ΔΗΜΗΤΡΗΣ ΑΡ. ΒΑΪΟΠΟΥΛΟΣ
Διδάκτωρ των Μαθηματικών
Καθηγητής Πανεπιστημίου Αθηνών

ΓΕΩΡΓΙΟΣ ΑΙΜ. ΣΚΙΑΝΗΣ
Διδάκτωρ Γεωλογίας
Επ. Καθηγητής Πανεπιστημίου Αθηνών

ΕΓΧΕΙΡΙΔΙΟ
ΣΤΑΤΙΣΤΙΚΗΣ - ΓΕΩΣΤΑΤΙΣΤΙΚΗΣ

(Βασισμένο αποκλειστικά: τα Κεφάλαια της Στατιστικής στο Βιβλίο «**Εισαγωγή στην Πληροφορική**» του Δ.Α.Βαϊόπουλου και το Κεφάλαιο της Γεωστατιστικής στα σχέδια μαθημάτων του Καθηγητή Δ.Α.Βαϊόπουλου και του Επ.Καθηγητή Γ. Σκιάνη)

- Γενικές Έννοιες Πιθανοτήτων και Στατιστικής
- Παράμετροι Κατανομής
- Στατιστικές Εκτιμήσεις – Διαστήματα Εμπιστοσύνης
- Ανάλυση Διασποράς
- Φίλτρα Επεξεργασίας Χρονοσειρών και Καμπυλών
- Ανάλυση Fourier
- Μέθοδος Ελαχίστων Τετραγώνων και Θεωρία Σφαλμάτων
- Γεωστατιστική
- Ανάλυση Χωρικών Προτύπων
- Μέθοδοι Χωρικής Παρεμβολής
- Χωρική Παρεμβολή με Kriging

ΑΘΗΝΑ 2007

Τα συμπεράσματα έχουν αξία και ισχύ, όταν τεκμηριώνονται.

Οι αποφάσεις είναι σωστές, όταν προκύπτουν από αποδεικτικά στοιχεία.

Η Στατιστική είναι το εργαλείο που περιβάλλει με επιστημονικό κύρος συμπεράσματα και αποφάσεις που αφορούν σε κάθε τομέα (επιστημονικό, επαγγελματικό, κ.λπ.) της ανθρώπινης δραστηριότητας.

ΠΕΡΙΕΧΟΜΕΝΑ

Σελίδα

ΠΡΟΛΟΓΟΣ

Ζούμε σε μια εποχή που η πληροφορία κατακλύζει τη ζωή μας. Η αθρόα παραγωγή γνώσης σε όλα τα πεδία των δραστηριοτήτων μας σε συνδυασμό με την αύξηση του πλήθους και των δυνατοτήτων των φορέων της πληροφορίας, καθώς και η δυνατότητα ταχύτατης μετάδοσής της σε ένα ραγδαία αυξανόμενο πλήθος αποδεκτών, που μπορεί να βρίσκονται σε κάθε γωνιά του πλανήτη μας, οδηγεί σε αυτό που ονομάζουμε πληροφοριακή έκρηξη.

Όμως, οι τεράστιοι όγκοι δεδομένων που αφορούν στους διάφορους επιστημονικούς κλάδους, αλλά και σε άλλες δραστηριότητες, επιχειρηματικές, πολιτικές, πολιτιστικές, καλλιτεχνικές, κ.λπ., για έχουν παραγωγική αξία χρήζουν αξιολόγησης, επεξεργασίας, αξιοποίησης.

Ευτυχώς, σήμερα, έχουμε στη διάθεσή μας τον Η/Υ, αλλά και επιστημονικές μεθοδολογίες που μας βοηθούν να αναδείξουμε και να παρουσιάσουμε την πληροφορία με αποτελεσματικό, αξιόπιστο, τεκμηριωμένο και δόκιμο τρόπο.

Ο Η/Υ, ημέρα με την ημέρα, αναδεικνύεται σε ένα ισχυρότατο και πολυδύναμο εργαλείο στο οποίο η μηχανολογική πολυπλοκότητα έχει αντικατασταθεί από την πολυπλοκότητα του λογισμικού.

Έτσι, ο επιστήμονας κάθε επιστημονικού κλάδου έχει στη διάθεσή του τον Η/Υ, που τον βοηθάει όχι μόνο να τεκμηριώσει καλύτερα και πειστικότερα τη νέα γνώση, αλλά και να προχωρήσει ταχύτερα και αποδοτικότερα και σε νέα πεδία επιστημονικής έρευνας. Μάλιστα, μερικά πεδία, όπως μαθηματικές μοντελοποιήσεις και προσομοιώσεις φυσικών συστημάτων, θα ήταν εντελώς ανέφικτο να προσεγγιστούν ή τουλάχιστον δεν θα είχαν πρακτική αξία, χωρίς τη βοήθεια του Η/Υ.

Όμως, για να δοθούν απαντήσεις, σήμερα, σε πολλά επιστημονικά, και όχι μόνο, προβλήματα απαιτείται το θεωρητικό υπόβαθρο της μαθηματικής Στατιστικής, καθώς και γνώσεις των αριθμητικών μεθόδων που με τη βοήθεια των Η/Υ υλοποιούν χρονοβόρες και δύσκολες, τις περισσότερες φορές, λύσεις, ύστερα από επεξεργασία ενός πολύ μεγάλου όγκου δεδομένων. Η επεξεργασία όμως ενός πολύ μεγάλου όγκου δεδομένων είναι, αν όχι αδύνατη, πολύ δύσκολη να πραγματοποιηθεί χωρίς τον Η/Υ. Έτσι, η Στατιστική, σήμερα, σχεδόν απαραίτητα, συνυπάρχει με τον Η/Υ και δεν είναι

καθόλου υπερβολή αν πούμε, ότι χωρίς τον Η/Υ είναι αδιανόητη στις ημέρες μας. Αναμφίβολα, βοηθάει στην καλύτερη εκτίμηση των προς ανάλυση και επεξεργασία στοιχείων και η χρησιμοποίησή της οδηγεί στην ασφαλέστερη εξαγωγή της πληροφορίας.

Τα τελευταία πενήντα χρόνια, η Στατιστική Επιστήμη έχει γνωρίσει μεγάλη ανάπτυξη και σήμερα δεν υπάρχει τομέας επιστημονικός ή επαγγελματικός να μην τη χρησιμοποιεί.

Στο βοήθημά μας αυτό, που απευθύνεται στους φοιτητές του Μεταπτυχιακού Ενδεικτικού Ωκεανογραφίας, στους φοιτητές του Μεταπτυχιακού Προγράμματος της Εφαρμοσμένης Περιβαλλοντικής Γεωλογίας και στους φοιτητές του Μεταπτυχιακού Προγράμματος Πρόβλεψης και Διαχείρισης Φυσικών Καταστροφών, αναφερόμαστε σε γενικά θέματα της Στατιστικής και Γεωστατιστικής, αλλά και σε στοιχεία της θεωρίας των πιθανοτήτων, επειδή αυτή αποτελεί το θεωρητικό υπόβαθρο της Στατιστικής, τα οποία πιστεύουμε ότι θα βοηθήσουν στην καλύτερη αξιοποίηση και την πληρέστερη κατανόηση των διαφόρων λογισμικών Στατιστικής που, ενδεχομένως, θα χρησιμοποιήσουν για να χειριστούν και να παρουσιάσουν τα δεδομένα, αλλά και για να εμφανίσουν σε καλύτερη μορφή τα αποτελέσματα της έρευνάς τους και, τέλος, να τεκμηριώσουν με πρόσφορο τρόπο τα επιστημονικά συμπεράσματά τους.

Ευνόητο είναι ότι ένα διδακτικό βοήθημα, όταν απευθύνεται σε φοιτητές μεταπτυχιακού κύκλου σπουδών, δεν στοχεύει να καλύψει πλήρως το υπόβαθρο για το γνωστικό αντικείμενο που πραγματεύεται. Όμως, μπορεί να αποβεί χρήσιμο, αν συντελέσει στο να κεντρίσει το ενδιαφέρον των νέων επιστημόνων για περαιτέρω μελέτη και έρευνα και ενισχύσει την έφεση για διείσδυση σε βαθύτερα πεδία της επιστήμης.

Ευχαριστώ τον Επιστημονικό Συνεργάτη του Εργαστηρίου Τηλεανίχνευσης Δρα Κων/νο Νικολακόπουλο, για την προθυμία του να με βοηθήσει στη διαμόρφωση του κειμένου.

Αθήνα, Νοέμβριος 2007 *Δημήτρης Αρ. Βαϊόπουλος*

1. ΣΤΑΤΙΣΤΙΚΗ - ΕΙΣΑΓΩΓΗ - ΓΕΝΙΚΕΣ ΕΝΝΟΙΕΣ

Η Στατιστική Επιστήμη αναπτύχθηκε τις τελευταίες δεκαετίες, όταν άρχισε να χρησιμοποιεί μαθηματικές μεθόδους ανάλυσης στατιστικών δεδομένων, με σκοπό να εξαχθούν πιο τεκμηριωμένα συμπεράσματα από τις ερευνητικές εργασίες στις φυσικές και λοιπές επιστήμες, καθώς και στον τομέα της τεχνολογίας.

Τα τελευταία χρόνια, που η χρήση των ΗΥ διευρύνθηκε με ταχύ ρυθμό σε κάθε τομέα της ανθρώπινης δραστηριότητας, η Στατιστική Επιστήμη γίνεται καθημερινά ολοένα και περισσότερο προσιτή σε περισσότερους επιστήμονες κάθε επιστημονικού κλάδου. Σε αυτό συμβάλλει η ανάπτυξη της πληθώρας λογισμικού (software) με στατιστικό περιεχόμενο.

Σήμερα, κυκλοφορούν στην αγορά ισχυρά στατιστικά πακέτα, π.χ., το SPSS (Statistical Package for the Social Sciences), Statgraphics, Statistika, Primer, κ.λπ., αλλά και Λογιστικά Φύλλα (Spreadsheets) με στατιστικές εφαρμογές, όπως το LOTUS, το EXCEL, κ.λπ., που η εκμάθησή τους είναι σχετικά εύκολη και η χρήση τους βοηθά τον κάθε επιστήμονα στην έρευνά του.

Δεν χρειάζεται να εμβαθύνει κανένας στις έννοιες του Λογισμού των Πιθανοτήτων και της Στατιστικής για να αξιολογήσει τα πειράματά του και τις μετρήσεις του, ώστε να εξαγάγει τεκμηριωμένα συμπεράσματα από την επιστημονική του έρευνα. Αρκεί, να χρησιμοποιήσει σωστά το κατάλληλο λογισμικό για τη Στατιστική και με τον ΗΥ θα πάρει γρήγορα και στην επιθυμητή μορφή (διαγράμματα, πίνακες, κ.λπ.) τα αποτελέσματά του.

Τα πλεονεκτήματα από τη χρήση τέτοιων στατιστικών πακέτων τα καταλαβαίνει ο χρήστης, αμέσως μόλις εξοικειωθεί λίγο με αυτά.

Στα παρακάτω θα αναφερθούμε με λίγα λόγια σε μερικές βασικές έννοιες της Στατιστικής.

1.1. Πληθυσμός και δείγμα

'Ενα σύνολο από άτομα ή αντικείμενα, ή ένα πλήθος παρατηρήσεων ή πειραμάτων που γίνονται κάτω από σταθερές συνθήκες, στη γλώσσα της Στατιστικής ονομάζεται **πληθυσμός.**

Κάθε τώρα άτομο ή αντικείμενο, ή, αντίστοιχα, κάθε παρατήρηση ή πείραμα είναι ένα **στοιχείο** ή **μέλος**, όπως λέγεται, του αντίστοιχου πληθυσμού. Κάθε στοιχείο (μέλος) ενός πληθυσμού είναι δυνατό να διερευνηθεί για περισσότερα από ένα χαρακτηριστικά, που θεωρούνται ως **τυχαίες μεταβλητές X, Y, Z,....** Στην παράγραφο 23.6 αναφέρεται διεξοδικότερα η έννοια της τυχαίας μεταβλητής.

Συχνά, στην καθημερινή πρακτική, θέλουμε να βγάλουμε συμπεράσματα για ένα σύνολο στοιχείων. Αυτό τις περισσότερες φορές είναι ανέφικτο ή τουλάχιστον πολύ επίπονο, χρονοβόρο και δαπανηρό.

Έτσι, κατά τη στατιστική διερεύνηση, αντί να μελετήσουμε το σύνολο των στοιχείων, δηλαδή ολόκληρο τον πληθυσμό, αρκούμαστε στην εξέταση ενός υποσυνόλου του πληθυσμού. Το υποσύνολο του πληθυσμού, που συνήθως αποτελεί ένα μικρό μέρος του, ονομάζεται **δείγμα** και συμβολίζεται με **n**.

Μέγεθος του πληθυσμού είναι το πλήθος των στοιχείων του που μπορεί να είναι πεπερασμένο ή άπειρο και συμβολίζεται με **N**.

ΠΑΡΑΔΕΙΓΜΑ 1. Εάν η τυχαία μεταβλητή X είναι το ύψος των οκτάχρονων παιδιών, τότε όλα τα οκτάχρονα παιδιά αποτελούν τον πληθυσμό. Οι μετρήσεις του ύψους των παιδιών σε διάφορους τόπους αποτελούν το δείγμα και κάθε παιδί (αντίστοιχη μέτρηση) ένα στοιχείο του πληθυσμού.

Το ύψος που αντιστοιχείται στη μεταβλητή X είναι χαρακτηριστικό κάθε στοιχείου του πληθυσμού.

Με τις μεταβλητές Y, Z, **...**, κ.λπ., μπορούμε να αντιστοιχήσουμε άλλα χαρακτηριστικά του πληθυσμού, π.χ., το βάρος, την περίμετρο του κεφαλιού, κ.λπ.

ΠΑΡΑΔΕΙΓΜΑ 2. Θέλουμε να έχουμε μια τεκμηριωμένη άποψη για το ποσοστό των ελαττωματικών λαμπτήρων από την πενθήμερη παραγωγή ενός εργοστασίου, εξετάζοντας 20 τυχαίους λαμπτήρες από την ημερήσια παραγωγή. 'Ετσι, το σύνολο των λαμπτήρων της πενθήμερης παραγωγής αποτελεί τον πληθυσμό, ενώ οι 100 (5 x 20) λαμπτήρες είναι το εξεταζόμενο δείγμα.

1.2. Προετοιμασία σε ένα πείραμα

Κατά την εφαρμογή στατιστικών μεθόδων για να λάβουμε όσο το δυνατό καλύτερα αποτελέσματα για το προς επεξεργασία πρόβλημα, πρέπει να τηρηθεί μια αυστηρή διαδικασία που αφορά στην προετοιμασία και κατάστρωση του σχεδίου εργασίας.

Τα κυριότερα σημεία που πρέπει να ληφθούν υπόψη είναι τα εξής:

- ΤΟ ΥΛΙΚΟ που εξετάζεται πρέπει να είναι **ομογενές**. Αυτό σημαίνει, ότι κατά τη διάρκεια της έρευνας δεν πρέπει να γίνουν αλλαγές στις μεθόδους λήψης του δείγματος και στα όργανα μέτρησης.
- Πρέπει να αποκλείονται σφάλματα και επιδράσεις που οφείλονται σε παράγοντες εκτός του πειράματος. Π.χ., αν χρειάζεται να γίνει σύγκριση δυο συνόλων παρασκευασμάτων πρέπει η προετοιμασία τους να έχει γίνει με την ίδια μεθοδολογία και τα ίδια όργανα.
- Πρέπει να τίθενται προδιαγραφές (τιμές σύγκρισης) για να μπορούν να συγκριθούν τα αποτελέσματα του πειράματος.
- Η εκλογή των δειγμάτων πρέπει να είναι **τυχαία** και, όπου χρειάζεται, **αντιπροσωπευτική**. Τυχαία θεωρείται η εκλογή, όταν κάθε στοιχείο έχει την ίδια πιθανότητα να περιληφθεί στο δείγμα. Η αντιπροσωπευτικότητα απαιτείται, όταν τα στοιχεία του πληθυσμού διαχωρίζονται σε υποσύνολα. Στην περίπτωση αυτή η τυχαία εκλογή στοιχείου από κάθε υποσύνολο του πληθυσμού εξασφαλίζει την

αντιπροσωπευτικότητα και τότε λέμε ότι το δείγμα είναι αντιπροσωπευτικό.

- Πρέπει να επιλέγεται το σωστό δείγμα.

Μεγάλο δείγμα ασφαλώς και οδηγεί σε ακριβέστερα συμπεράσματα για τον όλο πληθυσμό. 'Ομως τότε η έρευνα αποβαίνει χρονοβόρος και δαπανηρή. Γιαυτό πρέπει για την επιλογή του μεγέθους του δείγματος να λαμβάνονται υπόψη οι εξής παράμετροι:

- Η ακρίβεια των συμπερασμάτων.
- Η χρονική διάρκεια της έρευνας.
- Η δαπάνη της έρευνας.

Το βέλτιστο μέγεθος του δείγματος είναι εκείνο που συμβιβάζει τις τρεις παραπάνω παραμέτρους.

1.3. Πειράματα τύχης - Δειγματοχώροι - Γεγονότα

Στην επιστήμη και, ιδιαίτερα, στους πειραματικούς κλάδους της, ως γνωστό, τα πειράματα έχουν μεγάλη σπουδαιότητα. Επαληθεύουν ή διαψεύδουν επιστημονικές υποθέσεις και, φυσικά, ενισχύουν και τεκμηριώνουν επιστημονικές θεωρίες. Πολλές φορές στην ιστορία της επιστήμης αμφισβητούμενες επιστημονικές θεωρίες επικράτησαν ύστερα από επαλήθευσή τους από ακριβή και επιτυχή πειράματα.

Τα πειράματα όμως βασίζονται σε μια θεμελιώδη παραδοχή: Κάθε πείραμα που επαναλαμβάνεται κάτω από τις ίδιες συνθήκες, δίνει τα ίδια αποτελέσματα. Η παραδοχή αυτή, φυσικά, δεν είναι αυθαίρετη, αλλά είναι συνέπεια μιας γενικότερης αρχής, της **αιτιοκρατικής (deterministic)**, σύμφωνα με την οποία : το ίδιο αίτιο προκαλεί το ίδιο αποτέλεσμα.

Υπάρχουν όμως και πειράματα που δεν δίνουν τα ίδια αποτελέσματα, παρόλο που επαναλαμβάνονται κάτω από τις ίδιες συνθήκες. Τα πειράματα αυτά ονομάζονται **πειράματα τύχης**. Εδώ, ίσως αξίζει να δώσουμε ένα γενικό ορισμό της έννοιας τύχη, που μπορούμε να πούμε ότι είναι : σύνολο αστάθμητων και απρόβλεπτων

παραγόντων, που καθορίζουν χαοτικά την πορεία και την έκβαση ενός πειράματος ή ενός γεγονότος.

Παράδειγμα πειράματος τύχης είναι η ρίψη ενός νομίσματος, όπου το αποτέλεσμα θα είναι "**κορώνα**" ή "**γράμματα**", δηλαδή ένα στοιχείο του συνόλου {Κ, Γ} , αν με Κ συμβολίσουμε το στοιχείο "κορώνα" και με Γ το στοιχείο "γράμματα".

Άλλο παράδειγμα πειράματος τύχης είναι η ρίψη ενός ζαριού, όπου το αποτέλεσμα του πειράματος θα είναι να έρθει στην επάνω όψη του ζαριού ένας από τους αριθμούς: 1, 2, 3, 4, 5, 6 που αποτελούν το σύνολο όλων των δυνατών αποτελεσμάτων του εν λόγω πειράματος τύχης, κ.λπ.

'Οταν λοιπόν εκτελούμε ένα πείραμα τύχης, παίρνουμε διάφορα αποτελέσματα.

Το σύνολο **S** όλων των δυνατών αποτελεσμάτων σε ένα πείραμα τύχης ονομάζεται **δειγματοχώρος**.

Ο δειγματοχώρος **S** στη γλώσσα της Στατιστικής ονομάζεται **πληθυσμός**. Ο δειγματοχώρος εάν έχει πεπερασμένο πλήθος στοιχείων ονομάζεται **πεπερασμένος** δειγματοχώρος.

Εάν σε ένα επαναλαμβανόμενο πείραμα τύχης το πλήθος των δυνατών αποτελεσμάτων αυξάνεται απεριόριστα, ο δειγματοχώρος ονομάζεται **άπειρος**.

Κάθε αποτέλεσμα σε ένα επαναλαμβανόμενο πείραμα τύχης αποτελεί ένα **ενδεχόμενο** ή **γεγονός**. Επομένως τα γεγονότα είναι τα δυνατά αποτελέσματα σε ένα πείραμα τύχης και το σύνολό τους αποτελεί το δειγματοχώρο S για το υπόψη πείραμα τύχης.

Τα γεγονότα διακρίνονται σε **βέβαια, αδύνατα** και **τυχαία**.

Βέβαιο ονομάζεται ένα γεγονός όταν, κάτω από ορισμένες συνθήκες, εμφανίζεται πάντοτε.

Αδύνατο ονομάζεται ένα γεγονός, όταν δεν είναι δυνατόν να εμφανιστεί ποτέ.

'Οταν όμως υπάρχουν δυνατότητες ένα γεγονός άλλοτε να εμφανίζεται και άλλοτε να μην εμφανίζεται, τότε το γεγονός ονομάζεται **τυχαίο**.

Εάν σε μια κληρωτίδα τοποθετήσουμε τα γράμματα του Ελληνικού αλφαβήτου και αφαιρώντας από την κληρωτίδα ένα γράμμα ζητήσουμε να είναι αυτό γράμμα του Ελληνικού αλφαβήτου, το γεγονός αυτό είναι **βέβαιο**. Αν όμως, αφαιρώντας από την κληρωτίδα ένα γράμμα, ζητήσουμε αυτό να είναι το γράμμα **w** του Λατινικού αλφαβήτου το γεγονός αυτό, προφανώς, είναι **αδύνατο**. Εάν, τέλος, αφαιρώντας από την κληρωτίδα ένα γράμμα, ζητήσουμε να είναι αυτό φωνήεν, το γεγονός αυτό είναι **τυχαίο**.

Επειδή τα γεγονότα μπορούν να αντιστοιχηθούν με σύνολα, μπορούμε να έχουμε μια άλγεβρα γεγονότων αντίστοιχη με την άλγεβρα των συνόλων.

'Ετσι, το σύνολο **S**, που αποτελεί το δειγματοχώρο ενός πειράματος τύχης, επειδή περιέχει όλα τα δυνατά αποτελέσματα αυτού του πειράματος, δηλαδή πραγματοποιείται (εμφανίζεται) οπωσδήποτε ένα από τα στοιχεία του, είναι **βέβαιο γεγονός**.

Στη θεωρία των συνόλων ο δειγματοχώρος **S** αντιστοιχεί στο **βασικό** σύνολο **Ω**.

Το **κενό σύνολο** $\varnothing$ είναι ένα γεγονός που δεν περιέχει κανένα από τα δυνατά αποτελέσματα ενός πειράματος τύχης και γιαυτό είναι **αδύνατο γεγονός**.

Εάν κάνουμε πράξεις με γεγονότα του συνόλου S, δηλαδή με σύνολα που είναι υποσύνολα του δειγματοχώρου S, τότε παίρνουμε άλλα γεγονότα του S.

Επομένως, εάν A και B είναι δυο γεγονότα, τότε θα έχουμε για την ένωση, την τομή, τη διαφορά, κ.λπ. :

1. $A \cup B$ είναι το γεγονός ή A ή B ή και τα δύο A και B
2. $A \cap B$ είναι το γεγονός και A και B
3. A-B είναι το γεγονός A όχι όμως και B
4. A΄ είναι το γεγονός όχι A

Εάν τώρα δύο γεγονότα Α και Β είναι ξένα, δηλαδή εάν $A \cap B = \varnothing$, τότε λέμε ότι τα γεγονότα αυτά είναι **ασυμβίβαστα**, ήτοι το ένα αποκλείει το άλλο. Αυτό σημαίνει ότι δεν είναι δυνατόν να πραγματοποιηθούν και τα δύο. Δηλαδή, εάν πραγματοποιηθεί το ένα από αυτά το άλλο ασφαλώς δεν μπορεί να πραγματοποιηθεί. Εάν όμως το ένα από τα δύο υπόψη γεγονότα δεν πραγματοποιηθεί, το άλλο δυνατόν να πραγματοποιηθεί.

Δύο τώρα γεγονότα $Γ_1$ και $Γ_2$ για τα οποία ισχύουν: Αν πραγματοποιείται το $Γ_1$ να μη μπορεί να πραγματοποιηθεί το $Γ_2$ και αν δεν πραγματοποιείται το $Γ_1$ να πραγματοποιείται οπωσδήποτε το $Γ_2$ και αν πραγματοποιείται το $Γ_2$ να μη μπορεί να πραγματοποιηθεί το $Γ_1$ και αν δεν πραγματοποιείται το $Γ_2$ να πραγματοποιείται οπωσδήποτε το $Γ_1$, χαρακτηρίζονται ως **αντίθετα** γεγονότα. Δηλαδή, στα αντίθετα γεγονότα η πραγματοποίηση του ενός εκ των δύο αποκλείει την πραγματοποίηση του άλλου και η μη πραγματοποίηση του ενός εκ των δύο συνεπάγεται αναγκαστικά την πραγματοποίηση του άλλου.

1.4. Η Έννοια της Πιθανότητας

Ο κλασικός ορισμός της έννοιας της πιθανότητας προσπαθεί να δώσει μία ποσοτική έννοια του τυχαίου.

Κατ'αρχήν σε κάθε πείραμα τύχης όλες οι δυνατές εκβάσεις, δηλαδή όλα τα δυνατά αποτελέσματα, θεωρούνται εξίσου πιθανά.

Εάν το πλήθος των διαφορετικών δυνατών αποτελεσμάτων σε ένα πείραμα τύχης είναι ν και μ είναι το πλήθος των ευνοϊκών αποτελεσμάτων (που σημαίνει, ότι αν συμβεί ένα από αυτά, πραγματοποιείται το γεγονός Γ), τότε ο λόγος $\frac{\mu}{\nu}$ δίνει την πιθανότητα του γεγονότος Γ που συμβολίζεται με Ρ(Γ), δηλαδή $P(Γ) = \frac{\mu}{\nu}$.

Ο παραπάνω ορισμός της μαθηματικής πιθανότητας ισχύει για την περίπτωση πεπερασμένου πλήθους δυνατών ισοδύναμων περιπτώσεων και αφορά στην "a priori" πιθανότητα.

Από το μαθηματικό ορισμό της πιθανότητας γίνεται φανερό ότι η πιθανότητα P(Γ) εμφάνισης ενός γεγονότος Γ ικανοποιεί τα αξιώματα:

1. $0 \leq P(\Gamma) \leq 1$
2. Η πιθανότητα του βέβαιου γεγονότος είναι $P(\Omega) = \frac{\nu}{\nu} = 1$
3. Η πιθανότητα κάθε άλλου απλού γεγονότος (δηλαδή εμφάνιση μιας ευνοϊκής περίπτωσης) είναι $P(\Gamma_i) = \frac{1}{\nu}$, i=1,2,...,μ όπου $\Gamma=\{\Gamma_1,\Gamma_2,....\Gamma_\mu\}$ είναι το σύνολο των ευνοϊκών περιπτώσεων σε ένα πείραμα τύχης, τα στοιχεία του οποίου αποτελούν τα απλά γεγονότα του εν λόγω πειράματος τύχης.
4. Αν Γ^c είναι το συμπληρωματικό γεγονός του Γ, τότε ισχύει :

$$P(\Gamma^c) = \frac{\nu - \mu}{\nu}$$

και επομένως

$$P(\Gamma)+P(\Gamma^c)=\frac{\mu}{\nu}+\frac{\nu-\mu}{\nu}=1$$

Όταν όμως σε ένα πείραμα τύχης ο αριθμός των δοκιμών ή επαναλήψεων του πειράματος είναι πολύ μεγάλος, τότε καταφεύγουμε στη στατιστική ή εμπειρική ή “a posteriori” (εκ των υστέρων) πιθανότητα που ορίζεται από το λόγο του πλήθους Μ της εμφάνισης του γεγονότος Γ προς τον αρκούντως μεγάλο αριθμό Ν των πειραμάτων, ήτοι $P(\Gamma_i)=\frac{M}{N}$

Παράδειγμα: Σε 1000 ρίψεις του ζαριού ο αριθμός 3 εμφανίστηκε 183 φορές, άρα η εμπειρική ή εκ των υστέρων πιθανότητα πραγματοποίησης του γεγονότος Γ_3={εμφάνιση του αριθμού 3} είναι :

$P(3) = \frac{183}{1000} = 0.183.$

Η μαθηματική ή εκ των προτέρων (a priori) πιθανότητα πραγματοποίησης του ίδιου γεγονότος, όμως είναι :

$$P(3) = \frac{\pi\lambda\eta\theta o\varsigma\ \varepsilon\upsilon\nu o\iota\kappa\omega\nu\ \pi\varepsilon\rho\iota\pi\tau\omega\sigma\varepsilon\omega\nu}{\pi\lambda\eta\theta o\varsigma\ \delta\upsilon\nu\alpha\tau\omega\nu\ \pi\varepsilon\rho\iota\pi\tau\omega\sigma\varepsilon\omega\nu} = \frac{1}{6}$$

ήτοι, P(3)=0.166.

1.5. Η Έννοια της Συχνότητας

Εάν γίνει μια στατιστική έρευνα για την τιμή μιας μεταβλητής X και λάβουν χώρα N παρατηρήσεις, τότε οι τιμές της μεταβλητής X θα είναι :

$$x_1, x_2, \ldots, x_N$$

Έστω τώρα ότι από τις τιμές αυτές v_1 είναι ίσες με x_1, v_2 ίσες με x_2, ..., v_μ ίσες με x_μ. Τότε μπορούμε να καταρτίσουμε τον πίνακα :

x_1	x_2	...	x_μ
v_1	v_2	...	v_μ

Κάθε ένας από τους αριθμούς v_1, v_2, ..., v_μ λέγεται **απόλυτη συχνότητα** ή απλώς συχνότητα εμφάνισης της αντίστοιχης τιμής x_i, i=1, 2, ..., μ και συμβολίζεται με f_i.

Ο αριθμός $N = v_1 + v_2 + ... + v_\mu$ εκφράζει το συνολικό αριθμό των παρατηρήσεων (πληθικός αριθμός) και λέγεται **ολική συχνότητα,** συμβολίζεται δε με :

$$\Sigma f_i,\ i=1, 2, ..., \mu$$

Οι λόγοι v_1/N, v_2/N, ..., v_μ/N λέγονται **σχετικές συχνότητες** των τιμών x_1, x_2, ..., x_μ αντίστοιχα, ενώ οι λόγοι $100 \cdot v_1/N$, $100 \cdot v_2/N$, ..., $100 \cdot v_\mu/N$ εκφράζουν την επι τοις εκατό (%) σχετική συχνότητα.

$$\text{Άρα, σχετική συχνότητα} = \frac{\text{απόλυτη συχνότητα}}{\text{ολική συχνότητα}}$$

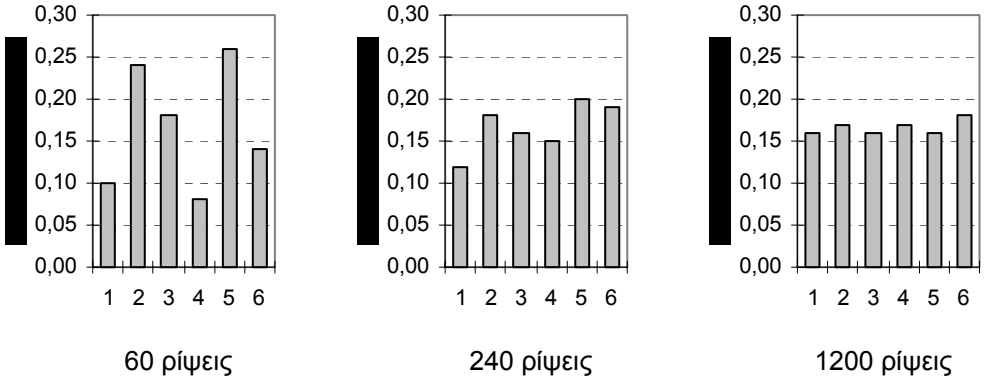
Από τα παραπάνω συνάγεται ότι: το άθροισμα όλων των σχετικών συχνοτήτων μιας στατιστικής έρευνας είναι ίσο με τη μονάδα, πράγματι :

$$\frac{v_1}{N} + \frac{v_2}{N} + \ldots + \frac{v_\mu}{N} = 1$$

Η πείρα δείχνει ότι στα διάφορα πειράματα τύχης είναι πρακτικά βέβαιο ότι η σχετική συχνότητα τείνει προς έναν ορισμένο αριθμό μ για ένα πείραμα τύχης που επαναλαμβάνεται απεριόριστα, τον ίδιο, όσες φορές το πείραμα εκτελείται κάτω από τις ίδιες συνθήκες. Αυτό αποτελεί το αξίωμα για την στατιστική πιθανότητα.

Έτσι, όταν ο αριθμός Ν των πειραμάτων τύχης είναι αρκούντως μεγάλος, τότε η σχετική συχνότητα δηλαδή ο αριθμός Μ εμφανίσεων του γεγονότος Γ μετά από Ν δοκιμές, ήτοι ο λόγος Μ/Ν τείνει στην αριθμητική τιμή της μαθηματικής πιθανότητας.

Στο σχήμα 13 βλέπουμε πώς η σχετική συχνότητα εμφάνισης των αριθμών 1,2,3,4,5,6 κατά τις ρίψεις του ζαριού διαφοροποιείται, όταν αυξάνει ο αριθμός των πειραμάτων (ρίψεων του ζαριού) και μετά από ένα μεγάλο αριθμό ρίψεων τείνει στην τιμή της μαθηματικής πιθανότητας $\frac{1}{6}$=0.166



Σχήμα 13. Σχετική συχνότητα εμφάνισης των αριθμών 1,2,3,4,5,6 κατά τη ρίψη ζαριού

Στη συνέχεια παραθέτουμε μερικά χρήσιμα θεωρήματα σε προβλήματα πιθανοτήτων :

1. Εάν είναι $\Gamma_1 \subseteq \Gamma_2 \rightarrow P(\Gamma_1) \leq P(\Gamma_2)$ και $P(\Gamma_2-\Gamma_1)=P(\Gamma_2)-P(\Gamma_1)$
2. Η πιθανότητα του αδύνατου γεγονότος είναι μηδέν, ήτοι $P(\varnothing)=0$
3. Εάν ένα γεγονός Γ συνίσταται από την ένωση $\Gamma_1,\Gamma_2,\ldots,\Gamma_v$ γεγονότων ασυμβίβαστων ανά δύο, δηλαδή αν $\Gamma=\Gamma_1 \cup \Gamma_2 \cup \ldots \cup \Gamma_v$, τότε ισχύει : $P(\Gamma)=P(\Gamma_1)+P(\Gamma_2)+\ldots+P(\Gamma_v)$ και, φυσικά, αν $\Gamma=\Omega$ όπου Ω είναι ο δειγματικός χώρος ≡ το βέβαιο γεγονός, τότε ισχύει: $P(\Gamma_1)+P(\Gamma_2)+\ldots+P(\Gamma_v)=1$
4. Εάν Γ_1 και Γ_2 είναι δύο οποιαδήποτε γεγονότα τότε ισχύει:

   $P(\Gamma_1 \cup \Gamma_2)=P(\Gamma_1)+P(\Gamma_2)-P(\Gamma_1 \cap \Gamma_2)$. Γενικότερα, αν $\Gamma_1,\Gamma_2,\Gamma_3$ είναι τρία οποιαδήποτε γεγονότα, τότε ισχύει:

   $P(\Gamma_1 \cup \Gamma_2 \cup \Gamma_3)=P(\Gamma_1)+P(\Gamma_2)+P(\Gamma_3)-P(\Gamma_1 \cap \Gamma_2)- P(\Gamma_2 \cap \Gamma_3) - P(\Gamma_3 \cap \Gamma_1)+ P(\Gamma_1 \cap \Gamma_2 \cap \Gamma_3)$.

1.6. Πιθανότητες υπό συνθήκη ή χωρίς συνθήκη

Η πιθανότητα που υπολογίζεται για ενδεχόμενα που εμφανίζονται σε πειράματα τύχης που εκτελούνται κάτω από σταθερές συνθήκες ονομάζεται πιθανότητα χωρίς συνθήκες.

Πολλές φορές όμως για τον υπολογισμό της πιθανότητας πραγματοποίησης ενός γεγονότος Γ πρέπει να ληφθεί υπόψη η συνθήκη ότι πρέπει πριν από το γεγονός Γ να προϋπάρξει η πραγματοποίηση του γεγονότος Α με μια ορισμένη πιθανότητα P(A). Η πιθανότητα αυτή ονομάζεται πιθανότητα υπό συνθήκη και συμβολίζεται με P(Γ/A)

Έτσι, αν θεωρήσουμε δύο γεγονότα Α και Γ με P(A)>0, τότε η πιθανότητα P(Γ/A), να συμβεί το γεγονός Γ με την προϋπόθεση (υπό συνθήκη) ότι έχει συμβεί το Α είναι ίση με

$$P(\Gamma/A) \equiv \frac{P(A \cap \Gamma)}{P(A)} \quad \text{ή}$$

$$P(A \cap \Gamma) \equiv P(A) \cdot P(\Gamma/A)$$

Γεγονότα Ανεξάρτητα

Δύο γεγονότα Γ και Α λέγονται ανεξάρτητα μεταξύ τους, όταν η πιθανότητα να συμβεί το ένα από αυτά, π.χ., το Γ, είναι ανεξάρτητη από την πραγματοποίηση ή μη του άλλου, π.χ., του Α.Ή συμβολικά, εάν P(Γ/A)=P(Γ)

Νόμος Πολλαπλασιασμού

Η πιθανότητα για τη σύγχρονη εμφάνιση δύο γεγονότων Α και Γ είναι ίση με το γινόμενο της πιθανότητας να εμφανιστεί το Α, ήτοι P(A), επί την πιθανότητα να εμφανιστεί το Γ, αφού προηγουμένως εμφανιστεί το Α, ήτοι επί P(Γ/A).

Εάν όμως τα γεγονότα Γ και Α είναι ανεξάρτητα, τότε σύμφωνα με τα παραπάνω θα είναι :

$$P(A \cap \Gamma) \equiv P(A) \cdot P(\Gamma)$$

που αποτελεί και το νόμο πολλαπλασιασμού πιθανοτήτων για ανεξάρτητα γεγονότα.

<u>Προγενέστερη και Μεταγενέστερη Πιθανότητα</u>

Όπως αναφέρθηκε και στα προηγούμενα διακρίνουμε, γενικά, δύο πιθανότητες για το γεγονός Γ. Μία πριν από την παρατήρηση της πραγματοποίησης του γεγονότος Α, την οποία ονομάζουμε προγενέστερη ή εκ των προτέρων ή a priori πιθανότητα και μία άλλη μετά την παρατήρηση της πραγματοποίησης του γεγονότος Α, την οποία ονομάζουμε μεταγενέστερη ή εκ των υστέρων ή a posteriori πιθανότητα. Δηλαδή:

Η πιθανότητα Ρ(Γ) πραγματοποίησης ενός γεγονότος Γ που σχετίζεται με ένα άλλο γεγονός (ή υπόθεση) Α λέγεται προγενέστερη πιθανότητα, όταν νοείται ότι αυτή υπολογίζεται πριν από την παρατήρηση της πραγματοποίησης του γεγονότος Α. Ενώ, η πιθανότητα Ρ(Γ/Α) πραγματοποίησης ενός γεγονότος Γ που σχετίζεται με ένα άλλο γεγονός (ή υπόθεση) Α λέγεται μεταγενέστερη πιθανότητα, όταν νοείται ότι αυτή υπολογίζεται μετά από την παρατηρηθείσα πραγματοποίηση του γεγονότος Α.

Μετά τους παραπάνω ορισμούς μπορούμε να διατυπώσουμε το ακόλουθο θεώρημα του BAYES που αφορά στις πιθανότητες των διαφόρων γεγονότων $\Gamma_1,\Gamma_2,\ldots\Gamma_\nu$, η πραγματοποίηση των οποίων έχουν σαν αποτέλεσμα την πραγματοποίηση του γεγονότος Α.

Το θεώρημα του BAYES δίνεται από τον τύπο:

$$P(\Gamma_\kappa/A)=\frac{P(\Gamma_\kappa)P(A/\Gamma_\kappa)}{P(\Gamma_1)P(A/\Gamma_1)+P(\Gamma_2)P(A/\Gamma_2)+\ldots+P(\Gamma_\nu)P(A/\Gamma_\nu)}$$

ήτοι :

$$P(\Gamma_\kappa/A)=\frac{P(\Gamma_\kappa)P(A/\Gamma_\kappa)}{\sum_{\kappa=1}^{\nu}P(\Gamma_\kappa)P(A/\Gamma_\kappa)}$$

Σημείωση : Τα γεγονότα Γ_1, Γ_2, …, Γ_v, είναι ασυμβίβαστα μεταξύ τους και η ένωσή τους αποτελεί το δειγματοχώρο S πράγμα που σημαίνει ότι οπωσδήποτε ένα και μόνο από τα γεγονότα αυτά θα πραγματοποιηθεί.

Τυχαία Μεταβλητή

Ένα μέγεθος ονομάζεται τυχαίο ή τυχαία μεταβλητή όταν, στα διάφορα πειράματα τύχης που διεξάγονται κάτω από τις ίδιες συνθήκες, λαμβάνει εντελώς τυχαίες τιμές.

Η τυχαία μεταβλητή είναι λοιπόν εξ ορισμού μια συνάρτηση ορισμένη στο δειγματοχώρο. Συνήθως, συμβολίζεται με ένα κεφαλαίο γράμμα Χ, Υ, κ.λπ., ενώ οι τιμές της συμβολίζονται με τα αντίστοιχα μικρά γράμματα x, y, κ.λπ.

Γενικά, μία τυχαία μεταβλητή παίρνει πεπερασμένο ή άπειρο πλήθος τιμών. Στην πρώτη περίπτωση ονομάζεται απαριθμητή ή διακριτή τυχαία μεταβλητή, ενώ στη δεύτερη, που παίρνει άπειρο πλήθος τιμών, συνεχής τυχαία μεταβλητή. Η ρίψη του ζαριού εμφανίζει τη μεταβλητή Χ να παίρνει, π.χ., μόνο τις διακριτές τιμές : x=1,2,3,4,5,6. Είναι επομένως η μεταβλητή Χ διακριτή τυχαία μεταβλητή. Όμως, η ταχύτητα V κάποιου μορίου αερίου, που με συγκρούσεις με άλλα μόρια μέσα σε δεδομένο χώρο μεταβάλλεται και μπορεί να πάρει οποιαδήποτε τιμή v_i, είναι συνεχής τυχαία μεταβλητή.

1.7. Νόμος Κατανομής

Ένα τυχαίο μέγεθος είναι εντελώς χαρακτηρισμένο εάν, εκτός από το πλήθος των τιμών τις οποίες μπορεί να πάρει, είναι γνωστή και η πιθανότητα κάθε τιμής.

Έτσι, αν μία διακριτή τυχαία μεταβλητή X μπορεί να πάρει τις τιμές $x_1,x_2,\ldots x_v$ και υποθέσουμε ότι οι πιθανότητες να πάρει η μεταβλητή τις τιμές αυτές είναι αντιστοίχως $P(x_1),P(x_2),\ldots,P(x_v)$ ή

$$P(X=x_\kappa)=f(x_\kappa),\ \kappa= 1,2,\ldots v \qquad (8)$$

μπορούμε να ορίσουμε μία συνάρτηση πιθανότητας ή συνάρτηση κατανομής

$$P(X=x)=f(x) \qquad (9)$$

τέτοια ώστε για $x=x_\kappa$ να μεταβαίνουμε στην (8), ενώ για τιμές του $x \neq x_\kappa$, $\kappa=1,2,3,\ldots v$ να είναι $f(x)=0$.

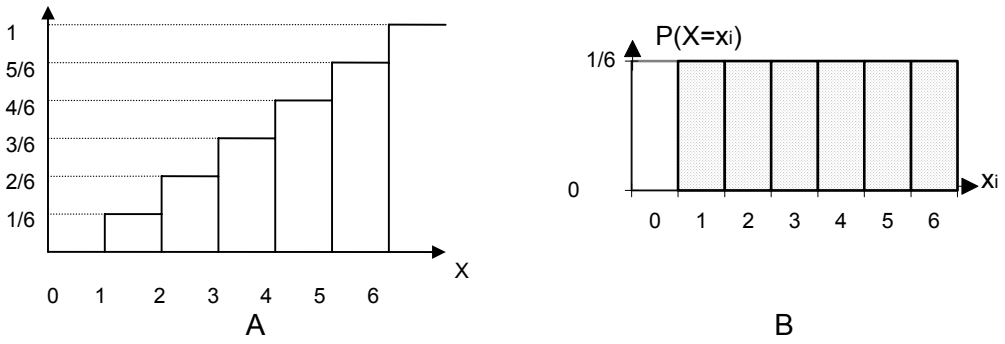
Κατά τη ρίψη ιδεωδώς κατασκευασμένου ζαριού, ο αριθμός που προκύπτει είναι τυχαία μεταβλητή X που παίρνει τις διακριτές τιμές $x_1=1$, $x_2=2$, $x_3=3$, $x_4=4$, $x_5=5$, $x_6=6$ και μάλιστα κάθε μία από αυτές με πιθανότητα $P(X=x_\kappa)=\frac{1}{6}$, κ=1, 2, 3, 4, 5, 6.

Επομένως ο νόμος κατανομής είναι x_κ: 1 2 3 4 5 6

$$P(X=x_\kappa): \quad \frac{1}{6}\ \frac{1}{6}\ \frac{1}{6}\ \ \frac{1}{6}\ \frac{1}{6}\ \frac{1}{6}$$

Η γραφική παράσταση του νόμου κατανομής γίνεται αν, ως τετμημένες θέσουμε τις δυνατές τιμές της τυχαίας μεταβλητής και ως τεταγμένες τις αντίστοιχες τιμές της πιθανότητας εμφάνισής τους (σχήμα 14).

Στο παραπάνω παράδειγμα οι στήλες που αντιστοιχούν στις τεταγμένες έχουν το ίδιο ύψος, γιατί $P(X=x_\kappa)=\frac{1}{6}$. Αυτό βέβαια, δεν συμβαίνει γενικώς. Εκείνο όμως που είναι ουσιώδες, είναι ότι το ολικό εμβαδό, που είναι το άθροισμα όλων των στηλών $P(X=x_\kappa)$, πρέπει να ισούται με τη μονάδα. Εφόσον το πλάτος της βάσης κάθε στήλης ληφθεί ίσο με τη μονάδα, προκύπτει ύψος $=\frac{1}{6}$ και επιφάνεια =1



Σχήμα 14. (Α) Παράδειγμα συνάρτησης κατανομής. (Β) Παράδειγμα νόμου κατανομής

Από το νόμο κατανομής κατασκευάζεται εύκολα η συνάρτηση κατανομής ως η συνάρτηση που δίνει την πιθανότητα, ο εμφανισθείς αριθμός να είναι μικρότερος ενός δοθέντος αριθμού. Ισχύει π.χ. f(1)=P(x<1)=0, ήτοι η πιθανότητα κατά τη ρίψη του ζαριού να εμφανιστεί αριθμός μικρότερος της μονάδας είναι ίση με μηδέν. Στη συνέχεια προκύπτουν για την κατανομή πιθανότητας :

f(2)=P(X<2)=P(X=1)= $\frac{1}{6}$

f(3)=P(X<3)=P(X=2)+P(X=1)= $\frac{1}{6}+\frac{1}{6}=\frac{1}{3}$

f(4)=P(X<4)=P(X=3)+P(X=2)+P(X=1)= $\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{1}{2}$

f(5)=P(X<5)=P(X=4)+P(X=3)+P(X=2)+P(X=1)= $\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{4}{6}=\frac{2}{3}$

f(6)=P(X<6)=P(X=5)+P(X=4)+P(X=3)+P(X=2)+P(X=1)=

$\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{5}{6}$

f(X>6)=P(X<x)=P(X=6)+P(X=5)+P(X=4)+P(X=3)+P(X=2)+P(X=1)=

$\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{6}{6}=1$

Γενικά, η f(x) είναι μία συνάρτηση πιθανότητας εάν ισχύει:

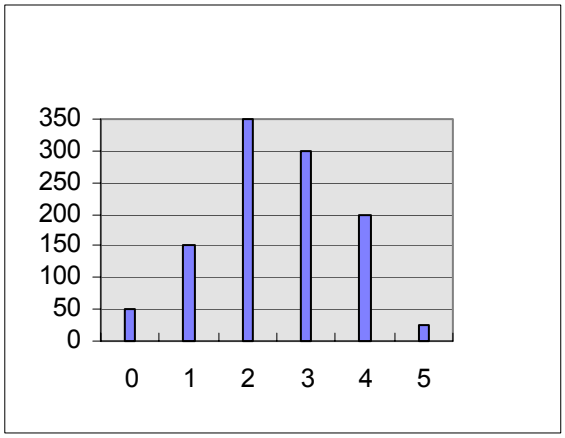
1. $f(x) \geq 0$ (σ_1)

2. $\sum_x f(x) = 1$ (σ_2)

Το άθροισμα στη 2 νοείται ως προς όλες τις δυνατές τιμές του x. Η γραφική παράσταση της f(x) ονομάζεται γραφική παράσταση της πιθανότητας.

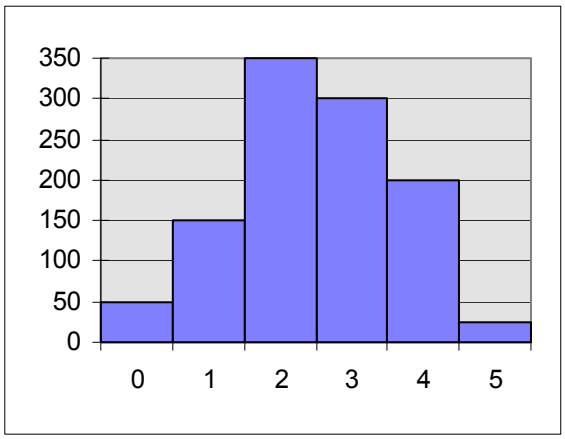
Μια γραφική παράσταση της συνάρτησης πιθανότητας μπορεί να δοθεί με ένα ραβδόγραμμα ή με ένα ιστόγραμμα.

Το ραβδόγραμμα είναι ένα διάγραμμα σε ορθογώνιο σύστημα αναφοράς που αποτελείται από μία διαδοχή ορθογωνίων με βάσεις ίσες (επάνω στον άξονα Ox ή Oy) και ύψη ανάλογα προς τις αντίστοιχες τιμές της μεταβλητής X (σχήμα 15).



Σχήμα 15. Ραβδόγραμμα

Το ιστόγραμμα είναι ένα διάγραμμα σε ορθογώνιο σύστημα αναφοράς που αποτελείται από ορθογώνια με βάσεις τα ίσα τμήματα του άξονα Ox στα οποία αντιστοιχεί το εύρος κάθε τάξης και ύψη τις αντίστοιχες συχνότητες (σχήμα 16).



Σχήμα 16. Ιστόγραμμα

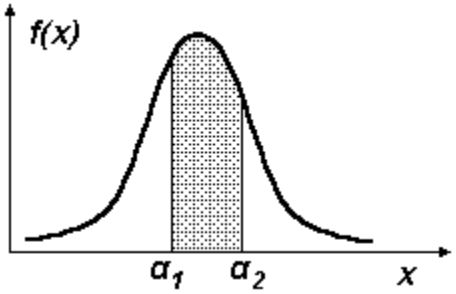
Στα δύο παραπάνω διαγράμματα οι τεταγμένες εκφράζουν το πλήθος των πειραμάτων.

Και από τα δύο διαγράμματα, τόσο από το ραβδόγραμμα όσο και από το ιστόγραμμα, αντλούμε την ίδια πληροφορία. Προτιμούμε ένα ραβδόγραμμα όταν η τυχαία μεταβλητή X παίρνει μόνο ακέραιες τιμές, ενώ χρησιμοποιούμε το ιστόγραμμα όταν οι τιμές της μεταβλητής είναι συνεχείς.

Όπως αναφέραμε και στα προηγούμενα μία συνάρτηση f(x) είναι μία συνάρτηση πιθανότητας εάν ικανοποιεί τις παραπάνω συνθήκες (σ_1) και (σ_2). Ακόμη, η συνάρτηση πιθανότητας ονομάζεται και κατανομή πιθανότητας για την αντίστοιχη συνεχή τυχαία μεταβλητή ή επίσης συνάρτηση πυκνότητας πιθανότητας ή απλά συνάρτηση πυκνότητας.

Μία συνεχής τυχαία μεταβλητή παίρνει σε δεδομένο διάστημα άπειρο πλήθος τιμών. Επομένως, η πιθανότητα να πάρει μία μεμονωμένη τιμή είναι ίση με μηδέν. Για το λόγο αυτόν δεν μπορούμε να ορίσουμε μία συνάρτηση πιθανότητας, όπως ακριβώς για μία διακριτή μεταβλητή.

Εκείνο όμως που μπορούμε να κάνουμε είναι να ορίσουμε την πιθανότητα να πάρει η συνεχής τυχαία μεταβλητή X τιμές μεταξύ δύο τιμών α_1 και α_2 ($\alpha_2>\alpha_1$). Ήτοι $P(\alpha_1<X<\alpha_2)$. Η πιθανότητα αυτή, που γραφικά παριστάνεται από τη σκιασμένη επιφάνεια του σχήματος 17,



Σχήμα 17. Παράσταση συνάρτησης πυκνότητας

είναι ίση με : $P(\alpha_1<X<\alpha_2)=\int_{a1}^{a2} f(x)dx = F(a_2) - F(a_1)$

1.8. Παράμετροι Κατανομής

Σκοπός της Στατιστικής Επιστήμης είναι η όσο το δυνατόν ταχύτερη εξαγωγή της πληροφορίας από μεγάλους όγκους στοιχείων.

Η γενική μεθοδολογία που ακολουθείται, μπορούμε να πούμε ότι περιλαμβάνει τη συλλογή των κάθε μορφής στατιστικών στοιχείων, την ταξινόμησή τους και στη συνέχεια την παρουσίασή τους σε πίνακες συχνοτήτων.

Όμως, και τα πινακοποιημένα στοιχεία, παρόλο που περιορίζουν κατά πολύ τον όγκο των πρωτογενών συγκεντρωθέντων στοιχείων, εξακολουθούν να μη δίνουν μια πλήρη εικόνα του μεγέθους στο οποίο αναφέρονται. Γιαυτό, θεωρείται πολλές φορές αναγκαίο να γίνει μια μεγαλύτερη ακόμη συμπύκνωση των στοιχείων και να αντικατασταθούν τελικά με ορισμένες παραμέτρους της κατανομής.

Τέτοιες παράμετροι κατανομής είναι : οι παράμετροι **θέσεως** και οι παράμετροι **διασποράς.**

1.8.1. Παράμετροι Θέσεως

Οι παράμετροι θέσεως εντοπίζουν τη "θέση" ενός πληθυσμού, γιατί απεικονίζουν στοιχεία γύρω από τα οποία βρίσκονται οι τιμές της μεταβλητής που ερευνούμε. Η σημαντικότερη από τις παραμέτρους θέσεως είναι η μέση τιμή:

Για μια διακριτή τυχαία μεταβλητή X που παίρνει τις τιμές $x_1,x_2,x_3,\ldots x_v$ η μέση τιμή δίνεται από την έκφραση:

$$M(X)=x_1P(X=x_1)+\ldots+x_vP(X=x_v)=\sum_{i=1}^{v} x_i P(X=x_i)$$

και αν θέσουμε $P(X=x_i)=f(x_i)$

$$M(X)=x_1f(x_1)+\ldots+x_vf(x_v)=\sum_{i=1}^{v} x_i f(x_i)$$

δηλαδή, η μέση τιμή M(X) μιας διακριτής τυχαίας μεταβλητής προκύπτει με πολλαπλασιασμό κάθε τιμής της επί την αντίστοιχη πιθανότητα και στη συνέχεια με άθροιση των επιμέρους γινομένων.

Η μέση αυτή τιμή δεν είναι απαραίτητα μία από τις τιμές της μεταβλητής X.

Στην ειδική περίπτωση που όλες οι πιθανότητες είναι ίσες, δηλαδή $f(x_1)=f(x_2)=\ldots=f(x_v)$ και επειδή $f(x_1)+f(x_2)+\ldots+f(x_v)=1$, αφού το να λάβει μία τιμή η τυχαία μεταβλητή είναι βέβαιο γεγονός, θα είναι $f(x_i)=\frac{1}{v}$, i=1,2,…v,

έχουμε :

$$M(X)=\frac{x_1+x_2+\ldots x_v}{v}$$

που ονομάζεται και αριθμητικός μέσος όρος ή απλά μέσος όρος των $x_1,x_2,x_3,\ldots x_v$.

Για συνεχή τυχαία μεταβλητή X με συνάρτηση πυκνότητας f(x), η μέση τιμή λαμβάνεται αν πολλαπλασιάσουμε τη συνάρτηση πυκνότητας f(x) επι x και κατόπιν ολοκληρώσουμε από $-\infty$ έως $+\infty$, δηλαδή προκύπτει από την έκφραση :

$$M(X)=\int_{-\infty}^{+\infty} xf(x)dx$$

Η μέση τιμή της X πολλές φορές συμβολίζεται με μ ή με μ_y ή με μ_z αν η τυχαία μεταβλητή είναι αντιστοίχως η Y ή η Z.

Για τη μέση τιμή ισχύουν οι ακόλουθες προτάσεις :

1. Εάν c είναι μία σταθερή, τότε M(cX)=cM(X)
2. Εάν X και Y είναι δύο τυχαίες μεταβλητές οπότε και το άθροισμά τους X+Y είναι επίσης τυχαία μεταβλητή, τότε M(X+Y)=M(X)+M(Y)
3. Εάν X και Y είναι δυο ανεξάρτητες τυχαίες μεταβλητές, τότε M(XY)=M(X)M(Y)

Σημείωση : Οι παραπάνω προτάσεις γενικεύονται και για περισσότερες τυχαίες μεταβλητές.

Συνοπτικά, μπορούμε να πούμε ότι η **μέση τιμή** (αριθμητικός μέσος) :

$$\bar{x} = \frac{\sum_{i=1}^{v} [x_i]}{v} \qquad \bar{x} = \frac{\sum_{i=1}^{v} [v_i x_i]}{\sum_{i=1}^{v} [v_i]}$$

είναι το άθροισμα των τιμών ενός συνόλου δεδομένων, διαιρεμένο δια του πλήθους τους.

Ο δεύτερος τύπος ισχύει, όταν τα δεδομένα μας είναι ήδη ομαδοποιημένα, δηλαδή, όταν είναι γνωστή η απόλυτη συχνότητα v_i με την οποία εμφανίζεται καθε ένα στο δείγμα. Ακριβώς, ο ίδιος τύπος ισχύει, αν θέλουμε να δώσουμε "βάρος" μεγαλύτερο από 1 σε ορισμένα στοιχεία του συνόλου (όπου, v_i είναι τώρα τα "βάρη").

Οι παραπάνω τρεις προτάσεις που ισχύουν για τη μέση τιμή σημαίνουν ότι :

i. Αν πολλαπλασιάσουμε όλα τα δεδομένα με μια σταθερή c, τότε πολλαπλασιάζεται και η μέση τιμή με c.

ii. Αν έχουμε δυο σειρές δεδομένων (x και y) τότε:

iia. Η μέση τιμή των αθροισμάτων (x+y) είναι ίση με το άθροισμα των μέσων τιμών των x και y.

iib. Η μέση τιμή των γινομένων (x·y) είναι ίση με το γινόμενο των μέσων τιμών των x και y.

Άλλες παράμετροι θέσεως είναι ακόμη :

Ο **γεωμετρικός μέσος :** $G^{\nu} = \sqrt[\nu]{x_1 \cdot x_2 \cdot \ldots \cdot x_{\nu}}$

Ο **αρμονικός μέσος :** $H = \dfrac{v}{\sum_{i=1}^{v} \left[\dfrac{1}{x_i}\right]}$

Η **διάμεσος τιμή :** Για να οριστεί η διάμεσος τιμή μιας μεταβλητής πρέπει οι τιμές της να διαταχθούν κατ' αύξουσα σειρά. Αν το πλήθος **v** των τιμών της μεταβλητής είναι περιττός αριθμός, τότε ως διάμεσος τιμή λαμβάνεται η κεντρική τιμή της διάταξης, δηλαδή η τιμή που βρίσκεται στην **(v+1)/2** θέση της διάταξης. Ενώ, αν το **v** είναι άρτιος, τότε ως διάμεσος τιμή λαμβάνεται η μέση τιμή των δυο κεντρικών τιμών της διάταξης, δηλαδή των τιμών που βρίσκονται στις θέσεις **(v/2)** και **(v/2)+1**.

Η **επικρατούσα τιμή :** Είναι η τιμή της μεταβλητής με τη μεγαλύτερη συχνότητα εμφάνισης (δεν αλλάζει, αν χρησιμοποιήσουμε απόλυτες ή σχετικές συχνότητες).

Η σημασία των παραμέτρων θέσεως είναι μεγάλη, όταν ο πληθυσμός είναι ομοιογενής. Στην περίπτωση που ο πληθυσμός είναι ανομοιογενής μειώνεται σημαντικά η σημασία τους. Ένα παράδειγμα

έχουμε, αν θεωρήσουμε δυο μεταβλητές Χ και Υ που παίρνουν αντίστοιχα τις τιμές :

Χ : 15, 38, 38, 43, 47, 47, 50, 51, 52, 55, 59

Υ : 5, 12, 13, 20, 35, 47, 52, 53, 77, 85, 96

Παρατηρούμε ότι παρόλο που οι δυο μεταβλητές έχουν τον ίδιο αριθμητικό μέσο 45 και την ίδια διάμεσο 47, οι δυο μεταβλητές διαφέρουν μεταξύ τους, γιατί οι τιμές της μεταβλητής Χ κυμαίνονται μεταξύ των αριθμών 15 και 59, ενώ οι τιμές της μεταβλητής Υ κυμαίνονται μεταξύ των αριθμών 5 και 96.

Βλέπουμε λοιπόν ότι οι δυο παράμετροι θέσεως, μέση τιμή και διάμεσος μιας μεταβλητής, δεν μας καθορίζουν πόσο συγκεντρωμένες ή πόσο διασπαρμένες είναι οι τιμές της μεταβλητής γύρω από αυτές. Έτσι, η εικόνα που μας δίνουν είναι ανεπαρκής και γιαυτό καταφεύγουμε και σε άλλες παραμέτρους που μπορούν να μας δώσουν το βαθμό συγκέντρωσης ή διασποράς των τιμών της μεταβλητής γύρω από κάποια παράμετρο θέσεως.

1.8.2. Παράμετροι Διασποράς

Οι αριθμοί που εκφράζουν πόσο συγκεντρωμένες ή πόσο διασπαρμένες είναι οι παρατηρήσεις (μετρήσεις) μας ονομάζονται παράμετροι διασποράς και είναι οι εξής :

Εύρος Μεταβολής

Υποθέτουμε ότι μια μεταβλητή Χ παίρνει τις τιμές : x_1, x_2, x_3,..., x_v και έστω ότι x_μ είναι η μικρότερη τιμή και x_M η μεγαλύτερη τιμή. Η διαφορά ανάμεσα στη μεγαλύτερη και τη μικρότερη τιμή, δηλαδή η διαφορά $x_M - x_\mu$ είναι το εύρος μεταβολής.

Το εύρος μεταβολής είναι η απλούστερη παράμετρος διασποράς και δίνει μια γενική εικόνα του βαθμού διασποράς των τιμών μιας μεταβλητής μεταξύ τους. Όμως, έχει το μειονέκτημα να μην εξαρτάται από όλες τις τιμές της μεταβλητής, αλλά μόνο από τις ακραίες

τιμές της, τη μέγιστη και την ελάχιστη. Έτσι, η παράμετρος αυτή μπορεί να οδηγήσει σε ψευδή εικόνα διασποράς των τιμών της μεταβλητής, αν μια απο τις ακραίες τιμές της μεταβλητής είναι υπερβολικά υψηλή ή χαμηλή ή αν προέρχεται από λανθασμένη παρατήρηση (μέτρηση).

Το εύρος μεταβολής χρησιμοποιείται κυρίως σε στατιστικές πολλών μεγεθών που εμφανίζονται με τη μορφή ανωτάτων και κατωτάτων τιμών, π.χ., θερμοκρασίες του αέρα στη Μετεωρολογία και τιμές των μετοχών στα Χρηματιστήρια.

Αν είναι γνωστές οι τιμές Q_1, Q_2, Q_3 που καθορίζουν τα τεταρτημόρια της κατανομής (στα σημεία 25%, 50%, 75% αντίστοιχα, δηλαδή έτσι ώστε 25%, 50% και 75% των στοιχείων να βρίσκονται κάτω από τα αντίστοιχα όρια Q_1, Q_2, Q_3) χρησιμοποιούμε καλύτερα το **μισό του ενδομοριακού εύρους μεταβολής $Q = (Q_3 - Q_1) / 2$** σαν τιμή που μπορεί να μας δώσει ακριβέστερη ένδειξη εύρους μεταβολής. Τα σημεία Q_1, Q_2, Q_3 υπολογίζονται είτε γραφικά είτε υπολογιστικά.

Μέση Απόκλιση

Αν μια μεταβλητή X παίρνει τις τιμές : x_1, x_2, x_3,..., x_v που έχουν μέση τιμή x, τότε η μέση τιμή των απολύτων τιμών όλων των διαφορών $x_1 - x$, $x_2 - x$, $x_3 - x$, ..., $x_v - x$, δηλαδή η θετική ποσότητα :

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

ονομάζεται **μέση απόκλιση.**

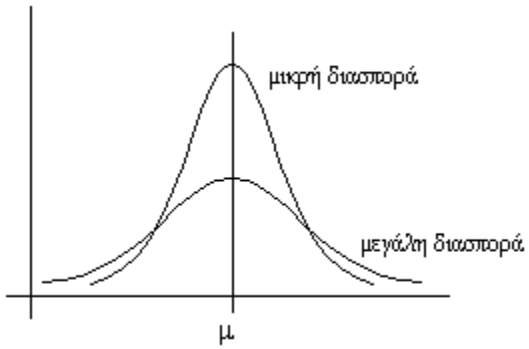
Η μέση απόκλιση δίνει το βαθμό συγκέντρωσης των μετρήσεών μας γύρω από τη μέση τιμή τους. Έχει το πλεονέκτημα ότι ο υπολογισμός της συνάγεται από όλες τις τιμές της μεταβλητής X και όχι μόνο από τις ακραίες τιμές της.

Διακύμανση ή Διασπορά και Τυπική Απόκλιση

Η βασικότερη παράμετρος διασποράς είναι η διακύμανση ή διασπορά. Η παράμετρος αυτή δείχνει το “άπλωμα” της κατανομής.

Δηλαδή, η διασπορά αποτελεί ένα μέτρο του πόσου “απλωμένες” είναι οι τιμές x_i της τυχαίας μεταβλητής X γύρω από τη μέση τιμή M(X). Εάν οι διάφορες τιμές x_i της τυχαίας μεταβλητής X είναι συγκεντρωμένες κοντά στη μέση τιμή M(X), τότε η διασπορά είναι μικρή, εάν όμως είναι διασπαρμένες, τότε η διασπορά είναι μεγάλη.

Στο σχήμα 18 απεικονίζονται δυο συνεχείς κατανομές που έχουν την ίδια μέση τιμή αλλά, διαφέρουν στις διασπορές, η μια έχει μικρή διασπορά και η άλλη μεγάλη.



Σχήμα 18. Συνεχείς κατανομές με την ίδια μέση τιμή αλλά διαφορετικές διασπορές

Για να ορίσουμε τη διασπορά των τιμών της τυχαίας μεταβλητής X θεωρούμε τη μέση τιμή της M(X). Εάν x_i είναι μία τυχούσα τιμή της μεταβλητής X, i=1,2,…,N, τότε η ποσότητα x_i-M(X) ονομάζεται απόκλιση της x_i τιμής της μεταβλητής.

Εάν θεωρήσουμε τώρα τις ποσότητες $[x_i-\mu(X)]^2$ για κάθε x_i, i=1,2,…,N και πάρουμε το μέσο όρο των ποσοτήτων αυτών, δηλαδή αν πάρουμε το μέσο όρο των τετραγώνων των αποκλίσεων των τιμών της μεταβλητής X, τοτε ορίζουμε ένα μη αρνητικό αριθμό σ^2 που ονομάζεται διασπορά της μεταβλητής X και είναι ίση με :

$$\sigma^2 = \frac{[x_1 - M(X)]^2 + [x_2 - M(X)]^2 + \ldots + [x_\nu - M(X)]^2}{N}$$

ή $\sigma^2 = \dfrac{\sum_{i=1}^{N}[x_i - M(X)]^2}{N}$

Τη θετική τετραγωνική ρίζα της παραπάνω ποσότητας, ήτοι την

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}[x_i - M(X)]^2}{N}}$$

ονομάζουμε **τυπική απόκλιση** της μεταβλητής Χ.

Εάν η Χ είναι μία διακριτή τυχαία μεταβλητή με συνάρτηση πιθανότητας f(x), τότε για τη διασπορά μπορούμε να γράψουμε και τη σχέση :

$$\sigma_x^2 = \frac{\sum_{i=1}^{N}[x_i - M(X)]^2}{N} = \sum_{i=1}^{N}[x_i - M(X)]^2 f(x_i)$$

και αν όλες οι πιθανότητες είναι ίσες, και η μέση τιμή Μ(Χ) της τυχαίας μεταβλητής Χ συμβολιστεί με μ, τότε έχουμε :

$$\sigma^2 = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \ldots + (x_N - \mu)^2}{N}$$

που δίνει τη διασπορά των Ν τιμών $x_1, x_2, \ldots x_N$ της τυχαίας μεταβλητής Χ.

Εάν η Χ είναι μία συνεχής τυχαία μεταβλητή με συνάρτηση πιθανότητας f(x), τότε η διασπορά δίνεται από τη σχέση:

$$\sigma_x^2 = \int_{-\infty}^{+\infty}(x - \mu)^2 f(x)dx$$

Από τους παραπάνω ορισμούς είναι φανερό ότι η τυπική απόκλιση σ εκφράζεται στις ίδιες μονάδες που εκφράζεται και η τυχαία μεταβλητή (π.χ. cm, sec κ.λπ.), ενώ η διασπορά εκφράζεται στα τετράγωνα των μονάδων της τυχαίας μεταβλητής (π.χ. cm^2, sec^2 κ.λπ.). Γι'αυτό και τις περισσότερες φορές προτιμάται η τυπική απόκλιση (standard deviation).

Εάν τη διασπορά ορίσουμε και με τον εξής βολικό τρόπο :

$Var(X)=M[(X-\mu)^2]$,

τότε εύκολα αποδεικνύεται ότι ισχύουν οι προτάσεις:

1. Ισχύει $\sigma^2=M[(X-\mu)^2]=M(X^2)-\mu^2=M(X^2)-[M(X)]^2$ όπου $\mu=M(X)$
2. Εάν c είναι μία σταθερή, τότε $Var(cX)=c^2Var(X)$
3. Η ποσότητα $M[(X-\alpha)^2]$ γίνεται ελάχιστη, όταν $\alpha=\mu=M(X)$
4. Εάν X και Y είναι δυο ανεξάρτητες τυχαίες μεταβλητές με διασπορά αντίστοιχα $Var(X)\equiv \sigma_X^2$ και $Var(Y)\equiv \sigma_Y^2$,τότε και η τυχαία μεταβλητή Z=X+Y είναι τυχαία μεταβλητή και η διασπορά της $Var(Z)=Var(X+Y)\equiv \sigma_Z^2 \equiv \sigma_{X+Y}^2$ είναι ίση με: Var(Z)=Var(X+Y)=Var(X)+Var(Y), ήτοι

$$\sigma_Z^2 \equiv \sigma_{X+Y}^2 \equiv \sigma_X^2 + \sigma_Y^2$$

δηλαδή η διασπορά (Variance) αθροίσματος δύο τυχαίων μεταβλητών ισούται προς το άθροισμα των διασπορών των επιμέρους μεταβλητών.

Η παραπάνω πρόταση ισχύει με την ίδια έκφραση και για τη διασπορά διαφοράς δύο ανεξάρτητων μεταβλητών, δηλαδή και Var(X-Y)=Var(X)+Var(Y), ή $\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2$

Σημείωση : Η τελευταία πρόταση 4 γενικεύεται και για περισσότερες από δύο ανεξάρτητες τυχαίες μεταβλητές.

Η τυπική απόκλιση **σ** είναι η κυριότερη παράμετρος διασποράς και φανερώνει πόσο διασπαρμένες γύρω από τη μέση τιμή της μεταβλητής είναι οι άλλες τιμές της. Δηλαδή, μεγάλη τιμή της τυπικής απόκλισης σημαίνει ότι οι τιμές της μεταβλητής X (παρατηρήσεις μας, μετρήσεις μας) είναι πολύ διασπαρμένες γύρω από τη μέση τιμή της μεταβλητής, ενώ μικρή τιμή της τυπικής απόκλισης σημαίνει ότι οι τιμές

της μεταβλητής Χ (παρατηρήσεις μας, μετρήσεις μας) είναι συγκεντρωμένες γύρω από τη μέση τιμή της.

Η χρησιμότητα και η σημασία της τυπικής απόκλισης στη Στατιστική είναι μεγάλη και συναντάται σε πολλά θέματά της.

Συντελεστής Μεταβλητότητας

Συντελεστής μεταβλητότητας μιας κατανομής, που έχει μέση τιμή $\bar{x}$ ονομάζεται το πηλίκο

$$CV = \frac{\sigma}{\bar{x}} \ ,$$

δηλαδή ο λόγος της τυπικής απόκλισης της κατανομής προς τη μέση τιμή της.

Επειδή η τυπική απόκλιση σ και η μέση τιμή $\bar{x}$ μιας μεταβλητής Χ εκφράζονται στις ίδιες μονάδες (τις μονάδες που εκφράζονται και οι τιμές της μεταβλητής Χ), ο συντελεστής μεταβλητότητας είναι καθαρός αριθμός, ανεξάρτητος από τις μονάδες μέτρησης και για λόγους καλύτερης κατανόησης εκφράζεται συνήθως επί τοις %.

Όταν οι τιμές δυο μεταβλητών Χ και Υ εκφράζονται σε διαφορετικές μονάδες, π.χ., αν η μεταβλητή Χ αφορά στο βάρος μιας ομάδας ατόμων θα εκφράζεται σε κιλά και αν η μεταβλητή Υ αφορά στο ύψος της ίδιας ομάδας θα εκφράζεται σε cm, η σύγκριση της διασποράς των τιμών των δυο κατανομών δεν είναι δυνατόν να γίνει παρά μόνο με τους συντελεστές μεταβλητότητας.

Παράδειγμα **:** Από την εξέταση μιας ομάδας ατόμων ως προς το βάρος και το ύψος τους βρέθηκε:

Η μέση τιμή του βάρους των μελών της ομάδας ήταν $\bar{x}$ = 70 κιλά

και η τυπική απόκλιση $\boldsymbol{\sigma_x}$ = 4 κιλά, ενώ

η μέση τιμή του ύψους των μελών της ομάδας ήταν $\bar{y}$ = 170 cm

και η τυπική απόκλιση $\boldsymbol{\sigma_y}$ = 7 cm.

Στο παράδειγμα δεν μπορούμε να αποφανθούμε ποια από τις δυο κατανομές μεταβλητών (Χ = βάρους και Υ = ύψους) παρουσιάζει

μεγαλύτερη διασπορά τιμών συγκρίνοντας τις τυπικές αποκλίσεις τους, γιατί οι τιμές του βάρους και του ύψους εκφράζονται σε διαφορετικές μονάδες μέτρησης. Αν όμως πάρουμε τους συντελεστές μεταβλητότητας που αντίστοιχα για τις δυο μεταβλητές Χ και Υ είναι:

για τη μεταβλητή Χ (βάρος): $CV = \frac{4}{70} = 0.057$, ήτοι 5.7% και

για τη μεταβλητή Υ (ύψος): $CV = \frac{7}{170} = 0.041$, ήτοι 4.1%,

μπορούμε να αποφανθούμε ότι η κατανομή του βάρους παρουσιάζει μεγαλύτερη διασπορά τιμών σε σύγκριση με την κατανομή του ύψους.

Γενικά, μπορούμε να πούμε ότι, όσο μικρότερος είναι ο συντελεστής μεταβλητότητας, τόσο μεγαλύτερη ομοιογένεια έχουν οι τιμές της μεταβλητής. Όμως, ο συντελεστής μεταβλητότητας, εξ ορισμού, επηρεάζεται από την τιμή της μέσης τιμής της μεταβλητής και μπορεί να πάρει πολύ μεγάλες τιμές στην περίπτωση που η μέση τιμή της μεταβλητής τείνει προς πολύ μικρές τιμές. Στην περίπτωση αυτή θα έχουμε ψευδή εικόνα της διασποράς των τιμών της υπόψη μεταβλητής. Έτσι, θα πρέπει να αποφεύγεται η χρησιμοποίηση αυτής της παραμέτρου σε περιπτώσεις που η μέση τιμή της υπό εξέταση μεταβλητής είναι πολύ μικρή.

Ασυμμετρία και Κύρτωση

Άλλες δύο παράμετροι κατανομής, που μας πληροφορούν για τη μορφή της καμπύλης κατανομής των τιμών της μεταβλητής (πόσο συμμετρικά ή πόσο συγκεντρωμένα κατανέμονται οι τιμές της μεταβλητής γύρω από τη μέση τιμή), αντίστοιχα είναι η ασυμμετρία και η κύρτωση.

$$\textbf{Ασυμμετρία :} \frac{\sum_{i=1}^{v}[x_i-\bar{x}]^3}{v\sigma^3}, \qquad \textbf{Κύρτωση :} \frac{\sum_{i=1}^{v}[x_i-\bar{x}]^4}{v\sigma^4}$$

Για περισσότερες πληροφορίες πάνω στις διάφορες μορφές στατιστικών κατανομών, στους πληθυσμούς (πεπερασμένους ή άπειρους) και στα δείγματα, μπορεί ο αναγνώστης να ανατρέξει σε εγχειρίδια Θεωρίας Στατιστικής και Πιθανοτήτων.

1.8.3. Παράμετροι δείγματος και παράμετροι πληθυσμού

Ορίζοντας προηγουμένως τη μέση τιμή και τη διασπορά, θεωρήσαμε σιωπηρά ότι το σύνολο των τιμών x_i αποτελεί το σύνολο των τιμών που μπορεί να πάρει η μεταβλητή X, οπότε οι τιμές τα μεγέθη $\bar{x}$ και σ^2 είναι η μέση τιμή και η διασπορά του πληθυσμού αντίστοιχα, ο οποίος εκφράζεται με τη μεταβλητή X. Ωστόσο, συχνά δε γνωρίζουμε όλες τις τιμές που μπορεί να πάρει η μεταβλητή X, αλλά τις τιμές $x_1,x_2,...,x_v$ ενός δείγματος του πληθυσμού, μεγέθους v. Σε μια τέτοια περίπτωση δε μιλάμε, κατ'αρχήν, για μέση τιμή και διασπορά πληθυσμού, αλλά για μέση τιμή και διασπορά δείγματος.

Η **δειγματική μέση τιμή** ορίζεται ως :

$$\bar{x}=(x_1+x_2+...+x_v)/v$$

δηλαδή με τον ίδιο τρόπο όπως και η μέση τιμή του πληθυσμού. Συχνά συμβολίζεται με μ η μέση τιμή του πληθυσμού και με $\bar{x}$ η δειγματική μέση τιμή.

Αποδεικνύεται μάλιστα ότι η μέση τιμή $E(\bar{X})$ των δειγματικών μέσων τιμών είναι ίση με τη μέση τιμή του πληθυσμού, δηλαδή :

$$E(\bar{X}) = \mu$$

Μπορεί να οριστεί μια **δειγματική διασπορά** s^2, με τον ίδιο τρόπο που ορίστηκε και η διασπορά του πληθυσμού σ^2, δηλαδή :

$$s^2 = [(x_1 - \bar{x})^2+(x_2 - \bar{x})^2+...+(x_v - \bar{x})^2] / v$$

όπου $\bar{x}$ η δειγματική μέση τιμή.

Με τα σύμβολα $\bar{X}$ και S υποδηλώνονται οι παράμετροι δειγματικής μέσης τιμής και διασποράς αντίστοιχα, ενώ με $\bar{x}$ και s υποδηλώνονται οι εκάστοτε τιμές των παραμέτρων αυτών.

Αποδεικνύεται ότι :

$$E(S^2) = [(v-1)/v]\cdot\sigma^2$$

που σημαίνει ότι η μέση τιμή της δειγματικής διασποράς διαφέρει από τη διασπορά του πληθυσμού.

Για το λόγο αυτό, προτιμάται, ενίοτε, να οριστεί η δειγματική διασπορά με βάση το μέγεθος s^2_{v-1}, όπου :

$$s^2_{v-1} = [(x_1- \bar{x})^2+(x_2- \bar{x})^2+...+(x_v- \bar{x})^2] / (v-1)$$

Στην περίπτωση αυτή αποδεικνύεται ότι η μέση τιμή των ποσοτήτων s^2_{v-1} είναι ίση με τη διασπορά του πληθυσμού, δηλαδή :

$$E(S^2_{v-1}) = \sigma^2$$

Όσο μεγαλύτερο είναι το μέγεθος του δείγματος, τόσο τείνει να εξισωθεί το s^2 με το s^2_{v-1}. Στις παρούσες σημειώσεις, η δειγματική διασπορά υπολογίζεται ως s^2.

Σε ένα δείγμα, που αποτελεί υποσύνολο του πληθυσμού, μπορούν να προσδιοριστούν η δειγματική μέση τιμή και η δειγματική διασπορά και από τα μεγέθη αυτά να εκτιμηθούν η μέση τιμή και η διασπορά του πληθυσμού.

Η δειγματική τυπική απόκλιση ορίζεται ως η τετραγωνική ρίζα της δειγματικής διασποράς.

Εφαρμογή

Σε 8 λεπτές τομές από πυριγενές πέτρωμα, μετρήθηκε η εκατοστιαία περιεκτικότητα σε χαλαζία και προέκυψαν τα παρακάτω αποτελέσματα:

Περιεκτικότητα (%) σε χαλαζία							
23.5	16.6	25.4	19.1	19.3	22.4	20.9	24.9

Να βρεθούν η μέση τιμή, η διασπορά και η τυπική απόκλιση του δείγματος.

Απάντηση

Το μέγεθος του δείγματος είναι v=8.

Η δειγματική μέση τιμή είναι :

$$\bar{x} = 23.5+16.6+25.4+19.1+19.3+22.4+20.9+24.9) / 8 = 21.5$$

Η δειγματική διασπορά s^2 είναι :

$$s^2 = [(23.5-21.5)^2+(16.6-21.5)^2+(25.4-21.5)^2+(19.1-21.5)^2+ \\ +(19.3-21.5)^2+(22.4-21.5)^2+(20.9-21.5)^2+(24.9-21.5)^2] / 8$$

Επομένως :

$$s^2 = 8.32$$

Η δειγματική διασπορά μπορεί να υπολογιστεί και ως :

$s^2_{v-1} = [(23.5-21.5)^2+(16.6-21.5)^2+(25.4-21.5)^2+(19.1-21.5)^2+$
$+(19.3-21.5)^2+(22.4-21.5)^2+(20.9-21.5)^2+(24.9-21.5)^2] / (8-1)$

Επομένως :

$s^2_{v-1} = 9.51$

Η δειγματική τυπική απόκλιση είναι :

$s = \sqrt{s^2} = \sqrt{8.32} = 2.88$

ή, εναλλακτικά,

$s_{v-1} = \sqrt{s^2_{v-1}} = \sqrt{9.51} = 3.08$

1.9. Η Ανισότητα του Tschebyschev

Στα προηγούμενα αναφερθήκαμε στις σχέσεις μεταξύ κατανομής, μέσης τιμής και διασποράς. Η γνώση της μέσης τιμής και της διασποράς μιας τυχαίας μεταβλητής X (διακριτής ή συνεχούς) δίνει, βέβαια, κάποια εποπτεία της τυχαίας μεταβλητής, δεν επαρκεί όμως για την εκτίμηση της πιθανότητας αποκλίσεων από τη μέση τιμή μ. Αυτήν την εκτίμηση όμως μας δίνει η ανισότητα του Tschebyschev που εκφράζεται ως εξής: Για δεδομένη τυχαία μεταβλητή X (διακριτή ή συνεχή) με τιμές x_i i=1,2,…. και μέση τιμή μ και διασπορά σ^2 πεπερασμένες, για κάθε θετικό αριθμό ε ισχύει:

$$P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

Δηλαδή, η πιθανότητα ότι η διαφορά (x-μ), της τιμής x της μεταβλητής X από τη μέση τιμή της μ, είναι κατά απόλυτη τιμή

μεγαλύτερη ή ίση με δεδομένο αριθμό ε είναι μικρότερη ή ίση του λόγου της διασποράς σ^2 προς το τετράγωνο του αριθμού ε.

Αυτό σημαίνει ότι μπορούμε να εκτιμήσουμε την πιθανότητα να διαφέρει η μεταβλητή Χ από τη μέση τιμή της, τόσο (ή και πλέον) όσο ορίζει ένας θετικός αριθμός ε.

1.10. Ο Νόμος των Μεγάλων Αριθμών

Πολλές φορές στην καθημερινή πρακτική απαιτείται η πραγματοποίηση ενός γεγονότος να είναι βέβαιη, π.χ. η μεταφορά επιβατών με μεταφορικό μέσο να είναι ασφαλής, οπότε η πιθανότητα πραγματοποίησης του γεγονότος, δηλαδή της ασφαλούς μεταφοράς πρέπει να είναι πάρα πολύ κοντά στη μονάδα. Άλλες φορές πάλι, προς αποφυγή μαζικών καταστροφών, απαιτείται η μη πραγματοποίηση ενός γεγονότος, δηλαδή η πραγματοποίηση του γεγονότος να είναι μηδενική, π.χ. η κατάρρευση ενός ουρανοξύστη, ενός φράγματος, κ.λπ. Το τελευταίο όμως ανάγεται στην ανάγκη πραγματοποίησης του αντιθέτου γεγονότος, δηλαδή της μη κατάρρευσης του ουρανοξύστη, του φράγματος, κ.λπ. με απαίτηση και εδώ η πιθανότητα να είναι πάρα πολύ κοντά στη μονάδα.

Έτσι, ένα από τα προβλήματα του λογισμού των πιθανοτήτων είναι η ανεύρεση νόμων για τους οποίους η πιθανότητα πραγματοποίησής τους είναι πάρα πολύ κοντά στη μονάδα. Μεγάλη σπουδαιότητα σ'αυτό παίζουν νόμοι που εμφανίζονται ως το αποτέλεσμα της επίδρασης μεγάλου πλήθους ανεξαρτήτων μεταξύ τους τυχαίων μεταβλητών. Ιδιαίτερη σημασία ανάμεσα σ'αυτούς τους νόμους είναι ο νόμος των μεγάλων αριθμών που είναι μια ενδιαφέρουσα συνέπεια της ανισότητας του Tschebyschev.

Θεωρούμε τις ανεξάρτητες τυχαίες μεταβλητές $X_1, X_2, \ldots, X_v$, με μέσες τιμές $\mu_1, \mu_2, \ldots, \mu_v$, και διασπορές $\sigma_1^2, \sigma_2^2, \ldots, \sigma_v^2$ μικρότερες από ένα φράγμα β^2. Εάν $A=\frac{1}{v}(\mu_1+\mu_2+\ldots+\mu_v)$ είναι ο αριθμητικός μέσος όρος

των μέσων τιμών, τότε εφαρμόζοντας την ανισότητα του Tschebyschev έχουμε:

$$P(\left|\frac{1}{\nu}\sum_{i=1}^{\nu} xi - A\right| < \varepsilon) \geq 1 - \frac{\beta^2}{\nu\varepsilon^2} \quad \forall \varepsilon > 0$$

Δηλαδή, ο νόμος των μεγάλων αριθμών με τη βοήθεια της ανισότητας του Tschebyschev λέγει: Για πάρα πολύ μεγάλα ν και με πιθανότητα που προσεγγίζει οσοδήποτε τη μονάδα, η αριθμητική μέση τιμή Α των μέσων τιμών ν τυχαίων μεταβλητών διαφέρει κατά απόλυτη τιμή λιγότερο του ε από την αριθμητική μέση τιμή των ν τυχαίων μεταβλητών.

Ακόμη, ο νόμος των μεγάλων αριθμών δίνεται και με τη βοήθεια του Bernoulli :

Θεωρούμε p την πιθανότητα εμφάνισης του γεγονότος Γ και έστω ότι κατά ν ανεξάρτητα πειράματα το γεγονός Γ εμφανίστηκε ν_i φορές, τότε $\forall \varepsilon > 0$ ισχύει :

$$P(\left|\frac{\nu_i}{\nu} - p\right| < \varepsilon) \geq 1 - \frac{1}{4\varepsilon^2 \nu}$$

Δηλαδή, ο νόμος των μεγάλων αριθμών κατά τον Bernoulli λέγει:

Για πάρα πολύ μεγάλα *ν* και με πιθανότητα που προσεγγίζει οσοδήποτε τη μονάδα, η σχετική συχνότητα εμφάνισης του γεγονότος Γ διαφέρει κατά απόλυτη τιμή λιγότερο του *ε* από την πιθανότητα p εμφάνισης του γεγονότος.

1.11. Μερικές Βασικές Κατανομές

Μερικές από τις κατανομές είναι πολύ σημαντικές για τις πρακτικές εφαρμογές. Τέτοιες είναι η διωνυμική κατανομή, η κατανομή του Poisson, η κανονική κατανομή, κ.ά.

Η διωνυμική κατανομή, που ονομάζεται και κατανομή του Bernoulli, εφαρμόζεται σ'όλα τα προβλήματα που βασίζονται σε πειράματα τύχης που επαναλαμβάνονται πολλές φορές και η

πιθανότητα να εμφανιστεί ένα γεγονός Γ δεν αλλάζει από πείραμα σε πείραμα, δηλαδή τα αποτελέσματα των διαφόρων πειραμάτων είναι ανεξάρτητα. Π.χ., η ρίψη ενός νομίσματος, ενός ζαριού, το τράβηγμα ενός κλήρου και στη συνέχεια η επανατοποθέτηση στην κληρωτίδα για νέο πάλι τράβηγμα, όταν επαναλαμβάνονται πολλές φορές, είναι πειράματα με ανεξάρτητα αποτελέσματα και ονομάζονται ανεξάρτητες δοκιμές ή δοκιμές Bernoulli.

Εάν p είναι η πιθανότητα να πραγματοποιηθεί ένα γεγονός Γ σε μια μόνο δοκιμή Bernoulli, τότε η πιθανότητα να μην πραγματοποιηθεί το γεγονός Γ είναι q=1-p. Η πιθανότητα p πραγματοποίησης του γεγονότος Γ λέγεται και πιθανότητα επιτυχίας, ενώ η πιθανότητα q μη πραγματοποίησης του γεγονότος Γ λέγεται και πιθανότητα αποτυχίας. Η πιθανότητα να πραγματοποιηθεί το γεγονός Γ ακριβώς *κ* φορές σε *ν* πειράματα, δηλαδή να έχουμε *κ* επιτυχίες και *ν-κ* αποτυχίες δίνεται από τη συνάρτηση πιθανότητας:

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

όπου η τυχαία μεταβλητή Χ παριστάνει το πλήθος των επιτυχιών σε *ν* πειράματα και *κ*=0,1,2,…,*ν*.

Η παραπάνω διακριτή συνάρτηση πιθανότητας f(x)=P(X=κ) καλείται διωνυμική κατανομή, γιατί για *κ*=0,1,2,…,*ν* δίνει τους όρους του αναπτύγματος του διωνύμου του Νεύτωνα:

$$(q+p)^{\nu}=q^{\nu}+\binom{\nu}{1} q^{\nu-1}p+\binom{\nu}{2} q^{\nu-2}p^2+\ldots+p^{\nu}=\sum_{\kappa=0}^{\nu}\binom{\nu}{\kappa} p^{\kappa}q^{\nu-\kappa}$$

Επίσης, καλείται και κατανομή του Bernoulli ή κατανομή του Νεύτωνα.

Η διωνυμική κατανομή έχει μέση τιμή μ=*ν*p, διασπορά $\sigma^2 = \nu pq$ και τυπική απόκλιση $\sigma=\sqrt{\nu pq}$.

Η διωνυμική κατανομή είναι πολύ χρήσιμη για μικρές τιμές των *ν* και *κ*. Για μεγάλες τιμές όμως ο υπολογισμός γίνεται επίπονος και γι'αυτό αντικαθίσταται, ανάλογα με το προς επίλυση πρόβλημα, με άλλες κατανομές (π.χ., κατανομή Gauss ή κατανομή Poisson, κ.λπ.)

1.11.1. Η Κατανομή του Poisson

Η κατανομή του Poisson εφαρμόζεται σε προβλήματα που βασίζονται σε δοκιμές (πειράματα τύχης) που ο αριθμός τους *v* είναι πάρα πολύ μεγάλος, ενώ η πιθανότητα p για την πραγματοποίηση ενός γεγονότος Γ είναι πάρα πολύ μικρή. Θα μπορούσαμε να πούμε ότι η κατανομή του Poisson είναι οριακή της διωνυμικής κατανομής, για πλήθος δοκιμών που τείνουν στο άπειρο, $v \to \infty$ και όταν η πιθανότητα p πραγματοποίησης ενός γεγονότος Γ τείνει στο μηδέν, ήτοι $p \to 0$, με μια πρόσθετη συνθήκη, ότι το γινόμενο της πιθανότητας πραγματοποίησης του γεγονότος Γ επί το πλήθος των πειραμάτων τύχης να είναι σταθερό, ήτοι p*v*=λ=σταθερό. Κατ'αυτόν τον τρόπο η πιθανότητα $p=P_v(X=\kappa)$ σε *v* πειράματα τύχης να πραγματοποιηθεί ένα γεγονός Γ *κ* φορές δίνεται από το όριο

$$\lim_{v \to \infty} P_v(X=\kappa) = \frac{\lambda^\kappa e^{-\lambda}}{\kappa!}$$

Ήτοι, εάν μία τυχαία μεταβλητή X μπορεί να πάρει τις τιμές 0, 1, 2, … με συνάρτηση πιθανότητας $f(\kappa)=P(X=\kappa)= \frac{\lambda^\kappa e^{-\lambda}}{\kappa!}$ *κ*=0, 1, 2, … , όπου λ είναι μία δεδομένη σταθερή μεγαλύτερη από το μηδέν λέμε ότι έχουμε μία κατανομή του Poisson, ή αλλιώς ότι η διακριτή τυχαία μεταβλητή X είναι κατανεμημένη κατά Poisson.

Τιμές της συνάρτησης πιθανότητας f(*κ*) μπορούν να υπολογιστούν από πίνακες στους οποίους δίνονται τιμές της ποσότητας $e^{-\lambda}$ για διάφορα λ ή απευθείας με τη βοήθεια των λογαρίθμων.

Η κατανομή του Poisson καθορίζεται πλήρως όταν είναι γνωστή η ποσότητα λ.

Βασικές ιδιότητες που χαρακτηρίζουν την κατανομή του Poisson είναι ότι η μέση τιμή της ισούται με λ, ήτοι μ=λ=pv, καθώς επίσης και η διασπορά της έχει την ίδια τιμή, ήτοι $\sigma^2=\lambda=pv$. Φυσικά,η τυπική της απόκλιση είναι $\sigma=\sqrt{\lambda}=\sqrt{pv}$.

Παλιότερα, η περιοχή εφαρμογής της κατανομής του Poisson περιοριζόταν σε σπάνια γεγονότα, π.χ., θανάτους από σπάνιες ασθένειες, παιδικές αυτοκτονίες κ.λπ. Τις τελευταίες δεκαετίες όμως η περιοχή εφαρμογής της κατανομής του Poisson αυξήθηκε σημαντικά. Ήδη, χρησιμοποιείται ευρύτατα στη βιολογία, μετεωρολογία, στις τηλεπικοινωνίες, στο στατιστικό έλεγχο ποιότητας διαφόρων αγαθών, κ.λπ.

Ακόμη, η κατανομή του Poisson χρησιμεύει για να προσεγγίσει τη διωνυμική κατανομή. Πράγματι εάν σε μία διωνυμική κατανομή το *v* είναι μεγαλο και η πιθανότητα επιτυχίας p να πραγματοποιηθεί ένα γεγονός Γ είναι πολύ μικρή, δηλαδή η πιθανότητα αποτυχίας q=1-p είναι περίπου ίση με τη μονάδα, τότε λέμε ότι έχουμε να κάνουμε με ένα σπάνιο γεγονός.

Στην πράξη, χαρακτηρίζουμε ένα γεγονός σπάνιο, εάν αναφερόμαστε τουλάχιστον σε 50 πειράματα τύχης ($v \geq 50$) και συγχρόνως το γινόμενο p*v* είναι μικρότερο του 5 ήτοι, p*v*<5. Στις περιπτώσεις αυτές αντί της διωνυμικής κατανομής μπορούμε να χρησιμοποιήσουμε την κατανομή του Poisson και να έχουμε μία πάρα πολύ καλή προσέγγιση.

1.11.2. Κανονική Κατανομή ή Κατανομή του Gauss

Η κανονική κατανομή είναι μία από τις πιο σημαντικές κατανομές του λογισμού των πιθανοτήτων και ονομάζεται και κατανομή του Gauss, από το όνομα του εισηγητή της ο οποίος τη βρήκε και τη χρησιμοποίησε σε μια προσπάθειά του να προσαρμόσει αποτελέσματα γεωδαιτικών μετρήσεων.

Η κανονική κατανομή προκύπτει από τη διωνυμική κατανομή, αν αυξηθεί ο αριθμός *v* των πειραμάτων τύχης προς το άπειρο, ενώ η πιθανότητα p πραγματοποίησης του γεγονότος Γ, δηλαδή η πιθανότητα επιτυχίας παραμένει σταθερή ίση με $\frac{1}{2}$ ήτοι p=$\frac{1}{2}$.

Ενώ η διωνυμική κατανομή ορίζεται μόνο για ακέραιες τιμές της τυχαίας μεταβλητής, η κανονική κατανομή ισχύει για συνεχείς μεταβλητές, δηλαδή εφαρμόζεται σε απείρως πυκνό πλήθος τιμών της τυχαίας μεταβλητής.

Η συνάρτηση πυκνότητας της κανονικής κατανομής είναι :

$$f(\chi)=\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \qquad -\infty < x < \infty$$

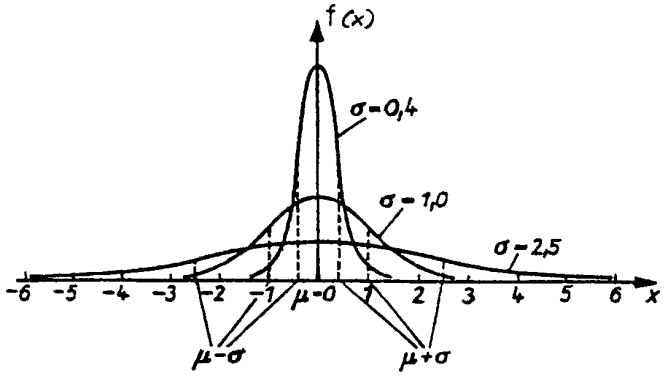
όπου μ και σ είναι σταθερές και μάλιστα ίσες αντίστοιχα με τη μέση τιμή και την τυπική απόκλιση.

Η μέση τιμή μ είναι :

$$\mu=\int_{-\infty}^{+\infty} x\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}\,dx$$

και η διασπορά σ^2 είναι :

$$\sigma^2=\int_{-\infty}^{+\infty} (x-\mu)^2\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}\,dx$$



Σχήμα 19. Γραφική παράσταση της συνάρτησης πυκνότητας f(x) για διάφορες τιμές της διασποράς

Η μέση τιμή και η διασπορά περιγράφουν πλήρως την κανονική κατανομή. Στο σχήμα 19 δίνεται η γραφική παράσταση της συνάρτησης πυκνότητας f(x) της κανονικής κατανομής για διάφορες τιμές της διασποράς. Οι καμπύλες έχουν τη μορφή κώδωνος. Η κορυφή κάθε

καμπύλης βρίσκεται στη μέση τιμή μ. Από αυτή την τιμή (θέση) η καμπύλη πέφτει συμμετρικά προς τις δυο πλευρές και πλησιάζει ασυμπτωτικά τον άξονα των x. Στην απόσταση $\pm$σ από τη μέση τιμή μ βρίσκονται τα σημεία καμπής της καμπύλης. Από το παραπάνω σχήμα φαίνεται η επίδραση της αύξησης της διασποράς στη μορφή της καμπύλης. Όσο αυξάνεται η τυπική απόκλιση σ (δηλ., η τετραγωνική ρίζα της διασποράς) οι καμπύλες γίνονται χαμηλότερες και ευρύτερες. Η συνάρτηση κατανομής της κατανομής του Gauss είναι:

$$F(x)=P(X\leq x)=\int_{-\infty}^{x} f(x)dx=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{26^2}}\,dt$$

και τότε λέμε ότι η τυχαία μεταβλητή X είναι κανονική ή κανονικά κατανενημένη με μέση τιμή μ και διασπορά σ^2.

Όμως, για μια τυχαία μεταβλητή X, που είναι κανονική ή κατανεμημένη κατά Gauss, όπως λέμε, με δεδομένη μέση τιμή μ και διασπορά σ^2, ο υπολογισμός επί μέρους τιμών της συνάρτησης πυκνότητας της κατανομής f(x) είναι επίπονος. Γι'αυτό η κατανομή της X ανάγεται σε μία κατανομή Gauss μιας τυποποιημένης, όπως ονομάζεται, μεταβλητής Z που έχει μέση τιμή μ=0 και διασπορά σ^2=1. Η συνάρτηση πυκνότητας της τυποποιημένης μεταβλητής Z είναι τότε:

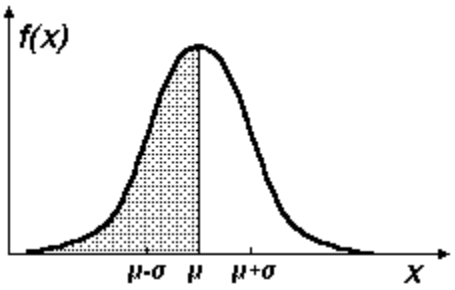
$$f(z)=\frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}$$

και ονομάζεται κανονικοποιημένη κατανομή Gauss ή τυποποιημένη κανονική κατανομή ή απλώς κανονική κατανομή.

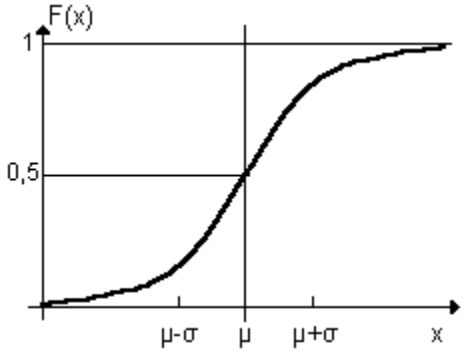
Η μετάβαση από κατανομή Gauss με μέση τιμή μ και διασπορά σ^2 στην κανονική κατανομή γίνεται με το μετασχηματισμό

$$Z=\frac{X-\mu}{\sigma} \qquad (M)$$

Η παραπάνω συνάρτηση κατανομής F(x) ονομάζεται ολοκλήρωμα Gauss ή ολοκλήρωμα σφάλματος. Παριστάνει το εμβαδό της επιφάνειας κάτω από την καμπύλη f(x) με όρια $-\infty$ και x (σχήμα 20). Η συνάρτηση F(x) έχει ως ασύμπτωτες τον άξονα των x και την ευθεία F(x)=1 και σημείο καμπής στο σημείο x=μ (σχήμα 21).



Σχήμα 20. Συνάρτηση πυκνότητας της κατανομής Gauss

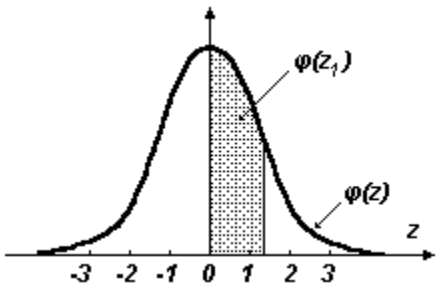


Σχήμα 21. Συνάρτηση κατανομής της κατανομής Gauss

Αν ληφθεί υπόψη η συμμετρία της συνάρτησης πυκνότητας της κανονικής κατανομής, δηλαδή η συμμετρία της f(x), η συνάρτηση κατανομής F(x) γράφεται σε κανονική μορφή για μ=0 και σ^2=1 ως εξής :

$$\Phi(x)=\int_0^z \phi(u)du = \frac{1}{\sqrt{2\pi}}\int_0^z e^{-\frac{u^2}{2}}du$$

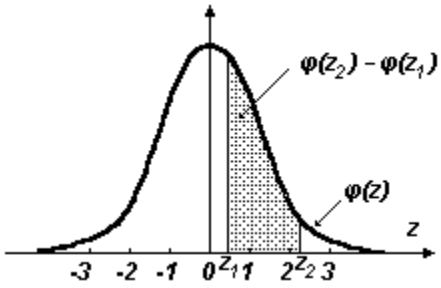
Οι τιμές της Φ(z) καταχωρίζονται σε πίνακες. Κάθε κατανομή Gauss με μέση τιμή μ και διασπορά σ^2 συνδέεται με τη Φ(z) με τον παραπάνω μετασχηματισμό (Μ). Στο σχήμα 22 δίνεται η γεωμετρική σημασία της Φ(z).



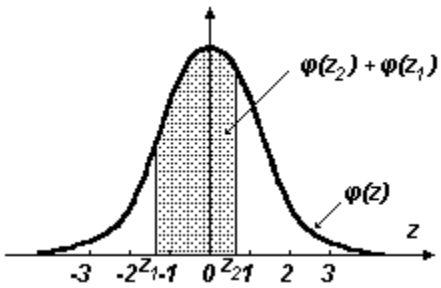
Σχήμα 22. Γεωμετρική σημασία της συνάρτησης Φ(z)

Από το μετασχηματισμό (Μ) μπορούμε να βρούμε τα z_1, z_2 που αντιστοιχούν στις δύο τιμές x_1 και x_2 της τυχαίας μεταβλητής Χ και να υπολογίσουμε την αντίστοιχη επιφάνεια. Για τον υπολογισμό διακρίνουμε δύο βασικές περιπτώσεις:

1. Εάν οι τιμές των z_1 και z_2 βρίσκονται στα δεξιά του μηδενός και είναι $z_2>z_1$, τότε η αντίστοιχη επιφάνεια είναι ίση με τη διαφορά $\Phi(z_2)$-$\Phi(z_1)$. Το ανάλογο ισχύει εάν οι δύο τιμές z_1 και z_2 βρίσκονται στα αριστερά του μηδενός (σχήμα 23).
2. Εάν η τιμή z_1 βρίσκεται αριστερά και η τιμή z_2 δεξιά του μηδενός,τότε η αντίστοιχη επιφάνεια δίνεται από το άθροισμα $\Phi(z_2)+\Phi(z_1)$ (σχήμα 24). Και στις δυο περιπτώσεις το μέτρο της υπολογισθείσας επιφανείας δίνει την πιθανότητα η τιμή της τυχαίας μεταβλητής να βρίσκεται μεταξύ των x_1 και x_2.

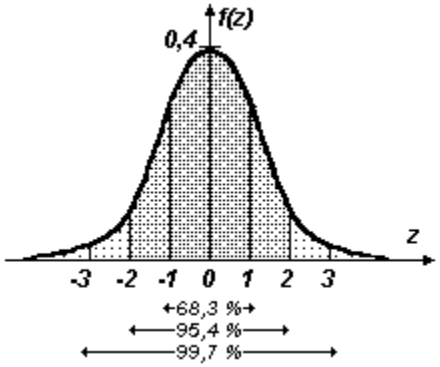


Σχήμα 23. Επιφάνεια ίση με $\Phi(z_2)$-$\Phi(z_1)$



Σχήμα 24. Επιφάνεια ίση με $\Phi(z_2)+\Phi(z_1)$

Μία γραφική παράσταση της συνάρτησης πυκνότητας της τυποποιημένης μεταβλητής Z, που πολλές φορές ονομάζεται τυπική ή τυποποιημένη κανονική καμπύλη δίνεται στο σχήμα 25. Στο σχήμα είναι σκιασμένες οι επιφάνειες που αντιστοιχούν σε τυπικές αποκλίσεις 1,2, και 3 από τη μέση τιμή. Οι επιφάνειες αυτές περιλαμβάνονται μεταξύ των τιμών z_1 και z_2 όπου, z_1 παίρνει τις τιμές -1,-2,-3 και z_2 τις τιμές 1,2,3 αντίστοιχα. Δηλαδή, οι επιφάνειες αυτές περιλαμβάνονται μεταξύ z=-1 και z=+1, z=-2 και z=+2, z=-3 και z=+3 και αποτελούν αντίστοιχα τα 68.3%,95.4% και 99.7% της ολικής επιφανείας που είναι ίση με τη μονάδα. Τα παραπάνω είναι ισοδύναμα αντίστοιχα με τις σχέσεις :



Σχήμα 25. Γραφική παράσταση της συνάρτησης πυκνότητας της τυποποιημένης μεταβλητής z

$$P(-1 \le Z \le 1) = 0.683\,,\ P(-2 \le Z \le 2) = 0.954\,,\ P(-3 \le Z \le 3) = 0.997$$

1.11.3 Οι κατανομές γάμα, χ^2, t, F.

Για στατιστικές εκτιμήσεις παραμέτρων πληθυσμού, καθώς και για τον έλεγχο στατιστικών υποθέσεων, χρησιμοποιούνται οι παρακάτω κατανομές:

Η κατανομή γάμα: Μια τυχαία μεταβλητή x ακολουθεί την κατανομή γάμα, αν η συνάρτηση πυκνότητας είναι :

$$f(x) = \begin{cases} \dfrac{x^{a-1} e^{-x/b}}{b^a \Gamma(a)} & x>0 \\ 0 & x\le 0 \end{cases}$$

όπου a, b > 0

Γ(a) είναι η **συνάρτηση γάμα**, που ορίζεται από τη σχέση :

$$\Gamma(a)=\int_0^\infty t^{a-1} e^{-t}\, dt \qquad a>0 .$$

Η μέση τιμή μ και η διασπορά σ^2 είναι, αντίστοιχα :

$\mu = ab$ **$\sigma^2 = ab^2$**

Η κατανομή χ^2: Η κατανομή χ^2, με ν βαθμούς ελευθερίας, έχει συνάρτηση πυκνότητας :

$$f(x) = \begin{cases} \dfrac{x^{\nu/2-1} e^{-x/2}}{2^{\nu/2} \Gamma(\nu/2)} & x>0 \\ 0 & x\le 0 \end{cases}$$

με μέση τιμή $\mu = \nu$ και διασπορά $\sigma^2 = 2\nu$.

Η κατανομή t του Student: Μια τυχαία μεταβλητή t με συνάρτηση πυκνότητας :

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}\,(1+t^2/\nu)e^{-(\nu+1)/2} \qquad -\infty < t < \infty$$

ακολουθεί την **κατανομή t με ν βαθμούς ελευθερίας.** Η κατανομή t έχει μέση τιμή μ = 0 και διασπορά $\sigma^2 = \nu/(\nu-2)$, με ν>2. Αν το ν είναι μεγάλο (ν≥30), η f(t) προσεγγίζει την κανονική (γκαουσιανή) κατανομή.

Η κατανομή F: Μια τυχαία μεταβλητή u ακολουθεί την **κατανομή F με ν_1 και ν_2 βαθμούς ελευθερίας**, αν έχει συνάρτηση πυκνότητας :

$$f(u)=\begin{cases} \dfrac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)}\nu_1^{\nu_1/2}\nu_2^{\nu_2/2}\;u^{(\nu_1/2)-1}(\nu_2+\nu_1 u)^{-(\nu_1+\nu_2)/2} & u>0 \\ 0 & u\le 0 \end{cases}$$

η μέση τιμή μ και η διασπορά σ^2 δίνονται αντίστοιχα από τις σχέσεις :

$$\mu = \frac{\nu_2}{\nu_2-2} \qquad \nu_2>2$$

και

$$\sigma^2 = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-4)(\nu_2-2)^2} \qquad \nu_2>4$$

1.12. Στατιστικές Εκτιμήσεις - Διαστήματα Εμπιστοσύνης

Αν η εκτίμηση μιας παραμέτρου Μ του πληθυσμού (για παράδειγμα της μέσης τιμής) εκφράζεται με ένα μόνο αριθμό, έχουμε μια **σημειακή εκτίμηση** της παραμέτρου. Αν όμως ο αριθμός αυτός

συνοδεύεται και με ένα διάστημα τιμών, μέσα στο οποίο πιστεύουμε ότι βρίσκεται η παράμετρος του πληθυσμού, τότε έχουμε μια **εκτίμηση διαστήματος** της παραμέτρου.

Κατά τη διαδικασία της εκτίμησης της παραμέτρου του πληθυσμού, υπεισέρχεται η **στατιστική συνάρτηση** (ή **δειγματοσυνάρτηση**) **S,** με τη βοήθεια της οποίας υπολογίζεται μια μέση τιμή M_s, που εξυπηρετεί στον υπολογισμό της παραμέτρου του πληθυσμού.

Παράδειγμα στατιστικής συνάρτησης S είναι η δειγματική μέση τιμή $\bar{x}$, που είναι ο μέσος όρος των τιμών του δείγματος x_1, x_2, ..., x_v, και που εξυπηρετεί στην εκτίμηση της μέσης τιμής M του πληθυσμού.

Αν η S ακολουθεί, έστω και κατά προσέγγιση, την κανονική κατανομή, τότε η εκτίμηση της παραμέτρου του πληθυσμού εκφράζεται από τη **μέση τιμή M_s** της S και τη **μέση απόκλιση σ_s**. Η μέση απόκλιση σ_s συνδέεται με την τυπική απόκλιση σ του πληθυσμού με τη σχέση :

$$\sigma_s = \sigma / \sqrt{v}$$

όπου v είναι το πλήθος των στοιχείων (μέγεθος) του δείγματος.

Αν η στατιστική συνάρτηση S είναι η δειγματική μέση τιμή $\bar{x}$, τότε ισχύει :

$$M_s = \bar{x}$$

Έχοντας λοιπόν προσδιορίσει τις ποσότητες M_s και σ_s, και θεωρώντας ότι η S ακολουθεί μια κανονική κατανομή, εκτιμούμε ότι η παράμετρος M του πληθυσμού βρίσκεται μέσα στο διάστημα $[M_s-\sigma_s, M_s+\sigma_s]$ με πιθανότητα 68.27%, στο διάστημα $[M_s-2\sigma_s, M_s+2\sigma_s]$ με πιθανότητα 95.45% και στο διάστημα $[M_s-3\sigma_s, M_s+3\sigma_s]$ με πιθανότητα 99.73%. Τα διαστήματα αυτά καλούνται 68.27%, 95.45%, 99.73% **διαστήματα εμπιστοσύνης** για την εκτίμηση της παραμέτρου του πληθυσμού. Τα άκρα των διαστημάτων αυτών είναι τα **όρια εμπιστοσύνης**. Οι αριθμοί 1, 2, 3, που προσδιορίζουν τα όρια των διαστημάτων εμπιστοσύνης, ως συντελεστές στην ποσότητα $\pm\sigma_s$, είναι οι **κρίσιμες τιμές** και συμβολίζονται με z_c. Για ένα διάστημα εμπιστοσύνης 95% το z_c είναι 1.96 και για 99% είναι 2.58. Οι τιμές z_c προσδιορίζονται από την καμπύλη Φ(z), που είναι το εμβαδόν του χωρίου που

περικλείεται από την κανονική κατανομή από 0 ως z. Τα ποσοστά 95%, 99%, κ.λπ., ονομάζονται **συντελεστές εμπιστοσύνης**.

Επειδή η τυπική απόκλιση **σ** του πληθυσμού συνήθως δεν είναι γνωστή, στις σχέσεις με τις οποίες προσδιορίζονται η μέση απόκλιση και τα διαστήματα εμπιστοσύνης, υπεισέρχεται η τυπική απόκλιση **s** του δείγματος, αντί του **σ**.

1.12.1. Διαστήματα εμπιστοσύνης για μέσες τιμές

Όταν η στατιστική συνάρτηση S είναι η δειγματική μέση τιμή $\bar{x}$, και **το μέγεθος του δείγματος είναι μεγάλο (v≥30)**, τότε η εκτίμηση για τη μέση τιμή Μ του πληθυσμού είναι :

$$\bar{x} \pm z_c \, s / \sqrt{v}$$

υποθέτοντας έναν άπειρο πληθυσμό.

Για μικρά δείγματα (v<30), τα διαστήματα εμπιστοσύνης καθορίζονται με βάση την κατανομή t, με v-1 βαθμούς ελευθερίας. Αν λοιπόν θέλουμε, για παράδειγμα, να καθορίσουμε τα όρια του διαστήματος εμπιστοσύνης 95%, θα πρέπει, με τη βοήθεια πινάκων, να προσδιορίσουμε την κρίσιμη τιμή t_c που αντιστοιχεί σε συντελεστή εμπιστοσύνης 97.5% (ώστε το εμβαδόν της t να χωριστεί σε τρία τμήματα που να καταλαμβάνουν το 2.5%, το 95% και το 2.5% αντίστοιχα), οπότε η εκτίμηση για τη μέση τιμή είναι :

$$\bar{x} \pm t_{.975} \, s / \sqrt{(v-1)}$$

Γενικά, η μέση τιμή για ένα διάστημα εμπιστοσύνης είναι :

$$\bar{x} \pm t_c \, s / \sqrt{(v-1)}$$

όπου t_c το αντίστοιχο όριο εμπιστοσύνης

Παράδειγμα 1 *(Spiegel, 1977) : Σε δείγμα 100 κιβωτίων φρούτων, από ένα πολύ μεγαλύτερο πληθυσμό, βρέθηκε αριθμητική μέση τιμή βάρους κιβωτίου* $\bar{x}$ *= 67.45 Kg με τυπική απόκλιση s = 2.93 Kg. Να*

υπολογιστούν: α) το 95%, β) το 99% διάστημα εμπιστοσύνης για τη μέση τιμή του βάρους των κιβωτίων.

α) Για 95% συντελεστή εμπιστοσύνης έχουμε $\mathbf{z_c = 1.96}$. Επομένως το μέσο βάρος είναι $\mathbf{67.45 \pm 1.96\ (2.93 / \sqrt{100}) = (67.45 \pm 0.57)\ kg}$, με διάστημα εμπιστοσύνης 95%.

β) Για συντελεστή εμπιστοσύνης 99% έχουμε $\mathbf{z_c = 2.58}$. Το μέσο βάρος είναι $\mathbf{67.45 \pm 2.58\ (2.93 / \sqrt{100}) = (67.45 \pm 0.76)\ kg}$, με διάστημα εμπιστοσύνης 99%.

Παράδειγμα 2 *(Spiegel, 1977) : Δέκα μετρήσεις της διαμέτρου μιας σφαίρας έδωσαν μέση τιμή* $\bar{x}$*=4.38 cm και τυπική απόκλιση s=0.06 cm. Να εκτιμηθεί η πραγματική διάμετρος της σφαίρας για 95% διάστημα εμπιστοσύνης.*

Επειδή v<30, η μέση τιμή $\bar{x}$ ακολουθεί μια κατανομή t με 10-1=9 βαθμούς ελευθερίας. Για 95% συντελεστή εμπιστοσύνης έχουμε κρίσιμη τιμή $\mathbf{t_c = t_{.975} = 2.26}$.

Οπότε, για διάστημα εμπιστοσύνης 95%, η διάμετρος της σφαίρας είναι $\mathbf{4.38 \pm 2.26\ (0.06 / \sqrt{(10-1)}) = (4.38 \pm 0.0452)\ cm}$.

1.12.2. Διαστήματα εμπιστοσύνης για αναλογίες

Αν η στατιστική συνάρτηση S παριστάνει την αναλογία επιτυχιών σε δείγμα μεγέθους $v \geq 30$, που προέρχεται από ένα διωνυμικό πληθυσμό στον οποίο p είναι η αναλογία των επιτυχιών, το διάστημα εμπιστοσύνης για το p είναι το $[P-z_c s,\ P+z_c s]$, όπου P η αναλογία επιτυχιών στο δείγμα μεγέθους v.

Αποδεικνύεται ότι η τυπική απόκλιση δείγματος s είναι ίση με $\sqrt{(p(1-p)/v)}$, οπότε τα όρια εμπιστοσύνης για την αναλογία του πληθυσμού είναι :

$$\mathbf{P \pm z_c \sqrt{(p(1-p)/v)}}$$

Το p, στην υπόρριζο ποσότητα, μπορεί να προσδιοριστεί από την εκτίμηση P για το δείγμα.

1.12.3. Διαστήματα εμπιστοσύνης για διαφορές και αθροίσματα

Αν S_1 και S_2 είναι στατιστικές συναρτήσεις που ακολουθούν κατά προσέγγιση μια κανονική κατανομή, τα όρια εμπιστοσύνης για τη διαφορά των παραμέτρων των πληθυσμών που αντιστοιχούν στις S_1 και S_2 είναι :

$$S_1 - S_2 \pm z_c\sqrt{(\sigma_{S1}^2 + \sigma_{S2}^2)}$$

και για το άθροισμα των παραμέτρων είναι :

$$S_1 + S_2 \pm z_c \sqrt{(\sigma_{S1}^2 + \sigma_{S2}^2)}$$

Αν οι S_1, S_2 είναι οι αριθμητικές τιμές $\bar{x}_1$, $\bar{x}_2$ των δειγμάτων από δύο άπειρους πληθυσμούς, τότε τα όρια εμπιστοσύνης για τη διαφορά και το άθροισμα των μέσων τιμών είναι :

$$\bar{x}_1 - \bar{x}_2 \pm z_c \sqrt{(s_1^2/v_1 + s_2^2/v_2)}$$

και

$$\bar{x}_1 + \bar{x}_2 \pm z_c \sqrt{(s_1^2/v_1 + s_2^2/v_2)}$$

αντίστοιχα, όπου σ_1, σ_2 είναι οι τυπικές αποκλίσεις των δειγμάτων.

1.13. Έλεγχοι υποθέσεων και επίπεδα σημαντικότητας

Συνηθίζουμε να εκτιμούμε τις παραμέτρους πληθυσμών, ή να κάνουμε υποθέσεις γι'αυτούς, μελετώντας δείγματα των πληθυσμών. Από τη μελέτη διαφορετικών δειγμάτων, συνάγουμε γενικά διαφορετικά αποτελέσματα. Το ερώτημα που τίθεται συχνά είναι το αν τα αποτελέσματα από δύο διαφορετικά δείγματα αντιστοιχούν σε δύο διαφορετικούς πληθυσμούς, ή αν αντιστοιχούν στον ίδιο πληθυσμό,

οπότε οι διαφορές στα αποτελέσματα των δειγμάτων οφείλονται σε τυχαίες αιτίες (στατιστικά σφάλματα). Αυτή η τελευταία υπόθεση (διαφορετικά δείγματα από τον ίδιο πληθυσμό, άρα οι μετρούμενες τιμές είναι κατ'ουσίαν ίδιες, εφόσον αντιστοιχούν στον ίδιο πληθυσμό) ονομάζεται **μηδενική υπόθεση** και συμβολίζεται με H_0. Μια υπόθεση που είναι ασυμβίβαστη με την H_0 (δηλαδή τα δύο δείγματα δεν ανήκουν στον ίδιο πληθυσμό, άρα δίνουν διαφορετικές παραμέτρους) είναι μια **εναλλακτική υπόθεση** και συμβολίζεται με H_1.

Για να αποφασίσουμε, αν θα δεχτούμε ή θα απορρίψουμε μια υπόθεση, ή για να αποφανθούμε για το αν τα αποτελέσματα των δειγμάτων διαφέρουν **σημαντικά** από αυτά που προβλέπει η υπόθεση, χρησιμοποιούμε διάφορες διαδικασίες, μεθόδους και κανόνες, που καλούνται γενικά **έλεγχοι υποθέσεων**, ή **έλεγχοι σημαντικότητας** ή **κανόνες απόφασης**.

Αν, με βάση τα δεδομένα που έχουμε, απορρίψουμε μια υπόθεση που στην πραγματικότητα αληθεύει και θα έπρεπε να τη δεχτούμε, **κάνουμε ένα σφάλμα τύπου** I (ή **πρώτου είδους**). Αντίθετα, αν δεχτούμε μια υπόθεση που θα έπρεπε να απορρίψουμε, τότε κάνουμε ένα **σφάλμα τύπου** II (ή **δεύτερου είδους**). Και στις δύο περιπτώσεις έχουμε πάρει μια λαθεμένη απόφαση.

Ο έλεγχος της υπόθεσης, γενικά δεν εξασφαλίζει την απόλυτη βεβαιότητα για το αν η υπόθεση ισχύει ή δεν ισχύει. Θα πρέπει λοιπόν να εισαγάγουμε ένα ποσοστό αβεβαιότητας ως προς την ορθότητα της απόφανσής μας.

Ονομάζουμε λοιπόν **επίπεδο** ή **στάθμη σημαντικότητας**, τη μέγιστη πιθανότητα με την οποία δεχόμαστε να κάνουμε σφάλμα τύπου I. Συνήθως, ως επίπεδο σημαντικότητας χρησιμοποιείται το 0.01 ή το 0.05. Αν λοιπόν πάρουμε, για έναν έλεγχο υπόθεσης, επίπεδο σημαντικότητας 0.05 και απορρίψουμε την υπόθεση, τότε από τις 100 όμοιες περιπτώσεις αναμένεται να σφάλλουμε μόνο στις 5, ή, με άλλα λόγια, είμαστε 95% βέβαιοι ότι πήραμε τη σωστή απόφαση. Σε μια τέτοια περίπτωση λέμε ότι **η υπόθεση απορρίπτεται σε επίπεδο εμπιστοσύνης 0.05** ή **5%**.

1.13.1. Έλεγχος μηδενικής υπόθεσης με την κανονική κατανομή

Ας υποθέσουμε ότι η δειγματοληπτική κατανομή μιας στατιστικής συνάρτησης S ακολουθεί την κανονική κατανομή, με μέση τιμή M_s και απόκλιση σ_s. Αυτό σημαίνει ότι η παράμετρος $z=(S-M_s)/\sigma_s$ ακολουθεί τυπική κανονική κατανομή (μέση τιμή 0, διασπορά 1). Αν η υπόθεση είναι αληθής, τότε μπορούμε να είμαστε 95% βέβαιοι ότι η μετρούμενη τιμή z βρίσκεται στο διάστημα [-1.96, +1.96]. Αν όμως σε ένα δείγμα βρούμε μια τιμή z έξω από το διάστημα [-1.96, +1.96], τότε έχουμε ένα γεγονός που έχει μόλις 5% πιθανότητα να συμβεί, οπότε η τιμή z διαφέρει σημαντικά από αυτήν που περιμένουμε. Κατά συνέπεια, η αρχική υπόθεση για μέση τιμή M_s και τυπική απόκλιση σ_s, θα πρέπει μάλλον να απορριφθεί.

Οι τιμές του z έξω από το διάστημα [-1.96, +1.96] αποτελούν την **κρίσιμη περιοχή**, ή **περιοχή απόρριψης**, ή **περιοχή σημαντικότητας** και καλύπτει το 5% του εμβαδού της κανονικής κατανομής. Οι τιμές του z μέσα στο διάστημα [-1.96, +1.96] αποτελούν την **περιοχή αποδοχής** της υπόθεσης, που καλύπτει το 95% του εμβαδού της κανονικής κατανομής.

Μπορούμε λοιπόν να διατυπώσουμε τον εξής κανόνα απόφασης :

Αν η τιμή z από ένα δείγμα βρίσκεται έξω από το διάστημα [-1.96,+1.96], τότε απορρίπτουμε την υπόθεση σε επίπεδο σημαντικότητας 0.05 (5%).

Αντίθετα, αν η z βρίσκεται μέσα στο διάστημα [-1.96,+1.96], τότε δεχόμαστε την υπόθεση, ή, αν θέλουμε, δεν παίρνουμε απόφαση.

Ο κανόνας απόφασης που διατυπώσαμε ισχύει για **δίπλευρο έλεγχο**, στον οποίο μας ενδιαφέρουν οι ακραίες τιμές της S, δεξιά και αριστερά από το 0. Συχνά όμως ενδιαφερόμαστε για ακραίες τιμές μόνο στο ένα άκρο της δειγματικής κατανομής της S, όπως για παράδειγμα όταν εξετάζουμε κατά πόσον η παράμετρος ενός πληθυσμού είναι μεγαλύτερη από μια τιμή α. Αυτό είναι διαφορετικό από το να ελέγξουμε την υπόθεση αν η παράμετρος είναι διάφορη του α. Ένας τέτοιος έλεγχος καλείται **μονόπλευρος**. Η αντίστοιχη κρίσιμη περιοχή είναι στο

ένα άκρο της περιοχής και έχει εμβαδόν ίσο με το επίπεδο σημαντικότητας.

1.13.2. Έλεγχοι υποθέσεων για μεγάλα δείγματα από άπειρους πληθυσμούς.

Για μεγάλα δείγματα, πολλές στατιστικές συναρτήσεις ακολουθούν κανονικές κατανομές (ή περίπου κανονικές) με μέση τιμή M_s και τυπική απόκλιση σ_s.

Είναι επομένως δυνατό να γίνουν έλεγχοι υποθέσεων με βάση την κανονική κατανομή. Παρακάτω παρουσιάζονται ορισμένες περιπτώσεις ελέγχου υπόθεσης με πρακτικό ενδιαφέρον.

Μέσες τιμές : Στην περίπτωση αυτή έχουμε μια τυποποιημένη μεταβλητή :

$$\mathbf{z = (\ \bar{x} - M) / (s / \sqrt{v})}$$

όπου $\bar{x}$ είναι η δειγματική μέση τιμή, M η μέση τιμή του πληθυσμού και σ η τυπική απόκλιση του πληθυσμού, που εκτιμάται από την τυπική απόκλιση του δείγματος. v είναι το πλήθος των στοιχείων του δείγματος.

Για να ελέγξουμε, για παράδειγμα, την υπόθεση ότι η μέση τιμή M είναι M = α (μηδενική υπόθεση), πραγματοποιούμε ένα δίπλευρο έλεγχο σε επίπεδο 0.05, ο οποίος αντιστοιχεί σε κρίσιμες τιμές $\pm$1.96 και εξετάζουμε αν ισχύει :

$$\mathbf{-1.96 \leq (\ \bar{x} - \alpha) / (s / \sqrt{v}) \leq 1.96}$$

Αν η παραπάνω ανίσωση ισχύει, τότε δεχόμαστε την υπόθεση M = α, σε επίπεδο σημαντικότητας 0.05.

Για να ελέγξουμε την υπόθεση M > α, χρησιμοποιούμε πάλι την υπόθεση M = α, και πραγματοποιούμε ένα μονόπλευρο έλεγχο, εξετάζοντας το κατά πόσον ισχύει η ανίσωση :

$$\mathbf{(\ \bar{x} - \alpha) / (s / \sqrt{v}) < 1.645}$$

Το 1.645 είναι η κρίσιμη τιμή του z για μονόπλευρο έλεγχο σε επίπεδο σημαντικότητας 0.05. Αν ισχύει η παραπάνω ανίσωση, θα πρέπει να απορριφθεί η υπόθεση $M>\alpha$.

Για να ελεγχθεί η υπόθεση $M<\alpha$, εξετάζουμε αν ισχύει η ανίσωση :

$$\mathbf{(\ \bar{x} - \alpha) / (s / \sqrt{v}) > -1.645}$$

Αν ισχύει, τότε απορρίπτεται η υπόθεση $M<\alpha$, σε επίπεδο σημαντικότητας 0.05.

Αναλογίες : Στην περίπτωση αυτή η στατιστική συνάρτηση είναι η p (αναλογία επιτυχιών σε δείγμα) και η μέση τιμή του πληθυσμού είναι η P. Ισχύει ότι $s = \sqrt{(p(1-p)/v)}$.

Η τυποποιημένη μεταβλητή είναι :

$$\mathbf{z = (p-P) / \sqrt{(p(1-p)/v)}}$$

Ο τρόπος εργασίας είναι ανάλογος με αυτόν για τις μέσες τιμές.

Διαφορές μέσων τιμών : Θεωρούμε ότι δεν υπάρχει διαφορά μέσων τιμών μεταξύ δύο πληθυσμών (μηδενική υπόθεση). Η τυποποιημένη μεταβλητή είναι :

$$\mathbf{z = (\ \bar{x}_1 - \bar{x}_2) / s_{1,2}}$$

$\bar{x}_1$ είναι η μέση τιμή του δείγματος μεγέθους v_1

$\bar{x}_2$ είναι η μέση τιμή του δείγματος μεγέθους v_2

$s_{1,2} = \sqrt{(s_1^2/v_1+s_2^2/v_2)}$

s_1, s_2 είναι οι τυπικές αποκλίσεις των δειγμάτων μεγέθους v_1, v_2, αντίστοιχα.

Διαφορές αναλογιών: Για να ελεγχθεί η μηδενική υπόθεση ότι δεν υπάρχει διαφορά μεταξύ των αναλογιών δύο πληθυσμών, χρησιμοποιούμε την τυποποιημένη μεταβλητή :

$$\mathbf{z = (p_1 - p_2) / s_{1,2}}$$

p_1, p_2 είναι οι αναλογίες επιτυχιών σε δύο δείγματα μεγέθους v_1, v_2 που προέρχονται από διαφορετικούς πληθυσμούς :

$\sigma_{1,2} = \sqrt{(p(1 - p) (1/v_1 + 1/v_2))}$

$p = (v_1p_1 + v_2p_2) / (v_1 + v_2)$

1.13.3. Έλεγχοι υποθέσεων για δείγματα μικρού μεγέθους

Όταν τα δείγματα είναι μικρού μεγέθους (v<30), οι έλεγχοι υποθέσεων πραγματοποιούνται με τη βοήθεια άλλων, μη κανονικών κατανομών, όπως η t, η F και η χ^2. Οι έλεγχοι που εφαρμόζονται στα μικρά δείγματα, μπορούν να εφαρμοστούν και για μεγάλα. Παρακάτω παρουσιάζονται κάποιες χαρακτηριστικές περιπτώσεις ελέγχου υπόθεσης.

Μέσες τιμές : Για να ελέγξουμε την υπόθεση ότι μια κανονική, ή κωδωνοειδής έστω, κατανομή έχει μέση τιμή M, χρησιμοποιούμε την τυποποιημένη κατανομή :

$$\mathbf{T = (\bar{x} - M) \sqrt{(v - 1)} / s}$$

που ακολουθεί την **κατανομή t του Student με v-1 βαθμούς ελευθερίας**. $\bar{x}$ είναι η μέση τιμή του δείγματος μεγέθους v και σ η τυπική απόκλιση του δείγματος. Όσο το v αυξάνει, τόσο η T προσεγγίζει την κανονική κατανομή z.

Διαφορές μέσων τιμών: Έστω δύο τυχαία δείγματα μεγέθους v_1 και v_2, που προέρχονται από δύο κανονικούς (ή περίπου κανονικούς) πληθυσμούς, οι οποίοι έχουν την ίδια διασπορά. Έστω $\bar{x}_1$, $\bar{x}_2$ και σ_1, σ_2 οι μέσες τιμές και οι τυπικές αποκλίσεις των δειγμάτων, αντίστοιχα. Για να ελέγξουμε τη μηδενική υπόθεση H_0 ότι τα δείγματα προέρχονται από τον ίδιο πληθυσμό, χρησιμοποιούμε τη μεταβλητή :

$$T = (\bar{x}_1 - \bar{x}_2) / (s\sqrt{(1/v_1 + 1/v_2)})$$

όπου

$$s = [(v_1 s_1^2 + v_2 s_2^2) / (v_1 + v_2 - 2)]^{½}$$

Η T ακολουθεί την κατανομή t του Student με v_1+v_2-2 βαθμούς ελευθερίας.

Διασπορές : Για να ελέγξουμε τη μηδενική υπόθεση H_0 ότι ένας πληθυσμός που ακολουθεί κανονική κατανομή έχει διασπορά σ_0^2, θεωρούμε την ποσότητα x^2, όπου :

$$x^2 = vs^2 / \sigma_0^2$$

Η ποσότητα αυτή ακολουθεί κατανομή χ^2 με v-1 βαθμούς ελευθερίας. s^2 είναι η διασπορά του δείγματος μεγέθους v.

Για να δεχτούμε την H_0, σε επίπεδο σημαντικότητας 0.05, θα εξετάσουμε αν ισχύει η ανισότητα :

$$\chi^2_{.025} \leq vs^2 / \sigma_0^2 \leq \chi^2_{.975}$$

Αν ισχύει, τότε δεχόμαστε ότι ο πληθυσμός έχει διασπορά σ_0. Αν όχι, τότε απορρίπτουμε την H_0.

Για να ελέγξουμε την υπόθεση ότι η διασπορά είναι μεγαλύτερη του σ_0, με επίπεδο σημαντικότητας 0.05, πραγματοποιούμε μονόπλευρο έλεγχο με βάση την ανισότητα :

$$vs^2 / \sigma_0^2 > \chi^2_{.95}$$

Ο μονόπλευρος έλεγχος της υπόθεσης ότι η διασπορά σ είναι μικρότερη του σ_0, πραγματοποιείται με βάση την ανισότητα :

$$vs^2 / \sigma_0^2 < \chi^2_{.95}$$

Λόγοι διασπορών: Έστω δύο δείγματα μεγέθους v_1 και v_2 με τυπικές αποκλίσεις s_1, s_2, αντίστοιχα. Αν θέλουμε να εξετάσουμε τη μηδενική υπόθεση ότι δεν υπάρχει διαφορά μεταξύ των δύο διασπορών, δηλαδή ότι τα δείγματα προέρχονται από τον ίδιο πληθυσμό, χρησιμοποιούμε την ποσότητα :

$$f = (s_1^2(v_2 - 1)v_1) / (s_2^2(v_1 - 1)v_2)$$

που ακολουθεί κατανομή F με βαθμούς ελευθερίας v_1-1, v_2-1.

Για να ελέγξουμε τη μηδενική υπόθεση σε επίπεδο σημαντικότητας 0.1, εξετάζουμε κατά πόσον ισχύει η ανίσωση :

$$F_{.05} \leq (s_1^2(v_2 - 1)v_1) / (s_2^2(v_1 - 1)v_2) \leq F_{.95}$$

Αν όντως ισχύει, τότε αποδεχόμαστε τη μηδενική υπόθεση. Αν όχι, την απορρίπτουμε.

1.13.4. Παραδείγματα

Μπορούμε να σχηματίσουμε μια πιο σαφή εικόνα για το πώς μπορούμε να χρησιμοποιήσουμε τα κριτήρια ελέγχου υποθέσεων που διατυπώθηκαν στις παραγράφους 23.13.2 και 23.13.3, με τα παραδείγματα που ακολουθούν.

Παράδειγμα 1 *(Spiegel, 1977) :* Η *μέση ζωή 100 τυχαίων λαμπτήρων φθορισμού βρέθηκε πως είναι 1570 ώρες με τυπική απόκλιση 120 ώρες. Αν Μ είναι η μέση ζωή των λαμπτήρων φθορισμού*

της εταιρείας, να ελέγξετε την υπόθεση M=1600, με εναλλακτική υπόθεση την M≠1600, σε επίπεδο σημαντικότητας 0.05.

Θα εφαρμόσουμε δίπλευρο έλεγχο, καθώς M≠1600 σημαίνει ότι είναι δεκτές τιμές μέσης ζωής μικρότερες και μεγαλύτερες των 1600 ωρών.

Για το σκοπό αυτόν εξετάζουμε αν ισχύει η ανισότητα :

$$\mathbf{-1.96 \leq z \leq 1.96}$$

Οι τιμές $\pm$1.96 είναι οι κρίσιμες τιμές της μεταβλητής z, που ακολουθεί την τυποποιημένη κατανομή, για επίπεδο σημαντικότητας 0.05. Στη συγκεκριμένη περίπτωση :

$$\mathbf{z = (1570 - 1600) / (120 / \sqrt{100}) = -2.50}$$

η τιμή αυτή βρίσκεται σαφώς έξω από το διάστημα [-1.96, +1.96], κατά συνέπεια απορρίπτουμε την υπόθεση M=1600.

Παράδειγμα 2 *(Spiegel, 1977) : Μια μηχανή γεμίζει σακουλάκια καφέ. Στην κανονική της λειτουργία βάζει σε κάθε σακουλάκι 40 gr καφέ με τυπική απόκλιση 0.25 gr. Σε τυχαίο δείγμα με 20 σακουλάκια μετρήθηκε τυπική απόκλιση 0.32 gr. Είναι σημαντική αυτή η διαφορά στην τυπική απόκλιση, σε επίπεδο σημαντικότητας 0.05 ;*

Χρησιμοποιώντας τη μεταβλητή $x^2 = n\sigma^2/\sigma_0^2$, που ακολουθεί κατανομή χ^2 με n-1 βαθμούς ελευθερίας (στη στατιστική, βαθμοί ελευθερίας *f* είναι μια παράμετρος που συνδέεται με το μέγεθος του δείγματος), και αντικαθιστώντας, έχουμε :

$$\mathbf{x^2 = 20(0.32)^2 / (0.25)^2 = 32.8}$$

χρησιμοποιώντας μονόπλευρο έλεγχο θα απορρίψουμε την H_0 σε επίπεδο σημαντικότητας 0.05, αν

$$x^2 > \chi^2_{.95}$$

Με τη βοήθεια πινάκων βρίσκουμε ότι :

$\chi^2_{.95} = 30.1$ όταν **v = 20 - 1 = 19** βαθμοί ελευθερίας

Και επειδή $x^2 = 32.8$, η ανισότητα ισχύει, οπότε δεχόμαστε ότι η τυπική απόκλιση του δείγματος είναι μεγαλύτερη από την κανονική, σε επίπεδο σημαντικότητας 0.05.

1.14. Ανάλυση διασποράς

Στην προηγούμενη ενότητα 26.13. παρουσιάσαμε κριτήρια και μεθόδους σύγκρισης μεταξύ δύο δειγματικών μέσων τιμών, που μας επιτρέπουν να αποφανθούμε, σε κάποιο επίπεδο σημαντικότητας, για το αν οι δύο δειγματικές τιμές αντιστοιχούν στην ίδια μέση τιμή πληθυσμού ή διαφέρουν μεταξύ τους. Είναι επίσης δυνατό να ελεγχθεί το κατά πόσον τρείς ή περισσότερες δειγματικές μέσες τιμές είναι ίσες ή διαφέρουν σημαντικά, με τη βοήθεια της **ανάλυσης διασποράς**, που εκτίθεται παρακάτω. Τόσο στην περίπτωση των δύο, όσο και των πολλών δειγματικών τιμών δεχόμαστε ότι οι πληθυσμοί από όπου προέρχονται τα δείγματα έχουν την ίδια διασπορά.

Έστω λοιπόν ότι έχουμε a ανεξάρτητα δείγματα με b τιμές (μετρήσεις) το καθένα. Τα αποτελέσματα των μετρήσεων παρουσιάζονται στον παρακάτω πίνακα.

Α/α δείγματος	τιμές δείγματος	μέση τιμή δείγματος
1	x_{11} x_{12} ... x_{1b}	$\bar{x}_1$
2	x_{21} x_{22}... x_{2b}	$\bar{x}_2$
...	...	...
a	x_{a1} x_{a2} ... x_{ab}	$\bar{x}_a$

Το x_{ij} είναι το στοιχείο της i γραμμής και j στήλης. Τα $\bar{x}_1$, $\bar{x}_2$, ..., $\bar{x}_a$ είναι οι δειγματικές μέσες τιμές για τα δείγματα 1,2,...,a αντίστοιχα. Το σύνολο των τιμών όλων των δειγμάτων έχει μια γενική μέση τιμή $\bar{x}$.

Επομένως, ισχύει ότι :

$$\overline{x}_i = \frac{1}{b}\sum_{j=1}^{b}[x_{ij}]$$

$$\overline{x} = \frac{1}{ab}\sum_{i,j}[x_{ij}]$$

όπου $i = 1,2,...,a$
$j = 1,2,...,b$

Το ερώτημα που τίθεται είναι : πώς μπορεί να ελεγχθεί η μηδενική υπόθεση H_0, ότι οι δειγματικές μέσες τιμές δε διαφέρουν σημαντικά μεταξύ τους, δηλαδή αντιστοιχούν στην ίδια μέση τιμή πληθυσμού. Για το σκοπό αυτόν υιοθετείται το **γραμμικό μοντέλο για την ανάλυση διασποράς,** σύμφωνα με το οποίο κάθε δειγματική τιμή x_{ij} μπορεί να γραφεί ως :

$$\mathbf{x_{ij} = M + \alpha_i + \Delta_j}$$

όπου M η μέση τιμή του (γενικού) πληθυσμού, α_i μια ποσότητα που εκφράζει τη διαφοροποίηση του κάθε δείγματος από τα υπόλοιπα, Δ_j μια ποσότητα που εκφράζει τη διαφοροποίηση της κάθε τιμής από τις άλλες, στο ίδιο δείγμα.

Αποδεικνύεται ότι η α_i εκπληρώνει τη σχέση :

$$\sum_i \alpha_i = 0 \ , \quad i = 1,2, \ldots ,a$$

Με βάση το γραμμικό μοντέλο ανάλυσης διασποράς, και με μια σειρά θεωρημάτων της στατιστικής, αποδεικνύεται ότι ο έλεγχος της στατιστικής υπόθεσης H_0 μπορεί να γίνει με τη στατιστική συνάρτηση :

$$f = S_b^2 / S_{ab}^2$$

που έχει κατανομή F με a-1 και a(b-1) βαθμούς ελευθερίας. Οι ποσότητες s_b^2 και s_{ab}^2 δίνονται από τις σχέσεις :

$$s_b^2 = \frac{b\sum_i [\bar{x}_i - \bar{x}]^2}{a - 1}$$

$$s_{ab}^2 = \frac{\sum_{i,j} [x_{ij} - \bar{x}_i]^2}{a(b - 1)}$$

Επομένως, για να ελεγχθεί, για κάποιο επίπεδο σημαντικότητας, η μηδενική υπόθεση H_0 για τις μέσες τιμές $\bar{x}_1$, $\bar{x}_2$, ..., $\bar{x}_a$, που προκύπτουν από τα ανεξάρτητα δείγματα 1, 2, ..., a αντίστοιχα, εφαρμόζεται η παρακάτω μεθοδολογία :

1. Υπολογισμός των ποσοτήτων $\bar{x}_i$ (i=1,2,...,v), καθώς και του $\bar{x}$.
2. Υπολογισμός των s_b, s_{ab} και, μέσω αυτών, της f.
3. Καθορισμός του επιπέδου εμπιστοσύνης και εύρεση του αντίστοιχου εκατοστιαίου σημείου της κατανομής F, για βαθμούς ελευθερίας $v_1 = a-1$ και $v_2 = a(b-1)$.
4. Σύγκριση του εκατοστιαίου σημείου της κατανομής F με την ποσότητα f. Αν η f είναι μεγαλύτερη του εκατοστιαίου σημείου της F, τότε η υπόθεση δεν ισχύει.

Η ανάλυση διασποράς μπορεί να επεκταθεί, με κατάλληλες τροποποιήσεις, και σε περιπτώσεις όπου τα μεγέθη των δειγμάτων διαφέρουν μεταξύ τους, καθώς και σε περιπτώσεις όπου μελετούμε δείγματα πληθυσμών με δύο ή περισσότερες παραμέτρους (ανάλυση διασποράς για πολλούς παράγοντες).

Παράδειγμα *(Spiegel, 1977) : Στον παρακάτω πίνακα δίνεται η απόδοση σε σιτάρι (μπούσελ ανά στρέμμα) δώδεκα ομοίων κομματιών γης, όπου βάλαμε τριών ειδών λιπάσματα Α, Β, Γ. α) Να υπολογίσετε τη μέση απόδοση για κάθε λίπασμα, καθώς και τη γενική μέση απόδοση. β) Να ελέγξετε τη μηδενική υπόθεση H_0 ότι οι μέσες αποδόσεις για τα αντίστοιχα λιπάσματα είναι ίδιες.*

Α	*48 49 50 49*
Β	*47 49 48 48*
Γ	*49 51 50 50*

α) Για το δείγμα Α: $\bar{x}_1 = (48+49+50+49)/4 = 49$

Για το δείγμα Β: $\bar{x}_2 = 48$

Για το δείγμα Γ: $\bar{x}_3 = 50$

Η γενική μέση τιμή είναι : $\bar{x} = (\bar{x}_1 + \bar{x}_2 + \bar{x}_3)/3 = 49$

β) Στο συγκεκριμένο πρόβλημα έχουμε a = 3 και b = 4

Αντικαθιστώντας στις εκφράσεις για τα s_b, s_{ab} βρίσκουμε ότι :

$$s_b^2 = 4[(49-49)^2+(48-49)^2+(50-49)^2]/(3-1) = 4$$

$$s_{ab}^2 = [(48-49)^2+(49-49)^2+(50-49)^2+(49-49)^2+(47-48)^2+(49-48)^2+$$
$$+(48-48)^2+(48-48)^2+(49-50)^2+(51-50)^2+(50-50)^2+(50-50)^2]/[3(4-1)]$$

Τελικά $s_{ab}^2 = 2/3$

Οπότε $f = s_b^2/s_{ab}^2 = 4/(2/3) = 6$

Η f ακολουθεί μια κατανομή F με βαθμούς ελευθερίας v_1=a-1=2 και $v_2 = 3(4-1) = 9$. Η αντίστοιχη εκατοστιαία τιμή $F_{.95}$, που αντιστοιχεί σε επίπεδο σημαντικότητας 0.05, για βαθμούς ελευθερίας 2, 9, είναι :

$F_{.95} = 4.26$

Έχουμε λοιπόν ότι $f>F_{.95}$ και κατά συνέπεια η μηδενική υπόθεση απορρίπτεται σε επίπεδο σημαντικότητας 0.05. Αυτό σημαίνει ότι οι μέσες τιμές των αποδόσεων, για κάθε λίπασμα, διαφέρουν σημαντικά μεταξύ τους.

1.15. Το κριτήριο χ^2 για καλή προσαρμογή

Στη στατιστική ανάλυση δεδομένων, τίθεται συχνά το ερώτημα κατά πόσον οι παρατηρούμενες συχνότητες εμφάνισης γεγονότων συμφωνούν με τις προβλέψεις της θεωρίας ή με τις εκτιμήσεις του ερευνητή. Για παράδειγμα, αν οι παρατηρούμενες φαινοτυπικές αναλογίες σε ένα δείγμα φυτικού οργανισμού ακολουθούν τις αναλογίες που προβλέπουν οι νόμοι του Μέντελ, ή αν το ιστόγραμμα τιμών περιεκτικότητας σε οργανικό υλικό που μετρήθηκαν σε τομές ενός πετρώματος ακολουθούν μια κανονική κατανομή.

Τέτοιου είδους ερωτήματα μπορούν να απαντηθούν αν ληφθεί υπόψη ότι η στατιστική συνάρτηση X^2, οι τιμές της οποίας ορίζονται ως :

$$x^2 = \sum_i \frac{(x_i - a_i)^2}{a_i}$$

ακολουθεί προσεγγιστικά τη γνωστή κατανομή χ^2 με βαθμούς ελευθερίας k = n-1-m, x_i και a_i είναι η παρατηρούμενη και η προβλεπόμενη συχνότητα εμφάνισης του γεγονότος i, αντίστοιχα, n είναι ο αριθμός των γεγονότων και m ο αριθμός των στατιστικών παραμέτρων που υπολογίζονται κατά τον προσδιορισμό των συχνοτήτων εμφάνισης x_i.

Αν ληφθεί υπόψη η παραπάνω παρατήρηση, μπορεί να γίνει ένας έλεγχος για το κατά πόσον προσαρμόζονται τα παρατηρησιακά δεδομένα με τη θεωρία ή τις εκτιμήσεις του ερευνητή, με βάση το **κριτήριο χ^2**.

Το κριτήριο χ^2 μπορεί να εφαρμοστεί είτε για δεδομένα που εκφράζουν διακριτά γεγονότα (όπως οι συχνότητες εμφάνισης φαινοτύπων) είτε για δεδομένα που εκφράζουν φυσικά μεγέθη με

συνεχές πεδίο τιμών (όπως οι συχνότητες εμφάνισης διαστημάτων τιμών σε ιστόγραμμα).

Παρακάτω παρουσιάζονται τα βήματα με τα οποία υλοποιείται ο έλεγχος με κριτήριο χ^2, για τις δυο περιπτώσεις.

1.15.1. Εφαρμογή του κριτηρίου χ^2 για διακριτά γεγονότα

Στην περίπτωση αυτή ελέγχεται η μηδενική υπόθεση H_0 (οι παρατηρούμενες συχνότητες εμφάνισης ακολουθούν τις προβλεπόμενες αναλογίες), με εναλλακτική υπόθεση H_1 (οι παρατηρούμενες συχνότητες εμφάνισης δεν ακολουθούν τις προβλεπόμενες αναλογίες). Για το σκοπό αυτό υπολογίζεται η ποσότητα x^2, με βάση τη σχέση :

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

και συγκρίνεται με την τιμή x^2_c της κατανομής χ^2 για δεδομένους βαθμούς ελευθερίας (που καθορίζονται από τον αριθμό των γεγονότων και από τον αριθμό των υπολογιζόμενων στατιστικών παραμέτρων) και δεδομένο επίπεδο σημαντικότητας.

Αν

$$x^2 > x^2_c$$

τότε απορρίπτεται η H_0 στο συγκεκριμένο επίπεδο σημαντικότητας. Αν δεν ισχύει η παραπάνω ανισότητα (δηλαδή αν $x^2 \leq x^2_c$), **τότε δεν απορρίπτεται η H_0**, στο ίδιο επίπεδο σημαντικότητας.
Τα παραπάνω αποσαφηνίζονται περισσότερο με το παράδειγμα που ακολουθεί.

Παράδειγμα (Spiegel 1977): Σε πείραμα γενετικής (διασταυρώσεις μπιζελιών) μετρήθηκαν 315 μπιζέλια στρογγυλά και κίτρινα (x_1), 108 στρογγυλά και πράσινα (x_2), 101 ζαρωμένα και κίτρινα (x_3) και 32 ζαρωμένα και πράσινα (x_4). Σύμφωνα με τους νόμους του Mendel η προβλεπόμενη αναλογία είναι 9:3:3:1. Να ελεγχθεί κατά πόσον συμφωνούν τα πειραματικά δεδομένα με την προβλεπόμενη φαινοτυπική αναλογία, σε επίπεδο σημαντικότητας 0.05.

Λύση: Ο συνολικός αριθμός των μπιζελιών είναι

$I = x_1 + x_2 + x_3 + x_4 = 556$

Το άθροισμα των όρων της αναμενόμενης αναλογίας είναι

$I' = 9 + 3 + 3 + 1 = 16$

Με βάση τους νόμους του Mendel, ο αναμενόμενος αριθμός στογγυλών και κίτρινων μπιζελιών a_1 είναι:

$a_1 = 556 \times (9/16) = 312.75$

Ο αναμενόμενος αριθμός στρογγυλών και πράσινων μπιζελιών a_2 είναι:

$a_2 = 556 \times (3/16) = 104.25$

Ο αναμενόμενος αριθμός ζαρωμένων και κίτρινων μπιζελιών a_3 είναι:

$a_3 = 556 \times (3/16) = 104.25$

Ο αναμενόμενος αριθμός ζαρωμένων και πράσινων μπιζελιών a_4 είναι:

$a_4 = 556 \times (1/16) = 34.75$

Η ποσότητα x^2 είναι:

$$x^2 = \Sigma_{i=1}^{4} [(x_i - a_i)^2/a_i] = [(315-312.75)^2/312.75] + [(108-104.25)^2/104.25] + [(101-104.25)^2/104.25] + [(32-34.75)^2/34.75]$$

Δηλαδή

$x^2 = 0.47$

Για 0.05 επίπεδο σημαντικότητας και k=n-1=4-1=3 βαθμούς ελευθερίας, η ποσότητα x_c^2 είναι:

$x_c^2 = 7.85$

Οπότε

$x^2 < x_c^2$

Αυτό σημαίνει ότι δεν απορρίπτεται η μηδενική υπόθεση H_o (τα πειραματικά δεδομένα συμφωνούν με την αναμενόμενη φαινοτυπική αναλογία) σε επίπεδο σημαντικότητας 0.05.

1.15.2. Εφαρμογή του κριτηρίου χ^2 για μετρήσεις μεγεθών με συνεχές πεδίο τιμών

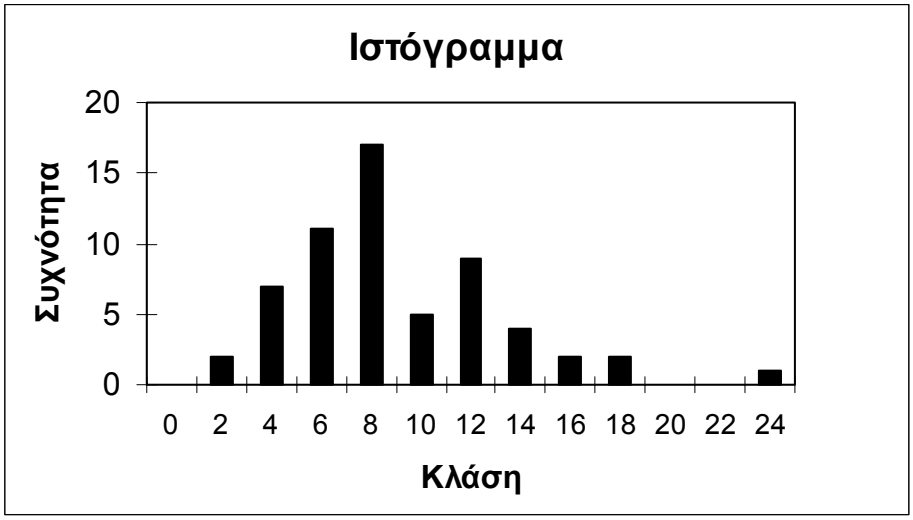
Οι διάμετροι κόκκων εδάφους, η περιεκτικότητα ιζημάτων σε κάποιο χημικό στοιχείο, καθώς και άλλες ποσότητες που μετρούνται στο ύπαιθρο ή στο εργαστήριο, είναι συνεχή μεγέθη, οι συχνότητες εμφάνισης των οποίων αναπαριστώνται σε ιστογράμματα τιμών. Εδώ τίθεται το ερώτημα αν το ιστόγραμμα τιμών του δείγματος ανήκει σε πληθυσμό που ακολουθεί μια συγκεκριμένη κατανομή (κανονική για παράδειγμα). Σε μια τέτοια περίπτωση μπορεί να εφαρμοστεί το κριτήριο χ^2, υπάρχει όμως το πρόβλημα ότι η υπολογιζόμενη από τα πειραματικά δεδομένα ποσότητα x^2 μπορεί να διαφέρει, ανάλογα με τα διαστήματα τιμών με τα οποία κατασκευάζεται το ιστόγραμμα. Για το λόγο αυτό, συχνά κρίνεται σκόπιμο να χωριστούν οι μετρήσεις σε διαστήματα τιμών τέτοια ώστε οι αντίστοιχες συχνότητες εμφάνισης είναι περίπου ίσες. Αυτά τα διαστήματα τιμών προσδιορίζονται με βάση την κατανομή που υποτίθεται ότι ακολουθεί ο πληθυσμός. Αυτές οι μετρούμενες συχνότητες συγκρίνονται με τις συχνότητες που προβλέπονται θεωρητικά, αν ο πληθυσμός ακολουθεί όντως τη συγκεκριμένη κατανομή (οι θεωρητικά προβλεπόμενες συχνότητες είναι μεταξύ τους ίσες). Η σύγκριση πραγματοποιείται με το κριτήριο χ^2 και με την ίδια κατά βάση μεθοδολογία με αυτήν που εφαρμόζεται για τα διακριτά γεγονότα.

Τα παραπάνω μπορούν να αποσαφηνιστούν μελετώντας το παράδειγμα που ακολουθεί.

Παράδειγμα (Swan & Sandilands 1995): Τα παρακάτω δεδομένα είναι τιμές περιεκτικότητας 60 δειγμάτων πετρώματος σε οργανική ύλη :

11.8, 6.5, 7.7, 22.8, 5.9, 5.4, 7.0, 7.8, 6.3, 3.2, 11.0, 7.2, 5.3, 3.5, 8.6, 2.3, 7.9, 2.0, 5.5, 4.9, 17.1, 16.2, 7.1, 11.5, 5.9, 15.2, 11.7, 7.6, 10.7, 10.1, 2.3, 13.9, 10.2, 7.7, 13.4, 8.0, 11.6, 5.9, 5.4, 7.6, 7.1, 11.7, 5.1, 12.6, 3.0, 14.1, 1.6, 9.3, 8.3, 12.2, 9.6, 2.4, 4.9, 8.3, 8.0, 3.6, 7.1, 6.5, 7.4, 7.4

Το ιστόγραμμα τιμών εμφανίζεται στο παρακάτω σχήμα.



Προέρχονται αυτά τα δεδομένα από κανονική κατανομή;

Λύση. Τα διαστήματα τιμών στο παραπάνω ιστόγραμμα ορίστηκαν αυθαίρετα. Θα μπορούσε να εφαρμοστεί το κριτήριο χ^2 με βάση το ιστόγραμμα αυτό, όμως η τιμή της x^2 θα μπορούσε να είναι διαφορετική για άλλο διάστημα τιμών, οπότε η εφαρμογή του ελέγχου χ^2 χάνει σε αξιοπιστία. Γι'αυτό το λόγο είναι προτιμότερο να χωριστούν οι τιμές των μετρήσεων σε 10 άνισα διαστήματα, το καθένα από τα οποία να αντιστοιχεί στο 10% του συνολικού εμβαδού του χωρίου που ορίζεται από την καμπύλη της τυποποιημένης κανονικής κατανομής.

Από σχετικούς πίνακες, ή με τη βοήθεια του λογισμικού Excel, βρίσκεται ότι τα δέκα αυτά διαστήματα ορίζονται από τους αριθμούς:

-1.2816, -0.8416, -0.5284, -0.2533, 0, 0.2533, 0.5284, 0.8416, 1.2816

Οι 9 αυτές τιμές ορίζουν 10 διαστήματα, συμπεριλαμβανομένων των δυο διαστημάτων που περιέχουν τιμές μικρότερες του –1.2816 και μεγαλύτερες του +1.2816.

Οι 9 παραπάνω αριθμοί ορίζουν τα 10 διαστήματα ίσου εμβαδού της τυποποιημένης κανονικής κατανομής. Τα αντίστοιχα διαστήματα των περίπου ίσων συχνοτήτων εμφάνισης μετρουμένων τιμών περιεκτικότητας οργανικής ύλης προσδιορίζονται με βάση τη σχέση:

Όριο διαστήματος τιμών δείγματος = (Όριο τυποποιημένης κανονικής κατανομής)·s + $\bar{x}$

Από τα δεδομένα των μετρήσεων βρίσκουμε :

$\bar{x} = 8.16$

$s = 4.13$

Οπότε τα αντίστοιχα όρια διαστημάτων τιμών δείγματος είναι :

2.87, 4.68, 5.98, 7.11, 8.16, 9.21, 10.34, 11.64, 13.45

Τα όρια αυτά ορίζουν 10 διαστήματα τιμών, για κάθε ένα από τα οποία η αναμενόμενη συχνότητα εμφάνισης είναι ίση με 60/10 = 6.

Οι πραγματικές αντίστοιχες συχνότητες εμφάνισης για τα 10 διαστήματα τιμών είναι:

5, 4, 11, 7, 10, 3, 4, 4, 6, 6

Επομένως η τιμή x^2 είναι :

$$x^2 = (5-6)^2/6 + (4-6)^2/6 + (11-6)^2/6 + (7-6)^2/6 + (10-6)^2/6 + (3-6)^2/6 + (4-6)^2/6 + (4-6)^2/6 + (6-6)^2/6 + (6-6)^2/6$$

Οπότε :

$x^2 = 10.67$

Για τον υπολογισμό του x_c^2 θα λάβουμε επίπεδο σημαντικότητας 5% και αριθμό βαθμών ελευθερίας ίσο με τον αριθμό των διαστημάτων τιμών μείον 1 μείον αριθμό υπολογισθεισών στατιστικών παραμέτρων.

Στη συγκεκριμένη εφαρμογή προσδιορίστηκαν 10 διαστήματα τιμών και υπολογίστηκαν 2 στατιστικές παράμετροι (μέση τιμή και τυπική απόκλιση δείγματος), οπότε:

Αριθμός βαθμών ελευθερίας = 10 – 1 – 2 = 7

Με τη βοήθεια πινάκων ή ενός προγράμματος λογισμικού, βρίσκεται ότι για 5% επίπεδο σημαντικότητας και 7 βαθμούς ελευθερίας η τιμή x_c^2 είναι:

$$x_c^2 = 14{,}07$$

οπότε

$$x^2 < x_c^2$$

και επομένως δεν απορρίπτεται η μηδενική υπόθεση (οι τιμές του δείγματος προέρχονται από πληθυσμό κανονικής κατανομής) για επίπεδο σημαντικότητας 5%.

Με το Excel δεν υπάρχει η δυνατότητα αυτόματης εκτέλεσης του ελέγχου με το κριτήριο χ^2. Αν όμως ο ερευνητής γνωρίζει τα βήματα της όλης εργασίας, οι σχετικοί υπολογισμοί μπορούν να γίνουν σύντομα και χωρίς κόπο. Σε άλλα, πιο εξειδικευμένα στατιστικά προγράμματα, ο έλεγχος με κριτήριο χ^2 μπορεί να πραγματοποιηθεί εντελώς αυτοματοποιημένα.

2. ΜΕΘΟΔΟΣ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ - ΘΕΩΡΙΑ ΣΦΑΛΜΑΤΩΝ

Μέχρι τώρα εξετάσαμε πώς συμπεριφέρονται οι τιμές μιας μεταβλητής ως προς μια ιδιότητά τους. Είναι όμως χρήσιμο να ιδούμε πώς σχετίζονται μεταξύ τους οι τιμές δυο μεταβλητών και στη συνέχεια να διατυπώσουμε με μαθηματικό τρόπο τη μεταξύ τους εξάρτηση.

Επίσης, στο κεφάλαιο αυτό θα αναφερθούμε σε γενικές γραμμές και στη Θεωρία Σφαλμάτων.

Η Θεωρία Σφαλμάτων ασχολείται με την ακρίβεια των μετρήσεων και των αποτελεσμάτων των υπολογισμών, ως επί το πλείστον, κατά τη μέτρηση μεγεθών, την εκτέλεση πειραμάτων και τη διεξαγωγή στατιστικών μελετών.

2.1. Ευθεία Ελαχίστων Τετραγώνων

Έστω, ότι έχουμε **v** αριθμό ζευγών δεδομένων, που τα μέλη τους είναι αντίστοιχα οι τιμές δύο μεταβλητών. Αυτό που θέλουμε να ελέγξουμε είναι, εάν η εξάρτηση των δυο μεγεθών ή μεταβλητών μεταξύ τους υπακούει σε κάποιο νόμο μεταβολής, (γραμμικής φύσης, δηλαδή βασισμένο σε εξίσωση 1ου βαθμού ή μη γραμμικής). Θεωρούμε, συνήθως, ότι η μια μεταβλητή (X) είναι ανεξάρτητη και η άλλη (Y) είναι εξαρτημένη από αυτήν.

Ο λόγος για τον οποίο η μέθοδος ονομάζεται "μέθοδος ελαχίστων τετραγώνων" είναι ο εξής : Έστω, ότι προβάλλουμε τα σημεία που ορίζονται από τα ζεύγη (x,y) σε ένα δισδιάστατο διάγραμμα. Σε

αυτό το διάγραμμα προσπαθούμε να χαράξουμε μια ευθεία (ή καμπύλη ανώτερου βαθμού, αν θέλουμε) η οποία να περνά από όσο το δυνατόν περισσότερα σημεία. Το πόσο καλά το επιτυγχάνουμε μπορούμε να το υπολογίσουμε από τις διαφορές των πραγματικών τιμών y από αυτές που αντιστοιχούν (y') στα ίδια x αλλά πάνω στην ευθεία (ή καμπύλη). Καλύτερα προσαρμοσμένη ευθεία (ή καμπύλη) ορίζεται αυτή για την οποία το άθροισμα των τετραγώνων των διαφορών y-y' για κάθε σημείο γίνεται το ελάχιστο δυνατό (δηλαδή, η ευθεία ή καμπύλη πλησιάζει ή εφάπτεται σε όσο το δυνατόν περισσότερα σημεία και άρα οι διαφορές y-y' είναι πολύ μικρές). Αποδεικνύεται, ότι αυτή η ευθεία, αν, βέβαια, μιλάμε για ευθεία και όχι για καμπύλη, είναι της μορφής:

$$y=a+bx \text{ (όπου a και b συντελεστές)}$$

Η παρα πάνω σχέση με άθροιση για v ζεύγη δίνει:

$$\sum_{i=1}^{v}[y] = va+b\sum_{i=1}^{v}[x]$$

ενώ, με πολλαπλασιασμό κάθε μέλους της επί x και στη συνέχεια άθροιση, μας δίνει:

$$\sum_{I=1}^{v}[xy] = a\sum_{i=1}^{v}[x]+b\sum_{i=1}^{v}[x^2]$$

Από τις δυο παρα πάνω σχέσεις, λύνοντας ως προς a και b (για συντομία, η απόδειξη παραλείπεται), έχουμε:

$$a = \frac{\sum_{i=1}^{v}[y]\sum_{i=1}^{v}[x^2] - \sum_{i=1}^{v}[x]\sum_{i=1}^{v}[xy]}{v\sum_{i=1}^{v}[x^2] - (\sum_{i=1}^{v}[x])^2}$$

$$b = \frac{v\sum_{i=1}^{v}[xy] - \sum_{i=1}^{v}[x]\sum_{i=1}^{v}[y]}{v\sum_{i=1}^{v}[x^2] - (\sum_{i=1}^{v}[x])^2}$$

Οπότε, έχουμε βρεί τους συντελεστές a (σημείο τομής της ευθείας με τον y άξονα) και b (κλίση της ευθείας) και, άρα, έχουμε καθορίσει και την εξίσωση, η οποία φαίνεται, ότι μας δίνει την εξάρτηση της μιας μεταβλητής από την άλλη.

Η παρα πάνω μέθοδος μπορεί να εφαρμοστεί και σε πολλές άλλες περιπτώσεις, όπου δεν φαίνεται να έχουμε γραμμική συσχέτιση, αλλά κάποιου άλλου είδους συσχέτιση, την οποία πολλές φορές μπορούμε να μετατρέψουμε σε αντίστοιχη μορφή. Για παράδειγμα, μια συσχέτιση της ανεξάρτητης μεταβλητής (συνόλου στοιχείων) x με την εξαρτημένη y της εκθετικής μορφής:

$$y = ax^b$$

μετασχηματίζεται με λογαρίθμηση σε:

$$\log(y) = \log(a)+b \log(x) \quad [= \text{γραμμική μορφή }]$$

που, βέβαια, είναι τώρα της μορφής:

$$Y = a'+b' X$$

και απο εδώ συνεχίζουμε με παρόμοιο τρόπο χρησιμοποιώντας σαν ανεξάρτητη μεταβλητή X τα log(x) και σαν εξαρτημένη Y τα log(y).

Υπολογίζουμε έτσι τους συντελεστές a' και b' και από αυτούς τους αρχικούς με:

$a' = \log(a) \Rightarrow a = 10^{a'}$ (αν πρόκειται για δεκαδικό λογάριθμο)

$b' = b$

Παρατήρηση : Αν θέλουμε να δώσουμε "βάρος" σε ένα σημείο, δηλαδή, να αναγκάσουμε την ευθεία να περνά πολύ κοντά από αυτό, επαναλαμβάνουμε τόσες φορές τις συντεταγμένες του, όσες φορές περισσότερο "βάρος" πρέπει να έχει το συγκεκριμένο σημείο από τα υπόλοιπα. Δηλαδή, είναι σαν να προσθέτουμε περισσότερα σημεία που προβάλλονται όμως στην ίδια φυσική θέση, και το ίδιο μπορεί να γίνει για όσα σημεία θέλουμε. Ο αριθμός των ζευγών **v**, βέβαια, αλλάζει αντίστοιχα.

Με τον παρα κάτω τύπο μπορούμε να βρούμε το συντελεστή συσχέτισης **r** μεταξύ των δυο συνόλων στοιχείων x και y, που μας δείχνει πόσο καλά συσχετίζονται μεταξύ τους και κατά πόσο η ευθεία των ελαχίστων τετραγώνων προσεγγίζει την κατανομή των σημείων στο δισδιάστατο διάγραμμα. Μόνο από εδώ μπορούμε τελικά να ιδούμε αν οι μεταβλητές συσχετίζονται γραμμικά ή όχι, καθώς σε κάθε περίπτωση είναι δυνατόν να κατασκευάσουμε ευθεία ελαχίστων τετραγώνων (που όμως δεν θα έχει καμία πραγματική αξία ή ερμηνεία, αν όλα τα σημεία κατανέμονται άτακτα πολύ μακριά από αυτή).

$$r = \frac{v\sum_{i=1}^{v}[xy] - \sum_{i=1}^{v}[x]\sum_{i=1}^{v}[y]}{\sqrt{v\sum_{i=1}^{v}[x^2] - (\sum_{i=1}^{v}[x])^2} \cdot \sqrt{v\sum_{i=1}^{v}[y^2] - (\sum_{i=1}^{v}[y])^2}}$$

Τιμές κοντά στο 0 σημαίνουν ελάχιστη γραμμική συσχέτιση (προσοχή: υπάρχει και μη γραμμική συσχέτιση που μπορεί να είναι μεγάλη), ενώ τιμές κοντα στο 1 και -1 σημαίνουν μεγάλη γραμμική θετική και αρνητική συσχέτιση, αντίστοιχα (δηλαδή, ταυτόχρονη αύξηση των

δύο μεταβλητών ή αύξηση της μιας και μείωση της άλλης, αντίστοιχα). Παρατηρούμε ότι σε Χ-Υ διάγραμμα έχουμε τοποθέτηση των σημείων κοντά σε ευθεία (την ευθεία που προσεγγίζει η μέθοδος των ελαχίστων τετραγώνων) με θετική ή αρνητική κλίση, αν έχουμε μεγάλη θετική ή αρνητική γραμμική συσχέτιση αντίστοιχα, ενώ όσο μικραίνει η συσχέτιση έχουμε μια γενική, πιο άτακτη εξάπλωση των σημείων στο χώρο, που δεν πλησιάζει πια σε ευθεία.

2.1.1. Παραβολή Ελαχίστων Τετραγώνων

Αν επιχειρήσουμε να προσεγγίσουμε ένα σύνολο ζευγών τιμών (x,y) με μια παραβολή της μορφής :

$$y = a+bx+cx^2$$

οι εξισώσεις για τον προσδιορισμό των συντελεστών της παραβολής, αποδεικνύεται πως είναι οι :

$$\Sigma[y] = va+b\Sigma[x]+c\Sigma[x^2]$$
$$\Sigma[xy] = a\Sigma[x]+b\Sigma[x^2]+c\Sigma[x^3]$$
$$\Sigma[x^2y] = a\Sigma[x^2]+b\Sigma[x^3]+c\Sigma[x^4]$$

Οι αθροίσεις των ποσοτήτων στις αγκύλες γίνονται για i=1,2,...,v, όπου i ο αύξων αριθμός του ζεύγους (x,y) και v το πλήθος των σημείων (x,y).

Το σημειοσύνολο (x, y) μπορεί επίσης να προσεγγιστεί με πολυώνυμα τρίτου ή μεγαλύτερου βαθμού, πάντα με το κριτήριο των ελαχίστων τετραγώνων.

2.1.2. Επίπεδο Ελαχίστων Τετραγώνων

Έστω ότι μια μεταβλητή z εξαρτάται από τις μεταβλητές x,y. Αν υπάρχει μια γραμμική σχέση που συνδέει τις x,y,z, τότε αναζητούμε μια εξίσωση της μορφής :

$$z = a+bx+cy$$

που να προσεγγίζει, με το κριτήριο των ελαχίστων τετραγώνων, το σύνολο των σημείων (x,y,z). Σ'αυτήν την περίπτωση, οι συντελεστές a, b, c ορίζουν το επίπεδο των ελαχίστων τετραγώνων. Οι εξισώσεις για τον προσδιορισμό των συντελεστών του επιπέδου είναι οι :

$$\Sigma[z] = va+b\Sigma[x]+c\Sigma[y]$$
$$\Sigma[xz] = a\Sigma[x]+b\Sigma[x^2]+c\Sigma[xy]$$
$$\Sigma[yz] = a\Sigma[y]+b\Sigma[xy]+c\Sigma[y^2]$$

Υπάρχει επίσης η δυνατότητα να προσδιοριστούν μη επίπεδες επιφάνειες ελαχίστων τετραγώνων, όπου το z συνδέεται με μη γραμμικές σχέσεις με τα x, y.

2.1.3. Τυπικό Σφάλμα Εκτίμησης

Έστω y' η εκτιμώμενη τιμή του y, για μια δεδομένη τιμή της ανεξάρτητης μεταβλητής x, με βάση την ευθεία ή την καμπύλη ελαχίστων τετραγώνων. Ένα μέτρο του πόσο διασπαρμένα είναι τα σημεία (x,y) γύρω από την καμπύλη, είναι το τυπικό σφάλμα εκτίμησης του y από το x, που συμβολίζεται με $\sigma_{y,x}$ και ορίζεται από τη σχέση :

$$\sigma_{y,x} = \sqrt{\frac{\Sigma[y-y']^2}{\nu}}$$

y είναι η πειραματική τιμή που αντιστοιχεί στη μεταβλητή x.

Το τυπικό σφάλμα εκτίμησης έχει ιδιότητες ανάλογες με αυτές της τυπικής απόκλισης, με την έννοια ότι αν φέρουμε δύο ευθείες παράλληλες προς την ευθεία των ελαχίστων τεραγώνων και σε κατακόρυφες από αυτήν αποστάσεις $\sigma_{y,x}$, $2\sigma_{y,x}$, $3\sigma_{y,x}$, θα δούμε ότι για ένα μεγάλο αριθμό δεδομένων, μεταξύ των δύο αυτών ευθειών

βρίσκεται αντίστοιχα το 68.3%, το 95.5% και το 99.7% του συνόλου των σημείων (x, y).

2.1.4. Γενίκευση της Έννοιας του Συντελεστή Συσχέτισης

Η έκφραση :

$$r = \frac{\nu\Sigma[xy]-\Sigma[x]\Sigma[y]}{\sqrt{\nu\Sigma[x^2]-(\Sigma[x])^2}\sqrt{\nu\Sigma[y^2]-(\Sigma[y])^2}}$$

δίνει το συντελεστή συσχέτισης για γραμμική σχέση μεταξύ των συντελεστών x και y. Γι'αυτόν το λόγο, ο r στην παρα πάνω σχέση ονομάζεται **γραμμικός συντελεστής συσχέτισης.**

Για μεγέθη που συνδέονται με μη γραμμική σχέση (για παράδειγμα με μια εξίσωση παραβολής), θα πρέπει να οριστεί ένας γενικευμένος συντελεστής συσχέτισης, σε τρόπο ώστε να λαμβάνεται υπόψη το είδος της συσχέτισης που επιθυμούμε να διερευνήσουμε.

Για το σκοπό αυτόν, είναι σημαντικό να ληφθεί υπόψη ότι ο συντελεστής συσχέτισης r για γραμμική σχέση μεταξύ x και y εκφράζεται και ως :

$$r^2 = 1-(\Sigma[y-y']^2)/(\Sigma[y-\bar{y}]^2)$$

$\bar{y}$ είναι η μέση τιμή του συνόλου των τιμών της μεταβλητής y.

Αποδεικνύεται επίσης ότι :

$$\Sigma[y-\bar{y}]^2 = \Sigma[y-y']^2+\Sigma[y'-\bar{y}]^2$$

Η ποσότητα στο αριστερό μέλος της παραπάνω εξίσωσης, καλείται **ολική μεταβολή**, ο πρώτος όρος του δεξιού μέλους καλείται **υπόλοιπη μεταβολή** και ο δεύτερος όρος **παλινδρομική** ή **παραγοντική μεταβολή** (Spiegel, 1977). Η ορολογία αυτή μπορεί να εξηγηθεί από το ότι η ποσότητα y-y' εκφράζει την τυχαία απόκλιση από τη προβλεπόμενη τιμή y', ενώ η ποσότητα y' - $\bar{y}$ εκφράζει μια

συστηματική συμπεριφορά που υποδηλώνεται από την ευθεία των ελαχίστων τετραγώνων.

Μπορούμε λοιπόν να γράψουμε :

$r^2 = (\Sigma[y' - \bar{y}]^2)/(\Sigma[y - \bar{y}]^2)$ = (Παλινδρομική μεταβολή)/(Ολική μεταβολή)

Δηλαδή, το r^2 εκφράζει το κλάσμα της ολικής μεταβολής που συνδέεται με την ευθεία ελαχίστων τετραγώνων. Από αυτήν την έκφραση για το r^2 μπορεί να οριστεί ένας γενικευμένος συντελεστής συσχέτισης r, για γραμμική ή μη γραμμική σχέση μεταξύ των x και y, με βάση τον τύπο :

$$r = \sqrt{\frac{\Sigma[y'-\bar{y}]^2}{\Sigma[y-\bar{y}]^2}}$$

Η τελευταία αυτή σχέση χρησιμοποιείται για τον υπολογισμό συντελεστών μη γραμμικής συσχέτισης, που εκφράζουν το βαθμό συσχέτισης των δεδομένων x, y με βάση μια γραμμική ή μη γραμμική σχέση.

Φυσικά, για να ληφθούν στατιστικές αποφάσεις από όλα τα παραπάνω, πρέπει να διερευνηθούν τα επίπεδα σημαντικότητας (όρια εμπιστοσύνης) των αποτελεσμάτων (των συντελεστών που εξήχθησαν), οι βαθμοί ελευθερίας, και να χρησιμοποιηθούν κατάλληλα διαγράμματα ή πίνακες.

Μπορούμε λοιπόν τώρα να εφαρμόσουμε τα παρα πάνω σε κάποια γλώσσα προγραμματισμού. Πολύ απλούστερα, μπορούμε να χρησιμοποιήσουμε ένα λογιστικό φύλλο π.χ. το Aseasy για περιβάλλον DOS, το Lotus 123 για Windows, το Excel για Windows κ.λπ. που κάνουν αυτόματο υπολογισμό των συντελεστών της ευθείας των ελαχίστων τετραγώνων και μπορούν να υπολογίσουν καμπύλες προσαρμογής βασισμένες σε εξισώσεις όχι μόνο πρώτου αλλά και ανωτέρου βαθμού.

2.2. Στοιχεία Θεωρίας Σφαλμάτων

Η μέτρηση ενός φυσικού μεγέθους ενέχει το στοιχείο της παρέμβασης του ανθρώπου-παρατηρητή πάνω στη φύση. Η παρέμβαση αυτή, σε ατομική κλίμακα, μπορεί να προξενήσει διαταραχές στη συμπεριφορά του συστήματος, άρα και αβεβαιότητα στον προσδιορισμό της τιμής του μετρούμενου μεγέθους (πρόκειται για το γνωστό πρόβλημα της κβαντομηχανικής για την επίδραση του παρατηρητή πάνω στο αντικείμενο παρατήρησης).

Σε ένα ευρύ φάσμα μετρήσεων στο ύπαιθρο και στο εργαστήριο, η αβεβαιότητα αυτή μπορεί να θεωρηθεί πολύ μικρή, δηλαδή αμελητέα, σε σχέση με την ακρίβεια που επιθυμούμε ή που μπορούμε να επιτύχουμε με το διαθέσιμο εξοπλισμό. Υπεισέρχονται όμως σφάλματα που έχουν να κάνουν με τη συμπεριφορά του ίδιου του παρατηρητή, τις ιδιότητες και επιδόσεις των χρησιμοποιούμενων οργάνων, καθώς και με άλλους, ανεξέλεγκτους παράγοντες, που σχετίζονται με το μέσο που παρεμβάλλεται μεταξύ αντικειμένου παρατήρησης και παρατηρητή. Τέτοια σφάλματα μπορεί να είναι σημαντικά, με την έννοια ότι οι μετρούμενες τιμές μπορούν να διαφέρουν σημαντικά από τις «αληθείς» τιμές που θα λαμβάνονταν από παρατηρήσεις πολύ υψηλής ακρίβειας.

Θα πρέπει λοιπόν να γίνει αντιληπτό ότι πάντα, σε οποιοδήποτε είδος μέτρησης, υπάρχει ένα σφάλμα (μια αβεβαιότητα) στην καταγραφόμενη τιμή, το μέτρο του οποίου μας δίνει την περιοχή τιμών στην οποία εκτιμούμε πως βρίσκεται η τιμή του μετρούμενου φυσικού μεγέθους. Γνωρίζοντας αυτό το σφάλμα, και συγκρίνοντάς το με την ακρίβεια που εμείς θεωρούμε ως ικανοποιητική, σχηματίζουμε μια εικόνα για την αξιοπιστία των μετρήσεών μας και για το τι αποκλίσεις αναμένεται να υπάρχουν μεταξύ των δικών μας προβλέψεων και της συμπεριφοράς του πραγματικού κόσμου.

Αντικείμενο της θεωρίας σφαλμάτων είναι ο προσδιορισμός του μέτρου της ακρίβειας, ή, ισοδύναμα, της αβεβαιότητας, στον υπολογισμό των παραμέτρων με τις οποίες περιγράφεται η συμπεριφορά ενός φυσικού συστήματος. Ως γνωστικό αντικείμενο, η θεωρία σφαλμάτων θεμελιώνεται πάνω στη θεωρία πιθανοτήτων και στη στατιστική και εστιάζεται στο πώς και σε ποιο βαθμό οι μετρήσεις και οι υπολογισμοί φυσικών μεγεθών προσεγγίζουν ή αποκλίνουν από τη φυσική πραγματικότητα.

Παρακάτω εκτίθενται συνοπτικά κάποιες όψεις της θεωρίας σφαλμάτων με άμεσο ενδιαφέρον στη διαδικασία μέτρησης και υπολογισμού μεγεθών.

Σφάλματα, όμως, που οφείλονται σε επιπολαιότητα η απροσεξία κατά την εκτέλεση των μετρήσεων και των υπολογισμών, καθώς επίσης σφάλματα που οφείλονται σε λανθασμένους αλγορίθμους ή στη μη τήρηση των κανόνων διεξαγωγής των πειραμάτων, δεν εξετάζονται από την θεωρία σφαλμάτων.

Για να αποφεύγουμε τέτοια σφάλματα, δηλ. σφάλματα που δεν κατηγοριοποιούνται και δεν ταξινομούνται, ώστε να μπορεί να αρθεί η επίδρασή τους με την ανάλογη μεθοδολογία, οφείλουμε να δείχνουμε μεγάλη προσοχή κατά την εκτέλεση των υπολογισμών και τη διεξαγωγή των πειραματικών και στατιστικών μελετών.

2.2.1. Είδη Σφαλμάτων

Τα σφάλματα που υπεισέρχονται στις μετρήσεις μεγεθών μπορούν να διακριθούν σε τρεις κατηγορίες (Κωτσάκης, 1972): **Φανερά** (χονδροειδή)**, συστηματικά** και **τυχαία.**

Τα **φανερά σφάλματα** οφείλονται σε λάθη του παρατηρητή κατά την ανάγνωση και καταγραφή των τιμών, στην ακαταλληλότητα του οργάνου μέτρησης, σε λάθη ή παρατηρήσεις στη διαδικασία εκτέλεσης του πειράματος, καθώς και σε κάποιο έκτακτο εξωτερικό παράγοντα (για παράδειγμα, καταιγίδα κατά τη διάρκεια μέτρησης του γήινου ηλεκτρομαγνητικού πεδίου). Τα φανερά σφάλματα μπορούν να

εντοπιστούν και να εξαλειφθεί η επίδρασή τους στο μετρούμενο μέγεθος.

Τα **συστηματικά σφάλματα** είναι εκείνα που επηρεάζουν τα αποτελέσματα μιας μέτρησης κάτω από τις ίδιες συνθήκες και με την ίδια φορά. Δηλαδή, το συγκεκριμένο σφάλμα ή θα μεγαλώνει ή θα μικραίνει το αποτέλεσμα της μέτρησης. Τα συστηματικά σφάλματα μπορούν να οφείλονται στη λανθασμένη ένδειξη οργάνων, στο αντικείμενο μέτρησης, στην επίδραση του τρόπου ή του οργάνου μέτρησης, στην παράλειψη στο να λάβουμε υπόψη εξωτερικές επιδράσεις, καθώς και στον υπολογισμό μεγεθών με βάση μαθηματικούς τύπους που είναι χονδροειδείς προσεγγίσεις της φυσικής πραγματικότητας και που δεν ενδείκνυνται για το φυσικό σύστημα που παρατηρούμε.

Τα συστηματικά σφάλματα έχουν μια αναγνωρίσιμη συμπεριφορά, που επιτρέπει να γίνουν οι απαραίτητες διορθώσεις στη μέτρηση και στον υπολογισμό των μεγεθών.

Τα **τυχαία σφάλματα** έχουν ακανόνιστο (στατιστικό) χαρακτήρα και μπορούν να επηρεάσουν το αποτέλεσμα μιας μέτρησης κατά τυχαία φορά (δηλαδή μπορούν να το μεγαλώσουν ή να το μικρύνουν).

Οφείλονται τόσο σε αντικειμενικούς παράγοντες (σφάλματα λόγω κυμαινόμενων συνθηκών θερμοκρασίας, πίεσης, ακτινοβολίας και άλλων διαταραχών), όσο και σε υποκειμενικούς (σφάλματα κρίσης που γίνονται κατά την ανάγνωση της κλασματικής υποδιαίρεσης μιας κλίμακας).

Ο τυχαίος χαρακτήρας αυτών των σφαλμάτων συνηγορεί στο ότι θα πρέπει να καταφύγουμε σε στατιστικές μεθόδους για τον υπολογισμό τους.

Στις επόμενες παραγράφους εξετάζεται το πώς μπορούν να υπολογιστούν τα τυχαία σφάλματα.

2.2.2. Υπολογισμός Μέσου Σφάλματος Παρατήρησης και Μέσου Σφάλματος Μέσης Τιμής

Έστω ότι κάποιο μέγεθος x, μετράται ν φορές με το ίδιο όργανο και υπό τις ίδιες συνθήκες, οπότε λαμβάνονται οι τιμές x_1, x_2, ..., x_v. Οι τιμές αυτές γενικά διαφέρουν μεταξύ τους, λόγω του τυχαίου σφάλματος που αναπόφευκτα υπεισέρχεται στη διαδικασία της μέτρησης.

Συνήθως θεωρούμε ότι οι τιμές x_1, x_2, ..., x_v ακολουθούν μια κανονική κατανομή γύρω από μια μέση τιμή Μ, που μπορεί να θεωρηθεί ότι είναι η αληθής τιμή και προσδιορίζεται, αν διαθέτουμε ένα πολύ μεγάλο αριθμό μετρήσεων. Από την τυπική απόκλιση σ της καμπύλης προσδιορίζεται η πιθανότητα να βρίσκεται μια τιμή x_i σε διάφορα διαστήματα τιμών γύρω από το Μ. Συγκεκριμένα, η πιθανότητα να βρίσκεται η τιμή x_i στο διάστημα [Μ-σ, Μ+σ] προσδιορίζεται από το εμβαδόν της καμπύλης από Μ-σ ως Μ+σ και είναι 0.683. Για το διάστημα [Μ-2σ, Μ+2σ] η πιθανότητα είναι 0.954 και για το διάστημα [Μ-3σ, Μ+3σ] είναι 0.997.

Στην πράξη όμως δε διαθέτουμε άπειρο, αλλά πεπερασμένο, πλήθος μετρήσεων ν, με αποτέλεσμα να μην υπολογίζουμε την αληθή τιμή Μ, αλλά μια αριθμητική μέση τιμή $\bar{x}$, με βάση τον τύπο :

$$\bar{x} = \Sigma[x_i] / v$$

όπου γενικά το $\bar{x}$ διαφέρει από το Μ. Από το σύνολο των μετρούμενων τιμών του x, υπολογίζουμε επίσης και μια **τυπική απόκλιση σ_v**, ή **μέσο σφάλμα παρατήρησης**, με βάση τον τύπο :

$$\sigma_v = \sqrt{\frac{\Sigma[\bar{x}-x_i]^2}{v-1}}$$

Το φαινομενικά παράδοξο να τοποθετηθεί παρονομαστής ν-1 στην έκφραση για το σ_v, αντί για ν, εξηγείται από το ότι οι διαφορές $\bar{x}-x_i$ δεν είναι οι αποκλίσεις από την αληθή τιμή Μ, αλλά από τον αριθμητικό μέσο $\bar{x}$. Αποδεικνύεται μάλιστα ότι

$$\Sigma[\bar{x}-x_i]^2 < \Sigma[M-x_i]^2 \qquad \text{(Κωτσάκης, 1972)}$$

Για το λόγο αυτόν, προκειμένου να έχουμε μια αξιόπιστη, κατά το δυνατόν, εκτίμηση του παρατηρησιακού σφάλματος, θέτουμε στον παρονομαστή ν-1 αντί για ν, οπότε αυξάνεται η αριθμητική τιμή του κλάσματος.

Ωστόσο, ένα διαφορετικό δείγμα τιμών x_i, με ν το πλήθος μετρήσεις, θα έδινε γενικά ένα διαφορετικό μέσο $\bar{x}$. Σύμφωνα με σχετικό θεώρημα της στατιστικής, οι διάφοροι αριθμητικοί μέσοι των αντίστοιχων δειγμάτων ακολουθούν επίσης μια κανονική κατανομή, η διασπορά της οποίας είναι η μέση τιμή των διασπορών των δειγμάτων, με αποτέλεσμα η αντίστοιχη τυπική απόκλιση της μέσης τιμής σ να δίνεται από τη σχέση :

$$\Delta\bar{x} = \sigma = \sqrt{\frac{\Sigma[\bar{x}-x_i]^2}{\nu(\nu-1)}}$$

Κατά συνέπεια, από ένα σύνολο μετρήσεων του μεγέθους x, υπολογίζουμε τον αριθμητικό μέσο $\bar{x}$, που ονομάζεται και **μέση τιμή** $\bar{x}$, και την τυπική απόκλιση της μέσης τιμής, ή **μέσο σφάλμα της μέσης τιμής**, που συμβολίζεται με **σ** ή με $\Delta\bar{x}$. Συχνά, το $\Delta\bar{x}$ ονομάζεται απλώς **σφάλμα** του x.

Το αποτέλεσμα των μετρήσεων του μεγέθους x αναγράφεται, μαζί με το σφάλμα, με τη μορφή

$$\bar{x} \pm \Delta\bar{x}$$

και με κατάλληλη "στρογγυλοποίηση", σε τρόπο ώστε το σημαντικό ψηφίο του $\bar{x}$ να είναι σύμφωνο με την τάξη μεγέθους του μέσου σφάλματος της μέσης τιμής. Αυτό μπορεί να γίνει κατανοητό με το παρακάτω παράδειγμα.

Παράδειγμα : Μετράμε 10 φορές το μήκος l ενός νήματος. Τα αποτελέσματα παρουσιάζονται στον παρακάτω πίνακα

l_i (cm)
97.30
97.40

97.35
97.45
97.30
97.40
97.35
97.40
97.30
97.35

Από τις τιμές αυτές βρίσκουμε μέση τιμή μήκους

$$\bar{l} = \Sigma[l_i]/10 = 97.36\text{cm}$$

Το μέσο σφάλμα της μέσης τιμής είναι

$$\sigma = \sqrt{\frac{\Sigma[\bar{l}\text{-}l_i]^2}{90}} = 0.016\text{cm}$$

Με κατάλληλη στρογγυλοποίηση καταλήγουμε ότι

$$l = (97.36 \pm 0.02)\ \text{cm}$$

2.2.3. Σφάλμα στον Υπολογισμό Ποσότητας που Προσδιορίζεται από Έμμεσες Παρατηρήσεις

Έστω u ποσότητα που δε μετράται άμεσα, αλλά υπολογίζεται από άλλες άμεσα μετρήσιμες ποσότητες a, b, c, ... με βάση τη μαθηματική σχέση :

$$u = f(a, b, c, ...)$$

Έστω $\Delta\bar{a}$, $\Delta\bar{b}$, $\Delta\bar{c}$, ..., τα σφάλματα στη μέτρηση των παραμέτρων $\bar{a}$, $\bar{b}$, $\bar{c}$, ..., αντίστοιχα. Αποδεικνύεται ότι το σφάλμα $\Delta\bar{u}$ στον υπολογισμό της ποσότητας u δίδεται από τη σχέση :

$$\Delta\bar{u} = \sqrt{(\frac{\partial f}{\partial a}\Delta\bar{a})^2 + (\frac{\partial f}{\partial b}\Delta\bar{b})^2 + (\frac{\partial f}{\partial c}\Delta\bar{c})^2 + ...}$$

Η μέση τιμή $\bar{u}$ υπολογίζεται από τη σχέση :

$$\bar{u} = f(\bar{a}, \bar{b}, \bar{c}, ...)$$

Παράδειγμα: Με τη βοήθεια του απλού εκκρεμούς μετράμε την επιτάχυνση της βαρύτητας g, με βάση τον τύπο :

$$g = 4\pi^2 l/T^2$$

όπου l το μήκος του εκκρεμούς και T η περίοδος ταλάντωσης.

Εκτελώντας επανειλημμένα άμεσες μετρήσεις των l και T, βρίσκουμε τις μέσες τιμές $\bar{l}$ και $\bar{T}$, και τα αντίστοιχα σφάλματα $\Delta\bar{l}$, $\Delta\bar{T}$. Έστω ότι βρίσκουμε :

$$\bar{l} \pm \Delta\bar{l} = (97.36 \pm 0.02)\ \text{cm}$$

$$\bar{T} \pm \Delta\bar{T} = (1.981 \pm 0.001)\ \text{sec}$$

Οπότε η μέση τιμή $\bar{g}$ είναι :

$$\bar{g} = 4\pi^2\ \bar{l}/\ \bar{T} = 979.43\ \text{cm/sec}^2$$

και το μέσο σφάλμα της μέσης τιμής $\Delta\bar{g}$ είναι :

$$\Delta\bar{g} = \sqrt{(\frac{\partial g}{\partial l}\Delta\bar{l})^2 + (\frac{\partial g}{\partial T}\Delta\bar{T})^2} = \sqrt{(\frac{4\pi^2}{\bar{T}^2}\Delta\bar{l})^2 + (\frac{8\pi^2\bar{l}}{\bar{T}^3}\Delta\bar{T}^2)}$$

και τελικά :

$$\Delta\bar{g} = 1.009\ \text{cm/sec}^2$$

Επομένως, στρογγυλοποιώντας το αποτέλεσμα, καταλήγουμε ότι η επιτάχυνση της βαρύτητας είναι :

$$\bar{g} \pm \Delta\bar{g} = (979 \pm 1)\ \text{cm/sec}^2$$

Αυτό σημαίνει ότι αν το πείραμα προσδιορισμού του g εκτελεστεί από 100 πειραματιστές, οι 68 θα υπολογίσουν επιτάχυνση βαρύτητας από 978 ως 980 cm/sec^2.

2.2.4. Υπολογισμός Σφαλμάτων των Συντελεστών της Ευθείας Ελαχίστων Τετραγώνων.

Αν υπάρχει μια γραμμική σχέση μεταξύ των μετρούμενων μεγεθών x και y, ώστε :

$$y = a+bx$$

οι συντελεστές a, b μπορούν να προσδιοριστούν από τα ζεύγη τιμών (x_i, y_i) με βάση τη μέθοδο των ελαχίστων τετραγώνων. Αν το σφάλμα στις τιμές x_i είναι μικρό σε σχέση με το σφάλμα των y_i, τότε μπορούν να προσδιοριστούν οι μέσες τιμές $\bar{a}$, $\bar{b}$, από τις οποίες λαμβάνονται οι εκτιμώμενες τιμές y', με βάση τη σχέση :

$$y' = \bar{a} + \bar{b}x$$

Όπως αναφέρθηκε στην ενότητα 23.15., οι συντελεστές $\bar{a}$, $\bar{b}$ υπολογίζονται από τις σχέσεις :

$$\bar{a} = \frac{\Sigma[y_i]\Sigma[x_i^2] - \Sigma[x_i]\Sigma[x_i y_i]}{\nu\Sigma[x_i^2] - (\Sigma[x_i])^2}$$

$$\bar{b} = \frac{\nu\Sigma[x_i y_i] - \Sigma[x_i]\Sigma[y_i]}{\nu\Sigma[x_i^2] - (\Sigma[x_i])^2}$$

Τα σφάλματα $\Delta\bar{a}$, $\Delta\bar{b}$ είναι (Κωτσάκης, 1972) :

$$\Delta\bar{a} = \Delta\bar{y}\sqrt{\frac{\Sigma[x_i^2]}{\nu\Sigma[x_i^2] - (\Sigma[x_i])^2}}$$

$$\Delta\bar{b} = \Delta\bar{y}\sqrt{\frac{\nu}{\nu\Sigma[x_i^2] - (\Sigma[x_i])^2}}$$

όπου :

$$\Delta\bar{y} = \sqrt{\frac{\Sigma[y_i - \bar{a} - \bar{b}x_i]^2}{\nu-2}}$$

2.3. Φίλτρα Επεξεργασίας Χρονοσειρών και Καμπυλών Μεταβολής Φυσικών Μεγεθών

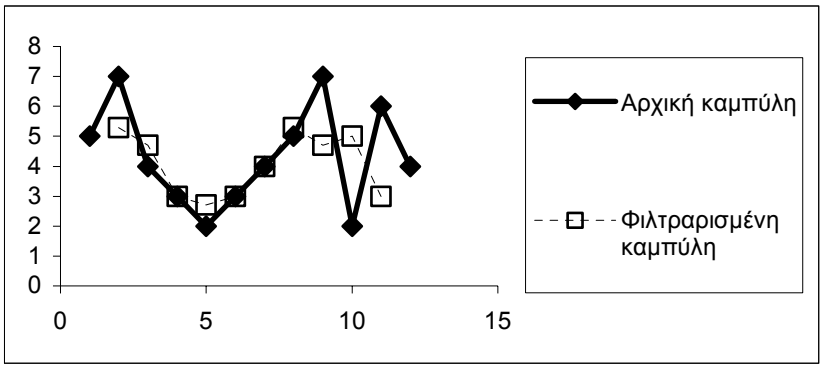
Συχνά, μια χρονοσειρά ή μια καμπύλη μεταβολής φυσικού μεγέθους, όπως π.χ. η μεταβολή του μαγνητικού πεδίου κατά μήκος μιας τομής στην επιφάνεια του εδάφους, μπορεί να θεωρηθεί ότι αποτελείται από μια σχετικά αργά μεταβαλλόμενη συνιστώσα και από μια ταχύτερα μεταβαλλόμενη συνιστώσα, που αποκαλείται **υψίσυχνος θόρυβος.** Αν ενδιαφέρει η ανάδειξη της σχετικά αργά μεταβαλλόμενης συνιστώσας, θα πρέπει να εξομαλυνθεί η καμπύλη, απομακρύνοντας τον υψίσυχνο θόρυβο. Η εξομάλυνση της καμπύλης μπορεί να πραγματοποιηθεί με την αντικατάσταση της κάθε τιμής y_i του μετρούμενου μεγέθους με μια ποσότητα y_i', που δίνεται από τη σχέση :

$$y_i' = (y_{i-m} + y_{i-m+1} + \ldots + y_i + y_{i+1} + \ldots + y_{i+m})/(2m+1)$$

όπου m φυσικός αριθμός.

Στην περίπτωση αυτή έχουμε έναν **κινούμενο μέσο όρο**, που λειτουργεί ως **φίλτρο εξομάλυνσης.** Το πλήθος των όρων στον αριθμητή της παραπάνω σχέσης ορίζει το **παράθυρο** του φίλτρου που εφαρμόζεται πάνω στην καμπύλη.

Στο παρακάτω σχήμα φαίνεται η επίδραση ενός φίλτρου εξομάλυνσης (κινούμενου μέσου όρου με m=1) σε καμπύλη. Μπορεί κανείς να δει μια μετατόπιση των κορυφών της φιλτραρισμένης καμπύλης σε σχέση με την αρχική. Αυτό μπορεί να έχει ανεπιθύμητα αποτελέσματα στην ερμηνεία των φιλτραρισμένων δεδομένων.



Επίδραση φίλτρου κινούμενου μέσου όρου σε καμπύλη.

Στο βαθμό που αυξάνεται το εύρος του παραθύρου (δηλαδή αυξάνεται το m), περιορίζονται οι διακυμάνσεις της καμπύλης, δηλαδή έχουμε μεγαλύτερη εξομάλυνση. Υπάρχει όμως ο κίνδυνος, μαζί με τον ανεπιθύμητο υψίσυχνο θόρυβο, να χαθεί και μέρος του σήματος που επιθυμούμε να αναδείξουμε. Για το λόγο αυτό, θα πρέπει το μέγεθος του παραθύρου να ρυθμίζεται κατάλληλα, για την κάθε καμπύλη που πρόκειται να υποστεί επεξεργασία. Για τον προσδιορισμό του εύρους του παραθύρου, επιστρατεύεται συχνά η διαίσθηση και η εμπειρία, καθώς δεν υπάρχει αξιόπιστος γενικός κανόνας που να υποδεικνύει το κατάλληλο φίλτρο για την εκάστοτε περίπτωση.

Συχνά, κατά την εξομάλυνση, είναι επιθυμητό να έχει το y_i μεγαλύτερη συμβολή στη διαμόρφωση του y_i' απ'όσο οι γειτονικές τιμές $y_{i\pm j}$. Για το σκοπό αυτό, το y_i πολλαπλασιάζεται με ένα σχετικά υψηλό συντελεστή βαρύτητας, ενώ οι συντελεστές βαρύτητας των $y_{i\pm j}$ είναι σχετικά μικρότεροι. Η τιμή του παρονομαστή του φίλτρου εξομάλυνσης είναι ίση με το αλγεβρικό άθροισμα των συντελεστών βαρύτητας. Οι σχέσεις :

$$y_i'=[17y_i+12(y_{i+1}+y_{i-1})-3(y_{i+2}+y_{i-2})]/35$$

και

$$y_i'=[7y_i+6(y_{i+1}+y_{i-1})+3(y_{i+2}+y_{i-2})-2(y_{i+3}+y_{i-3})]/21$$

εκφράζουν δυο φίλτρα με 5 και 7 όρους, αντίστοιχα.

Γενικά, **φίλτρο** ονομάζεται η μαθηματική σχέση με την οποία προσδιορίζεται το y_i', ενώ η **απόκριση του φίλτρου** προσδιορίζεται από τις τιμές των συντελεστών βαρύτητας.

Υπάρχουν περιπτώσεις όπου δεν επιθυμούμε να εξομαλύνουμε την καμπύλη, αλλά αντίθετα να την οξύνουμε, δίνοντας έμφαση στις διακυμάνσεις της. Για το σκοπό αυτό, μπορεί να υπολογιστεί η **βαθμίδα** της καμπύλης σε κάθε σημείο της, με βάση το φίλτρο :

$$y_i' = y_i - y_{i-1}$$

ή με το φίλτρο :

$$y_i' = 2y_{i+2} + y_{i+1} - y_{i-1} - 2y_{i-2}$$

το παράθυρο του οποίου αποτελείται από 5 όρους.

Αξίζει να σημειωθεί ότι στα φίλτρα εξομάλυνσης το αλγεβρικό άθροισμα των συντελεστών βαρύτητας είναι ένας θετικός αριθμός, ενώ στα φίλτρα βαθμίδας το αλγεβρικό άθροισμα είναι μηδέν. Γενικά, στα φίλτρα όξυνσης μιας καμπύλης, το αλγεβρικό άθροισμα των συντελεστών βαρύτητας βρίσκεται κοντά στο μηδέν, αν δεν είναι μηδέν.

Είναι προφανές ότι αν το παράθυρο του φίλτρου αποτελείται από 2m+1 όρους, τα πρώτα m και τα τελευταία m σημεία της χρονοσειράς, ή της καμπύλης μεταβολής φυσικού μεγέθους, δεν είναι δυνατό να φιλτραριστούν. Τα σημεία αυτά, ή θα διατηρήσουν τις τιμές τους ή θα υποστούν επεξεργασία με φίλτρο, το παράθυρο του οποίου έχει μικρότερο εύρος.

Τα φίλτρα επινοήθηκαν αρχικά από τους ηλεκτρολόγους, για την επεξεργασία ηλεκτρικών σημάτων. Στη συνέχεια όμως εφαρμόστηκαν σε μεγάλη κλίμακα σε διαφόρους τομείς των γεωεπιστημών, όπως επεξεργασία σεισμικών σημάτων, μετρήσεων βαρυτικού, μαγνητικού ή ηλεκτρομαγνητικού πεδίου, καθώς και δεδομένων από γεωτρήσεις.

2.4. Ανάλυση Fourier

Έστω η περιοδική συνάρτηση g(x) (σήμα g(x)) με περίοδο T. Το σήμα αυτό μπορεί να αναλυθεί σε άθροισμα τριγωνομετρικών συναρτήσεων της μορφής :

$$g(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos\frac{2n\pi x}{T} + b_n \sin\frac{2n\pi x}{T})$$

Τα a_0, a_n, b_n προσδιορίζονται από τις σχέσεις :

$$a_0 = \frac{2}{T}\int_{-T/2}^{T/2} g(x)dx$$

$$a_n = \frac{2}{T}\int_{-T/2}^{T/2} g(x)\cos\frac{2n\pi x}{T}dx$$

$$b_n = \frac{2}{T}\int_{-T/2}^{T/2} g(x)\sin\frac{2n\pi x}{T}dx$$

αντίστοιχα.

Κάθε ζεύγος a_n, b_n εκφράζει ένα κύμα, που περιγράφεται από τη σχέση:

$y_n(x) = a_n\cos(2n\pi x/T) + b_n\sin(2n\pi x/T)$

Η διαδικασία προσδιορισμού των συντελεστών a_0, a_n, b_n καλείται **ανάλυση Fourier**, ή **ανάλυση συχνοτήτων.** Στο κάθε ζεύγος τιμών a_n, b_n αντιστοιχεί ένας ακέραιος αριθμός n, που είναι η **τάξη αρμονικής** και συνδέεται με τη συχνότητα f_n του αντίστοιχου κύματος με τη σχέση :

$f_n = n/T$

Για κάθε τάξη αρμονικής n, και αντίστοιχη συχνότητα f_n, αντιστοιχεί ένα **πλάτος** A_n, μια **φάση** φ_n και μια **ισχύς** P_n, που ορίζονται από τις σχέσεις :

$A_n = (a_n^2 + b_n^2)^{1/2}$

$\varphi_n = \arctan(b_n / a_n)$

$P_n = a_n^2 + b_n^2$

Το πλάτος και η ισχύς εκφράζουν το πόσο συμβάλλει το κύμα στη διαμόρφωση του σήματος και η φάση εκφράζει τη σχετική μετατόπιση του κύματος ως προς τα άλλα κύματα. Ειδικά, η ισχύς εκφράζει την ενέργεια που περιέχει το κύμα που αντιστοιχεί σε μια συγκεκριμένη συχνότητα.

Ένα μη περιοδικό σήμα g(x), μπορεί να θεωρηθεί ότι έχει περίοδο $T \to \infty$ και να αναλυθεί κατά Fourier με βάση τη σχέση :

$$G(f) = \int_{-\infty}^{\infty} g(x).e^{-i2\pi fx}dx$$ (μετασχηματισμός Fourier)

όπου f η συχνότητα.

Το g(x) μπορεί να θεωρηθεί ότι είναι ένα "άθροισμα" κυμάτων της μορφής $g(f)\cdot e^{i2\pi fx}$, το οποίο εκφράζεται ως γενικευμένο ολοκλήρωμα:

$$g(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(f).e^{i2\pi fx}df$$ (αντίστροφος μετασχηματισμός Fourier)

Ως τώρα, έγινε λόγος για συνεχή σήματα, δηλαδή για συναρτήσεις συνεχών μεταβλητών. Ωστόσο, ένα ψηφιακά αποθηκευμένο σεισμογράφημα, ή ένα σύνολο τιμών του βαρυτικού πεδίου κατά μήκος μιας τομής στο ύπαιθρο, είναι συναρτήσεις διακριτών μεταβλητών. Στην πρώτη περίπτωση, η καταγραφή του φυσικού μεγέθους γίνεται ανά συγκεκριμένο χρονικό διάστημα Δt και στη δεύτερη

ανά συγκεκριμένη απόσταση Δx. Το διακριτό σήμα g(x), που έχει ένα πεπερασμένο μήκος, δε θεωρείται περιοδικό και η ανάλυση Fourier πραγματοποιείται με αριθμητικές μεθόδους, με τις οποίες προσδιορίζεται το ολοκλήρωμα G(f), που ορίστηκε παρα πάνω. Όπως και στην περίπτωση του συνεχούς περιοδικού σήματος, κι'εδώ μπορούν, με βάση τις ίδιες σχέσεις, να υπολογιστούν το πλάτος, η φάση και η ισχύς για την κάθε συχνότητα f_n. Η συχνότητα f_n είναι κι'εδώ :

$f_n=n/T$

όπου όμως το T είναι το συνολικό μήκος του σήματος g(x). Το n είναι, και σ'αυτήν την περίπτωση, η τάξη αρμονικής και λαμβάνει τιμές :

n=0, 1, 2,..., N-1

όπου N είναι το πλήθος των καταγραφών από τις οποίες αποτελείται το σήμα.

Κατά την ανάλυση Fourier ενός σήματος, λαμβάνεται το **φάσμα συχνοτήτων** του. Το σύνολο των τιμών A_n είναι το **φάσμα πλατών** (amplitude spectrum) και το σύνολο των τιμών P_n είναι το **φάσμα ισχύων** (power spectrum). Ένα φάσμα πλατών, όπως και ένα φάσμα ισχύων, χαρακτηρίζεται από κορυφές (σημεία μεγίστου), που παρέχουν πληροφορίες για το ποιές κυρίως συχνότητες διαμορφώνουν το σήμα. Αν, για παράδειγμα, υπάρχει κορυφή σε τάξη αρμονικής n, τότε η συχνότητα $f_n=n/T$ συμβάλλει σημαντικά στη διαμόρφωση του σήματος και η περίοδος του αντίστοιχου κύματος είναι T/n.

Η μέγιστη συχνότητα που μπορεί να αναγνωριστεί σε ένα σήμα που αναλύεται κατά Fourier ονομάζεται **συχνότητα Nyquist** (f_N) και είναι ίση με το μισό της **συχνότητας δειγματοληψίας** (sampling frequency) f_s, με την οποία καταγράφηκε το σήμα. Αυτό οφείλεται στο ότι για να αναγνωριστεί ένα κύμα χρειάζονται τουλάχιστον δυο μετρήσεις (μια στην κορυφή και μια στην κοιλάδα του κύματος). Επομένως, αν για το σήμα g(x) υπάρχουν N καταγραφές, η ερμηνεία της καμπύλης που προκύπτει από την ανάλυση Fourier πραγματοποιείται με τα πρώτα N/2 σημεία μόνο. Ωστόσο, συχνότητες μεγαλύτερες του f_N μπορούν να εκφραστούν στο φάσμα Fourier μέσα από μικρότερες συχνότητες και να

οδηγήσουν σε εσφαλμένη ερμηνεία του φάσματος. Το φαινόμενο αυτό ονομάζεται **φαινόμενο επικάλυψης** (aliasing effect) και για να αποφευχθεί θα πρέπει να είναι η συχνότητα Nyquist αρκετά μεγάλη, ώστε οι συχνότητες με σημαντική συνεισφορά στη διαμόρφωση του σήματος να βρίσκονται μέσα στο διάστημα [0, f_N]. Για να εξασφαλιστεί μια τέτοια συνθήκη χρειάζεται μια a priori γνώση για το φασματικό περιεχόμενο του σήματος, καθώς και ένα αρκετά μικρό διάστημα δειγματοληψίας, που να αντιστοιχεί στην κατάλληλη συχνότητα Nyquist.

Η ανάλυση Fourier, εκτός από την αναγνώριση χαρακτηριστικών συχνοτήτων του σήματος, μπορεί επίσης να αξιοποιηθεί και στην επεξεργασία του σήματος, με σκοπό να απαλλαγεί αυτό από τον ανεπιθύμητο θόρυβο. Αυτό μπορεί να γίνει αν, στην καμπύλη που προκύπτει από την ανάλυση Fourier, αναγνωριστούν οι συχνότητες που αντιστοιχούν στο θόρυβο και αποκοπούν από τις υπόλοιπες, οπότε μπορεί να αποκατασταθεί το αρχικό σήμα, με έναν αντίστροφο μετασχηματισμό Fourier.

2.4.1. Παράδειγμα

Μια χρονοσειρά αποτελείται από N = 256 καταγραφές, οι οποίες ελήφθησαν με διάστημα δειγματοληψίας (sampling interval) Δt = 2 sec. Η χρονοσειρά αναλύθηκε κατά Fourier και βρέθηκε ότι στο φάσμα πλατών της υπάρχει κορυφή που αντιστοιχεί σε τάξη αρμονικής n = 30. Να βρεθούν: α) Το μήκος T της χρονοσειράς, β) η συχνότητα δειγματοληψίας f_s, γ) η συχνότητα Nyquist f_N, δ) η συχνότητα f_3 του κύματος που αντιστοιχεί στον κύκλο τάξης 30, ε) η περίοδος T_{30} του κύματος που αντιστοιχεί στην τάξη αρμονικής 30.

Λύση:

α) Το μήκος της χρονοσειράς είναι :

$T = (N-1)\cdot\Delta t = (256-1)\cdot 2 = 510$ sec

β) Η συχνότητα δειγματοληψίας είναι :

$f_s = 1/\Delta t = 1/2 = 0.5$ c/sec

γ) Η συχνότητα Nyquist είναι :

$f_N = f_s/2 = 0.5/2 = 0.25$ c/sec

δ) η συχνότητα του κύματος που αντιστοιχεί στην τάξη αρμονικής 30 είναι :

$f_{30} = 30/T = 30/510 = 0.059$ c/sec

ε) η περίοδος του κύματος που αντιστοιχεί στην τάξη αρμονικής 30 είναι:

$T_{30} = T/30 = 510/30 = 17$ sec

3. ΓΕΩΣΤΑΤΙΣΤΙΚΗ

3.1. Εισαγωγή

Στη στατιστική της μιας μεταβλητής, εξετάζεται η συμπεριφορά ενός πληθυσμού ή ενός δείγματος μετρήσεων ενός μόνο φυσικού μεγέθους. Στατιστική μιας μεταβλητής εφαρμόζεται, για παράδειγμα, στην περίπτωση όπου έχουμε να επεξεργαστούμε μετρήσεις θερμοκρασίας ενός μετεωρολογικού σταθμού σε έναν τόπο. Υπάρχουν όμως περιπτώσεις όπου στα δεδομένα των μετρήσεων υπεισέρχεται και ο γεωγραφικός παράγοντας, που συνήθως εκφράζεται με τις συντεταγμένες Χ και Υ των σταθμών όπου μετράται το φυσικό μέγεθος. Για παράδειγμα, οι μετρήσεις θερμοκρασίας σε διάφορους σταθμούς, με διαφορετικές χωρικές συντεταγμένες, συνιστούν ένα σύνολο δεδομένων όπου, πέρα από τις τιμές του μετρούμενου φυσικού μεγέθους, υπεισέρχεται και η γεωγραφική διάσταση. Η επεξεργασία δεδομένων με χωρικές συντεταγμένες αποτελεί αντικείμενο της *γεωστατιστικής*.

Οι θέσεις όπου καταγράφεται ένα γεγονός, ή μετράται ένα φυσικό μέγεθος, συχνά αναπαριστάνονται ως σημεία στο γεωγραφικό χώρο των δυο διαστάσεων. Ένα τέτοιο σημειοσύνολο μπορεί να περιέχει μόνο *γεωγραφική πληροφορία* (π.χ. συντεταγμένες ηφαιστείων σε έναν ευρύτερο γεωγραφικό χώρο), ή να περιέχει γεωγραφική πληροφορία μαζί με πληροφορία για την τιμή φυσικών μεγεθών (π.χ. τιμές ραδιενέργειας σε δίκτυο σταθμών μέτρησης που καλύπτει μια ευρύτερη περιοχή).

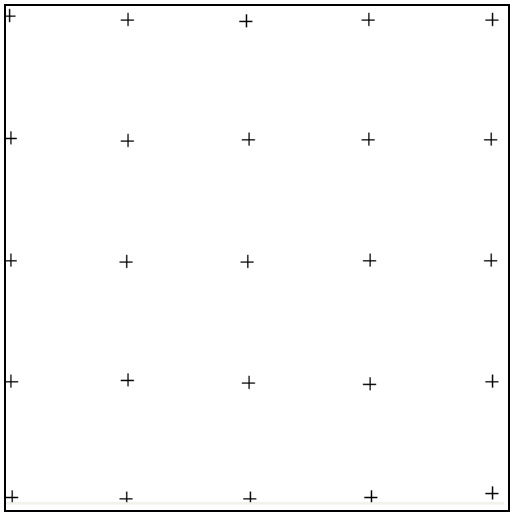
Στην πρώτη περίπτωση (σημειοσύνολα με χωρική μόνο πληροφορία), η γεωστατιστική εξετάζει προβλήματα του τύπου: κατά πόσον η δεδομένη κατανομή σημείων στο χώρο είναι ομοιόμορφη (ή τυχαία, ή ομαδοποιημένη); Τέτοιου είδους προβλήματα εμπίπτουν στην *ανάλυση χωρικών προτύπων*.

Στη δεύτερη περίπτωση (σημειοσύνολα με τιμές φυσικού μεγέθους και με γεωγραφικές συντεταγμένες), η γεωστατιστική εξετάζει προβλήματα *χωρικής παρεμβολής* του τύπου: πώς μπορούν να υπολογιστούν οι τιμές ενός φυσικού μεγέθους σε όλες τις θέσεις μιας περιοχής, αν

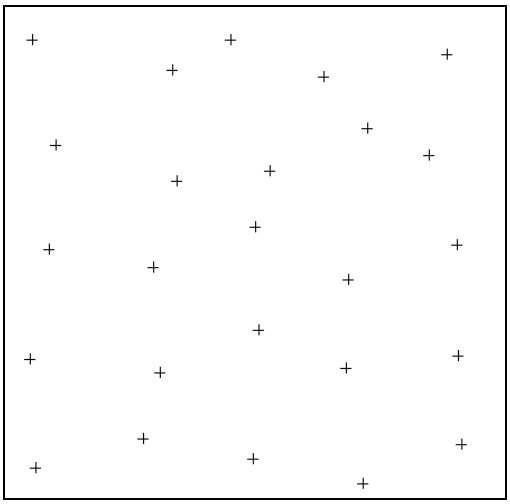
έχουμε στη διάθεσή μας τιμές μόνο σε ορισμένες θέσεις και πόσο ακριβείς θα είναι οι υπολογιζόμενες τιμές;
Η μεθοδολογία της γεωστατιστικής θεμελιώνεται πάνω στη στατιστική της μιας μεταβλητής (κατανομές, στατιστικές παράμετροι, διαστήματα εμπιστοσύνης, έλεγχοι υποθέσεων). Μια καλή γνώση της στατιστικής της μιας μεταβλητής είναι απαραίτητη για να κατανοήσει κανείς τις μεθόδους της γεωστατιστικής.

3.2. Ανάλυση χωρικών προτύπων

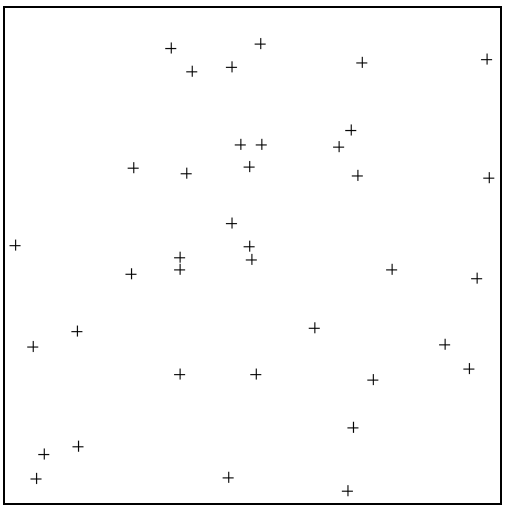
Χωρικό πρότυπο είναι ο τρόπος με τον οποίο τα σημεία κατανέμονται στο χώρο. Τα σημεία μπορούν να εκφράζουν θέσεις όπου εξελίσσονται ανθρωπογενείς δραστηριότητες (αστικά κέντρα, εμπορικά καταστήματα, εγκλήματα), φυσικά φαινόμενα (επίκεντρα σεισμών, ηφαίστεια, εστίες πυρκαϊών), γεωλογικές διεργασίες (θέσεις απολιθωμάτων, φλεβίδια ορυκτών σε πετρώματα-ξενιστές, πετρελαϊκά κοιτάσματα σε μια ευρύτερη περιοχή) και πολλές άλλες διαδικασίες.
Η χωρική κατανομή των σημείων μπορεί να είναι *κανονική* (κανονικό χωρικό πρότυπο, σχ. 1), *ομοιόμορφη* (σχ. 2), *τυχαία* (σχ. 3), *ομαδοποιημένη* (σχ. 4) και *ανισότροπη* (σχ. 5). Σε μια κανονική χωρική κατανομή τα σημεία απέχουν μεταξύ τους καθορισμένες αποστάσεις. Σε μια ομοιόμορφη κατανομή η πυκνότητα των σημείων είναι σταθερή, σε όλη την επιφάνεια.



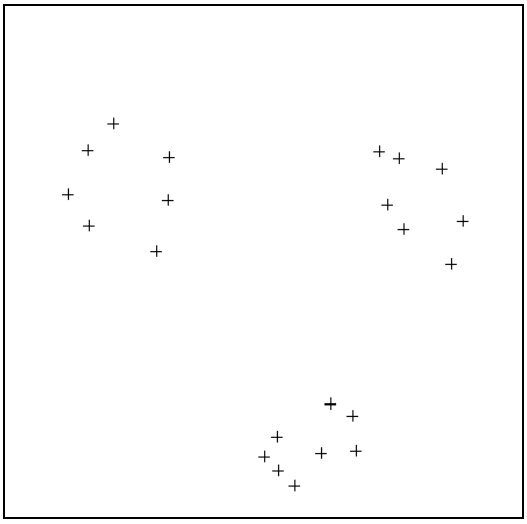
Σχ. 1. Κανονική χωρική κατανομή



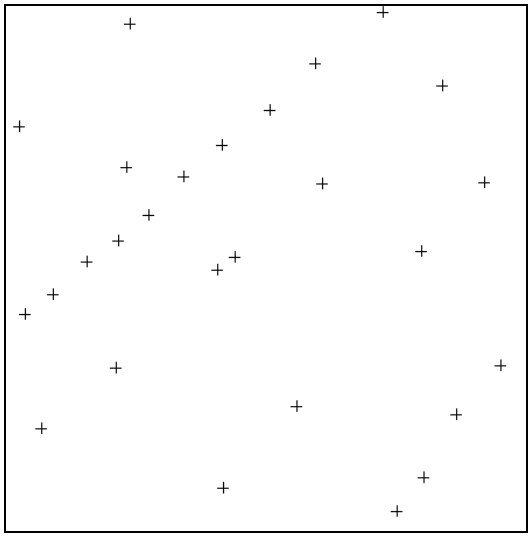
Σχ. 2. Ομοιόμορφη χωρική κατανομή



Σχ. 3. Τυχαία χωρική κατανομή



Σχ. 4. Ομαδοποιημένη χωρική κατανομή



Σχ. 5. Ανισότροπη χωρική κατανομή

Σε μια τυχαία χωρική κατανομή, κάθε σημείο έχει την ίδια πιθανότητα εμφάνισης σε οποιαδήποτε θέση και, επιπλέον, η θέση ενός σημείου είναι ανεξάρτητη από τη θέση των άλλων σημείων. Σε μια ομαδοποιημένη χωρική κατανομή τα σημεία είναι επιλεκτικά συγκεντρωμένα σε ορισμένες περιοχές. Τέλος, σε μια ανισότροπη

κατανομή τα σημεία τείνουν να σχηματίζουν γραμμικούς σχηματισμούς, ή η πυκνότητα σημείων διαφέρει με το αζιμούθιο.
Αντικείμενο της ανάλυσης χωρικών προτύπων είναι η αναγνώριση του προτύπου που ακολουθεί μια δεδομένη κατανομή σημείων. Για το σκοπό αυτό ακολουθούνται δυο βασικές κατηγορίες μεθόδων: οι *μέθοδοι καννάβου* και οι *μέθοδοι απόστασης* (Κουτσόπουλος 2002).

3.2.1. Μέθοδοι ανάλυσης καννάβου χωρικών κατανομών

Στις μεθόδους ανάλυσης *καννάβου*, η επιφάνεια με τα σημεία χωρίζεται σε ορθογώνια φατνία. Για το κάθε φατνίο *j* μετράται και καταγράφεται ο αριθμός x_j των σημείων που βρίσκονται μέσα σ'αυτό, ή το πλήθος των φατνίων με ένα δεδομένο αριθμό σημείων. Στη συνέχεια, συγκρίνονται οι τιμές x_j με τις θεωρητικά προβλεπόμενες (αναμενόμενες) τιμές α_j, που αντιστοιχούν σε μια δεδομένη χωρική κατανομή (τυχαία, ομοιόμορφη, ή κάποιου άλλου τύπου). Αν, με κάποιο στατιστικό κριτήριο (συνήθως χ^2, t ή z), συμφωνούν οι τιμές x_j και α_j, τότε εκτιμάται ότι η εξεταζόμενη χωρική κατανομή με τιμές x_j ακολουθεί ένα καθορισμένο πρότυπο (π.χ. τυχαίο ή ομοιόμορφο, ανάλογα με βάση ποια χωρική κατανομή υπολογίστηκαν τα α_j). Αν υπάρχει σημαντική απόκλιση μεταξύ τιμών x_j και α_j, τότε εκτιμάται ότι η εξεταζόμενη χωρική κατανομή *δεν* ακολουθεί το χωρικό πρότυπο με βάση το οποίο υπολογίστηκαν οι τιμές α_j. Παρακάτω παρουσιάζονται δυο μέθοδοι ανάλυσης *καννάβου*, χωρικού προτύπου.

3.2.1.α. Έλεγχος χ^2 ομοιομορφίας χωρικής κατανομής

Στον έλεγχο χ^2, η βασική υπόθεση H_0 είναι: η υπό εξέταση χωρική κατανομή είναι ομοιόμορφη.
Η εναλλακτική υπόθεση H_1 είναι: η υπό εξέταση χωρική κατανομή δεν είναι ομοιόμορφη.
Το μέγεθος μέσω του οποίου πραγματοποιείται ο έλεγχος χ^2 είναι η ποσότητα x^2 που υπολογίζεται από τη σχέση:

$$x^2 = \sum_{j=1}^{T} \frac{(x_j - a)^2}{a} \qquad (1)$$

T είναι το πλήθος των φατνίων. α είναι η πυκνότητα σημείων ανά φατνίο για μια πρότυπη ομοιόμορφη χωρική κατανομή και υπολογίζεται από τη σχέση:

$$\alpha = n/T \qquad (2)$$

όπου n είναι ο αριθμός των σημείων της υπό εξέταση χωρικής κατανομής.

Εφαρμογή

Δείγμα πετρώματος φωτογραφήθηκε και η ψηφιακή εικόνα χωρίστηκε σε T = 15 φατνία, για το καθένα από τα οποία μετρήθηκε το πλήθος x_j των απολιθωμάτων, που είναι μικρών διαστάσεων και μπορούν να θεωρηθούν σημειακά. Το πλήθος των σημείων n = 255. Να εξεταστεί, σε επίπεδο σημαντικότητας 5%, το κατά πόσον η χωρική κατανομή των απολιθωμάτων είναι ομοιόμορφη.

Λύση: Στον παρακάτω πίνακα εμφανίζονται οι τιμές x_j, οι θεωρητικά προβλεπόμενες τιμές α_j που είναι ίσες με τη μέση πυκνότητα ανά φατνίο α, όπου α = 255/15 = 17, καθώς και οι ποσότητες $(x_j-\alpha_j)^2/\alpha_j$ (οι υπολογισμοί έγιναν με το Excel).
Από τη σχέση (1), υπολογίζεται ότι

$$x^2 = 5{,}64$$

Οι βαθμοί ελευθερίας v είναι:

$$v = T - 1 = 14$$

Για 14 βαθμούς ελευθερίας (β.ε), και για πιθανότητα 0,05, η κρίσιμη τιμή x_c^2 είναι:

$$x_c^2 = 23{,}68$$

Επομένως $x^2 < x_c^2$ και αυτό σημαίνει ότι ισχύει η υπόθεση περί ομοιόμορφης χωρικής κατανομής, σε επίπεδο σημαντικότητας 5%.

xj	aj=a	(xj-aj)^2/aj
20	17	0,5294118
19	17	0,2352941
18	17	0,0588235
17	17	0
13	17	0,9411765
21	17	0,9411765
15	17	0,2352941
19	17	0,2352941
16	17	0,0588235
17	17	0
15	17	0,2352941
19	17	0,2352941
15	17	0,2352941
19	17	0,2352941
12	17	1,4705882

3.2.1.β. Έλεγχος χ^2 τυχαίας χωρικής κατανομής

Προκειμένου να ελεγχθεί αν μια χωρική κατανομή ακολουθεί τυχαίο χωρικό πρότυπο, το οποίο περιγράφεται με μια κατανομή Poisson, διαιρείται η επιφάνεια σε *T* ορθογώνια φατνία, στο κάθε ένα από τα οποία μετράται το πλήθος των περιεχόμενων σημείων, που συμβολίζεται με *j*. Για κάθε τιμή *j*, μετράται το πλήθος x_j, των φατνίων που καθένα από αυτά περιέχει *j* το πλήθος σημεία.
Στη συνέχεια, για κάθε *j* υπολογίζεται το πλήθος των φατνίων α_j, με βάση το τυχαίο χωρικό πρότυπο. Τα α_j υπολογίζονται από τη σχέση:

$$a_j = Te^{-n/T}\frac{(n/T)^j}{j!} \qquad (3)$$

που είναι η κατανομή Poisson, πολλαπλασιασμένη επί το πλήθος των φατνίων. *n* είναι το πλήθος των σημείων.
Η βασική υπόθεση H_0 είναι: Η χωρική κατανομή ακολουθεί το τυχαίο πρότυπο.
Η εναλλακτική υπόθεση H_1 είναι: Η χωρική κατανομή δεν ακολουθεί το τυχαίο πρότυπο.

Ο έλεγχος της υπόθεσης πραγματοποιείται με βάση την κατανομή χ^2, μέσω του υπολογισμού της ποσότητας x^2, όπου:

$$x^2 = \sum_{j=0}^{\infty} \frac{(x_j - a_j)^2}{a_j} \qquad (4)$$

Εφαρμογή

Να εξεταστεί, σε επίπεδο σημαντικότητας 5%, η υπόθεση ότι η χωρική κατανομή της εφαρμογής της παραγράφου 2. 1. α ακολουθεί τυχαίο χωρικό πρότυπο.

Λύση: Για να ελεγχθεί η υπόθεση περί τυχαίου χωρικού προτύπου, η εξεταζόμενη χωρική κατανομή χωρίζεται σε T = 60 φατνία. Σε κάθε φατνίο μετρήθηκε το πλήθος j των σημείων που περιέχει αυτό. Στη συνέχεια, για κάθε j, μετρήθηκε το πλήθος x_j των φατνίων που περιέχουν j σημεία και υπολογίστηκαν οι αντίστοιχες τιμές α_j που αντιστοιχούν σε τυχαίο χωρικό πρότυπο, με βάση τη σχέση (3). Τα αποτελέσματα των μετρήσεων και υπολογισμών εμφανίζονται στον παρακάτω πίνακα:

j	xj	aj		(xj-aj)^2/aj
0	0	0,855854	4,493234	4,493234
1	0	3,63738		
2	1	7,729432		5,858807
3	10	10,95003		0,082425
4	29	11,63441		25,92001
5	13	9,889244		0,978518
6	7	7,004881		3,4E-06
7	0	4,252964		4,252964
8	0	2,259387	4,045807	4,045807
9-infinite	0	1,78642		

Οι μικρές τιμές a_j αθροίστηκαν ώστε να μην υπολείπονται κατά πολύ από το 5.
Το x^2 υπολογίστηκε με βάση τη σχέση (4) και είναι:

$x^2 = 45,63$

Οι β. ε. *v* είναι:

$v = 8 - 2 = 6$

Το x_c^2, για πιθανότητα 0,05 και 6 β. ε. είναι:

$x_c^2 = 12,59$

Επομένως $x^2 > x_c^2$ και αυτό σημαίνει ότι απορρίπτεται η υπόθεση περί τυχαίας χωρικής κατανομής, σε επίπεδο σημαντικότητας 5%.

3.2.1.γ. Έλεγχος t χωρικής κατανομής με βάση το δείκτη $s^2/\bar{x}$

Μια τυχαία χωρική κατανομή περιγράφεται από την κατανομή Poisson που, όπως είναι γνωστό, το πηλίκο διασποράς προς μέση τιμή είναι ίσο με τη μονάδα. Με βάση αυτό το δεδομένο, ένας τρόπος να εξεταστεί κατά πόσο μια χωρική κατανομή σημείων είναι τυχαία, είναι να υπολογιστεί το πηλίκο διασποράς s^2 προς τη μέση τιμή $\bar{x}$. Αν το πηλίκο $s^2/\bar{x}$ είναι κοντά στη μονάδα, τότε εκτιμάται ότι η εξεταζόμενη χωρική κατανομή είναι τυχαία. Αν το $s^2/\bar{x}$ είναι σημαντικά μικρότερο ή μεγαλύτερο από τη μονάδα, τότε η κατανομή μπορεί να είναι ομοιόμορφη ή ομαδοποιημένη, αντίστοιχα. Η ποσότητα που καθορίζει το κατά πόσον η εξεταζόμενη κατανομή είναι ή δεν είναι τυχαία, είναι η ποσότητα *t*, που υπολογίζεται από τη σχέση:

$$t = \frac{\frac{s^2}{\bar{x}} - 1}{\sigma} \qquad (5)$$

όπου το *σ* προσδιορίζεται από τη σχέση:

$$\sigma = \sqrt{\frac{2}{T-1}} \qquad (6)$$

T είναι το πλήθος των φατνίων.
Αυτό που εκφράζεται μέσω των σχέσεων (5) και (6), είναι ότι σε μια χωρική κατανομή που ακολουθεί το τυχαίο πρότυπο, ο λόγος διασποράς προς μέση τιμή μείον τη μονάδα, $s^2/\bar{x}$ - 1, ακολουθεί κατανομή t (Student), με τυπική απόκλιση σ που δίδεται από τη σχέση (6) και με T-1 β.ε.
Επομένως, στον έλεγχο χωρικής κατανομής με βάση το δείκτη $s^2/\bar{x}$, η βασική υπόθεση H_0 είναι: η χωρική κατανομή ακολουθεί το τυχαίο πρότυπο.
Η εναλλακτική υπόθεση H_1 είναι: η χωρική κατανομή δεν ακολουθεί το τυχαίο πρότυπο.
Η ποσότητα t που κρίνει την ισχύ της βασικής υπόθεσης, υπολογίζεται από τη σχέση (5).

Εφαρμογή

Να ελεγχθεί, με βάση το δείκτη $s^2/\bar{x}$, το κατά πόσον η χωρική κατανομή της παραγράφου 3. 2. 1. β., όπου τα σημεία περιέχονται σε T = 60 φατνία, ακολουθεί το τυχαίο χωρικό πρότυπο, σε επίπεδο σημαντικότητας 5%.

Λύση: Η μέση τιμή $\bar{x}$, με βάση τον πίνακα της παραγράφου 2. 1. β., είναι:

$$\bar{x} = (\Sigma j x_j)/T = 4{,}25$$

Η διασπορά s^2, με βάση τον ίδιο πίνακα, είναι:

$$s^2 = [\Sigma(j - \bar{x})^2 x_j]/T = 0{,}854$$

Επομένως

$$s^2/\bar{x} = 0{,}201$$

Επίσης, από τη σχέση (6) και με T = 60, έχουμε

σ = 0,184

Με βάση τη σχέση (5) υπολογίζουμε την ποσότητα *t*, που είναι:

t = -4,340

Οι β.ε. είναι *T*-1=59 και η πιθανότητα είναι 0,05. Το κρίσιμο σημείο t_c, γι'αυτές τις παραμέτρους και για δίπλευρο έλεγχο, είναι:

t_c = 2,001

Συγκρίνοντας τα *t* και t_c, είναι φανερό ότι το *t* βρίσκεται εκτός του διαστήματος [$-t_c$, t_c].
Αυτό σημαίνει ότι σε επίπεδο σημαντικότητας 5% εκτιμούμε πως η εξεταζόμενη χωρική κατανομή δεν ακολουθεί το τυχαίο πρότυπο.
Η αρνητική τιμή *t* φαίνεται να υποδηλώνει ομοιόμορφη χωρική κατανομή. Για να συναχθούν όμως οριστικά συμπεράσματα, θα πρέπει, σε δεδομένο επίπεδο σημαντικότητας, να γίνει μονόπλευρος έλεγχος της βασικής υπόθεσης H_0 (τυχαία κατανομή), με εναλλακτική υπόθεση H_1: ομοιόμορφη χωρική κατανομή. Η εξεταζόμενη ανισότητα είναι η: $-t_c \leq t$.
Αν το *t* είχε μια σημαντική θετική τιμή, αυτό θα ήταν ένδειξη για ομαδοποίηση της χωρικής κατανομής. Σε αυτήν την περίπτωση, ενδείκνυται ο μονόπλευρος έλεγχος της βασικής υπόθεσης H_0 (τυχαία χωρική κατανομή), με εναλλακτική υπόθεση H_1: ομαδοποιημένη χωρική κατανομή. Η εξεταζόμενη ανισότητα είναι η: $t \leq t_c$.

3.2.1.δ. Ένα ακανθώδες ζήτημα

Στην παράγραφο 3.2.1.α., όπου η βασική υπόθεση ήταν αυτή της ομοιόμορφης χωρικής κατανομής, και αυτό που καταγράφονταν ήταν ο αριθμός σημείων σε

κάθε φατνίο, η επιφάνεια με τα 255 σημεία χωρίστηκε σε 15 φατνία. Στις παραγράφους 3.2.1.β. και 3.2.1.γ, όπου η βασική υπόθεση ήταν αυτή της τυχαίας χωρικής κατανομής και αυτό που καταγράφονταν ήταν ο αριθμός των φατνίων που περιείχαν ένα δεδομένο πλήθος σημείων, η επιφάνεια χωρίστηκεσε 60 φατνία (4 φορές περισσότερα από όσο

προηγουμένως). Σύμφωνα με τους Swan & Sandilands 1995, οι αναλύσεις καννάβου με τη βασική υπόθεση του τυχαίου χωρικού προτύπου πραγματοποιούνται με αρκετά μεγαλύτερο αριθμό φατνίων σε σχέση με αυτόν που αντιστοιχεί στις αναλύσεις καννάβου με τη βασική υπόθεση της ομοιομορφίας.
Ωστόσο, η επιλογή του μεγέθους του φατνίου, και συνακόλουθα του πλήθους των φατνίων στα οποία χωρίζονται τα σημεία μιας χωρικής κατανομής, έχει μεγάλη σημασία, καθώς διαφορετικές διαστάσεις φατνίων μπορούν να οδηγήσουν σε διαφορετικές στατιστικές εκτιμήσεις. Είναι λοιπόν σημαντικό να έχει κανείς ένα αντικειμενικό κριτήριο προσδιορισμού του βέλτιστου μεγέθους φατνίου. Ένα τέτοιο κριτήριο, μπορεί να είναι αυτό της μεγιστοποίησης της ισχύος του ελέγχου χ^2, που σημαίνει ότι το μέγεθος και το πλήθος των φατνίων θα πρέπει να είναι τέτοια ώστε, για ένα δεδομένο επίπεδο σημαντικότητας, οι τιμές x^2 και x_c^2 θα πρέπει να διαφέρουν όσο το δυνατόν περισσότερο. Υπάρχουν επίσης και εμπειρικοί κανόνες, σύμφωνα με τους οποίους ένα αξιόπιστο εμβαδόν φατνίου θα πρέπει να κυμαίνεται μεταξύ A/n και $2A/n$, όπου A η συνολική επιφάνεια και n το πλήθος των σημείων της χωρικής κατανομής (Κουτσόπουλος 2002).

3.2.2. Χωρική ανάλυση με μεθόδους απόστασης

Η ανάλυση καννάβου επικεντρώνεται στην καταγραφή του αριθμού σημείων ανά φατνίο. Η ανάλυση με μεθόδους απόστασης, επικεντρώνεται στο πόσο απέχει, από κάθε σημείο, το πλησιέστερο σε αυτό (κριτήριο απόστασης γειτονικού σημείου). Σε μια ομαδοποιημένη χωρική κατανομή, η απόσταση από το γειτονικό σημείο είναι σχετικά μικρή, ενώ σε μια ομοιόμορφη χωρική κατανομή η απόσταση είναι μεγάλη. Σε μια τυχαία χωρική κατανομή, η απόσταση από το γειτονικό σημείο λαμβάνει ενδιάμεσες τιμές. Παρακάτω εξετάζονται δυο συνήθεις μέθοδοι ανάλυσης χωρικής κατανομής με το κριτήριο της απόστασης.

3.2.2.α. Έλεγχος εγγύτερου γείτονα

Στον έλεγχο εγγύτερου γείτονα, υπολογίζονται οι ποσότητες d και δ, που προσδιορίζονται από τις σχέσεις:

$$d = \frac{1}{n}\sum_{i=1}^{n} d_i \qquad (7)$$

$$\delta = \frac{1}{2}\sqrt{\frac{A}{n}} \qquad (8)$$

n είναι ο αριθμός των σημείων της χωρικής κατανομής. d_i είναι η απόσταση σημείου *i* από το πλησιέστερο γειτονικό σημείο. *A* είναι το εμβαδόν της επιφάνειας της χωρικής κατανομής.
Το *d* είναι η παρατηρούμενη μέση απόσταση από το πλησιέστερο γειτονικό σημείο και το *δ* είναι η αναμενόμενη μέση απόσταση από το πλησιέστερο γειτονικό σημείο.
Οι τιμές *d* ακολουθούν μια γκαουσιανή κατανομή με μέση τιμή *δ* και με τυπική απόκλιση *σ* που υπολογίζεται από τη σχέση (Κουτσόπουλος 2002):

$$\sigma = \frac{0{,}26136}{\sqrt{n^2 / A}} \qquad (9)$$

Οπότε η ποσότητα *z*, που ορίζεται ως:

$$z = \frac{d - \delta}{\sigma} \qquad (10)$$

ακολουθεί μια τυποποιημένη κανονική κατανομή. Με βάση το *z* μπορεί να γίνει ένας έλεγχος υπόθεσης με κατανομή z, όπου η βασική υπόθεση H_0 είναι: η χωρική κατανομή ακολουθεί το τυχαίο πρότυπο και η εναλλακτική υπόθεση H_1 είναι: η χωρική κατανομή δεν ακολουθεί το τυχαίο πρότυπο.

Εφαρμογή

Η επιφάνεια του πετρώματος της εφαρμογής της παραγράφου 2.1.α. είναι *A* = 10000m^2 και ο αριθμός των σημείων είναι *n* = 255. Η ανάλυση απόστασης από τον πλησιέστερο γείτονα έδωσε *d* = 6,253m. Να εξεταστεί, σε επίπεδο σημαντικότητας 5%, η υπόθεση ότι η χωρική κατανομή των απολιθωμάτων, που εκφράζονται ως σημεία στην ψηφιακή εικόνα, είναι τυχαία.

Λύση: Από τη σχέση (8) υπολογίζεται το δ, που είναι:

$\delta = 0,5\sqrt{(10000/255)} = 3,131$

Από τη σχέση (9) υπολογίζεται το σ, που είναι:

$\sigma = 0,26136 \times \sqrt{(10000/255^2)} = 0,1025$

Από τη σχέση (10) υπολογίζεται το z, που είναι:

$z = (6,253 - 3,131)/0,1025 = 30,46$

Για δίπλευρο έλεγχο σε επίπεδο σημαντικότητας 5%, το z_c είναι:

$z_c = 1,96$

Δηλαδή $z > z_c$, οπότε απορρίπτεται η υπόθεση περί τυχαίας χωρικής κατανομής, σε επίπεδο σημαντικότητας 5%.

3.2.2.β. Έλεγχος χωρικής κατανομής με το λόγο *d/δ*

Έχει διαπιστωθεί (Κουτσόπουλος 2002), ότι όταν ο λόγος *d/δ* είναι μικρότερος της μονάδας, η χωρική κατανομή είναι ομαδοποιημένη και όταν το *d/δ* είναι μεγαλύτερο της μονάδας η χωρική κατανομή είναι ομοιόμορφη, και μάλιστα όταν αυτός ο λόγος είναι μεγαλύτερος ή ίσος του 2, η χωρική κατανομή παρουσιάζει μια κανονικότητα. Όταν το *d/δ* είναι περίπου ίσο με τη μονάδα, τότε η χωρική κατανομή είναι τυχαία.
Ο λόγος *d/δ* μπορεί, επομένως, να αξιοποιηθεί στην αναγνώριση χωρικών προτύπων. Στην εφαρμογή της παραγράφου 3.2.2.α, για παράδειγμα, έχουμε *d/δ* = 6,253/3,131 = 1,997. Αυτό σημαίνει ότι η εξεταζόμενη χωρική κατανομή τείνει προς την κανονικότητα.

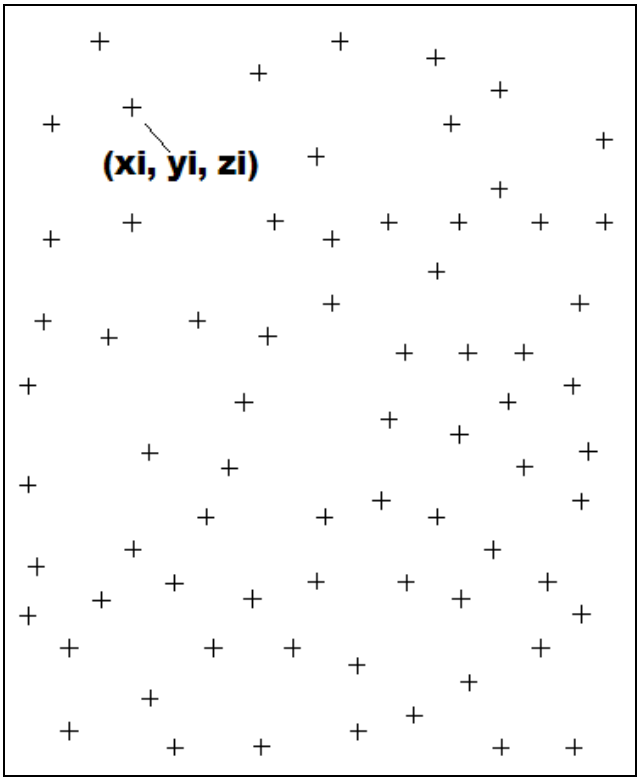
3.2.3. Έλεγχος ανισοτροπίας

Σε μια ευρύτερη περιοχή με ηφαιστειακή δραστηριότητα, τα ηφαίστεια μπορεί να μην κατανέμονται ομοιόμορφα αλλά κατά μήκος γραμμών που αντιστοιχούν σε τεκτονικές δομές. Ακόμα, σε μια περιοχή όπου γίνονται μετρήσεις σε διάφορες θέσεις, η δειγματοληψία μπορεί να μην είναι τυχαία, ούτε ομοιόμορφη, αλλά να ακολουθεί επιλεκτικά γραμμές

ορισμένης διεύθυνσης. Αν το μετρούμενο μέγεθος (π.χ. συγκεντρώσεις μετάλλων) συνδέεται με τον τεκτονισμό της περιοχής (κοίτασμα που αναπτύσσεται σε ρηγματική ζώνη), και αν οι θέσεις δειγματοληψίας είναι σε διευθύνσεις παράλληλες προς τον τεκτονισμό, τότε οι μετρήσεις μπορούν να περιέχουν συστηματικά σφάλματα.
Τίθεται επομένως το πρόβλημα της αναγνώρισης ανισότροπου χωρικού προτύπου, όπως αυτού που εμφανίζεται στο (σχ. 5). Για το σκοπό αυτό μπορεί να πραγματοποιηθεί ένας έλεγχος χ^2, με βασική υπόθεση H_0: κανένας επιλεκτικός προσανατολισμός των σημείων (ομοιόμορφο χωρικό πρότυπο) και H_1: επιλεκτικός προσανατολισμός των σημείων (μη ομοιόμορφο χωρικό πρότυπο). Το μετρούμενο μέγεθος είναι το αζιμούθιο του ευθύγραμμου τμήματος που συνδέει δυο σημεία. Από το σύνολο των τιμών αζιμουθίου, και από την τιμή του μέσου αζιμουθίου, μπορεί να υπολογιστεί η ποσότητα x^2 και να συγκριθεί με την τιμή x_c^2, για δεδομένο επίπεδο εμπιστοσύνης.

3.3. Μέθοδοι χωρικής παρεμβολής

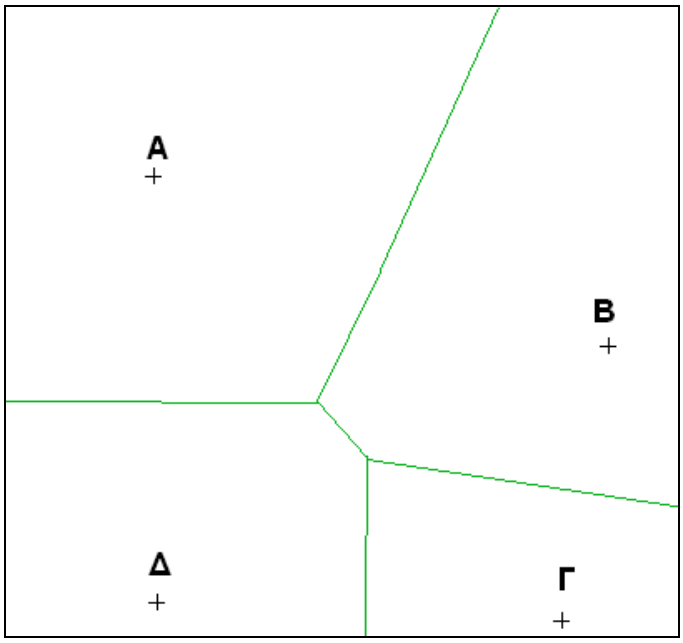
Οι μετρήσεις ενός φυσικού μεγέθους στο ύπαιθρο (π.χ. μετρήσεις θερμοκρασίας, συγκέντρωσης μετάλλων, μαγνητικού πεδίου και άλλες) γίνονται σε διακριτές θέσεις με συντεταγμένες (x_i, y_i). Στην κάθε θέση (x_i, y_i), αντιστοιχεί μια τιμή z_i, του μετρούμενου φυσικού μεγέθους z. Στο (σχ. 6) εμφανίζονται οι θέσεις γεωτρήσεων για τον προσδιορισμό του πάχους z επιφανειακού στρώματος αλλουβιακών αποθέσεων (τα δεδομένα ελήφθησαν από το εκπαιδευτικό υλικό του λογισμικού ILWIS 3. 3). Το πρόβλημα που ανακύπτει είναι πώς μπορούν να υπολογιστούν οι τιμές του z στις άλλες θέσεις της περιοχής έρευνας, με βάση τις τιμές z_i των *σημείων ελέγχου* (x_i, y_i). Για το σκοπό αυτό έχουν αναπτυχθεί διάφορες *μέθοδοι χωρικής παρεμβολής*, που αναπτύσσονται παρακάτω.



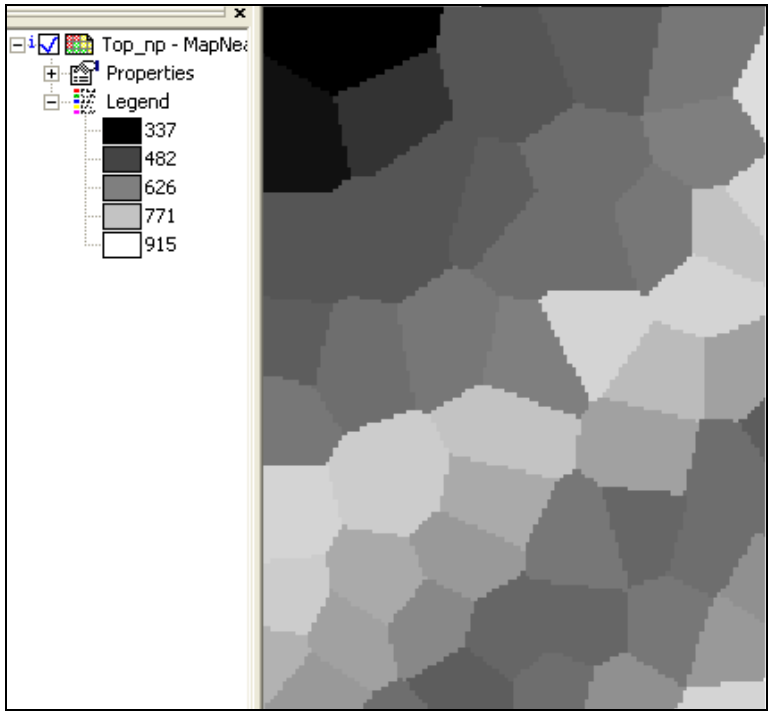
Σχ. 6. Σημεία λήψης μετρήσεων (σημεία ελέγχου) με γεωγραφικές συντεταγμένες (x_i, y_i) και με αντίστοιχη τιμή μετρούμενου μεγέθους z_i.

3.3. 1. Χωρική παρεμβολή με πολύγωνα Thiessen

Σε αυτή τη μέθοδο χωρικής παρεμβολής, η επιφάνεια τιμών *z* θεωρείται ασυνεχής και αποτελούμενη από πολυγωνικές επιφάνειες σταθερής τιμής z_i, που εκτείνονται γύρω από τα σημεία (x_i, y_i). Η διαδικασία σχηματισμού των πολυγωνικών επιφανειών παρουσιάεται στο (σχ. 7). Οι μεσοκάθετοι των γειτονικών σημείων ελέγχου (x_i, y_i) και (x_j, y_j) ορίζουν τα πολύγωνα που τα περιβάλλουν. Σε κάθε πολύγωνο, η τιμή του μετρούμενου μεγέθους είναι σταθερή και ίση με την τιμή του περιεχόμενου σημείου ελέγχου. Σε όλα τα σημεία του πολυγώνου Α, για παράδειγμα, η τιμή του μεγέθους *z* είναι ίση με την τιμή του περιεχόμενου σημείου ελέγχου.
Αν εφαρμοστεί αυτή η μέθοδος χωρικής παρεμβολής στα δεδομένα των μετρήσεων *z* στην περιοχή που αναπαριστάνεται στο (σχ. 6), προκύπτει ο ψηφιδωτός χάρτης του (σχ. 8).



Σχ. 7. Προσδιορισμός πολυγώνων Thiessen.



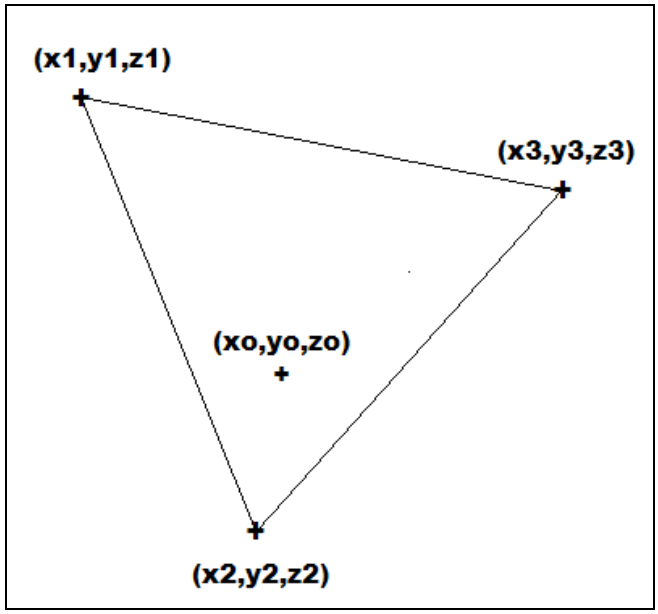
Σχ. 8. Χωρική παρεμβολή με πολύγωνα Thiessen

Στο κάθε πολύγωνο του (σχ. 8), η τιμή του μεγέθους *z* είναι σταθερή. Η μέθοδος χωρικής παρεμβολής με πολύγωνα Thiessen, που στη βιβλιογραφία αναφέρεται και ως *παρεμβολή με πολύγωνα Voronoi*, ως *ψηφιδοποίηση Dirichlet*, ή ακόμα ως μέθοδος του *εγγύτερου γείτονα* (nearest neighbor),

εφαρμόζεται όταν το καταγραφόμενο μέγεθος *z* δεν είναι μια συνεχής και ομαλά μεταβαλλόμενη ποσότητα, αλλά μια ταξινομητική κατηγορία (π.χ. τύπος φυτοκάλυψης, βαθμός διαβρωσιμότητας εδάφους, εισοδηματική κλίμακα και άλλα ασυνεχή μεγέθη που υποδηλώνουν ταξινόμηση του παρατηρησιακού υλικού).
Στις μεθόδους χωρικής παρεμβολής που παρουσιάζονται παρακάτω, το μετρούμενο φυσικό μέγεθος είναι μια ποσότητα συνεχής και ομαλά μεταβαλλόμενη (π.χ. υψόμετρο, θερμοκρασία, συγκέντρωση ρύπων και άλλα μεγέθη).

3.3.2. Χωρική παρεμβολή με τριγωνισμό

Στη μέθοδο χωρικής παρεμβολής με τριγωνισμό, τα σημεία ελέγχου συνδέονται μεταξύ τους με ευθύγραμμα τμήματα, ώστε να σχηματιστεί ένα *τριγωνικό δίκτυο* (Triangular Irregular Network, TIN). Ο συνηθέστερος αλγόριθμος σχηματισμού τριγωνικού δίκτύου είναι ο *τριγωνισμός κατά Delauney*, μέσω του οποίου επιδιώκεται να σχηματιστούν οξυγώνια τρίγωνα, που τείνουν να είναι ισόπλευρα (Isaaks & Srivastava 1989). Η τιμή z_0, σε μια θέση με συντεταγμένες (x_0, y_0), που βρίσκεται μέσα σε τρίγωνο με συντεταγμένες (x_1, y_1), (x_2, y_2), (x_3, y_3) (βλ. σχ. 9), προσδιορίζεται από την εξίσωση του επιπέδου που ορίζεται από τις τρεις κορυφές του τριγώνου.



Σχ. 9. Χωρική παρεμβολή με τη μέθοδο του τριγωνισμού

Η εξίσωση του επιπέδου είναι:

$$z = ax + by + c \qquad (11)$$

a, b, c, είναι συντελεστές που υπολογίζονται από την επίλυση του γραμμικού συστήματος:

$$\begin{aligned} ax_1 + by_1 + c &= z_1 \\ ax_2 + by_2 + c &= z_2 \\ ax_3 + by_3 + c &= z_3 \end{aligned} \qquad (12)$$

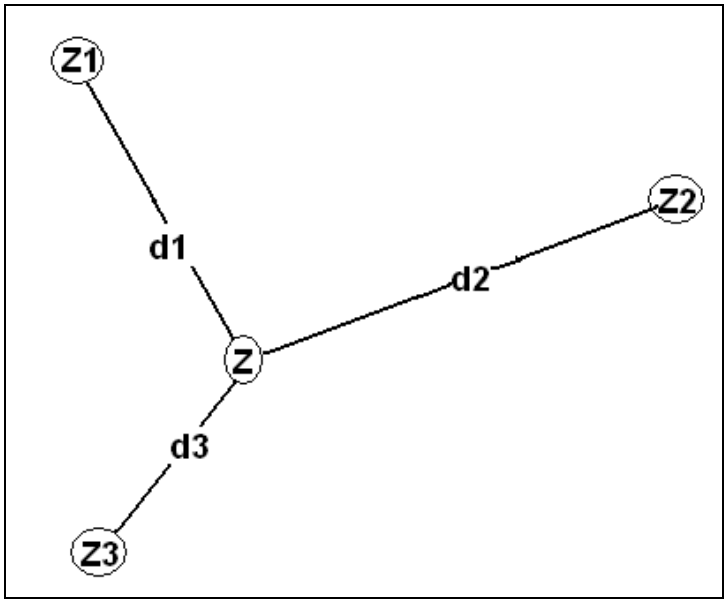
Έχοντας υπολογίσει τις παραμέτρους a, b, c, είναι δυνατό να υπολογιστεί η τιμή z_0, με γεωγραφικές συντεταγμένες (x_0, y_0), από τη σχέση:

$$z_0 = ax_0 + by_0 + c \qquad (13)$$

Η ίδια διαδικασία επαναλαμβάνεται σε όλα τα σημεία του γεωγραφικού χώρου, σε όλα τα τρίγωνα.
Η μέθοδος του τριγωνισμού χρησιμοποιείται, κυρίως, στην κατασκευή τοπογραφικών χαρτών. Γνωρίζοντας το ανάγλυφο της περιοχής έρευνας, είναι δυνατό να καταγραφούν με ένα GPS τα υψόμετρα διαφόρων σημείων, που να καλύπτουν ικανοποιητικά τις δομές που εμφανίζονται στο ανάγλυφο (κορυφογραμμές, χαράδρες, επίπεδες περιοχές). Έχοντας ένα λεπτομερές δίκτυο τιμών υψομέτρου, μπορούν να γίνουν αξιόπιστοι υπολογισμοί των υψομέτρων όλων των σημείων της περιοχής, με βάση τις σχέσεις (12) και (13).
Σε μετρήσεις άλλων μεγεθών, όπως συγκέντρωση μετάλλων ή θερμοκρασία, δεδομένου ότι δεν είναι εκ των προτέρων γνωστό πού είναι οι υψηλές τιμές, πού είναι οι χαμηλές τιμές και πού είναι μεγάλη η χωρική μεταβολή του μετρούμενου μεγέθους, χρησιμοποιούνται άλλες μέθοδοι χωρικής παρεμβολής, που παρουσιάζονται στις επόμενες παραγράφους.

3.3.3. Χωρική παρεμβολή με κινούμενους μέσους όρους

Στη χωρική παρεμβολή με κινούμενους μέσους όρους (moving averages), η τιμή του μεγέθους z σε σημείο με συντεταγμένες (x, y) προσδιορίζεται από ένα γραμμικό συνδυασμό των γνωστών τιμών z_i γειτονικών σημείων ελέγχου. Οι συντελεστές των τιμών z_i στο γραμμικό μετασχηματισμό, που ονομάζονται *συντελεστές βαρύτητας* (weighting coefficients) είναι, συνήθως, αντιστρόφως ανάλογοι της απόστασης των σημείων ελέγχου από το σημείο (x, y), έτσι ώστε η τιμή z να διαμορφώνεται, κυρίως, από τα εγγύτερα σημεία ελέγχου.



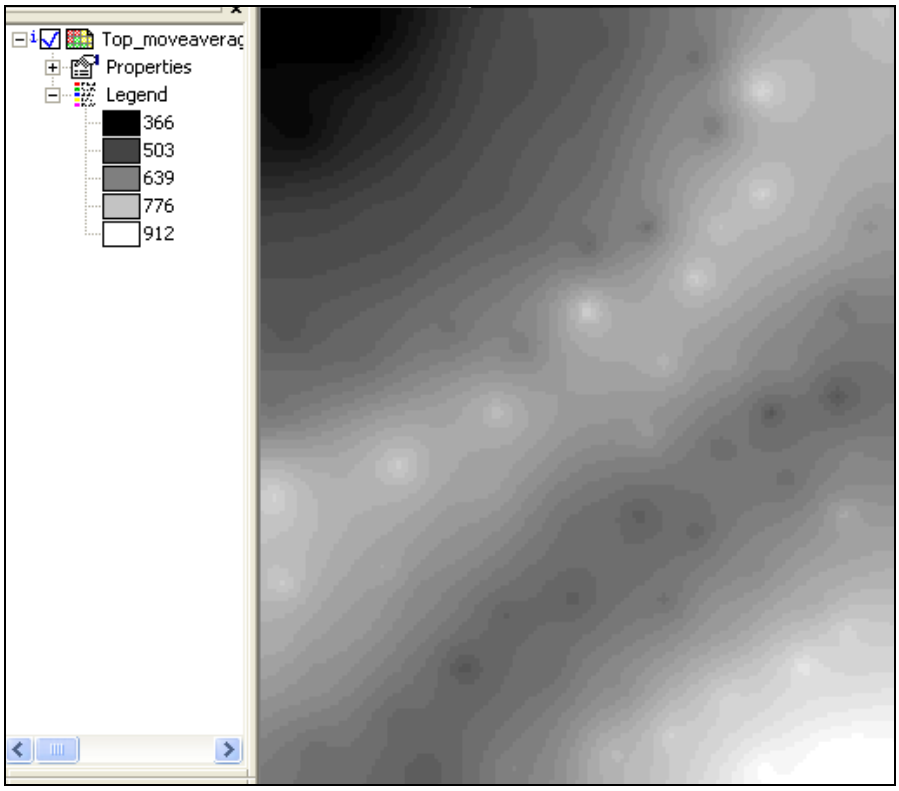
Σχ. 10. Χωρική παρεμβολή με κινούμενο μέσο όρο.

Τα παραπάνω εκτεθέντα μπορούν να γίνουν περισσότερο κατανοητά, παρατηρώντας το (σχ. 10). Το σημείο στο οποίο επιχειρείται να υπολογιστεί η τιμή του φυσικού μεγέθους z, απέχει αποστάσεις d_1, d_2 και d_3, από τα σημεία ελέγχου με γνωστές τιμές z_1, z_2 και z_3, αντίστοιχα. Με βάση τη μέθοδο του κινούμενου μέσου όρου, η τιμή του z είναι:

$$z = \frac{z_1 / d_1 + z_2 / d_2 + z_3 / d_3}{1/d_1 + 1/d_2 + 1/d_3} \qquad (14)$$

Ο παρονομαστής στο δεξιό μέλος της σχέσης (14) εξυπηρετεί στην κανονικοποίηση των συντελεστών βαρύτητας (άθροισμα αυτών ίσο με τη μονάδα).
Γενικότερα, αν το σημείο (x, y) περιβάλλεται από n το πλήθος γειτονικά σημεία ελέγχου με τιμές z_i και αποστάσεις d_i, η τιμή z στη θέση (x, y) υπολογίζεται από τη σχέση:

$$z = \frac{\sum_{i=1}^{n} z_i / d_i}{\sum_{i=1}^{n} 1 / d_i} \qquad (15)$$



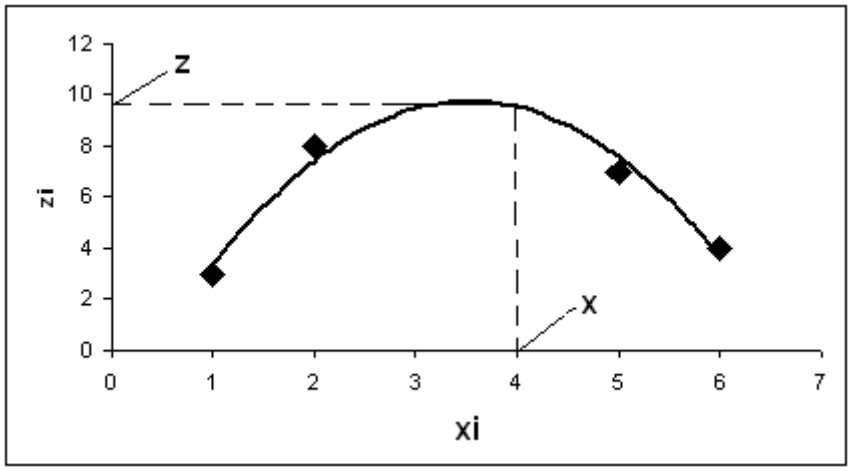
Σχ. 11. Χάρτης τιμών *z* που παράχθηκε από χωρική παρεμβολή με κινούμενο μέσο όρο

Στο (σχ. 11) εμφανίζεται ο χάρτης τιμών πάχους επιφανειακού στρώματος, που παράχθηκε από χωρική παρεμβολή, με κινούμενο μέσο όρο, στα δεδομένα του χάρτη του (σχ. 7) (παράγραφος 3. 1). Παρατηρούμε διάφορες φωτεινές κηλίδες, που εκφράζουν υψηλές τιμές *z*, να είναι διατεταγμένες διαγωνίως, από την κάτω αριστερή προς την άνω δεξιά πλευρά του χάρτη. Οι κηλίδες αυτές υποδηλώνουν μια επιμήκη περιοχή με μέγιστες τιμές πάχους *z*, που εκτιμάται να έχει μια πιο ομαλή συμπεριφορά, αλλά που παρουσιάζει τοπικές κορυφές σε θέσεις που υπάρχουν σημεία ελέγχου. Είναι χαρακτηριστικό της

μεθόδου του κινούμενου μέσου όρου να παράγει χάρτες με τοπικές κορυφές σε σημεία ελέγχου, χωρίς αυτές να ανταποκρίνονται στη φυσική πραγματικότητα.
Σε πολλά προγράμματα χωρικής παρεμβολής, ο αριθμός των γειτονικών σημείων ελέγχου που διαμορφώνουν την τιμή *z* σε μια θέση (*x*, *y*), διαμορφώνεται από την ακτίνα επιρροής με κέντρο το σημείο (*x*, *y*), το μήκος της οποίας επιλέγει ο χρήστης. Όσο μεγαλύτερη είναι η ακτίνα, τόσο αυξάνεται το πλήθος των σημείων ελέγχου που διαμορφώνουν την τιμή *z*.

3.3.4.α. Χωρική παρεμβολή με τοπικές επιφάνειες τάσης

Στη *χωρική παρεμβολή με τοπικές επιφάνειες τάσης* (local trends), προσδιορίζεται η επιφάνεια ελαχίστων τετραγώνων που προσεγγίζει τα γειτονικά σημεία ελέγχου γύρω από τη θέση (*x*, *y*). Η επιφάνεια αυτή είναι συνήθως επίπεδη ($1^{ου}$ βαθμού) ή $2^{ου}$ βαθμού. Έχοντας προσδιορίσει τους συντελεστές της επιφάνειας ελαχίστων τετραγώνων, μπορεί να υπολογιστεί η τιμή *z* στο (*x*, *y*), από την εξίσωση της επιφάνειας.



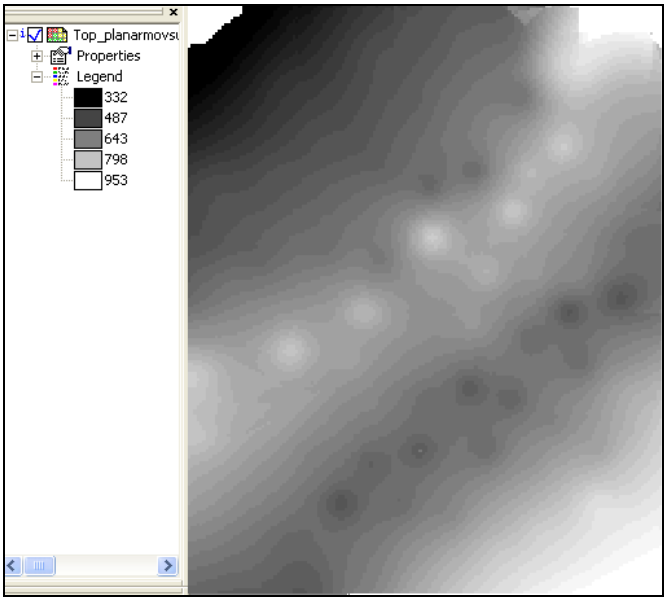
Σχ. 12. Τοπική καμπύλη τάσης και χωρική παρεμβολή στη θέση *x*.

Στο (σχ. 11), παρουσιάζεται το πώς λειτουργεί, στις δυο διαστάσεις, η μέθοδος των τοπικών τάσεων. Εδώ, έχουμε τέσσερα συνευθειακά σημεία ελέγχου x_i, με γνωστές τιμές z_i. Με τη μέθοδο των ελαχίστων

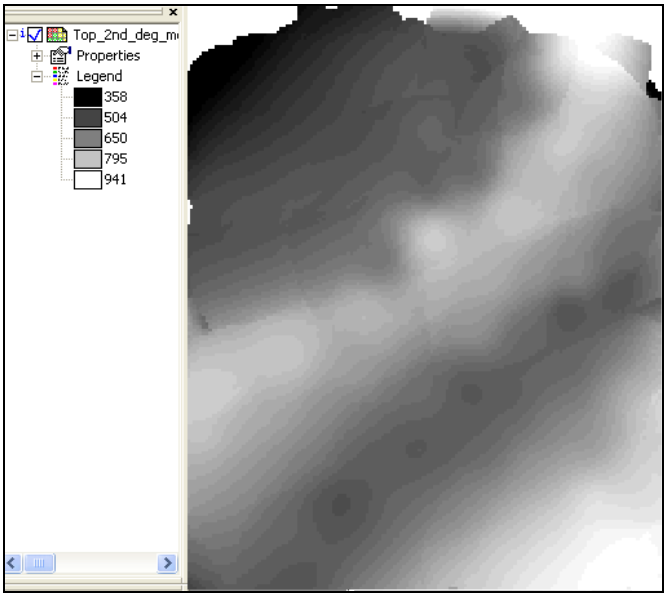
τετραγώνων, προσδιορίζεται η καμπύλη $2^{ου}$ βαθμού, που προσεγγίζει τις τέσσερις τιμές (x_i, z_i), και στη συνέχεια υπολογίζεται η τιμή *z* για τη θέση *x*. Η ίδια λογική ισχύει και για το γεωγραφικό χώρο, όπου εδώ τα σημεία ελέγχου δεν είναι συνευθειακά αλλά ομοεπίπεδα και αντί για καμπύλες ελάχιστων τετραγώνων έχουμε επιφάνειες ελάχιστων τετραγώνων.

Στα (σχ. 13) και (σχ. 14) παρουσιάζονται οι χάρτες που προέκυψαν από χωρική παρεμβολή των δεδομένων του χάρτη του (σχ. 6), με τη μέθοδο των τοπικών τάσεων επίπεδων επιφανειών και δευτεροβάθμιων επιφανειών, αντίστοιχα. Παρατηρούμε ότι στο χάρτη που προέκυψε από τις επίπεδες επιφάνειες, εμφανίζονται κηλίδες υψηλής φωτεινότητας σε διαγώνια διάταξη, όπως και στο χάρτη του (σχ. 11) (χωρική παρεμβολή με κινούμενο μέσο όρο), ενώ στο χάρτη που παράχθηκε από δευτεροβάθμιες επιφάνειες τάσης δεν εκδηλώνεται αυτό το φαινόμενο. Ωστόσο και στους δυο χάρτες χωρικής παρεμβολής με τοπικές επιφάνειες τάσης, εκδηλώνονται πλευρικά φαινόμενα (edge effects), κυρίως στην πάνω πλευρά, όπου δεν κατέστη δυνατό να γίνουν αξιόπιστοι υπολογισμοί του πάχους *z* του επιφανειακού στρώματος.

Για χωρική παρεμβολή, αντί για πρωτοβάθμιες ή δευτεροβάθμιες επιφάνειες ελαχίστων τετραγώνων, είναι δυνατό να προσδιοριστούν *κυβικές splines*, που είναι τριτοβάθμιες επιφάνειες με ομαλή συμπεριφορά και με την ιδιότητα της πολλαπλής διαφορισιμότητας στα σημεία ελέγχου (περισσότερες πληροφορίες για τις συναρτήσεις spline και τις ιδιότητές τους μπορεί κανείς να βρει στον Kreyszig 1979, ή σε άλλα συγγράμματα εφαρμοσμένων μαθηματικών).



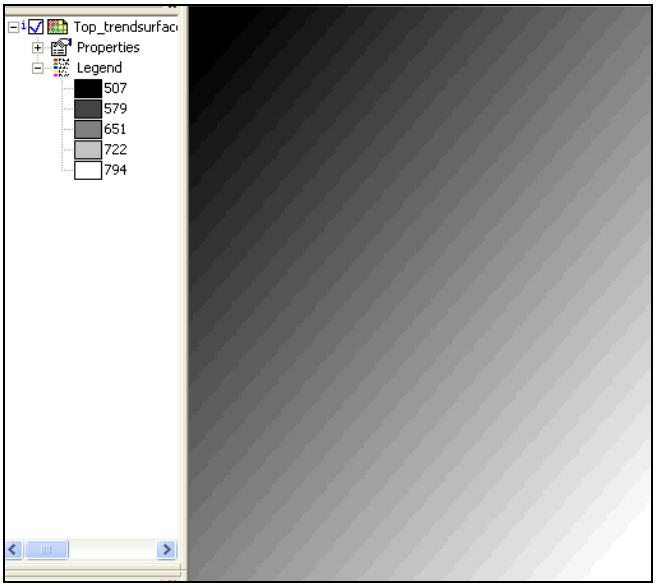
Σχ. 13. Χωρική παρεμβολή με τοπικές επίπεδες επιφάνειες



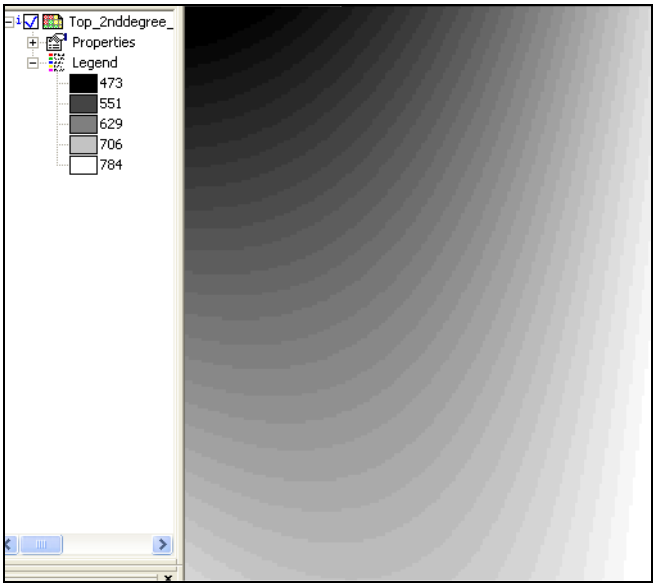
Σχ. 14. Χωρική παρεμβολή με τοπικές δευτεροβάθμιες επιφάνειες

3.3.4.β. Χωρική παρεμβολή με εκτίμηση συνολικής επιφάνειας τάσης

Όταν η επιφάνεια τάσης δεν προσδιορίζεται μόνο από ορισμένα σημεία ελέγχου που βρίσκονται σε κύκλο πεπερασμένης ακτίνας γύρω από κάποια θέση, αλλά από το σύνολο των σημείων ελέγχου της περιοχής έρευνας, τότε μιλάμε για χωρική παρεμβολή με τη *συνολική επιφάνεια τάσης*. Προσδιορίζοντας την επιφάνεια τάσης όλων των σημείων ελέγχου, υπολογίζονται, στη συνέχεια, οι τιμές του μεγέθους *z* σε κάθε θέση (*x*, *y*) της περιοχής έρευνας.



Σχ. 15. Προσέγγιση τιμών σημείων ελέγχου με επίπεδη επιφάνεια ελαχίστων τετραγώνων



Σχ. 16. Προσέγγιση τιμών σημείων ελέγχου με δευτεροβάθμια επιφάνεια ελαχίστων τετραγώνων

Η συνολική επιφάνεια τάσης είναι, συνήθως, επίπεδη ή δευτεροβάθμια, και σπανιότερα τριτοβάθμια. Όσο μεγαλύτερος είναι ο βαθμός της επιφάνειας προσέγγισης, τόσο πλησιάζουν οι τιμές *z'* της επιφάνειας προσέγγισης τις τιμές *z* των σημείων ελέγχου. Όμως οι επιφάνειες μεγάλου βαθμού παρουσιάζουν

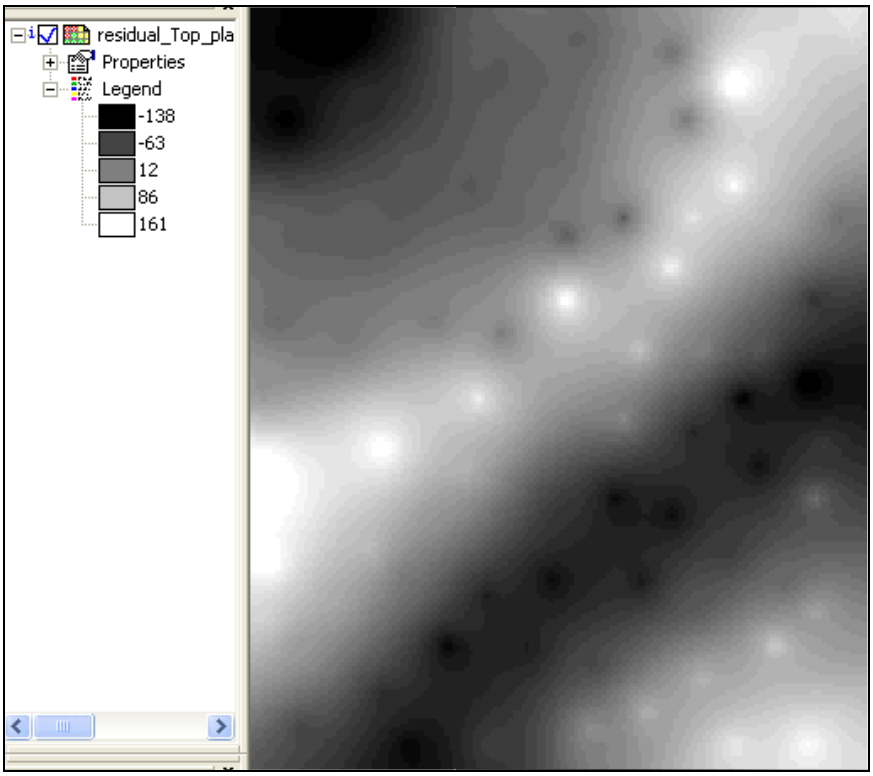
τοπικά ελάχιστα ή μέγιστα που δεν είναι καθόλου βέβαιο ότι ανταποκρίνονται στη φυσική πραγματικότητα.
Στα (σχ. 15) και (σχ. 16), εμφανίζονται η επίπεδη επιφάνεια ελαχίστων τετραγώνων και η δευτεροβάθμια επιφάνεια ελαχίστων τετραγώνων, αντίστοιχα, των δεδομένων του χάρτη του (σχ. 6) της παραγράφου 3.3. 1. Η εξίσωση της επίπεδης επιφάνειας ελαχίστων τετραγώνων, όπως προσδιορίστηκε από το λογισμικό ILWIS 3.3, είναι:

$$z = 656{,}08 + 0{,}1045x - 0{,}0893y \qquad (16)$$

Η εξίσωση της δευτεροβάθμιας επιφάνειας ελαχίστων τετραγώνων, που προσδιορίστηκε από το ίδιο λογισμικό, είναι:

$$z = 642{,}34 + 0{,}1107x - 0{,}0866y + 9{,}193.e^{-5}\,x^2 + 7{,}792.e^{-5}xy - 9{,}654.e^{-6}y^2 \qquad (17)$$

Παρατηρώντας τα (σχ. 15) και (σχ. 16) διαπιστώνουμε ότι οι ζώνες ίδιου διαστήματος τιμών *z* της δευτεροβάθμιας επιφάνειας είναι καμπυλωμένες, ενώ οι αντίστοιχες της επίπεδης επιφάνειας είναι ευθύγραμμες. Η δευτεροβάθμια επιφάνεια παρουσιάζει μεγαλύτερη κλίση από πάνω αριστερά προς κάτω δεξιά, από όσο από πάνω προς τα κάτω.



Σχ. 17. Υπολοιπόμενη επιφάνεια που προκύπτει από την αφαίρεση της επίπεδης επιφάνειας ελαχίστων τετραγώνων (σχ. 15) από την επιφάνεια της χωρικής παρεμβολής με κινούμενο μέσο όρο (σχ. 11).

Η επιλογή του βαθμού της πολυωνυμικής επιφάνειας προσέγγισης, μπορεί να γίνει με τη βοήθεια της *ανάλυσης διασποράς* (ANOVA, Swan & Sandilands 1995). Γενικά, όταν οι τιμές *z* μεταβάλλονται από το κέντρο προς την περιφέρεια, ενδείκνυται μια δευτεροβάθμια επιφάνεια

ελαχίστων τετραγώνων. Όταν το *z* τείνει να μεταβάλλεται με σταθερό ρυθμό προς μια κατεύθυνση, τότε είναι καλύτερο να προσεγγιστούν οι τιμές των σημείων ελέγχου με μια επίπεδη επιφάνεια ελαχίστων τετραγώνων.

Στην επεξεργασία χωρικών δεδομένων, η γενική επιφάνεια τάσης δεν αξιοποιείται τόσο για τον υπολογισμό των τιμών του μεγέθους *z* στις διάφορες θέσεις εκτός των σημείων ελέγχου, όσο για τη δημιουργία χαρτών *υπολοιπόμενων ανωμαλιών (residuals)*. Η υπολοιπόμενη ανωμαλία είναι η ανωμαλία περιορισμένης χωρικής κλίμακας, που μπορεί να προσδιοριστεί αν από το χάρτη τιμών *z* αφαιρεθεί η ανωμαλία ευρείας κλίμακας, που μπορεί να περιγραφεί από την εξίσωση της γενικής επιφάνειας τάσης.

Στο (σχ. 17), εμφανίζεται ο χάρτης των υπολοιπόμενων ανωμαλιών z_{res} που προκύπτουν αφαιρώντας τη γενική επίπεδη επιφάνεια τάσης της σχέσης (16), από τις τιμές *z* της χωρικής παρεμβολής με κινούμενο μέσο όρο, οι οποίες υπολογίστηκαν με βάση τη σχέση (15) και εμφανίζονται στο χάρτη του (σχ. 11).

Επίσης, η επιφάνεια γενικής τάσης αξιοποιείται και στη χωρική παρεμβολή με Kriging, όπως θα δούμε παρακάτω.

3.3.5. Εκτίμηση χωρικής συσχέτισης τιμών φυσικού μεγέθους

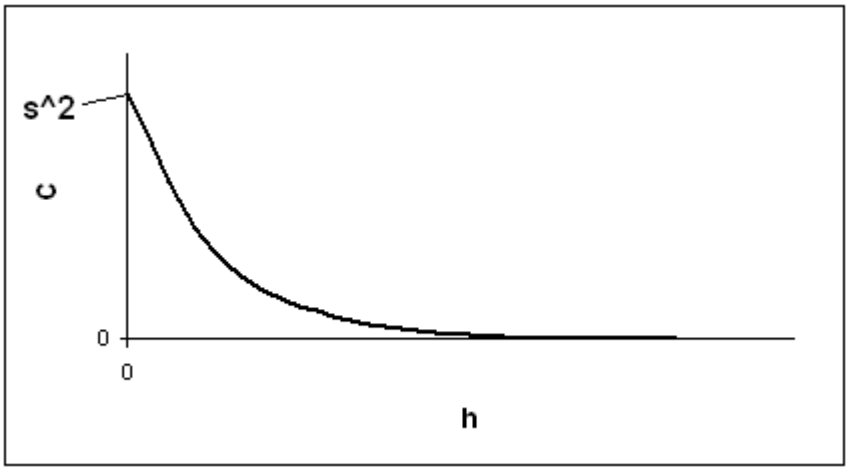
Στις μεθόδους χωρικής παρεμβολής με κινούμενους μέσους όρους και με τοπικές επιφάνειες τάσης, σημαντικό ρόλο παίζει η ακτίνα επιρροής των γειτονικών σημείων στον υπολογισμό της τιμής $z(x, y)$, δηλαδή μέχρι ποια απόσταση από τη θέση (x, y) θα πρέπει να ληφθούν υπόψη γειτονικές τιμές $z_i(x_i, y_i)$ για χωρική παρεμβολή στη θέση (x, y). Ένα ποσοτικό κριτήριο για τον προσδιορισμό της ακτίνας επιρροής, είναι αυτό της θέσης ουσιαστικού μηδενισμού της καμπύλης *συνδιασποράς* (*covariance*) των τιμών $z_i(x_i, y_i)$.

Ως συνδιασπορά $C(h)$ τιμών $z_i(x_i, y_i)$ σε απόσταση h, ορίζεται η μέση τιμή της ποσότητας $[(z_i - \mu(z_i)] \times [(z_j - \mu(z_j)]$, όπου z_i, z_j τιμές μετρούμενου μεγέθους σε θέσεις που απέχουν απόσταση h και $\mu(z_i)$, $\mu(z_j)$ οι τοπικές μέσες τιμές των z_i και z_j, αντίστοιχα (Κουτσόπουλος 2002). Ο αριθμητικός υπολογισμός του $C(h)$, για ένα σύνολο τιμών *z* σημείων ελέγχου σε μια περιοχή έρευνας, μπορεί να πραγματοποιηθεί με βάση τη σχέση (Isaaks & Srivastava 1989):

$$C(h) = \frac{1}{N(h)} \cdot \sum_{(i,j)|h} z_i z_j - \left(\frac{1}{n} \cdot \sum_{k=1}^{n} z_k \right)^2 \qquad (18)$$

N(*h*) είναι το πλήθος των ζευγών σημείων ελέγχου που απέχουν μεταξύ τους απόσταση *h*. *N* είναι το πλήθος των σημείων ελέγχου της περιοχής έρευνας. Η παραπάνω σχέση ισχύει αν η χωρική κατανομή των τιμών *z* είναι *στάσιμη* και *ισοτροπική*. Ο όρος *στάσιμη* σημαίνει ότι οι στατιστικές παράμετροι της κατανομής είναι ανεξάρτητες από την απόλυτη θέση στην περιοχή έρευνας. Οι

διαφορές στατιστικών παραμέτρων εξαρτώνται μόνο από το διάνυσμα της σχετικής θέσης των σημείων, πάνω στα οποία μετρούνται. Ο όρος *ισοτροπική* υποδηλώνει ότι οι διαφορές στατιστικών παραμέτρων εξαρτώνται μόνο από την απόσταση μεταξύ των σημείων και όχι από το αζιμούθιο της σχετικής θέσης αυτών.
Αν το *z* μεταβάλλεται ομαλά ως προς τη θέση, είναι εύλογο να εκτιμούμε ότι για μικρές τιμές *h* τα z_i, z_j δεν διαφέρουν σημαντικά, ενώ για μεγάλα *h* η διαφορά είναι σημαντική. Κατά συνέπεια, με βάση τη σχέση (18), το *C* παρουσιάζει μια πτωτική τάση ως προς *h*, όπως φαίνεται στο (σχ. 18). Το *C*(0) είναι ίσο με την τυπική απόκλιση s^2 των τιμών $z_i(x_i, y_i)$, αν το σφάλμα μέτρησης του μεγέθους *z* είναι μηδενικό. Αν το σφάλμα δεν είναι μηδενικό, τότε το *C*(0) διαφέρει από το s^2 .
Από το *διάγραμμα συνδιασποράς* (*συνβαριόγραμμα*) του (σχ. 18), μπορεί να εκτιμηθεί η απόσταση h_0, στον οριζόντιο άξονα *h*, για την οποία η τιμή *C* πρακτικά μηδενίζεται. Αυτή η τιμή h_0 είναι ίση με την ακτίνα επιρροής των γειτονικών σημείων, για χωρική παρεμβολή με κινούμενους μέσους όρους ή τοπικές επιφάνειες τάσης.



Σχ. 18. Διάγραμμα συνδιασποράς

Το πηλίκο $C(h)/C(0)$ είναι η *χωρική συσχέτιση* (spatial correlation) $\rho(h)$, που λαμβάνει τιμές από 0 ως 1 και, από ποιοτική άποψη, έχει τη συμπεριφορά της καμπύλης του (σχ. 18).
Άμεση σχέση με τη συνάρτηση $C(h)$ έχει η συνάρτηση $\gamma(h)$, που ονομάζεται *βαριόγραμμα* ή *ημιβαριόγραμμα* (*variogram, semivariogram*). Η συνάρτηση $\gamma(h)$ ορίζεται ως το ήμισυ της διασποράς της ποσότητας ($z_i - z_j$), όπου z_i και z_j τιμές z σημείων ελέγχου που απέχουν μεταξύ τους απόσταση h (Κουτσόπουλος 2002). Το $\gamma(h)$ μπορεί να υπολογιστεί από τη σχέση (Isaaks & Srivastava 1989):

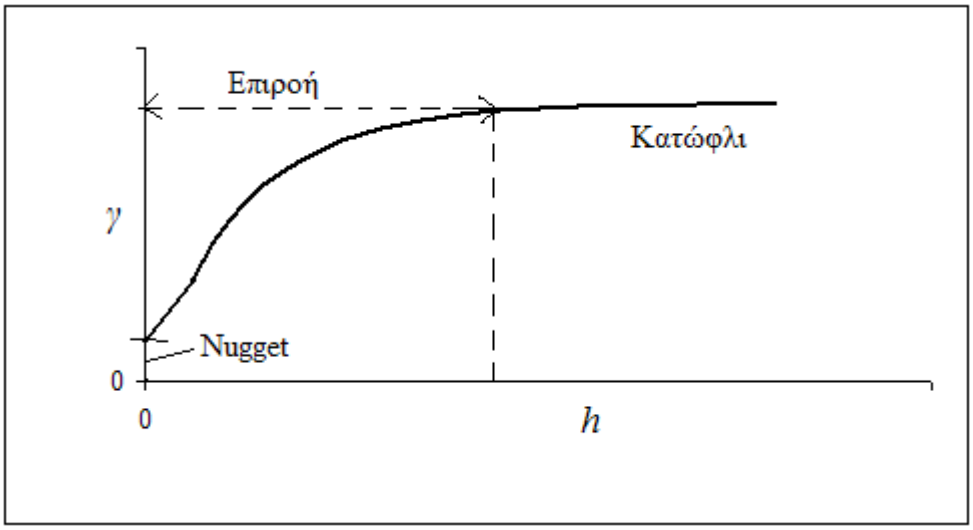
$$\gamma(h) = \frac{1}{2N(h)} \cdot \sum_{(i,j)|h} (z_i - z_j)^2 \qquad (19)$$

Η σημασία των συμβόλων N και h είναι η ίδια όπως και στη σχέση (18).
Οι συναρτήσεις $\gamma(h)$ και $C(h)$ συνδέονται μεταξύ τους με τη σχέση (Κουτσόπουλος 2002):

$$\gamma(h) = C(0) - C(h) \qquad (20)$$

Στο (σχ. 19) εμφανίζεται η γραφική παράσταση του βαριογράμματος γ ως προς h. Παρατηρούμε ότι η καμπύλη $\gamma(h)$ τείνει προς μια σταθερή

τιμή *κατωφλίου* (*sill*), που αποκαθίσταται από μια τιμή του *h* ίση με h_0 και μετά. Το h_0 είναι επίσης η τιμή του *h* για την οποία η συνδιασπορά *C* πρακτικά μηδενίζεται. Η ζώνη τιμών *h* από 0 ως h_0, κατά την οποία το βαριόγραμμα παρουσιάζει μια σαφή ανοδική τάση, ονομάζεται *ζώνη επιρροής* (*range*) του βαριογράμματος, και είναι ίση με την τιμή. Για *h* = 0, όταν το σφάλμα στη μέτρηση του μεγέθους *z* είναι μηδενικό, το *γ* είναι επίσης μηδενικό. Αν το σφάλμα δεν είναι μηδενικό, τότε η καμπύλη του βαριογράμματος τέμνει τον κατακόρυφο άξονα σε μια θετική τιμή *γ*, που ορίζει το *nugget* του βαριογράμματος. Αν το nugget είναι μηδέν, τότε η τιμή κατωφλίου είναι ίση με τη διασπορά s^2 του *z* στην περιοχή έρευνας.



Σχ. 19. Η συνάρτηση βαριόγραμμα

Το βαριόγραμμα που παράγεται από δεδομένα υπαίθρου δεν έχει τόσο ομαλή συμπεριφορά όπως στο (σχ. 19), μπορεί όμως να προσεγγιστεί από χαρακτηριστικές συναρτήσεις (μοντέλα), όπως το *σφαιρικό μοντέλο,* το *εκθετικό μοντέλο,* το *κανονικό μοντέλο* και το *γραμμικό μοντέλο*.
Το σφαιρικό (spherical) μοντέλο ορίζεται από τη σχέση:

$$\gamma(h)=\begin{cases} a+(s^2-a)(\dfrac{3h}{2h_0}-\dfrac{h^3}{2h_0{}^3}) & 0<h<h_0 \\ 0 & h=0 \\ s^2 & h\geq h_0 \end{cases} \qquad (21)$$

a είναι η τιμή nugget, s^2 είναι η τιμή κατωφλίου (διασπορά) και h_0 είναι το μήκος της ζώνης επιρροής.
Το εκθετικό (exponential) μοντέλο είναι:

$$\gamma(h)=\begin{cases} a+(s^2-a)\cdot[1-\exp(-3h/h_0)] & h>0 \\ 0 & h=0 \end{cases} \qquad (22)$$

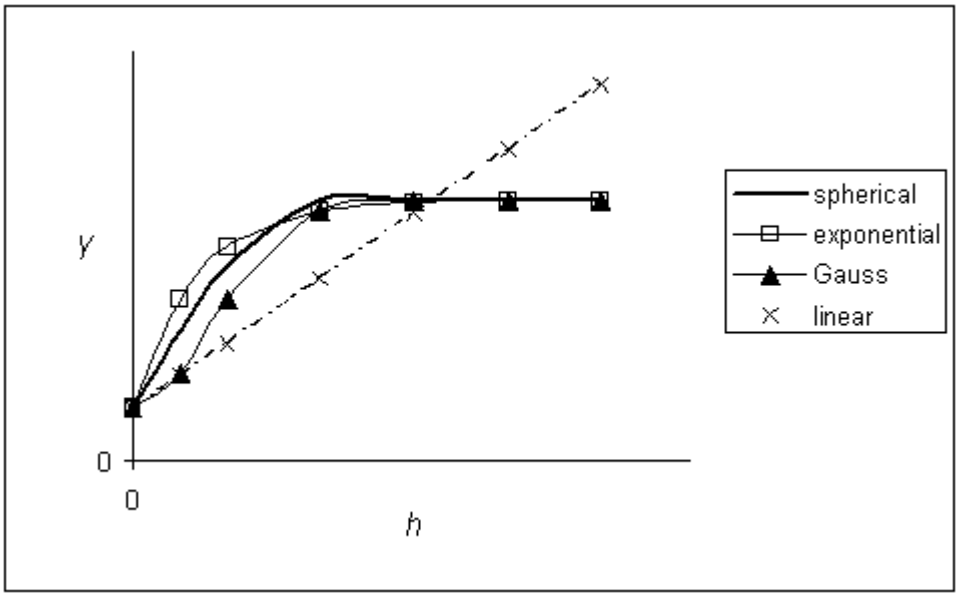
Το κανονικό μοντέλο, ή μοντέλο Gauss, είναι:

$$\gamma(h)=\begin{cases} a+(s^2-a)\cdot[1-\exp(-3h^2/h_0{}^2)] & h>0 \\ 0 & h=0 \end{cases} \qquad (23)$$

Το γραμμικό (linear) μοντέλο είναι:

$$\gamma(h)=\begin{cases} a+bh & h>0 \\ 0 & h=0 \end{cases} \qquad (24)$$

b είναι η κλίση της ευθείας του βαριογράμματος.
Στο (σχ. 20) παρουσιάζεται η μεταβολή του *γ* ως προς *h* για κάθε μοντέλο βαριογράμματος.



Σχ. 20. Βαριογράμματα με βάση το σφαιρικό, το εκθετικό, το κανονικό και το γραμμικό μοντέλο

Άλλα μοντέλα βαριογράμματος παρουσιάζονται από τον Liang 2004. Η επιλογή του κατάλληλου μοντέλου για την προσέγγιση του βαριογράμματος των δεδομένων υπαίθρου, εξαρτάται από τη συμπεριφορά του ως προς *h*. Ο υπολογισμός των παραμέτρων του μοντέλου, μπορεί να γίνει είτε με τη μέθοδο δοκιμής-σφάλματος, είτε με αυτοματοποιημένους αλγορίθμους (GSLIB Conventions), και αποτελεί ένα από τα βήματα της διαδικασίας χωρικής παρεμβολής με Kriging.
Για την εκτίμηση της χωρικής συσχέτισης τιμών ***z*** ενός φυσικού μεγέθους, χρησιμοποιούνται επίσης οι *δείκτες χωρικής αυτοσυσχέτισης Moran* και *Geary*.
Ο δείκτης Moran *I* ορίζεται ως:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(z_i - \bar{z})(z_j - \bar{z})}{\left(\sum_{i=1}^{n}(z_i - \bar{z})^2\right)\left(\sum_{i\neq j}\sum w_{ij}\right)} \qquad (25)$$

n είναι ο αριθμός των σημείων

z^- είναι η μέση τιμή
z_i και z_j είναι οι παρατηρούμενες τιμές στις θέσεις (x_i, y_i) και (x_j, y_j), αντίστοιχα
w_{ij} είναι μια ποσότητα που εκφράζει τη χωρική εγγύτητα μεταξύ (x_i, y_i) και (x_j, y_j) και ορίζεται ως:

w_{ij} = 1, αν η απόσταση μεταξύ (x_i, y_i) και (x_j, y_j) είναι μέσα στα όρια ενός συγκεκριμένου διαστήματος τιμών απόστασης [$h - \Delta h$, $h + \Delta h$]
w_{ij} = 0, αν η απόσταση μεταξύ (x_i, y_i) και (x_j, y_j) είναι μέσα στο διάστημα τιμών [$h - \Delta h$, $h + \Delta h$]

Ο δείκτης Geary *c* ορίζεται ως:

$$c = \frac{(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(z_i - z_j)^2}{2\left(\sum_{i=1}^{n}(z_i - \bar{z})^2\right)\left(\sum\sum_{i\neq j} w_{ij}\right)} \qquad (26)$$

Ο δείκτης *I* λαμβάνει θετικές τιμές για σημαντική χωρική συσχέτιση και τιμές κοντά στο μηδέν, ή και αρνητικές, όταν δεν υπάρχει χωρική συσχέτιση.
Ο δείκτης *c* είναι κοντά στο μηδέν για υψηλή χωρική συσχέτιση και λαμβάνει θετικές τιμές όταν δεν υπάρχει σημαντική χωρική συσχέτιση.

3.3.6. Χωρική παρεμβολή με Kriging

Στην παράγραφο 3.3.3 παρουσιάσαμε τη μέθοδο της χωρικής παρεμβολής με κινούμενο μέσο όρο. Επισημάναμε ότι υπάρχει ένας υποκειμενισμός ως προς την επιλογή της ακτίνας του κύκλου με κέντρο το σημείο (*x*, *y*), όπου επιχειρείται ο υπολογισμός του μεγέθους *z*, μέσα στον οποίο συμπεριλαμβάνονται τα σημεία ελέγχου που διαμορφώνουν το μέσο όρο (ακτίνα επιρροής). Γενικότερα, δεν είδαμε ένα κριτήριο που να προσδιορίζει πόσα σημεία ελέγχου συμμετέχουν στη διαμόρφωση του μέσου όρου. Επίσης, η παραδοχή ότι οι συντελεστές βαρύτητας των τιμών των σημείων ελέγχου είναι αντιστρόφως ανάλογοι της απόστασης από το (*x*, *y*), είναι επίσης αυθαίρετη, έστω και αν διαισθητικά φαίνεται

να στέκει. Επί πλέον, με τους κινούμενους μέσους όρους, δεν υπάρχει δυνατότητα υπολογισμού του σφάλματος, ή του επί τοις εκατό διαστήματος εμπιστοσύνης, του υπολογιζόμενου μεγέθους.
Ανάλογες παρατηρήσεις μπορούν να διατυπωθούν και για τις μεθόδους χωρικής παρεμβολής με τοπικές επιφάνειες τάσης ή με splines, κυρίως σε ό,τι αφορά την ακτίνα επιρροής και το πλήθος των συμμετεχόντων σημείων ελέγχου, το είδος της τοπικής επιφάνειας τάσης και το σφάλμα στο υπολογιζόμενο μέγεθος.
Στην περίπτωση χωρικής παρεμβολής με κινούμενο μέσο όρο, διατυπώθηκε, στην παράγραφο 3.3.5., ένα κριτήριο προσδιορισμού της ακτίνας επιρροής, με τη βοήθεια του συνβαριογράμματος, ή ισοδύναμα, του βαριογράμματος των τιμών των σημείων ελέγχου. Μια τέτοια πρακτική δεν συνηθίζεται σε αυτό το είδος χωρικής παρεμβολής, δείχνει όμως την ανάγκη και τη δυνατότητα ανάπτυξης μιας εναλλακτικής μεθοδολογίας εκτίμησης τιμών *z* από σημεία ελέγχου, που να αξιοποιεί την πληροφορία που δίνει η χωρική συσχέτιση του μετρούμενου μεγέθους. Αυτή η εναλλακτική μεθοδολογία χωρικής παρεμβολής, όπου αξιοποιείται η χωρική συσχέτιση για να αντιμετωπιστούν οι αδυναμίες των μεθόδων που εκτέθηκαν στις παραγράφους 3.3.3. και 3.3.4., έχει το γενικό όνομα *Kriging* (Krige: ο μηχανικός μεταλλείων που ανακάλυψε τη μέθοδο).
Η μεθοδολογία Kriging, αναπτύσσεται στη βάση της παραδοχής ότι το μετρούμενο μέγεθος *z* είναι μια *περιφερειοποιημένη μεταβλητή* (*regionalized variable*), δηλαδή μια μονοσήμαντη και ομαλά μεταβαλλόμενη μεταβλητή, που δεν λαμβάνει τυχαίες τιμές, ούτε όμως περιγράφεται με απόλυτη ακρίβεια από μια γεωμετρική επιφάνεια που να ορίζεται από μια μαθηματική συνάρτηση. Η περιφερειοποιημένη μεταβλητή έχει μια *συνιστώσα τάσης* (*drift component*), που μπορεί να είναι μια σταθερή μέση τιμή ή μια επιφάνεια πρώτου, δεύτερου ή τρίτου βαθμού. Η δεύτερη συνιστώσα της περιφερειοποιημένης μεταβλητής είναι η *υπολοιπόμενη* (*residual*) *συνιστώσα*, που προσδιορίζεται αφαιρώντας τη συνιστώσα τάσης από την τιμή της περιφερειοποιημένης μεταβλητής.
Υπάρχουν διάφορες επί μέρους εκδοχές Kriging, ανάλογα με τη συμπεριφορά της συνιστώσας τάσης, την ανισοτροπία των τιμών *z* των σημείων ελέγχου, το αν επιχειρείται ο υπολογισμός του *z* σε σημείο ή σε πολύγωνο, καθώς και με το αν επιχειρείται ο υπολογισμός τιμών δυο ή περισσότερων περιφερειοποιημένων μεταβλητών.

3.3.6.1. Σύνηθες Kriging

Το *σύνηθες* (*ordinary*) Kriging πραγματοποιείται με βάση την παραδοχή, ότι η συνιστώσα τάσης είναι σταθερή σε όλη την περιοχή έρευνας, με τιμή *μ*. Αυτό σημαίνει, ότι το μετρούμενο μέγεθος (περιφερειοποιημένη μεταβλητή) *z*, σε κάθε θέση (*x*, *y*), μπορεί να εκφραστεί ως:

$$z(x, y) = \mu + u(x, y) \quad (27)$$

u είναι η υπολοιπόμενη μεταβλητή.
Το ζητούμενο είναι ο προσδιορισμός του *u* σε κάθε θέση (*x*, *y*), μέσω ενός γραμμικού συνδυασμού της μορφής:

$$u' = \sum_{i=1}^{n} w_i u(x_i, y_i) \quad (28)$$

$u(x_i, y_i)$ είναι οι υπολοιπόμενες τιμές των σημείων ελέγχου. *N* είναι το πλήθος όλων των σημείων ελέγχου. w_i είναι οι συντελεστές βαρύτητας, που το άθροισμά τους είναι ίσο με τη μονάδα. *u'* είναι η εκτιμώμενη υπολοιπόμενη τιμή στο σημείο (*x*, *y*), που, γενικά, διαφέρει από την αληθή τιμή *u*. Αν προσδιοριστούν τα *u'*(*x*, *y*), τότε μπορούν να βρεθούν οι εκτιμώμενες τιμές *z'* από τη σχέση:

$$z'(x, y) = \mu + u'(x, y) \quad (29)$$

Οι τιμές w_i εκπληρώνουν τη συνθήκη της ελάχιστης μέσης τιμής *Q* του τετραγώνου της διαφοράς $(u - u')^2$ στη θέση (*x*, *y*).
Το *Q* εκφράζεται, κατ'αρχήν, ως:

$$Q = \mathrm{E}\left[\left(u(x,y) - u'(x,y)\right)^2\right] = \mathrm{E}\left[\left(u(x,y) - \sum_{i=1}^{n} w_i u(x_i, y_i)\right)^2\right] \quad (30)$$

Το σύμβολο Ε υποδηλώνει μέση τιμή.
Το *Q* δεν αλλάζει, αν τα *u*(*x*, *y*) και *u'*(*x*, *y*) αντικατασταθούν από τα *u*(*x*, *y*) – *μ* και *u'*(*x*, *y*) – *μ*, αντίστοιχα.
Οπότε

$$Q = \mathrm{E}\left[\left([u(x,y)-\mu]-\sum_{i=1}^{n} w_i[u(x_i,y_i)-\mu]\right)^2\right] \quad (31)$$

Μετά από κάποια αλγεβρική επεξεργασία, η σχέση (31) γίνεται:

$$Q = C_0 + \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j C_{ij} - 2\sum_{i=1}^{n} w_i C_i \quad (32)$$

C_0 είναι η συνδιασπορά για απόσταση h ίση με μηδέν, $C(0)$.
C_{ij} είναι η συνδιασπορά $C(h_{ij})$ για απόσταση σημείων ελέγχου (x_i, y_i) και (x_j, y_j) ίση με h_{ij}, με $i \neq j$.
C_i είναι η συνδιασπορά $C(h_i)$ μεταξύ σημείου (x, y), όπου πρόκειται να υπολογιστεί το *u'*, και σημείου ελέγχου (x_i, y_i), σε απόσταση h_i από το (x, y).
Οι ποσότητες C_0, C_{ij} και C_i υπολογίζονται με βάση κάποιο από τα μοντέλα προσέγγισης βαριογράμματος των σχέσεων (21) ως (24), με κατάλληλες παραμέτρους a και h_0 που να προσεγγίζουν το βαριόγραμμα της υπό μελέτη περιοχής. Έχοντας προσδιορίσει τη συνάρτηση $\gamma(h)$, είναι δυνατό να υπολογιστούν οι τιμές $C(h)$, με βάση τη σχέση (20), θέτοντας $C(0)$ ίση με την τιμή κατωφλίου της $\gamma(h)$.
Η ελαχιστοποίηση του Q, στη σχέση (32), εξασφαλίζεται με μηδενισμό των παραγώγων $\partial Q/\partial w_i$. Αυτό, σε συνδυασμό με τον περιορισμό $\Sigma_{i=1}^{n} w_i = 1$, οδηγεί σε σύστημα $n+1$ εξισώσεων με n αγνώστους. Για να βρεθεί η ακριβής λύση του συστήματος, η σχέση (32) γράφεται ως:

$$Q = C_0 + \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j C_{ij} - 2\sum_{i=1}^{n} w_i C_i + 2\lambda\left(\sum_{i=1}^{n} w_i - 1\right) \quad (33)$$

Η σχέση (32) είναι ισοδύναμη με την (31), καθώς το άθροισμα των w_i ισούται με τη μονάδα.
Το λ είναι ο *πολλαπλασιαστής Lagrange*, που είναι μια άγνωστη πραγματική ποσότητα, η οποία λειτουργεί «βοηθητικά» για να προσδιοριστούν επακριβώς οι τιμές w_i (περισσότερα για το πώς ορίζεται και πώς αξιοποιείται ο πολλαπλασιαστής Lagrange, μπορεί κανείς να δει στον Logan 1999).
Δεδομένου ότι για κάθε w_i ισχύει, λόγω της συνθήκης ελαχιστοποίησης του Q, ότι:

$$\frac{\partial Q}{\partial w_i} = 2\sum_{j=1}^{n} w_j C_{ij} - 2C_i + 2\lambda = 0 \qquad (34)$$

καθώς και ότι:

$$\sum_{i=1}^{n} w_i = 1 \qquad (35)$$

καταλήγουμε στο παρακάτω γραμμικό σύστημα n+1 εξισώσεων με αγνώστους w_1, w_2,…,w_n, λ:

$$\begin{aligned} &w_1C_0 + w_2C_{12} + \ldots + w_nC_{1n} + \lambda = C_1 \\ &w_1C_{12} + w_2C_0 + \ldots + w_nC_{2n} + \lambda = C_2 \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &w_1C_{1n} + w_2C_{2n} + \ldots + w_nC_0 + \lambda = C_n \\ &w_1 + w_2 + \ldots + w_n + 0\lambda = 1 \end{aligned} \qquad (36)$$

Το γραμμικό σύστημα (36), επιλύεται με τη βοήθεια της άλγεβρας πινάκων. Στη συνέχεια, γνωρίζοντας τους συντελεστές βαρύτητας w_i, καθώς και τη μέση τιμή μ, υπολογίζονται οι εκτιμώμενες υπολοιπόμενες τιμές $u'(x, y)$ μέσω της σχέσης (28), καθώς και οι εκτιμώμενες τιμές της περιφερειοποιημένης μεταβλητής z', μέσω της σχέσης (29).
Αντικαθιστώντας τις λύσεις w_i του γραμμικού συστήματος (36) στη σχέση (32), και μετά από κάποια αλγεβρική επεξεργασία, βρίσκεται η ελάχιστη τιμή σ_k^2 της ποσότητας Q. Το σ_k^2 εκφράζει, μέσω της σχέσης (30), την ελάχιστη μέση τετραγωνική απόκλιση μεταξύ αληθούς και εκτιμώμενης υπολοιπόμενης τιμής στη θέση (x, y) (διασπορά).
Η έκφραση για το σ_k^2 είναι (Isaaks & Srivastava 1989):

$$\sigma_k{}^2 = C_0 - \sum_{i=1}^{n} w_i C_i - \lambda \qquad (37)$$

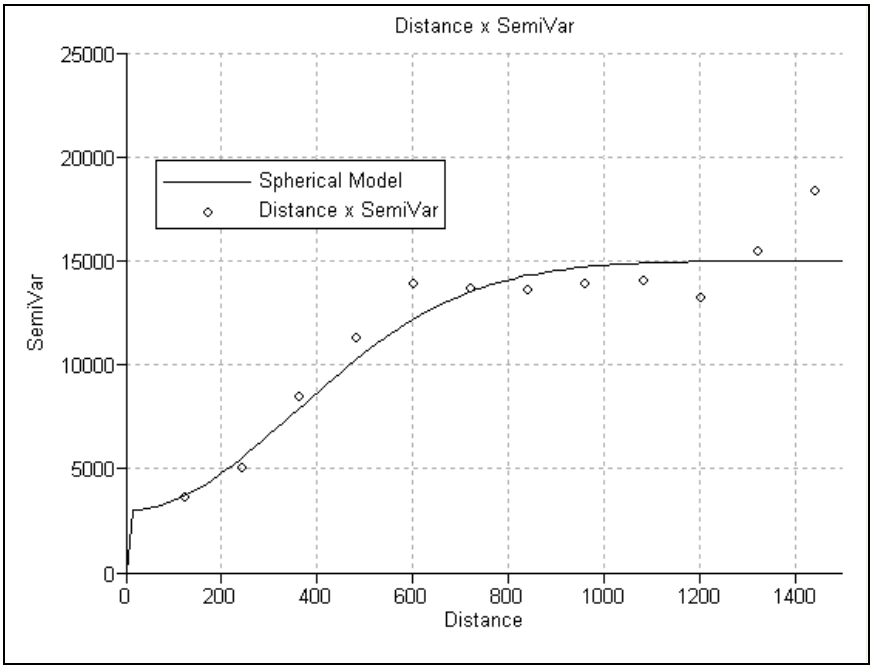
και το *σφάλμα Kriging* σ_k είναι:

$$\sigma_k = \sqrt{C_0 - \sum_{i=1}^{n} w_i C_i - \lambda} \qquad (38)$$

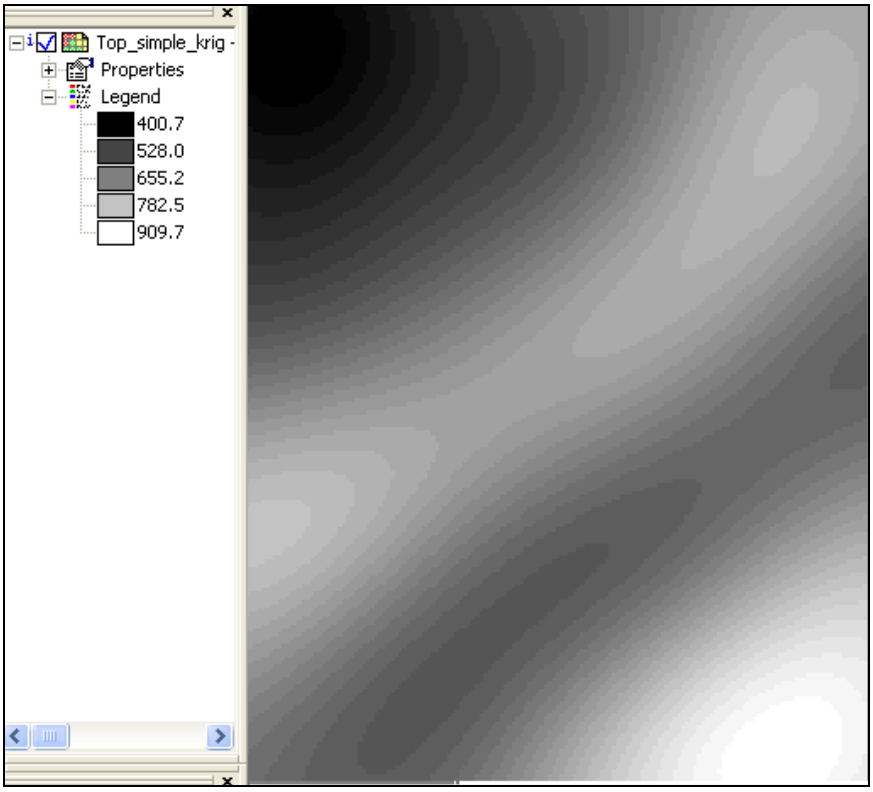
Με βάση το σ_k μπορούν να υπολογιστούν τα επί τοις εκατό διαστήματα εμπιστοσύνης της εκτιμώμενης τιμής του μεγέθους *z* στη θέση (*x*, *y*). Με βάση τις σχέσεις (28), (29) και (38), το διάστημα εμπιστοσύνης είναι:

$$z' \pm z_c \sigma_k = \left[\mu + \sum_{i=1}^{n} w_i u'_i - z_c \sqrt{C_0 - \sum_{i=1}^{n} w_i C_i - \lambda},\ \mu + \sum_{i=1}^{n} w_i u'_i + z_c \sqrt{C_0 - \sum_{i=1}^{n} w_i C_i - \lambda} \right] \qquad (39)$$

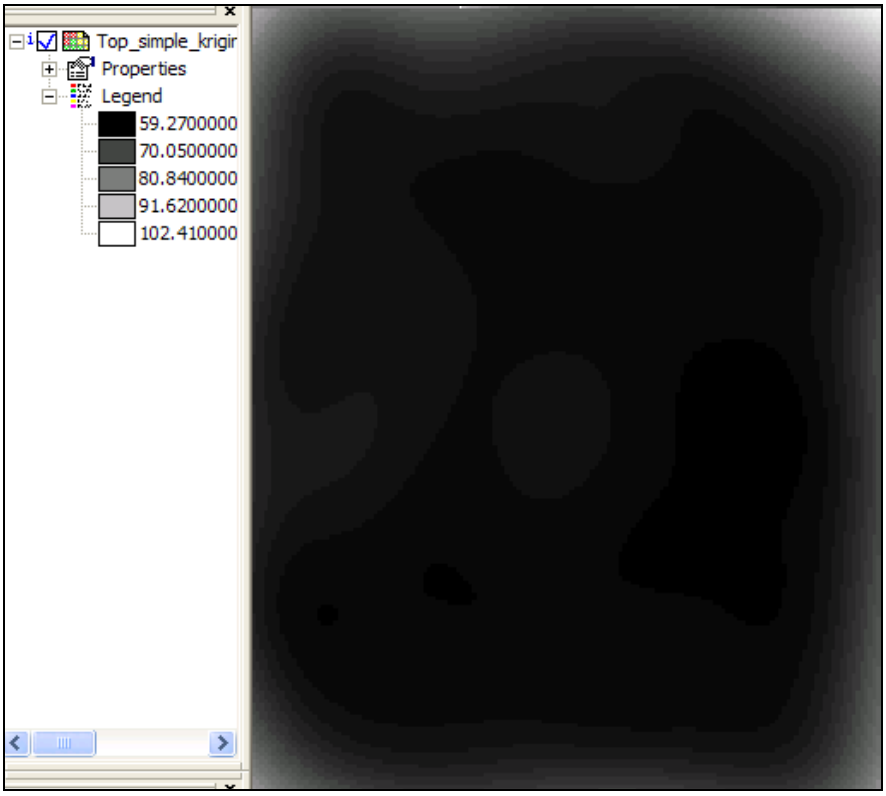
z_c είναι το εκατοστιαίο κρίσιμο σημείο της τυποποιημένης κανονικής κατανομής, που αντιστοιχεί στο επί τοις εκατό διάστημα εμπιστοσύνης.
Το σύνηθες Kriging εφαρμόστηκε στις τιμές *z* της χωρικής κατανομής της παραγράφου (3. 1). Πρώτα, προσδιορίστηκε το βαριόγραμμα της περιοχής μελέτης, που εμφανίζεται στο (σχ. 21). Το βαριόγραμμα αυτό προσεγγίστηκε με το γκαουσιανό μοντέλο της σχέσης (23), θέτοντας nugget *a* = 3000, κατώφλι *s* = 15000 και ζώνη επιρροής h_0 = 500. Η καμπύλη του βαριογράμματος του μοντέλου αυτού, παρουσιάζεται επίσης στο (σχ. 21).
Από τις τιμές *C*(*h*) του γκαουσιανού μοντέλου βαριογράμματος, προσδιορίστηκαν τα C_0, C_{ij} και C_i και, μέσω του γραμμικού συστήματος της σχέσης (36), τα w_i και *λ*. Στη συνέχεια, από τις σχέσεις (28) και (29), υπολογίστηκαν οι τιμές χωρικής παρεμβολής *z'*(*x*, *y*) (σχ. 22) και, μέσω των σχέσεων (38) και (39), τα σφάλματα Kriging (z_c = 1). Ο χάρτης των σφαλμάτων εμφανίζεται στο (σχ. 23).



Σχ. 21. Βαριόγραμμα των τιμών *z* της χωρικής κατανομής του (σχ. 6)



Σχ. 22. Χάρτης χωρικής παρεμβολής με σύνηθες Kriging



Σχ. 23. Χάρτης σφαλμάτων Kriging

3.3.6.2. Άλλες μορφές Kriging

Εκτός από τη συνήθη Kriging, υπάρχουν και άλλες μέθοδοι χωρικής παρεμβολής, που είναι παραλλαγές αυτής και ανταποκρίνονται σε διαφορετικές συμπεριφορές μετρούμενων τιμών ή και σε διαφορετικά ζητούμενα. Αυτές οι εναλλακτικές μέθοδοι Kriging παρουσιάζονται παρακάτω συνοπτικά. Για περισσότερες πληροφορίες, μπορεί κανείς να ανατρέξει στους Isaaks & Srivastava 1989.

3.3.6.2.α. *Universal Kriging*

Η μέθοδος αυτή ενδείκνυται για περιπτώσεις, όπου υπάρχει μια ευρύτερη τάση χωρικής μεταβολής της τοπικής μέσης τιμής. Στη χωρική παρεμβολή με Universal Kriging, προσδιορίζεται η επιφάνεια τάσης και αφαιρείται από τις τιμές της περιφερειοποιημένης μεταβλητής. Στις

υπολοιπόμενες συνιστώσες εφαρμόζεται μια συνήθης χωρική παρεμβολή Kriging και στις τιμές που προκύπτουν προστίθεται η επιφάνεια τάσης. Universal Kriging εφαρμόστηκε στα δεδομένα της χωρικής κατανομής του (σχ. 6) και οι τιμές πάχους *z*, που προέκυψαν, διαφέρουν σημαντικά από αυτές της συνήθους Kriging, σε ποσοστά που φτάνουν το 10%. Και τα σφάλματα, επίσης, είναι κάπως μεγαλύτερα, σε σχέση με αυτά που εκτιμήθηκαν με συνήθη Kriging.

3.3.6.2.β. *Ανισοτροπικό Kriging*

Ανισοτροπία ως προς τη συμπεριφορά των τιμών μετρούμενου μεγέθους *z*, εκδηλώνεται στο βαριόγραμμα, η συμπεριφορά του οποίου είναι διαφορετική με το αζιμούθιο. Αυτή η ανισοτροπία στη συμπεριφορά του βαριογράμματος, λαμβάνεται υπόψη στο ανισοτροπικό Kriging.

3.3.6.2.γ. *Block Kriging*

Η μέθοδος αυτή εφαρμόζεται, όταν το ζητούμενο δεν είναι η χωρική παρεμβολή ανά σημείο, αλλά η εκτίμηση της μέσης τιμής της περιφερειοποιημένης μεταβλητής *z* σε διαφορετικές υποπεριοχές της ευρύτερης περιοχής έρευνας.

3.3.6.2.δ. *Co Kriging*

Εφαρμόζεται, όταν σε κάθε σημείο της περιοχής έρευνας μετρώνται δυο ή περισσότερες περιφερειοποιημένες μεταβλητές.

ΒΙΒΛΙΟΓΡΑΦΙΑ

ΑΛΕΞΑΝΔΡΟΠΟΥΛΟΣ ΑΝ. - ΚΑΤΩΠΟΔΗΣ ΕΠ. - ΠΑΛΙΑΤΣΟΣ ΑΘ. - ΠΡΕΒΕΖΑΚΟΣ Ν. 1994. Στατιστική. Σύγχρονη Εκδοτική. Αθήνα

APOSTOL TOM M.1962. Διαφορικός και Ολοκληρωτικός Λογισμός. Τομος ΙΙ. Μ. Πεχλιβανίδης & Σια Α.Ε. Αθήνα.

ΒΑΪΟΠΟΥΛΟΣ, Δ.Α. 1995. Στοιχεία Στατιστικής. Σημειώσεις για το μάθημα Μεταπτυχιακού Ωκεανογραφίας. Αθήνα.

ΒΑΪΟΠΟΥΛΟΣ, Δ.Α. 2004. Εισαγωγή στην Πληροφορική. Αθήνα.

ΕΓΚΥΚΛΟΠΑΙΔΕΙΑ ΜΑΘΗΜΑΤΙΚΩΝ.1974. Λογισμός Πιθανοτήτων και Στατιστική. Τόμος Ε. Παγουλάτος. Αθήνα.

FRANCIS, A. 1988. Advanced level Statistics. 2nd ed. Stanley Thornes (Publishers). G. Britain.

FREUND, I. E. 1979. Modern Elementary Statistics. 5th ed. Prentice Hall. London.

GSLIB CONVENTIONS: Geostatistical Simulations for the Mining Industry.

ILWIS 3. 3. Software package. ITC.

ISAAKS, E. H., SRIVASTAVA, R. M., 1989: An Introduction to Applied Geostatistics. Oxford University Press.

ΚΑΚΟΥΛΛΟΣ, Θ.Ν. 1972. Στατιστική, Θεωρία και Εφαρμογαί.

ΚΙΟΧΟΣ, Π.Α. 1993 Στατιστική. Ίδρυμα Ευγενίδου. Αθήνα.

ΚΙΤΣΟΣ Χ.Π. 1992. Θέματα Εφαρμοσμένης Στατιστικής. Εκδόσεις Νέων Τεχνολογιών. Αθ.

ΚΟΥΤΣΟΠΟΥΛΟΣ, Κ., 2002.Γεωγραφικά Συστήματα Πληροφοριών και Ανάλυση Χώρου. Εκδόσεις Παπασωτηρίου.

KREYSZIG, B., 1979: Advanced Engineering Mathematics. John Wiley & Sons.

ΚΥΡΙΑΚΟΥΣΗΣ Γ.Α. 1993. Βιοστατιστική. Αθήνα.

LIANG, S., 2004: Quantitative Remote Sensing of Land Surfaces. Wiley.

LOGAN, D. J., 2002: Εφαρμοσμένα Μαθηματικά. Πανεπιστημιακές Εκδόσεις Κρήτης

ΜΑΚΡΗΣ,Π. 1988. Ο Computer με απλά λόγια. Εκδόσεις Πέρσοναλ. Αθήνα.

PEKELIS, VICTOR. 1986. Κυβερνητική. Από το Α ως το Ω. Εκδόσεις Gutenberg. Αθήνα.

ΣΑΡΑΝΤΟΠΟΥΛΟΣ ΣΠ. Β. 1961. Λογισμός των Πιθανοτήτων και Στατιστική.Τόμος Α'. Αθήνα.

SPIEGEL,MURRAY.R.1977. Πιθανότητες και Στατιστική. McGraw-Hill New York, ΕΣΠΙ Αθήνα.

SWAN, A.R.H., AND SANDILANDS, M., 1995: Introduction to Geological Data Analysis. Blackwell Science.

WOLF F.L. 1974. Elements of Probability and Statistics. 2nd ed. Mc Graw-Hill.USA.

ΦΟΥΝΤΑΣ ΓΡ.Χ. Μεγάλο Μαθηματικό Τυπολόγιο. Γρ. Φούντας.

.