2177	.2052	.1931	.1815	.1703	.1596	.1494
2	.2700	.2681	.2652	.2613	.2565	.2510	.2450	.2384	.2314	.2240
3	.1890	.1966	.2033	.2090	.2138	.2176	.2205	.2225	.2237	.2240
4	.0992	.1082	.1169	.1254	.1336	.1414	.1488	.1557	.1622	.1680
5	.0417	.0476	.0538	.0602	.0668	.0735	.0804	.0872	.0940	.1008
6	.0146	.0174	.0206	.0241	.0278	.0319	.0362	.0407	.0455	.0504
7	.0044	.0055	.0068	.0083	.0099	.0118	.0139	.0163	.0188	.0216
8	.0011	.0015	.0019	.0025	.0031	.0038	.0047	.0057	.0068	.0081
9	.0003	.0004	.0005	.0007	.0009	.0011	.0014	.0018	.0022	.0027
10	.0001	.0001	.0001	.0002	.0002	.0003	.0004	.0005	.0006	.0008
11	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0002	.0002
12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

						λ				
x	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
0	.0450	.0408	.0369	.0334	.0302	.0273	.0247	.0224	.0202	.0183
1	.1397	.1304	.1217	.1135	.1057	.0984	.0915	.0850	.0789	.0733
2	.2165	.2087	.2008	.1929	.1850	.1771	.1692	.1615	.1539	.1465
3	.2237	.2226	.2209	.2186	.2158	.2125	.2087	.2046	.2001	.1954
4	.1734	.1781	.1823	.1858	.1888	.1912	.1931	.1944	.1951	.1954
5	.1075	.1140	.1203	.1264	.1322	.1377	.1429	.1477	.1522	.1563
6	.0555	.0608	.0662	.0716	.0771	.0826	.0881	.0936	.0989	.1042
7	.0246	.0278	.0312	.0348	.0385	.0425	.0466	.0508	.0551	.0595
8	.0095	.0111	.0129	.0148	.0169	.0191	.0215	.0241	.0269	.0298
9	.0033	.0040	.0047	.0056	.0066	.0076	.0089	.0102	.0116	.0132
10	.0010	.0013	.0016	.0019	.0023	.0028	.0033	.0039	.0045	.0053
11	.0003	.0004	.0005	.0006	.0007	.0009	.0011	.0013	.0016	.0019
12	.0001	.0001	.0001	.0002	.0002	.0003	.0003	.0004	.0005	.0006
13	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0002	.0002
14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΠΙΝΑΚΑΣ ΙΙ: Κατανομή Poisson (Συνέχεια)

						λ				
x	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0
0	.0166	.0150	.0136	.0123	.0111	.0101	.0091	.0082	.0074	.0067
1	.0679	.0630	.0583	.0540	.0500	.0462	.0427	.0395	.0365	.0337
2	.1393	.1323	.1254	.1188	.1125	.1063	.1005	.0948	.0894	.0842
3	.1904	.1852	.1798	.1743	.1687	.1631	.1574	.1517	.1460	.1404
4	.1951	.1944	.1933	.1917	.1898	.1875	.1849	.1820	.1789	.1755
5	.1600	.1633	.1662	.1687	.1708	.1725	.1738	.1747	.1753	.1755
6	.1093	.1143	.1191	.1237	.1281	.1323	.1362	.1398	.1432	.1462
7	.0640	.0686	.0732	.0778	.0824	.0869	.0914	.0959	.1002	.1044
8	.0328	.0360	.0393	.0428	.0463	.0500	.0537	.0575	.0614	.0653
9	.0150	.0168	.0188	.0209	.0232	.0255	.0280	.0307	.0334	.0363
10	.0061	.0071	.0081	.0092	.0104	.0118	.0132	.0147	.0164	.0181
11	.0023	.0027	.0032	.0037	.0043	.0049	.0056	.0064	.0073	.0082
12	.0008	.0009	.0011	.0014	.0016	.0019	.0022	.0026	.0030	.0034
13	.0002	.0003	.0004	.0005	.0006	.0007	.0008	.0009	.0011	.0013
14	.0001	.0001	.0001	.0001	.0002	.0002	.0003	.0003	.0004	.0005
15	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0001	.0002

						λ				
x	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6.0
0	.0061	.0055	.0050	.0045	.0041	.0037	.0033	.0030	.0027	.0025
1	.0311	.0287	.0265	.0244	.0225	.0207	.0191	.0176	.0162	.0149
2	.0793	.0746	.0701	.0659	.0618	.0580	.0544	.0509	.0477	.0446
3	.1348	.1293	.1239	.1185	.1133	.1082	.1033	.0985	.0938	.0892
4	.1719	.1681	.1641	.1600	.1558	.1515	.1472	.1428	.1383	.1339
5	.1753	.1748	.1740	.1728	.1714	.1697	.1678	.1656	.1632	.1606
6	.1490	.1515	.1537	.1555	.1571	.1584	.1594	.1601	.1605	.1606
7	.1086	.1125	.1163	.1200	.1234	.1267	.1298	.1326	.1353	.1377
8	.0692	.0731	.0771	.0810	.0849	.0887	.0925	.0962	.0998	.1033
9	.0392	.0423	.0454	.0486	.0519	.0552	.0586	.0620	.0654	.0688
10	.0200	.0220	.0241	.0262	.0285	.0309	.0334	.0359	.0386	.0413
11	.0093	.0104	.0116	.0129	.0143	.0157	.0173	.0190	.0207	.0225
12	.0039	.0045	.0051	.0058	.0065	.0073	.0082	.0092	.0102	.0113
13	.0015	.0018	.0021	.0024	.0028	.0032	.0036	.0041	.0046	.0052
14	.0006	.0007	.0008	.0009	.0011	.0013	.0015	.0017	.0019	.0022
15	.0002	.0002	.0003	.0003	.0004	.0005	.0006	.0007	.0008	.0009
16	.0001	.0001	.0001	.0001	.0001	.0002	.0002	.0002	.0003	.0003
17	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0001

						λ				
x	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6 8	6.9	7.0
0	.0022	.0020	.0018	.0017	.0015	.0014	.0012	.0011	.0010	.0009
1	.0137	.0126	.0116	.0106	.0098	.0090	.0082	.0076	.0070	.0064
2	.0417	.0390	.0364	.0340	.0318	.0296	.0276	.0258	.0240	.0223
3	.0848	.0806	.0765	.0726	.0688	.0652	.0617	.0584	.0552	.0521
4	.1294	.1249	.1205	.1162	.1118	.1076	.1034	.0992	.0952	.0912
5	.1579	.1549	.1519	.1487	.1454	.1420	.1385	.1349	.1314	.1277
6	.1605	.1601	.1595	.1586	.1575	.1562	.1546	.1529	.1511	.1490
7	.1399	.1418	.1435	.1450	.1462	.1472	.1480	.1486	.1489	.1490
8	.1066	.1099	.1130	.1160	.1188	.1215	.1240	.1263	.1284	.1304
9	.0723	.0757	.0791	.0825	.0858	.0891	.0923	.0954	.0985	.1014
10	.0441	.0469	.0498	.0528	.0558	.0588	.0618	.0649	.0679	.0710
11	.0245	.0265	.0285	.0307	.0330	.0353	.0377	.0401	.0426	.0452
12	.0124	.0137	.0150	.0164	.0179	.0194	.0210	.0227	.0245	.0264
13	.0058	.0065	.0073	.0081	.0089	.0098	.0108	.0119	.0130	.0142
14	.0025	.0029	.0033	.0037	.0041	.0046	.0052	.0058	.0064	.0071
15	.0010	.0012	.0014	.0016	.0018	.0020	.0023	.0026	.0029	.0033
16	.0004	.0005	.0005	.0006	.0007	.0008	.0010	.0011	.0013	.0014
17	.0001	.0002	.0002	.0002	.0003	.0003	.0004	.0004	.0005	.0006
18	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002	.0002
19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΠΙΝΑΚΑΣ ΙΙ: Κατανομή Poisson (Συνέχεια)

						λ				
x	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	8.0
0	.0008	.0007	.0007	.0006	.0006	.0005	.0005	.0004	.0004	.0003
1	.0059	.0054	.0049	.0045	.0041	.0038	.0035	.0032	.0029	.0027
2	.0208	.0194	.0180	.0167	.0156	.0145	.0134	.0125	.0116	.0107
3	.0492	.0464	.0438	.0413	.0389	.0366	.0345	.0324	.0305	.0286
4	.0874	.0836	.0799	.0764	.0729	.0696	.0663	.0632	.0602	.0573
5	.1241	.1204	.1167	.1130	.1094	.1057	.1021	.0986	.0951	.0916
6	.1468	.1445	.1420	.1394	.1367	.1339	.1311	.1282	.1252	.1221
7	.1489	.1486	.1481	.1474	.1465	.1454	.1442	.1428	.1413	.1396
8	.1321	.1337	.1351	.1363	.1373	.1382	.1388	.1392	.1395	.1396
9	.1042	.1070	.1096	.1121	.1144	.1167	.1187	.1207	.1224	.1241
10	.0740	.0770	.0800	.0829	.0858	.0887	.0914	.0941	.0967	.0993
11	.0478	.0504	.0531	.0558	.0585	.0613	.0640	.0667	.0695	.0722
12	.0283	.0303	.0323	.0344	.0366	.0388	.0411	.0434	.0457	.0481
13	.0154	.0168	.0181	.0196	.0211	.0227	.0243	.0260	.0278	.0296
14	.0078	.0086	.0095	.0104	.0113	.0123	.0134	.0145	.0157	.0169
15	.0037	.0041	.0046	.0051	.0057	.0062	.0069	.0075	.0083	.0090
16	.0016	.0019	.0021	.0024	.0026	.0030	.0033	.0037	.0041	.0045
17	.0007	.0008	.0009	.0010	.0012	.0013	.0015	.0017	.0019	.0021
18	.0003	.0003	.0004	.0004	.0005	.0006	.0006	.0007	.0008	.0009
19	.0001	.0001	.0001	.0002	.0002	.0002	.0003	.0003	.0003	.0004
20	.0000	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002
21	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001

						λ				
x	8.1	8.2	8.3	8.4	8.5	8.6	8.7	8.8	8.9	9.0
0	.0003	.0003	.0002	.0002	.0002	.0002	.0002	.0002	.0001	.0001
1	.0025	.0023	.0021	.0019	.0017	.0016	.0014	.0013	.0012	.0011
2	.0100	.0092	.0086	.0079	.0074	.0068	.0063	.0058	.0054	.0050
3	.0269	.0252	.0237	.0222	.0208	.0195	.0183	.0171	.0160	.0150
4	.0544	.0517	.0491	.0466	.0443	.0420	.0398	.0377	.0357	.0337
5	.0882	.0849	.0816	.0784	.0752	.0722	.0692	.0663	.0635	.0607
6	.1191	.1160	.1128	.1097	.1066	.1034	.1003	.0972	.0941	.0911
7	.1378	.1358	.1338	.1317	.1294	.1271	.1247	.1222	.1197	.1171
8	.1395	.1392	.1388	.1382	.1375	.1366	.1356	.1344	.1332	.1318
9	.1256	.1269	.1280	.1290	.1299	.1306	.1311	.1315	.1317	.1318
10	.1017	.1040	.1063	.1084	.1104	.1123	.1140	.1157	.1172	.1186
11	.0749	.0776	.0802	.0828	.0853	.0878	.0902	.0925	.0948	.0970
12	.0505	.0530	.0555	.0579	.0604	.0629	.0654	.0679	.0703	.0728
13	.0315	.0334	.0354	.0374	.0395	.0416	.0438	.0459	.0481	.0504
14	.0182	.0196	.0210	.0225	.0240	.0256	.0272	.0289	.0306	.0324
15	.0098	.0107	.0116	.0126	.0136	.0147	.0158	.0169	.0182	.0194
16	.0050	.0055	.0060	.0066	.0072	.0079	.0086	.0093	.0101	.0109
17	.0024	.0026	.0029	.0033	.0036	.0040	.0044	.0048	.0053	.0058
18	.0011	.0012	.0014	.0015	.0017	.0019	.0021	.0024	.0026	.0029
19	.0005	.0005	.0006	.0007	.0008	.0009	.0010	.0011	.0012	.0014
20	.0002	.0002	.0002	.0003	.0003	.0004	.0004	.0005	.0005	.0006
21	.0001	.0001	.0001	.0001	.0001	.0002	.0002	.0002	.0002	.0003
22	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0001	.0001

						λ				
x	9.1	9.2	9.3	9.4	9.5	9.6	9.7	9.8	9.9	10
0	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0000
1	.0010	.0009	.0009	.0008	.0007	.0007	.0006	.0005	.0005	.0005
2	.0046	.0043	.0040	.0037	.0034	.0031	.0029	.0027	.0025	.0023
3	.0140	.0131	.0123	.0115	.0107	.0100	.0093	.0087	.0081	.0076
4	.0319	.0302	.0285	.0269	.0254	.0240	.0226	.0213	.0201	.0189
5	.0581	.0555	.0530	.0506	.0483	.0460	.0439	.0418	.0398	.0378
6	.0881	.0851	.0822	.0793	.0764	.0736	.0709	.0682	.0656	.0631
7	.1145	.1118	.1091	.1064	.1037	.1010	.0982	.0955	.0928	.0901
8	.1302	.1286	.1269	.1251	.1232	.1212	.1191	.1170	.1148	.1126
9	.1317	.1315	.1311	.1306	.1300	.1293	.1284	.1274	.1263	.1251

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

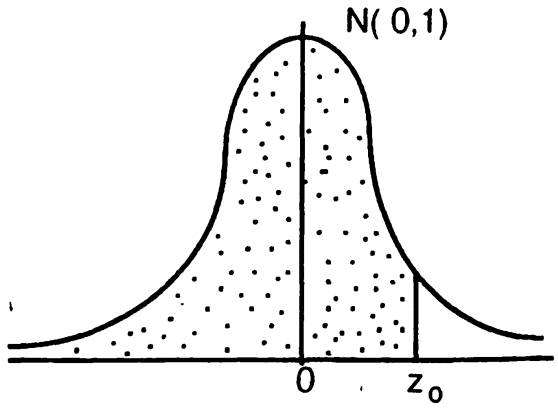
ΠΙΝΑΚΑΣ ΙΙ: Κατανομή Poisson (Συνέχεια)

						λ				
x	9.1	9.2	9.3	9.4	9.5	9.6	9.7	9.8	9.9	10
10	.1198	.1210	.1219	.1228	.1235	.1241	.1245	.1249	.1250	.1251
11	.0991	.1012	.1031	.1049	.1067	.1083	.1098	.1112	.1125	.1137
12	.0752	.0776	.0799	.0822	.0844	.0866	.0888	.0908	.0928	.0948
13	.0526	.0549	.0572	.0594	.0617	.0640	.0662	.0685	.0707	.0729
14	.0342	.0361	.0380	.0399	.0419	.0439	.0459	.0479	.0500	.0521
15	.0208	.0221	.0235	.0250	.0265	.0281	.0297	.0313	.0330	.0347
16	.0118	.0127	.0137	.0147	.0157	.0168	.0180	.0192	.0204	.0217
17	.0063	.0069	.0075	.0081	.0088	.0095	.0103	.0111	.0119	.0128
18	.0032	.0035	.0039	.0042	.0046	.0051	.0055	.0060	.0065	.0071
19	.0015	.0017	.0019	.0021	.0023	.0026	.0028	.0031	.0034	.0037
20	.0007	.0008	.0009	.0010	.0011	.0012	.0014	.0015	.0017	.0019
21	.0003	.0003	.0004	.0004	.0005	.0006	.0006	.0007	.0008	.0009
22	.0001	.0001	.0002	.0002	.0002	.0002	.0003	.0003	.0004	.0004
23	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002
24	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001

						λ				
x	11	12	13	14	15	16	17	18	19	20
0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1	.0002	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2	.0010	.0004	.0002	.0001	.0000	.0000	.0000	.0000	.0000	.0000
3	.0037	.0018	.0008	.0004	.0002	.0001	.0000	.0000	.0000	.0000
4	.0102	.0053	.0027	.0013	.0006	.0003	.0001	.0001	.0000	.0000
5	.0224	.0127	.0070	.0037	.0019	.0010	.0005	.0002	.0001	.0001
6	.0411	.0255	.0152	.0087	.0048	.0026	.0014	.0007	.0004	.0002
7	.0646	.0437	.0281	.0174	.0104	.0060	.0034	.0018	.0010	.0005
8	.0888	.0655	.0457	.0304	.0194	.0120	.0072	.0042	.0024	.0013
9	.1085	.0874	.0661	.0473	.0324	.0213	.0135	.0083	.0050	.0029
10	.1194	.1048	.0859	.0663	.0486	.0341	.0230	.0150	.0095	.0058
11	.1194	.1144	.1015	.0844	.0663	.0496	.0355	.0245	.0164	.0106
12	.1094	.1144	.1099	.0984	.0829	.0661	.0504	.0368	.0259	.0176
13	.0926	.1056	.1099	.1060	.0956	.0814	.0658	.0509	.0378	.0271
14	.0728	.0905	.1021	.1060	.1024	.0930	.0800	.0655	.0514	.0387
15	.0534	.0724	.0885	.0989	.1024	.0992	.0906	.0786	.0650	.0516
16	.0367	.0543	.0719	.0866	.0960	.0992	.0963	.0884	.0772	.0646
17	.0237	.0383	.0550	.0713	.0847	.0934	.0963	.0936	.0863	.0760
18	.0145	.0256	.0397	.0554	.0706	.0830	.0909	.0936	.0911	.0844
19	.0084	.0161	.0272	.0409	.0557	.0699	.0814	.0887	.0911	.0888
20	.0046	.0097	.0177	.0286	.0418	.0559	.0692	.0798	.0866	.0888
21	.0024	.0055	.0109	.0191	.0299	.0426	.0560	.0684	.0783	.0846
22	.0012	.0030	.0065	.0121	.0204	.0310	.0433	.0560	.0676	.0769
23	.0006	.0016	.0037	.0074	.0133	.0216	.0320	.0438	.0559	.0669
24	.0003	.0008	.0020	.0043	.0083	.0144	.0226	.0328	.0442	.0557
25	.0001	.0004	.0010	.0024	.0050	.0092	.0154	.0237	.0336	.0446
26	.0000	.0002	.0005	.0013	.0029	.0057	.0101	.0164	.0246	.0343
27	.0000	.0001	.0002	.0007	.0016	.0034	.0063	.0109	.0173	.0254
28	.0000	.0000	.0001	.0003	.0009	.0019	.0038	.0070	.0117	.0181
29	.0000	.0000	.0001	.0002	.0004	.0011	.0023	.0044	.0077	.0125
30	.0000	.0000	.0000	.0001	.0002	.0006	.0013	.0026	.0049	.0083
31	.0000	.0000	.0000	.0000	.0001	.0003	.0007	.0015	.0030	.0054
32	.0000	.0000	.0000	.0000	.0001	.0001	.0004	.0009	.0018	.0034
33	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0005	.0010	.0020
34	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0006	.0012
35	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0007
36	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0004
37	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002
38	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
39	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΠΙΝΑΚΑΣ ΙΙΙ: Τυπική Κανονική Κατανομή

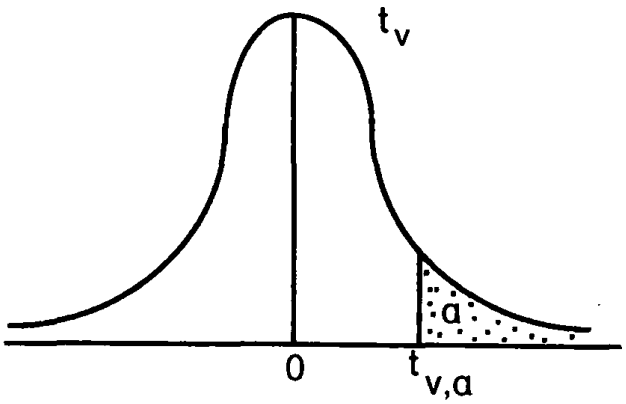


Ο πίνακας δίνει τις πιθανότητες: $P(-\infty < Z \leq z_0)$, $Z \sim N(0,1)$

Z_0	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΠΙΝΑΚΑΣ IV: Κατανομή t



Ο πίνακας δίνει τις τιμές $t_{v,\alpha}$ με $P(t_v \geq t_{v,\alpha}) = \alpha$.

v	α=.10	α=.05	α=.025	α=.01	α=.005	v
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
inf.	1.282	1.645	1.960	2.326	2.576	inf.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ ΒΙΒΛΙΟΘΗΚΗ

ΑΠΑΝΤΗΣΕΙΣ ΣΤΙΣ ΑΣΚΗΣΕΙΣ

ΚΕΦΑΛΑΙΟ 2 & 3:

2-3.1: $\bar{X} = 13.167$, $S^2 = 9.186$, M = 13, K = 12.5, C.V = 23.018%,
$x_{25} = 10.9$, $x_{50} = 13$, $x_{75} = 15.33$.

2-3.2: $\bar{X} = 54.5$, $S^2 = 226.95$, M = 55.5, K = 54.5, C.V = 27.642%
$x_{25} = 44$, $x_{50} = 55.5$, $x_{75} = 66.167$.

2-3.4: $\bar{X} = 71$, $S^2 = 245.9$, M = 72.5, K = 71.5, C.V = 22.086%,
$x_{25} = 58.5$, $x_{50} = 72.5$, $x_{75} = 84$.

2-3.6: $\bar{X} = 146$, $S^2 = 173.077$, M = 145.42, K = 143.5, C.V. = 9.01%
$x_{25} = 137.07$, $x_{50} = 145.42$, $x_{75} = 155.67$.

ΚΕΦΑΛΑΙΟ 4:

4.1: i) 7/16, ii) 6/16, iii) 3/16, iv) 13/16.

4.2: i) 30/75, ii) 21/75, iii) 24/54.

4.3: α) i) 50/100, ii) 40/100, iii) 40/60, iv) 30/50.
β) Εξαρτημένα.

4.4: Ναι.

ΚΕΦΑΛΑΙΟ 5:

5.1: i) 0.0031, ii) 0.0038, iii) 0.0001, iv) 0.5446.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

5.2: i) 0.8965, ii) 0.0879.

5.3: i) 0.1044, ii) 0.0318, iii) 0.0067, iv) 0.4405.

5.4: i) 0.905, ii) 0.1587, iii) 0.1587, iv) 0.2514, v) 0.9888.

5.5: i) 68.26%, ii) 13.59%, iii) 6.06%.

5.6.: α) 3.83%, β) 0.0289.

5.7: Αν X το κέρδος τότε $p_X(x) = \begin{cases} 3/8, & x= 0,2 \\ 1/4, & x= -1 \end{cases}$

και E(X) = 0.5, Var(X) = 1.5.

ΚΕΦΑΛΑΙΟ 6 & 7

6-7.1: α) i) (165.41, 169.79), ii) H_o: μ = 165, H_a: μ> 165, Απορ. H_o

β) i) (165.74, 169.46), ii) H_o: μ = 165, H_a: μ > 165, Απορ. H_o.

6-7.2: α) i) $\bar{X} = 7$, ii) (6.18, 7.82), iii) H_o: μ = 6.5, H_a: μ ≠ 6.5
Αποδ. H_o.

β) i) $\bar{X} = 7$, ii) (6.04, 7.96), iii) H_o : μ = 6.5, H_a: μ ≠ 6.5
Αποδ. H_o .

6-7.3: H_o: μ = 180, H_a: μ > 180, Αποδοχή H_o .

6-7.4: α) i) $\bar{X}_1 - \bar{X}_2 = 3$, ii) (-2.33, 8.33), iii) H_o: $\mu_1=\mu_2$, H_a: $\mu_1 > \mu_2$,
Αποδ. H_o.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

β) i) $\bar{X}_1 - \bar{X}_2 = 3$, iii) (-1.698, 7.698), iii) $H_o: \mu_1 = \mu_2$, $H_a: \mu_1 > \mu_2$, Απορ. H_o.

6-7.5: $H_o: \mu = 0$, $H_a: \mu > 0$ (δίαιτα αποτελεσματική), Απορ. H_o.
$H_o: \mu = 3$, $H_a: \mu \neq 3$, Αποδ. H_o δηλ. η δίαιτα προσφέρει μείωση βάρους κατά 3 κιλά.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

ΒΙΒΛΙΟΓΡΑΦΙΑ

Α. ΕΛΛΗΝΙΚΗ

Κάκουλλος, Θ. (1972). *Στατιστική - Θεωρία και Εφαρμογαί*, Αθήνα.

Κουνιάς, Σ., Καλύβα-Μαχαίρα, Φ., Μπαγιάτης, Κ. και Μπόρα-Σέντα, Ε. (1985). *Εισαγωγή στη Στατιστική*, Θεσσαλονίκη.

Λαμπράκης, Δ. (1983). *Στοιχεία Στατιστικής*, Αθήνα.

Λουκάς, Σ. (1989). *Μια Εισαγωγή στη Στατιστική*. Ιωάννινα.

Παπαϊωάννου, Τ. και Λουκάς, Σ. (1990). *Εισαγωγή στη Στατιστική*. Ιωάννινα.

Παπαϊωάννου, Τ. και Φερεντίνος, Κ. (1981). *Βιομαθηματικά*. Ιωάννινα.

Παπαναστασίου, Κ. (1977). *Η Στατιστική εις την Εκπαίδευσιν*. Λευκωσία.

Τσάντα, Ν. (1988). *Μαθήματα Θεωρίας Πιθανοτήτων και Στατιστικής*. Θεσσαλονίκη.

Β. ΞΕΝΗ

Armitage, P. and Berry, G. (1987). *Statistical methods in medical research*, Second Edition. Blackwell Scientific Publications.

Daniel, W., W. (1974). *Biostatistics: A foundation for analysis in the health sciences*. John Wiley and Sons, Inc.

Ferguson, G. A. (1966). *Statistical analysis in Psychology and Education*. Second Edition. McGraw-Hill.

Freund, J. (1971). *Mathematical Statistics*, Second Edition, Prentice- Hall, Inc. New Jersey.

Koosis, D. (1985). *Statistics - A self teaching guide*, Third Edition, John Wiley and Sons.

Leaverton, P. E. (1983). *Στοιχεία Βιοστατιστικής - Ένα πρόγραμμα αυτοδιδασκαλίας*. Μετάφραση - Επιμέλεια: Παπαϊωάννου, Τ. και Φερεντίνος, Κ. Ιατρικές Εκδόσεις Λίτσας. Αθήνα.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Mood, A., Graybill, F. and Boes, D. (1974). *Introduction to the theory of statistics*. Third Edition. McGraw-Hill.

Mueller, J. H., Schuessler, K. F. and Costner, H. L. (1961). *Statistical reasoning in sociology*. Second Edition. Houghton Mifflin Company.

Noether, G. E. (1991). *Introduction to Statistics. The nonparametric way*. Springer - Verlag.

Phillips, D. S. (1978). *Basic Statistics for Health Science Students*. W. H. Freeman and Company. San Francisco.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ

Τυπώθηκε στο Πανεπιστημιακό Τυπογραφείο με δαπάνη του Πανεπιστημίου Ιωαννίνων

Κ.Α. Πανεπιστημιακού Τυπογραφείου. .....Ε

ΔΙΑΝΕΜΕΤΑΙ ΔΩΡΕΑΝ στους φοιτητές.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΗ ΒΙΒΛΙΟΘΗΚΗ ΙΩΑΝΝΙΝΩΝ