ΕΝΟΤΗΤΑ Α

ΣΤΑΤΙΣΤΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΑΝΑΛΥΤΙΚΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

Καθηγητή Κωνσταντίνου Η. Ευσταθίου,
Εργαστήριο Αναλυτικής Χημείας
Πανεπιστημίου Αθηνών

"Το αναλυτικό αποτέλεσμα δεν έχει την παραμικρή αξία αν δεν συνοδεύεται από κάποια εκτίμηση του εμπεριεχόμενου πιθανού σφάλματος"

1. ΣΦΑΛΜΑΤΑ ΣΤΙΣ ΜΕΤΡΗΣΕΙΣ

1.1. Αναλυτικό σφάλμα

Κάθε μέτρηση μιας φυσικής ή χημικής ποσότητας υπόκειται σε σφάλματα, τα οποία προσδίδουν στο αποτέλεσμα μια αβεβαιότητα. Η αβεβαιότητα μπορεί να εκτιμηθεί, μπορεί να ελαχιστοποιηθεί, αλλά ποτέ δεν μπορεί να μηδενιστεί.

Το **σφάλμα** (error) Ε παρέχεται από τη σχέση

$$E = x_i - \mu \qquad (1\text{-}1)$$

όπου x_i είναι η πειραματική τιμή και μ η πραγματική τιμή.

Το **σχετικό σφάλμα** (relative error) E_r παρέχεται από τη σχέση

$$E_r = E/\mu = (x_i - \mu) / \mu \qquad (1\text{-}2)$$

και συνήθως εκφράζεται ως **σχετικό σφάλμα επί τοις εκατό** (relative error per cent), $E_r\%$:

$$E_r\% = E_r \times 100 \qquad (1\text{-}3)$$

1.2. Τύποι αναλυτικών σφαλμάτων

Υπάρχουν δύο τύποι αναλυτικών σφαλμάτων:

1. Τα **συστηματικά** (systematic) ή **καθορισμένα** (determinate), που οφείλονται σε ένα ή περισσότερους λόγους γνωστής προέλευσης (π.χ. ακαθαρσίες αντιδραστηρίων, προσωπικά σφάλματα, οργανολογικά προβλήματα, ατέλεια της χρησιμοποιούμενης αναλυτικής τεχνικής).
2. Τα **τυχαία** (random) ή **ακαθόριστα** (indeterminate) σφάλματα, που οφείλονται σε απροσδιόριστα αίτια και στο "στατιστικό" θόρυβο που ενυπάρχει σε κάθε διαδικασία μέτρησης.

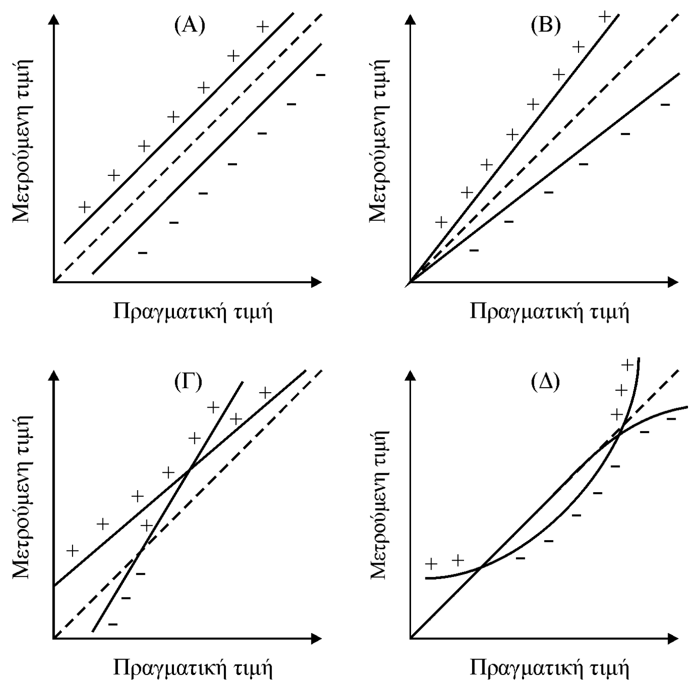
Η σύγχρονη παρουσία συστηματικών και τυχαίων σφαλμάτων, που είναι πάντοτε ανεξάρτητα μεταξύ τους, είναι μία σχετικά συνήθης περίπτωση, οπότε το ολικό αναλυτικό σφάλμα μπορεί να αναλυθεί σε δύο συστατικά σύμφωνα με την εξίσωση:

$$\text{Ολικό αναλυτικό σφάλμα} = \text{συστηματικό σφάλμα} + \text{τυχαίο σφάλμα} \qquad (1\text{-}4)$$

1.3. Εισαγωγική θεώρηση συστηματικών σφαλμάτων

Η εξακρίβωση αρχικά της παρουσίας και στη συνέχεια του τύπου του συστηματικού σφάλματος είναι τα πρώτα βήματα που θα οδηγήσουν στην ανεύρεση της πηγής του. Μόλις βρεθεί η πηγή, εξετάζεται το κατά πόσο είναι δυνατόν αυτή να εκλείψει. Εάν τούτο είναι αδύνατον εξ αιτίας παραγόντων όπως η φύση της μεθόδου, η χρησιμοποιούμενη τεχνική, ο τύπος του δείγματος, ή κάποιου συνδυασμού των παραγόντων αυτών, εξετάζεται (ως ύστατη λύση) η έμμεση διόρθωση με μαθηματική επέμβαση στο αποτέλεσμα.

Η μαθηματική επέμβαση πραγματοποιείται αφού προηγουμένως εξακριβωθεί πλήρως ο τύπος του συστηματικού σφάλματος και διερευνηθούν προσεκτικά τα μαθηματικά του χαρακτηριστικά και οι τυχόν εξαρτήσεις του (π.χ. από τη φύση του δείγματος, των υπόλοιπων συστατικών). Εάν η μαθηματική επέμβαση είναι εφικτή, τότε αυτή καθίσταται μέρος της μεθόδου. Εάν ούτε αυτή είναι εφικτή πρέπει να αρχίσει πλέον η αναζήτηση άλλης αναλυτικής μεθόδου.



Σχήμα 1-1. *Διαγράμματα συσχέτισης για τον καθορισμό του τύπου του συστηματικού σφάλματος μιας αναλυτικής μεθόδου (οι καμπύλες με διακεκομμένη γραμμή αποδίδουν την ιδανική συσχέτιση). (Α) σταθερό σφάλμα, (Β) αναλογικό σφάλμα, (Γ) σύνθετο σφάλμα, (Δ) περίπλοκο/μικτό σφάλμα.*

Ο προσδιορισμός του είδους του συστηματικού σφάλματος πραγματοποιείται με θεώρηση του **διαγράμματος συσχέτισης** (correlation plot) πραγματικών - μετρούμενων τιμών χημικών μονάδων (συγκέντρωσης ή μάζας) και τυπικές μορφές τους δείχνονται στο Σχήμα 1-1. Απαραίτητη προϋπόθεση για τη χάραξη ενός διαγράμματος συσχέτισης είναι η μέτρηση ικανού αριθμού δειγμάτων με γνωστή συγκέντρωση ή ποσότητα του προσδιοριζόμενου συστατικού. Τα δείγματα αυτά είτε παρασκευάζονται (συνθετικά), είτε είναι φυσικά και έχουν προηγουμένως μετρηθεί με τεχνική γνωστής αξιοπιστίας και απαλλαγμένη από συστηματικά σφάλματα.

Τα δείγματα που χρησιμοποιούνται για την κατασκευή των διαγραμμάτων συσχέτισης πρέπει να καλύπτουν ευρεία περιοχή τιμών περιεκτικότητας ή συγκέντρωσης. Ιδανικά πρέπει να καλύπτουν ολόκληρη την περιοχή των αναμενόμενων τιμών. Σημεία των διαγραμμάτων που ομαδοποιούνται σε περιοχές εκτός της καμπύλης συσχέτισης, συνήθως αποκαλύπτουν ειδικού τύπου παρεμποδίσεις της εξεταζόμενης μεθόδου.

Τα συστηματικά σφάλματα ταξινομούνται στις ακόλουθες κατηγορίες, ανάλογα με τη μορφή της καμπύλης του διαγράμματος συσχέτισης:

1. **Σταθερό σφάλμα** (constant error): Αποδίδεται από τις καμπύλες του διαγράμματος συσχέτισης 1-1Α. Η καμπύλη συσχέτισης είναι παράλληλη προς την ιδανική και δεν διέρχεται από την αρχή των αξόνων. Οδηγεί σε μονοκατευθυνόμενο (θετικό ή αρνητικό) αναλυτικό σφάλμα (π.χ. με πρόσημα σφαλμάτων του τύπου +++++ ή − − − − −). Είναι προφανές ότι το σχετικό σφάλμα είναι μεγαλύτερο στα "χαμηλά" δείγματα. Η διαπίστωση ικανοποιητικής **ανάκτησης** (recovery) κατά τον έλεγχο της μεθόδου δεν αποκαλύπτει την παρουσία σταθερού σφάλματος.

Εάν δεν είναι δυνατή η εξάλειψη του σταθερού συστηματικού σφάλματος, μπορεί απλά να διαπιστωθεί το μέγεθός του και εφόσον είναι σταθερό σε όλα τα δείγματα τα αποτελέσματα μπορούν να υποστούν μαθηματική διόρθωση. Είναι προφανές ότι τούτο δεν ισχύει, όταν το σταθερό σφάλμα οφείλεται σε παρεμποδίζουσες ουσίες των οποίων η συγκέντρωση ποικίλει από δείγμα σε δείγμα.

Η σύγχρονη παρουσία σταθερού συστηματικού σφάλματος και έντονου τυχαίου σφάλματος περιορίζει σημαντικά τη δυνατότητα μετρήσεων χαμηλών συγκεντρώσεων ή ποσοτήτων, αφού αυξάνεται σημαντικά η αβεβαιότητα του αποτελέσματος.

Η παρουσία σταθερού αλλά άγνωστου σε τιμή συστηματικού σφάλματος (τυφλή μέτρηση διαφορετική του μηδενός) στο στάδιο της μέτρησης (μόνο) αποκλείει τη **γνωστή προσθήκη** (standard known addition) ως τεχνική ποσοτικοποίησης.

2. **Αναλογικό σφάλμα** (proportional error). Αποδίδεται από τις καμπύλες του διαγράμματος συσχέτισης 1-1Β. Η καμπύλη συσχέτισης διέρχεται από την αρχή των αξόνων (σημείο 0,0). Οδηγεί σε μονοκατευθυνόμενο αναλυτικό σφάλμα (π.χ. +++++ ή – – – –) με αυξανόμενο μέγεθος, όσο αυξάνει η περιεκτικότητα του δείγματος στο προσδιοριζόμενο συστατικό, ενώ το σχετικό σφάλμα διατηρείται σταθερό. Διαπιστώνεται εύκολα με πειράματα ανάκτησης, η οποία εμφανίζεται συστηματικά μεγαλύτερη ή μικρότερη του 100%.

Η τεχνική της γνωστής προσθήκης δεν επηρεάζεται από αυτό το αναλογικό σφάλμα και αποτελεί την πλέον ενδεδειγμένη τεχνική ποσοτικοποίησης, εάν δεν είναι δυνατή η εξάλειψη της αιτίας αυτού του σφάλματος και εφόσον το σφάλμα αυτό υπεισέρχεται στο τελικό στάδιο της μέτρησης (π.χ. κατά τη φωτομέτρηση του επεξεργασμένου δείγματος).

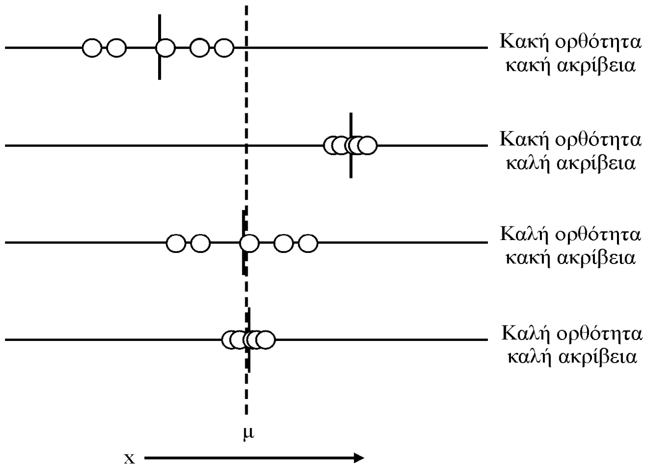
3. **Συνδυασμός σταθερού - αναλογικού σφάλματος** (combined error). Αποδίδεται από τις καμπύλες του διαγράμματος συσχέτισης (1-1Γ). Οδηγεί σε σφάλματα μονοκατευθυνόμενα ή με μία μόνο χαρακτηριστική αλλαγή προσήμου του σφάλματος (π.χ. +++++ ή +++ – – – –). Εάν υπεισέρχεται μόνο στο στάδιο της τελικής μέτρησης δεν μπορεί να αντιμετωπιστεί παρά μόνο με χρησιμοποίηση καμπύλης αναφοράς ως τεχνικής ποσοτικοποίησης.

4. **Περίπλοκο/μικτό σφάλμα**. Αποτελεί μερική περίπτωση του προηγούμενου τύπου και αποδίδεται από μη ευθύγραμμες καμπύλες συσχέτισης. Συνήθως οφείλεται στη χημεία της μέτρησης (π.χ. μικρές σταθερές ισορροπίας, βραδεία αποκατάσταση ισορροπίας) και σε οργανολογικά προβλήματα. Τα σφάλματα μπορεί να είναι μονοκατευθυνόμενα ή ομαδοποιούνται σε περιορισμένο αριθμό ομάδων με το ίδιο πρόσημο (π.χ. +++ ––––++). Το τελευταίο αυτό χαρακτηριστικό καθιστά το είδος αυτό του σφάλματος ως το πλέον "ύπουλο", αφού μπορεί (συμπτωματικά) να υπάρχουν περιοχές πρακτικά μηδενικού σφάλματος, ή μίας "καθησυχαστικής" εναλλαγής των προσήμων του σφάλματος. Εάν σφάλμα αυτού του είδους υπεισέρχεται μόνο στο στάδιο της μέτρησης, δεν μπορεί να αντιμετωπιστεί παρά μόνο με χρησιμοποίηση καμπύλης αναφοράς ως τεχνικής ποσοτικοποίησης, όπου τα πρότυπα θα πρέπει να έχουν υπόβαθρο περιεχόμενο παραπλήσιο των δειγμάτων.

1.4. Τυχαία σφάλματα

Μετρήσεις με αποκλειστικώς τυχαία σφάλματα χαρακτηρίζονται από τυχαιότητα στη συχνότητα και αλληλουχία του πρόσημου του σφάλματος (π.χ. ++–+ —+–++–+). Τα μικρά (απολύτως) σφάλματα είναι περισσότερο πιθανά από τα μεγάλα και γενικά χαρακτηρίζονται από κάποιο μέτρο της **διασποράς** (dispersion ή scattering) των λαμβανόμενων τιμών (π.χ. εύρος τιμών, τυπική απόκλιση). Το μέτρο διασποράς επιλέγεται ανάλογα με το είδος της στατιστικής κατανομής τους.

Εξ ορισμού τα τυχαία σφάλματα έχουν ως πραγματική μέση τιμή το μηδέν και ως εκ τούτου αποτελεσματική μείωσή τους μπορεί να επιτευχθεί με αύξηση του αριθμού των μετρήσεων. Τούτο βέβαια δεν είναι πάντοτε εφικτό λαμβανομένων υπόψη παραγόντων, όπως η ποσότητα του διατιθέμενου δείγματος, το κόστος, ο χρόνος και ο φόρτος εργασίας. Σε αντίθεση με τα τυχαία σφάλματα, τα συστηματικά σφάλματα δεν μειώνονται με αύξηση του αριθμού μετρήσεων.



Σχήμα 1-2. *Σημειογράμματα αξιολόγησης της ορθότητας και ακρίβειας μετρήσεων.*

1.5. Αξιολόγηση μετρήσεων

Οι επόμενοι όροι χαρακτηρίζουν μία αναλυτική τεχνική, αλλά και τη γενικότερη απόδοση ενός εργαστηρίου, όπως και την προσωπική ικανότητα του αναλυτή:

1. **Ορθότητα** (accuracy). Είναι μέτρο της εγγύτητας της μετρούμενης τιμής προς την πραγματική και αποδίδεται αποκλειστικά από το σφάλμα (Εξίσωση 1-1) ή τη μέση τιμή των απόλυτων τιμών του σφάλματος (ή του σχετικού σφάλματος) σε περίπτωση γενικής αξιολόγησης πολλών μετρήσεων σε πολλά δείγματα.

2. **Ακρίβεια** (precision). Είναι το μέτρο της διασποράς μιας σειράς μετρήσεων από **μέτρηση σε μέτρηση** και μέσα στην **ίδια σειρά μετρήσεων** (within-run precision) και εκφράζεται με κάποιο στατιστικό μέτρο της διασποράς (π.χ. με την τυπική απόκλιση).

Όροι που σχετίζονται με την ακρίβεια είναι η **επαναληψιμότητα** (repeatability) και η **αναπαραγωγιμότητα**. (reproducibility). Γενικά, η επαναληψιμότητα αποτελεί μέτρο της διασποράς ομοειδών μετρήσεων που εκτελούνται στο ίδιο εργαστήριο και από το ίδιο προσωπικό, ενώ η αναπαραγωγιμότητα αποτελεί ανάλογο μέτρο της διασποράς ομοειδών μετρήσεων, που έχουν ληφθεί σε διάφορα εργαστήρια.

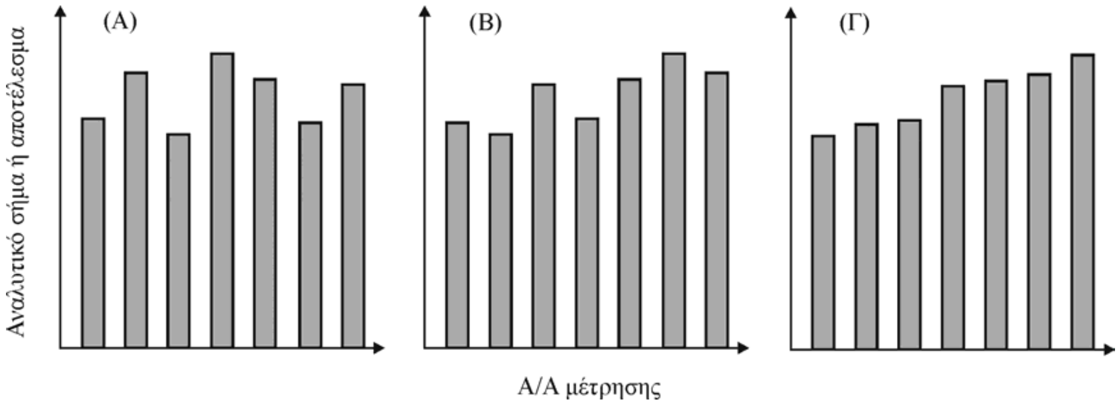
Στα σημειογράμματα του Σχήματος 1-2 δείχνονται τυπικές περιπτώσεις συνδυασμών ορθότητας και ακρίβειας.

3. **Τάση** (trend). Εκφράζει τον τρόπο κατά τον οποίο εξαρτάται το αποτέλεσμα της μέτρησης ενός μεγέθους από την προηγούμενη μέτρηση του ίδιου μεγέθους (στο ίδιο δείγμα) και επομένως εμπεριέχει τη διάσταση του χρόνου. Ιδανικά δεν πρέπει να υπάρχει καμία εξάρτηση, άσχετα από το είδος του ενυπάρχοντος σφάλματος (συστηματικού ή τυχαίου). Ή δεν θα πρέπει να υπάρχει διαφορά (μηδενικό ή αποκλειστικά συστηματικό σφάλμα) ή το εκάστοτε αποτέλεσμα θα πρέπει να εμφανίζεται τυχαία μεγαλύτερο ή μικρότερο από το προηγούμενο. Εάν διαπιστωθεί συστηματική εμφάνιση συνεχώς αυξανόμενων ή μειούμενων τιμών, τότε υπάρχει τάση (ή ολίσθηση) τιμών (Σχήμα 1-3).

Η τάση συχνά παραβλέπεται ως παράγοντας αξιολόγησης σειρά μετρήσεων και δεν ελέγχεται η ύπαρξή της. Παρ' όλα αυτά μπορεί να ανιχνευθεί με σχετικά απλές στατιστικές δοκιμασίες[1], οι οποίες επιβάλλεται να προηγηθούν από κάθε άλλη στατιστική δοκιμασία αξιολόγησης της ορθότητας και της ακρίβειας της εξεταζόμενης μεθόδου.

---

[1] 'Ενας απλός τρόπος ανίχνευσης τάσης βασίζεται στην εξέταση της τυπικής απόκλισης (s_a) της κλίσης a της ευθείας που περιγράφει το αποτέλεσμα κάθε μέτρησης (Y) ως συνάρτηση του αύξοντα αριθμού της (i), όπως λαμβάνεται με τη μέθοδο των ελαχίστων τετραγώνων ($Y = a\,i + b$). Εάν η κλίση a βρεθεί **σημαντικά** διαφορετική από το **μηδέν** (π.χ. με τη δοκιμασία Student's t), τότε θεωρείται ότι υπάρχει τάση στο δεδομένο επίπεδο εμπιστοσύνης.



Σχήμα 1-3. *Τάση μετρήσεων: (Α) απουσία συγκεκριμένης τάσης, (Β) ενδείξεις αυξητικής τάσης, (Γ) σαφής ύπαρξη αυξητικής τάσης.*

Η ύπαρξη τάσης (γενικά) είναι "ό,τι χειρότερο" μπορεί να συμβεί σε μία μέτρηση, αφού φέρει την όλη διαδικασία **εκτός ελέγχου** (out of control). Η οποιαδήποτε αξιολόγηση ως προς την ορθότητα ή ακρίβεια, οσωνδήποτε μετρήσεων με τάση, δεν έχει νόημα. Η τάση καθιστά αδύνατο τον υπολογισμό οποιουδήποτε στατιστικού στοιχείου (ακόμη και της μέσης τίμης) και διορθώνεται μόνο με δραστικό έλεγχο της χημείας (π.χ. αστάθεια κάποιου αντιδραστηρίου) και των χρησιμοποιούμενων οργάνων (π.χ. σωρευτική μόλυνση ή ρύπανση ενός ανιχνευτή, θερμική αστάθεια).

Η τεχνική του **εσωτερικού προτύπου** (internal standard), ως τεχνική ποσοτικοποίησης, πολλές φορές διορθώνει τα αποτελέσματα και εφαρμόζεται, όπου είναι δυνατόν (ιδιαίτερα στις χρωματογραφικές τεχνικές). Τυπικά η τεχνική αυτή εφαρμόζεται, εφόσον έχουμε τη δυνατότητα σύγχρονης μέτρησης δύο αναλυτικών σημάτων, δηλ. της μετρούμενης ουσίας και του εσωτερικού προτύπου.

1.6. Μερικοί ορισμοί κατά ISO 5725

Ο Διεθνής Οργανισμός Προτυποποίησης (International Organization for Standardization) στην έκδοσή του 5725 με τίτλο “Accuracy (Trueness and Precision) of measurement methods and results” καθιερώνει ένα πρότυπο τρόπο προσδιορισμού της ορθότητας και ακρίβειας νέων επίσημων αναλυτικών μεθόδων ελέγχου, μέσω **διεργαστηριακών** (interlaboratory) αναλύσεων προτύπων δειγμάτων. Στην πρώτη ενότητα (5725-1) του προτύπου αυτού παρέχονται ορισμένοι βασικοί ορισμοί ως οι ακόλουθοι:

- **Παρατηρούμενη τιμή** (observed value): η τιμή του χαρακτηριστικού που λαμβάνεται ως αποτέλεσμα μιας παρατήρησης.

- **Αποτέλεσμα δοκιμασίας** (test result): η τιμή ενός χαρακτηριστικού που λαμβάνεται με εκτέλεση μιας καθορισμένης μεθόδου δοκιμασίας[2].

- **Επίπεδο δοκιμασίας σε πείραμα επαναληψιμότητας** (level of the test in a precision experiment): Η **γενική μέση τιμή** (general average) των αποτελεσμάτων δοκιμασίας από όλα τα εργαστήρια για ένα συγκεκριμένο δοκιμαζόμενο υλικό ή δείγμα.

- **Στοιχείο σε πείραμα επαναληψιμότητας** (cell in a precision experiment): Τα αποτελέσματα δοκιμασίας σε απλό επίπεδο δοκιμασίας που λαμβάνονται από ένα εργαστήριο.

[2] *[Σημείωση ISO]* Η μέθοδος δοκιμασίας πρέπει να καθορίζει τον αριθμό των πραγματοποιούμενων παρατηρήσεων και τον τρόπο με το οποίο από τις παρατηρήσεις προκύπτει και πώς παρουσιάζεται το αποτέλεσμα (με λήψη μέσης τιμής, διάμεσης τιμής, υπολογισμό τυπικής απόκλισης κ.λπ.). Μπορεί επιπλέον να απαιτούνται κάποιες τυπικές διορθώσεις (π.χ. ως προς πίεση και θερμοκρασία σε περιπτώσεις τιμών όγκου αερίων). Η τιμή μπορεί να υπολογισθεί από κάποιο αριθμό παρατηρήσεων και στην απλούστερη περίπτωση μόνο από μία.

• **Αποδεκτή τιμή αναφοράς** (accepted reference value): Μια τιμή που εξυπηρετεί ως "εκ συμφωνίας" αναφορά για σύγκριση και η οποία προκύπτει ως:

α) Θεωρητική ή καθιερωμένη τιμή με βάση επιστημονικές αρχές.
β) **Αποδιδόμενη** (assigned) ή **διακριβωμένη** (certified) τιμή με βάση πειραματική εργασία ενός εθνικού ή διεθνούς οργανισμού.
γ) **Εκ συγκατάθεσης** (consensus) ή διακριβωμένη τιμή βασιζόμενη σε πειραματική συνεργασία υπό την γενική επίβλεψη μιας επιστημονικής ή τεχνικής ομάδας.
δ) Σε περίπτωση που δεν υπάρχουν τα α), β) και γ), η **αναμενόμενη τιμή** (expected value) της (μετρούμενης) ποσότητας, δηλ. η μέση τιμή ενός καθορισμένου πληθυσμού μετρήσεων.

• **Ορθότητα** (accuracy): Η **εγγύτητα** (closeness) της συμφωνίας μεταξύ του αποτελέσματος δοκιμασίας και της αποδεκτής τιμής αναφοράς[3].

• **Αληθότητα** (trueness): Η εγγύτητα της συμφωνίας μεταξύ της μέσης τιμής που λαμβάνεται από μια μεγάλη σειρά αποτελεσμάτων δοκιμασίας και της αποδεκτής τιμής αναφοράς[4].

• **Συστηματικό σφάλμα** (bias): Η διαφορά μεταξύ της αναμενόμενης τιμής μιας σειράς αποτελεσμάτων δοκιμασίας και της αποδεκτής τιμής αναφοράς[5, 6].

• **Συστηματικό σφάλμα (ή μεροληψία) εργαστηρίου** (laboratory bias): Η διαφορά μεταξύ της αναμενόμενης τιμής μιας σειράς αποτελεσμάτων δοκιμασίας ενός συγκεκριμένου εργαστηρίου και της αποδεκτής τιμής αναφοράς.

• **Συστηματικό σφάλμα μεθόδου μέτρησης** (bias of the measurement method): Η διαφορά μεταξύ της αναμενόμενης τιμής μιας σειράς αποτελεσμάτων δοκιμασίας από όλα τα εργαστήρια που χρησιμοποιούν τη μέθοδο αυτή και της αποδεκτής τιμής αναφοράς.

• **Συστατικό συστηματικού σφάλματος εργαστηρίου** (laboratory component of bias): Η διαφορά μεταξύ του συστηματικού σφάλματος του εργαστηρίου και του συστηματικού σφάλματος μεθόδου μέτρησης.

• **Ακρίβεια** (precision): Η εγγύτητα της συμφωνίας μεταξύ των ανεξάρτητων αποτελεσμάτων δοκιμασίας που λαμβάνονται κάτω από αυστηρά 'επιβαλλόμενες' (stipulated) συνθήκες[7].

---

[3] *[Σημείωση ISO]* Ο όρος αυτός, όταν εφαρμόζεται σε μια **ομάδα** (set) αποτελεσμάτων δοκιμασίας, περιλαμβάνει συνδυασμό τυχαίων συστατικών και ένα κοινό συστηματικό σφάλμα ή συστατικό συστηματικού σφάλματος.

[4] *[Σημείωση ISO]*
α) Το συστηματικό σφάλμα αποτελεί συνήθως το μέτρο της ορθότητας.
β) Η αληθότητα έχει αναφερθεί και ως "ορθότητα της μέσης τιμής". Δεν συνιστάται η χρήση του όρου αυτού.

[5] *[Σημείωση ISO]* Το συστηματικό σφάλμα αποτελεί ουσιαστικά το "ολικό συστηματικό σφάλμα" σε αντίθεση με το τυχαίο σφάλμα. Μπορεί να υπάρχουν ένα ή περισσότερα συστατικά που συνεισφέρουν στο συστηματικό σφάλμα. Μία μεγαλύτερη συστηματική διαφορά από την αποδεκτή τιμή αναφοράς απεικονίζεται από μια αντίστοιχα μεγαλύτερη τιμή συστηματικού σφάλματος.

[6] *[Σημείωση]* Ο όρος bias στην Ελληνική βιβλιογραφία Αναλυτικής Χημείας αναφέρεται και ως "προκατάληψη" ή "μεροληψία" , που αποτελεί και μια πιο άμεση και καθημερινή απόδοση αυτής της Αγγλικής λέξης στην Ελληνική.

[7] *[Σημείωση ISO]*
α) Η ακρίβεια εξαρτάται από την κατανομή των τυχαίων λαθών και δεν σχετίζεται με την αληθή τιμή ή την καθορισμένη τιμή.
β) Η ακρίβεια μπορεί να εκφρασθεί και ως **ανακρίβεια** (imprecision) και υπολογίζεται ως η τυπική απόκλιση των αποτελεσμάτων δοκιμασίας. Όσο μικρότερη είναι η ακρίβεια, τόσο μεγαλύτερη είναι τυπική απόκλιση.
γ) Ο όρος "*ανεξάρτητα αποτελέσματα δοκιμασίας*" υποδηλώνει αποτελέσματα που ελήφθησαν κατά τρόπο που δεν επηρεάζεται από ένα προηγούμενο αποτέλεσμα πάνω στο ίδιο το δοκιμαζόμενο αντικείμενο. Τα μέτρα της ακρίβειας εξαρτώνται κρίσιμα από τις επιβαλλόμενες συνθήκες.

• **Επαναληψιμότητα** (repeatability): Η ακρίβεια κάτω από συνθήκες ενδοεργαστηριακής επαναληψιμότητας.

• **Συνθήκες επαναληψιμότητας** (repeatability conditions): Συνθήκες κάτω από τις οποίες λαμβάνονται ανεξάρτητα αποτελέσματα δοκιμασιών με την ίδια μέθοδο, επάνω στα ίδια δείγματα, στο ίδιο εργαστήριο, από τον ίδιο αναλυτή, με τα ίδια σκεύη και μεταξύ σύντομων χρονικών διαστημάτων.

• **Τυπική απόκλιση επαναληψιμότητας** (repeatability standard deviation): Η τυπική απόκλιση των αποτελεσμάτων δοκιμασίας που λαμβάνονται κάτω από συνθήκες επαναληψιμότητας[8].

• **Όριο επαναληψιμότητας** (repeatability limit): Τιμή ως προς την οποία η διαφορά δύο αποτελεσμάτων δοκιμασιών κάτω από συνθήκες επαναληψιμότητας αναμένεται να είναι ίση ή μικρότερη στο 95% των περιπτώσεων[9].

• **Αναπαραγωγιμότητα** (reproducibility): Ακρίβεια κάτω από συνθήκες αναπαραγωγιμότητας.

• **Συνθήκες αναπαραγωγιμότητας** (reproducibility conditions): Συνθήκες κάτω από τις οποίες λαμβάνονται αποτελέσματα δοκιμασιών με την ίδια μέθοδο, επάνω στα ίδια δείγματα σε διαφορετικά εργαστήρια, από διαφορετικούς αναλυτές, που χρησιμοποιούν διαφορετικά σκεύη.

• **Τυπική απόκλιση αναπαραγωγιμότητας** (reproducibility standard deviation): Η τυπική απόκλιση των αποτελεσμάτων δοκιμασίας που λαμβάνονται κάτω από συνθήκες αναπαραγωγιμότητας[10].

• **Όριο αναπαραγωγιμότητας** (reproducibility limit): Τιμή προς την οποία η διαφορά δύο αποτελέσματων δοκιμασιών κάτω από συνθήκες αναπαραγωγιμότητας αναμένεται να είναι ίση ή μικρότερη στο 95% των περιπτώσεων[11].

• **Έκτοπη τιμή ή μέτρηση** (outlier): Μέλος ομάδας τιμών το οποίο είναι ασυνεπές ως προς τα άλλα μέλη της ομάδας[12].

---

[8] *[Σημείωση ISO]*
α) Αποτελεί μέτρο της διασποράς κάτω από συνθήκες επαναληψιμότητας.
β) Αντίστοιχα θα μπορούσαν να ορισθούν οι όροι “*διακύμανση επαναληψιμότητας*” (repeatability variance) και “*συντελεστής διακύμανσης επαναληψιμότητας*” (repeatability coefficient of variation), ως μέτρα της διασποράς κάτω από συνθήκες ενδοεργαστηριακής επαναληψιμότητας.

[9] *[Σημείωση ISO]* Το χρησιμοποιούμενο σύμβολο είναι *r*.

[10] *[Σημείωση ISO]*
α) Αποτελεί μέτρο της διασποράς κάτω από συνθήκες αναπαραγωγιμότητας.
β) Αντίστοιχα θα μπορούσαν να ορισθούν οι όροι “διακύμανση αναπαραγωγιμότητας” (reproducibility variance) και “συντελεστής διακύμανσης αναπαραγωγιμότητας” (reproducibility coefficient of variation), ως μέτρα της διασποράς κάτω από συνθήκες αναπαραγωγιμότητας.

[11] *[Σημείωση ISO]* Το χρησιμοποιούμενο σύμβολο είναι *R*.

[12] *[Σημείωση ISO]* Στο ISO 5725-2 καθορίζονται οι στατιστικές δοκιμασίες και το επίπεδο σημαντικότητας που πρέπει να χρησιμοποιούνται για τον εντοπισμό εκτρόπων τιμών σε πειράματα ορθότητας και ακρίβειας.

Παρατήρηση επί της ορολογίας:

Μία απλή μέτρηση μπορεί να θεωρηθεί ότι παρέχεται από τη σχέση:

$$\underset{\text{(μετρούμενη τιμή)}}{x_i} = \underset{\text{(αποδεκτή τιμή)}}{\mu} + \underset{\text{(συστημ. σφάλμα, bias)}}{B} + \underset{\text{(τυχαίο σφάλμα)}}{e}$$

Αν έχουν εκτελεστεί πολλές μετρήσεις και έχει ληφθεί η μέση τιμή μετρήσεων $\bar{x}$ κάτω από συνθήκες επαναληψιμότητας, οι τιμές B και e αποκτούν πλέον μέτρα και αντικαθίστανται ως εξής:

$$\underset{\text{(μετρούμενη τιμή)}}{x_i} = \underset{\text{(αποδεκτή τιμή)}}{\mu} + \underset{\text{(συστημ. σφάλμα, bias)}}{(\bar{x} - \mu)} + \underset{\text{(τυχαίο σφάλμα)}}{(x_i - \bar{x})}$$

η τελευταία μπορεί να γραφεί ως εξής:

$$\underset{\text{(ολικό σφάλμα μίας μέτρησης)}}{x_i - \mu} = \underset{\text{(συστημ. σφάλμα, bias)}}{(\bar{x} - \mu)} + \underset{\text{(τυχαίο σφάλμα)}}{(x_i - \bar{x})}$$

τα παραπάνω αποτελούν μέτρα των εννοιών:

$$\underset{\text{(accuracy)}}{\text{Ορθότητα}} = \underset{\text{(trueness)}}{\text{Αληθότητα}} + \underset{\text{(precision)}}{\text{Ακρίβεια}}$$

2. ΚΑΤΑΝΟΜΕΣ - ΣΤΑΤΙΣΤΙΚΑ ΣΤΟΙΧΕΙΑ ΠΛΗΘΥΣΜΩΝ

2.1. Κατανομές

Η διαπίστωση ενός ενδεικτικού μεγέθους του τυχαίου σφάλματος μπορεί να πραγματοποιηθεί μόνο με πολλαπλές πάντοτε μετρήσεις ενός δείγματος ή (πληρέστερα) περισσότερων δειγμάτων διαφορετικών περιεκτικοτήτων, για να διαπιστωθεί η εξάρτηση των στατιστικών ιδιοτήτων του τυχαίου σφάλματος από τη δεδομένη αναλυτική περιοχή (τυπικά: στα 10%, 50% και 90% της συνολικής περιοχής).

Οι αριθμητικές τιμές πλήθους μετρήσεων μπορούν να καταταχθούν σε διαδοχικές **κλάσεις** (classes). Κάθε κλάση περιλαμβάνει τις μετρήσεις που βρίσκονται μεταξύ δύο οριακών τιμών, η διαφορά των οποίων Δx είναι η ίδια (κατά κανόνα) σε όλες τις κλάσεις. Οι μετρήσεις μπορούν να πραγματοποιούνται στο ίδιο δείγμα, όταν εξετάζεται η κατανομή των αποτελεσμάτων μιας μεθόδου ή σε πλήθος ομοειδών δειγμάτων, όταν εξετάζεται η κατανομή των τιμών της μετρούμενης παραμέτρου. Σε κάθε περίπτωση, για να έχει νόημα η επίπονη αυτή διαδικασία, πρέπει να εξασφαλίζεται η ομοιογένεια του πληθυσμού.

Το ποιος πρέπει να είναι ο αριθμός των κλάσεων (επομένως και το εύρος τους Δx) για ένα δεδομένο αριθμό μετρήσεων ρυθμίζεται από τον ερευνητή σύμφωνα με την πείρα του. Προφανώς, όσο μεγαλύτερος είναι ο αριθμός των κλάσεων, τόσο λιγότερες μετρήσεις θα περιλαμβάνονται σε κάθε μία εξ αυτών και ενδέχεται να μην υπάρχει ικανοποιητική απεικόνιση της κατανομής. Για μικρό αριθμό κλάσεων, ενώ θα υπάρχει ικανοποιητικός πλέον αριθμός μετρήσεων για κάθε κλάση, πάλι δεν θα υπάρχει ικανοποιητική απεικόνιση της κατανομής. Υπάρχει ένας εμπειρικός κανόνας για τον καλύτερο αριθμό κλάσεων (κανόνας του Sturges) σύμφωνα με τον οποίο ο άριστος αριθμός κλάσεων για N μετρήσεις είναι ο πλησιέστερος ακέραιος αριθμός προς τον αριθμό k, που παρέχεται από την εξίσωση:

$$\text{Κανόνας Sturges:} \quad k = 1 + 3{,}322\ \log_{10}N \tag{2-1}$$

Στον Πίνακα 2.1 δίνονται τιμές 50 διαδοχικών προσδιορισμών θειικών ιόντων σε δείγμα φυσικού ύδατος και στον Πίνακα 2.2 δείχνεται η κατάταξη τους σε διαφορετικές κλάσεις.

Το πλήθος των μετρήσεων που ανήκει σε κάθε κλάση ονομάζεται **συχνότητα** (frequency, f) της κλάσης, το δε πηλίκο της συχνότητας δια του συνολικού αριθμού των μετρήσεων (πληθυσμιακό κλάσμα) αποτελεί τη **σχετική συχνότητα** (relative frequency, f_R) της κλάσης. Η απεικόνιση του πληθυσμιακού κλάσματος ή **συχνότητας** (frequency, f) των τιμών ενός μεγάλου αριθμού (N) (ιδανικά άπειρων) μετρήσεων στο ίδιο δείγμα, σε n κλάσεις αποδίδει χονδρικά την **κατανομή** (distribution) των μετρήσεων με τη μορφή ιστογραμμάτων (Σχήμα 2-1A). Είναι προφανής η ισχύς των σχέσεων:

$$\sum_{i=1}^{n} f(x) = N \tag{2-2}$$

$$\sum_{i=1}^{n} f_R(x) = 1 \tag{2-3}$$

Όσο αυξάνει ο αριθμός των μετρήσεων N τείνει προς το άπειρο και το εύρος Δx των κλάσεων τείνει προς το μηδέν, το περίγραμμα του ιστογράμματος τείνει να αποκτήσει τη μορφή συνεχούς καμπύλης (Σχήμα 2-1B), οπότε η **διακριτή** (discrete) Εξίσωση 2-2 αποκτά τη **συνεχή** (continuous) μορφή της:

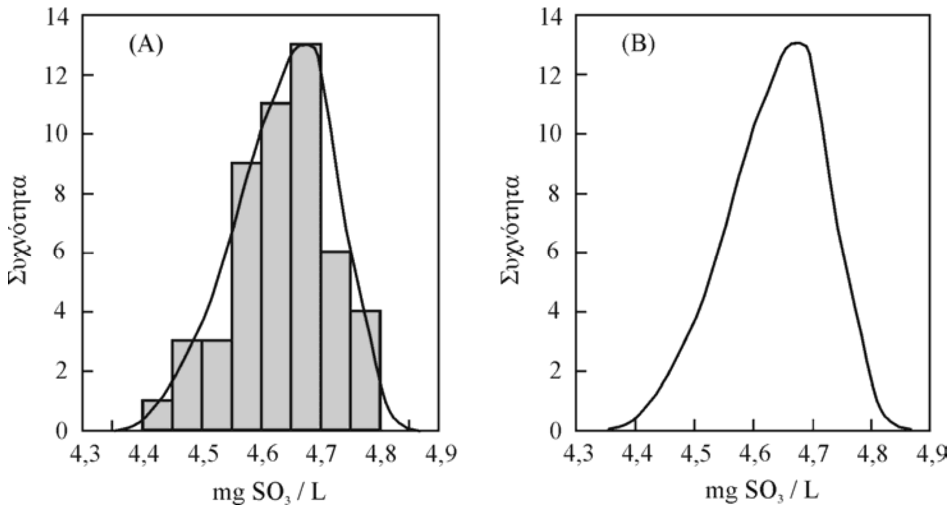
$$\int_{-\infty}^{+\infty} f(x)dx = 1 \qquad (\Delta x \to 0,\ N \to \infty) \tag{2-4}$$

Πίνακας 2.1. Αποτελέσματα 50 διαδοχικών μετρήσεων θειικών σε δείγμα φυσικού ύδατος με μία φωτομετρική μέθοδο.

Προσδιορισθείσα συγκέντρωση (mg SO_3/L)									
4,59	4,66	4,76	4,65	4,64	4,56	4,57	4,63	4,66	4,49
4,62	4,71	4,51	4,72	4,44	4,68	4,66	4,69	4,62	4,56
4,68	4,58	4,67	4,79	4,68	4,55	4,46	4,75	4,67	4,57
4,63	4,77	4,65	4,70	4,64	4,67	4,74	4,76	4,65	4,69
4,56	4,73	4,48	4,63	4,73	4,65	4,58	4,70	4,59	4,54

Πίνακας 2.2. Κατάταξη σε κλάσεις των μετρήσεων του Πίνακα 2.1.

Περιγραφή κλάσης (mg SO_3/L)	Συχνότητα, f(x)	Σχετική συχνότητα, $f_R(x)$
$4,40 < x \leq 4,45$	1	0,02
$4,45 < x \leq 4,50$	3	0,06
$4,50 < x \leq 4,55$	3	0,06
$4,55 < x \leq 4,60$	9	0,18
$4,60 < x \leq 4,65$	11	0,22
$4,65 < x \leq 4,70$	13	0,26
$4,70 < x \leq 4,75$	6	0,12
$4,75 < x \leq 4,80$	4	0,08
$4,40 < x \leq 4,80$	50	1,00



Σχήμα 2-1. *(A): Πληθυσμιακή "διακριτή" κατανομή (ιστόγραμμα) του συνόλου των 50 μετρήσεων του Πίνακα 2.1. (B): Πιθανή μορφή της αντίστοιχης συνεχούς καμπύλης κατανομής.*

Η τιμή του ολοκληρώματος μεταξύ δύο ορίων (α, β) εκφράζει την πιθανότητα (P) του να βρεθεί τιμή x στην περιοχή αυτή ($P(\alpha \leq x \leq \beta)$). Η τιμή 1 (P=1 σημαίνει βεβαιότητα) στην περιοχή ολοκλήρωσης ($-\infty$, $+\infty$) εκφράζει την προφανή βεβαιότητα ανεύρεσης μιας τιμής x μεταξύ $-\infty$ και $+\infty$, δηλ. πάντοτε $P(-\infty \leq x \leq +\infty) = 1$.

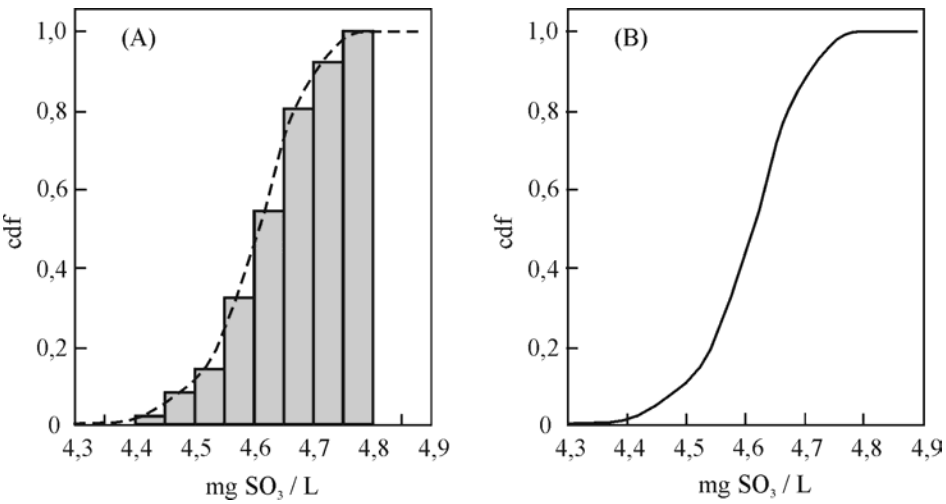
Η **καμπύλη κατανομής** (distribution curve), γενικά, αποτελεί την πληρέστερη απεικόνιση της στατιστικής κατανομής ενός πλήθους τιμών (π.χ. φυσιολογικών τιμών ενός συστατικού του ορού).

Εναλλακτικό τρόπο παρουσίασης της στατιστικής κατανομής ενός πλήθους τιμών αποτελούν τα διαγράμματα της συνάρτησης **σωρευτικής κατανομής** (cumulative distribution function, cdf). Η cdf (στη συνεχή μορφή της) ουσιαστικά αποδίδει το ολοκλήρωμα της συνάρτησης κατανομής $f_R(x)$ από την τιμή $-\infty$ έως μια τιμή α, δηλαδή ουσιαστικά είναι γραφική παράσταση της πιθανότητας $P(-\infty \le \alpha)$.

$$\text{cdf:}\ P(-\infty \le \alpha) = \sum_{x=-\infty}^{x=\alpha} f_R(x) \qquad (2\text{-}5)$$

$$\text{cdf:}\ P(-\infty \le \alpha) = \int_{-\infty}^{\alpha} f_R(x)\,dx \qquad (2\text{-}6)$$

Η cdf είναι πάντοτε μονότονη συνάρτηση (η τιμή της αυξάνει, όσο αυξάνει η τιμή α) με δυνατή περιοχή τιμών από 0 έως 1. Στο Σχήμα 2-2 δείχνονται οι καμπύλες cdf που αντιστοιχούν στις κατανομές του Σχήματος 2-1 στη διακριτή και στη συνεχή μορφή τους.



Σχήμα 2-2. *Καμπύλες cdf: (A) της διακριτής και (B) συνεχούς κατανομής του Σχήματος 2-1.*

Με εξέταση της μορφής της διάκριτης καμπύλης cdf ενός πεπερασμένου πλήθους τιμών διαπιστώνεται ο τύπος κατανομής (διαγράμματα cdf σε χάρτες "κανονικών πιθανοτήτων", δοκιμασία Kolmogorov - Smirnov κ.α.).

2.2. Χαρακτηριστικές παράμετροι κατανομών και πληθυσμιακών δειγμάτων

Επειδή μια οποιαδήποτε καμπύλη κατανομής δύσκολα μπορεί να αποδοθεί με τρόπο διαφορετικό από το ίδιο το σχήμα της, καταβάλλεται προσπάθεια να αποδοθεί με ένα περιορισμένο αριθμό χαρακτηριστικών αριθμητικών παραμέτρων.

Οι ακριβείς τιμές των χαρακτηριστικών παραμέτρων μιας κατανομής ενός πληθυσμού είναι επακριβώς γνωστές στο σημείο που είναι διαθέσιμος ολόκληρος ο πληθυσμός των στοιχείων που απαρτίζουν την κατανομή. Κάτι τέτοιο βέβαια δεν είναι δυνατόν. Στην πράξη οι οποιεσδήποτε στατιστικές μελέτες πραγματοποιούνται σε περιορισμένο αριθμό στοιχείων του πληθυσμού, γνωστό ως **πληθυσμιακό δείγμα** (population sample).

Όσο αυξάνει το μέγεθος (αριθμός στοιχείων, μετρήσεων κ.λπ.) του πληθυσμιακού δείγματος, τόσο ακριβέστερα μπορούν να εκτιμηθούν οι χαρακτηριστικές παράμετροι του πληθυσμού. Τα μεγέθη αυτά, ως υπολογίζονται από το πληθυσμιακό δείγμα, ονομάζονται γενικά **εκτιμήτριες** (estimators) παράμετροι των πληθυσμιακών κατανομών.

Οι κυριότερες χαρακτηριστικές παράμετροι των κατανομών, που παρέχονται από πληθυσμιακό δείγμα (x_1, x_2, ..., x_n) (π.χ. τα αναλυτικά αποτελέσματα μετρήσεων ενός συστατικού στο ίδιο προς ανάλυση δείγμα) χωρίζονται σε τρεις κατηγορίες:

1. Παράμετροι θέσης.
2. Παράμετροι διασποράς.
3. Παράμετροι συμμετρίας.

και περιγράφονται στη συνέχεια. Οι παράμετροι αυτές συλλογικά ονομάζονται **περιγραφικά στατιστικά στοιχεία** (descriptive statistics), εφόσον οι τιμές τους περιγράφουν κατά κάποιο τρόπο τη μορφή της κατανομής του δείγματος.

2.2.1. Παράμετροι θέσης

Οι παράμετροι θέσης, γνωστές και ως στατιστικά στοιχεία **κεντρικής τάσης** (central tendency) είναι οι παράμετροι εκείνες που σχετίζονται με τη θέση του κέντρου της κατανομής και εκφράζονται πάντοτε σε μονάδες της μετρούμενης ποσότητας.

1. **Μέση τιμή** (mean). Η μέση τιμή είναι η κυριότερη παράμετρος θέσης και παρέχεται από τη σχέση:

$$\overline{x} = (x_1 + x_2 + ... + x_n)/n = \sum_{i=1}^{n} x_i/n \qquad (2\text{-}7)$$

Εφόσον έχει αποκλειστεί η παρουσία συστηματικού σφάλματος, η μέση τιμή προσεγγίζει την πραγματική και το τυχαίο σφάλμα περιορίζεται, όσο αυξάνει το μέγεθος του δείγματος. Η μέση τιμή πληθυσμού ($n = \infty$) συμπίπτει με την πραγματική μ, δηλαδή ισχύει:

$$\overline{x} \rightarrow \mu, \quad \text{εάν} \quad n \rightarrow \infty \qquad (2\text{-}8)$$

Σε όλες τις πληθυσμιακές κατανομές, η κάθετος που φέρεται στη μέση τιμή χωρίζει την περιβαλλόμενη επιφάνεια σε δύο μέρη ίσου εμβαδού (σημείο Α, Σχήμα 2-1β).

2. **Διάμεση τιμή** (median). Μετά τη διάταξη των τιμών κατά αυξανόμενη τιμή ($x_1 < x_2 < ... < x_n$) (**διατεταγμένες τιμές**, ordered values), ως διάμεση (ή διχοτόμος) τιμή median(x_i) παρέχεται από τις σχέσεις:

$$\text{median}(x_i) = x_{(n+1)/2} \qquad (\text{n: περιττός}) \qquad (2\text{-}9\alpha)$$

$$\text{median}(x_i) = \frac{x_{n/2} + x_{n/2+1}}{2} \qquad (\text{n: άρτιος}) \qquad (2\text{-}9\beta)$$

Η διάμεση τιμή συχνά προτιμάται ως **ανθεκτικότερο στατιστικό στοιχείο** (robust statistic ή robust estimator) σε σχέση με τη μέση τιμή, ιδιαίτερα σε περιπτώσεις μικρών πληθυσμιακών δειγμάτων, διότι δεν επηρεάζεται από την παρουσία ισχυρώς αποκλινουσών ή **εκτόπων μετρήσεων** (outliers) στο πληθυσμιακό δείγμα.

Η μόνη δυσκολία στον υπολογισμό της διάμεσης τιμής είναι ο αλγοριθμικός χαρακτήρας του προσδιορισμού της, δηλ. είναι αποτέλεσμα αλληλουχίας διαδικασιών ως: διάταξη τιμών κατά αυξανόμενη τιμή, ανεύρεση των μεσαίων στοιχείων και επιλογή της Εξίσωσης 2-9α ή της Εξίσωσης 2-9β) και δεν υπολογίζεται από μια συγκεκριμένη εξίσωση (όπως π.χ. η μέση τιμή).

3. **Ρυθμισμένη μέση τιμή** (trimmed mean value). Πολλές φορές, για να αποφευχθεί η επίδραση ισχυρώς αποκλεινουσών μετρήσεων στη μέση τιμή, απορρίπτονται οι ακραίες τιμές μιας διατεταγμένης σειράς μετρήσεων (δηλ. οι μικρότερες και οι μεγαλύτερες τιμές σε ίσο αριθμό) και υπολογίζεται η μέση τιμή των υπολοίπων.

Ως **εκατοστιαίο ποσοστό ρύθμισης** (trimming percentage) ορίζεται η ποσότητα 100r/n, όπου r ο αριθμός των απορριπτόμενων μετρήσεων και n ο αρχικός αριθμός μετρήσεων. Τα συνήθη ποσοστά ρύθμισης είναι 10 έως 25%.

4. **Επικρατούσα τιμή** (mode, MO). Αντιστοιχεί στην τιμή που εμφανίζει τη μεγαλύτερη συχνότητα εμφάνισης (τιμή κορυφής). Δεν συμπίπτει κατ' ανάγκη με τη μέση ή διάμεση τιμή (σημείο Β, Σχήμα 2-1Β). Είναι σύνηθες το φαινόμενο κατανομών με περισσότερες από μία κορυφές (πολυκόρυφη κατανομή), π.χ. με δύο κορυφές, (**δικόρυφη κατανομή**, bimodal distribution) κατά την εξέταση της κατανομής των τιμών (όχι σφαλμάτων) δειγμάτων κλινικού ενδιαφέροντος. Τούτο συνήθως είναι ένδειξη ανομοιογένειας του πληθυσμιακού δείγματος από το οποίο δεν θα πρέπει να αναμένονται στατιστικώς χρήσιμα αποτελέσματα, δεδομένου ότι η ανάμιξη πληθυσμιακών δειγμάτων καταργεί την οποιαδήποτε αξία των στατιστικών δοκιμασιών.

Εάν η κατανομή είναι διακριτή (δηλ. πληθυσμιακά στοιχεία είναι διατεταγμένα σε τάξεις), μπορεί να αναδειχθεί μόνο η επικρατούσα κλάση (εκείνη με τη μέγιστη συχνότητα). Στην περίπτωση αυτή η επικρατούσα τιμή ορίζεται από τη σχέση:

$$MO = L + c\frac{L_1}{L_1 + L_2} \qquad (2\text{-}10)$$

όπου L είναι η μικρότερη τιμή της επικρατούσας κλάσης, L_1 και L_2 οι απόλυτες τιμές των διαφορών των συχνοτήτων της επικρατούσας κλάσης από την προηγούμενη και την επόμενη, αντιστοίχως, κλάση και c το εύρος της επικρατούσας κλάσης.

2.2.2. Παράμετροι διασποράς

1. **Εύρος** ή **περιοχή τιμών** (range, R). Στη διατεταγμένη ομάδα τιμών ($x_1 < x_2 < ... < x_n$) αποδίδεται από τη διαφορά:

$$R = x_n - x_1 \qquad (2\text{-}11)$$

Το εύρος τιμών έχει περιορισμένη στατιστική αξία ως ελάχιστα ανθεκτικό στατιστικό στοιχείο, εφόσον είναι προφανές ότι επηρεάζεται δραματικά από την παρουσία εκτόπων μετρήσεων.

2. **Τυπική απόκλιση** (standard deviation, s). Παρέχεται από την εξίσωση:

$$s = \sqrt{\frac{\sum_{i=1}^{n}\left(\bar{x} - x_i\right)^2}{n-1}} \qquad (2\text{-}12)$$

και αποτελεί το συνηθέστερο τρόπο έκφρασης της διασποράς των τιμών και επομένως της ακρίβειας μίας μέτρησης. Η s αποτελεί εκτιμήτρια της **τυπικής απόκλισης πληθυσμού** (population standard deviation, σ) και ισχύει

$$s \to \sigma, \quad \text{εάν} \quad n \to \infty \qquad (2\text{-}13)$$

Η τυπική απόκλιση (πληθυσμιακού δείγματος) αποτελεί πολύτιμο στατιστικό στοιχείο, αφού πλήθος στατιστικών δοκιμασιών μπορεί να διευκολυνθεί από τη γνώση της τιμής της. Πρέπει να τονιστεί ότι οι στατιστικές δοκιμασίες αυτές προϋποθέτουν κανονικές κατανομές πληθυσμού **κάτι που συχνά παραβλέπεται**.

Μια ισοδύναμη αλλά πιο εύχρηστη μορφή της Εξίσωσης 2-12 είναι η ακόλουθη:

$$s = \sqrt{\frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}} \qquad (2\text{-}14)$$

Η τελευταία εξίσωση επιτρέπει τον υπολογισμό της τρέχουσας τυπικής απόκλισης πληθυσμιακού δείγματος (καθώς τούτο αυξάνει, π.χ. κατά τη συλλογή μετρήσεων μέσω ηλεκτρονικών υπολογιστών), χωρίς να είναι αναγκαίος ο προηγούμενος υπολογισμός της μέσης τιμής.

3. **Σχετική τυπική απόκλιση** (relative standard deviation, s_r), η οποία παρέχεται από την εξίσωση:

$$s_r = s / \bar{x} \qquad (2\text{-}15)$$

Συχνά η σχετική τυπική απόκλιση αναφέρεται και ως **συντελεστής μεταβλητότητας** (coefficient of variation, CV), αλλά η χρήση του όρου αυτού δεν συνιστάται πλέον.

4. **Σχετική τυπική απόκλιση επί τοις εκατό** (relative standard deviation per cent, RSD %), η οποία παρέχεται από την εξίσωση:

$$\text{RSD,\%} = (s / \bar{x}) \times 100 \qquad (2\text{-}16)$$

Η έκφραση αυτή της σχετικής τυπικής απόκλισης είναι αρκετά συνηθισμένη αλλά χρειάζεται ιδιαίτερη προσοχή, όταν οι ίδιες οι μετρήσεις εκφράζονται επί τοις εκατό και καλό θα ήταν γενικά να αποφεύγεται (προτιμότερο είναι να δίνεται απλά η τιμή s_r ως δεκαδικό κλάσμα και μόνο). Για παράδειγμα: Αναφέρεται ανάλυση ορυκτού σιδήρου με αποτέλεσμα: Fe: 15,20% και τυπική απόκλιση 0,25%, η σχετική τυπική απόκλιση είναι s_r = 0,25% / 15,20% = 0,0164 ή 1,64%. Το αποτέλεσμα μπορεί να δοθεί ως:

15,20% ± 0,25% (± τυπική απόκλιση)

ή

15,20% ± 1,64% (± σχετική τυπική απόκλιση επί τοις εκατό)

Είναι προφανές ότι αν δεν δοθούν οι εντός παρενθέσεων διευκρινίσεις τα συμπεράσματα για την επαναληψιμότητα των μετρήσεων μπορεί να είναι τελείως διάφορα. Ο μόνος ορθός τρόπος παρουσίασης του προηγούμενου αποτελέσματος, που δεν απαιτεί επιπλέον διευκρινίσεις είναι: (15,20±0,25) %.

5. **Διακύμανση** (variance). Η διακύμανση (γνωστή και ως **διασπορά** ή **μεταβλητότητα**) (τελεστής: Var) είναι ίση με το τετράγωνο της τυπικής απόκλισης ($v = s^2$ ή σ^2). Θεωρείται ως πρωτογενές στατιστικό στοιχείο σε σχέση με την τυπική απόκλιση. Σημασία έχει στη διάδοση σφαλμάτων, όπου η χρήση της διευκολύνει τους υπολογισμούς, εφόσον ισχύουν σχέσεις ως (A, B: δύο διαφορετικοί πληθυσμοί, όπως π.χ. αποτελέσματα μετρήσεων διαφορετικών μεγεθών, με αντίστοιχες διακυμάνσεις Var(A), Var(B)).

$$\text{Var}(A \pm B) = \text{Var}(A) + \text{Var}(B) \qquad (2\text{-}17\alpha)$$

$$[\text{Var}(A \cdot B)] / (A \cdot B) = [\text{Var}(A)]/A + [\text{Var}(B)]/B \qquad (2\text{-}17\beta)$$

6. **Μέση απόλυτη απόκλιση** (absolute mean deviation, amd). Παρέχεται από την εξίσωση:

$$\text{amd} = \frac{\sum_{i=1}^{n} \left| \bar{x} - x_i \right|}{n} \qquad (2\text{-}18)$$

και είναι ανθεκτικότερο στατιστικό στοιχείο και μέτρο της διασποράς, σε σχέση με την τυπική απόκλιση (με περιορισμένη όμως χρήση για περαιτέρω δοκιμασίες), επηρεαζόμενο λιγότερο από την παρουσία εκτόπων μετρήσεων.

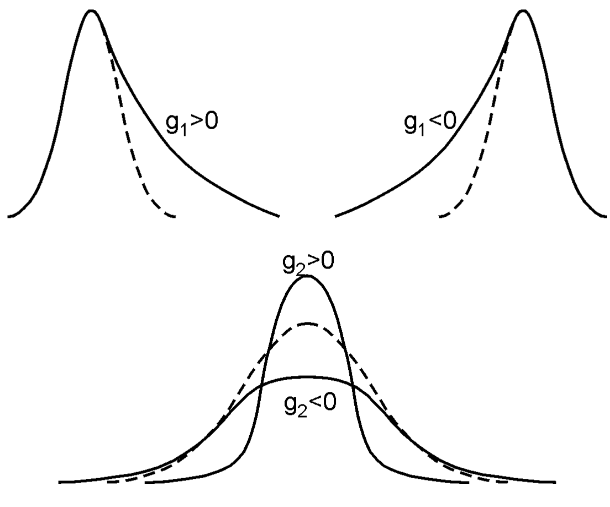
7. **Τυπικό σφάλμα** (standard error, S.E.). Αποτελεί μέτρο της διασποράς της μέσης τιμής από πληθυσμιακό δείγμα σε πληθυσμιακό δείγμα που έχει ληφθεί από την ίδια κατανομή. Ουσιαστικά πρόκειται για την τυπική απόκλιση της μέσης τιμής (όπως και συχνά αναφέρεται: standard deviation of the mean) και παρέχεται από την εξίσωση:

$$\mathrm{S.E.} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n(n-1)}} = \frac{s}{\sqrt{n}} \tag{2-19}$$

8. **Εύρος μεταξύ ποσοστημοριακών σημείων** (interpercentile range). Είναι ίσως ο καταλληλότερος τρόπος έκφρασης της διασποράς των τιμών σε περιπτώσεις πληθυσμών με άγνωστη κατανομή και εκφράζει την περιοχή τιμών που χωρίζουν δύο ποσοστά του διατεταγμενου συνόλου των μετρήσεων. Γενικά, εάν το στοιχείο του δείγματος με τιμή x_α χωρίζει το διατεταγμένο σύνολο των τιμών σε ομάδες που περιέχουν αντίστοιχα το 100α% και 100(1–α)% του συνόλου των στοιχείων και η τιμή $x_{1-\alpha}$ χωρίζει το διατεταγμένο σύνολο των μετρήσεων αντίστοιχα σε 100(1–α)% και 100α%, τότε η διαφορά $x_{1-\alpha} - x_\alpha$ αποδίδει το εύρος μεταξύ των 100α% και 100(1–α)% σημείων.

Ουσιαστικά το εύρος μεταξύ ποσοστημοριακών σημείων, αντιστοιχεί στο εύρος τιμών του πληθυσμιακού δείγματος, αφού απορριφθούν τα πρώτα $\alpha \times n$ και τα τελευταία $\alpha \times n$ από τα διατεταγμένα σημεία. Είναι προφανές ότι το εύρος μεταξύ αυτών των σημείων, ως εκτιμήτρια ποσότητα είναι αρκετά ανθεκτικό, εφόσον αποκλείονται οι έκτοπες τιμές (outliers).

Συνηθέστερη μορφή είναι η **ενδοτεταρτημοριακή περιοχή** (interquartile range), η οποία αντιστοιχεί στην περιοχή μεταξύ των διατεταγμένων τιμών που αντιστοιχούν στο 25% και 75% του συνολικού πληθυσμού. Χρησιμοποιείται κυρίως για την περιγραφή διασποράς πλήθους τιμών διαφορετικών αναλυτικών δειγμάτων και όχι τόσο στην περιγραφή της διασποράς μετρήσεων ενός δείγματος για το χαρακτηρισμό τυχαίου σφάλματος.



Σχήμα 2-3. Τυπικές περιπτώσεις κατανομών με διαφορετικές τιμές παραμέτρων ασυμμετρίας (με διακεκομμένες γραμμές δείχνεται συγκριτικά καμπύλη κανονικής κατανομής).

2.2.3. Παράμετροι συμμετρίας

Οι παράμετροι συμμετρίας έχουν περιορισμένη μόνο χρησιμότητα και στην ουσία παρέχουν πληροφορίες ενδεικτικές του κατά πόσο η παρατηρούμενη κατανομή πλησιάζει ή όχι την κανονική κατανομή.

1. **Ασυμμετρία** (skewness, g_1). Η παράμετρος αυτή αποτελεί μέτρο της συμμετρίας της κατανομής του πληθυσμιακού δείγματος και της απόκλισής της από την κανονική κατανομή. Η ασυμμετρία παρέχεται από την εξίσωση:

$$g_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{(n-1)(n-2)s^3} \tag{2-20}$$

Θετική τιμή ασυμμετρίας υποδηλώνει ασυμμετρία ("ουρά" ή "άπλωμα" σε σχέση με ότι θα αναμενόταν σε περίπτωση κανονικής κατανομής) προς τα δεξιά, αρνητική υποδηλώνει ασυμμετρία προς τα αριστερά, ενώ μηδενική τιμή υποδηλώνει συμμετρική κατανομή. Ενδεικτικά, για δείγματα μεγέθους 20 και 90 στοιχείων, τιμές g_1 μέχρι 0,8 και 0,4, αντιστοίχως, υποδηλώνουν κανονική κατανομή (σε στάθμη εμπιστοσύνης 95%).

2. **Κύρτωση** (kurtosis, g_2). Η παράμετρος αυτή αποτελεί μέτρο της επιμήκυνσης ή διαπλάτυνσης της κορυφής της κατανομής σε σχέση με κορυφή της κανονικής κατανομής και παρέχεται από την εξίσωση:

$$g_2 = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{x_i - \bar{x}}{s}\right]^4 - 3 \tag{2-21}$$

Πίνακας 2.3. Περιγραφικά στατιστικά πληθυσμιακού δείγματος Πίνακα 2.1.

Στατιστικό στοιχείο	Τιμή
Γενικά στοιχεία	
Μέγεθος δείγματος (size)	50
Παράμετροι θέσης	
Μικρότερη τιμή (minimum)	4,4400
Μεγαλύτερη τιμή (maximum)	4,7900
Μέση τιμή (mean)	4,6382
Διάμεση τιμή (median)	4,6500
Επικρατούσα τιμή (mode)	4,6611
Παράμετροι διασποράς	
Εύρος τιμών (range)	0,3500
Τυπική απόκλιση (standard deviation)	0,0829
Σχετική τυπική απόκλιση (relative stand. deviation)	0,0179
Σχετική τυπική απόκλιση % (rel. stand. deviation %)	1,79
Διακύμανση (variance)	0,0069
Μέση απόλυτη απόκλιση (mean absol. deviation)	0,0657
Τυπικό σφάλμα (standard error)	0,0117
Περιοχή μεταξύ τεταρτημορίων (interquartile range)	0,115
Παράμετροι συμμετρίας	
Ασυμμετρία (skewness)	–0,4358
Κύρτωση (kurtosis)	2,7768

Αρνητικές και θετικές τιμές της παραμέτρου αυτής υποδηλώνουν σχετικά πεπλατυσμένες και σχετικά οξύτερες κορυφές, αντιστοίχως. Μηδενική τιμή υποδηλώνει κορυφή αντίστοιχη εκείνης της κανονικής κατανομής. Ενδεικτικά, για δείγματα μεγέθους 20 και 100 στοιχείων, τιμές g_2 μέχρι 1,18 και 0,77, αντιστοίχως, υποδηλώνουν κορυφή η οποία δεν αποκλίνει από εκείνη της κανονικής κατανομής (σε στάθμη εμπιστοσύνης 95%).

Στον Πίνακα 2.3 δίνονται τα κυριότερα περιγραφικά στατιστικά στοιχεία που αφορούν το πληθυσμιακό δείγμα του παραδείγματος κατανομής (στοιχεία Πίνακα 2.1).

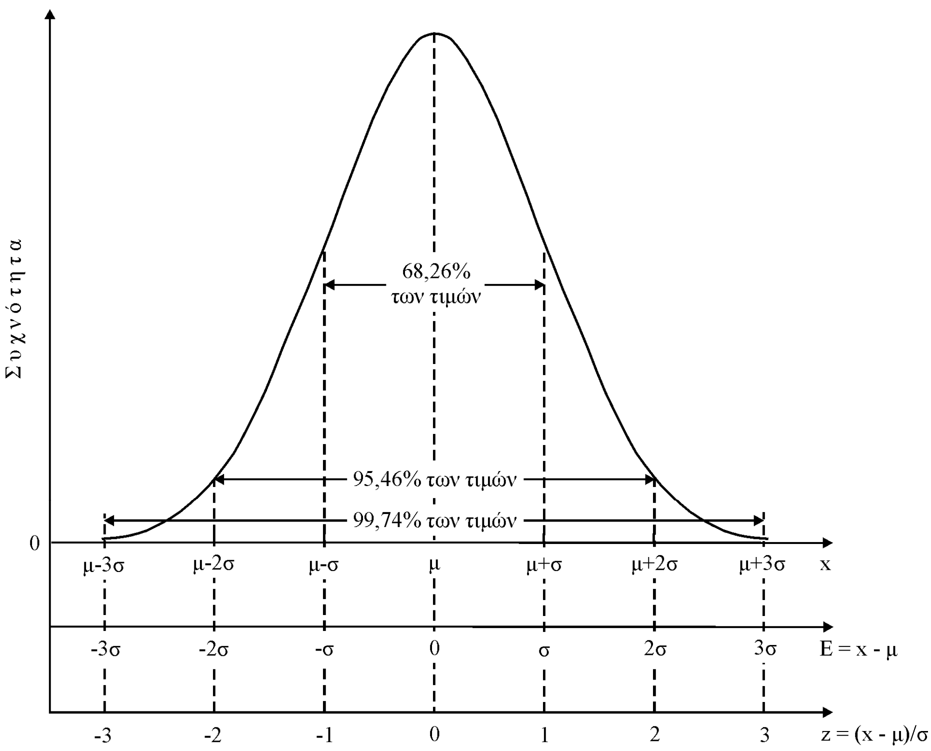
2.3. Κανονική κατανομή

Η **κανονική κατανομή** (normal distribution) ή **κατανομή κατά Gauss** (Gaussian distribution) αποτελεί τον πλέον συνήθη τύπο κατανομής και αποδίδεται από την καμπύλη του Σχήματος 2-4. Ισχύει η εξίσωση

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad (2\text{-}22)$$

όπου σ είναι η **τυπική απόκλιση** (standard deviation) του πληθυσμού και ο συντελεστής $1/\sigma\sqrt{2\pi}$ δρα ως συντελεστής κανονικοποίησης, ο οποίος εξασφαλίζει εμβαδόν κάτω από την καμπύλη (ολοκλήρωμα της f(x) από $-\infty$ έως $+\infty$) ίσο προς τη μονάδα, ως επιβάλλεται από την Εξίσωση 2-3.

Χαρακτηριστική είναι η συμμετρία της κατανομής αυτής και συνεπώς η σύμπτωση της μέσης τιμής (μ) με την επικρατέστερη. Στην περίπτωση πληθυσμών τυχαίων αναλυτικών σφαλμάτων θεωρείται ως η επικρατέστερη και πλέον πιθανή κατανομή.



Σχήμα 2-4. *Τυπική καμπύλη κανονικής κατανομής (καμπύλη Gauss) και διάφορα χαρακτηριστικά σημεία της.*

Σε ένα κανονικό πληθυσμό τυπικές χαρακτηριστικές περιοχές της κατανομής είναι οι εξής:

$$P(\mu-\sigma \leq x \leq \mu+\sigma) \quad ή \quad P(-1 \leq z \leq +1) = 0{,}683$$

$$P(\mu-2\sigma \leq x \leq \mu+2\sigma) \quad ή \quad P(-2 \leq z \leq +2) = 0{,}955$$

$$P(\mu-3\sigma \leq x \leq \mu+3\sigma) \quad ή \quad P(-3 \leq z \leq +3) = 0{,}997$$

δηλαδή η πιθανότητα μιας μέτρησης να βρεθεί στην περιοχή $\mu-\sigma$ έως $\mu+\sigma$ είναι περίπου 68%, στην περιοχή $\mu-2\sigma$ έως $\mu+2\sigma$ περίπου 95% (ή -ισοδύναμα-μία στις 20 μετρήσεις αναμένεται να αποκλίνει απολύτως περισσότερο από 2σ), κ.ο.κ. Πληρέστεροι πίνακες παρέχουν τις τιμές της πιθανότητας $P(\mu-z\sigma \leq x \leq \mu+z\sigma)$ για διάφορες τιμές του z. Τυπικό παράδειγμα αποτελεί ο Πίνακας 2.4, όπου παρέχονται οι πιθανότητες με ακρίβεια τεσσάρων δεκαδικών ψηφίων για τιμές z που καλύπτουν την περιοχή 0,00 - 3,09 σε βήματα 0,01.

Πολλές δοκιμασίες μπορούν να πραγματοποιηθούν με βάση τα παραπάνω ή τα δεδομένα των πληρέστερων πινάκων. Π.χ. τιμές που αποκλείνουν από τη μέση τιμή περισσότερο από 2σ μπορούν να θεωρηθούν ως **έκτοπες** (outliers) και να απορριφθούν ως προερχόμενες από άλλο πληθυσμό, διακινδυνεύοντας όμως πιθανότητα 5% να απορριφθεί τιμή προερχόμενη από τον αρχικό πληθυσμό ("νόμιμη" τιμή).

Πίνακας 2.4. Κανονική (Gaussian) κατανομή. Τιμές του ολοκληρώματος:

$$P = \frac{1}{\sqrt{2\pi}} \int_{-z}^{+z} e^{-t^2/2} dt, \qquad όπου \qquad z = \frac{x-\mu}{\sigma}$$

z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0080	0,0160	0,0239	0,0319	0,0399	0,0478	0,0558	0,0638	0,0717
0,1	0,0797	0,0876	0,0955	0,1034	0,1113	0,1192	0,1271	0,1350	0,1428	0,1507
0,2	0,1585	0,1663	0,1753	0,1819	0,1897	0,1974	0,2051	0,2127	0,2205	0,2282
0,3	0,2358	0,2434	0,2510	0,2586	0,2661	0,2737	0,2812	0,2886	0,2961	0,3035
0,4	0,3108	0,3182	0,3255	0,3328	0,3401	0,3473	0,3545	0,3616	0,3688	0,3759
0,5	0,3829	0,3899	0,3969	0,4039	0,4177	0,4108	0,4245	0,4313	0,4381	0,4448
0,6	0,4515	0,4581	0,4647	0,4713	0,4778	0,4843	0,4907	0,4971	0,5035	0,5098
0,7	0,5161	0,5223	0,5285	0,5346	0,5407	0,5467	0,5527	0,5587	0,5646	0,5705
0,8	0,5763	0,5821	0,5878	0,5935	0,5991	0,6047	0,6102	0,6157	0,6210	0,6265
0,9	0,6319	0,6372	0,6424	0,6476	0,6528	0,6579	0,6629	0,6680	0,6729	0,6779
1,0	0,6827	0,6875	0,6923	0,6970	0,7017	0,7063	0,7109	0,7154	0,7199	0,7243
1,1	0,7287	0,7330	0,7373	0,7415	0,7457	0,7499	0,7540	0,7580	0,7620	0,7660
1,2	0,7699	0,7737	0,7775	0,7813	0,7850	0,7887	0,7923	0,7959	0,7995	0,8029
1,3	0,8064	0,8098	0,8132	0,8165	0,8198	0,8230	0,8262	0,8293	0,8324	0,8355
1,4	0,8385	0,8415	0,8444	0,8473	0,8501	0,8529	0,8557	0,8584	0,8611	0,8638
1,5	0,8664	0,8690	0,8715	0,8740	0,8764	0,8789	0,8812	0,8836	0,8859	0,8882
1,6	0,8904	0,8926	0,8926	0,8990	0,9011	0,9031	0,9051	0,9051	0,9070	0,9090
1,7	0,9109	0,9127	0,9146	0,9164	0,9181	0,9199	0,9216	0,9233	0,9249	0,9265
1,8	0,9281	0,9297	0,9312	0,9327	0,9342	0,9357	0,9371	0,9385	0,9399	0,9412
1,9	0,9426	0,9439	0,9451	0,9464	0,9476	0,9488	0,9500	0,9512	0,9523	0,9534
2,0	0,9545	0,9556	0,9566	0,9576	0,9586	0,9596	0,9606	0,9615	0,9625	0,9633
2,1	0,9643	0,9651	0,9660	0,9668	0,9676	0,9684	0,9692	0,9700	0,9707	0,9715
2,2	0,9722	0,9729	0,9736	0,9743	0,9749	0,9756	0,9762	0,9768	0,9774	0,9780
2,3	0,9786	0,9791	0,9797	0,9802	0,9807	0,9812	0,9817	0,9822	0,9827	0,9832
2,4	0,9836	0,9840	0,9845	0,9849	0,9853	0,9857	0,9861	0,9865	0,9869	0,9872
2,5	0,9876	0,9879	0,9883	0,9886	0,9889	0,9892	0,9895	0,9898	0,9901	0,9904
2,6	0,9907	0,9909	0,9912	0,9915	0,9917	0,9920	0,9922	0,9924	0,9926	0,9929
2,7	0,9931	0,9933	0,9935	0,9937	0,9939	0,9940	0,9942	0,9944	0,9946	0,9947
2,8	0,9949	0,9950	0,9952	0,9953	0,9955	0,9956	0,9958	0,9959	0,9960	0,9961
2,9	0,9963	0,9964	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972
3,0	0,9973	0,9974	0,9975	0,9976	0,9976	0,9977	0,9978	0,9979	0,9979	0,9980

Παράδειγμα 2-1. Μέθοδος προσδιορισμού συστατικού X σε δείγμα, απαλλαγμένη από συστηματικά σφάλματα παρουσιάζει τυπική απόκλιση 2,0 mg/g δείγματος. Δείγμα περιέχει 40,0 mg X / g. Ποια η πιθανότητα ανάλυση του δείγματος να έχει ως αποτέλεσμα μια τιμή στις περιοχές: α) 39,0 - 41,0 mg X / g και β) 41,5 - 42,0 mg X / g.

Λύση. Σε όλες τις περιπτώσεις η περιοχή ανάγεται σε αδιάστατες μονάδες τυπικής απόκλισης z ($z = (x - \mu)/\sigma$) και υπολογίζεται (από τις τιμές του Πίνακα 2.4) το εμβαδόν (ολοκλήρωμα) της καμπύλης Gauss που αντιστοιχεί στην εκάστοτε περιοχή:

α) Οι τιμές 39,0 και 41,0 mg X / g αντιστοιχούν στις τιμές z:
$z_1 = (39{,}0 - 40{,}0) / 2{,}0 = -0{,}50$ και $z_2 = (41{,}0 - 40{,}0) / 2{,}0 = +0{,}50$. Εφόσον η περιοχή είναι συμμετρική ως προς την κεντρική τιμή ($|z_1| = |z_2|$) η πιθανότητα παρέχεται από την τιμή P για z = 0,50 και είναι 0,3829. Επομένως η πιθανότητα (%) να βρεθεί η τιμή στην αναφερόμενη περιοχή είναι **38,3%**.

β) Οι τιμές 41,5 και 42,0 mg X / g αντιστοιχούν στις τιμές z:
$z_1 = (41{,}5 - 40{,}0) / 2{,}0 = 0{,}75$ και $z_2 = (42{,}0 - 40{,}0) / 2{,}0 = 1{,}00$. Από τον Πίνακα 2.4 υπολογίζεται πιθανότητα για να βρεθεί η τιμή στην περιοχή +0,75z έως –0,75z ($P_{|z1|}$) είναι 0,5467. Ανάλογα και για να βρεθεί η τιμή στην περιοχή +1,00z έως –1,00z ($P_{|z2|}$) είναι 0,6827. Η διαφορά $(P_{|z2|}) - (P_{|z1|}) = 0{,}6827 - 0{,}5467 = 0{,}1360$ παρέχει την πιθανότητα να βρεθεί η τιμή οπουδήποτε στις περιοχές –1,00z έως –0,75z ή +0,75z έως +1,00z. Επειδή ενδιαφερόμαστε μόνο για την μία εξ αυτών και λόγω της συμμετρικότητας της κανονικής καμπύλης κατανομής η ζητούμενη πιθανότητα θα είναι η μισή της προηγουμένης δηλαδή 0,1360 / 2 = 0,068 ή **6,8%**.

Θεώρημα κεντρικού ορίου. Χαρακτηριστική και μάλλον απρόσμενη ιδιότητα της κανονικής κατανομής είναι το ότι αποτελεί την οριακή κατανομή των μέσων τιμών πολλαπλών μετρήσεων, προερχόμενων από πληθυσμούς με οποιαδήποτε κατανομή. Η ιδιότητα αυτή εκφράζεται από το ακόλουθο σημαντικότατο θεώρημα, που είναι γνωστό ως **θεώρημα κεντρικού ορίου** (central limit theorem):

ΘΕΩΡΗΜΑ ΚΕΝΤΡΙΚΟΥ ΟΡΙΟΥ
Το άθροισμα τιμών προερχομένων από ένα ή περισσότερους πληθυσμούς με οποιαδήποτε κατανομή τείνει να αποτελέσει στοιχείο ενός πληθυσμού με κανονική κατανομή, όσο αυξάνει ο αριθμός των αθροιζόμενων τιμών.

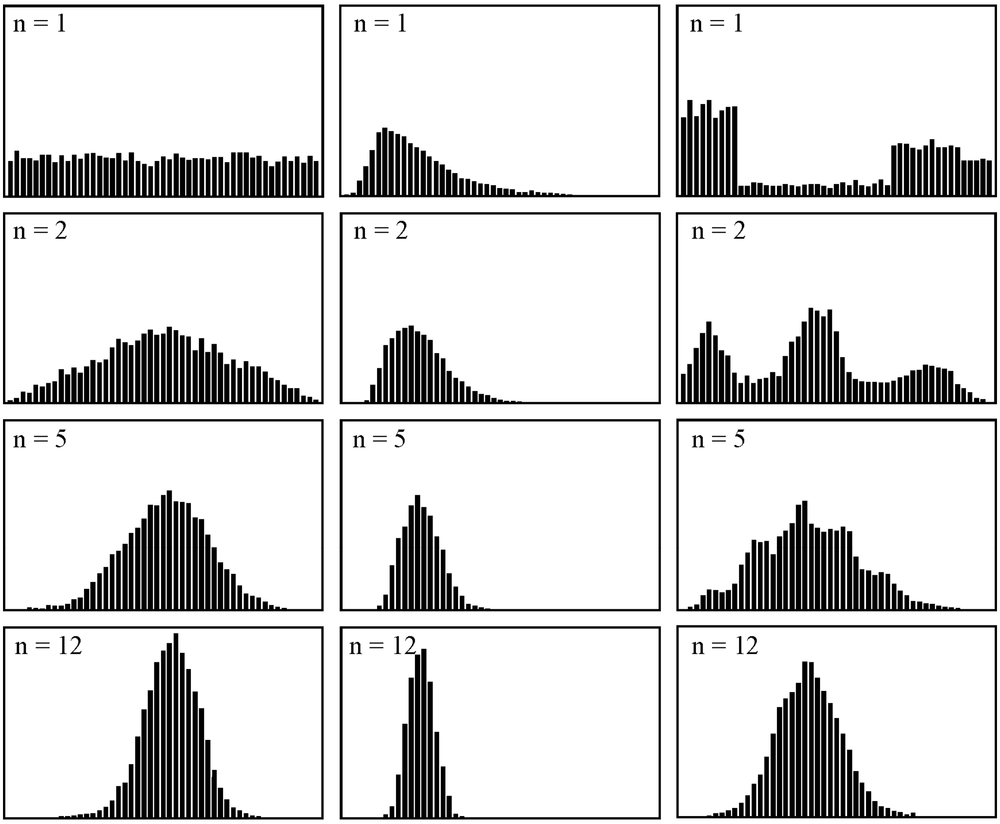
Στο Σχήμα 2-5 δείχνονται τα ιστογράμματα τριών κατανομών διαφορετικού τύπου (Α: ομοιογενής κατανομή, Β: λογαριθμοκανονική, Γ: ασυνεχούς τύπου -κατά τόπους ορισμένη-) και τα ιστογράμματα των κατανομών των μέσων τιμών 3, 5 και 12 αρχικών τιμών της κάθε κατανομής. Όλα τα ιστογράμματα έχουν κατασκευαστεί από ένα πλήθος 5000 ψευδοτυχαίων τιμών, οι οποίες δημιουργήθηκαν με τη βοήθεια μικροϋπολογιστή. Είναι προφανές ότι ανεξάρτητα από το σχήμα της κατανομής των αρχικών (απλών) τιμών, η κατανομή των μέσων τιμών σε όλες τις περιπτώσεις τείνει προς την κανονική.

Δύο επιπτώσεις του θεωρήματος του κεντρικού ορίου στη χημική ανάλυση είναι οι εξής:

1) Όταν το αποτέλεσμα είναι η μέση (ή διάμεση) τιμή n μετρήσεων, τότε τα αποτελέσματα (ως τιμές) τείνουν να αποκτήσουν (όσο η τιμή του n αυξάνει) κανονική κατανομή ανεξάρτητα από το είδος της κατανομής του πληθυσμού των απλών μετρήσεων.

2) Εάν στις απλές μετρήσεις το τυχαίο του σφάλματος είναι συνέπεια διάδοσης και άθροισης καθ' οιονδήποτε τρόπο πολλών τυχαίων σφαλμάτων (π.χ. στη φωτομετρική ανάλυση: σφάλμα ζύγισης προτύπου, σφάλμα λήψης όγκου αγνώστου, σφάλματα αραίωσης προτύπων, σφάλμα αραίωσης αγνώστου, ανοχές στην τοποθέτηση κυψελίδας, και όσο αυξάνει ο αριθμός των

πιθανών πηγών τυχαίου σφάλματος, τόσο περισσότερο το *τελικό αποτέλεσμα* τείνει να αποτελέσει στοιχείο πληθυσμού με κανονική κατανομή.



Σχήμα 2-5. *Ιστογράμματα 5000 τιμών απλών (n=1) και μέσων τιμών 2, 5 και 12 μετρήσεων από πληθυσμούς τριών διαφορετικών κατανομών.*

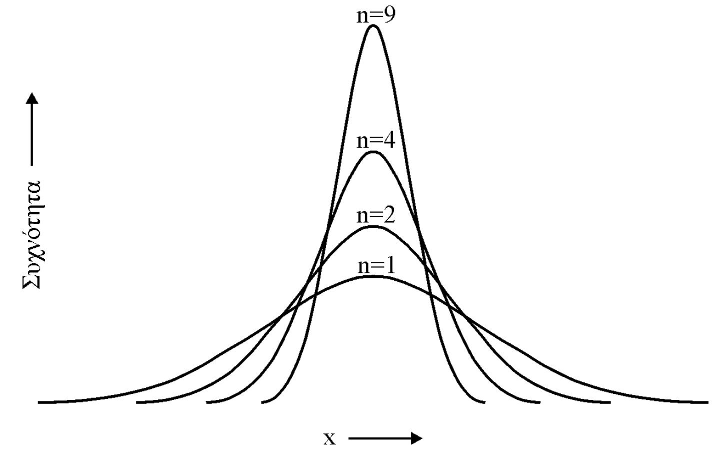
2.4. Κατανομή μέσων τιμών

Εφόσον το θεώρημα κεντρικού ορίου ισχύει για κατανομές οποιουδήποτε τύπου, πολύ περισσότερο θα ισχύει και για κατανομές που είναι ήδη κανονικές: η κατανομή της μέσης τιμής δείγματος πληθυσμού κανονικής κατανομής θα είναι επίσης κανονική, χωρίς να είναι προϋπόθεση η άθροιση μεγάλου αριθμού στοιχείων, όπως στις κατανομές τυχαίας μορφής. Έτσι, η μέση τιμή n τιμών προερχόμενων από κανονικό πληθυσμό αποτελεί στοιχείο ενός άλλου κανονικού πληθυσμού με τυπική απόκλιση που παρέχεται από τη σχέση

$$\sigma_{\bar{x}} = \sigma_x / \sqrt{n} \qquad (2\text{-}23)$$

Έτσι, για να υποδιπλασιαστεί, να υποτριπλασιαστεί κ.ο.κ. η διασπορά των μετρήσεων και επομένως το τυχαίο σφάλμα θα πρέπει αντιστοίχως να τετραπλασιαστεί, εννεαπλασιαστεί, κ.ο.κ. ο αριθμός των μετρήσεων, κάτι που βέβαια δεν είναι πάντοτε εφικτό.

Στο Σχήμα 2-6 δείχνονται οι κατανομές του αρχικού κανονικού πληθυσμού (n=1) και οι κατανομές πληθυσμού μέσων τιμών 2, 4 και 9 στοιχείων του αρχικού πληθυσμού. Είναι χαρακτηριστική η μείωση του εύρους της κατανομής, όσο αυξάνει ο αριθμός των αθροιζόμενων τιμών για τον υπολογισμό της μέσης τιμής.



Σχήμα 2-6. *Κατανομές κανονικού πληθυσμού (n=1) και των μέσων τιμών 2, 4 και 9 τιμών στοιχείων του.*

Παράδειγμα 2-2. Να υπολογιστεί η πιθανότητα του να βρεθεί η μέση τιμή 3 μετρήσεων του συστατικού X (Παράδειγμα 2-1) στην περιοχή 39,5 - 40,5 mg X/g δείγματος.

Λύση. Η τυπική απόκλιση των απλών μετρήσεων είναι 2,0 mg X/g. Η τυπική απόκλιση του νέου πληθυσμού (μέσες τιμές 3 μετρήσεων) είναι: $2{,}0/\sqrt{3} = 1{,}387$ mg X/g, επομένως η ζητούμενη περιοχή (σε αδιάστατες μονάδες τυπικής απόκλισης z) είναι: (39,5– 40,0)/1,387 = –0,36 έως (40,5–40,0)/1,387 = +0,36 και από τον Πίνακα 2.4 προκύπτει ότι για z = 0,36 η ζητούμενη πιθανότητα είναι P = 0,2812 ή **28,1%**.

Παράδειγμα 2-3. Να υπολογιστεί ο απαιτούμενος αριθμός n απλών μετρήσεων του συστατικού X (Παράδειγμα 2-1) έτσι, ώστε να υπάρχει πιθανότητα 97% η μέση τιμή των μετρήσεων αυτών να βρεθεί στην περιοχή 39,0 - 41,0 mg X/g δείγματος.

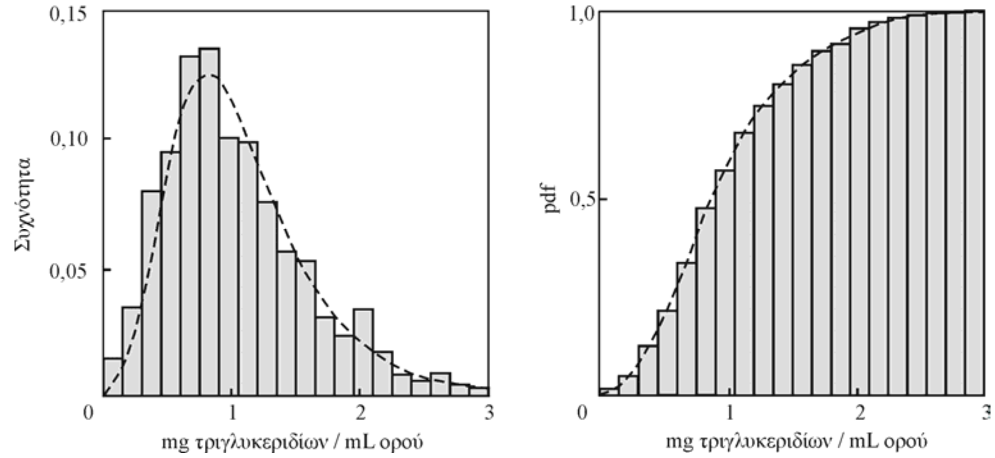
Λύση. Η περιοχή 39,0-41,0 mg X/g είναι συμμετρική ως προς την κεντρική (ορθή) τιμή των 40,0 mg X/g. Από τον Πίνακα 2-4 προκύπτει ότι το 97% του εμβαδού της καμπύλης Gauss οριοθετείται από την περιοχή ±2,17z. Επομένως, εάν η τυπική απόκλιση της μέσης τιμής των n μετρήσεων είναι $\sigma_{\bar{x}}$, τότε $1{,}0 = 2{,}17\sigma_{\bar{x}}$ ή $\sigma_{\bar{x}} = 0{,}4608$ mg X/g και από την Εξίσωση 2-23 έχουμε: $n = (\sigma / \sigma_{\bar{x}})^2 = (2{,}0 / 0{,}4608)^2 = 18{,}8$, επομένως απαιτούνται 19 μετρήσεις.

2.5. Λογαριθμοκανονική κατανομή

Στους **λογαριθμοκανονικούς** (log-normal) πληθυσμούς τιμών x_1, x_2, ... οι λογάριθμοι των τιμών αυτών ($\log x_1$, $\log x_2$, ...) αποτελούν στοιχεία κανονικών πληθυσμών. Η λογαριθμοκανονική κατανομή αποτελεί το πλέον τυπικό παράδειγμα ασύμμετρης κατανομής.

Η λογαριθμοκανονική κατανομή, ενώ είναι σχετικά σπάνια ως κατανομή τυχαίων σφαλμάτων, είναι συνήθης ως κατανομή των τιμών συγκεντρώσεων συστατικών βιολογικών δειγμάτων ομογενών πληθυσμών (Σχήμα 2-7).

Λογαριθμοκανονικά δεδομένα κανονικοποιούνται με λογαρίθμηση και μπορούν στη συνέχεια να υποστούν την ίδια στατιστική επεξεργασία με τα διάφορα κριτήρια σημαντικότητας που ισχύουν στις περιπτώσεις κανονικών πληθυσμών.



Σχήμα 2-7. *Λογαριθμοκανονική κατανομή δείγματος 500 τιμών τριγλυκεριδίων στον ορό και το αντίστοιχο διάγραμμα cdf.*

3. ΣΤΑΤΙΣΤΙΚΕΣ ΔΟΚΙΜΑΣΙΕΣ ΣΗΜΑΝΤΙΚΟΤΗΤΑΣ

Τα πειραματικά δεδομένα σπάνια συμφωνούν απόλυτα με τα αναμενόμενα, τα οποία βασίζονται κάποιο θεωρητικό μοντέλο. Επιπροσθέτως, κατά τη σύγκριση δύο ομάδων μετρήσεων της ίδιας ποσότητας, κατά κανόνα, οι τιμές δεν βρίσκονται σε συμφωνία. Στις περιπτώσεις αυτές, η αριθμητική διαφορά μεταξύ των τιμών μπορεί να αποδοθεί, είτε σε συστηματικά σφάλματα, οπότε η διαφορά κρίνεται ως **σημαντική** (significant), ή σε τυχαία και μόνο σφάλματα. Ο έλεγχος της σημαντικότητας των παρατηρούμενων διαφορών πραγματοποιείται με ειδικές στατιστικές δοκιμασίες γνωστές ως **δοκιμασίες σημαντικότητας** (significance tests)[13].

3.1. Μηδενική και εναλλακτική υπόθεση

Σε όλες τις στατιστικές δοκιμασίες εξετάζεται η πιθανότητα ισχύος κάποιας υπόθεσης. Για παράδειγμα, μια υπόθεση μπορεί να είναι ότι οι παρατηρούμενες διαφορές στα αποτελέσματα αναλύσεων δύο δειγμάτων είναι τυχαίες και στην πραγματικότητα τα δύο δείγματα είναι τα ίδια.

Η υπόθεση εκκίνησης ή βασική υπόθεση είναι η **μηδενική υπόθεση** (null hypothesis), που συμβολίζεται ως H_o. Το επίθετο "μηδενική" υποδηλώνει την ανυπαρξία πραγματικής (συστηματικής) διαφοράς μεταξύ των εξεταζομένων ποσοτήτων, π.χ., ότι οι διαφορές μεταξύ πραγματικής τιμής μ και μέσης τιμής μετρήσεων $\bar{x}$, μεταξύ δύο μέσων τιμών μετρήσεων $\bar{x}_A$ και $\bar{x}_B$, μεταξύ δύο τυπικών αποκλίσεων s_A και s_B πληθυσμιακών δειγμάτων Α και Β ή μετρήσεων του ίδιου πληθυσμιακού δείγματος με δύο διαφορετικές μεθόδους Α και Β, είναι καθαρά τυχαίες.

Εάν η μηδενική υπόθεση θεωρηθεί ότι δεν ισχύει, τότε ισχύει η συμπληρωματική της υπόθεση, γνωστότερη ως **εναλλακτική υπόθεση** (alternative hypothesis), που συμβολίζεται ως H_a. Τονίζεται ο συμπληρωματικός χαρακτήρας των υποθέσεων, δηλαδή δεν υφίσταται άλλη (τρίτη) υπόθεση και ισχύει πάντοτε η εξίσωση:

---

[13] Τονίζεται με ιδιαίτερη έμφαση ότι το επίθετο **σημαντικός** (significant) στη στατιστική δεν έχει την καθημερινή έννοια του "μεγάλου", "υπέρμετρου", "σοβαρού" κ.λπ., αλλά την έννοια της **ιδιαίτερης σημασίας**. Έτσι, π.χ., **σημαντική διαφορά** σε κάποια αναλυτικά αποτελέσματα δεν σημαίνει "μεγάλη διαφορά", αλλά διαφορά που σε καθορισμένα πλαίσια πιθανοτήτων (επίπεδο εμπιστοσύνης), οφείλεται σε παρουσία συστηματικού σφάλματος, διαφορετικά δείγματα, διαφορετική επιδεξιότητα των αναλυτών ή άλλους λόγους, όχι όμως σε τυχαίες και μόνο διαφορές.

$$P(H_o) + P(H_a) = 1 \qquad (3\text{-}1)$$

όπου $P(H_o)$ και $P(H_a)$ είναι η πιθανότητες ισχύος της μηδενικής και εναλλακτικής υπόθεσης, αντίστοιχα. Η Εξίσωση 3-1 δηλώνει ως βεβαιότητα (P=1) το ότι θα ισχύει πάντοτε η μία ή άλλη υπόθεση και τίποτα το ενδιάμεσο.

Τυπικά παραδείγματα μηδενικής και της αντίστοιχης εναλλακτικής υπόθεσης των πλέον συνηθισμένων δοκιμασιών σημαντικότητας δίνονται στον Πίνακα 3.1.

3.2. Σφάλματα 1ου και 2ου είδους

Το αποτέλεσμα μιας στατιστικής δοκιμασίας μπορεί να είναι ορθό (να ανταποκρίνεται στην πραγματικότητα) ή όχι. Μπορεί να θεωρηθεί ότι υπάρχουν δύο τύποι σφαλμάτων στα συμπεράσματα μιας στατιστικής δοκιμασίας. Τα σφάλματα αυτά συνδέονται άμεσα με την αποδοχή ή όχι της μηδενικής υπόθεσης:

Πίνακας 3.1. Τυπικές περιπτώσεις μηδενικής και εναλλακτικής υπόθεσης

Ερώτημα	Μηδενική υπόθεση H_o	Εναλλακτική υπόθεση H_a
Διαφέρει η μέση τιμή μιας σειράς μετρήσεων από την αποδεκτή τιμή ;	$\bar{x} = \mu$	$\bar{x} \neq \mu$
Διαφέρουν οι μέσες τιμές δύο σειρών μετρήσεων Α και Β ;	$\mu_A = \mu_B$ *	$\mu_A \neq \mu_B$ *
Διαφέρουν οι τυπικές αποκλίσεις δύο σειρών μετρήσεων Α και Β ;	$\sigma_A = \sigma_B$ *	$\sigma_A \neq \sigma_B$ *
Μία απέχουσα τιμή x ανήκει στον πληθυσμό Α(μ, σ) ή είναι τιμή που ανήκει σε άλλο πληθυσμό;	$x \in A(\mu,\sigma)$	$x \notin A(\mu,\sigma)$

* Χρησιμοποιούνται χαρακτηριστικά πληθυσμού (μ, σ) και όχι πληθυσμιακού δείγματος (x, s), διότι τα συμπεράσματα των δοκιμασιών αφορούν πλέον τους ίδιους τους πληθυσμούς.

Σφάλμα 1ου είδους (error of 1st kind ή error α):
Απόρριψη της μηδενικής υπόθεσης, ενώ αυτή ισχύει.

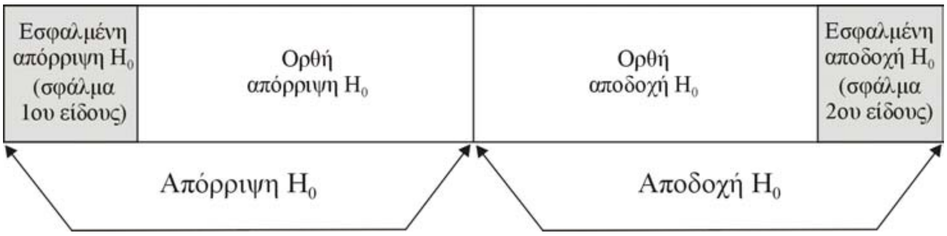
Σφάλμα 2ου είδους (error of 2nd kind ή error β):
Η μηδενική υπόθεση γίνεται αποδεκτή, ενώ αυτή δεν ισχύει.

Είναι προφανές ότι σε περίπτωση σφάλματος 1ου είδους δεν ισχύει η εναλλακτική υπόθεση, ενώ σε περίπτωση σφάλματος 2ου είδους ισχύει η εναλλακτική υπόθεση. Στη στατιστική που σχετίζεται με τεχνικοοικονομικά θέματα το σφάλμα 1ου είδους χαρακτηριστικά ονομάζεται "**κίνδυνος του παραγωγού**" (producer's risk), επειδή μπορεί π.χ. να θεωρηθεί ότι ο παραγωγός ζημιώνεται απορρίπτοντας παρτίδα παραγωγής δεχόμενος (εσφαλμένα) ότι αποκλίνει από τις προδιαγραφές. Αντίστοιχα, το σφάλμα 2ου είδους ονομάζεται "**κίνδυνος του καταναλωτή**" (consumer's risk), επειδή μπορεί π.χ. να θεωρηθεί ότι ο καταναλωτής ζημιώνεται εξαιτίας παρτίδας παραγωγής, που ενώ απέκλινε από τις προδιαγραφές (εσφαλμένα) πέρασε στην κατανάλωση.

Στην ιδανική περίπτωση οι πιθανότητες σφάλματος 1ου και 2ου είδους πρέπει να είναι όσο το δυνατόν μικρότερες. Μπορούν βέβαια να μηδενιστούν μόνο εάν εξεταστεί το σύνολο του πληθυσμού κάτι φυσικά αδύνατον. Στον Πίνακα 3.2 δείχνονται συγκριτικά οι γενικές περιπτώσεις εμφάνισης σφάλματος 1ου και 2ου είδους και στο Σχήμα 3-1 παρουσιάζεται γραφική παράσταση των πιθανών εκβάσεων μιας στατιστικής δοκιμασίας σημαντικότητας.

Πίνακας 3.2. Γενικές περιπτώσεις σφαλμάτων 1ου και 2ου είδους.

Συμπέρασμα στατιστικής δοκιμασίας	Πραγματική κατάσταση	
	Ισχύει η H_o	Ισχύει η H_a
Απόρριψη H_o	**Σφάλμα 1ου είδους**	Ορθό συμπέρασμα
Απόρριψη H_a	Ορθό συμπέρασμα	**Σφάλμα 2ου είδους**



Σχήμα 3-1. *Γραφική παράσταση των πιθανών εκβάσεων μιας στατιστικής δοκιμασίας σημαντικότητας.*

Για να αποφεύγει κανείς **πάντοτε** σφάλματα 1ου είδους, θα πρέπει μόνιμα να δέχεται ότι η μηδενική υπόθεση ισχύει, π.χ. πάντοτε θα αποκλείει τον χαρακτηρισμό τιμών ως εκτόπων και όλες τις απέχουσες μετρήσεις θα τις θεωρεί έγκυρες, πάντοτε θα θεωρεί ότι οι μέσες τιμές ομάδων μετρήσεων είναι ουσιαστικά ίδιες. Δηλαδή θα δέχεται πάντοτε ότι οι όποιες διακυμάνσεις και αποκλίσεις οφείλονται αποκλειστικά και μόνο σε τυχαία σφάλματα.

Αντίστοιχα, αν θέλει κανείς να αποφεύγει **πάντοτε** σφάλματα 2ου είδους, θα πρέπει μόνιμα να απορρίπτει τη μηδενική υπόθεση, π.χ. πάντοτε θα αποδίδει τις όποιες αποκλίσεις σε συστηματικά σφάλματα, θα θεωρεί εκ των προτέρων κάθε αποκλίνουσα μέτρηση ως έκτοπη εξωγενούς χαρακτήρα (δηλ. ως προερχόμενη από άλλο “ξένο” πληθυσμό). Δηλαδή θα δέχεται πάντοτε ότι οι όποιες διακυμάνσεις και αποκλίσεις οφείλονται μόνο συστηματικά σφάλματα.

Είναι προφανές ότι και οι δύο παραπάνω παραδοχές είναι **παράλογες** και ότι κάθε προσπάθεια μείωσης της πιθανότητας του ενός τύπου σφάλματος, αυξάνει την πιθανότητα σφάλματος του άλλου τύπου. Επομένως είναι αναγκαία η εκτέλεση των στατιστικών δοκιμασιών σε μια λογική και όχι ακραία στάθμη εμπιστοσύνης.

3.3. Στάθμη εμπιστοσύνης - Πιθανότητα σφάλματος

Όλες οι στατιστικές δοκιμασίες καταλήγουν σε συμπεράσματα που αναπόφευκτα δεν είναι πάντοτε αλάνθαστα. Τα συμπεράσματα αυτά μπορεί να ανταποκρίνονται στην πραγματικότητα, μπορεί και όχι. Σκοπός των δοκιμασιών αυτών είναι το να αποφευχθεί η εισαγωγή υποκειμενικού χαρακτήρα υποθέσεων και συμπερασμάτων (που συνήθως ξεκινούν με φράσεις όπως: “μου φαίνεται”, “μάλλον”, “νομίζω”, “κατά πάσα πιθανότητα”) κατά την εξέταση και σύγκριση πειραματικών δεδομένων, αλλά αυτές να γίνονται κατά αντικειμενικό, ενιαίο και επιστημονικά αποδεκτό τρόπο.

Ως **στάθμη** (ή **επίπεδο**) **εμπιστοσύνης** (confidence level, CL) μιας στατιστικής δοκιμασίας ονομάζεται η ελάχιστη πιθανότητα αποφυγής σφάλματος 1ου είδους. Η πλέον συνηθισμένη στάθμη εμπιστοσύνης που χρησιμοποιείται τουλάχιστον σε θέματα χημικής ανάλυσης είναι το 0,95 (ή 95%). Σε κάθε περίπτωση, η στάθμη εμπιστοσύνης μιας δοκιμασίας **πρέπει να δηλώνεται**.

Εναλλακτικά χρησιμοποιείται και ο όρος **πιθανότητα σφάλματος** (error probability, P), που υποδηλώνει την πιθανότητα σφάλματος 1ου είδους. Έτσι, στάθμη εμπιστοσύνης 0,95 (ή 95%) υποδηλώνει μέγιστη πιθανότητα (εννοείται πάντοτε: σφάλματος 1ου είδους): 1–0,95 = 0,05 (ή 5%) [14].

[14] Ένα συνηθισμένο εκφραστικό λάθος (που συχνά οδηγεί και σε νοηματικό λάθος) στη θεώρηση της στατιστικής ανάλυσης σημαντικότητας των αναλυτικών αποτελεσμάτων, η οποία διεξάγεται σε καθο-

Στάθμη εμπιστοσύνης 95% (ή ισοδύναμα: μέγιστη πιθανότητα σφάλματος 5%) σημαίνει αποδοχή του ότι κατά μέσο όρο 1 απόρριψη της μηδενικής υπόθεσης στις 20 είναι εσφαλμένη. Είναι προφανές ότι όσο αυξάνει η στάθμη εμπιστοσύνης (ή όσο ελαττώνεται η πιθανότητα σφάλματος), τόσο ελαττώνεται η πιθανότητα σφάλματος 1ου είδους, αλλά αντίστοιχα αυξάνει η πιθανότητα σφάλματος 2ου είδους.

Εργασία σε στάθμη εμπιστοσύνης 100% σημαίνει ότι ποτέ δεν θα υπάρχει περίπτωση σφάλματος 1ου είδους (όλα τα λάθη, όποτε προκύπτουν, θα είναι αποκλειστικά 2ου είδους). Από πρακτική άποψη αυτό σημαίνει (σε διάφορες δοκιμασίες) π.χ. ότι ποτέ δεν θα διαπιστωθούν συστηματικά λάθη, ότι όλες οι αποκλίσεις των μέσων τιμών ομάδων μετρήσεων θεωρούνται τυχαίες (όσο μεγάλες και να είναι), ότι όλες οι τιμές θεωρούνται έγκυρες και δεν υπάρχουν **έκτοπες τιμές** (outliers) για απόρριψη, όσο και να απέχουν από την κύρια μάζα των τιμών. Είναι ευνόητο ότι δοκιμασίες σε τόσο υψηλές στάθμες εμπιστοσύνης ουσιαστικά δεν προσφέρουν τίποτα.

Οι περισσότερες δοκιμασίες βασίζονται σε σύγκριση των **πειραματικών τιμών** (experimental values, R_{exp}) ποικιλίας στατιστικών κριτηρίων με **θεωρητικές** ή **κρίσιμες τιμές** (theoretical ή critical values, R_{theor}) των κριτηρίων αυτών που βρίσκονται σε στατιστικούς πίνακες. Οι κρίσιμες τιμές των πινάκων έχουν υπολογιστεί σε καθορισμένες στάθμες εμπιστοσύνης, όπως π.χ. 0,90, 0,95 και 0,99 ή στις αντίστοιχες πιθανότητες σφάλματος (ορθότερα: μέγιστες πιθανότητες σφάλματος 1ου είδους) που θα είναι: 0,10, 0,05 και 0,01, όπως και για διάφορα μεγέθη πληθυσμιακών δειγμάτων ή (γενικότερα) βαθμών ελευθερίας (δισδιάστατοι πίνακες). Η ανάγκη χρήσης πινάκων οφείλεται στο ότι δεν υπάρχουν απλές αναλυτικές εκφράσεις των θεωρητικών τιμών R_{theor}, ως συναρτήσεις της στάθμης εμπιστοσύνης και των βαθμών ελευθερίας. Εάν οι τιμές των στατιστικών κριτηρίων υπερβαίνουν αυτές τις κρίσιμες τιμές, τότε (ανάλογα με τη φύση της στατιστικής δοκιμασίας) γίνεται αποδεκτή ή όχι, η μηδενική υπόθεση που χαρακτηρίζει τη δοκιμασία.

Κατά κανόνα στις μετρήσεις χρησιμοποιείται η στάθμη εμπιστοσύνης 95%. Ωστόσο, αν οι επιπτώσεις σφάλματος 1ου είδους είναι κατά πολύ δραματικότερες (πλέον ανεπιθύμητες) από εκείνες του σφάλματος 2ου είδους, τότε μπορεί να χρησιμοποιηθεί η στάθμη εμπιστοσύνης 99%, ενώ εάν ισχύει το αντίστροφο, τότε μπορεί να χρησιμοποιηθεί η στάθμη εμπιστοσύνης του 90%.

Στατιστικά προγράμματα υπολογιστών και η πιθανότητα σφάλματος. Ο σύγχρονος τρόπος εκτέλεσης των στατιστικών δοκιμασιών γίνεται πλέον με τη βοήθεια υπολογιστών και στατιστικών προγραμμάτων που ουσιαστικά καθιστούν άχρηστους τους πίνακες κρίσιμων τιμών. Τα προγράμματα, λαμβάνοντας υπόψη όλες τις απαραίτητες παραμέτρους, όπως πλήθος μετρήσεων, βαθμούς ελευθερίας κ.λπ., υπολογίζουν όχι μόνο την πειραματική τιμή του εκάστοτε στατιστικού κριτηρίου, αλλά (και κυρίως) την οριακή πιθανότητα σφάλματος που αντιστοιχεί στην τιμή αυτή. Η ίδια η τιμή του στατιστικού κριτηρίου εκ των πραγμάτων πλέον δεν χρησιμεύει σε τίποτα. Συνήθως στην απάντηση δεν αναφέρεται ρητά το εάν απορρίπτεται ή όχι η (όποια) μηδενική υπόθεση. Την απόφαση την αφήνει στον χρήστη του προγράμματος, ο οποίος θα κρίνει ανάλογα με την επιθυμητή στάθμη εμπιστοσύνης. Έτσι, π.χ. εάν το πρόγραμμα υπολογίσει για κάποια δοκιμασία σημαντικότητας ως αποτέλεσμα:

ρισμένη στάθμη εμπιστοσύνης (π.χ. 95%), είναι το να λέγεται γενικά και αόριστα "*η πιθανότητα σφάλματος είναι 5%*". Υπάρχουν δύο λάθη στην έκφραση:

(1) Δεν δηλώνεται το είδος του σφάλματος, οπότε ο χρήστης δημιουργεί την εντύπωση ότι γενικά η γνωμάτευση του είναι κατά 95% ορθή. Αν όμως ήταν έτσι, θα μπορούσε να χρησιμοποιήσει στάθμη εμπιστοσύνης 99% για να είναι 99% βέβαιος ότι η γνωμάτευση είναι ορθή !!

(2) Ορίζει επακριβώς την πιθανότητα σφάλματος ως 5%, ενώ αυτή είναι η μέγιστη πιθανή.

Επομένως, η σωστή έκφραση είναι: "*η πιθανότητα σφάλματος 1ου είδους δεν υπερβαίνει το 5%*".

P (error probability) = 0,024 τούτο σημαίνει απλά ότι: *“Ως έχουν τα δεδομένα, η μηδενική υπόθεση μπορεί να απορριφθεί με πιθανότητα σφάλματος πρώτου είδους (εσφαλμένης απόρριψης) 2,4%”*. Σε περίπτωση που ο χρήστης επιθυμεί το συμπέρασμα να βασίζεται σε προκαθορισμένες (τυποποιημένες) στάθμες εμπιστοσύνης, τότε:

Σε στάθμη εμπιστοσύνης 90% (P=0,10) η μηδενική υπόθεση **απορρίπτεται**.
Σε στάθμη εμπιστοσύνης 95% (P=0,05) η μηδενική υπόθεση **απορρίπτεται**.
Σε στάθμη εμπιστοσύνης 99% (P=0,01) η μηδενική υπόθεση **δεν απορρίπτεται**.

3.4. Δοκιμασίες ενός και δύο άκρων.

Οι δοκιμασίες σημαντικότητας διακρίνονται γενικά σε δύο τύπους: Σε δοκιμασίες **ενός άκρου** ή **μονόπλευρες** (one-tailed tests) και σε δοκιμασίες **δύο άκρων** ή **δίπλευρες** (two-tailed tests).

Στις **δοκιμασίες ενός άκρου** εξετάζεται π.χ. το εάν οι διαφορές μεταξύ των αποτελεσμάτων A και B είναι τυχαίες ή εάν το αποτέλεσμα A είναι **συστηματικά μεγαλύτερο** (ή **μικρότερο**) του αποτελέσματος B, δηλ. η μηδενική υπόθεση και η εναλλακτική υπόθεση θα είναι:

$$H_0 : A = B$$
$$H_a : A > B \quad \text{ή} \quad H_a : A < B$$

Στις **δοκιμασίες δύο άκρων** εξετάζεται π.χ. το εάν οι διαφορές μεταξύ των μετρήσεων A και B είναι τυχαίες ή εάν τιμές A και B **διαφέρουν συστηματικά**, δηλ. η μηδενική υπόθεση και η εναλλακτική υπόθεση θα είναι:

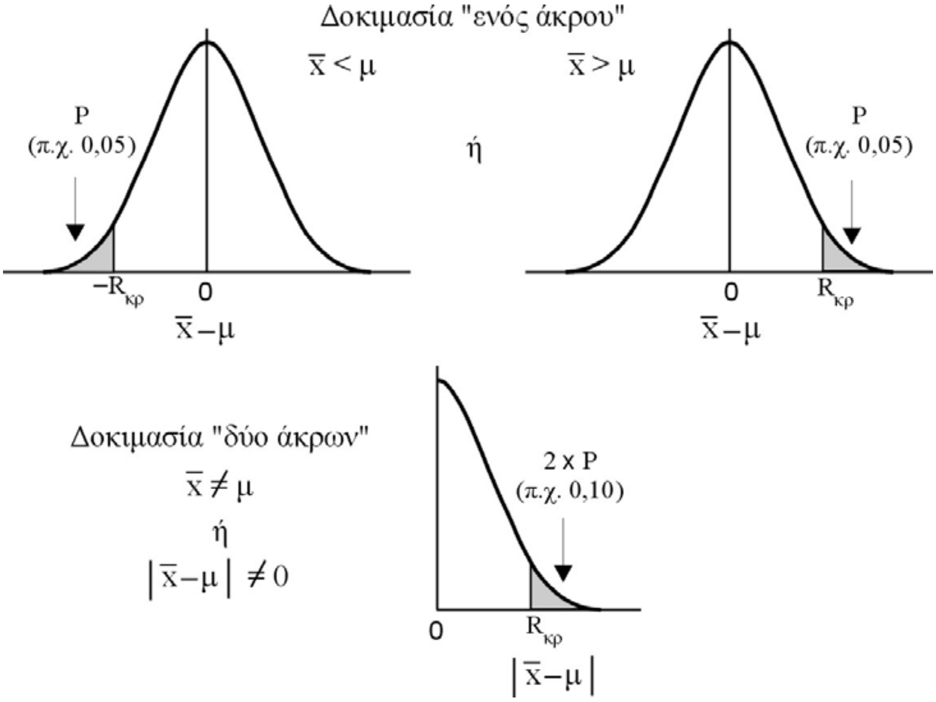
$$H_0 : A = B$$
$$H_a : A \neq B \ \ (\text{ή ισοδύναμα: } |A - B| \neq 0)$$

Έχει ιδιαίτερη σημασία να γνωρίζουμε ποιου τύπου δοκιμασία θα πρέπει να εφαρμόσουμε, διότι αυτό θα καθορίσει και το ποιες κρίσιμες τιμές θα χρησιμοποιήσουμε, δεδομένου ότι οι κρίσιμες τιμές των στατιστικών στοιχείων που βρίσκονται σε στατιστικούς πίνακες δίνονται για δοκιμασίες του ενός μόνο τύπου (συνήθως για δοκιμασίες “δύο άκρων”).

Τυπικό παράδειγμα όπου απαιτείται στατιστική **δοκιμασία ενός άκρου**: Βρέθηκε ότι η μέση περιεκτικότητα τοξικής ουσίας σε τρόφιμο είναι ίση προς 0,12% (με s = 0,05 και n = 4). Το ανώτερο επιτρεπτό όριο είναι 0,10%. Η τοξική ουσία υπερβαίνει πραγματικά το ανώτατο όριο ή η διαφορά αυτή οφείλεται σε τυχαίους λόγους;

Τυπικό παράδειγμα όπου απαιτείται στατιστική **δοκιμασία δύο άκρων**: Σε δείγμα ορυκτού προσδιορίζεται η περιεκτικότητα του σε ένα δευτερεύον συστατικό με δύο διαφορετικές μεθόδους, την A και τη B. Η μέθοδος A παρέχει αναλυτικό αποτέλεσμα 1,35 (με s = 0,20 και n = 3), ενώ η μέθοδος B παρέχει αναλυτικό αποτέλεσμα 1,42 (με s = 0,25 και n = 4). Οι μέθοδοι παρέχουν διαφορετικά αποτελέσματα ή όχι; ΠΡΟΣΟΧΗ: εάν το ερώτημα ήταν το εάν η B παρέχει συστηματικώς μεγαλύτερα αποτελέσματα από την A, τότε η δοκιμασία θα πρέπει να είναι ενός άκρου.

Στο Σχήμα 3-2 δείχνεται η διαφορά μεταξύ των δύο τύπων δοκιμασιών, στην περίπτωση σύγκρισης μιας μέσης τιμής μετρήσεων με την πραγματική για τη διαπίστωση ύπαρξης συστηματικού σφάλματος στη μέθοδο μέτρησης.



Σχήμα 3-2. *Σχηματική επεξήγηση (βλ. κείμενο) των διαφορών των στατιστικών δοκιμασιών "ενός άκρου" (εάν υπάρχει συστηματικό ή αρνητικό ή θετικό σφάλμα) και "δύο άκρων"(εάν υπάρχει γενικώς σφάλμα) στην περίπτωση σύγκρισης μιας μέσης τιμής μετρήσεων ($\bar{x}$) και της πραγματικής τιμής (μ).*

Οι δύο επάνω καμπύλες κατανομής (Gauss) του σφάλματος δείχνουν την περίπτωση **δοκιμασίας ενός άκρου**. Η μία (επάνω αριστερά) αφορά την περίπτωση κατά την οποία η μέση τιμή εξετάζεται εάν είναι μικρότερη της πραγματικής (αρνητικό σφάλμα). Η άλλη (επάνω δεξιά) την περίπτωση κατά την οποία η μέση τιμή εξετάζεται εάν είναι μεγαλύτερη της πραγματικής (θετικό σφάλμα). Η επιλεγόμενη κρίσιμη τιμή $R_{κρ}$ οριοθετεί (είτε στη μια είτε στην άλλη) "ουρά" πιθανότητας P (π.χ. 0,05, δηλ. το 5% της συνολικής επιφάνειας).

Στην κάτω καμπύλη δείχνεται η περίπτωση **δοκιμασίας δύο άκρων**. Τώρα, στην καμπύλη Gauss λείπει το αριστερό μισό, διότι χρησιμοποιείται η απόλυτη τιμή της διαφοράς, οπότε η κατανομή αφορά αποκλειστικώς θετικές τιμές. Η ίδια κρίσιμη τιμή $R_{κρ}$ είναι προφανές ότι οριοθετεί "ουρά" πιθανότητας 2×P (π.χ. 0,10, δηλ. το 10% της συνολικής επιφάνειας), αφού η συνολική επιφάνεια είναι η μισή σε σχέση με την επιφάνεια των επάνω καμπυλών Gauss.

Έστω ότι διαθέτουμε πίνακα κρίσιμων τιμών δοκιμασιών "δύο άκρων" σε στάθμες εμπιστοσύνης 90% (ή P = 0,10), 95% (ή P = 0,05) και 99% (ή P = 0,01). Εάν για κάποιο λόγο χρειαστεί να πραγματοποιήσουμε δοκιμασία "ενός άκρου" σε στάθμη εμπιστοσύνης 95% (ή P = 0,05), τότε θα χρησιμοποιήσουμε τις τιμές των κρίσιμων τιμών του πίνακα που αντιστοιχούν σε στάθμη εμπιστοσύνης 90% (ή P = 0,10).

Γενικά, εάν θέλουμε να πραγματοποιήσουμε στατιστική δοκιμασία "ενός άκρου" με μέγιστη πιθανότητα σφάλματος (1ου είδους πάντοτε) P, πρέπει να χρησιμοποιήσουμε τις κρίσιμες τιμές του στατιστικού στοιχείου δοκιμασίας "δύο άκρων" που αντιστοιχούν σε πιθανότητα σφάλματος 2×P).

3.5. Δοκιμασίες Student ή δοκιμασίες t

Οι δοκιμασίες του Student[15] βασίζονται στην ομώνυμη κατανομή (Student's t distribution) και απαιτούν μόνο εκτίμηση της τυπικής απόκλισης (s) (από μικρό πληθυσμιακό δείγμα) και όχι την πραγματική τιμή της τυπικής απόκλισης (σ) [16]. Υποτίθεται πάντοτε ότι **οι πληθυσμοί είναι κανονικοί**. Η ομάδα των δοκιμασιών αυτού του τύπου, γνωστές και ως t-tests είναι από τις πλέον απλές στατιστικές δοκιμασίες και συγχρόνως είναι οι πλέον συνηθισμένες στην αξιολόγηση αναλυτικών μεθόδων και μετρήσεων. Με τις δοκιμασίες αυτές παρέχονται στατιστικώς αντικειμενικές απαντήσεις σε ερωτήματα όπως:

1. Η μέση τιμή μιας σειράς μετρήσεων διαφέρει σημαντικά από την αποδεκτή (ή πραγματική) τιμή του δείγματος;

2. Οι μέσες τιμές δύο σειρών μετρήσεων διαφέρουν σημαντικά;

Σύγκριση πειραματικής μέσης τιμής ($\bar{x}$) και πραγματικής τιμής (μ). Η δοκιμασία της αξιοπιστίας μιας αναλυτικής μεθόδου ξεκινάει συνήθως με πολλαπλές μετρήσεις σε ένα δείγμα γνωστής περιεκτικότητας και σύγκριση της μέσης τιμής των αποτελεσμάτων με την πραγματική τιμή. Υπολογίζεται η πειραματική τιμή του στατιστικού στοιχείου t (t_{exp}), από την ακόλουθη εξίσωση:

$$t_{exp} = \frac{|\mu - \bar{x}|\sqrt{N}}{s} \qquad (3\text{-}2)$$

όπου μ είναι η πραγματική (αποδεκτή) μέση τιμή (ή η γνωστή τιμή σε συνθετικό δείγμα), $\bar{x}$ η πειραματική μέση τιμή που υπολογίστηκε από N μετρήσεις και s η τυπική απόκλιση των μετρήσεων που υπολογίστηκε με ν βαθμούς ελευθερίας (π.χ. ν = N–1 εάν έχει υπολογιστεί από τις N μετρήσεις).

Εάν είναι $t_{exp} \geq t_{theor}$ τότε η μηδενική υπόθεση απορρίπτεται στη δεδομένη στάθμη εμπιστοσύνης, δηλαδή η μέση τιμή θεωρείται σημαντικά διαφορετική από την πραγματική (ή αποδεκτή τιμή) και τούτο μπορεί να αποδοθεί σε ανεπάρκεια (συστηματικό σφάλμα) της εξεταζόμενης μεθόδου. Σε αντίθετη περίπτωση, οι όποιες διαφορές θεωρούνται απλώς τυχαίες.

Οι θεωρητικές τιμές του t (t_{theor}) σε διάφορες στάθμες εμπιστοσύνης και βαθμούς ελευθερίας μέτρησης της s δίνονται στον Πίνακα 3.3.

Στο επόμενο παράδειγμα εξετάζεται λεπτομερειακά η επίδραση της απόκλισης της μέσης τιμής από την πραγματική τιμή, όπως επίσης και τα στατιστικά συμπεράσματα που εξάγονται σε κάθε περίπτωση σε διάφορες στάθμες εμπιστοσύνης.

---

[15] Ο W.S. Gossett το 1905 δημοσίευσε τη σχετική μελέτη με το ψευδώνυμο Student, επειδή ως υπάλληλος της Guinness Breweries της Ιρλανδίας, δεν θα έπρεπε να δημοσιεύσει αποτελέσματα, που τυπικά αποτελούσαν ιδιοκτησία της εταιρίας. Η σπουδαιότητα των αποτελεσμάτων όμως της πνευματικής του εργασίας ήταν τέτοια, που ο ίδιος έκρινε ότι θα έπρεπε να τα δημοσιεύσει, έστω και με ψευδώνυμο.

[16] Η κατανομή t αφορά στην κατανομή πληθυσμών των τιμών $t = (\mu - \bar{x})\sqrt{N} / s_x$, όπου $\bar{x}$ η μέση τιμή n τιμών που έχει παρθεί από κανονικό πληθυσμό μέσης τιμής μ. s_x είναι η υπολογιζόμενη τυπική απόκλιση από τις n τιμές. Η κατανομή t αντικαθιστά την κατανομή z (κανονικοποιημένη κατανομή Gauss), όταν έχουμε μόνο εκτιμήτρια (s_x) της πραγματικής τυπικής απόκλισης (σ_x) των τιμών του πληθυσμού και ως εκ τούτου εκ της κατανομής αυτής εξάγονται όλα τα συμπεράσματα σε περιπτώσεις χειρισμού μικρών πληθυσμιακών δειγμάτων.

Πίνακας 3.3. Θεωρητικές τιμές t ("δύο άκρων") σε διάφορους βαθμούς ελευθερίας και στάθμες εμπιστοσύνης.

Βαθμοί ελευθερίας	Στάθμη εμπιστοσύνης			
$\nu = N - 1$	80%	90%	95%	99%
1	3,078	6,314	12,706	63,657
2	1,886	2,920	4,303	9,925
3	1,638	2,353	3,182	5,841
4	1,533	2,132	2,776	4,604
5	1,476	2,015	2,571	4,032
6	1,440	1,943	2,447	3,707
7	1,415	1,895	2,365	3,500
8	1,397	1,860	2,306	3,355
9	1,383	1,833	2,262	3,250
10	1,372	1,812	2,228	3,169
11	1,363	1,796	2,201	3,106
12	1,356	1,782	2,179	3,055
15	1,341	1,753	2,131	2,947
20	1,325	1,725	2,086	2,845
30	1,310	1,697	2,042	2,787
∞*	1,282	1,645	1,960	2,576

(*) Όταν οι βαθμοί ελευθερίας καταστούν άπειροι, οι κρίσιμες τιμές t ουσιαστικά καθίστανται ίσες προς τις προβλεπόμενες από την κανονική κατανομή, η οποία είναι η οριακή κατανομή της κατανομής Student. Έτσι, από τον Πίνακα 2.4 μπορεί να διαπιστωθεί ότι η πιθανότητα μιας τιμής (κανονικού πληθυσμού) να βρεθεί στην περιοχή ±1,282, ±1,645, ±1,960 και ±2,576 τυπικών αποκλίσεων (σε μονάδες z) είναι αντίστοιχα: 0,80, 0,90, 0,95 και 0,99.

Παράδειγμα 3-1. Μια υπό δοκιμή αναλυτική μέθοδος προσδιορισμού μιας ουσίας Α παρέχει τα ακόλουθα αναλυτικά αποτελέσματα (π.χ. % περιεκτικότητα):

5,42, 5,67, 5,75, 5,51, 5,82, 5,55

Σε στάθμη εμπιστοσύνης 95%, να εξεταστεί εάν στη δοκιμαζόμενη μέθοδο ενυπάρχει συστηματικό σφάλμα, εάν η πραγματική τιμή (Α%) στο εξεταζόμενο δείγμα είναι:

α) 5,64, β) 5,76, γ) 5,83 και δ) 6,12

Λύση. Η μέση τιμή $\bar{x}$ και η τυπική απόκλιση s είναι 5,62 και 0,1526 αντιστοίχως, οπότε για τις τρεις περιπτώσεις οι αντίστοιχες τιμες t_{exp} είναι:

α) $t_{exp} = |5{,}64 - 5{,}62|\sqrt{6}\big/0{,}1526 = 0{,}321$

β) $t_{exp} = |5{,}76 - 5{,}62|\sqrt{6}\big/0{,}1526 = 2{,}247$

γ) $t_{exp} = |5{,}83 - 5{,}62|\sqrt{6}\big/0{,}1526 = 3{,}371$

δ) $t_{exp} = |6{,}12 - 5{,}62|\sqrt{6}\big/0{,}1526 = 8{,}026$

Σε όλες τις περιπτώσεις, από τον Πίνακα 3.3 ισχύει ότι σε στάθμη εμπιστοσύνης 95% και για $\nu = 6 - 1 = 5$ η θεωρητική τιμή t_{theor} είναι 2,571. Επομένως οι απαντήσεις είναι:

α) Για πραγματική τιμή Α% 5,64: Επειδή $t_{exp} < t_{theor}$ τότε σε στάθμη εμπιστοσύνης 95% (δηλαδή, διακινδυνεύοντας πιθανότητα σφάλματος 1ου είδους μικρότερη από 0,05) διαπιστώνεται ότι **δεν υπάρχει συστηματικό σφάλμα**. [**Παρατήρηση**: Ακόμη και σε στάθμη εμπιστοσύνης

90%, όπου η απόρριψη της μηδενικής υπόθεσης είναι ευκολότερη, το συμπέρασμα είναι το ίδιο, είναι άλλωστε προφανές ότι η απόκλιση από την πραγματική τιμή είναι ελάχιστη].

β) Για πραγματική τιμή Α% 5,76: Επειδή $t_{exp}<t_{theor}$ τότε σε στάθμη εμπιστοσύνης 95% (δηλαδή, διακινδυνεύοντας πιθανότητα σφάλματος 1ου είδους μικρότερη από 0,05) διαπιστώνεται ότι **δεν υπάρχει συστηματικό σφάλμα**. [**Παρατήρηση**: Σε στάθμη εμπιστοσύνης 90%, όπου η απόρριψη της μηδενικής υπόθεσης είναι ευκολότερη, το συμπέρασμα είναι ότι **υπάρχει** συστηματικό σφάλμα].

γ) Για πραγματική τιμή Α% 5,83: Επειδή $t_{exp}>t_{theor}$ τότε σε στάθμη εμπιστοσύνης 95% (δηλαδή διακινδυνεύοντας πιθανότητα σφάλματος 1ου είδους μικρότερη από 0,05) διαπιστώνεται ότι **υπάρχει συστηματικό σφάλμα**. [**Παρατήρηση**: Σε στάθμη εμπιστοσύνης 99%, όπου η απόρριψη της μηδενικής υπόθεσης είναι δυσκολότερη, το συμπέρασμα είναι ότι **δεν υπάρχει** συστηματικό σφάλμα].

δ) Για πραγματική τιμή Α% 6,12: Επειδή $t_{exp}>t_{theor}$ τότε σε στάθμη εμπιστοσύνης 95% (δηλαδή διακινδυνεύοντας πιθανότητα σφάλματος 1ου είδους μικρότερη από 0,05) διαπιστώνεται ότι **υπάρχει συστηματικό σφάλμα**. [**Παρατήρηση**: Είναι πλέον τόσο μεγάλη η διαφορά της πραγματικής τιμής, που ακόμη και σε στάθμη εμπιστοσύνης 99%, όπου η απόρριψη της μηδενικής υπόθεσης είναι δυσκολότερη, το συμπέρασμα εξακολουθεί να είναι ότι **υπάρχει** συστηματικό σφάλμα].

Τα συνολικά συμπεράσματα συνοψίζονται στον ακόλουθο πίνακα, όπου παρέχονται και οι οριακές τιμές της πιθανότητας σφάλματος (P_{theor}) που αντιστοιχούν στις τιμές t_{exp}, όπως θα υπολογιζόντουσαν από ένα στατιστικό πακέτο προγράμματος.

Περίπτωση	$\bar{x}$	μ	t_{exp}	Συμπέρασμα σε στάθμη εμπιστοσύνης			P_{theor}
				90%	95%	99%	
α	5,62	5,64	0,321	$\bar{x} = \mu$	$\bar{x} = \mu$	$\bar{x} = \mu$	0,7611
β	5,62	5,76	2,247	$\bar{x} \neq \mu$	$\bar{x} = \mu$	$\bar{x} = \mu$	0,0745
γ	5,62	5,83	3,371	$\bar{x} \neq \mu$	$\bar{x} \neq \mu$	$\bar{x} = \mu$	0,0199
δ	5,62	6,12	8,026	$\bar{x} \neq \mu$	$\bar{x} \neq \mu$	$\bar{x} \neq \mu$	0,0005

Έτσι, εξετάζοντας τις ακραίες περιπτώσεις **α** και **δ**, φαίνεται ότι στην **α**, όπου είναι προφανές ότι η μηδενική υπόθεση είναι ισχυρότατη, σε περίπτωση που επιμείνουμε στην απόρριψή της, διακινδυνεύουμε απαράδεκτα μεγάλη πιθανότητα σφάλματος 1ου είδους (76%). Στην περίπτωση **δ**, όπου είναι προφανές ότι η μηδενική υπόθεση είναι εξαιρετικά ανίσχυρη, θα τη δεχόμαστε μόνο αν θέλαμε υπερβολικά μικρή πιθανότητα 0,05% σφάλματος 1ου είδους (αντίστοιχη στάθμη εμπιστοσύνης: 99,95%), κάτι που βέβαια δεν θα είχε κανένα νόημα.

Παράδειγμα 3-2. Τιτλοδοτούνται 10,00 mL διαλύματος ασθενούς βάσης (π.χ. ενός αλκαλοειδούς) 0,1000 M σε άνυδρο οξικό οξύ με πρότυπο διάλυμα $HClO_4$ 0,1000 M (επίσης σε άνυδρο οξικό οξύ) και υπάρχει η υποψία ότι με τον χρησιμοποιούμενο δείκτη λαμβάνονται συστηματικώς θετικά σφάλματα. Να εξεταστεί το αληθές της υπόθεσης σε στάθμη εμπιστοσύνης 95%, εάν για 5 τιτλοδοτήσεις χρειάστηκαν: 10,22, 10,08, 10,15, 9,97 και 10,12 mL του διαλύματος $HClO_4$.

Λύση. Η πειραματική τιμή t_{exp} βρίσκεται ίση προς 2,404 (Εξίσωση 3-2). Εφόσον ζητείται να εξετάσουμε το αν η μέση τιμή **υπερβαίνει** (και όχι απλώς ότι **διαφέρει**) σημαντικά από τη θεωρητική, θα χρησιμοποιήσουμε **δοκιμασία "ενός άκρου"**. Για τη ζητούμενη στάθμη εμπιστοσύνης 95% (ή P = 0,05) χρησιμοποιείται η θεωρητική τιμή t "δύο άκρων" 90% (ή P = 0,10) (ν = N–1 = 4) που είναι 2,132. Επειδή $t_{exp}>t_{theor}$ η μηδενική υπόθεση απορρίπτεται και επομένως στη ζητούμενη στάθμη εμπιστοσύνης διαπιστώνεται ότι **υπάρχει συστηματικό θετικό σφάλμα** στην τιτλοδότηση. [**Παρατήρηση:** Σε πρώτη προσέγγιση φαίνεται παράδοξο, το ότι στην ίδια στάθμη εμπιστοσύνης δεν διαπιστώνεται συστηματικό σφάλμα, αφού t_{theor}

(95%, ν = 4) = 2,776. Άλλη όμως η υπόθεση “υπάρχει σημαντική θετική διαφορά” και άλλη η υπόθεση “υπάρχει σημαντική διαφορά”].

Σύγκριση δύο πειραματικών μέσων τιμών. Συχνά απαιτείται σύγκριση των τιμών δύο διαφορετικών μεθόδων (με ανάλυση του ίδιου δείγματος) ή σύγκριση δύο δειγμάτων (αναλυόμενων με την ίδια μέθοδο). Η σύγκριση γίνεται για να διαπιστωθεί αν υπάρχει σημαντική διαφορά στα αποτελέσματα μεταξύ των δύο αναλυτικών μεθόδων ή αν τα εξεταζόμενα δείγματα είναι διαφορετικά ή όχι, αντίστοιχα. Γενικά, η στατιστική σύγκριση δύο μέσων τιμών αναφέρεται ως **μη κατά ζεύγη δοκιμασία t** (unpaired t-test).

Η δοκιμασία βασίζεται στον υπολογισμό της πειραματικής τιμής t_{exp}, που παρέχεται από την εξίσωση:

$$t_{exp} = \frac{\left|\bar{x}_A - \bar{x}_B\right|}{s_{A,B}\sqrt{\dfrac{1}{N_A} + \dfrac{1}{N_B}}} \qquad (3\text{-}3)$$

όπου $s_{A,B}$ είναι η **συνδυασμένη τυπική απόκλιση** (pooled ή combined standard deviation), που υπολογίζεται και από τις δύο ομάδες δεδομένων και παρέχεται από την εξίσωση:

$$s_{A,B} = \sqrt{\frac{\sum_{i=1}^{N_A}\left(\bar{x}_A - x_{i,A}\right)^2 + \sum_{i=1}^{N_B}\left(\bar{x}_B - x_{i,B}\right)^2}{N_A + N_B - 2}} \qquad (3\text{-}4)$$

όπου $\bar{x}_A$ η μέση τιμή της πρώτης σειράς μετρήσεων ($x_{1,A}$, $x_{2,A}$, … $x_{NA,A}$) και $\bar{x}_B$ η μέση τιμή της δεύτερης σειράς μετρήσεων ($x_{1,B}$, $x_{2,B}$, … $x_{NB,B}$).

Η t_{exp} συγκρίνεται με τη θεωρητική τιμή t_{theor} (για $N_A + N_B - 2$ βαθμούς ελευθερίας) (τιμές Πίνακα 3.3) και εάν $t_{exp} \leq t_{theor}$ η μηδενική υπόθεση θεωρείται ότι ισχύει (π.χ. ότι οι μέσες τιμές δεν διαφέρουν σημαντικά ή ότι τα δείγματα είναι ίδια).

Βασική προϋπόθεση εφαρμογής των δοκιμασιών t, σε περιπτώσεις σύγκρισης πειραματικών μέσων τιμών είναι η ομοσκεδαστικότητα των μετρήσεων, δηλ. δεν πρέπει να υπάρχει σημαντική διαφορά στις τυπικές αποκλίσεις των ομάδων μετρήσεων. Ο στατιστικός έλεγχος πραγματοποιείται με τη δοκιμασία F (Snedecor’s F test) “δύο άκρων” που περιγράφεται αμέσως μετά τις δοκιμασίες Student.

Παράδειγμα 3-3. Συστατικό του ίδιου δείγματος προσδιορίστηκε με δύο μεθόδους A και B και ελήφθησαν οι ακόλουθες τιμές.

Μέθοδος A : 3,23, 3,45, 3,48, 3,67, 3,71
Μέθοδος B : 3,51, 3,71, 3,79, 3,98

Να εκτιμηθεί το εάν τα αποτελέσματα διαφέρουν σημαντικά σε στάθμη εμπιστοσύνης 95% .

Λύση. Για να εκτιμηθεί το εάν υπάρχει ή όχι σημαντική διαφορά στα αποτελέσματα σε στάθμη εμπιστοσύνης 95%, υπολογίζονται οι πειραματικές μέσες τιμές των δύο μεθόδων και η συνδυασμένη τυπική απόκλιση (Εξίσωση 3-3) και βρίσκονται αντιστοίχως:

$\bar{x}_A = 3{,}508$, $\bar{x}_B = 3{,}7475$ και $s_{A,B} = 0{,}1935$

[**Παρατήρηση:** Τα παραπάνω αποτελέσματα θεωρούνται ως ενδιάμεσα των υπολογισμών και για τον λόγο αυτό αναγράφονται με μεγάλο αριθμό σημαντικών ψηφίων και όχι τον αρμόζοντα για τη διαφαινόμενη επαναληψιμότητα των μεθόδων] και είναι

$$t_{exp} = \frac{|3{,}508 - 3{,}7475|}{0{,}1935\sqrt{(1/5)+(1/4)}} = 1{,}8425$$

Επειδή $t_{exp} < t_{theor}$ (= 2,365, για ν = (5+4) – 2 = 7) **η μηδενική υπόθεση δεν απορρίπτεται** (δηλ. τα αποτελέσματα θεωρούνται ουσιαστικά ίδια). [**Παρατήρηση**: Η οριακή πιθανότητα σφάλματος που αντιστοιχεί στην τιμή t_{exp} και υπολογίζεται από στατιστικό πρόγραμμα είναι 0,1080, δηλαδή, -σχεδόν οριακά- η μηδενική υπόθεση δεν απορρίπτεται ούτε σε στάθμη εμπιστοσύνης 90%].

Δοκιμασία t για παρατηρήσεις κατά ζεύγη. Κατά τη σύγκριση της ισοδυναμίας δύο αναλυτικών μεθόδων συχνά αναλύεται μια σειρά δειγμάτων, που καλύπτουν μία σχετικώς ευρεία περιοχή τιμών, οπότε διατίθεται μια σειρά "ζευγών" αναλυτικών αποτελεσμάτων. Περιπτώσεις όπως αυτές που αναφέρονται στη συνέχεια ουσιαστικά υποχρεώνουν τον αναλυτή σε στατιστική δοκιμασία κατά ζεύγη:

1) Η ποσότητα των διατιθέμενων δειγμάτων δεν επιτρέπει παρά μόνο ένα προσδιορισμό με κάθε μέθοδο.

2) Οι μέθοδοι συγκρίνονται με ποικιλία δειγμάτων διαφόρων προελεύσεων και ευρείας περιοχής συγκεντρώσεων.

3) Χρησιμοποιούνται αποτελέσματα που ελήφθησαν σε διαφορετικές χρονικές περιόδους, κάτω από διαφορετικές εργαστηριακές συνθήκες.

Τα αποτελέσματα συγκρίνονται με μία δοκιμασία t που είναι γνωστή και ως **κατά ζεύγη δοκιμασία t** (paired t-test). Η τιμή t_{exp} παρέχεται από τη σχέση:

$$t_{exp} = |\bar{x}_A - \bar{x}_B| \sqrt{\frac{N(N-1)}{\sum_{i=1}^{N}(d_i - \bar{d}_i)}} \qquad (3\text{-}5)$$

όπου $\bar{x}_A$, $\bar{x}_B$ οι μέσες τιμές των δύο ομάδων μετρήσεων (με τη μέθοδο A και B), N ο αριθμός των ζευγών, d_i η θετική ή αρνητική διαφορά $x_{iA} - x_{iB}$ και $\bar{d}_i$ η μέση τιμή των διαφορών αυτών.

Η Εξίσωση 3-5 γράφεται και ως

$$t_{exp} = \frac{|\bar{d}|\sqrt{N}}{s_d} \qquad (3\text{-}6)$$

όπου s_d είναι η τυπική απόκλιση της διαφοράς των τιμών.

Η τιμή t_{exp} συγκρίνεται, όπως και στις προηγούμενες δοκιμασίες t, με τη θεωρητική τιμή t_{theor} που αντιστοιχεί σε ν = N–1 βαθμούς ελευθερίας.

Παράδειγμα 3-4. Σε σειρά 6 αναλυτικών δειγμάτων προσδιορίστηκε η % περιεκτικότητα σε συστατικό X με δύο διαφορετικές μεθόδους. Τα αποτελέσματα είχαν ως εξής:

Μέθοδος A, % X :	0,34	0,61	1,24	1,67	2,12	3,51
Μέθοδος B, % X :	0,36	0,70	1,34	1,58	2,21	3,38

Να εκτιμηθεί εάν σε στάθμη εμπιστοσύνης 95% διαφαίνεται σημαντική διαφορά στα αναλυτικά αποτελέσματα των δύο μεθόδων.

Λύση. Χρησιμοποιείται η Εξίσωση 3-6 με υπολογιζόμενες τιμές:

$\bar{d} = 0{,}01333$ και $s_d = 0{,}10053$, οπότε: $t_{exp} = 0{,}01333\sqrt{6} / 0{,}10053 = 0{,}3249$

Η θεωρητική τιμή t_{theor} για $\nu = 6 - 1 = 5$ και στάθμη εμπιστοσύνης 95% είναι 2,571, οπότε εφόσον $t_{exp}<t_{theor}$, **δεν διαπιστώνεται σημαντική διαφορά στα αποτελέσματα** που παρέχονται από τις δύο μεθόδους. [**Παρατήρηση**: Η συγκριτικά πολύ μικρή της τιμής t_{exp} δείχνει ότι η μηδενική υπόθεση είναι ισχυρότατη. Η οριακή πιθανότητα σφάλματος που αντιστοιχεί στην τιμή t_{exp} και υπολογίζεται από στατιστικό πρόγραμμα είναι 0,7584, δηλαδή, όχι μόνο η μηδενική υπόθεση δεν απορρίπτεται σε στάθμη εμπιστοσύνης 90%, αλλά ακόμη και σε πολύ μικρότερες στάθμες εμπιστοσύνης -μέχρι περίπου 24%- η μηδενική υπόθεση φαίνεται ότι ισχύει].

3.6. Σύγκριση διακυμάνσεων (δοκιμασία F)

Η δοκιμασία F (F του Snedecor) χρησιμοποιείται για να εξαχθούν συμπεράσματα ως προς το αν οι διακυμάνσεις δύο πληθυσμών είναι ίδιες ή όχι. Το συμπέρασμα θα εξαχθεί με εξέταση των διακυμάνσεων πληθυσμιακών δειγμάτων. Στη χημική ανάλυση μπορεί να χρησιμοποιηθεί για τη σύγκριση της επαναληψιμότητας δύο αναλυτικών μεθόδων ή δύο διαφορετικών αναλυτών, όταν εργάζονται με την ίδια μέθοδο κ.λπ. Εάν δειχθεί ότι στη δεδομένη στάθμη εμπιστοσύνης υπάρχει σημαντική διαφορά, τότε οι επιμέρους μέσες τιμές δεν μπορούν να συγκριθούν με τη δοκιμασία t.

Το στατιστικό στοιχείο που χρησιμοποιείται και στις δύο περιπτώσεις είναι η πειραματική τιμή F_{exp}, που ορίζεται ως

$$F_{exp} = \frac{s_A^2}{s_B^2} \qquad (3\text{-}7)$$

όπου s_A και s_B είναι οι τυπικές αποκλίσεις των δύο ομάδων τιμών (πάντοτε πρέπει να είναι $s_A \geq s_B$, ώστε αντιστοίχως να είναι $F_{exp} \geq 1$). Οι θεωρητικές τιμές F (F_{theor}) παρέχονται από πίνακες (Πίνακας 3.4) για διάφορους βαθμούς ελευθερίας (ν_A και ν_B) προσδιορισμού των επιμέρους τυπικών αποκλίσεων και στάθμες εμπιστοσύνης. Όταν $F_{exp}>F_{theor}$ τότε στη δεδομένη στάθμη εμπιστοσύνης συμπεραίνουμε ότι οι διακυμάνσεις (ή οι ακρίβειες) των δύο μεθόδων διαφέρουν.

Είναι χαρακτηριστικό το ότι για άπειρους βαθμούς ελευθερίας και για τις δύο διακυμάνσεις, η θεωρητική τιμή F είναι ακριβώς 1 για οποιαδήποτε στάθμη εμπιστοσύνης. Τούτο θα έπρεπε να είναι αναμενόμενο, εφόσον οι άπειρες μετρήσεις παρέχουν τις πραγματικές αποκλίσεις (σ) και εφόσον αυτές ταυτίζονται ($\sigma_B = \sigma_A$) θα πρέπει να είναι ακριβώς F=1.

Πίνακας 3.4. Θεωρητικές τιμές F σε διάφορους βαθμούς ελευθερίας και στάθμες εμπιστοσύνης [1].

ν_A / ν_B	2	3	4	5	6	10	∞ [2]	Στάθμη εμπιστοσύνης δοκιμασίας: ενός άκρου	δύο άκρων
	9,00	9,16	9,24	9,29	9,33	9,39	9,49	90%	80%
2	19,00	19,16	19,25	19,30	19,33	19,37	19,50	95%	90%
	39,00	39,17	39,25	39,30	39,33	39,40	39,50	97,5%	95%
	99,00	99,17	99,25	99,30	99,33	99,36	99,50	99%	98%
	5,46	5,39	5,34	5,31	5,28	5,23	5,13	90%	80%
3	9,55	9,28	9,12	9,01	8,98	8,78	8,53	95%	90%
	16,04	15,44	15,10	14,88	14,73	14,42	13,90	97,5%	95%
	30,82	29,46	28,71	28,24	27,91	27,23	26,12	99%	98%
	4,32	4,19	4,11	4,05	4,01	3,92	3,76	90%	80%
4	6,94	6,59	6,39	6,26	6,16	5,96	5,63	95%	90%
	10,65	9,98	9,60	9,36	9,20	8,84	8,26	97,5%	95%
	18,00	16,69	15,98	15,52	15,21	14,54	13,46	99%	98%
	3,78	3,62	3,52	3,45	3,40	3,30	3,10	90%	80%
5	5,79	5,41	5,19	5,05	4,95	4,74	4,36	95%	90%
	8,43	7,76	7,39	7,15	6,98	6,61	6,02	97,5%	95%
	13,27	12,06	11.39	10,97	10,67	10,05	9,02	99%	98%
	3,46	3,29	3,18	3,11	3,05	2,94	2,72	90%	80%
6	5,14	4,76	4,53	4,39	4,28	4,06	3,67	95%	90%
	7,26	6,60	6,23	5,99	5,82	5,46	4,85	97,5%	95%
	10,92	9,78	9,15	8,75	8,47	7,87	6,88	99%	98%
	2,92	2,73	2,61	2,52	2,46	2,32	2,06	90%	80%
10	4,10	3,71	3,48	3,33	3,22	2,97	2,54	95%	90%
	5,46	4,83	4,47	4,24	4,07	3,72	3,08	97,5%	95%
	7,56	6,55	5,99	5,64	5,39	4,85	3,91	99%	98%
	2,70	2,49	2,36	2,27	2,21	2,06	1,76	90%	80%
15	3,68	3,29	3,06	2,90	2,79	2,54	2,07	95%	90%
	4,77	4,15	3,80	3,58	3,41	3,06	2,40	97,5%	95%
	6,36	5,42	4,89	4,36	4,32	3,80	2,87	99%	98%
[2]	2,30	2,08	1,94	1,85	1,77	1,60	**1,00**	90%	80%
∞	3,00	2,60	2,37	2,21	2,10	1,83	**1,00**	95%	90%
	3,69	3,12	2,79	2,57	2,41	2,05	**1,00**	97,5%	95%
	4,61	3,78	3,32	3,02	2,80	2,32	**1,00**	99%	98%

(1) ν_A, ν_B: βαθμοί ελευθερίας διακύμανσης αριθμητή και παρανομαστή, αντίστοιχα.
(2) Οι τιμές αυτές θα χρησιμοποιηθούν για σύγκριση μιας πειραματικής τιμής τυπικής απόκλισης (s_A ή s_B) με την πραγματικές τιμές τυπικής απόκλισης (σ_B και σ_A, αντιστοίχως), εάν βέβαια οι τελευταίες είναι γνωστές.

Παράδειγμα 3-5. Να εξεταστεί εάν υπάρχει στατιστικώς σημαντική διαφορά στα αποτελέσματα που παρέχει η μέθοδος Α σε σχέση με εκείνα που παρέχει η μέθοδος Β σε στάθμη εμπιστοσύνης 95%, όταν κατά τη μέτρηση του ίδιου δείγματος ελήφθησαν οι ακόλουθες τιμές:

Μέθοδος Α: 4,41, 4,54, 4,89, 4,89, 5,12
Μέθοδος Β: 4,52, 4,62, 4,68, 4,73

Λύση. Οι τυπικές αποκλίσεις (σχέση 2-12) είναι 0,2888 και 0,09032 αντιστοίχως, οπότε είναι $F_{exp} = 0{,}2888^2 / 0{,}09032^2 = 10{,}22 < F_{theor}$ (=15,10, για $\nu_A = 4$, $\nu_B = 3$). Επομένως, σε στάθμη εμπιστοσύνης 95% η μηδενική υπόθεση ισχύει και μπορούμε να αποφανθούμε ότι οι μέθοδοι είναι ισοδύναμες από άποψη ακρίβειας. [**Παρατήρηση**: Η οριακή πιθανότητα σφάλματος που αντιστοιχεί στην τιμή F_{exp} και υπολογίζεται από στατιστικό πρόγραμμα είναι 0,0856, δηλαδή, η μηδενική υπόθεση απορρίπτεται σε στάθμη εμπιστοσύνης 90%, κάτι που φαίνεται και από την αντίστοιχη κρίσιμη τιμή του πίνακα]

3.7. Δοκιμασίες ανίχνευσης εκτόπων τιμών (outliers)

Σε μία σειρά επαναλαμβανόμενων μετρήσεων της ίδιας ποσότητας, που συνήθως έχει ως σκοπό τον υπολογισμό της μέσης τιμής ή άλλων στατιστικών χαρακτηριστικών του πληθυσμιακού δείγματος (π.χ. της τυπικής απόκλισης), συχνά εμφανίζονται τιμές απομακρυσμένες από την κύρια μάζα των μετρήσεων. Οι τιμές αυτές θεωρούνται ύποπτες, ενδεχομένως είναι φορείς κάποιου συστηματικού σφάλματος και υπάρχει μια φυσική τάση του αναλυτή να μην τις συμπεριλάβει στους υπολογισμούς, κάτι που πρέπει να αποφύγει πριν δικαιολογήσει με στατιστικά στοιχεία την απόφαση αυτή.

Αντίστοιχες περιπτώσεις "ύποπτων" ή "κακών" τιμών εμφανίζονται κατά τη χάραξη καμπύλης αναφοράς, όπου π.χ. ενώ η πλειονότητα των τιμών των μετρήσεων των προτύπων εμφανίζουν την αναμενόμενη συμπεριφορά (π.χ. βρίσκονται σε ευθεία) μερικές τιμές φαίνεται να εκτρέπονται αναιτιολόγητα από αυτήν. Δημιουργείται πάλι το ερώτημα εάν οι τιμές (ή σημεία) αυτές θα πρέπει να περιληφθούν ή όχι στο διάγραμμα, οπότε ενδεχομένως θα επηρεάσουν τη θέση της καμπύλης και φυσικά όλα τα αποτελέσματα των μετρήσεων που θα ακολουθήσουν με βάση την προκύπτουσα καμπύλη αναφοράς.

> Εάν είναι γνωστό ότι κάτι "δεν πήγε καλά" κατά τη διάρκεια των μετρήσεων που έδωσαν τις "ύποπτες" τιμές (π.χ. κακή αραίωση ή απώλεια δείγματος, μια παροδική αστάθεια στην τάση τροφοδοσίας των συσκευών) οι τιμές αυτές, ανεξαρτήτως του εάν φαίνονται "καλές" ή "κακές", "ταιριάζουν" ή "δεν ταιριάζουν" με τις υπόλοιπες, θα πρέπει να απορριφθούν χωρίς δισταγμό και χωρίς να προηγηθεί στατιστική δοκιμασία.

Εάν δεν υπάρχουν ενδείξεις, ως προς το εάν οι "ύποπτες" τιμές είναι φορείς συστηματικού σφάλματος, τότε η απόρριψή τους σε καμία περίπτωση δεν θα πρέπει να γίνει με βάση υποκειμενικά κριτήρια.

Οι στατιστικώς ορθότερες δοκιμασίες απόρριψης είναι εκείνες που χαρακτηρίζουν τις μετρήσεις απορριπτέες σε προκαθορισμένες στάθμες εμπιστοσύνης (π.χ. 95%) ή -ισοδύναμα-πιθανότητας σφάλματος (π.χ. 0,05 ή 5%). Η μηδενική υπόθεση στις περιπτώσεις αυτές είναι ότι "η ύποπτη τιμή δεν διαφέρει σημαντικά από τις υπόλοιπες".

3.7.1. Ορισμοί

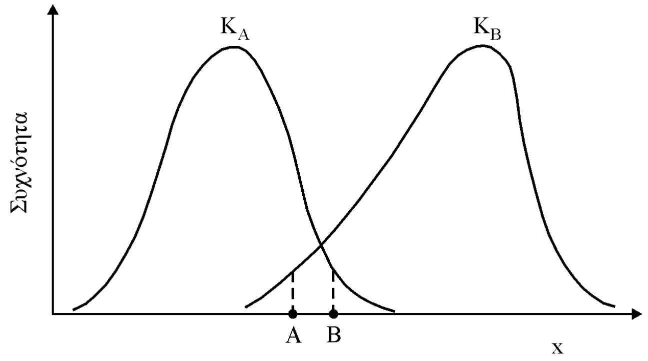
> Ως **έκτοπη** τιμή (outlier[17]) σε ένα πληθυσμιακό δείγμα χαρακτηρίζεται κάθε τιμή, που ανήκει σε διαφορετικό πληθυσμό από τον πληθυσμό της πλειονότητας των μετρήσεων.

Η διαφορά μεταξύ των πληθυσμών μπορεί να εντοπίζεται στη μέση τιμή, στη διασπορά (π.χ. στην τυπική απόκλιση) ή και στα δύο. Φαίνεται αρκετά παράδοξο το ότι με βάση τον παρα-

---

[17] Από την Αγγλική λέξη outlying που σημαίνει: “αυτός που βρίσκεται μακριά από το κέντρο”. Άλλες αποδόσεις που θα μπορούσαν να δοθούν στην Ελληνική για τον όρο outlier είναι: “απόκεντρος”, “αποκλίνουσα”, “απόμακρη”. Οι όροι αυτοί ως έννοιες έχουν συνεχή χαρακτήρα (π.χ. λίγο-, πολύ-, εξαιρετικά-αποκλίνουσα). Ο όρος “εξωγενής” θα απέδιδε καλύτερα την ουσία του όρου: τιμή προερχόμενη από άλλο είδος πληθυσμού και ως έννοια έχει δυαδικό χαρακτήρα (ή είναι ή δεν είναι εξωγενής). Όλες οι σειρές μετρήσεων περιέχουν αποκλίνουσες τιμές, αυτό όμως δεν σημαίνει ότι υποχρεωτικά αυτές είναι και απορριπτέες. Απορριπτέες θα είναι αν αποδειχθούν “εξωγενείς” με βάση κάποιο στατιστικό στοιχείο και σε προκαθορισμένο επίπεδο εμπιστοσύνης.

πάνω ορισμό είναι δυνατόν μία αποκλίνουσα τιμή να φαίνεται "καλύτερη" από μία πραγματική τιμή, όπως π.χ. τα σημεία A και B (Σχήμα 3-3), των πληθυσμών με κατανομές K_A (του μετρούμενου πληθυσμιακού δείγματος των "ορθών" τιμών) και K_B (των αποκλινουσών τιμών). Καμία στατιστική δοκιμασία για την απαλλαγή του εξεταζόμενου πληθυσμού K_A από έκτοπες τιμές, δεν μπορεί να κρατήσει την "νόμιμη" τιμή B και να απορρίψει την "εξωγενή" τιμή A, έτσι οι δοκιμασίες απόρριψης εκ των πραγμάτων περιορίζονται στις **ασύμφωνες**, **εκτρεπόμενες** ή **ακραίες** τιμές (discordant, deviant, extreme values), μικρών ομάδων δεδομένων.



Σχήμα 3-3. *Παράδειγμα επικάλυψης διαγραμμάτων κατανομής εξεταζόμενου πληθυσμού (K_A) και ενός άλλου πληθυσμού (K_B).*

Τα στατιστικά στοιχεία που επηρεάζονται ελάχιστα από την παρουσία εκτόπων μετρήσεων στο εξεταζόμενο δείγμα, ονομάζονται **ανθεκτικά** (robust), όπως επίσης ανθεκτικές ονομάζονται και οι αντίστοιχες δοκιμασίες, που κατά τα τελευταία χρόνια γίνονται συνεχώς και δημοφιλέστερες. Στο τελευταίο βοήθησε και η μεγάλη διαθεσιμότητα υπολογιστών, επειδή οι σχετικοί υπολογισμοί είναι περισσότερο σύνθετοι και απαιτούν περισσότερο χρόνο, σε σχέση με τις κλασσικές στατιστικές δοκιμασίες.

3.7.2. Κριτήρια απόρριψης τιμών

Το σύνολο σχεδόν των κριτηρίων απόρριψης ακραίων τιμών βασίζεται στο να θεωρούνται a priori οι τιμές, που βρίσκονται στα άκρα των κατανομών του πραγματικού πληθυσμού, ως έκτοπες μετρήσεις. Έτσι, π.χ. οι τιμές που βρίσκονται στο κατώτερο 2,5% (και κάτω) εκατοστημόριο και ανώτερο 97,5% (και άνω) εκατοστημόριο της κατανομής θεωρούνται ως αποκλίνουσες με προφανή πιθανότητα σφάλματος 1ου είδους 5%, δηλαδή για να να αποφευχθεί η παρουσία αποκλινουσών μετρήσεων προαποφασίζεται η απόρριψη μίας στις είκοσι (κατά μέσο όρο) κανονικών μετρήσεων ("ψαλίδισμα κατανομής").

Εάν είναι γνωστή η μορφή του διαγράμματος κατανομής ή η οποιαδήποτε περιγραφή του (π.χ. ότι πρόκειται για κανονική κατανομή με γνωστή τυπική απόκλιση σ), τότε είναι εύκολο με θεώρηση του διαγράμματος κατανομής (αν πρόκειται για κατανομή μη συγκεκριμένου τύπου) ή των σχετικών πινάκων (αριθμός σ ως συνάρτηση πληθυσμιακού κλάσματος στην περίπτωση κανονικής κατανομής) να απορριφθεί ή όχι. Η μηδενική υπόθεση παραμένει πάντοτε η ίδια: η εξεταζόμενη τιμή δεν διαφέρει από τις υπόλοιπες (οι όποιες διαφορές είναι μόνο τυχαίες).

Γενικά, απαιτείται ιδιαίτερη προσοχή στην απόρριψη μετρήσεων. Καλό θα ήταν για κάθε απορριπτόμενη τιμή να υπήρχε και μια ερμηνεία ως προς το πώς προέκυψε (αναζήτηση συστηματικού σφάλματος). Στους διεργαστηριακούς ελέγχους αναλυτικών μεθόδων (ISO-5725) υπάρχει τυποποιημένο και εξαιρετικά αυστηρό πρωτόκολλο χαρακτηρισμού μετρήσεων ως απορριπτέων. Τιμές που χαρακτηρίζονται ως έκτοπες (με στατιστικές δοκιμασίες ως αυτές που θα περιγραφούν στη συνέχεια) σε στάθμη εμπιστοσύνης 95%, αλλά όχι σε στάθμη 99% χαρακτηρίζονται ως **ακατάστατες** (stragglers). Κάτω από ορισμένες προϋποθέσεις και ανάλογα με τη συχνότητα εμφάνισής τους οι τιμές αυτές δεν απορρίπτονται και περιλαμβάνονται στους παραπέρα υπολογισμούς.

Τα κριτήρια που περιγράφονται στη συνέχεια (όπως και τα προηγούμενα) θεωρούν ότι η κατανομή του πληθυσμού είναι κανονική, επομένως μπορούν να εφαρμοστούν σε πληθυσμούς αναλυτικών σφαλμάτων (που κατά κανόνα ακολουθούν την κανονική κατανομή), όχι όμως σε πληθυσμούς, όπως π.χ. τιμές συγκεντρώσεων συστατικών δειγμάτων κλινικού ενδιαφέροντος, που ακολουθούν κατά κανόνα διαφορετικές ή και απρόβλεπτες κατανομές.

3.7.3. Μέθοδος Neir

Η μέθοδος αυτή βασίζεται στην εκτίμηση της τυπικής απόκλισης από περιορισμένο αριθμό μετρήσεων μεγαλύτερου όμως από εκείνο της εξεταζόμενης ομάδας δεδομένων (κατανομές Student).

Εάν x_m είναι η μικρότερη ή μεγαλύτερη τιμή x δείγματος μεγέθους N, υπολογίζεται η λόγος

$$r_N = \frac{\left| x_m - \bar{x} \right|}{s_e} \qquad (3\text{-}8)$$

όπου $\bar{x}$ είναι μέση τιμή του εξεταζόμενου δείγματος (**συμπεριλαμβανόμενης** της ύποπτης μέτρησης) και s_e η εκτιμούμενη τιμή της τυπικής απόκλισης πληθυσμιακού δείγματος που έχει υπολογιστεί με ν βαθμούς ελευθερίας (π.χ. με ν+1 ομοειδείς, επαναλαμβανόμενες μετρήσεις). Η τιμή r_N συγκρίνεται με κρίσιμες τιμές (θεωρητικές τιμές r_N, $r_{N,theor}$), που έχουν υπολογιστεί για διάφορες στάθμες εμπιστοσύνης. Εάν η πειραματική τιμή είναι μεγαλύτερη η τιμή x_m μπορεί να απορριφθεί στη δεδομένη στάθμη εμπιστοσύνης. Απόσπασμα πινάκων θεωρητικών τιμών r_N δίνεται στον Πίνακα 3.5.

Πίνακας 3.5. Θεωρητικές τιμές $r_{N,theor}$ για ανίχνευση εκτόπων τιμών κατά Neir.

N:	3	4	5	6	7	8	9
ν	Στάθμη εμπιστοσύνης 95%						
10	2,02	2,29	2,49	2,63	2,75	2,85	2,93
11	1,99	2,26	2,44	2,58	2,70	2,79	2,87
12	1,97	2,22	2,40	2,54	2,65	2,75	2,83
15	1,92	2,16	2,33	2,46	2,56	2,65	2,73
20	1,87	2,10	2,26	2,38	2,48	2,56	2,63
30	1,82	2,04	2,20	2,31	2,40	2,48	2,55
∞	1,74	1,94	2,08	2,18	2,27	2,33	2,39
ν	Στάθμη εμπιστοσύνης 99%						
10	2,76	3,05	3,25	3,39	3,50	3,59	3,67
11	2,71	3,00	3,19	3,33	3,44	3,53	3,61
12	2,67	2,95	3,14	3,28	3,39	3,48	3,55
15	2,57	2,84	3,02	3,16	3,27	3,35	3,43
20	2,47	2,73	2,91	3,04	3,14	3,22	3,29
30	2,38	2,62	2,79	2,91	3,01	3,08	3,15
∞	2,22	2,43	2,57	2,68	2,76	2,83	2,88
ν	Στάθμη εμπιστοσύνης 99,9%						
10	3,54	3,84	4,04	4,17	4,28	4,35	4,40
11	3,49	3,80	3,99	4,12	4,23	4,30	4,36
12	3,45	3,75	3,94	4,07	4,19	4,26	4,31
15	3,35	3,64	3,83	3,96	4,06	4,15	4,21
20	3,23	3,51	3,70	3,83	3,93	4,01	4,08
30	3,08	3,36	3,53	3,66	3,76	3,84	3,90
∞	2,78	3,01	3,17	3,28	3,36	3,43	3,48

3.7.4. Δοκιμασία Q του Dixon

Στη μέθοδο Neir απαιτείται να είναι γνωστή η εκτίμηση της τυπικής αποκλίσης (s_e) και τούτο αποτελεί ένα σαφές μειονέκτημα. Στη δοκιμασία Q κάτι τέτοιο δεν είναι απαραίτητο, αλλά πάλι θεωρείται ότι η ομάδα των δεδομένων προέρχεται από κανονικό πληθυσμό.

Η δοκιμασία Q (quotient) του Dixon βασίζεται στην κατανομή του λόγου **υποπεριοχών** (sub-range ratio) μιας διατεταγμένης σειράς μετρήσων ($x_1 > x_2 > ...> x_n$), που έχει ληφθεί από το ίδιο πληθυσμιακό δείγμα. Το στατιστικό στοιχείο είναι ο λόγος:

$$Q_{exp} = \frac{x_2 - x_1}{x_n - x_1} \qquad \text{(ύποπτη τιμή η } x_1\text{)} \qquad (3\text{-}9)$$

$$Q_{exp} = \frac{x_n - x_{n-1}}{x_n - x_1} \qquad \text{(ύποπτη τιμή η } x_n\text{)} \qquad (3\text{-}10)$$

Το γεγονός ότι οι τιμές Q είναι λόγος διαφορών, καθιστά τις τιμές αυτές ανεξάρτητες από τη μέση τιμή και την τυπική απόκλιση των κανονικών πληθυσμών και αυτό αποτελεί το κύριο πλεονέκτημα του κριτηρίου.

Πίνακας 3.6. Θεωρητικές τιμές Q σε διάφορες στάθμες εμπιστοσύνης και αριθμούς μετρήσεων (n).

n	80%	90%	95%	99%
3	0,886	0,941	0,970	0,994
4	0,679	0,765	0,829	0,926
5	0,557	0,642	0,710	0,821
6	0,482	0,560	0,625	0,740
7	0,434	0,507	0,568	0,680
8	0,399	0,468	0,526	0,634
9	0,370	0,437	0,493	0,598
10	0,349	0,412	0,466	0,568
11	0,332	0,392	0,444	0,542
12	0,318	0,376	0,426	0,522

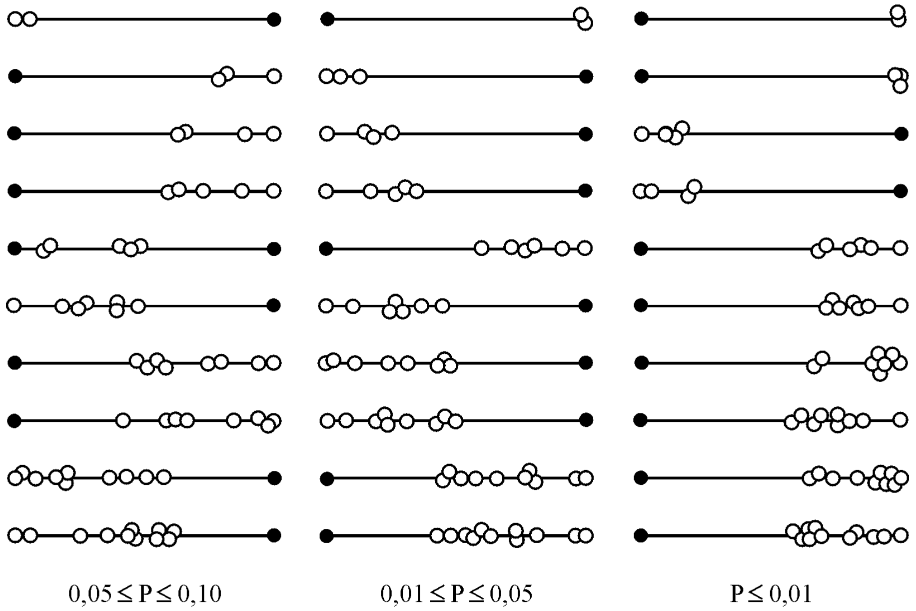
Εάν η τιμή Q_{exp} υπερβαίνει τη θεωρητική τιμή Q_{theor} (Πίνακας 3.6), τότε η ύποπτη τιμή (η x_1 ή η x_n) μπορεί να θεωρηθεί ως έκτοπη και να απορριφθεί σε δεδομένη στάθμη εμπιστοσύνης. Εάν απορριφθεί η ύποπτη ακραία τιμή, το κριτήριο Q δεν μπορεί να εφαρμοστεί στην ομάδα των υπολοίπων τιμών για την εξέταση επιπλέον τιμών. Τούτο οφείλεται στο ότι η αρχική δοκιμασία και απόρριψη βασιζόταν στην υπόθεση ύπαρξης μίας ή καμίας έκτοπης μέτρησης και έτσι η παραπέρα εφαρμογή απλά παραβίαζε την αρχική υπόθεση.

Στον Πίνακα 3.6 δίνονται οι θεωρητικές τιμές του λόγου Q σε διάφορες στάθμες εμπιστοσύνης. Η αναζήτηση μίας έκτοπης τιμής σε δείγματα με περισσότερα από 10 -12 στοιχεία **δεν έχει νόημα**, γιατί η τιμή αυτή θα επηρεάσει ελάχιστα μόνο τις στατιστικές ιδιότητες του δείγματος (μέση τιμή, τυπική απόκλιση).

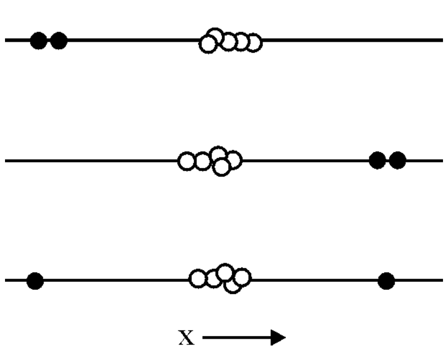
Στο Σχήμα 3-4 δείχνονται σημειογράμματα ομάδων 3 έως 12 δεδομένων με απορριπτόμενες τιμές σε διάφορες στάθμες εμπιστοσύνης. Από τα σημειογράμματα αυτά είναι φανερή η χαρακτηριστική ευκολία απόρριψης τιμών σε χαμηλή στάθμη εμπιστοσύνης (περιοχή 90-95% ή $0,05 \leq P \leq 0,10$) και η δυσκολία απόρριψης σε στάθμη εμπιστοσύνης 99% ή μεγαλύτερη (ή $P \leq 0,01$).

Παράδειγμα 3-6. Η ομάδα 5 μετρήσεων της ίδιας ποσότητας είναι η ακόλουθη (κατά αυξανόμενη σειρά): 5,32, 5,36, 5,41, 5,43 και 5,66. Να εξεταστεί εάν στην ομάδα τιμών μπορεί καμία να χαρακτηριστεί ως έκτοπη και να απορριφθεί.

Λύση. Προφανώς πιθανή έκτοπη τιμή είναι η 5,66. Από την Εξίσωση 3-10 υπολογίζεται η πειραματική τιμή του Q_{exp} = (5,66 – 5,43) / (5,66 – 5,32) = 0,676. Επειδή $Q_{exp} > Q_{theor,0,90}$ (=0,642) η τιμή 5,66 μπορεί να απορριφθεί με πιθανότητα σφάλματος 1ου είδους ≤10% (στάθμη εμπιστοσύνης: 90%). Εάν θέλουμε να αποφύγουμε σφάλμα 1ου είδους με πιθανότητα σφάλματος 1ου είδους ≤ 5% (στάθμη εμπιστοσύνης: 95%), επειδή $Q_{exp} < Q_{theor,0,95}$ (=0,710) η τιμή 5,66 δεν θα πρέπει να απορριφθεί.



Σχήμα 3-4. *Σημειογράμματα ομάδων σημείων (n=3 έως 12) με μία ανιχνευόμενη έκτοπη τιμή (μαύροι κύκλοι). Στην πρώτη στήλη η ακραία τιμή απορρίπτεται σε στάθμη εμπιστοσύνης 90% όχι όμως σε στάθμη εμπιστοσύνης 95%, στη δεύτερη στήλη η ακραία τιμή απορρίπτεται σε στάθμη εμπιστοσύνης 95% όχι όμως σε στάθμη εμπιστοσύνης 99% και στη τρίτη στήλη η ακραία τιμή απορρίπτεται σε στάθμη εμπιστοσύνης 99%..Είναι χαρακτηριστικό το ότι σε στάθμη εμπιστοσύνης 99%, όπου η απόρριψη της μηδενικής υπόθεσης γίνεται "δυσκολότερα", η παρουσία έκτοπης τιμής είναι προφανέστερη από ό,τι σε στάθμη εμπιστοσύνης 95% και αντίστοιχα σε στάθμη εμπιστοσύνης 95% η παρουσία έκτοπης τιμής είναι προφανέστερη από ό,τι σε στάθμη εμπιστοσύνης 90%..*



Σχήμα 3-5. *Σημειογράμματα ομάδων δεδομένων που δείχνουν τις θέσεις δύο πιθανώς έκτοπων τιμών (μαύροι κύκλοι).*

3.7.5. Καλυπτική δράση και επέκταση κριτηρίου Q για περισσότερες αποκλίνουσες τιμές

Η σύγχρονη παρουσία δύο ή περισσότερων αποκλινουσών τιμών σε μία μικρή ομάδα δεδομένων μπορεί να έχει ως αποτέλεσμα την **κάλυψη** (masking) της πλέον αποκλίνουσας τιμής, καθιστώντας αδύνατο τον χαρακτηρισμό της ως έκτοπης τιμής. Τούτο φαίνεται από τα σημειογράμματα του Σχήματος 3-5, όπου η παρουσία μία δεύτερης αποκλίνουσας τιμής υποβιβάζει σημαντικά τις πειραματικές τιμές Q.

Για τις περιπτώσεις αυτές υπάρχουν δοκιμασίες σύγχρονης ανίχνευσης ομάδων εκτόπων μετρήσεων, αρκετά σύνθετων στην εφαρμογή τους. Ένα απλό κριτήριο, επέκταση της δοκιμασίας Q, για τη σύγχρονη ανίχνευση δύο εκτόπων μετρήσεων σε μικρές ομάδες τιμών, βασίζεται στη χρήση του γινομένου λόγων υποπεριοχών QP (Quotient Product) ως στατιστικού στοιχείου:

$$QP = \frac{(x_3 - x_1)}{(x_n - x_1)} \cdot \frac{(x_3 - x_2)}{(x_n - x_2)} \qquad \text{(ύποπτο ζεύγος: } x_1, x_2\text{)} \qquad (3\text{-}11)$$

$$QP = \frac{(x_n - x_{n-2})}{(x_n - x_1)} \cdot \frac{(x_{n-1} - x_{n-2})}{(x_{n-1} - x_1)} \qquad \text{(ύποπτο ζεύγος: } x_{n-1}, x_n\text{)} \qquad (3\text{-}12)$$

$$QP = \frac{(x_2 - x_1)}{(x_{n-1} - x_1)} \cdot \frac{(x_n - x_{n-1})}{(x_n - x_2)} \qquad \text{(ύποπτο ζεύγος: } x_1, x_n\text{)} \qquad (3\text{-}13)$$

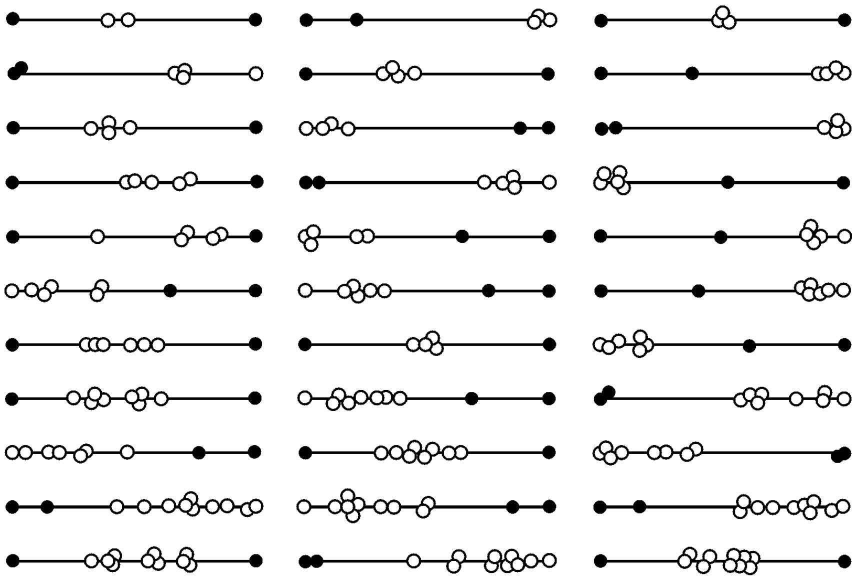
Κάθε τιμή QP είναι το γινόμενο δύο διαφορετικών τιμών Q (Dixon's Q). Κάθε τιμή Q υπολογίζεται από την ομάδα μετρήσεων λαμβάνοντας υπόψη τη μία και αγνοώντας την άλλη από τις ύποπτες μετρήσεις.

Οι κρίσιμες τιμές QP σε τρεις στάθμες εμπιστοσύνης υπολογίστηκαν από την κατανομή ενός μεγάλου πλήθους προσομοιώσεων (τυπικά 10000) με τη βοήθεια υπολογιστή και δίνονται στον Πίνακα 3.7. Στο Σχήμα 3-5 δείχνονται σημειογράμματα ομάδων 5 έως 12 τιμών με απορριπτόμενα ζεύγη τιμών σε διάφορες στάθμες εμπιστοσύνης.

Το κριτήριο QP μπορεί να επεκταθεί για την ανίχνευση τριάδων, τετράδων κ.λπ. εκτόπων τιμών. Πρέπει να τονιστεί ότι η παρουσία μίας έκτοπης τιμής δεν μπορεί να θεωρηθεί εκ των προτέρων πιθανότερη από την παρουσία δύο ή περισσότερων εκτόπων τιμών σε μια μικρή ομάδα δεδομένων.

Πίνακας 3.7. Θεωρητικές τιμές QP σε διάφορες στάθμες εμπιστοσύνης.

n	90%	95%	99%
5	0,644	0,732	0,867
6	0,448	0,532	0,695
7	0,336	0,410	0,555
8	0,265	0,325	0,458
9	0,220	0,271	0,384
10	0,185	0,231	0,332
11	0,163	0,204	0,297
12	0,145	0,181	0,266
13	0,129	0,162	0,235
14	0,118	0,149	0,219



Σχήμα 3-6. *Σημειογράμματα ομάδων σημείων (n = 5 έως 12) με χαρακτηριζόμενες ως έκτοπες τιμές (μαύροι κύκλοι) σε διάφορα υποδιαστήματα στάθμης εμπιστοσύνης.*

Παράδειγμα 3-7. Η ομάδα 8 μετρήσεων της ίδιας ποσότητας είναι η ακόλουθη (κατά αυξανόμενη τιμή): 5,62, 7,31, 7,42, 7,66, 7,91, 8,01, 8,22 και 9,55. Να εξεταστεί η ενδεχόμενη παρουσία μίας ή δύο εκτόπων τιμών σε στάθμη εμπιστοσύνης 95%.

Λύση. Πρώτα εξετάζεται η υπόθεση της παρουσίας μίας έκτοπης τιμής (η πλέον απέχουσα τιμή είναι η 5,62). Είναι Q = (7,31–5,62) / (9,55–5,62) = 0,430 < $Q_{theor,0,95}$ (= 0,526 για n = 8) και επομένως η 5,62 **δεν μπορεί να απορριφθεί**. Είναι σχεδόν εμφανές ότι η τιμή 5,62 “καλύπτεται” από την τιμή 9,55, απουσία της οποίας θα μπορούσε να απορριφθεί αφού: (7,31 –5,62) / (8,22–5,62) = 0,650 > $Q_{theor,0,95}$ (= 0,568 για n = 7).

Επειδή απορρίπτεται η υπόθεση της (αποκλειστικά) μίας έκτοπης μέτρησης, μπορεί να εξεταστεί πλέον η υπόθεση παρουσίας (αποκλειστικά) δύο εκτόπων μετρήσεων. Χρησιμοποιώντας την κατάλληλη για την περίπτωση αυτή Εξίσωση 3-13, έχουμε: QP = [(7,31–5,62) / (8,22–5,62)] × [(9,55–8,22) / (9,55–7,31)] = 0,386 > $QP_{0,95}$ (= 0,325 για n = 8), οπότε και οι δύο τιμές 5,62 και 9,55 **μπορούν να απορριφθούν** με πιθανότητα ≥95% ότι δεν απορρίπτεται ζεύγος "έγκυρων" τιμών.

---

4. ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ (ANOVA)

4.1. Εισαγωγή

Πολύ συχνά παρουσιάζεται η ανάγκη σύγκρισης των μέσων τιμών αναλυτικών αποτελεσμάτων για ανεύρεση σημαντικής διαφοράς (σε κάποια δεδομένη στάθμη εμπιστοσύνης) μεταξύ ομάδων μετρήσεων. Τυπικές περιπτώσεις:

α) Σύγκριση μέσων τιμών μετρήσεων που ελήφθησαν από το ίδιο δείγμα αλλά με διαφορετικές αναλυτικές τεχνικές.

β) Σύγκριση μέσων τιμών μετρήσεων από διαφορετικά δείγματα με την ίδια αναλυτική τεχνική, προκειμένου να διαπιστωθεί ότι τα δείγματα διαφέρουν ή όχι ως προς την περιεκτικότητα του μετρούμενου συστατικού.

Υπενθυμίζεται ότι εάν οι ομάδες μετρήσεων είναι μόνο δύο (με n_1 και n_2 μετρήσεις και αντίστοιχες μέσες τιμές $\overline{x}_1$ και $\overline{x}_2$) και με την προϋπόθεση ότι οι διακυμάνσεις των μετρήσεων δεν διαφέρουν σημαντικά (τούτο διαπιστώνεται με την δοκιμασία F και είναι πολύ πιθανό να μην ισχύει ιδιαίτερα στην (α) περίπτωση), τότε εφαρμόζεται η αντίστοιχη t-δοκιμασία: Υπολογίζεται η **συνδυασμένη τυπική απόκλιση** $s_{1\text{-}2}$ (pooled standard deviation) από τη σχέση:

$$s_{1\text{-}2}^2 = [\ (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\ ] / (n_1 + n_2 - 2) \qquad (4\text{-}1)$$

στη συνέχεια υπολογίζεται η πειραματική τιμή t_{exp} από τη σχέση

$$t_{exp} = \frac{\left|\overline{x}_1 - \overline{x}_2\right|}{s_{1-2}\sqrt{1/n_1 + 1/n_2}} \qquad (4\text{-}2)$$

η οποία συγκρίνεται κατά τα γνωστά με την θεωρητική τιμή t_{theor} για τη δεδομένη στάθμη εμπιστοσύνης και $n_1 + n_2 - 2$ βαθμούς ελευθερίας (υποκεφ. 3-4).

Η κατάσταση φαίνεται πιο πολύπλοκη εάν πρόκειται να συγκριθούν μεταξύ τους μέσες τιμές περισσότερων από δύο ομάδων μετρήσεων. Στις περιπτώσεις αυτές χρησιμοποιείται μία πανίσχυρη γενική στατιστική δοκιμασία (στην πραγματικότητα οικογένεια δοκιμασιών) γνωστή ως **ανάλυση διακύμανσης** (*AN*alysis *O*f *VA*riance), γνωστότερη ως **ANOVA**. Γενικός σκοπός της ANOVA είναι η ανεύρεση και η στατιστική σύγκριση των επιμέρους πηγών διακύμανσης (ή αβεβαιότητας), που συνεισφέρουν στην ολική διακύμανση (ή αβεβαιότητα) αποτελέσματος το οποίο προκύπτει από συνδυασμό ομάδων μετρήσεων.

Η ANOVA θα απαντήσει στο ερώτημα: **"Εάν δεχτούμε ότι όλες οι μετρήσεις προέρχονται από τον ίδιο πληθυσμό, τότε η διακύμανση που παρατηρείται μεταξύ των μέσων τιμών των επιμέρους ομάδων μετρήσεων είναι η στατιστικώς αναμενόμενη με βάση τις διακυμάνσεις των μετρήσεων σε κάθε ομάδα ή μήπως είναι μεγαλύτερη;"** Είναι προφανές ότι εάν η διακύμανση των μέσων τιμών είναι μεγαλύτερη, τότε η παραδοχή ότι όλες οι μετρήσεις προέρχονται από τον ίδιο πληθυσμό δεν ισχύει και επομένως υπάρχει κάποιος παράγοντας που διαφοροποιεί τις μετρήσεις από ομάδα σε ομάδα.

Τυπικές περιπτώσεις εφαρμογής της ANOVA:

- Σύγκριση μέσων συγκεντρώσεων ενός συστατικού σε διάφορα δείγματα.
- Σύγκριση μέσων συγκεντρώσεων ενός συστατικού από ομάδες μετρήσεων που ελήφθησαν από διαφορετικούς αναλυτές ή από διάφορα εργαστήρια (αλλά με την ίδια τεχνική).
- Σύγκριση μέσων συγκεντρώσεων ενός συστατικού σε δείγμα που υπέστη διάφορους τρόπους συντήρησης ή προεπεξεργασίας πριν από το στάδιο της μέτρησης.

Σε όλες τις περιπτώσεις δεχόμαστε ότι οι πηγές που συνεισφέρουν στη διακύμανση των μετρούμενων μέσων τιμών είναι:

1. Τα **τυχαία σφάλματα** κατά τη συνολική διαδικασία των μετρήσεων.

2. Οι **ελεγχόμενοι** (controlled) ή **καθορισμένου αποτελέσματος** (fixed-effect) **παράγοντες** (factors).

Η πρώτη πηγή είναι η βασική αιτία που προκαλεί τις διαφοροποιήσεις των μετρήσεων στην ίδια ομάδα μετρήσεων και είναι πάντοτε παρούσα. Η δεύτερη πηγή είναι εκείνη που εξετάζεται εάν υπάρχει και επιδρά σημαντικά στις λαμβανόμενες μέσες τιμές, δηλαδή εάν διαφοροποιεί πληθυσμιακά το σύνολο των μετρήσεων. Στις προηγούμενες τυπικές περιπτώσεις εφαρμογής της ANOVA, οι ελεγχόμενοι παράγοντες μπορεί να θεωρηθεί ότι είναι:

- Η διαφορετικότητα των δειγμάτων οφειλόμενη π.χ. σε δείγματα που ελήφθησαν από ανομοιογενές υλικό ή από διαφορετικές ημερομηνίες παραγωγής.
- Σε διαφορές της ικανότητας κάθε αναλυτή ή στο γενικό πρωτόκολλο ανάλυσης που ακολουθεί κάθε εργαστήριο.
- Ο τρόπος προεπεξεργασίας ή συντήρησης του δείγματος πριν από την ανάλυση.

Είναι προφανές ότι ο ελεγχόμενος παράγοντας δεν είναι κατ'ανάγκη **αριθμήσιμος** (π.χ. τυπικοί μη αριθμήσιμοι παράγοντες είναι: ο αναλυτής, το εργαστήριο, η χρησιμοποιούμενη μέθοδος, ενώ τυπικοί αριθμήσιμοι παράγοντες είναι: η θερμοκρασία φύλαξης ή ο χρόνος ζωής ενός αντιδραστηρίου, το βάθος λήψης δείγματος).

Η ANOVA επιτρέπει τη διάκριση της διακύμανσης, που οφείλεται στον ελεγχόμενο παράγοντα, από εκείνη που προέρχεται από καθαρά τυχαία σφάλματα. Ακόμη επιτρέπει τη διαπίστωση του εάν μια διαφοροποίηση στον κάθε ελεγχόμενο παράγοντα επιφέρει στατιστικά σημαντική διαφοροποίηση των λαμβανόμενων μέσων τιμών.

Η τεχνική ANOVA μπορεί να επεκταθεί και σε περιπτώσεις που υπάρχει πιθανή αλληλεπίδραση μεταξύ των παραγόντων αυτών, π.χ. διαφορετικός τρόπος συντήρησης δείγματος, σε συνδυασμό με τον τρόπο προεπεξεργασίας του δείγματος και τις συνθέσεις ενός ή περισσότερων από τα χρησιμοποιούμενα αντιδραστήρια. Η τεχνική που ακολουθείται στις περιπτώσεις αυτές ονομάζεται **παραγοντική ανάλυση** (factorial analysis).

Εάν κάθε φορά υπάρχει μόνο ένας ελεγχόμενος παράγοντας η τεχνική ονομάζεται **μονόδρομη** (one-way) ANOVA και αποτελεί την απλούστερη μορφή της δοκιμασίας.

4.2. Μονόδρομη ANOVA

Η απλούστερη περίπτωση εφαρμογής ANOVA αφορά σύγκριση m μέσων τιμών που προκύπτουν από m ομάδες μετρήσεων (ή δείγματα) κάθε μία των οποίων αποτελείται από n_j επιμέρους μετρήσεις (ο αριθμός μετρήσεων n_j μπορεί να είναι διαφορετικός σε κάθε ομάδα, δηλ. η 1η ομάδα (δείγμα) μπορεί να αποτελείται από 5 μέτρησεις, η 2η ομάδα από 3 μετρήσεις κλπ.) όπως φαίνεται στον παρακάτω πίνακα:

	Ομάδα μέτρησης				
	1	2	..	..	m
Μέτρηση 1	x_{11}	x_{12}	..	..	x_{1m}
Μέτρηση 2	x_{21}	x_{22}	..	..	x_{2m}
..	..	..		..	..
..	..	..		..	..
..	..	..	..	..	..
Μέτρηση n_j	x_{nj1}	x_{nj2}	..	..	x_{njm}

[**Παρατήρηση:** Αν η τιμή n_j ήταν ίδια σε όλες τις ομάδες (περίπτωση πλήρους πίνακα) θα μπορούσαμε να καταλήξουμε σε συγκεκριμένους τύπους. Ωστόσο είναι προτιμότερο να εξετάσουμε τη γενικότερη περίπτωση διαφορετικού αριθμού μετρήσεων ανά ομάδα, διότι κάτι τέτοιο είναι πολύ πιο πιθανό να συμβεί στη πράξη.]

Η μηδενική υπόθεση είναι:

$$H_o: \ \mu_1 = \mu_2 = \mu_3 = ... = \mu_m \tag{4-3}$$

δηλαδή: τα αποτελέσματα όλων των ομάδων μετρήσεων συμπίπτουν και ο ελεγχόμενος παράγοντας που διαφοροποιεί τη μία ομάδα μετρήσεων από την άλλη, δεν διαφοροποιεί σημαντικά τις μέσες τιμές (οι όποιες παρατηρούμενες διαφορές οφείλονται καθαρά και μόνο στα τυχαία σφάλματα κατά τις μετρήσεις).

Η εναλλακτική υπόθεση είναι:

$$\begin{aligned} H_a: \ & \mu_1 \neq \mu_2 = \mu_3 = ... \ \mu_m \\ & \mu_1 = \mu_2 \neq \mu_3 \neq ... \ \mu_m \quad \text{ή} \quad (\text{κ.λπ.}) \\ & \mu_1 \neq \mu_2 \neq \mu_3 \neq ... \ \mu_m \end{aligned} \tag{4-4}$$

δηλαδή υπάρχει τουλάχιστον μία διαφοροποίηση στις μέσες τιμές, η οποία οφείλεται σε επίδραση του ελεγχόμενου παράγοντα.

Η σύγκριση των μέσων τιμών ανά ζεύγη με δοκιμασία t είναι εφικτή αλλά επίπονη (απαιτούνται συνολικά $(m-1) \times m / 2$ δοκιμασίες t) και επιπλέον δεν θα παρείχε την πληρότητα των απαντήσεων που παρέχει η μονόδρομη ANOVA. Η τελευταία πραγματοποιείται "μια και έξω" στο σύνολο των ως άνω μετρήσεων και διαπιστώνει άμεσα την αποδοχή ή όχι της μηδενικής υπόθεσης (και όχι μόνο).

Δοθέντων των αποτελεσμάτων υπό την μορφή του ως άνω πίνακα $m \times n$ (όχι κατ' ανάγκη πλήρους) υπολογίζονται οι ακόλουθες ποσότητες που ομαδοποιούνται ως εξής:

- Παράμετροι που αφορούν κάθε ομάδα μετρήσεων
- Παράμετροι που αφορούν το σύνολο των μετρήσεων

και είναι οι ακόλουθες:

A) ΓΙΑ ΚΑΘΕ ΟΜΑΔΑ n_j ΜΕΤΡΗΣΕΩΝ ($1 \leq j \leq m$)

1) Το **άθροισμα** (sum) των μετρήσεων:

$$A_j = \sum_{i=1}^{n_j} x_i \tag{4-5}$$

2) Η **μέση τιμή της ομάδας** (group average):

$$\bar{x}_j = A_j / n_j \tag{4-6}$$

3) Η **διακύμανση της ομάδας** (group variance):

$$v_j = s_j^2 = \sum_{i=1}^{n_j} (\bar{x}_j - x_i)^2 / (n_j - 1) \tag{4-7}$$

4) Το **άθροισμα των τετραγώνων των υπολοίπων** (sum of squared residuals):

$$\sum_{i=1}^{n_j} r_i^2 = \sum_{i=1}^{n_j} (\bar{x}_j - x_i)^2 = v_j (n_j - 1) \tag{4-8}$$

B) ΓΙΑ ΤΟ ΣΥΝΟΛΟ ΤΩΝ ΜΕΤΡΗΣΕΩΝ

1) Ο συνολικός αριθμός μετρήσεων:

$$N = n_1 + n_2 + ... + n_m \tag{4-9}$$

2) Το **μεγάλο άθροισμα** (grand sum):

$$A = A_1 + A_2 + ... + A_m \tag{4-10}$$

3) Η **μεγάλη μέση τιμή** (grand average):

$$\bar{\bar{x}} = A/N \tag{4-11}$$

4) Το **άθροισμα των τετραγώνων των υπολοίπων** (sum of squared residuals):

$$S_1 = \sum_{i=1}^{n_{j1}} r_1^2 + \sum_{i=1}^{n_{j2}} r_2^2 + \ldots + \sum_{i=1}^{n_{jm}} r_m^2 = \sum_{k=1}^{m}\left(\sum_{i=1}^{n_k} r_i^2\right) \tag{4-12}$$

5) Το **άθροισμα των ζυγισμένων τετραγώνων** (sum of weighted squares):

$$S_2 = n_1(\bar{\bar{x}} - \bar{x}_1)^2 + n_2(\bar{\bar{x}} - \bar{x}_2)^2 + \ldots + n_m(\bar{\bar{x}} - \bar{x}_m)^2 = \sum_{k=1}^{m} n_k(\bar{\bar{x}} - \bar{x}_k)^2 \tag{4-13}$$

6) Το **ολικό άθροισμα τετραγώνων** (total sum of squares)

$$S_T = S_1 + S_2 \tag{4-14}$$

[**Παρατήρηση:** Το ολικό άθροισμα των τετραγώνων συμπίπτει και θα μπορούσε να υπολογιστεί με άθροιση των τετραγώνων των διαφορών όλων τιμών x (N τιμές) από την μεγάλη μέση τιμή ($\bar{\bar{x}}$), τούτο όμως δεν είναι απαραίτητο, αφού ήδη είναι στη διάθεσή μας (για τις ανάγκες της δοκιμασίας) οι μερικές τιμές S_1 και S_2.]

Γ) ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ANOVA

Συμπληρώνεται ο ακόλουθος πίνακας ANOVA :

Πίνακας 4.1. Βασικές υπολογιζόμενες παράμετροι μονόδρομης ANOVA

Άθροισμα τετραγώνων	Βαθμοί ελευθερίας	Διακύμανση	Πηγή διακύμανσης - Ονοματολογία
S_1	$f_1 = N - m$	$V_1 = S_1 / f_1$	Ενδοδειγματική διακύμανση (within-sample variance)
S_2	$f_2 = m - 1$	$V_2 = S_2 / f_2$	Διαδειγματική διακύμανση (between-sample variance)
$S_T = S_1 + S_2$	$f_T = N - 1$	$V_T = S_T / f_T$	Ολική διακύμανση (total variance)

Στη συνέχεια εξετάζεται η ισχύς της μηδενικής υπόθεσης (4-1) με αναζήτηση στατιστικώς σημαντικής διαφοράς μεταξύ της **ενδοδειγματικής** (within sample) διακύμανσης V_1 και της **διαδειγματικής** (between samples) διακύμανσης V_2.

Η δοκιμασία γίνεται κατά τα γνωστά (δοκιμασία Snedecor ή δοκιμασία F) σε μία στάθμη εμπιστοσύνης (συνήθως 95%) (ή -ισοδύναμα- με πιθανότητα P σφάλματος 1ου είδους 0,05). Η πειραματική τιμή F (λόγος V_2/V_1) συγκρίνεται με τη θεωρητική τιμή F με δοκιμασία “ενός άκρου” (one tailed F-test). Η δοκιμασία F “ενός άκρου” (και όχι της συνήθους δοκιμής “δύο άκρων”) επιβάλλεται, επειδή εξετάζεται το εάν η διαδειγματική διακύμανση είναι σημαντικώς μεγαλύτερη (και όχι απλώς διαφορετική) από την ενδοδειγματική διακύμανση ή εάν οι δύο διακυμάνσεις είναι ίσες (και οι όποιες διαφορές μπορούν να θεωρηθούν μόνο τυχαίες).

Οι επιπτώσεις που θα έχει η θετική ή αρνητική έκβαση της δοκιμασίας F συνοψίζονται στον Πίνακα 4.2.

Πίνακας 4.2. Σύνοψη συμπερασμάτων σε περίπτωση ισχύος ή μη της μηδενικής υπόθεσης στη μονόδρομη ANOVA

Η μηδενική υπόθεση ισχύει ($V_1 = V_2$)	**Η μηδενική υπόθεση δεν ισχύει ($V_1 < V_2$)**
Ο ελεγχόμενος παράγοντας δεν επιδρά στο αποτέλεσμα.	Ο ελεγχόμενος παράγοντας επιδρά στο αποτέλεσμα.
Όλες οι μετρήσεις προέρχονται από τον ίδιο ("ενιαίο") πληθυσμό και οι διαφορές στις μέσες τιμές οφείλονται μόνο σε τυχαία σφάλματα στο στάδιο των μετρήσεων.	Οι μετρήσεις δεν προέρχονται από τον ίδιο πληθυσμό και οι διαφορές στις μέσες τιμές οφείλονται σε επίδραση του ελεγχόμενου παράγοντα.
Όλες οι μετρήσεις μπορούν να συνενωθούν σε μια ενιαία ομάδα (ως προερχόμενες από τον ίδιο πληθυσμό).	Δεν επιτρέπεται η συνένωση των μετρήσεων σε μία ενιαία ομάδα, αλλά παραμένουν στις αρχικές ομάδες, μέχρις ότου διαπιστωθούν οι ομάδες στις οποίες επέδρασε ο ελεγχόμενος παράγοντας, όποτε μπορεί να ακολουθήσει νέα ομαδοποίηση.
Η μεγάλη μέση τιμή ($\bar{\bar{x}}$) μπορεί να δοθεί ως τελικό (ενιαίο) αποτέλεσμα της ανάλυσης.	Η μεγάλη μέση τιμή ($\bar{\bar{x}}$) και η ολική διακύμανση (V_T) δεν έχουν καμία στατιστική σημασία, αφού έχει απορριφθεί η ιδέα του "ενιαίου" πληθυσμού και δεν τις λαμβάνουμε πλέον υπόψη.
Η ολική διακύμανση V_T αποτελεί και τη διακύμανση του δείγματος και εκτιμήτρια παράμετρο διακύμανσης του πληθυσμού. Αντίστοιχα, η τυπική απόκλιση του δείγματος μπορεί να υπολογιστεί από τη σχέση: $s_T = \sqrt{V_T}$.	Πρέπει να δοθούν χωριστά οι τιμές V_1 και V_2 (ή αντίστοιχα, οι τιμές των τυπικών αποκλίσεων: $s_1 = \sqrt{V_1}$. $s_2 = \sqrt{V_2}$).

Παράδειγμα 4-1. Σε δείγματα Α, Β, Γ και Δ θαλασσίου ύδατος που ελήφθησαν στην ίδια τοποθεσία, αλλά σε βάθη: 1, 5, 20 και 100 μέτρων προσδιορίστηκε η περιεκτικότητά τους σε Br (με την ίδια φωτομετρική μέθοδο) και τα αποτελέσματα (σε mg Br/L) πολλαπλών μετρήσεων σε κάθε δείγμα ήταν τα ακόλουθα:

Δείγμα Α (βάθος: 1 m)	Δείγμα Β (βάθος: 5 m)	Δείγμα Γ (βάθος: 20 m)	Δείγμα Δ (βάθος: 100 m)
69,1	63,5	64,0	69,9
70,9	68,4	61,3	71,2
67,3	63,1	64,2	66,5
65,8	69,9	61,4	
69,7		64,6	
		69,1	
		69,2	

Να εξεταστεί εάν το βάθος δειγματοληψίας (ελεγχόμενος παράγοντας) έχει κάποια επίδραση στην τιμή της συγκέντρωσης βρωμίου.

Λύση. Από τις παραπάνω τιμές υπολογίζονται οι ακόλουθες τιμές κατά ομάδα: [**Παρατήρηση:** Σε όλους τους υπολογισμούς χρησιμοποιούνται 9-10 σημαντικά ψηφία, αλλά τα αποτελέσματα χάριν ευκολίας και καλύτερης παρουσίασης αναγράφονται εδώ στρογγυλεμένα στα 3-5 σημαντικά ψηφία. Για τον λόγο αυτό στο κείμενο αυτό μπορεί να παρουσιάζονται κάποιες μικρές αποκλίσεις στους υπολογισμούς των αθροισμάτων κ.α., οι οποίες βέβαια στην πραγματικότητα δεν υφίστανται].

Παράμετρος	Δείγμα Α	Δείγμα Β	Δείγμα Γ	Δείγμα Δ
Άθροισμα :	342,8	264,9	453,8	207,6
Μέση τιμή :	68,56	66,225	64,829	69,2
Διακύμανση :	4,068	11,809	10,449	5,89
Άθρ. τετραγώνων υπολοίπων :	16,272	35,4275	62,6943	11,78

και στη συνέχεια οι ακόλουθες γενικές ποσότητες:

Συνολικός αριθμός μετρήσεων: $N = 5 + 4 + 7 + 3 = \mathbf{19}$

Μεγάλο άθροισμα: $A = 342{,}8 + 264{,}9 + 453{,}8 + 207{,}6 = \mathbf{1269{,}1}$

Μεγάλη μέση τιμή: $\bar{\bar{x}} = 1269{,}1/19 = \mathbf{66{,}79}$

Άθροισμα τετραγ. υπολοίπων: $S_1 = 16{,}272+35{,}4275+62{,}6943+11{,}78 = \mathbf{126{,}17}$

Άθροισμα ζυγισμ.τετραγώνων: $S_2 = 5(66{,}79 - 68{,}56)^2 + 4(66{,}79 - 66{,}225)^2 + 7(66{,}79 - 64{,}829)^2 + 3(66{,}79 - 69{,}2)^2 = \mathbf{61{,}284}$

Ολικό άθροισμα τετραγώνων: $S_T = 126{,}17 + 61{,}284 = \mathbf{187{,}45}$

Ενδοδειγματική διακύμανση: $V_1 = S_1 / (19 - 4) = \mathbf{8{,}411}$ (within samples variance)

Διαδειγματική διακύμανση: $V_2 = S_2 / (4 - 1) = \mathbf{20{,}428}$ (between samples variance)

Ολική διακύμανση: $V_T = S_T / (19 - 1) = \mathbf{10{,}414}$ (total variance)

Η πειραματική τιμή F_{exp} είναι:

$F_{exp} = V_2 / V_1 = 20{,}428 / 8{,}411 = \mathbf{2{,}429}$

Από τους πίνακες αναζητείται η τιμή $F_{theor(3,\ 15,\ 0,05)}$ “ενός άκρου”, δηλ. η θεωρητική τιμή F για 15 βαθμούς ελευθερίας (διακύμανση παρανομαστή) και 3 βαθμούς ελευθερίας (διακύμανση αριθμητή) και με μέγιστη πιθανότητα σφάλματος (υπενθυμίζεται ότι ως σφάλμα εννοείται η εσφαλμένη απόρριψη της μηδενικής υπόθεσης -σφάλμα 1ου είδους-) 0,05 (ή ισοδύναμα: στάθμη εμπιστοσύνης: 95%)

Είναι $F_{theor(3,\ 15,\ 0,05)} = 3{,}29$ ($>F_{exp} = 3{,}093$) οπότε η μηδενική υπόθεση ισχύει και επομένως συμπεραίνουμε ότι:

Η περιεκτικότητα των δειγμάτων σε βρώμιο δεν εξαρτάται από τον ελεγχόμενο παράγοντα (βάθος δειγματοληψίας), η μέση περιεκτικότητα είναι 66,8 mg/L (δηλαδή ίση με τη μεγάλη μέση τιμή) και η τυπική απόκλιση των μετρήσεων είναι $\sqrt{V_T} = \sqrt{10{,}414} = 3{,}3$ mg/L.

Σε περίπτωση που διαπιστωθεί ότι η μηδενική υπόθεση απορρίπτεται, η ANOVA μπορεί να συνεχιστεί έτσι, ώστε να διαπιστωθεί σε ποια ή ποιες περιπτώσεις υπάρχει σημαντική διαφοροποίηση στις μέσες τιμές. Στη συνέχεια μπορούν να εξαχθούν τα σχετικά συμπεράσματα ως προς την επίδραση του ελεγχόμενου παράγοντα και ανάλογα με τη φύση (αριθμήσιμη ή μη) του τελευταίου. Ένας σχετικά απλός τρόπος δείχνεται στο επόμενο παράδειγμα.

Παράδειγμα 4-2. Να εξεταστεί το παράδειγμα 4-1 αν οι μετρήσεις των δειγμάτων Α και Β ήσαν:

Δείγμα Α : 62,1 64,3 62,1 65,8 61,7
Δείγμα Β : 63,7 62,4 64,1 63,3

(των υπολοίπων δειγμάτων οι μετρήσεις μένουν ως έχουν).

Λύση. Πραγματοποιώντας τους υπολογισμούς (δεν αναγράφονται τα επιμέρους στοιχεία) όπως στο παράδειγμα 4-1, προκύπτουν τα ακόλουθα στοιχεία:

Δείγμα	Μέση τιμή (mg Br/L)	Τυπική απόκλιση = $\sqrt{V}$ (mg Br/L)
Α	63,2	1,778
Β	63,375	0,7274
Γ	64,829	3,232
Δ	69,2	2,427

Άθροισμα τετραγώνων υπολοίπων: $S_1 = 88{,}7$

Άθροισμα ζυγισμένων τετραγώνων: $S_2 = 79{,}0$

Ολικό άθροισμα τετραγώνων: $S_T = 167{,}0$

Ενδοδειγματική διακύμανση: $V_1 = S_1 / (19 - 4) = 5{,}91$

Διαδειγματική διακύμανση: $V_2 = S_2 / (4 - 1) = 26{,}33$

Ολική διακύμανση: $V_T = S_T / (19 - 1) = 9{,}32$

Η πειραματική τιμή F είναι: $F_{exp} = V_2 / V_1 = 26{,}33 / 5{,}91 = 4{,}453$

Επειδή είναι $F_{exp} > F_{theor(3,\ 15,\ 0{,}05)} = 3{,}29$ η μηδενική υπόθεση απορρίπτεται (διακινδυνεύοντας πιθανότητα εσφαλμένης απόρριψης -δηλ., σφάλμα 1ου είδους- μικρότερη από 0,05), επομένως οι μέσες τιμές διαφέρουν σημαντικά.

Στη συνέχεια πρέπει να εξεταστεί που παρουσιάζεται η διαφοροποίηση αυτή. Για τον σκοπό αυτό οι μέσες τιμές αναγράφονται κατά αύξουσα σειρά. Στο Παράδειγμα 4-2 η σειρά αυτή συμπίπτει με τη σειρά βάθους δειγματοληψίας (που είναι ο ελεγχόμενος παράγοντας και στο συγκεκριμένο παράδειγμα είναι αριθμήσιμος), δεν σημαίνει όμως ότι τούτο είναι απαραίτητο να συμβαίνει πάντοτε. Επιπλέον υπολογίζονται οι μεταξύ τους διαφορές:

Δείγμα	Μέση τιμή	Διαφορά
Α	63,2	
		0,2
Β	63,4	
		1,4
Γ	64,8	
		4,4
Δ	69,2	

Υπολογίζεται η ελάχιστη σημαντική διαφορά (least significant difference, lsd) που παρέχεται από την εξίσωση:

$$\mathrm{lsd} = s_1 \sqrt{2/f_2} \times t_{f1} \qquad (4\text{-}13)$$

όπου s_1 είναι η ενδοδειγματική τυπική απόκλιση, f_1 και f_2 οι ενδοδειγματικοί και οι διαδειγματικοί βαθμοί ελευθερίας ($f_1 = N - m$, $f_2 = m - 1$) και t_{f1} η θεωρητική τιμή t για f_1 βαθμούς ελευθερίας στην προκαθορισμένη στάθμη εμπιστοσύνης (95% ή -ισοδύναμα- μέγιστη πιθανότητα σφάλματος 1ου είδους 0,05).

Στο συγκεκριμένο παράδειγμα είναι:

$$\mathbf{lsd} = 2{,}43\sqrt{2/3} \times 2{,}13 = 4{,}23$$

Στη συνέχεια εξετάζεται μεταξύ ποιών ομάδων (δειγμάτων) μετρήσεων η διαφορά των μέσων τιμών είναι ίση ή μεγαλύτερη από την **lsd**. Διαπιστώνεται ότι η διαφορά μεταξύ των μέσων τιμών των ομάδων Γ και Δ υπερβαίνει την **lsd**, οπότε καταλήγουμε στο εξής συμπέρασμα:

Η συγκέντρωση βρωμίου στο δείγμα Δ (69,2 mg/L) είναι σημαντικά διαφορετική (μεγαλύτερη στο συγκεκριμένο παράδειγμα) από εκείνη των δειγμάτων Α, Β, Γ, στα οποία μπορεί να θεωρηθεί ότι η συγκέντρωση βρωμίου είναι 64,0 mg/L (μετά τη συνένωση των τριών ομάδων μετρήσεων σε μία ενιαία ομάδα).

[**Παρατήρηση**: Υπάρχουν περιπτώσεις όπου απορρίπτεται μεν η μηδενική υπόθεση, αλλά στη συνέχεια διαπιστώνεται ότι καμία διαφορά δεν είναι μεγαλύτερη από την **lsd**. Για τις περιπτώσεις αυτές έχουν προταθεί πλέον σύνθετες δοκιμασίες.]

Υπολογισμοί ANOVA με τη βοήθεια υπολογιστή. Οι σχετικά επίπονοι υπολογισμοί για την εξαγωγή των συμπερασμάτων αποφεύγονται εφ' όσον πραγματοποιηθεί η ANOVA με τη βοήθεια υπολογιστή. Παρακάτω δείχνονται τα αποτελέσματα (για τα δύο προηγούμενα παραδείγματα) όπως παρουσιάζονται με το στατιστικό πρόγραμμα (STATMOST, DataMost Co., Salt Lake City, UT, USA):

Έκθεση αποτελεσμάτων ANOVA (δεδομένα Παραδείγματος 4-1)

```
StatMost for Windows        Tuesday, October 29, 1996    7:09:17 PM
--------------------------------------------------------------------
            ColName      Count    Mean      Std.Dev.    Std.Err.
         ---------------+-----+----------+-----------+----------
              A             5    68.5600     2.0169      0.9020
              B             4    66.2250     3.4364      1.7182
              C             7    64.8286     3.2325      1.2218
              D             3    69.2000     2.4269      1.4012
         ---------------+-----+----------+-----------+----------
                          One-Way ANOVA Results
====================================================================
     Source        DF        SS          MS          F          P
 ---------------+-----+----------+-----------+----------+----------
 Between Groups    3    61.2957    20.4319     2.4290     0.1057
 Within Groups    15   126.1738     8.4116
 ---------------+-----+----------+-----------+----------+----------
       Total      18   187.4695

**************************** The End ****************************
```

Έκθεση αποτελεσμάτων ANOVA (δεδομένα Παραδείγματος 4-2)

```
StatMost for Windows        Tuesday, October 29, 1996    7:11:31 PM
--------------------------------------------------------------------
            ColName      Count    Mean      Std.Dev.    Std.Err.
         ---------------+-----+----------+-----------+----------
              A             5   63.2000     1.7776      0.7950
              B             4   63.3750     0.7274      0.3637
              C             7   64.8286     3.2325      1.2218
              D             3   69.2000     2.4269      1.4012
         ---------------+-----+----------+-----------+----------
                          One-Way ANOVA Results
====================================================================
     Source        DF        SS          MS          F          P
  ---------------+-----+----------+-----------+----------+---------
  Between Groups    3    79.0035    26.3345     4.4533     0.0199
  Within Groups    15    88.7018     5.9135
  ---------------+-----+----------+-----------+----------+---------
       Total       18   167.7053

**************************** The End ****************************
```

Το αποτέλεσμα της ANOVA ουσιαστικά εντοπίζεται στην τιμή που αναγράφεται κάτω από το γράμμα P (: probability), που δείχνει την τιμή πιθανότητας σφάλματος 1ου, η οποία αντιστοιχεί στην πειραματική τιμή F (που παρέχεται επίσης μαζί με τις υπόλοιπες παραμέτρους). Έτσι, στην περίπτωση του παραδείγματος 4-1 είναι P = 0,1057, γεγονός που υποδηλώνει ότι η αν επιμείνουμε σε απόρριψη της μηδενικής υπόθεσης διακινδυνεύουμε πιθανότητα σφάλματος 1ου είδους 10,57% έναντι του μεγίστου 5% που επιβάλλει η επιλεγείσα στάθμη εμπιστοσύνης 95% (είναι φανερό ότι η μηδενική υπόθεση "αντέχει" οριακά και σε στάθμη εμπιστοσύνης 90%).

Στην περίπτωση ANOVA επί των δεδομένων του Παραδείγματος 4-2 είναι P = 0,0199. Τούτο σημαίνει ότι η μηδενική υπόθεση απορρίπτεται με πιθανότητα 1,99% σφάλματος 1ου είδους, δηλ. η μηδενική υπόθεση "δεν αντέχει" σε στάθμη εμπιστοσύνης 95% (αντίθετα, θα άντεχε σε πιθανότητα 1,00%, δηλ. αν είχε επιλεχθεί στάθμη εμπιστοσύνης 99%).

5. ΕΙΣΑΓΩΓΗ ΣΤΗ ΣΤΑΤΙΣΤΙΚΗ ΓΡΑΜΜΙΚΩΝ ΔΙΑΓΡΑΜΜΑΤΩΝ

Η γραμμική καμπύλη βαθμονόμησης ή καμπύλη αναφοράς (calibration curve ή calibration curve), όπως και τα διαγράμματα συσχέτισης αναλυτικών αποτελεσμάτων, τις περισσότερες φορές είναι γραμμικά διαγράμματα (της μορφής $y = αx + β$) και σχεδιάζονται συνήθως από ικανό αριθμό πειραματικών σημείων, τα οποία παρουσιάζουν μια διασπορά, με αποτέλεσμα μια αβεβαιότητα ως προς τις τιμές των παραμέτρων α και β. Η γνώση των στατιστικώς πιθανότερων τιμών α και β, όπως και των διακυμάνσεών τους έχει ιδιαίτερη σημασία κατά την αξιολόγηση των διαφόρων αναλυτικών μεθόδων.

5.1. Μέθοδος ελαχίστων τετραγώνων

Πολλές φορές είναι αναγκαία η γνώση της σχέσης $y = f(x, α, β, γ, ...)$, όπου y είναι η εξαρτημένη μεταβλητή, x η ανεξάρτητη μεταβλητή και α, β, γ, ... είναι παράμετροι που υπεισέρχονται στη σχέση αυτή. Εάν είναι γνωστή η μορφή της σχέσης, όπως και οι αριθμητικές τιμές των παραμέτρων, μπορούμε α) να σχεδιάσουμε την αντίστοιχη γραφική παράσταση και β) να υπολογίσουμε τις τιμές y που αντιστοιχούν σε ορισμένες τιμές x (ή αντιστρόφως), χωρίς να είναι αναγκαίο να καταφύγουμε, για το σκοπό αυτό, στην αντίστοιχη γραφική παράσταση.

Εάν είναι γνωστή μόνο η γενική μορφή της σχέσης f(x,α,β,γ,...) ή f(x) (χάριν απλότητας), τότε πολλές φορές από N πειραματικά ζεύγη μετρήσεων (y_1, x_1), (y_2, x_2), ..., (y_N, x_N), ζητείται ο υπολογισμός των παραμέτρων α, β, γ, ... Αποδεικνύεται ότι όταν ισχύουν οι εξής προϋποθέσεις:

1) Οι μετρούμενες τιμές του y εμπεριέχουν τυχαία (μόνο) σφάλματα με κανονική (κατά Gauss) κατανομή με τυπική απόκλιση ανεξάρτητη από τις τιμές x και

2) Οι τιμές x είναι απαλλαγμένες από σφάλματα,

τότε για τις στατιστικώς ορθότερες τιμές των παραμέτρων α, β, γ,... θα ισχύει η γενική σχέση

$$S = \sum_{i=1}^{N} \left[f(x_i) - y_i \right]^2 = \text{minimum} \qquad (5\text{-}1)$$

δηλαδή το άθροισμα των τετραγώνων των αποκλίσεων μεταξύ των υπολογιζόμενων (θεωρητικών) τιμών y από τη σχέση f(x) και των πειραματικών τιμών πρέπει να αποκτήσει την ελάχιστη δυνατή τιμή. Με βάση τη γενική αυτή αρχή είναι δυνατόν να ληφθούν ακριβείς σχέσεις, που συνδέουν τις πειραματικές τιμές y_1, y_2, ..., y_n και x_1, x_2,..., x_N με τις τιμές των παραμέτρων α, β, γ, ...

Η μέθοδος υπολογισμού των τιμών των υπεισερχόμενων παραμέτρων ονομάζεται **μέθοδος προσαρμογής ελάχιστων τετραγώνων** (least squares fit method).

Ιδιαίτερο ενδιαφέρον παρουσιάζει η εφαρμογή της μεθόδου ελαχίστων τετραγώνων στην περίπτωση των γραμμικών σχέσεων, δηλαδή εξισώσεων με τη μορφή

$$y = αx+β \qquad (5\text{-}2)$$

λόγω της συχνά απαντούμενης γραμμικής σχέσης μεταξύ της αναλυτικής πληροφορίας και της μετρούμενης παραμέτρου στην Ενόργανη Ανάλυση. Στην περίπτωση της παραπάνω γραμμικής εξίσωσης, η μέθοδος ονομάζεται **γραμμική προσαρμογή ελάχιστων τετραγώνων** (linear least squares fit) και η ευθεία που προκύπτει, γνωστή ως **ευθεία παλινδρόμησης** ή **συμμεταβολής** (regression line), παρέχει την άριστη γραμμική προσαρμογή των σημείων.

Η τιμή α ονομάζεται **κλίση** (slope) της ευθείας και η τιμή β ονομάζεται **τομή στον άξονα των τεταγμένων** ή **τομή "επί της αρχής"** (intercept) και παρέχονται από τις σχέσεις:

$$\alpha = \frac{\sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i \right)^2} \tag{5-3}$$

$$\beta = \frac{\sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} x_i y_i}{N \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i \right)^2} \tag{5-4}$$

5.2. Συντελεστής συσχέτισης

Από τις εξισώσεις (5-3) και (5-4) υπολογίζονται οι στατιστικώς ορθότερες τιμές των α (κλίση της ευθείας παλινδρόμησης) και β (τομή στον άξονα των τεταγμένων). Εάν ισχύει η Εξίσωση 5-2 θα πρέπει να ισχύει και η αντίστροφή της:

$$x = (1/\alpha)\, y - (\beta/\alpha) \tag{5-5}$$

Για τον υπολογισμό της τιμής (1/α) με τη μέθοδο των ελάχιστων τετραγώνων θα χρησιμοποιηθεί η Εξίσωση 5-3 με μόνη διαφορά την αμοιβαία αντικατάσταση των τιμών x και y, οπότε θα αποκτήσει την ακόλουθη μορφή

$$\alpha = \frac{\sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N \sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i \right)^2} \tag{5-6}$$

Προφανώς, εάν οι τιμές y ήσαν απαλλαγμένες από σφάλματα, θα ίσχυε ότι:

$$r = \left[\alpha(1/\alpha)\right]^{1/2} = 1 \tag{5-7}$$

Η παράμετρος r ονομάζεται **συντελεστής συσχέτισης** του Pearson (Pearson's correlation coefficient). Εάν υπεισέρχονται σφάλματα στις τιμές y, τότε η συσχέτιση των τιμών y και x γίνεται πιο "χαλαρή" και θα είναι $|r| < 1$, ενώ, όταν δεν υπάρχει καμία συσχέτιση, θα είναι $|r| = 0$.

Με συνδυασμό των σχέσεων (5-3), (5-6) και (5-7) προκύπτει η ακόλουθη αναλυτική έκφραση του συντελεστή συσχέτισης:

$$r = \frac{\sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{\sqrt{\left(N \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i \right)^2 \right)\left(N \sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i \right)^2 \right)}} \tag{5-8}$$

Ο συντελεστής συσχέτισης μπορεί να υπολογιστεί για μία καμπύλη αναφοράς, για να εξακριβωθεί ο βαθμός συσχέτισης μεταξύ της μετρούμενης φυσικής παραμέτρου και της συγκέντρωσης των πρότυπων διαλυμάτων ή μεταξύ των αποτελεσμάτων δύο μεθόδων.

Για τιμές 0,95 < | r | < 0,99 η συσχέτιση θεωρείται ικανοποιητική, ενώ τιμές | r | > 0,99 κατά κανόνα υποδηλώνουν πολύ καλή συσχέτιση. Με μετρήσεις ικανοποιητικής ακρίβειας και επαναληψιμότητας μπορούν να επιτευχθούν τιμές | r | >0,999. Δοκιμασία σημαντικότητας (τύπου Student's t) μπορεί να εφαρμοστεί με χρήση του ακόλουθου στατιστικού στοιχείου:

$$t = \frac{r}{\sqrt{(1-r^2)/(N-2)}} \qquad (5\text{-}9)$$

Η πειραματική τιμή t συγκρίνεται με τις θεωρητικές τιμές t, στην επιθυμητή στάθμη εμπιστοσύνης, με βαθμούς ελευθερίας $v = N - 2$.

Στην αναφερόμενη τιμή του r συνήθως κρατείται ένα στρογγυλεμένο ψηφίο μετά το τελευταίο 9, έτσι, π.χ., εάν οι επακριβώς υπολογιζόμενες τιμές του r είναι 0,9947281, 0,9987201 και 0,9999038, οι παρουσιαζόμενες τιμές είναι 0,995, 0,999 και 0,99990, αντιστοίχως.

Από τα προηγούμενα προκύπτει ότι η τιμή του συντελεστή συσχέτισης αποτελεί ενδεικτικό μέτρο του βαθμού διασποράς των πειραματικών σημείων ενός γραμμικού διαγράμματος, ωστόσο θα πρέπει να τονιστεί ότι η τιμή του r εξαρτάται και από τη σχετική κατανομή των σημείων στο διάγραμμα και σε καμία περίπτωση δεν πρέπει να χρησιμοποιείται ως ενδεικτικό στατιστικό στοιχείο της γραμμικότητας που συνδέει τα συσχετιζόμενα μεγέθη.

5.3. Αδυναμίες της μεθόδου ελαχίστων τετραγώνων

Πριν από την εφαρμογή της μεθόδου ελαχίστων τετραγώνων **συνιστάται πάντοτε η σχεδίαση του διαγράμματος με τα πειραματικά σημεία**, για τους ακόλουθους λόγους:

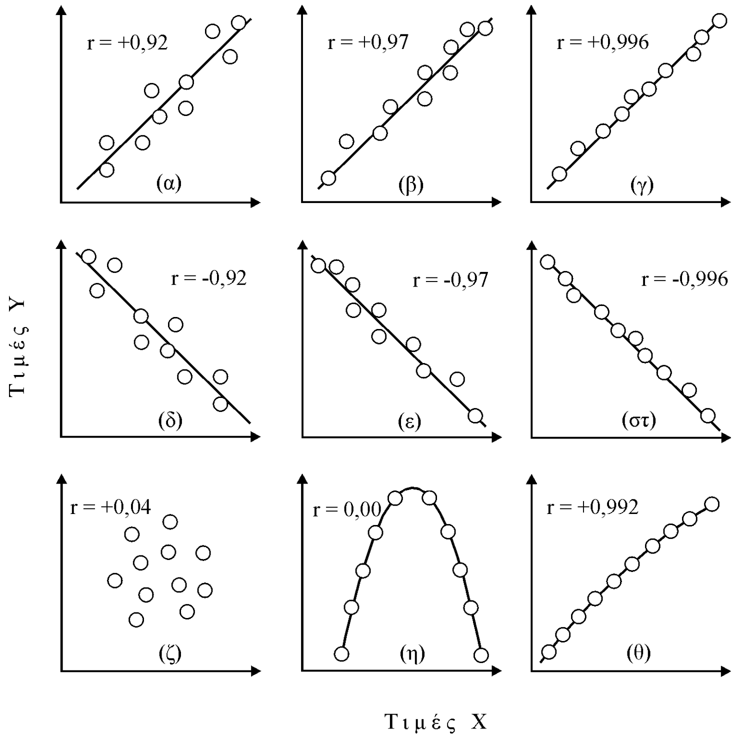
1) Από το διάγραμμα αποκαλύπτονται τα σημεία (ή ομάδες σημείων) τα οποία σαφώς εκτρέπονται από την αναμενόμενη ευθεία και απορρίπτονται πριν χρησιμοποιηθούν στους υπολογισμούς.

2) Αποκαλύπτεται τυχόν μη γραμμική συσχέτιση των μεγεθών. Οι παραπάνω διεργασίες έχουν σαφώς υποκειμενικό χαρακτήρα. Στατιστικές δοκιμασίες αποκλεισμού σημείων υπάρχουν, αλλά είναι εξαιρετικά δύσχρηστες και απαιτούν τη χρήση υπολογιστών.

Η μέθοδος των ελαχίστων τετραγώνων είναι εξαιρετικά ευαίσθητη στην παρουσία ισχυρώς αποκλινουσών μετρήσεων, που πρέπει να απομακρύνονται πριν από την εφαρμογή της. Έχουν προταθεί ανθεκτικές μέθοδοι προσαρμογής, όπως π.χ. εκείνες που βασίζονται στις διάμεσες τιμές των "μερικών κλίσεων" (ανά ζεύγη μετρήσεων), οι οποίες όμως μπορούν να πραγματοποιηθούν με τη βοήθεια υπολογιστών λόγω του μεγάλου αριθμού των υπολογισμών και διατάξεων των τιμών κατά αυξανόμενο μέγεθος (μέθοδος Theil). Στις μεθόδους αυτές δεν χρησιμοποιούνται "τύποι" αλλά υπολογιστικοί "αλγόριθμοι".

Ένα σημείο το οποίο απαιτεί ιδιαίτερη προσοχή είναι το ότι ικανοποιητική τιμή του r δεν επιβεβαιώνει γραμμική σχέση, αλλά ικανοποιητική συσχέτιση των μεγεθών x και y. Έτσι, π.χ. διάγραμμα με σημεία που αποδίδουν μία γραμμή με πολύ μικρή διασπορά, αλλά με μια μικρή καμπύλωση μπορεί να παρουσιάζει καλύτερη τιμή r, από ένα διάγραμμα σημείων μιας πραγματικά γραμμικής σχέσης με μεγαλύτερη διασπορά.

Στο Σχήμα 5-1 δείχνονται εννέα διαγράμματα με διαφορετικό βαθμό διασποράς των πειραματικών σημείων (x, y) και οι αντίστοιχες τιμές του συντελεστή συσχέτισης. Τα διαγράμματα (α)-(στ) αποδίδουν γραμμικές συσχετίσεις με θετική (α)-(γ) και αρνητική κλίση (δ)-(στ) με διάφορους βαθμούς διασποράς των σημείων γεγονός, που επηρεάζει τον συντελεστή συσχέτισης. Στο διάγραμμα (ζ) δείχνεται μια τυπική περίπτωση σχεδόν πλήρους έλλειψης συσχέτισης (r=0,04). Στο διάγραμμα (η) πάλι εμφανίζεται η τιμή r ίση προς το μηδέν, ενώ υπάρχει σαφής συσχέτιση μεταξύ των τιμών x και y. Τούτο οφείλεται στο ότι χρησιμοποιείται ο συντελεστής γραμμικής συσχέτισης σε μια συσχέτιση που είναι 2ου βαθμού (παραβολική) και στη διάταξη των σημείων. Το γεγονός αυτό αποδεικνύει ότι ποτέ δεν πρέπει η τιμή r να οδηγεί σε συμπεράσματα χωρίς παράλληλα να έχει σχεδιασθεί το αντίστοιχο διάγραμμα. Ανάλογα, στο διάγραμμα (θ) δείχνεται περίπτωση όπου υπάρχει μία σαφώς μη γραμμική συσχέτιση, παρ'όλα αυτά η τιμή r φαίνεται ικανοποιητικότερη γραμμική συσχέτιση σε σχέση με εκείνες των διαγραμμάτων (α), και (β) γεγονός που καταδεικνύει ότι η τιμή r δεν μπορεί να χρησιμεύσει ως τεκμήριο γραμμικότητας μεταξύ των συσχετιζόμενων ποσοτήτων.



Σχήμα 5-1. *Υπολογιζόμενοι συντελεστές γραμμικής συσχέτισης για διάφορους τύπους διαγραμμάτων.*

Παράδειγμα 5-1. Να βρεθεί η εξίσωση που προσαρμόζεται στα παρακάτω πειραματικά δεδομένα και να υπολογιστεί ο συντελεστής συσχέτισης

Τιμές x:	1,0	2,0	3,0	4,0	5,0	6,0	7,0	8,0
Τιμές y:	3,1	6,0	8,7	12,9	15,3	17,9	22,0	23,7

υποθέτοντας ότι η σχέση μεταξύ y και x είναι γραμμική.

Λύση. Υπολογίζουμε τις τιμές των x_i^2, y_i^2 και x_iy_i (προκειμένου να υπολογιστούν και τα αντίστοιχα αθροίσματα) και αναγράφουμε τα αποτελέσματα στον ακόλουθο πίνακα:

i	x_i	y_i	x_i^2	y_i^2	x_iy_i
1	1,0	3,1	1,0	9,61	3,1
2	2,0	6,0	4,0	36,00	12,0
3	3,0	8,7	9,0	75,69	26,1
4	4,0	12,9	16,0	166,41	51,6
5	5,0	15,3	25,0	234,09	76,5
6	6,0	17,9	36,0	320,41	107,4
7	7,0	22,0	49,0	484,00	154,0
8	8,0	23,7	64,0	561,69	189,6
N=8	$\Sigma x_i =$ 36,0	$\Sigma y_i =$ 109,6	$\Sigma x_i^2 =$ 204,0	$\Sigma y_i^2 =$ 1887,90	$\Sigma x_iy_i =$ 620,3

[**Παρατήρηση:** Κατά το στάδιο υπολογισμού των παραπάνω αθροισμάτων πρέπει να αποφεύγονται τα "στρογγυλεύματα" και να χρησιμοποιούνται τα ενδιάμεσα αριθμητικά αποτελέσματα με όσο το δυνατόν περισσότερα σημαντικά ψηφία].

Με αντικατάσταση των τιμών του παραπάνω πίνακα στις εξισώσεις (5-3), (5-4) και (5-8) έχουμε:

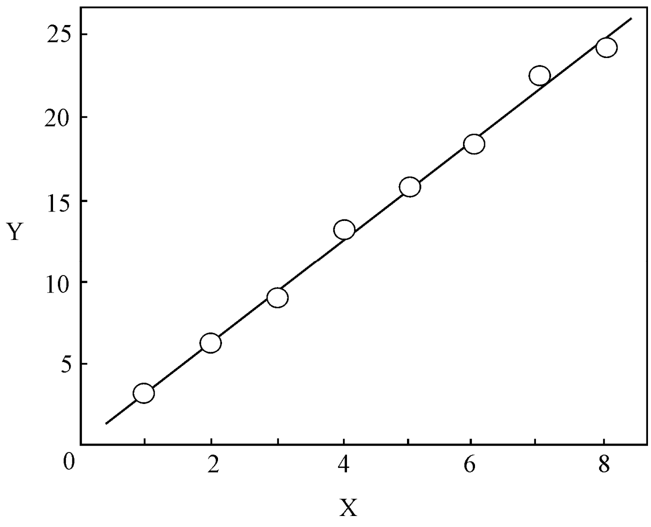
$$\alpha = \frac{(8\times 620{,}3)\text{-}(36{,}0\times 109{,}6)}{(8\times 204{,}0)-(36{,}0)^2} = 3{,}026 \qquad \beta = \frac{(204{,}0\times 109{,}6)\ \text{-}\ (36{,}0\times 620{,}3)}{(8\times 204{,}0)-(36{,}0)^2} = 0{,}0821$$

$$r = \frac{(8\times 620{,}3)\text{-}(36{,}0\times 109{,}6)}{\sqrt{[(8\times 204{,}0)-(36{,}0)^2][(8\times 1887{,}90)-(109{,}6)^2]}} = 0{,}9977 \approx \mathbf{0{,}998}$$

Επομένως η ζητούμενη εξίσωση είναι

$$\mathbf{y = 3{,}026x + 0{,}0821}$$

Η ευθεία παλινδρόμησης και τα πειραματικά σημεία δείχνονται στο Σχήμα 5-2. Είναι προφανής η πολύ καλή συσχέτιση των τιμών y και x, όπως αυτό καταφαίνεται και από την τιμή του συντελεστή συσχέτισης.



Σχήμα 5-2. *Ευθεία παλινδρόμησης των δεδομένων του παραδείγματος.*

5.4. Τυπική απόκλιση των παραμέτρων α και β

Σε πολλές περιπτώσεις εξίσου ή ακόμη και περισσότερο χρήσιμη από την ίδια την τιμή των παραμέτρων α και β είναι η τυπική τους απόκλιση, που εξαρτάται ως αναμένεται από τον διασκορπισμό (λόγω τυχαίων πάντοτε σφαλμάτων) των πειραματικών σημείων (x_i, y_i). Οι τυπικές αποκλίσεις της κλίσης α και της τομής στον άξονα των τεταγμένων β, παρέχονται από τις ακόλουθες εξισώσεις:

$$s_\alpha = s_{y/x}\sqrt{\frac{N}{N\sum_{i=1}^{N}x_i^2-\left(\sum_{i=1}^{N}x_i\right)^2}} \qquad (5\text{-}10)$$

$$s_\beta = s_{y/x}\sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}} \tag{5-11}$$

όπου

$$s_{y/x} = \sqrt{\frac{\sum_{i=1}^{N} y_i^2 - \frac{\left(\sum_{i=1}^{N} y_i\right)^2}{N} - \frac{1}{N}\cdot\frac{\left(N\sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i\right)^2}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}}{N-2}} \tag{5-12}$$

Το στατιστικό στοιχείο $s_{y/x}$ είναι η τυπική απόκλιση των υπολοίπων $y_i - y_{\theta,i}$, δηλαδή των διαφορών μεταξύ πειραματικών και θεωρητικών τιμών (που υπολογίζονται από την εξίσωση παλινδρόμησης, $y_{\theta,i} = \alpha x_i + \beta$).

5.5. Εφαρμογές στατιστικής γραμμικών διαγραμμάτων

5.5.1. Σύγκριση μεθόδων.

Τα διαγράμματα συσχέτισης των αναλυτικών αποτελεσμάτων μιας δοκιμαζόμενης μεθόδου ως προς εκείνα μιας αποδεκτής μεθόδου, ιδανικά θα πρέπει να παρουσιάζουν κλίση α = 1 και τομή στον άξονα των τεταγμένων β = 0. Στην πράξη η τιμές α και β διαθα φέρουν από τις ιδανικές τιμές.

Εφόσον είναι γνωστές οι τυπικές αποκλίσεις και των δύο τιμών, είναι δυνατή η εφαρμογή της δοκιμασίας t (σύγκριση πειραματικής μέσης τιμής με πραγματική), για να διαπιστωθεί εάν υπάρχει στατιστικώς σημαντική διαφορά της α από τη μονάδα (ένδειξη αναλογικού σφάλματος) ή της β από το μηδέν (ένδειξη σταθερού σφάλματος) ή και τα δύο (μικτό σφάλμα).

Οι βαθμοί ελευθερίας για την επιλογή της τιμής t_{theor} για τις παραπάνω δοκιμασίες είναι πάντοτε ν = Ν – 2 (Ν: ο αριθμός των πειραματικών ζευγών).

Παράδειγμα 5-2. Να συγκριθούν τα ακόλουθα αναλυτικά αποτελέσματα που ελήφθησαν με 12 διαφορετικά δείγματα με τη μέθοδο Α (εξακριβωμένα αξιόπιστη μέθοδο) με τα αντίστοιχα αποτελέσματα που ελήφθησαν με τη μέθοδο Β (υπό δοκιμασία νέα μέθοδος).

Μέθ. Α (x_i)	2,0	3,0	4,0	4,5	5,2	5,5	5,9	6,9	7,3	7,5	8,2	9,4
Μέθ. Β (y_i)	1,7	3,0	3,8	4,6	5,1	5,5	6,0	6,5	6,8	7,5	8,0	9,2

Λύση. Με εφαρμογή των προηγουμένων εξισώσεων υπολογίζονται οι συντελεστές της εξίσωσης παλινδρόμησης και οι αντίστοιχες τυπικές αποκλίσεις:

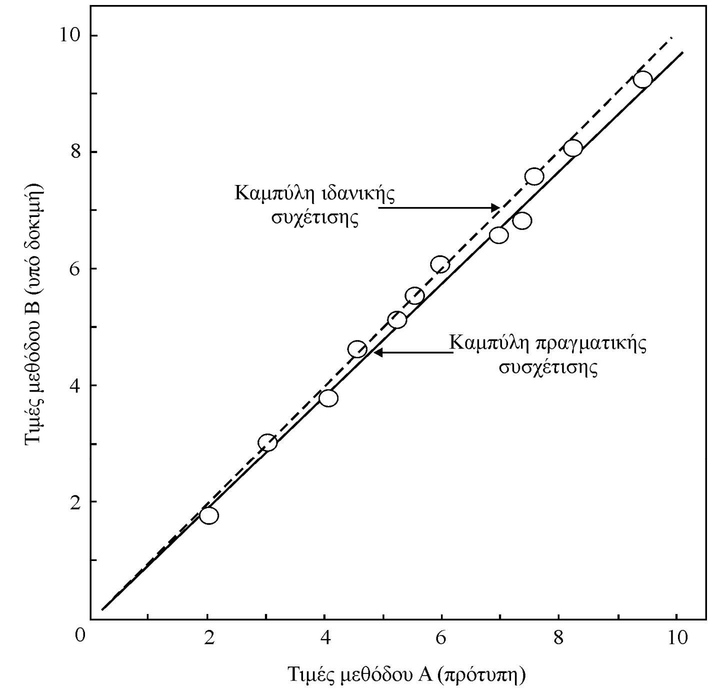
$\alpha = 0{,}9815 \quad s_\alpha = 0{,}0272$ και $\beta = -0{,}0344 \quad s_\beta = 0{,}1675$

Για να εξακριβωθεί εάν οι τιμές α και β διαφέρουν σημαντικά από τις επιθυμητές τιμές 1 και 0, αρκεί να υπολογιστεί εάν οι επιθυμητές (για ικανοποιητική συσχέτιση των μεθόδων Α και Β) τιμές βρίσκονται στο **διάστημα εμπιστοσύνης** (confidence interval) των α και β. Το διάστημα εμπιστοσύνης υπολογίζεται (σε στάθμη εμπιστοσύνης 95%) ως εξής:

Διάστημα εμπιστοσύνης (95%) α :
$\alpha \pm t_{theor}s_a = 0{,}9815 \pm (2{,}179\times0{,}0272) = 0{,}9815 \pm 0{,}0592$ δηλ. : 0,9222 έως 1,09997

Διάστημα εμπιστοσύνης (95%) β :
$\beta \pm t_{theor}s_b = -0{,}0344 \pm (2{,}179\times0{,}1675) = -0{,}0344 \pm 0{,}3650$ δηλ. : –0,3994 έως 0,3306

ως t_{theor} χρησιμοποιείται η τιμή t (Πίνακας 3.3) που αντιστοιχεί σε $\nu = N - 2 = 12 - 2 = 10$ βαθμούς ελευθερίας.



Σχήμα 5-3. *Καμπύλη συσχέτισης αναλυτικών αποτελεσμάτων των μεθόδων Α και Β.*

Από τα ως άνω αποτελέσματα προκύπτει ότι σε στάθμη εμπιστοσύνης 95% τα αποτελέσματα των δύο μεθόδων είναι ισοδύναμα, δηλ. δεν ανιχνεύεται αναλογικό ή σταθερό σφάλμα μεταξύ των δύο σειρών μετρήσεων. [**Παρατήρηση**: Εναλλακτικά θα μπορούσε να υπολογιστεί οι πειραματικές τιμές t (π.χ. για την κλίση: $t_{exp} = |(\alpha - 1)/s_a|$ και την τομή επί την αρχή: $t_{exp} = |(\beta - 0)\ /s_a| =|\ \beta/s_b\ |$ και σύγκριση των τιμών αυτών με τις θεωρητικές τιμές, στην επιθυμητή στάθμη εμπιστοσύνης και για $\nu = N - 2$ βαθμούς ελευθερίας]

5.5.2. Ανίχνευση τάσης

Μια ομάδα τιμών διαδοχικών μετρήσεων στο ίδιο δείγμα εξετάζεται για την ύπαρξη τάσης με ανάλογη δοκιμασία. Στην περίπτωση αυτή εξετάζεται αν σε διάγραμμα: μετρούμενη τιμή = α×(αύξοντας αριθμός μέτρησης) + β, η παράμετρος α διαφέρει σημαντικά από την ιδανική τιμή 0.

5.5.3. Ανίχνευση εκτόπων μετρήσεων σε γραμμικό διάγραμμα

Χρησιμοποιείται η τιμή $s_{y/x}$, που θεωρείται η ίδια σε όλη τη χρήσιμη περιοχή του διαγράμματος, εκτός εάν με επαναλαμβανόμενες μετρήσεις εκτιμηθεί τυπική απόκλιση διαφορετική από περιοχή σε περιοχή, κάτι αρκετά συνηθισμένο. Εάν η ύποπτη τιμή απέχει σημαντικά από τη θεωρητικά προβλεπόμενη τιμή, τότε μπορεί να εφαρμοστεί η δοκιμασία t για την απόρριψη της αποκλίνουσας τιμής ως έκτοπης και επανεκτίμηση των παραμέτρων α και β του διαγράμματος.

5.5.4. Σύγκριση κλίσεων δύο γραμμικών διαγραμμάτων

Πολλές φορές απαιτείται να συγκριθεί η κλίση δύο γραμμικών διαγραμμάτων για να διαπιστωθεί αν υπάρχει σημαντική διαφορά μεταξύ τους. Τυπική περίπτωση αποτελεί η σύγκριση των κλίσεων α_1, α_2 δύο καμπυλών αναφοράς με τα ίδια πρότυπα και το ίδιο όργανο, προκειμένου να διαπιστωθεί η σταθερότητα των προτύπων ή/και η οργανολογική σταθερότητα της μετρητικής διάταξης. Μια ταχεία δοκιμασία τύπου Student βασίζεται στον προσδιορισμό της στατιστικού στοιχείου t_{exp}

$$t_{exp} = \frac{|\alpha_1 - \alpha_2|}{\sqrt{s_{\alpha 1}^2 + s_{\alpha 2}^2}} \qquad (5\text{-}13)$$

Όπου $s_{\alpha1}$, $s_{\alpha2}$ οι τυπικές αποκλίσεις των κλίσεων α_1, α_2. Η πειραματική τιμή t_{exp} συγκρίνεται με την θεωρητική τιμή t_{theor}, που αντιστοιχεί στην επιθυμητή στάθμη εμπιστοσύνης και $\nu = (n_1 - 2) + (n_2 - 2) = n_1 + n_2 - 4$ βαθμούς ελευθερίας, όπου n_1 και n_2 ο αριθμός των ζευγών τιμών (x,y) που χρησιμοποιείται σε κάθε ένα από τα δύο διαγράμματα για τον προσδιορισμό των κλίσεων α_1 , α_2.

5.5.5. Ορισμός ευαισθησίας, ορίων ανίχνευσης και προσδιορισμού

Σε μια ενόργανη τεχνική ανάλυσης, στην οποία χρησιμοποιείται καμπύλη αναφοράς P = αC + β, για την ποσοτικοποίηση των αποτελεσμάτων ορίζουμε ως:

Ευαισθησία (sensitivity). Ορίζεται ως η κλίση της καμπύλης αναφοράς dP/dC, π.χ. για μια βολταμμετρική τεχνική ομιλούμε για ευαισθησία 15 μA/μM, για μια φωτομετρική ευαισθησίας 0,25 μονάδων απορρόφησης/(μg/mL) κ.ο.κ. Η ευαισθησία δεν πρέπει να συγχέεται με τα (κατώτερα) όρια ανίχνευσης και προσδιορισμού. Για να αποφευχθεί τέτοιου είδους σύγχυση συχνά αναφέρεται ως **οργανολογική ευαισθησία** (instrumental sensitivity)

Όριο ανίχνευσης (detection limit). Ορίζεται ως η συγκέντρωση C που παρέχει σήμα P ίσο προς $\beta + 3{,}3s_B$, όπου s_B είναι η τυπική απόκλιση του τυφλού (blank) και σε περίπτωση αδυναμίας μέτρησης του τυφλού, ως s_B μπορεί να χρησιμοποιηθεί η τυπική απόκλιση του αραιότερου προτύπου ή και η τιμή $s_{y/x}$.

Όριο προσδιορισμού (quantification limit). Ορίζεται ανάλογα ως η συγκέντρωση C που παρέχει σήμα ίσο P προς $\beta + 10s_B$. Πρέπει να τονιστεί ότι και οι δύο ορισμοί δεν είναι γενικοί και συνιστάται κάθε φορά που αναφέρονται αριθμητικές τιμές ενδεικτικές των προηγούμενων ορίων να αναφέρεται και ο εκάστοτε τρόπος ορισμού τους.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. A. M. Nevill, J. B. Kennedy, "Basic Statistical Methods for Engineers and Scientists", Intertext Books, London, 1964.
2. K. Eckschlager, "Errors, Measurement and Results in Chemical Analysis", Van Nostrand Reinhold Co., London, 1972.
3. H. M. Wadsworth, "Handbook of Statistical Methods for Engineers and Scientists", McGraw Hill, New York, 1989.
4. J. C. Miller, J. N. Miller, "Statistics for Analytical Chemistry", Ellis Horwood, Chichester, 3rd ed., 1993
5. Μ. Ι. Καραγιάννη, "Επεξεργασία, αξιολόγηση και παρουσίαση αναλυτικών δεδομένων", Αθήνα, 1978.
6. Θ. Π. Χατζηιωάννου, Κ. Η. Ευσταθίου, Δ. Π. Νικολέλη, "Προβλήματα Αναλυτικής Χημείας", Αθήνα, 1984.
7. Θ. Π. Χατζηιωάννου, Μ. Α. Κουππάρη, "Ενόργανη Ανάλυση", Αθήνα, 1990.
8. Π. Μαχαίρα, "Βιοφαρμακευτική", Αθήνα, 1992.