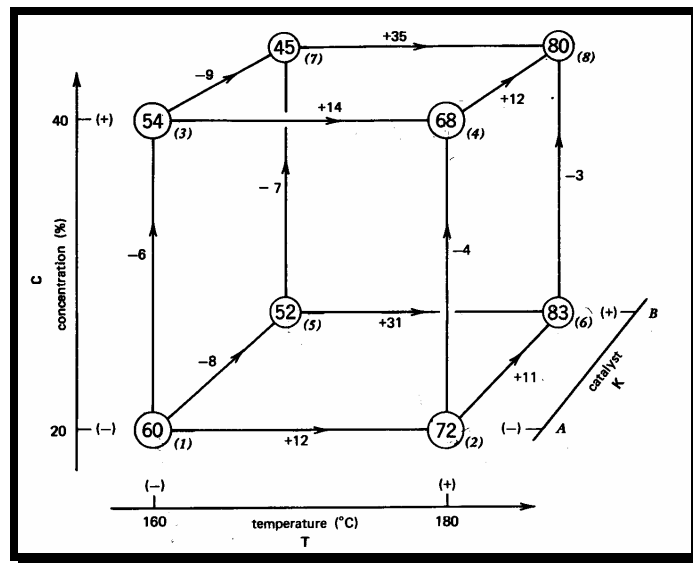


ΠΕΙΡΑΜΑΤΙΚΟΣ ΣΧΕΔΙΑΣΜΟΣ ΚΑΙ ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ



ΔΗΜΗΤΡΙΟΣ Π. ΚΟΜΙΛΗΣ
Λέκτορας

ΠΕΡΙΕΧΟΜΕΝΑ

1. Εισαγωγή	3
2. Επιστημονική σκέψη και πειραματικός σχεδιασμός	5
3. Βασικές αρχές στατιστικής ανάλυσης	18
4. Βασικές κατανομές στη περιβαλλοντική μηχανική	31
5. Αναγκαιότητα των γραφικών στην ανάλυση δεδομένων	41
6. Τεστ σημαντικότητας και έννοια της υπόθεσης	49
7. Σύγκριση πολλαπλών μέσων τιμών (k)	63
8. Ανάλυση διασποράς για k δείγματα (ANOVA)	68
9. Μετατροπή δεδομένων	76
10. Πλήρη παραγοντικά πειράματα	81
11. Κλασματικά παραγοντικά πειράματα	102
12. Πειράματα με μίγματα	115
13. Ανάπτυξη εμπειρικών μοντέλων μέσω γραμμικής παλινδρόμησης	133
14. Ο συντελεστής απόφασης R^2	150

1. Εισαγωγή

Η αναγκαιότητα των πειραμάτων στον κόσμο μας είναι αποτέλεσμα της Νευτώνειας προσέγγισης για την ερμηνεία και εξήγηση όλων των φυσικών συμβάντων. Τα πειράματα είναι ένα παράθυρο μέσα από το οποίο παρατηρούμε τη φύση (Box, 1974). Δηλαδή, δεν είναι η ίδια η φύση, η οποία είναι αδύνατο να μελετήσουμε. Με βάση τη λογική της παρατήρησης, της λογικής εξήγησης (ανάπτυξη θεωρίας) και των πειραματικών επαληθεύσεων, (το ORE¹ όπως ορίστηκε από τον σημαντικό επιστήμονα και φυσικό Richard Feynman), προσπαθούμε να ερμηνεύσουμε το τι συμβαίνει στη φύση. Αυτή εξάλλου είναι η βάση της επιστημονικής μεθόδου. Οι παρατηρήσεις και τα πειράματα μας παρέχουν ενδείξεις του τι συμβαίνει και όχι την απόλυτη γνώση του τι συμβαίνει. Το τελευταίο μπορεί να μην το γνωρίσουμε ποτέ ακριβώς.

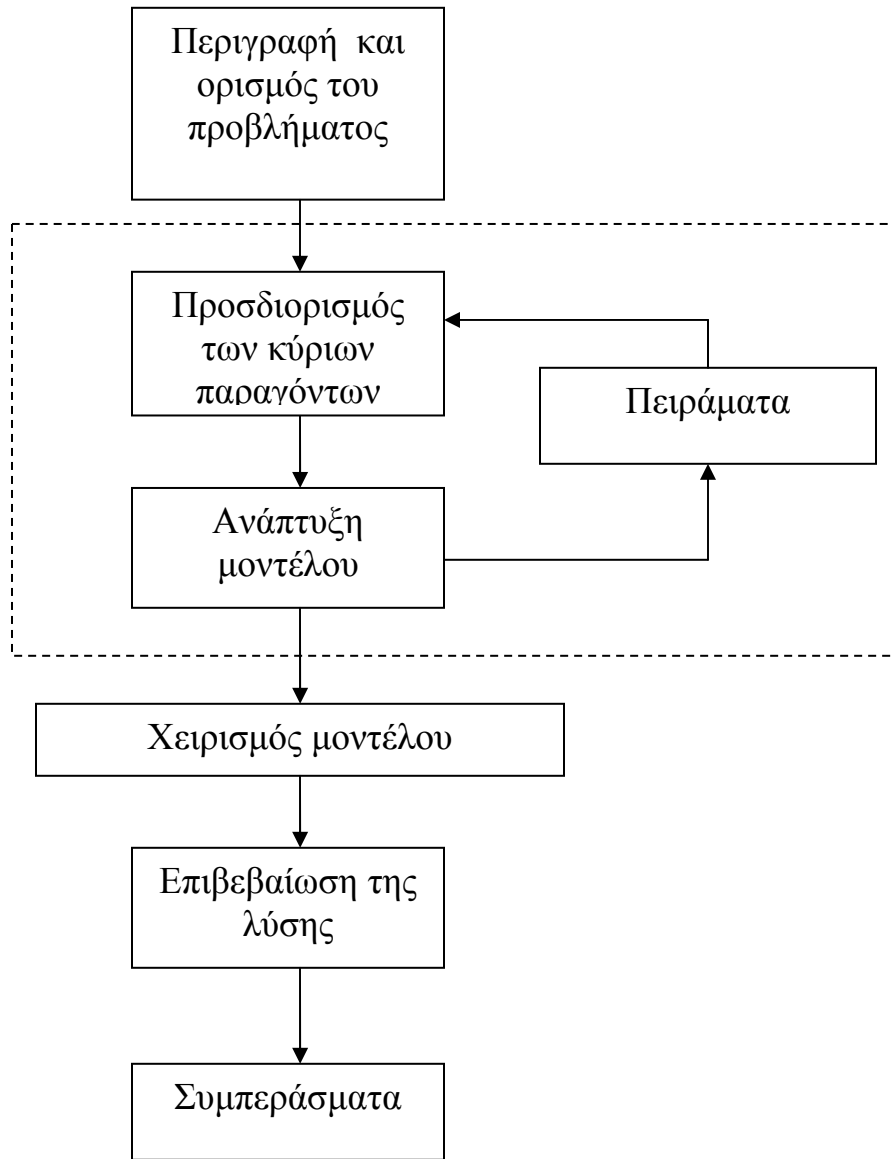
Όταν παρατηρούμε ένα φαινόμενο, το οποίο θέλουμε να ερμηνεύσουμε, αναπτύσσουμε μία λογική εξήγηση του τι μπορεί να συμβαίνει. Αυτή η λογική εξήγηση συχνά αποτελεί το αρχικό θεωρητικό υπόβαθρο (θεωρία) του φαινομένου. Τη θεωρία λοιπόν αυτή θέλουμε να επαληθεύσουμε (να ελέγξουμε) κάνοντας πειράματα βασισμένα σε έμμεσες (ή άμεσες) συνέπειες της θεωρίας. Αν κάποιο πείραμα δεν εξάγει τα αποτελέσματα που αναμένουμε βάσει της θεωρίας που αναπτύσσουμε, αυτό αποτελεί ένδειξη ότι η θεωρία πιθανά δεν ισχύει. Φυσικά πρέπει να επαναλάβουμε το πείραμα με περισσότερες ίσως μεταβλητές ή σε άλλες περιοχές τιμών των μεταβλητών. Πάντως αν υπάρχει τουλάχιστον μία εξαίρεση στον κανόνα (θεωρία) και αν η εξαίρεση αυτή μπορεί να επαληθευτεί πειραματικά, τότε η θεωρία ακυρώνεται, δηλ. **δεν ισχύει**. Αυτό είναι επιστημονική αρχή.

Γενικά η γνώση ενός φυσικού συμβάντος εξαρτάται από τον αριθμό και την ακρίβεια των πειραμάτων που θα διεξαχθούν με στόχο την κατανόησή του. Ποτέ δεν μπορούμε να γνωρίζουμε επακριβώς όλη την αλήθεια (η έννοια αλήθεια μπορεί να ταυτιστεί εδώ και με την έννοια φύση). Εξάλλου για αυτό και χρησιμοποιούμε τη στατιστική, που βασίζεται στην ιδέα ότι ο πραγματικός πληθυσμός δεν μπορεί να είναι ποτέ γνωστός (είναι πολύ μεγάλος γενικά και πολλές φορές πιθανά άπειρος) και συνεπώς πάντα καταλήγουμε στη παρατήρηση και μελέτη ενός κλάσματος του πραγματικού πληθυσμού (δηλ. του δείγματος).

Αφού λοιπόν τα πειράματα αποτελούν το μόνο τρόπο να δούμε τη φύση (αλήθεια) και συνεπώς να επαληθεύσουμε (ή όχι) κάποια θεωρία (υπόθεση), η άρτια πειραματική σχεδίαση πρέπει να αποτελέσει δομικό λίθο αυτής της προσπάθειας κατανόησης. Ως άμεση συνέπεια πρέπει να είναι κατανοητή και η ορθή ανάλυση των αποτελεσμάτων των πειραμάτων. Στόχος λοιπόν του βιβλίου αυτού είναι να παρουσιάσει τις βασικές αρχές πειραματικής σχεδίασης και στατιστικής ανάλυσης των πειραματικών αποτελεσμάτων ώστε να δούμε καλύτερα μέσα από το παράθυρο της φύσης (της αλήθειας), πραγματοποιώντας ακριβή πειράματα με την ελάχιστη δυνατή προσπάθεια.

Στόχος του παρόντος συγγράμματος είναι να ασχοληθεί με τις διεργασίες που περιγράφονται μεταξύ των διακεκομμένων γραμμών του παρακάτω σχήματος.

¹ ORE: observation (παρατήρηση), reason (λογική), experiment (πείραμα).

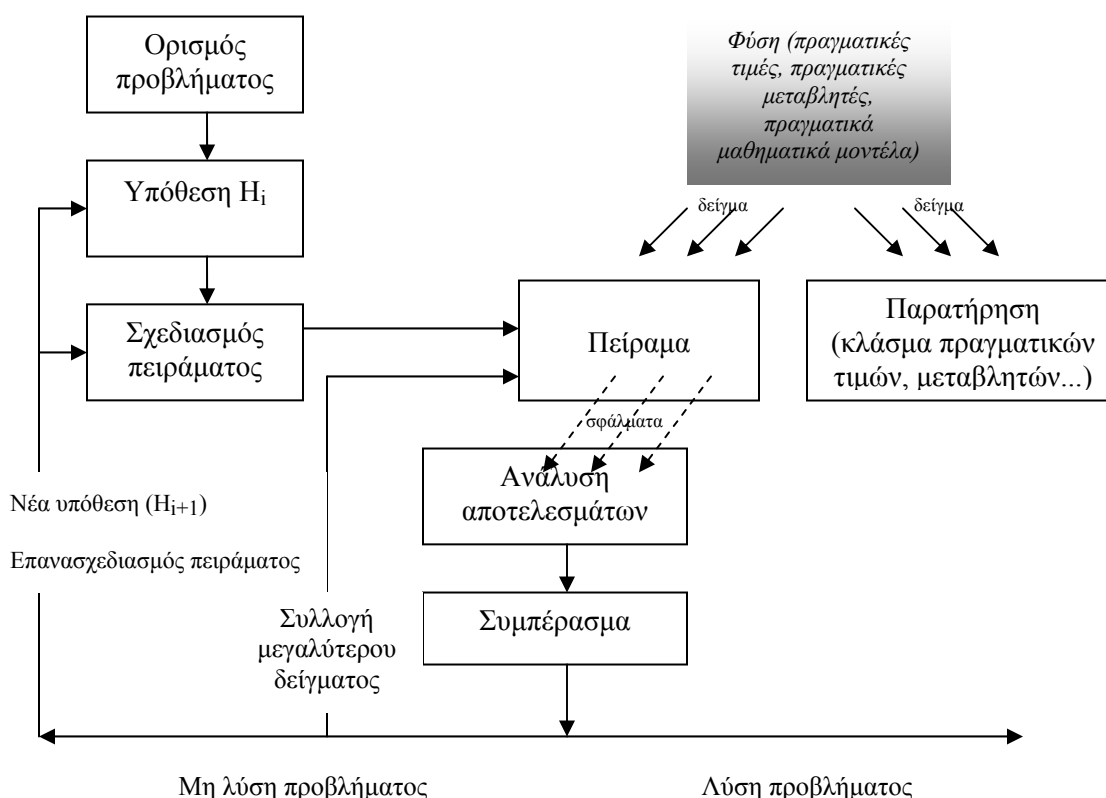


Σχήμα 1.1. Βασικές αρχές στην επιστημονική σκέψη

2. Επιστημονική σκέψη και πειραματικός σχεδιασμός

2.1 Επιστημονική σκέψη

Το σχήμα 2.1 παρακάτω απεικονίζει τις έννοιες της υπόθεσης, του πειραματικού σχεδιασμού και της διεξαγωγής του πειράματος, δηλ. της επιστημονικής σκέψης, όπως αναπτύχθηκαν παραπάνω.



Σχήμα 2.1. Απεικόνιση της λειτουργίας των πειραμάτων (βασισμένο σε Berthouex & Brown, 2002 & Box et al., 1978)

Σύμφωνα με το σχήμα 2.1, ορίζεται κατ'αρχάς ένα πρόβλημα που χρήζει λύσης². Για την αντιμετώπιση του προβλήματος ορίζεται κάποια υπόθεση - που συνήθως είναι κάποιο μαθηματικό μοντέλο -, που πρέπει να ελεγχθεί για την ισχύ της ή την μη ισχύ της. Για τον έλεγχο της υπόθεσης σχεδιάζεται κάποιο πείραμα. Ο στόχος του πειράματος είναι να ελέγξει την υπόθεση και, από μία άποψη, «να της δώσει κάθε ευκαιρία για να μην την επιβεβαιώσει». Το πείραμα αντλεί πεπερασμένο αριθμό δεδομένων (δείγμα) από τη φύση (πραγματικές τιμές, πραγματικές μεταβλητές). Ομοίως, ένα τμήμα (κλάσμα) των

² Αυτό αποτελεί και τον κύριο στόχο των μηχανικών, δηλαδή το να «λύνουν» προβλήματα μέσω κατασκευών και της τεχνολογίας. Πιο συγκεκριμένα, η αντιμετώπιση προβλημάτων που χρήζουν λύσης στον τομέα των τεχνολογιών περιβάλλοντος είναι και ο κύριος λόγος «ύπαρξης» των μηχανικών περιβάλλοντος

πραγματικών τιμών ή μεταβλητών της φύσης (όπως είναι το δείγμα) είναι διαθέσιμο προς παρατήρηση. Δηλαδή, η κατανόηση της λειτουργίας ενός συστήματος ή μίας διεργασίας γίνεται με συλλογή στοιχείων είτε μέσω μίας **μελέτης παρατήρησης** ή ενός **σχεδιασμένου πειράματος**. Στην πρώτη περίπτωση, ο μηχανικός μπορεί απλά να παρατηρήσει το σύστημα – βάσει παρελθοντικών δεδομένων – χωρίς τη δυνατότητα να επέμβει σε αυτό. Παράδειγμα, σε μία μονάδα επεξεργασίας λυμάτων υπάρχουν καταγεγραμμένες επί σειρά ετών οι ροές εισόδου λυμάτων στο εργοστάσιο. Ο μηχανικός προσπαθεί να συσχετίσει αυτές με τις βροχοπτώσεις των τελευταίων ετών και παρατηρεί λοιπόν τις δύο σειρές δεδομένων. Από τη συσχέτιση αυτή, μπορεί να προκύψει ένα εμπειρικό μοντέλο (δες ορισμό στη συνέχεια) που να περιγράφει τη σχέση βροχής και παροχής λυμάτων. Αντίθετα, στο σχεδιασμένο πείραμα, ο μηχανικός (ή επιστήμονας) πραγματοποιεί σκόπιμες αλλαγές στους ελεγχόμενους παράγοντες της διεργασίας, παρατηρεί το αποτέλεσμα και στη συνέχεια κάνει κάποια υπόθεση σχετικά με το ποιοι παράγοντες επηρεάζουν περισσότερο την απόκριση του συστήματος ή της διεργασίας που μελετάται. Προφανώς, τα σχεδιασμένα πειράματα αποτελούν την καλύτερη οδό εξαγωγής συμπερασμάτων για ένα μηχανικό. Περιορίζονται όμως από το κόστος και το χρόνο, και – προφανώς – δεν είναι εφικτά σε όλες τις περιπτώσεις. Παράδειγμα, επιδημιολογικά συμπεράσματα μπορούν να βασιστούν μόνο σε επιδημιολογικές μελέτες - που είναι εξ ορισμού μελέτες παρατήρησης - αφού δεν είναι δυνατό να επαναληφθούν πειραματικά (π.χ. μεγάλες καταστροφές – όπως η πτώση της ατομικής βόμβας στην Ιαπωνία - παρέχουν τέτοια επιδημιολογικά δεδομένα και είναι διαθέσιμες μόνο για παρατήρηση αλλά δεν μπορούν να επαναληφθούν πειραματικά).

Η φύση (ή η αλήθεια), λοιπόν, διοχετεύει δείγματα στον/στην πειραματιστή επιστήμονα για τη διεξαγωγή των πειραμάτων του/της, με στόχο την επαλήθευση ή όχι της αρχικής υπόθεσής του/της. Σε κάθε περίπτωση, όσο μεγάλο και να είναι ένα δείγμα, οι πραγματικές τιμές δεν είναι δυνατό να είναι διαθέσιμες, και συνεπώς πρέπει να λέμε πάντα ότι *προσεγγίζουμε* την αλήθεια. Δηλαδή, η αλήθεια κάποιας υπόθεσης δεν μπορεί με σιγουριά να καθοριστεί. Μπορούμε απλά να κάνουμε τεστ (δοκιμές) ώστε να δούμε αν τα δεδομένα έρχονται σε αντίθεση με την πιθανότητα να είναι η υπόθεση αληθής.

Τα πειράματα παράγουν αποτελέσματα, τα οποία αναλύονται. Η ανάλυση γίνεται με στατιστικές μεθόδους, αφού δεν μπορούμε ποτέ να υλοποιήσουμε το τέλει πείραμα και πάντα υπάρχουν πειραματικά σφάλματα. Οι στατιστικές μέθοδοι βοηθάνε στην κατανόηση της ποικιλότητας (variability) του συστήματος, δηλ. ότι διαδοχικές μετρήσεις ενός συστήματος ή ενός φαινομένου δεν παράγουν ίδια αποτελέσματα. Οι στατιστικές τιμές λοιπόν που εξάγονται από τις στατιστικές μεθόδους μπορούν να μας βοηθήσουν να κάνουμε επιστημονικές δηλώσεις σχετικά με την πιθανότητα κάποιων υποθέσεων να είναι αληθείς.

Ως αποτέλεσμα της ανάλυσης εξάγονται συμπεράσματα. Τα συμπεράσματα μπορούν να επαληθεύσουν την αρχική υπόθεση και συνεπώς το ορισμένο πρόβλημα να αντιμετωπιστεί. Αν η υπόθεση δεν επαληθευτεί, τότε ιεραρχικά πρέπει να:

1. Αυξήσουμε μέγεθος δείγματος και επαναλάβουμε το πείραμα, και αν αποτύχουμε στην επαλήθευση της αρχικής υπόθεσης τότε να

2. Επανασχεδιάσουμε πείραμα, και αν αποτύχουμε στην επαλήθευση της αρχικής υπόθεσης, τελικώς να,

3. Ορίσουμε διαφορετικά την υπόθεσή μας

Παράδειγμα λογικής πειραματικού σχεδιασμού.

Υπόθεση 1: Λόγω κάποιων ιδιοτήτων ενός νέου καταλύτη, η παρουσία του σε ένα μίγμα αναμένεται πιθανά να οδηγήσει στην αντίδραση της χημικής ένωσης Α με τη χημική ένωση Β για τη δημιουργία ενός νέου χρήσιμου προϊόντος Γ.

Αφού δεν υπάρχουν δεδομένα που να επαληθεύουν την παραπάνω υπόθεση, αποφασίζεται η διεξαγωγή πειραμάτων για τον έλεγχό της.

Πειραματικός σχεδιασμός (Π.Σ.) 1: Τα Α και Β αναμιγνύονται σε θερμοκρασία 600°C

Αποτελέσματα του (Π.Σ.) 1: Ένα μαύρο υλικό, με χαρακτηριστικά πίσσας παράγεται με ποσοστό Γ μικρότερο του 1%.

Η υπόθεση 1 και τα αποτελέσματα του ΠΣ 1 προφανώς δεν συμφωνούν. Εδώ μπαίνει η λογική σκέψη για την εξήγηση του αποτελέσματος του ΠΣ 1 σε συνδυασμό πάντα με την αρχική υπόθεση 1. Πιθανά η θερμοκρασία να ήταν πολύ υψηλή (δηλ. να μην επηρεάζουν μόνο οι ιδιότητες του Α & Β), που να οδήγησε σε σύντομη δημιουργία και μετέπειτα αποσύνθεση του Γ. Αυτό οδηγεί σε αλλαγή της υπόθεσης και επανασχεδιασμό του πειράματος.

Υπόθεση 2: Οι συνθήκες του πειράματος ήταν ακραίες – μία χαμηλότερη θερμοκρασία μπορεί να παράγει το προϊόν Γ (δεν είναι δηλαδή μόνο θέμα ιδιοτήτων των Α & Β όπως όρισε η υπόθεση 1).

Πειραματικός σχεδιασμός (Π.Σ.) 2: Τα Α και Β αναμιγνύονται σε δύο διαφορετικές μικρότερες θερμοκρασίες, ήτοι 550°C και 500°C.

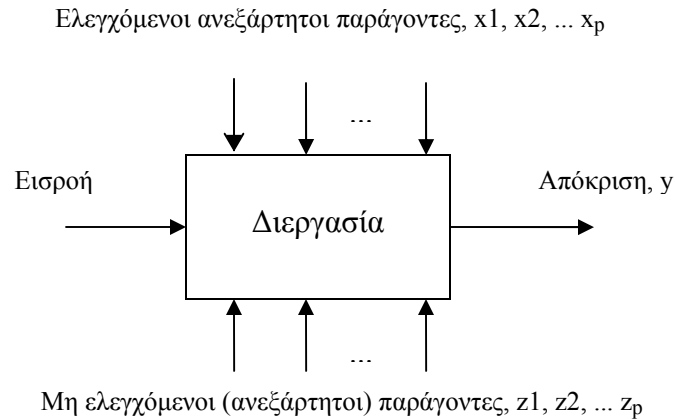
Αποτελέσματα του ΠΣ 2: Τα προϊόντα που παράγονται είναι λιγότερα πηκτώδη. Το προϊόν της θερμοκρασίας 550°C περιέχει το Γ σε ποσοστό 4% και της θερμοκρασίας 500°C σε ποσοστό 17%.

Τα παραπάνω δείχνουν την λογική του επαναπροσδιορισμού της αρχικής υπόθεσης, του επανασχεδιασμού του πειράματος και την αναγκαιότητα της λογικής ανάλυσης των αποτελεσμάτων.

2.2 Πειραματικός σχεδιασμός

Η πιο κοινή διεξαγωγή πειραμάτων αφορά στην μελέτη της επίδρασης παραγόντων (factors) σε κάποιο αποτέλεσμα ή στην «απόκριση» ενός συστήματος ή μίας διεργασίας. Υπάρχουν ανεξάρτητοι παράγοντες ελεγχόμενοι και μη ελεγχόμενοι, που επηρεάζουν τη

διεργασία, ενώ, φυσικά, κατά τη διεργασία απαιτούνται εισροές που παράγουν τουλάχιστον μία εκροή, που είναι το αποτέλεσμα ή η απόκριση. Ο όρος «απόκριση» αντιστοιχεί στον κοινά χρησιμοποιούμενο όρο «response» και συνεπώς θα χρησιμοποιείται στη συνέχεια. Με βάση τα παραπάνω, το γενικό μοντέλο μίας διεργασίας ή ενός συστήματος φαίνεται στο σχήμα 2.2.



Σχήμα 2.2. Εισροές και εκροές σε μία διεργασία

Οι στόχοι ενός πειράματος λοιπόν μπορεί να περιλαμβάνουν τα παρακάτω:

1. Κατανόηση για το ποιοι παράγοντες επηρεάζουν περισσότερο την απόκριση y
2. Απόφαση για το σε ποιές τιμές θα θέσουμε τα πιο σημαντικά x (ελεγχόμενες παράμετροι) έτσι ώστε το y να είναι σχεδόν πάντα κοντά στην επιθυμητή (βέλτιστη) τιμή του y
3. Απόφαση για το σε ποιές τιμές θα τεθούν τα x ώστε η διασπορά του y να είναι μικρή
4. Απόφαση για το σε ποιές τιμές θα τεθούν τα σημαντικά x ώστε οι επιδράσεις από τους μη ελεγχόμενους παράγοντες στην απόκριση y να είναι μικρές.

Η στατιστική σχεδίαση των πειραμάτων αναφέρεται στην διεργασία του σχεδιασμού πειραμάτων με τέτοιο τρόπο ώστε αποτελέσματα, τα οποία θα μπορούν να αναλυθούν με στατιστικές μεθόδους, θα συλλέγονται με στόχο την εξαγωγή έγκυρων και αντικειμενικών συμπερασμάτων. Εφόσον τα δεδομένα που παράγονται περιέχουν πάντα πειραματικά σφάλματα (θα οριστούν αυτά στη συνέχεια), οι στατιστικές μέθοδοι είναι η μόνη αντικειμενική προσέγγιση στην ανάλυση των δεδομένων.

Με βάση τα παραπάνω, υπάρχουν δύο πλευρές σε ένα πειραματικό πρόβλημα:

1. Η σχεδίαση του πειράματος
2. Η στατιστική ανάλυση των δεδομένων

Γενικά, όταν κάποιος/α υλοποιήσει το τέλειο πείραμα, η στατιστική ανάλυση δεν χρειάζεται. Τα αποτελέσματα είναι ξεκάθαρα και μπορούμε να εξάγουμε συμπεράσματα χωρίς στατιστική ανάλυση. Σπάνια όμως διεξάγουμε τα τέλεια πειράματα. Γενικά, το τέλειο πείραμα αποτελείται από άμεσες συγκρίσεις και έχει επανάληψη των δεδομένων.

Άμεσες συγκρίσεις σημαίνει ότι από το πείραμα έχουν πλήρως εξαιρεθεί οι μη ελεγχόμενοι παράγοντες (μπορούν να χαρακτηριστούν και ως ανεπιθύμητοι παράγοντες), ενώ *επανάληψη* (replication) σημαίνει απλώς ότι τα ίδια αποτελέσματα θα εξαχθούν με επανάληψη του πειράματος, δηλ. ότι τα αποτελέσματα που εξήχθησαν δεν είναι αποτέλεσμα απλώς τύχης.

2.3 Μαθηματικά μοντέλα

Ο πειραματισμός αποτελεί σημαντικό μέρος της σκέψης των μηχανικών³ ή επιστημόνων. Ο στόχος του πειράματος λοιπόν είναι ο εντοπισμός βέλτιστων τιμών καθώς και η μαθηματική περιγραφή του φαινομένου που μελετάται. Η μαθηματική περιγραφή έχει σαν στόχο τη ανάπτυξη μαθηματικών εξισώσεων, ώστε η εξαγωγή αποτελεσμάτων σε κάποια μελλοντική στιγμή να μην χρίζει επαναληπτικά κάποιο πείραμα. Συνήθως το πείραμα έχει κόστος οικονομικό και είναι χρονοβόρο ενώ πάντα η δυνατότητα χρήσης των «φθηνών» μαθηματικών εξισώσεων ή μοντέλων προτιμάται.

Τα μαθηματικά μοντέλα που ακολουθούν τους φυσικούς νόμους ή κάποια θεωρία (π.χ. $B=m g$) λέγονται θεωρητικά ή φυσικά⁴ μοντέλα. Βέβαια πάντα υπάρχει κάποιο σφάλμα, οπότε θα ήταν καλύτερο να κάναμε χρήση της θεωρητικής εξίσωσης με προσθήκη του πειραματικού σφάλματος, δηλ. $B=m g + \epsilon$.

Παρόλα αυτά τα περισσότερα μοντέλα των μηχανικών, που εξαρτώνται συνήθως στον πειραματισμό, βασίζονται, ακριβώς, στα αποτελέσματα των πειραμάτων αυτών. Είναι τα εμπειρικά μοντέλα και το σύγγραμμα θα ασχοληθεί με αυτά κατά κύριο λόγο καθώς και με τη δυνατότητα της μετατροπής των αποτελεσμάτων ενός (καλά σχεδιασμένου) πειράματος σε εμπειρικό μοντέλο. Παραδειγματικά, η σχέση μίας απόκρισης y από τους παράγοντες X, Z, Y – δηλ. $y=f(X,Z,Y)$ – μπορεί να εκφραστεί από το εμπειρικό (γραμμικό) μοντέλο: $y=b_1 X+b_2 Z + b_3 Y$ ή από το μη γραμμικό εμπειρικό μοντέλο $y=b_1 X+b_2 Z + b_3 Y + b_4 XZ$. Οι συντελεστές b_1, b_2, b_3, b_4 θα πρέπει να προσδιοριστούν κατά τη διαδικασία προσαρμογής του μοντέλου στα δεδομένα. Όμοια με πριν, θα πρέπει να προστίθεται το πειραματικό σφάλμα στη μαθηματική εξίσωση, δηλ. η σωστή έκφραση του γραμμικού μοντέλου είναι: $y=b_1 X+b_2 Z + b_3 Y + \epsilon$.

2.4 Βασικές αρχές πειραματισμού

Οι τέσσερις βασικές αρχές του (καλού) πειραματικού σχεδιασμού είναι:

- *Επανάληψη (replication)*

Επανάληψη (να μην συγχέεται με την επαναληψιμότητα, που είναι η δυνατότητα επανάληψης) σημαίνει την ανεξάρτητη επανάληψη των συνδυασμών των παραγόντων και αποτελεί μία εκτίμηση του τυχαίου πειραματικού σφάλματος. Η επίδραση του σφάλματος στην επίδραση ενός παράγοντα εκτιμάται με τον υπολογισμό του τυπικού

³ Ως μηχανική ορίζεται η εφαρμοσμένη επιστήμη που έχει στόχο τη λύση προβλημάτων. Κατά συνέπεια, ο μηχανικός περιβάλλοντος ασχολείται με τη λύση περιβαλλοντικών προβλημάτων για τη διασφάλιση της δημόσιας υγείας και του περιβάλλοντος.

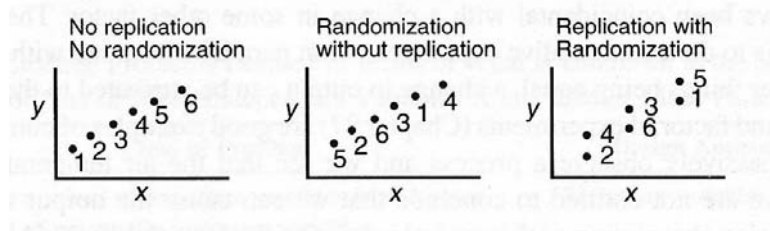
⁴ Ο όρος «μηχανιστικά μοντέλα» επίσης χρησιμοποιείται. Ο αγγλικός όρος είναι mechanistic models.

σφάλματος. Το τυπικό σφάλμα μειώνεται όσο ο αριθμός των παρατηρήσεων και επαναλήψεων αυξάνεται. Οι «γνήσιες επαναλήψεις» σημαίνουν απλά ότι επαναλαμβάνουμε το πείραμα διατηρώντας τις τιμές των ελεγχόμενων μεταβλητών x στην ίδια ακριβώς τιμή. Είναι συχνά καλύτερο να αυξήσουμε τον αριθμό των πειραμάτων με επαναλαμβανόμενες παρατηρήσεις στην ίδια τιμή παρά να προσθέτουμε νέες παρατηρήσεις σε διαφορετικές τιμές.

- *Τυχαία επιλογή δειγμάτων (randomization)*

Η τυχαία κατανομή είναι βάση του καλού πειραματικού σχεδιασμού. Ουσιαστικά σημαίνει απλώς ότι τόσο η κατανομή των υλικών του πειράματος, αλλά και η σειρά με την οποία θα γίνουν τα πειράματα, αποφασίζονται τυχαία. Γενικά οι στατιστικές μέθοδοι απαιτούν τις παρατηρήσεις που κάνουμε να είναι ανεξάρτητες και τυχαίες μεταβλητές. Γενικά η τυχαία κατανομή υποβοηθάται αρκετά από τη χρήση ηλεκτρονικών υπολογιστών που παρέχουν τη δυνατότητα εμφάνισης τόσο τυχαίων αριθμών, αλλά και κατανομή των πειραμάτων (runs) με τυχαία σειρά. Η συνάρτηση RAND() του EXCEL βοηθάει αρκετά στην υλοποίηση τυχαίας κατανομής υλικών και πειραμάτων.

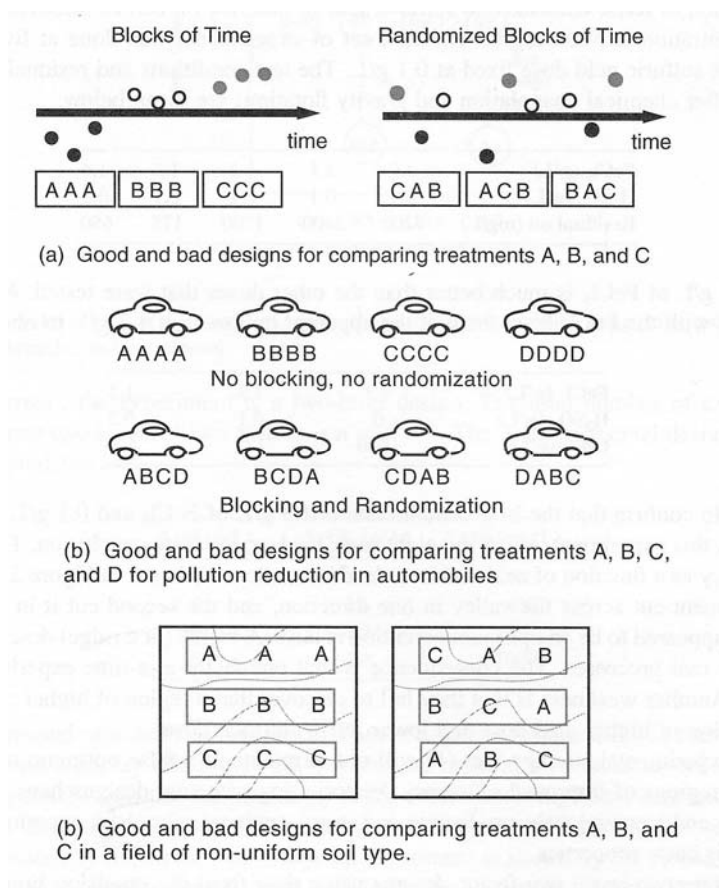
Το σχήμα 2.3 δείχνει ότι η πειραματική σχεδίαση βελτιώνεται από αριστερά στα δεξιά αφού αυξάνεται η τυχαία κατανομή (τυχαία σειρά) των πειραμάτων και οι επαναλήψεις. Οι αριθμοί δηλώνουν την σειρά των παρατηρήσεων.



Σχήμα 2.3. Επιρροές επανάληψης και τυχαίας κατανομής σε ένα πείραμα

- *Ομαδοποίηση (blocking)*

Η ομαδοποίηση βοηθάει στην ανάλυση των δεδομένων και στη μείωση του πειραματικού λάθους. Ο στόχος της ομαδοποίησης είναι ουσιαστικά η κατάταξη του συνόλου των δεδομένων σε «ομάδες» που να είναι όσο το δυνατό περισσότερο ομογενοποιημένες. Αυτό βοηθάει στο ότι το πειραματικό λάθος θα υπολογιστεί για κάθε ομάδα ξεχωριστά και στη συνέχεια όλα τα πειραματικά λάθη θα αθροιστούν για το σύνολο του πειράματος. Με αυτό τον τρόπο υπολογίζεται μικρότερο πειραματικό λάθος σε σχέση με τον υπολογισμό όταν δεν γίνει ομαδοποίηση. Το σχήμα 2.4. δείχνει κάποια παραδείγματα ομαδοποίησης που διευκολύνει στην ανάλυση των δεδομένων.



Σχήμα 2.4. Παραδείγματα ομαδοποιήσεων στη σχεδίαση πειραμάτων

Στην πρώτη περίπτωση, αν γίνουν τα πειράματα με τη σειρά A, B, Γ (με 3 επαναλήψεις) ανά σειρά, τότε πιθανά να εισερχόταν λάθος στις μετρήσεις λόγω των διαφοροποιήσεων στο χρόνο. Αν γίνουν 3 ομάδες που να περιέχουν (τυχαία) την σειρά των A,B και Γ, τότε θα μειωνόταν αρκετά η πιθανότητα αυτή. Το παράδειγμα εδώ βέβαια είναι συνδυασμός ομαδοποίησης και τυχαίας σειράς. Στο παράδειγμα με τα αυτοκίνητα, έστω οτι θέλουμε να ελέγξουμε 4 διεργασίες (π.χ. τεχνικές μείωσης συγκεκριμένου ρύπου) σε αυτοκίνητα. Αν η κάθε διεργασία (με τις επαναλήψεις) γίνει σε ένα αυτοκίνητο ξεχωριστά, τότε πιθανά να έχουμε αυξημένα λάθη λόγω της χρήσης διαφορετικών αυτοκινήτων (αν και μπορεί να είναι αυτοκίνητα της ίδιας μάρκας, πάντα υπάρχουν διαφοροποιήσεις μεταξύ «ομοίων» αυτοκινήτων). Η επιθυμητή σχεδίαση θα ήταν να χρησιμοποιηθεί το κάθε αυτοκίνητο ως «ομάδα» ή «κατηγορία» και σε κάθε μία να ελεγχθούν – με τυχαία σειρά – οι τεχνικές A,B,Γ,Δ. Βέβαια, επίσης αποδεκτή σχεδίαση θα ήταν η χρήση ενός αυτοκινήτου για τον έλεγχο όλων των διαφορετικών τεχνικών. Στο τελευταίο παράδειγμα, έστω οτι έχουμε ένα έδαφος ρυπασμένο με ρύπο, η κατανομή του οποίου φαίνεται από τις ισοπληθείς του σχεδίου. Ο έλεγχος μίας τεχνικής απορρύπανσης A μόνο στο πάνω τμήμα του πεδίου θα αμεροληπτούσε έναντι των τεχνικών B, Γ. Πάλι ένας συνδυασμός ομαδοποίησης και τυχαίας κατανομής των A,B,Γ από το πάνω στο κάτω τμήμα του χάρτη, όπως φαίνεται, αποτελεί αρτιότερη πειραματική σχεδίαση.

- *Συγκριτική σχεδίαση*

Η συγκριτική σχεδίαση στηρίζεται απλά στην αρχή του ότι αν θέλουμε να δούμε πως επηρεάζει ένας παράγοντας μία διεργασία, πρέπει να παρατηρήσουμε τι συμβαίνει στη διεργασία όταν αλλάζουν οι τιμές του εν λόγω παράγοντα, διατηρώντας τους υπόλοιπους ελεγχόμενους και μη ελεγχόμενους παράγοντες σταθερούς. Παραδειγματικά, αν προσθέσουμε ποσότητα X σε μία διεργασία και η εκροή (απόκριση) αυξηθεί ή βελτιωθεί, τείνουμε να πούμε ότι η βελτίωση αυτή οφείλεται στην προσθήκη του X . Αυτό όμως μπορεί να οδηγήσει σε λάθος συμπεράσματα, αφού η αλλαγή στην απόκριση μπορεί να ήταν συμπτωματική και να οφειλόταν στην (μη ελεγχόμενη) αλλαγή κάποιου άλλου παράγοντα που δεν λάβαμε υπόψη μας. Πρέπει λοιπόν να υλοποιηθούν συγκριτικά πειράματα, δηλ. η διεργασία να λειτουργήσει χωρίς την προσθήκη του X , και μετά ξεχωριστά με διαφορετικές προσθήκες του X , διατηρώντας τις άλλες παραμέτρους σταθερές. Πρακτικά, δεν είναι δυνατό να ελέγξουμε κάποιες παραμέτρους (π.χ. περιβαλλοντική θερμοκρασία). Σε κάθε περίπτωση, η ανάγκη των συγκριτικής σχεδίασης είναι η βάση ενός πειράματος και θα δούμε ότι εμφανίζεται σε πειραματικούς σχεδιασμούς, όπως παραγοντικές σχεδιάσεις και δοκιμή t σε ζεύγη.

2.5 Προσεγγίσεις πειράματος

- *Καλύτερη υπόθεση (best guess)*

Μία από τις πρώτες προσεγγίσεις ενός πειράματος είναι η προσέγγιση της «καλύτερης υπόθεσης – best guess approach». Δηλαδή επιλέγω τυχαία ένα συνδυασμό παραγόντων, διεξάγω το πείραμα και εξάγω αποτελέσματα. Στη συνέχεια, κάνω μία άλλη επιλογή, πιθανά αλλάζοντας ένα παράγοντα και κρατώντας τους άλλους σταθερούς, και παρατηρώ πιθανές αλλαγές στο αποτέλεσμα. Με αυτό τον τρόπο μπορώ να συνεχίσω σχεδόν απεριόριστες φορές. Το βασικό μειονέκτημα έχει να κάνει με το πόσο σωστή (δηλ. κοντά στο επιθυμητό βέλτιστο) είναι η αρχική υπόθεση. Επίσης, αν κάποια επόμενη υπόθεση δώσει όντως μία επιθυμητή τιμή – κοντά στο βέλτιστο – δεν είναι γνωστό αν αυτή τιμή είναι η τελικώς βέλτιστη και συνεπώς αν πρέπει να συνεχίσει ή σταματήσει τα πειράματά του ο πειραματιστής.

- *Μεταβολή ενός παράγοντα κάθε φορά (One Factor at a Time)*

Ελαφρώς διαφορετική από την παραπάνω στρατηγική είναι η στρατηγική της αλλαγής ενός παράγοντα κάθε φορά. Είναι κοινά χρησιμοποιούμενη στρατηγική μεταξύ επιστημόνων και μηχανικών. Βασίζεται στην επιλογή μίας βάσης συνδυασμών παραγόντων (baseline) και στη συνέχεια στη μεταβολή ενός παράγοντα σε διαφορετικές τιμές διατηρώντας τους υπολοίπους στην τιμή βάσης τους. Το βασικό μειονέκτημα της στρατηγικής αυτής είναι ότι δεν μελετάει αλληλεπιδράσεις. Ως αλληλεπίδραση ορίζεται «η αποτυχία ενός παράγοντα να παράξει την ίδια επίπτωση (effect) στην απόκριση σε διαφορετικά επίπεδα ενός άλλου παράγοντα».

- *Παραγοντικά πειράματα (factorial experiments)*

Αποτελεί τον σωστό τρόπο σχεδίασης πειραμάτων αφού με τον μικρότερο αριθμό πειραμάτων – σε σχέση με τα άλλες τεχνικές – παράγεται η μεγαλύτερη πληροφορία.

Επίσης, είναι δυνατή η ποσοτικοποίηση μελέτη των κυρίων επιδράσεων αλλά και **των αλληλεπιδράσεων** μεταξύ των παραγόντων.

Στη συνέχεια παρουσιάζεται ένα παράδειγμα χρήσης της προσέγγισης της μεταβολής ενός παράγοντα κάθε φορά και συγκρίνεται με αντίστοιχη παραγοντική σχεδίαση.

Παράδειγμα (μη καλής) πειραματικής σχεδίασης

Έστω ότι ο στόχος μας είναι η απομάκρυνση πετρελαϊκών καταλοίπων με χρήση του κροκιδωτικού FeCl_3 και H_2SO_4 . Η καλύτερη μελέτη της κροκίδωσης σε κάποιο υδατικό διάλυμα γίνεται μέσω των jar tests. Έστω λοιπόν ότι προσθέτουμε με διάφορες συγκεντρώσεις τα παραπάνω χημικά και ο στόχος μας είναι να βρούμε τη βέλτιστη δόση ώστε να έχουμε τη μεγαλύτερη καθίζηση της πετρελαϊκού υπολείμματος. Πως βρίσκουμε αυτή το βέλτιστο συνδυασμό δόσεων των δύο χημικών? Ο πρώτος πίνακας δείχνει την προσέγγιση της μεταβολής ενός παράγοντα κάθε φορά. Διατηρούμε το θεικό οξύ σταθερό σε συγκέντρωση 0.1 g/L (αυτό βασίστηκε καταρχάς σε μία αρχική υπόθεση) και μεταβάλλουμε ο FeCl_3 από 1.0 σε 1.4. Παρατηρούμε ότι η μικρότερη υπολειμματική συγκέντρωση πετρελαίου στην υδατική φάση είναι 175 mg/L. Άρα, είναι ο καλύτερος συνδυασμός των δυο χημικών οι συγκεντρώσεις 1.3 g/L & 0.1 g/L?

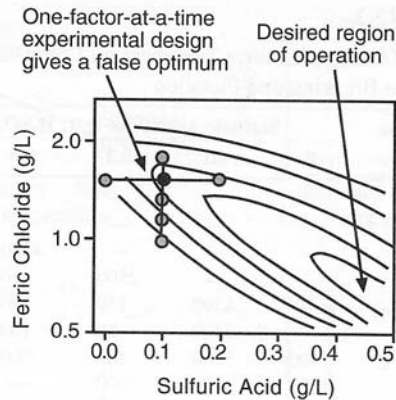
FeCl_3 (g/L)	1.0	1.1	1.2	1.3	1.4
H_2SO_4 (g/L)	0.1	0.1	0.1	0.1	0.1
Residual oil (mg/L)	4200	2400	1700	175	650

Στη συνέχεια διατηρούμε το FeCl_3 στην τιμή 1.3 g/L – που είχαμε δει ότι ανήκει στο βέλτιστο συνδυασμό στο πρώτο πείραμα – και μεταβάλλουμε το θεικό οξύ με άλλα 3 πειράματα.

FeCl_3 (g/L)	1.3	1.3	1.3
H_2SO_4 (g/L)	0	0.1	0.2
Oil (mg/L)	1600	175	500

Πάλι παρατηρούμε ότι ο βέλτιστος συνδυασμός είναι ο 1.3 g/L & 0.1 g/L για το FeCl_3 & H_2SO_4 αντίστοιχα.

Το παρακάτω γράφημα – οι ισοπληθείς του οποίου είχαν προκύψει από μεγάλο αριθμό πειραμάτων σε διαφορετικούς συνδυασμούς - δείχνουν ότι τελικά ο βέλτιστος συνδυασμός βρίσκεται αρκετά μακριά από τις παραπάνω τιμές. Συγκεκριμένα, ο βέλτιστος συνδυασμός βρίσκεται στο «νοτιοανατολικό» τμήμα του γραφήματος, υποδεικνύοντας τιμές H_2SO_4 κοντά στο 0.45 g/L και τιμές του FeCl_3 κοντά στο 0.7 g/L. Συνεπώς, ο πειραματικός σχεδιασμός με μεταβολή ενός παράγοντα κάθε φορά έδωσε ένα λάθος βέλτιστο.

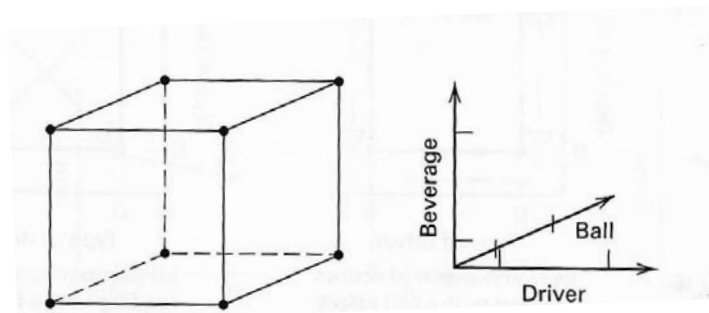


Σχήμα 2.5. Χρήση επιφάνειας απόκρισης για πειραματισμό σε βέλτιστη περιοχή

Η βασική αρχή στα παραγοντικά πειράματα είναι ότι μεταβάλλουμε η αριθμό παραγόντων σε 2 επίπεδα και με χρήση 2^n πειραμάτων υλοποιούμε όλους τους συνδυασμούς πειραμάτων μεταξύ των παραγόντων αυτών.

Σε περίπτωση δύο παραγόντων σε δύο επίπεδα (2^2 παραγοντική σχεδίαση) μπορούμε να φτιάξουμε ένα γράφημα, όπως το παραπάνω, όπου ο κάθε παράγοντας να βρίσκεται στον οριζόντιο και κατακόρυφο άξονα αντίστοιχα, ενώ εντός του γραφήματος «σχηματίζεται» μία επιφάνεια απόκρισης (response surface), που ουσιαστικά σχηματίζεται από τις διάφορες τιμές της απόκρισης y που μελετούμε στο συγκεκριμένο πείραμα. Μεταβολές των ελεγχόμενων παραγόντων είναι υπεύθυνες για τις διαφορετικές τιμές του y και κατά συνέπεια τη μορφή της επιφάνειας απόκρισης.

Σε περίπτωση 3 παραγόντων σε δύο επίπεδα σχηματίζεται ένας κύβος, οι γωνίες του οποίου αποτελούν τα σημεία όπου θα «τρέξουμε» το πείραμα. Αποτελείται δηλαδή από 8 πειράματα τουλάχιστον (εφόσον δεν έχουμε επανάληψη σε κάθε γωνία του κύβου, που σαφώς συνίσταται). Το σχήμα 2.6. δείχνει το χαρακτηριστικό «κύβος» ενός 2^3 παραγοντικού πειράματος. Οι τρεις παράγοντες που φαίνονται είναι ενδεικτικοί και προέρχονται από πείραμα που αφορούσε στην επίδραση 3 παραγόντων, ήτοι του είδους μπαστουνιού (driver), είδος μπάλας και είδος αναψυκτικού που καταναλώνεται κατά τη διάρκεια ενός παιχνιδιού γκολφ. Η απόκριση ήταν το σκορ.



Σχήμα 2.6. Γραφική απεικόνιση πειράματος με 3 παράγοντες που μεταβάλλονται σε 2 τιμές

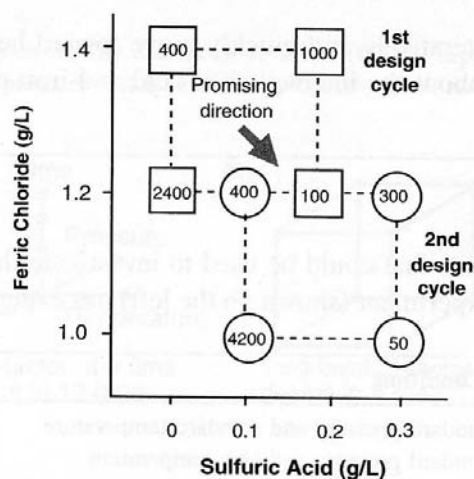
Παράδειγμα 2² παραγοντικού πειράματος

Με τους δύο παράγοντες στο παραπάνω jar test, θέτω 2 επίπεδα ανά παράγοντα. Συγκεκριμένα, το οξύ τίθεται στα επίπεδα 0 & 0.2 g/L και ο FeCl₃ στα επίπεδα 1.2 και 1.4 g/L. Οι (4) συνδυασμοί μεταξύ τους φαίνονται στον παρακάτω πίνακα. Έστω ότι θέλουμε να διερευνήσουμε το προς ποία περιοχή κινούμαστε. Οι αποκρίσεις (υπολειμματικό πετρέλαιο) επίσης φαίνονται στην τρίτη στήλη.

1^{ος} κύκλος πειραμάτων 2² παραγοντικής ανάλυσης		
Οξύ (g/L)	FeCl ₃ (g/L)	Oil (mg/L)
0	1.2	2400
0	1.4	400
0.2	1.2	100
0.2	1.4	1000

2^{ος} κύκλος πειραμάτων 2² παραγοντικής ανάλυσης κοντύτερα στη βέλτιστη περιοχή		
Οξύ (g/L)	FeCl ₃ (g/L)	Oil (mg/L)
0.1	1.0	4200
0.1	1.2	400
0.3	1.0	50
0.3	1.2	300

Το γράφημα στη συνέχεια δείχνει τις θέσεις της απόκρισης στον πρώτο και δεύτερο κύκλο πειραμάτων. Ο 1^{ος} κύκλος μας έδειξε ότι αξίζει να κινηθούμε προς τα «νοτιοανατολικά» του γραφήματος ώστε να εντοπίσουμε την βέλτιστη (ελάχιστη) τιμή υπολειμματικού πετρελαίου. Για αυτό το λόγο υλοποιούμε το β' κύκλο πειραμάτων.



Σχήμα 2.7. Χρήση παραγοντικών πειραμάτων με 2 παράγοντες σε διαφορετικές περιοχές απόκρισης

Είναι σημαντικό να τονιστεί ότι με 8 πειράματα ήταν δυνατό να εντοπίσουμε τη βέλτιστη περιοχή, ενώ ο πειραματιστής είχε χρησιμοποιήσει αρχικώς ένα σετ 28 πειραμάτων (δες παρακάτω πίνακα) για να εντοπίσει τη βέλτιστη περιοχή. Η αποδοτικότητα (efficiency) των παραγοντικών πειραμάτων είναι σαφώς το μεγάλο τους ατού!

FeCl ₃ Dose (g/L)	Sulfuric Acid Dose (g/L H ₂ SO ₄)				
	0	0.1	0.2	0.3	0.4
0.6	—	—	—	—	600
0.7	—	—	—	—	50
0.8	—	—	—	4200	50
0.9	—	—	2500	50	150
1.0	—	4200	150	50	200
1.1	—	2400	50	100	400
1.2	2400	1700	100	300	700
1.3	1600	175	500	—	—
1.4	400	650	1000	—	—
1.5	350	—	—	—	—
1.6	1600	—	—	—	—

Συμπεράσματα

Τα παραγοντικά πειράματα λοιπόν – πέραν της εκτίμησης των αλληλεπιδράσεων - μας βοηθούν στον να σχηματίσουμε την επιφάνεια απόκρισης (όταν έχουμε δύο παράγοντες) με την οποία, έστω και οπτικά, μπορούμε να εντοπίσουμε τις βέλτιστες περιοχές (μέγιστα ή ελάχιστα). Με 3 παράγοντες μιλάμε για σφαίρα απόκρισης, αλλά επειδή αυτή είναι δύσκολο να σχηματιστεί (ακόμα και με υπολογιστικά προγράμματα) σταματάμε την ανάλυση μέσω εικόνας (όπως στην περίπτωση της επιφάνειας απόκρισης που μπορούμε να δούμε στο χαρτί) και χρησιμοποιούμε απλώς την παραγόμενη εξίσωση, την οποία μπορούμε – με χρήση γραμμικού ή μη γραμμικού προγραμματισμού – να ελαχιστοποιήσουμε ή μεγιστοποιήσουμε καταλήγοντας στην επιθυμητή περιοχή απόκρισης και φυσικά στις επιθυμητές περιοχές των υπό μελέτη παραγόντων. Η ίδια μαθηματική (και όχι οπτική) προσέγγιση ισχύει για άνω των 3 παραγόντων. Οι

παραγοντικές αναλύσεις με τις οποίες ασχολείται το παρόν βιβλίο αφορά σε 2 επίπεδα. Τα παραγοντικά πειράματα θα αναλυθούν με λεπτομέρεια σε επόμενο κεφάλαιο.

3. Βασικές αρχές στατιστικής ανάλυσης

Όπως συζητήθηκε στο προηγούμενο κεφάλαιο, το δείγμα αποτελείται από έναν (πεπερασμένο) αριθμό από n παρατηρήσεις (ή δεδομένα), οι οποίες είναι διαθέσιμες στον επιστήμονα για μέτρηση και διεξαγωγή του πειράματος. Ο πληθυσμός είναι ένας πολύ μεγάλος αριθμός από N παρατηρήσεις (ή δεδομένα), που αποτελούν και την αλήθεια ή την πραγματικότητα και πρακτικώς δεν είναι ποτέ διαθέσιμες. Άρα οι τιμές τους δεν είναι γνωστές στον επιστήμονα, ακριβώς λόγω του μεγάλου (πιθανά και άπειρου) αριθμού τους. Το δείγμα συνεπώς προέρχεται από τον πληθυσμό, είναι κλάσμα αυτού, είναι ότι καλύτερο μπορούμε να συλλέξουμε, και συνεπώς μπορούμε να πούμε απλά ότι $n \ll N$.

Τα δεδομένα ενός δείγματος πρέπει να είναι τυχαία, δηλαδή τυχαίες μεταβλητές. Ένας πρακτικός ορισμός της τυχαίας μεταβλητής, που αποτελεί θεμελιώδη όρο στη διεξαγωγή πειραμάτων, είναι απλά: «η τιμή της επόμενης παρατήρησης σε ένα πείραμα». Ένας εναλλακτικός – και πιθανά περισσότερο κατανοητός – ορισμός είναι: «τυχαία μεταβλητή είναι η μεταβλητή που η μετρηθείσα της τιμή μπορεί να μεταβάλλεται μεταξύ επαναληπτικών μετρήσεων του πειράματος».

Παραδειγματικά λοιπόν, αν θέλουμε να μετρήσουμε το BOD ενός ποταμού, τότε θα πάρουμε έναν αριθμό n δειγμάτων (πεπερασμένος αριθμός που είναι οικονομικά και τεχνικά εφικτός να συλλεχθεί), που θα μας βοηθήσουν να προσεγγίσουμε (ή γνωρίσουμε) το πραγματικό BOD του ποταμού, το οποίο πρακτικά δεν θα μας είναι ποτέ γνωστό ακριβώς, αφού απαιτείται ένας τεράστιος αριθμός δειγμάτων (πιθανά άπειρος για την περίπτωση του ποταμού, ενώ αν μιλούσαμε για λίμνη να ήταν απλά ένας τεράστιος αριθμός δειγμάτων), πρακτικώς αδύνατο να συλλεχθεί και μετρηθεί.

Η διαφοροποίηση λοιπόν μεταξύ του πληθυσμού (συνήθως άγνωστου) και του δείγματος (μετρήσεων με τις οποίες πάμε να προσεγγίσουμε τον πληθυσμό) είναι χαρακτηριστική στην στατιστική. Οι παράμετροι του πληθυσμού συνήθως συμβολίζονται με ελληνικά σύμβολα και οι παράμετροι του δείγματος (στατιστικές παράμετροι) με λατινικά σύμβολα.

3.1 Βασικές στατιστικές παράμετροι

Σε ένα σύνολο n μετρήσεων με τιμές y_i από ένα δείγμα που προέρχεται από ένα πληθυσμό με (πολύ) μεγάλο αριθμό δεδομένων (ή παρατηρήσεων) N , ορίζονται οι παρακάτω «κλασσικές» παράμετροι:

- **Μέση τιμή⁵ πληθυσμού:** $\mu = \frac{\sum_{i=1}^N y_i}{N}$. Είναι η πραγματική τιμή του πληθυσμού που πιθανά ποτέ να μην γνωρίσουμε. Μόνο σε περιπτώσεις που «ετοιμάζουμε» ένα standard με γνωστή συγκέντρωση μπορούμε να πούμε ότι γνωρίζουμε τον τιμή του πληθυσμού.

⁵ Προσοχή: Η μέση τιμή (ή μέσος όρος) διαφέρουν από τον μέσο. Οι αντίστοιχοι αγγλικοί όροι είναι average (ή mean) & median, αντίστοιχα.

Επίσης, σε περιπτώσεις ακεραίων παρατηρήσεων, η τιμή μ είναι πάντα γνωστή. Π.χ. όταν μετράμε ένα πληθυσμό με 203 άλογα, τότε η πραγματική τιμή του πληθυσμού είναι πάντα 203 άλογα!

- **Διασπορά πληθυσμού:** $\sigma^2 = \frac{\sum_{i=1}^N (y_i - \mu)^2}{N}$. Εκφράζει την απόλυτη απόσταση των

επιμέρους τιμών του πληθυσμού από την πραγματική τιμή του πληθυσμού. Υψώνεται στο τετράγωνο διότι αλλιώς όλες οι αποκλίσεις $y_i - \mu$ αθροίζουν στο 0. Προφανώς έχει μονάδες μ^2 .

- **Τυπική απόκλιση πληθυσμού:** $\sigma = \sqrt{\frac{\sum_{i=1}^N (y_i - \mu)^2}{N}}$. Απλά αποτελεί τη ρίζα της

διασποράς για να έχουμε την εν λόγω τιμή σε μονάδες μέσου όρου και να μπορούμε να γράψουμε μία μέση τιμή ως $\mu \pm \sigma$.

- **Μέσος όρος δείγματος:** $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$. Εκτίμηση της πραγματικής τιμής του πληθυσμού

όπως βασίζεται στις n παρατηρήσεις με τιμή y_i .

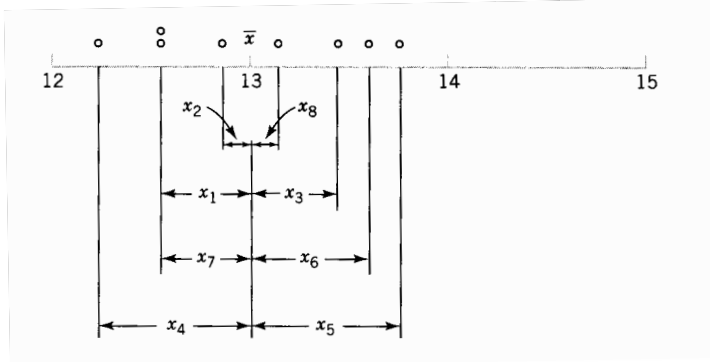
- **Διασπορά δείγματος:** $s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}$. Εδώ η διαφοροποίηση είναι στον

παρανομαστή. Χρησιμοποιούμε $n-1$ και όχι n , διότι θέλουμε να «μεγαλώσουμε» τη διασπορά, αφού η διασπορά των τιμών του δείγματος γύρω από τη μέση τιμή αναμένεται να είναι μεγαλύτερη από τη διασπορά των παρατηρήσεων από την πραγματική τιμή του

πληθυσμού. Εναλλακτικά χρησιμοποιείται και η σχέση: $s^2 = \frac{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n - 1}$. Όπως

είναι προφανές, οι μονάδες της διασποράς είναι οι μονάδες της μέτρησης υψωμένες στο τετράγωνο.

Το σχήμα 3.1 δείχνει το πως η διασπορά του δείγματος χρησιμοποιείται για την μέτρηση της απόκλισης 8 μετρήσεων από τη μέση τιμή $\bar{x} = 13$.



Σχήμα 3.1. Γραφική απεικόνιση της απόκλισης μετρήσεων από τη μέση τιμή (ή πραγματική τιμή). Εδώ η μέση τιμή είναι η τιμή 13.

Το -1 στις παραπάνω σχέσεις που αφορούν σε δείγματα εκφράζει του βαθμούς ελευθερίας του δείγματος. Οι βαθμοί ελευθερίας ορίζονται ως: οι πλεονάζουσες παρατηρήσεις για τον υπολογισμό μίας παραμέτρου. Αυτό σημαίνει ότι για να υπολογίσω την τιμή μίας παραμέτρου θέλω τουλάχιστον μία μέτρηση. Με δύο μετρήσεις της ίδιας παραμέτρου, έχω +1 βαθμό ελευθερίας, δηλ. μία παραπάνω μέτρηση από αυτή που κατ' ελάχιστο απαιτείται για να εκτιμήσω την παράμετρο. Με αριθμό μετρήσεων n , οι βαθμοί ελευθερίας είναι πάντα $n-1$. Αυτό σημαίνει επίσης ότι αν έχω τη μέση τιμή \bar{y} , που υπολογίζεται από n αριθμό y_i , τότε γνωρίζοντας τις $n-1$ διαφορές $y_i - \bar{y}$, μπορώ πάντα να υπολογίσω την εναπομένουσα τιμή y_i , αφού το άθροισμα όλων των n διαφορών $y_i - \bar{y}$ είναι πάντα 0.

- **Τυπική απόκλιση δείγματος:** $s = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}$. Εδώ, αφού μιλάμε πάλι για δείγμα,

κάνουμε χρήση των \bar{y} και $n-1$ αντί για μ και N , αντίστοιχα. Η τυπική απόκλιση χρησιμοποιείται διότι απλά έχει τις ίδιες μονάδες με τις μετρήσεις. Συνεπώς, μπορούμε να πούμε ότι η μέση τιμή κάποιων μετρήσεων είναι $\bar{y} \pm s$ (στη συνέχεια θα δούμε ότι το s αποτελεί ένδειξη της ακρίβειας μετρήσεων).

Οι παραπάνω 6 σχέσεις είναι οι βασικότερες στις στατιστικές αναλύσεις. Παρόλα αυτά, η παρακάτω είναι εξίσου σημαντική αν και καμιά φορά όχι τόσο κατανοητή. Μιλάμε για την τυπική διασπορά του μέσου και το τυπικό σφάλμα του μέσου, τα οποία ορίζονται ως εξής:

- **Τυπική διασπορά μέσου (για δείγμα):** $s_{\bar{y}}^2 = \frac{s^2}{n}$

- **Τυπικό σφάλμα του μέσου (για δείγμα):** $s_{\bar{y}} = \frac{s}{\sqrt{n}}$

Οι 2 παραπάνω παράμετροι έχουν την εξής σημασία: Αν θεωρήσουμε ότι από ένα πληθυσμό N , λάβω πολλά δείγματα μεγέθους n και λάβω τους μέσους όρους \bar{y} των δειγμάτων αυτών, τότε οι πολλοί αυτοί μέσοι όροι θα βρίσκονται κοντύτερα στην πραγματική τιμή του πληθυσμού μ από ότι ο μέσος όρος του κάθε δείγματος με μέγεθος n ξεχωριστά. Επίσης, οι μέσοι αυτοί όροι αναμένεται να κατανέμονται κανονικά με μέση τιμή μ και διασπορά (της μέσης τιμής) σ^2/n . Το τυπικό σφάλμα της μέσης τιμής σ/\sqrt{n} ουσιαστικά περιγράφει το «άπλωμα» των πολλών μέσων τιμών των πολλών δειγμάτων μεγέθους n γύρω από το μ , ενώ η διασπορά σ περιγράφει το «άπλωμα» των πολλών τιμών των μεμονωμένων δειγμάτων γύρω από το μ . Προφανώς, επειδή οι μέσοι όροι των πολλών δειγμάτων αναμένονται να είναι κοντύτερα στον μέσο μ από ότι οι μεμονωμένες παρατηρήσεις y , το τυπικό σφάλμα του μέσου $s_{\bar{y}}$ είναι πάντα μικρότερο της τυπικής απόκλισης s (για αυτόν τον λόγο στον παρονομαστή εμφανίζεται το \sqrt{n}).

Παράδειγμα

Το δείγμα των 27 μετρήσεων νιτρικών έχει μέση τιμή $\bar{y} = 7.51$, όπως είδαμε και τυπική απόκλιση δείγματος ίση με $s = 1.38$ (μη ξεχνάτε ότι οι μονάδες είναι οι ίδιες σε μ και s αλλά προφανώς όχι στη διασπορά s^2). Το τυπικό σφάλμα της μέσης τιμής είναι ίσο με

$$s_{\bar{y}} = 1.38 / \sqrt{27} = 0.27 \frac{mg}{L} \quad (\text{βλέπετε ότι } s_{\bar{y}} < s).$$

3.2 Πειραματικά σφάλματα

Αν τελικώς συλλεχθούν n δείγματα και μετρηθεί κάποια παράμετρος για κάθε δείγμα (y_i), με $i = 1 \dots n$, τότε μπορεί να υπολογιστεί μία μέση τιμή \bar{y} του δείγματος. Η μέση αυτή τιμή συνοδεύεται από ένα απόλυτο ή πειραματικό σφάλμα (δηλαδή μία απόκλιση ή διαφορά) από μία τιμή μ που ορίζεται ως η πραγματική (και περιέργως άγνωστη σε εμάς) τιμή της παραμέτρου που μετράται. Επίσης και οι μεμονωμένες μετρήσεις y_i θα αποκλίνουν από την πραγματική τιμή του πληθυσμού. Το πειραματικό (ή απόλυτο) σφάλμα (ε_{π}) ορίζεται ως:

$$\begin{aligned} \varepsilon_{\pi} &= \bar{y} - \mu \quad \text{ή} \\ \varepsilon_{\pi i} &= y_i - \mu \end{aligned}$$

, όπου το i συμβολίζει τον αριθμό των μεμονωμένων μετρήσεων.

Το πειραματικό σφάλμα, που είναι η απόκλιση από την πραγματική τιμή, (θα καλείται απόλυτο ή πειραματικό σφάλμα από εδώ και τώρα, ενώ συχνά ονομάζεται και «θόρυβος»), προϋποθέτει γνώση της πραγματικής τιμής, και αποτελείται από τα:

- τυχαία σφάλματα (random errors ή ε_{τ}) και,
- συστηματικά σφάλματα (systematic errors ή ε_{σ})⁶.

Δηλαδή,

$$\begin{aligned} \varepsilon_{\pi} &= \varepsilon_{\sigma} + \varepsilon_{\tau} \\ \text{ή } \varepsilon_{\pi i} &= \varepsilon_{\sigma i} + \varepsilon_{\tau i} \end{aligned}$$

Γενικά, ο άρτιος τρόπος αναφοράς οποιουδήποτε πειραματικού αποτελέσματος ή μέτρησης είναι το να συνοδεύεται από το αντίστοιχο «σφάλμα» του, ε_{π} . Δηλαδή για το BOD του ποταμού, θα πρέπει να αναφέρουμε ότι η μέση τιμή του **BOD** είναι **20 mg/L \pm 1 mg/L**, όπου το 1 mg/L είναι το πειραματικό σφάλμα που συνοδεύει τη μέτρησή μας.

Συχνά δεν μπορούμε να διαφοροποιήσουμε μεταξύ τυχαίων και συστηματικών σφαλμάτων, οπότε μιλάμε συνολικά για ένα πειραματικό σφάλμα. Σε κάθε περίπτωση, ο στόχος μας σε ένα εργαστηριακό πείραμα είναι να ελαχιστοποιήσουμε όσο μπορούμε το συστηματικό

⁶ Υπάρχει και τρίτο είδος λαθών – τα χοντρικά σφάλματα – που οφείλονται σε απροσεξίες, ατυχήματα στο εργαστήριο, χρήση μη σωστών αντιδραστηρίων κ.λ.π. Τα χοντρικά σφάλματα καταλήγουν σε τιμές που απέχουν σημαντικά από τις υπόλοιπες επαναληπτικές τιμές (outliers) και γενικώς αγνοούνται. Τα χοντρικά σφάλματα δεν αναλύονται παραπάνω στο παρόν βιβλίο.

σφάλμα, ώστε το πειραματικό σφάλμα να ισούται, όσο αυτό είναι δυνατό, με το τυχαίο σφάλμα, το οποίο δεν μπορεί να αποφευχθεί ποτέ.

1. Συστημικά σφάλματα (*systematic errors ή bias, ϵ_σ*)

Τα συστημικά σφάλματα έχουν καθορισμένη τιμή και ουσιαστικά ταυτίζονται με την έννοια της απόκλισης (*bias*), όπως αυτή θα οριστεί στη συνέχεια. Τα συστημικά σφάλματα λοιπόν είναι υπεύθυνα για την απόκλιση μετρήσεων από την πραγματική τιμή και δεν ελαχιστοποιούνται με επαναλαμβανόμενες μετρήσεις. Ελαχιστοποιούνται μόνο αν ελέγξουμε και βαθμονομήσουμε τα όργανα και είμαστε προσεκτικοί στη διεξαγωγή του πειράματος.

Τα συστημικά σφάλματα αποτελούνται από *σφάλματα οργάνων* (π.χ. διαρροές, επιδράσεις θερμοκρασίας σε ανιχνευτές κ.λ.π.), *προσωπικά σφάλματα* (λόγω επεμβάσεων από τον πειραματιστή, όπως η εκτιμήσεις δεικτών σε ένα μετρητή, εκτίμηση του τελικού χρώματος μετά από εισαγωγή δείκτη, προκατάληψη) και *σφάλματα της μεθόδου* (όπως μη ιδανική συμπεριφορά αντιδραστηρίων, προσροφήσεις υγρών σε στερεά, μη σταθερότητα αντιδραστηρίων, χημικές αλληλεπιδράσεις). Τα *σφάλματα μεθόδου* είναι συνήθως πιο δύσκολα να ανιχνευτούν και διορθωθούν σε σχέση με τις άλλες δύο κατηγορίες λαθών. Ο μόνος τρόπος να ανιχνευτούν είναι με την σύγκριση των αποτελεσμάτων της μεθόδου με πρότυπα υλικά. Σε κάθε περίπτωση, με προσεκτικές αναλύσεις, τα συστημικά σφάλματα μπορούν να ελαχιστοποιηθούν ή και να μηδενιστούν.

2. Τυχαία σφάλματα (*random errors, ϵ_r*)

Τα τυχαία σφάλματα οφείλονται απλά στην ύπαρξη μη προσδιορισμένων λαθών (τυχαίων σφαλμάτων) που εμφανίζονται σε οποιαδήποτε μέτρηση⁷. Η κατανομή των τυχαίων σφαλμάτων είναι κανονική, εφόσον και η κατανομή των μετρούμενων τιμών είναι (αν και δεν είναι πάντα) κανονική (το γνωστό σχήμα της καμπάνας). Οι πολλές επαναλαμβανόμενες μετρήσεις του ίδιου δείγματος είναι ο τρόπος ποσοτικοποίησης των τυχαίων σφαλμάτων.

3. Πρακτικός κανόνας που αφορά σε τυχαία σφάλματα

Όπως περιγράφηκε παραπάνω, η πραγματική τιμή μ είναι η μέση τιμή ενός άπειρου αριθμού δεδομένων N , που αποτελούν τον πληθυσμό. Στην περίπτωση αυτή δεν υπάρχει τυχαίο σφάλμα, αφού η μέση τιμή των ‘πολλών’ δεδομένων (δηλ. η \bar{y}) μπορεί να ταυτιστεί με την πραγματική μέση τιμή του πληθυσμού (δηλαδή $\bar{y} - \mu = 0$). Πρακτικά, σπάνια πραγματοποιούμε άνω των 20 – 30 επαναληπτικών μετρήσεων (συνήθως λόγω κόστους και χρόνου), ενώ συνήθως μπορεί να κάνουμε 3 (το πολύ 5) επαναληπτικές μετρήσεις (όσο περισσότερες τόσο περισσότερο προσεγγίζουμε την πραγματική μέση τιμή). Αν λοιπόν υπερβούμε τις 20-30 επαναληπτικές μετρήσεις, μπορούμε να πούμε πρακτικά ότι προσεγγίζουμε τον πληθυσμό και συνεπώς το τυχαίο σφάλμα πρακτικά μπορεί να μηδενιστεί, και άρα – με την προϋπόθεση ότι έχουν εξαλειφθεί τα συστημικά

⁷ Μόνο στις μετρήσεις ακέραιων αντικειμένων δεν έχουμε τυχαίο σφάλμα, δηλαδή αν κάποιος μετρήσει δέκα κουτιά και μετά τα ξαναμετρήσει, θα καταλήγει πάντα στον αριθμό δέκα.

σφάλματα – η μέση τιμή των δεδομένων ταυτίζεται με την πραγματική μέση τιμή του πληθυσμού. Αυτή είναι μία πρακτική προσέγγιση, όπως φαίνεται και στο παράδειγμα στη συνέχεια.

Παράδειγμα 1

Έστω ότι επαναλαμβάνουμε 50 μετρήσεις ενός μόνο δείγματος (π.χ. συγκέντρωση κάποιας παραμέτρου που μετράται γρήγορα και εύκολα με ένα φασματοφωτόμετρο) και βρούμε μία μέση τιμή ίση με 0.482 ppm. Λόγω του μεγάλου μεγέθους του δείγματος, θεωρούμε ότι η μέση τιμή $\bar{y} = 0.482$ ppm είναι ίση με την πραγματική μέση τιμή μ , και συνεπώς το τυχαίο σφάλμα $\varepsilon_t = \bar{y} - \mu = 0$. Αν επιλέξω περιορισμένο αριθμό μετρήσεων (π.χ 3 μετρήσεις με τιμές 0.488, 0.480, 0.486 τότε η μέση τους τιμή είναι 0.485. Το τυχαίο σφάλμα της μέσης τιμής 0.485 είναι $\varepsilon_t = 0.485 - 0.482 = +0.003$. Αυτή η προσέγγιση θεωρεί ότι έχουν εξαλειφθεί (με πολύ προσοχή) και τα συστηματικά σφάλματα. Αν δεν εξαλειφθούν τα συστηματικά σφάλματα, τότε, πολύ απλά το απόλυτο σφάλμα της μέσης τιμής των 3 μετρήσεων περιλαμβάνει τα τυχαία σφάλματα και τα συστηματικά σφάλματα, ώστε να πούμε τελικά ότι $\varepsilon_{\pi} = \varepsilon_t + \varepsilon_{\sigma} = 0.003$.

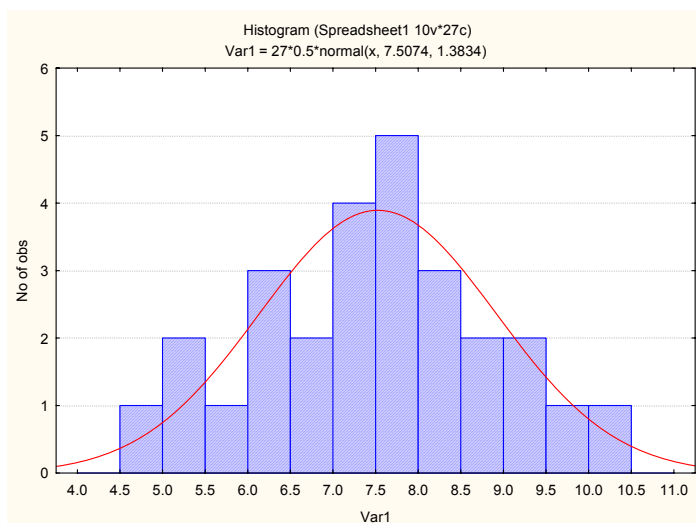
Παράδειγμα 2

Τοποθετώ 27 δείγματα σε ένα εργαστήριο, τα οποία περιέχουν όλα νιτρικά σε γνωστή συγκέντρωση $\mu = 8.0$ mg/L (το δείγματα δηλαδή είναι standard δείγματα). Τα δείγματα τοποθετήθηκαν μαζί με άλλα δείγματα στο εργαστήριο και μετρήθηκαν από χημικό που δεν γνώριζε φυσικά ότι τα συγκεκριμένα 27 δείγματα περιείχαν την παράμετρο αυτή στη γνωστή συγκέντρωση. Γίνονται οι μετρήσεις των 27 αυτών δειγμάτων, και οι τιμές που λαμβάνονται είναι 6.9, 7.8, 8.9, 5.2, 7.7, 9.6, 8.7, 6.7, 4.8, 8.0, 10.1, 8.5, 6.5, 9.2, 7.4, 6.3, 5.6, 7.3, 8.3, 7.2, 7.5, 6.1, 9.4, 5.4, 7.6, 8.1 και 7.9 mg/L.

Ο πραγματική τιμή του πληθυσμού είναι 8.0 mg/L, όπως γνωρίζαμε εξ αρχής. Το δείγμα είναι οι 27 παρατηρήσεις, δηλ. με μέγεθος δείγματος $n=27$. Η τυχαία μεταβλητή είναι η μετρούμενη συγκέντρωση σε κάθε δείγμα που έχει τη γνωστή συγκέντρωση των 8.0 mg/L. Το πειραματικό σφάλμα αναγκάζει τις τιμές που τελικώς μετρήθηκαν (ή παρατηρήθηκαν) να κυμαίνονται γύρω από την πραγματική τιμή των 8.0 και όχι να είναι ίσες με 8.0. Τα πειραματικά σφάλματα για κάθε μέτρηση είναι λοιπόν: 6.9-8.0=-1.1, 7.8-8.0=-0.2, +0.9, -2.8 κ.λ.π. Το πειραματικό αυτό σφάλμα είναι άθροισμα του τυχαίου και του συστηματικού λάθους.

Η μέση τιμή (\bar{y}) των νιτρικών των 27 δειγμάτων είναι: 7.51 mg/L. Σύμφωνα με τον Taylor (1987), η μέση τιμή πρέπει να αναφέρεται με τουλάχιστον ένα επιπρόσθετο σημαντικό ψηφίο ($\Sigma.\Psi.$) συγκρινόμενο με τα $\Sigma.\Psi.^8$ των μετρήσεων.

⁸ Δες παράρτημα για τον ορισμό και ιδιότητες των σημαντικών ψηφίων



Σχήμα 3.2. Διάγραμμα συχνότητας (ή ιστόγραμμα) και κανονική καμπύλη κατανομής για τις 27 μετρήσεις συγκέντρωσης νιτρικών (το διάστημα του ιστογράμματος είναι 0.5 mg/L, π.χ. το ιστόγραμμα μας λέει ότι υπάρχουν 3, εκ των 27, μετρήσεις με συγκέντρωση μεταξύ 6.0 και 6.5 mg/L. Το διάστημα ενός ιστογράμματος το ορίζουμε ανάλογα με τις μετρήσεις. Επίσης στον

κατακόρυφο άξονα μπορεί να μπει το ποσοστό των παρατηρήσεων, για το συγκεκριμένο διάστημα, επί του συνόλου των παρατηρήσεων. Η κόκκινη καμπύλη είναι η προσέγγιση της καμπύλης κανονικής κατανομής για τις 27 αυτές μετρήσεις. Το παραπάνω γράφημα έγινε με χρήση του προγρ. STATISTICA 6.0).

3.3 Ορθότητα, ακρίβεια, απόκλιση

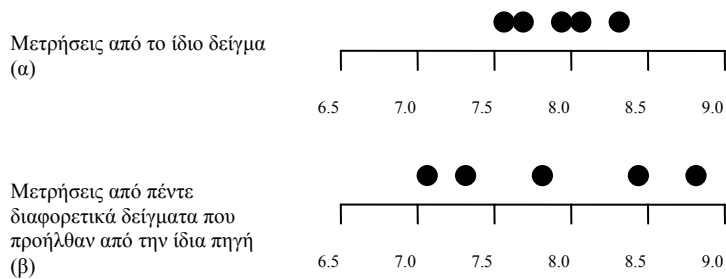
Με βάση τους ορισμούς των παραπάνω σφαλμάτων, η ορθότητα (accuracy) των μετρήσεων ορίζεται από τις παρακάτω δύο βασικές έννοιες:

1. την ακρίβεια (precision)
2. την απόκλιση (bias).

Ακρίβεια (precision)

Ο όρος αυτός εκφράζει την επαναληψιμότητα των αποτελεσμάτων, η οποία επιτυγχάνεται με απλή επανάληψη των μετρήσεων του ίδιου δείγματος. Εκφράζει δηλαδή τις διαφορές των μεμονωμένων μετρήσεων μεταξύ τους. Οι μετρήσεις επαναλαμβάνονται είτε για το ίδιο δείγμα είτε για πολλαπλά δείγματα τα οποία έχουν συλλεχθεί με τον ίδιο τρόπο. Σε κάθε περίπτωση αυτό πρέπει να διευκρινίζεται με σαφήνεια.

Παραδειγματικά, αν πάρουμε πολλά δείγματα νερού από το ίδιο σημείο μίας λίμνης στην ίδια χρονική στιγμή (π.χ. γεμίζοντας πολλά πλαστικά δοχεία), τότε η ακρίβεια των μετρήσεων θα προκύψει ως αποτέλεσμα των μεμονωμένων μετρήσεων σε κάθε ένα δοχείο (μία μέτρηση ανά δοχείο). Η ακρίβεια επίσης μπορεί να εκφραστεί μετά από συλλογή δείγματος νερού *σε ένα δοχείο και πολλαπλές μετρήσεις* της επιθυμητής παραμέτρου από *το ίδιο το δοχείο* (εφόσον η ποσότητα του δείγματος στο εν λόγω δοχείο επαρκεί για τις μετρήσεις, κάτι φυσικά για το οποίο έχουμε θα έχουμε προνοήσει). Πιθανά να υπάρχει διαφοροποίηση των παραπάνω, όπως εκφράζεται και στο διπλανό διάγραμμα σημείων (dot diagram). Σε κάθε περίπτωση, καλό είναι να γίνεται διευκρίνιση για το τι από τα δύο ισχύει.



Σχήμα 3.3. Επαναληπτικές μετρήσεις από (α) το ίδιο δείγμα που προήλθε από μία πηγή (β) από πέντε διαφορετικά δείγματα που προήλθαν από την ίδια αρχική πηγή. Είναι προφανές ότι η ακρίβεια στην (α) περίπτωση είναι μεγαλύτερη σε σχέση με την (β)

Η ακρίβεια θεωρείται ότι εκφράζει το τυχαίο σφάλμα, που αποτελεί το ένα συστατικό του πειραματικού σφάλματος, και το οποίο θα υπάρχει πάντα στις μετρήσεις διότι δεν έχουμε ποτέ τέλειες μετρήσεις. Παράδειγμα, σε ένα σύνολο 5 μετρήσεων ενός δείγματος [π.χ. 38.2, 39.7, 37.1, 39.0, 38.6], με μέση τιμή $\bar{y} = 38.5$, η τυπική απόκλιση s είναι 0.97. Το s αποτελεί εκτίμηση της ακρίβειας των μετρήσεων και ουσιαστικά είναι μέτρο της διασποράς των μετρήσεων γύρω από τη μέση τιμή του δείγματος. Η ακρίβεια όμως του μέσου όρου μετράται από την *τυπική απόκλιση του μέσου όρου* και είναι $s_{\bar{y}} = s/\sqrt{n}$ (επίσης ονομάζεται τυπικό σφάλμα του μέσου όρου). Αυτό αποτελεί μέτρο της διασποράς πολλών μέσων όρων προερχόμενων από πολλά δείγματα μεγέθους n γύρω από τον **πραγματικό μέσο όρο** της διεργασίας.

Οι 3 κύριες παράμετροι που χρησιμοποιούνται για την έκφραση της *ακρίβειας* είναι:

- η τυπική απόκλιση (standard deviation)
- η διασπορά (variance) και,
- ο συντελεστής μεταβλητότητας (coefficient of variation, CV).

Είναι προφανές ότι όσο μεγαλύτερη η ακρίβεια κάποιων μετρήσεων, τόσο μικρότερες οι τιμές των παραπάνω 3 παραμέτρων. Επίσης υπάρχουν και άλλες δύο παράμετροι, όπως φαίνεται στον πίνακα 1. Γενικά, ο συντελεστής μεταβλητότητας (CV) είναι από τους πιο κοινά χρησιμοποιούμενους δείκτες ακρίβειας, γιατί απλά εκφράζεται ως ποσοστό επί της μέσης τιμής.

Σε συνέχεια του σχολίου στη λεζάντα του σχήματος 3, η τυπική απόκλιση στην περίπτωση (α) είναι σαφώς μικρότερη από αυτής της περίπτωσης (β) και άρα η ακρίβεια των μετρήσεων στο (α) μεγαλύτερη από την ακρίβεια στο (β).

Οι σχέσεις υπολογισμού της ακρίβειας συμπεριλαμβάνονται στον πίνακα παρακάτω:

Τυπική απόκλιση	$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$ (έχει τις ίδιες τιμές με τον μέσο όρο και τις μετρήσεις σε αντίθεση με τη διασπορά του δείγματος που είναι s^2)
Σχετική τυπική απόκλιση	$RSD = \frac{s}{\bar{x}} \cdot 100$
Τυπικό σφάλμα του μέσου	$s_y = \frac{s}{\sqrt{n}}$
Συντελεστής μεταβλητότητας (coefficient of variation)	$CV = \frac{s}{x} \cdot 100\%$
Διασπορά δείγματος (variance)	s^2

Απόκλιση (Bias)⁹

Η μέση τιμή των δειγμάτων, \bar{y} , όπως αναφέρθηκε, αναμένεται να διαφέρει από την πραγματική μέση τιμή (μ) του πληθυσμού, αφού η τελευταία εξάλλου δεν είναι συχνά ποτέ γνωστή. Η διαφορά αυτή είναι η απόκλιση (bias), η οποία όμως οφείλεται στα συστηματικά σφάλματα, όπως αυτά ορίστηκαν παραπάνω. Επίσης, ως απόκλιση ορίζεται η διαφορά της τιμής ενός μεμονωμένου δείγματος (y_i) από την πραγματική τιμή του πληθυσμού (μ). Δηλαδή, η απόκλιση τελικά είναι:

$$\text{αποκλιση} = \mu - \bar{y} \quad \text{ή} \quad \text{αποκλιση}_i = \mu - y_i$$

Η πραγματική τιμή μ δεν είναι ποτέ γνωστή γιατί απαιτεί θεωρητικά ένα πολύ μεγάλο αριθμό δειγμάτων. Πρακτικά όμως, ένας πρακτικός κανόνας λέει ότι η πραγματική τιμή του πληθυσμού προσεγγίζεται σε μεγάλο βαθμό αν χρησιμοποιηθούν άνω των 20 ή 30 δειγμάτων. Συνεπώς, η μέση τιμή \bar{x} του δείγματος των 20 ή 30 μετρήσεων μπορεί να ταυτιστεί με τη πραγματική τιμή μ του πληθυσμού (δείτε παράδειγμα σε προηγούμενο κεφάλαιο).

Υπάρχουν δύο τρόποι «χειρισμού» της απόκλισης, οι οποίες είναι:

A) οι μετρήσεις μας είναι ορθές, εφόσον ο μέσος όρος των μετρήσεων βρίσκεται κοντά στην πραγματική τιμή ανεξάρτητα από την απόκλιση των μεμονωμένων μετρήσεων από την πραγματική τιμή.

B) κάθε μεμονωμένη μέτρηση πρέπει να βρίσκεται κοντά στην πραγματική τιμή.

Γενικά, η προσέγγιση B κρίνεται ως η ορθότερη και η περισσότερο χρησιμοποιημένη.

⁹ Ο όρος απόκλιση (bias) δεν πρέπει να συγχέεται με την τυπική απόκλιση (standard deviation)

Η μέτρηση της απόκλισης στο εργαστήριο προαπαιτεί την προετοιμασία και ανάλυση τυποποιημένων δειγμάτων αναφοράς (standards). Τα δείγματα αναφοράς μπορούν να προετοιμαστούν στο εργαστήριο ή να αγοραστούν έτοιμα από τους προμηθευτές χημικών ουσιών. Στη συνέχεια, μετρώνται τουλάχιστον 20 δείγματα των τυποποιημένων δειγμάτων και λαμβάνεται η μέση τιμή των 20 δειγμάτων (μ) – που θεωρείται ότι ταυτίζεται με την πραγματική τιμή του πληθυσμού. Στη συνέχεια υπολογίζεται για την κάθε μέτρηση x_i η απόκλιση από την πραγματική τιμή, ως η διαφορά $\mu - x_i$. Συνεπώς, εξάγονται 20 τέτοιες διαφορές που πρέπει να αθροίζονται στο 0 αλλά και να έχουν μέση τιμή κοντά στο 0.

Με βάση τα παραπάνω λοιπόν, η απόκλιση (bias) ταυτίζεται με το συστηματικό σφάλμα (ϵ_s). Σε αντίθεση με τα τυχαία σφάλματα, τα συστηματικά σφάλματα (δηλ. η απόκλιση) δεν μπορούν να μειωθούν κάνοντας επαναλαμβανόμενες μετρήσεις. Επίσης, για να προσδιοριστούν πρέπει να είναι γνωστή η πραγματική τιμή του πληθυσμού.

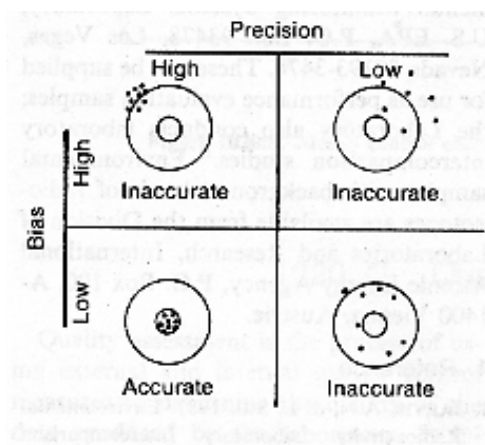
Παράδειγμα

Για τις 27 μετρήσεις των νιτρικών που προαναφέρθηκαν, η απόκλιση (bias) του μέσου όρου των μετρήσεων από την πραγματική τιμή είναι:

$$\text{Απόκλιση} = \bar{y} - \mu = 7.51 - 8.00 = -0.49 \text{ mg/L}$$

Ορθότητα (accuracy)

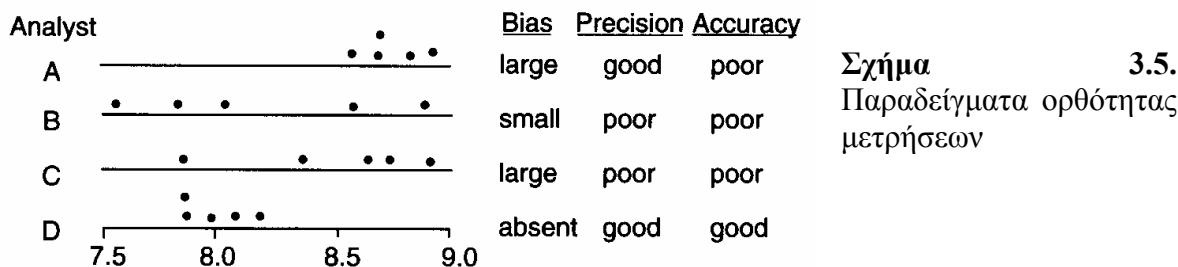
Η ορθότητα των μετρήσεων λοιπόν απαιτεί συνδυαστικά υψηλή ακρίβεια και χαμηλή απόκλιση, όπως φαίνεται στο παρακάτω σχήμα.



Σχήμα 3.4. Απεικόνιση των εννοιών της ορθότητας (accuracy), ακρίβειας (precision) και απόκλισης (bias). Μια ορθή μέτρηση απαιτεί υψηλή ακρίβεια και χαμηλή απόκλιση όπως δηλ. στο κάτω – αριστερά τμήμα του σχήματος. Το κέντρο του κύκλου συμβολίζει την πραγματική τιμή των μετρήσεων.

Το σχήμα 3.5 απεικονίζει όμοια τη σχέση μεταξύ ορθότητας (accuracy), απόκλισης (bias) και ακρίβειας (precision) για μετρήσεις με πραγματική τιμή 8.0 mg/L. Ορθές μετρήσεις προέρχονται τελικά μόνο από τον αναλυτή D (τελευταία σειρά). Ο αναλυτής A έχει μέν υψηλή ακρίβεια αλλά υψηλή απόκλιση (από την πραγματική τιμή των 8.0 mg/L), ενώ ο

αναλυτής Β έχει σχετικά μικρή απόκλιση αλλά χαμηλή ακρίβεια. Τέλος ο αναλυτής Γ έχει υψηλή απόκλιση και επίσης χαμηλή ακρίβεια.



Το παράδειγμα στη συνέχεια είναι καλό για να δείξει τις διαφοροποιήσεις μεταξύ της ακρίβειας των μετρήσεων και της απόκλισης.

Παράδειγμα

Έστω ότι σε δύο εργαστήρια δίνονται 14 δείγματα ενός τυποποιημένου διαλύματος που περιέχει συγκέντρωση ιόντων χλωρίου ίση με 2.50 mg/L. Επίσης, σε κανένα εργαστήριο δεν γνωστοποιήθηκε η γνωστή αυτή συγκέντρωση του χλωρίου. Τα αποτελέσματα των 2 εργαστηρίων ήταν:

Εργαστήριο A: $\bar{y}_A = 2.66$ mg/L, $s = 0.32$ mg/L, Απόκλιση = $2.66 - 2.50 = 0.16$ mg/L

Εργαστήριο B: $\bar{y}_B = 4.38$ mg/L, $s = 0.50$ mg/L, Απόκλιση = $4.38 - 2.50 = 1.88$ mg/L

Επειδή οι κατανομές των δεδομένων θεωρούνται κανονικές (και κατανομονται βάσει της κατανομής t, αφού το δείγμα είναι μικρό), θα ελέγξουμε το διάστημα εμπιστοσύνης¹⁰ (με 95% εμπιστοσύνη) για την απόκλιση του κάθε εργαστηρίου ξεχωριστά. Δηλαδή θα ελέγξουμε αν το διάστημα εμπιστοσύνης για τις διαφορές $\bar{y}_A - \mu$ και $\bar{y}_B - \mu$ ξεχωριστά, περιέχουν το 0 ή όχι. Συγκεκριμένα για 13 βαθμούς ελευθερίας και $\alpha=0.05$, έχω τιμή $t_{13,0.025}=2.160$. Τα τυπικά σφάλματα των μέσων τιμών για το εργαστήριο A και B αντίστοιχα είναι:

$$s_{\bar{y}_A} = s_A / \sqrt{14} = 0.32 / 3.74 = 0.18 \text{ \&}$$

$$s_{\bar{y}_B} = s_B / \sqrt{14} = 0.50 / 3.74 = 0.29 .$$

Τελικά τα διαστήματα εμπιστοσύνης των αποκλίσεων για τα εργ. A και B είναι:

$$(\bar{y}_A - 2.50) \pm t_{13,0.025} s_{\bar{y}_A} = 0.16 \pm 0.18 = (-0.03, 0.34)$$

$$(\bar{y}_B - 2.50) \pm t_{13,0.025} s_{\bar{y}_B} = 1.88 \pm 0.29 = (1.58, 2.16)$$

Στο εργαστήριο A λοιπόν μπορούμε να πούμε ότι η πραγματική τιμή της απόκλισης περιέχεται κατά 95% στο διάστημα -0.03 και 0.34, διάστημα που περιέχει το 0. Άρα με πιθανότητα 95% η απόκλιση είναι μηδενική για το εργαστήριο A (άρα μετράει καλά).

¹⁰ Η έννοια του διαστήματος εμπιστοσύνης θα οριστεί σε επόμενο κεφάλαιο λεπτομερώς.

Για το B όμως, το διάστημα εμπιστοσύνης είναι σταθερά θετικό, που σημαίνει ότι το B εξάγει σταθερά τιμές άνω της πραγματικής τιμής (άρα δεν μετράει καλά).

Το αρχικό συμπέρασμα που βγαίνει είναι ότι η απόκλιση του B εργαστηρίου είναι μεγαλύτερη του A. Οι ακρίβειες των μετρήσεων είναι περίπου οι ίδιες. Συγκεκριμένα, θα κάναμε χρήση του F-test βάσει του οποίου θα απορρίπταμε την υπόθεση ότι οι ακρίβειες είναι ίδιες, όταν η τιμή F ήταν μεγαλύτερη του 3.0. Η τιμή F είναι ο λόγος των διασπορών s^2_A/s^2_B , που είναι ίσος με 2.5. Επειδή, ο λόγος δεν είναι μεγαλύτερος από 3, τότε οι διασπορές, και άρα οι τυπικές αποκλίσεις (και συνεπώς οι ακρίβειες) είναι όμοιες.

3.4 Επαναληψιμότητα και αναπαραγωγικότητα

Η ορθότητα (accuracy) αποτελείται από 2 επιπλέον έννοιες, την *επαναληψιμότητα* (repeatability) και την *αναπαραγωγικότητα* (reproducibility).

Η *επαναληψιμότητα* (repeatability) περιγράφει την απόκλιση μεταξύ αποτελεσμάτων ανεξαρτήτων μετρήσεων εντός του ίδιου εργαστηρίου, που υλοποιούνται σε σχετικά σύντομο χρονικό διάστημα από τον ίδιο εργαστηριακό υπάλληλο με ένα συγκεκριμένο σετ χημικού εξοπλισμού που χρησιμοποιεί δείγματα που λαμβάνονται από μία συγκεκριμένη ποσότητα ομογενούς υλικού.

Η *αναπαραγωγικότητα* (reproducibility) περιγράφει την απόκλιση μεταξύ μεμονωμένων μετρήσεων που λήφθηκαν τυχαία και υπό διαφορετικές συνθήκες από το ίδιο ομογενές υλικό με χρήση της ίδιας μεθόδου μέτρησης. Οι διαφορετικές συνθήκες μπορούν πιθανά να είναι διαφορετικοί εργαστηριακοί υπάλληλοι εντός του ίδιου εργαστηρίου ή μπορούν να είναι μετρήσεις που πραγματοποιήθηκαν σε διαφορετικά εργαστήρια (πάντα με το ίδιο αρχικό υλικό).

Η *αναπαραγωγικότητα* αποτελεί εξαιρετικά βασική επιστημονική αρχή και είναι πίο σημαντική από την επαναληψιμότητα. Οι ορθές και αποδεκτές μετρήσεις, όπως και ένα ορθό και αποδεκτό πείραμα, πρέπει να είναι αναπαραγωγίσιμες/ο. Για αυτό το λόγο σε επιστημονικά άρθρα, οι εργαστηριακές μετρήσεις πρέπει να περιγράφονται με σαφήνεια και λεπτομέρεια, ώστε να είναι δυνατή η «αναπαραγωγή» τους (ή η μη αναπαραγωγή τους) σε διαφορετικό εργαστήριο ή/και από διαφορετικό χημικό.

Παράδειγμα 3

Έστω ότι από ένα δείγμα 500 g ξηραμένων και κονιορτοποιημένων στερεών αποβλήτων λαμβάνουμε τυχαία 5 δείγματα αρχικής μάζας 1 g. Κάνουμε μέτρηση των πτητικών στερεών και εξάγουμε τις τιμές 58,3%, 59,1%, 56,4%, 56,5%, 57,2% (επί ξηρού βάρους). Ποία είναι η ακρίβεια της μέτρησης?

Αρχικά υπολογίζεται η τυπική απόκλιση (απόλυτη) βάσει της σχέσης (1)¹¹. Για το συγκεκριμένο παράδειγμα: $s=1.17$, $x=57.50$. Σημειώνεται, όπως είναι προφανές, ότι η τυπική απόκλιση και η μέση τιμή έχουν τις ίδιες μονάδες (ενώ δεν ισχύει το ίδιο φυσικά για τη διασπορά s^2). Το CV είναι τελικά: 2.04%. Πρακτικά, αν και αυτό ορίζεται από το κάθε εργαστήριο ξεχωριστά, CV μικρότερα του 5% θεωρούνται χοντρικά ως ενδείξεις ομοιογένειας στο δείγμα, όπως επίσης και ενδείξεις ότι τα όργανα κάνουν συνεπείς (αμετάβλητες) αναλύσεις κατά τη διάρκεια των συγκεκριμένων μετρήσεων.

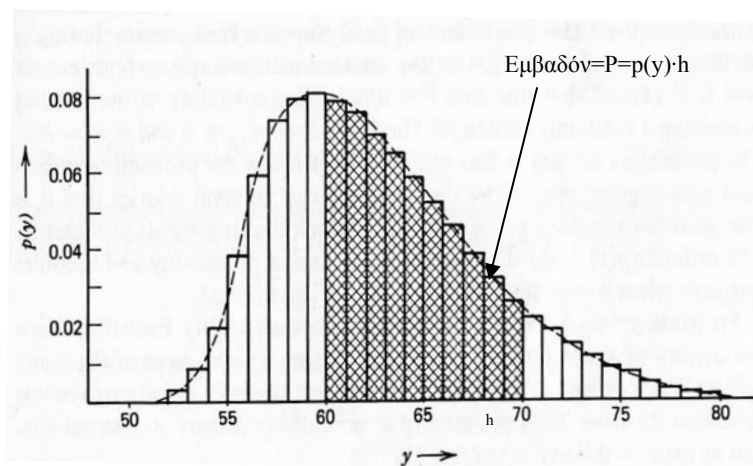
¹¹ Μπορεί να χρησιμοποιηθεί η συνάρτηση STDEV του EXCEL που εξάγει τα ίδια αποτελέσματα. Για τη μέση τιμή μπορεί να χρησιμοποιηθεί η συνάρτηση AVERAGE του EXCEL. Ο συντελεστής διασποράς είναι το κλάσμα STDEV προς AVERAGE.

4. Βασικές κατανομές στη περιβαλλοντική μηχανική

Κάθε πληθυσμός μπορεί να περιγραφεί από μία κατανομή πιθανότητας με συνάρτηση $p(y)$. Η έννοια της κατανομής πιθανότητας είναι ότι το γινόμενο της τιμής της συνάρτησης της καμπύλης στον άξονα y για την τιμή y , δηλ. το $p(y)$, επί του (απειρο)ελαχίστου διαστήματος h στο οποίο «ανήκει» η μέτρηση, αποτελεί την πιθανότητα P - σε % - της συγκεκριμένης παρατήρησης να εξαχθεί όταν τυχαία «συλλέξουμε» ένα δείγμα από τον πληθυσμό. Το $p(y)$ ονομάζεται πυκνότητα της πιθανότητας. Δεν είναι το ίδιο πιθανότητα.

Έτσι, έχοντας την καμπύλη πιθανότητας για ένα πληθυσμό, μπορούμε να βρούμε τις πιθανότητες να είναι κάποια τιμή του πληθυσμού μεγαλύτερη ή μικρότερη από κάποια γνωστή τιμή y_0 ή μεταξύ κάποιων τιμών y_1 & y_2 , μεγαλύτερη από y_1 , μικρότερη από y_2 κ.λ.π., βασιζόμενοι πάντα στο ότι το συνολικό εμβαδόν κάτω από την καμπύλη είναι 1 ή 100%.

Αφού όμως ποτέ δεν έχουμε ένα πληθυσμό διαθέσιμο, και πάντα έχουμε δείγματα του πληθυσμού, τότε το ιστόγραμμα (ή κατανομή συχνότητας) αποτελεί το εργαλείο για να «δούμε» ή καλύτερα προσεγγίσουμε την καμπύλη πιθανότητας του αντίστοιχου πληθυσμού. Αυτό φαίνεται και στο παρακάτω σχήμα.



Σχήμα 4.1 Τυπική κατανομή σε περιβαλλοντικά συστήματα (που όπως φαίνεται δεν είναι κανονική). Ο οριζόντιος άξονας έχει τις τιμές των παρατηρήσεων (μετρήσεων) και συνεπώς αντίστοιχες μονάδες, ενώ ο κατακόρυφος άξονας $p(y)$ είναι μία συνάρτηση που αν πολλαπλασιαστεί με κάποιο διάστημα h επί

του άξονα y , αποτελεί την πιθανότητα (επιφάνεια κάτω από την καμπύλη) που έχει το y να επιλεγεί μέσω κάποιου δείγματος από τον πληθυσμό.

Δηλαδή, η τιμή της συνάρτησης $p(y)$ δεν είναι πιθανότητα. Γίνεται πιθανότητα P μόνο όταν πολλαπλασιαστεί με κάποιο διάστημα h του άξονα y . Δηλαδή, $P = p(y) \cdot h$. Γενικά, η επιφάνεια κάτω από την καμπύλη είναι η πιθανότητα. Παραδειγματικά, η πιθανότητα ώστε μία τιμή του πληθυσμού να ανήκει σε όλον τον πληθυσμό είναι η επιφάνεια κάτω από το σύνολο της καμπύλης, άρα είναι 1. Η παραπάνω καμπύλη (ή το ιστόγραμμα), μας βοηθάει να δούμε τα παρακάτω:

1. Η πιθανότητα οτι κάποια τιμή y του πληθυσμού είναι μικρότερη κάποιας γνωστής τιμής y_0 είναι ίση με την επιφάνεια κάτω από την καμπύλη αριστερά του y_0
2. Η πιθανότητα οτι κάποια τιμή y του πληθυσμού είναι μεγαλύτερη κάποιας γνωστής τιμής y_0 είναι ίση με την επιφάνεια κάτω από την καμπύλη δεξιά του y_0
3. Η πιθανότητα οτι κάποια τιμή y του πληθυσμού είναι μεταξύ κάποιων γνωστών τιμών y_0, y_1 είναι ίση με την επιφάνεια κάτω από την καμπύλη μεταξύ των y_0 & y_1

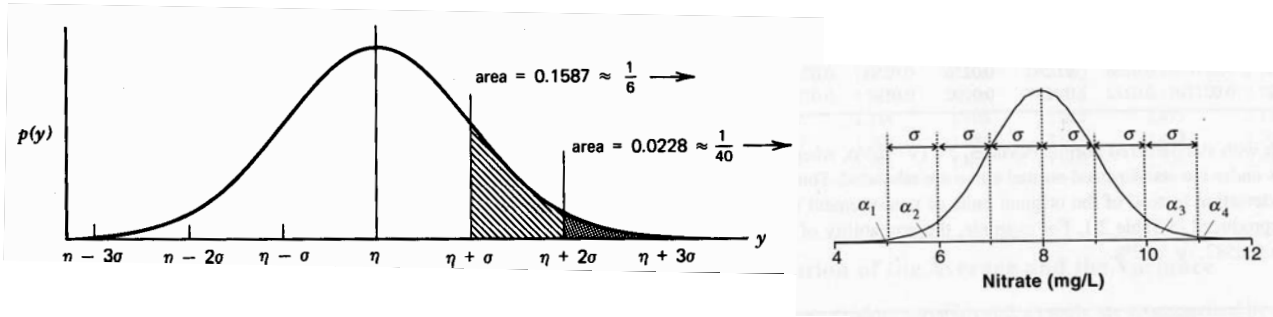
Τελικώς, μιλάμε για ολοκληρώματα της $p(y)$ μεταξύ κάποιων τιμών y_i , που αποτελούν τις πιθανότητες να ανήκουν οι εν λόγω τιμές στον πληθυσμό.

Στο ερώτημα πως φτιάχνουμε πρακτικά μία τέτοια καμπύλη, η απάντηση είναι οτι γενικά μας βολεύει να έχουμε κατ' αρχάς όσο το δυνατό περισσότερες τιμές. Στη συνέχεια κατασκευάζουμε ένα ιστόγραμμα – βάσει επιλεγόμενου διαστήματος h - το οποίο πιθανά να μας δώσει μία ένδειξη του ποία μπορεί να είναι κατανομή του πληθυσμού από την οποία προέρχεται το δείγμα το οποίο παρατηρήσαμε.

Αν και η περισσότερο κοινή κατανομή είναι η κανονική κατανομή, αυτό δεν είναι απαραίτητα αληθές για τις περιβαλλοντικές μετρήσεις. Το σχήμα 4.1 αποτελεί μια τυπική κατανομή στην περιβαλλοντική επιστήμη και μηχανική, που χαρακτηρίζεται από μία κορύφωση στα αριστερό περίπου τμήμα της καμπύλης (και όχι στο μέσο, όπως συμβαίνει με την κανονική κατανομή).

4.1 Κανονική κατανομή

Η κανονική κατανομή προέρχεται από ένα πληθυσμό με μέσο όρο μ και τυπική απόκλιση σ . Έχει το γνωστό σχήμα της «καμπάνας», όπως φαίνεται στο παρακάτω σχήμα.



(α)

(β)

Σχήμα 4.2. (α) Χαρακτηριστικές διαστάσεις σε μία κανονική κατανομή (προσοχή: η μέση τιμή συμβολίζεται εδώ με η αντί για μ), (β) κανονική κατανομή για το δείγμα των νιτρικών με μέση τιμή = 8 mg/L και $\sigma = 1$ mg/L.

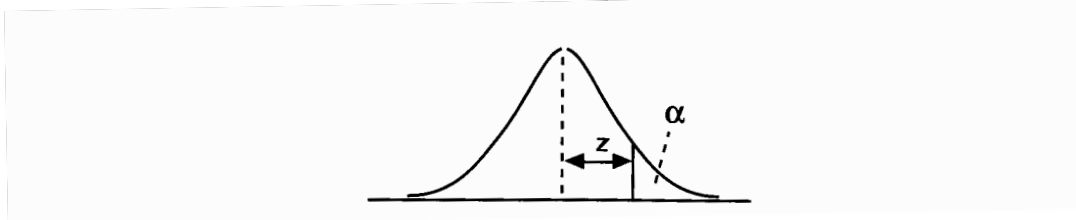
Για τον πληθυσμό των δειγμάτων νιτρικών με γνωστή συγκέντρωση 8 mg/L, τότε $\mu = 8$ mg/L, σ είναι η τυπική απόκλιση, που τυχαίνει να είναι ίση με 1 mg/L (βλέπε σχήμα 4.2β). Θεωρώντας οτι η κανονική κατανομή είναι γνωστή, υπενθυμίζονται τα εξής .

- Το σύνολο της επιφάνειας κάτω από την καμπύλη είναι ίσο με 1, που υποδηλώνει την πιθανότητα (100%) που έχει μία οποιαδήποτε μέτρηση από τον πληθυσμό να ανήκει στον πληθυσμό (δηλ. να είναι 1)
- Η μέση τιμή του πληθυσμού είναι η μ (εδώ 8 mg/L) και η τυπική απόκλιση σ (εδώ 1 mg/L). Η πιθανότητα μία τυχαία τιμή του πληθυσμού να είναι μεγαλύτερη (ή μικρότερη) από συν μία τυπική απόκλιση από τον μέσο όρο - δηλαδή μεγαλύτερη του 9 mg/L - (ή μικρότερη του 7 mg/L) είναι 15.87%, που είναι και η επιφάνεια δεξιά του 9 mg/L (ή αριστερά του 7 mg/L).
- Η πιθανότητα μία τυχαία μέτρηση του πληθυσμού να βρίσκεται κοντά στον μέσο όρο και συγκεκριμένα στο διάστημα $\mu \pm \sigma$ είναι 68.26%.
- Η πιθανότητα ότι μία τιμή θα είναι μεγαλύτερη κατά 2σ (ή μικρότερη κατά 2σ) από την μέση τιμή είναι 4.56% (εμβαδό α_3 ή α_2).
- Συνήθως τυποποιούμε την κανονική κατανομή ενός πληθυσμού μετατρέποντας την σε τιμές $z=(y-\mu)/\sigma$, ώστε η μέση τιμή να είναι το 0 και η τυπική απόκλιση το 1. Μιλάμε τότε για μία κανονική κατανομή με $N(0,1)$. Το σ βέβαια σπάνια είναι γνωστό και συνεπώς συνήθως χρησιμοποιούμε το s , που μας οδηγεί στην κατανομή t .

Για μία κανονική κατανομή ισχύουν τα παρακάτω:

- Η πιο συχνά παρατηρούμενη τιμή στην κανονική καμπύλη κατανομής των μετρήσεων είναι η μέση τιμή των δεδομένων.
- Τα αποτελέσματα κυμαίνονται συμμετρικά γύρω από τη μέση τιμή

Με βάση το παρακάτω σχήμα για την κανονική κατανομή $N(0,1)$, οι τιμές z, α μπορούν να αναζητηθούν σε «κλασσικό» πίνακα του παραρτήματος.



Σχήμα 4.3. Χαρακτηριστικές παράμετροι σε μία κανονική κατανομή (τιμή z, α)

4.2 Κατανομή t

Για λόγους ευκολίας, και επειδή το σ συνήθως δεν είναι γνωστό (το μ παρόλα αυτά μπορεί να είναι), τότε χρησιμοποιούμε τη συνάρτηση t με γενική μορφή:

$$t = (y - \mu) / s$$

δηλαδή όπου σ βάζουμε s που είναι και το γνωστό. Το t έχει και αυτό μία κατανομή που λέγεται t -κατανομή, που ομοιάζει με την κανονική, και από μία άποψη την αντικαθιστά στην πράξη, αφού η κανονική κατανομή είναι θεωρητική. Στην πράξη δηλαδή έχουμε πάντα κάποιο δείγμα με γνωστό μέγεθος ($n =$ κάποιος αριθμός ακέραιος και s γνωστό)

και με αυτό τον τρόπο προσπαθούμε να μάθουμε την κατανομή του πληθυσμού από όπου προέρχεται το δείγμα.

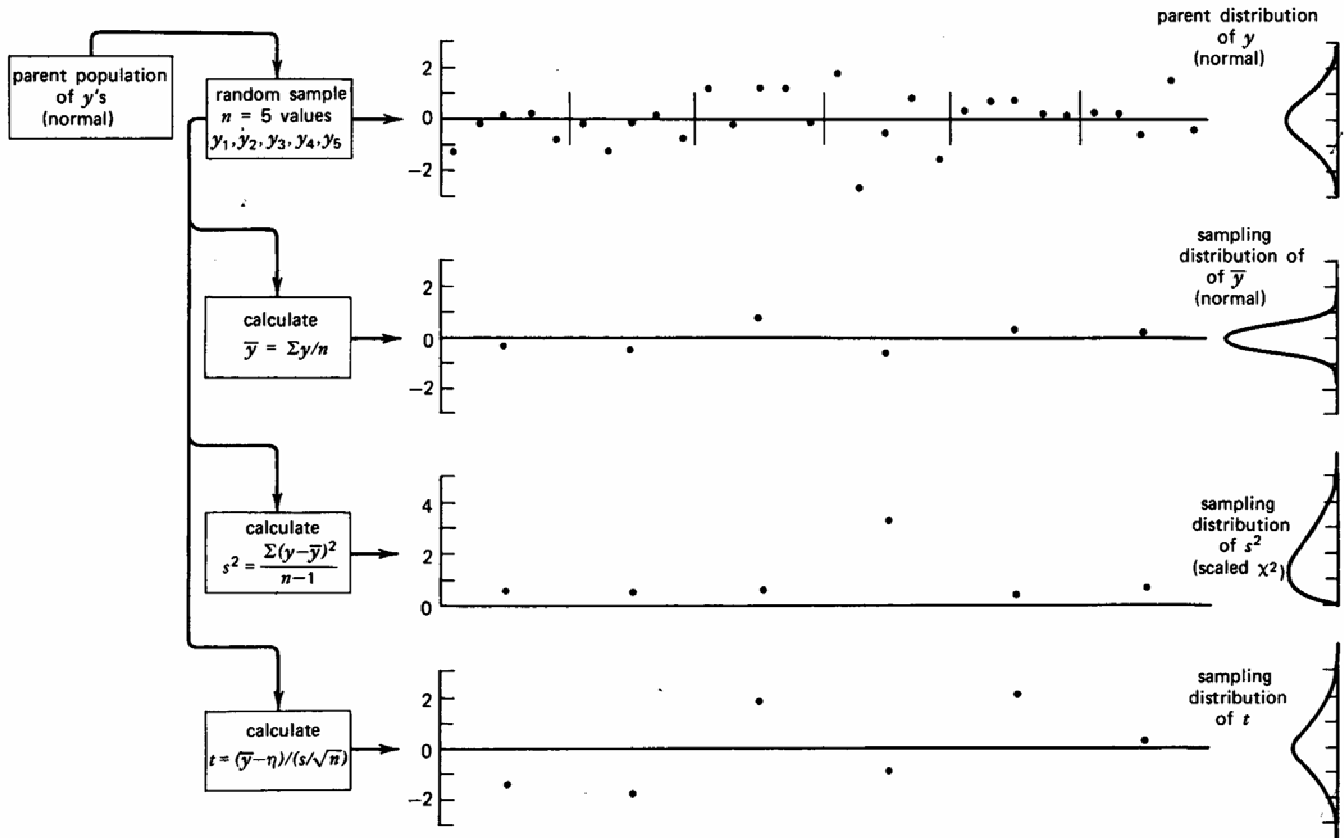
Η μορφή της κατανομής t εξαρτάται από το μέγεθος του δείγματος. Όταν το δείγμα είναι άπειρο, τότε δεν υπάρχει αβεβαιότητα στη διασπορά s^2 ($s^2 = \sigma^2$) και τότε η t είναι η κανονική κατανομή. Παρόλα αυτά, όταν ο αριθμός των μετρήσεων δεν είναι μεγάλος (ώστε να προσεγγίζει την κανονική κατανομή) τότε οι άκρες της καμπύλης t είναι μακρύτερες σε σχέση με την κανονική. Το πλάτος λοιπόν της t εξαρτάται από το s^2 , το οποίο με τη σειρά του εξαρτάται και διαφοροποιείται ανάλογα με τους βαθμούς ελευθερίας (ν).

Η πρακτική χρήση της κατανομής t είναι να μας δείξει την πιθανότητα να ανήκει μία τιμή κάπου. Συγκεκριμένα η επιφάνεια της ουράς (στα δεξιά) της καμπύλης της κατανομής t (επιφάνεια α) αποτελεί την πιθανότητα του t να ξεπεράσει κάποιο συγκεκριμένη τιμή.

Είναι σημαντικό ότι η τιμή t ακολουθεί την κατανομή t , υπό τις παρακάτω προϋποθέσεις:

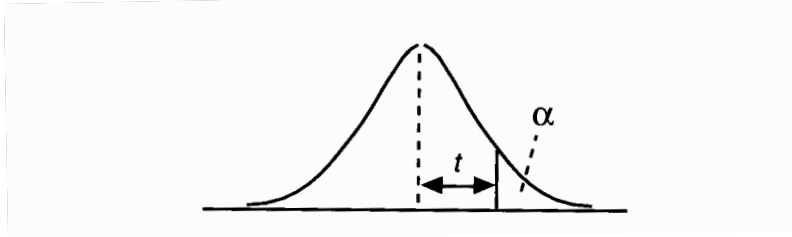
- Όταν το y κατανέμεται κανονικά γύρω από την τιμή μ με διασπορά σ^2
- Όταν η κατανομή του s είναι ανεξάρτητη της κατανομής του μ . Δηλαδή, όταν η διασπορά του δείγματος δεν αυξάνεται, ούτε μειώνεται όταν η μέση τιμή αυξάνεται ή μειώνεται αντίστοιχα.
- Όταν η ποσότητα s^2 υπολογίζεται από μετρήσεις που κατανέμονται κανονικά και ανεξάρτητα με διασπορά σ^2 . Σημειώνεται ότι η κατανομή του s^2 είναι η κατανομή χ^2 , που δεν είναι κανονική κατανομή, και έχει χαρακτηριστικά διαφορετικό σχήμα από την κανονική κατανομή.
- Όταν ισχύουν τα παραπάνω, τότε οι μέσες τιμές \bar{y} επίσης κατανέμονται κανονικά με διασπορά ίση με σ^2/n .
-

Οι παραπάνω προϋποθέσεις – όπως και οι τύποι υπολογισμού των αναγραφόμενων παραμέτρων - φαίνονται στο (κλασσικό) σχήμα από τους Box, Hunter & Hunter στη συνέχεια.



Σχήμα 4.4. Τυχαία δειγματοληψία από μία κανονική κατανομή $N(0,1)$ για την απεικόνιση των κατανομών δειγματοληψίας \bar{y} , s^2 & t (Box et al., 1978, pp. 71).

Με βάση το παρακάτω σχήμα για την κατανομή t με μέση τιμή 0 , οι τιμές t , α μπορούν να αναζητηθούν σε επίσης «κλασσικό» πίνακα του παραρτήματος. Εδώ η διαφορά είναι ότι οι τιμές t διαφοροποιούνται ανάλογα με τους βαθμούς ελευθερίας (άρα τους αριθμούς των μετρήσεων). Επίσης, ο πίνακας περιέχει τις τιμές t για συγκεκριμένα εμβαδά (τιμές α), σε αντίθεση με τον πίνακα της κανονικής κατανομής.



Σχήμα 4.5. Χαρακτηριστικές παράμετροι σε μία t κατανομή (τιμή t , α)

Η τιμή t υπολογίζεται τελικά με βάση την παρακάτω σχέση:

$$t = \frac{\bar{y} - \mu}{s / \sqrt{n}} = \frac{\bar{y} - \mu}{SE_{\bar{y}}}$$

Εδώ, ίσως να αποτελεί απορία η χρήση του τυπικού σφάλματος του μέσου στον παρονομαστή, αντί για την τιμή s . Αυτό εξηγείται ως εξής: η κατανομή των μέσων όρων \bar{y} προσεγγίζει την κανονική κατανομή, ακόμα και αν η κατανομή των πρωτογενών δεδομένων, του πληθυσμού, δεν είναι κανονική. Επίσης, όσο ο αριθμός των παρατηρήσεων αυξάνεται, τόσο η κατανομή του \bar{y} γίνεται περισσότερο κανονική. Συνεπώς μπορεί να χρησιμοποιηθεί η κανονική κατανομή με μέση τιμή μ και διασπορά (του μέσου) σ^2/n για να περιγράψουμε την \bar{y} . Επειδή η διασπορά σ^2 δεν είναι γνωστή, κάνουμε χρήση s^2 . Τελικώς, αφού μιλάμε για κατανομή του \bar{y} , θα κάνουμε χρήση του $s_{\bar{y}}$. Τελικά, αν η αρχική κατανομή είναι κανονική και η διασπορά του πληθυσμού εκτιμάται από το s^2 , τότε η τιμή $t = (\bar{y} - \mu) : s / \sqrt{n}$ (ο τυποποιημένος μέσος όρος) θα ακολουθεί μία κατανομή t με n βαθμούς ελευθερίας.

Παράδειγμα

Για τα δεδομένα των νιτρικών, η μέση τιμή \bar{y} του δείγματος είναι 7.51 mg/L που απέχει (αρκετά?) από την πραγματική τιμή των 8.0 mg/L. Το αν είναι αρκετά μακριά ή όχι από την πραγματική τιμή θα το μελετήσουμε. Στόχος λοιπόν είναι να δούμε ποιες είναι οι πιθανότητες μία μέση τιμή ίση με 7.51 να μετρηθεί στο εν λόγω εργαστήριο. Κάνουμε χρήση της συνάρτησης t :

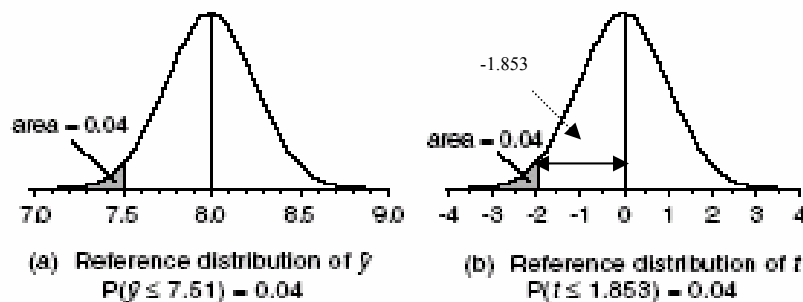
$$t = \frac{\bar{y} - \mu}{s / \sqrt{n}} = t = \frac{7.51 - 8.00}{1.38 / \sqrt{27}} = -1.842$$

με 26 βαθμούς ελευθερίας (δηλ. 27-1). Η κατανομή t έχει συγκεκριμένη μορφή ανάλογα με τους βαθμούς ελευθερίας. Αυτά έχουν καταχωρηθεί σε ειδικούς πίνακες, οπότε πάντα όταν εξάγουμε μία τέτοια τιμή, κάνουμε χρήση των πινάκων αυτών (βλέπε παράρτημα βιβλίου). Αυτά τα πινάκια ουσιαστικά υπολογίζουν την επιφάνεια κάτω από την καμπύλη της κατανομής t μεταξύ κάποιων σημείων ή αριστερά ή δεξιά κάποιου σημείου και συνεπώς την πιθανότητα να είναι κάποια τιμή σε κάποια περιοχή (αφού το συνολικό εμβαδό κάτω από την καμπύλη είναι πάντα 1 που αντιστοιχεί με πιθανότητα 100%).

Στο παράδειγμά μας τώρα, ψάχνουμε την τιμή 1.842 στη στήλη με τους 26 βαθμούς ελευθερίας. Δεν υπάρχει ακριβώς η ίδια τιμή, αλλά υπάρχει η 1.706 για πιθανότητα 0.05, η 2.056 για πιθανότητα 0.025. Άρα σίγουρα η πιθανότητα που συνδέεται με το 1.842 (το + και το - είναι το ίδιο) βρίσκεται μεταξύ των τιμών 5% και 2.5%, και μάλλον πιο κοντά στο 5%. Με μία χοντρική γραμμική παλινδρόμηση μεταξύ των τιμών αυτών βλέπουμε ότι για το 1.842 έχουμε πιθανότητα περίπου 0.04 ή 4%. Αυτό σημαίνει ότι η πιθανότητα να εμφανιστεί η τιμή 7.51 σε ένα πληθυσμό τιμών με γνωστή μέση τιμή ίση με 8.0 και με n και s όπως δίνονται παραπάνω, είναι περίπου ίση με 4%. Αυτή κρίνεται μικρή πιθανότητα (το μικρό ή μεγάλο τελικά είναι υποκειμενικό και πρέπει να καθορίζεται από την αρχή). Άρα μάλλον υπάρχει πρόβλημα με το εργαστήριο.

ΠΡΟΣΟΧΗ: Η έννοια «πιθανότητα να εμφανιστεί η τιμή 7.51, σημαίνει ουσιαστικά την πιθανότητα να εμφανιστούν τιμές 7.51 και μικρότερες, που είναι το εμβαδόν κάτω από την καμπύλη t και αριστερά του 7.51. Παρόλα αυτά, λέμε «πιθανότητα να εμφανιστεί το 7.51» διότι αυτή είναι η μέγιστη από τις τιμές που έχουν μέγιστη πιθανότητα 4% να εμφανιστούν.

Το σχήμα 4.6. δείχνει την επιφάνεια αριστερά της κανονικής καμπύλης αναφοράς του \bar{y} (με μέση τιμή τα 8.0 mg/L) και της κανονικής καμπύλης αναφοράς της αντίστοιχης τιμής t . Γενικά, θεωρούμε ότι επειδή η επιφάνεια είναι μικρότερη του 5% (δηλ. για επίπεδο σημαντικότητας 95%), η τιμή 7.51 δεν θεωρείται ότι ανήκει στον πληθυσμό. Αν όμως, αλλάζαμε το επίπεδο σημαντικότητάς μας (ή τελικά αποδοχής μας) σε 99%, τότε η τιμή θα ήταν αποδεκτή ως ανήκουσα στον πληθυσμό.



Σχήμα 4.6. Οι κατανομές \bar{y} και t (με γνωστό μ) για τα 27 δεδομένα των νιτρικών. Στο σχήμα (b), θα ήταν ορθότερο αν λέγαμε $t \leq -1.853$.

Ουσιαστικά η t είναι μία κατανομή αναφοράς με το 0 ως την μέση τιμή του πληθυσμού και s/\sqrt{n} ως την τυπική απόκλιση.

ΠΡΟΣΟΧΗ: Για να τυποποιήσω οποιοσδήποτε σειρές μετρήσεων σε κατανομή (0,1), τότε αφαιρώ από κάθε μέτρηση τη μέση τιμή και διαιρώ με το τυπικό σφάλμα της μέσης τιμής (κατανομή t) ή την τυπική απόκλιση.

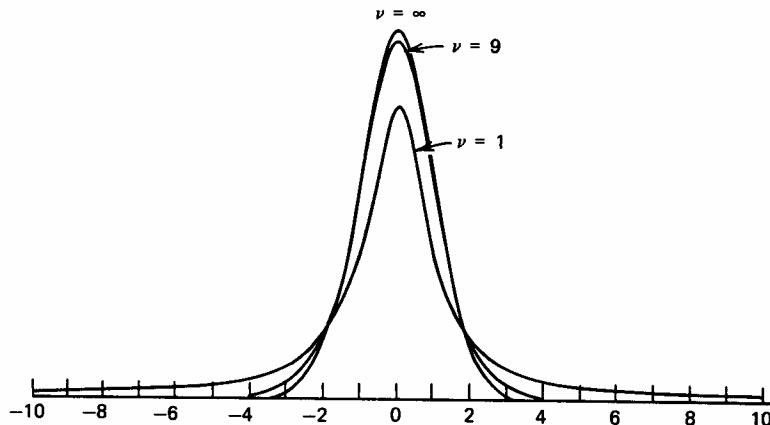
Συμπέρασμα των παραπάνω:

1. Η κανονική κατανομή είναι μία θεωρητική κατανομή που αφορά σε πληθυσμούς με πραγματική τιμή μ και τυπική απόκλιση σ .
2. Πρακτικά, για n δείγματα ενός πληθυσμού με τυπική απόκλιση s (άγνωστο σ) και γνωστή μέση τιμή μ χρησιμοποιούμε την κατανομή t και τη συνάρτηση t , με $v=n-1$ βαθμούς ελευθερίας. Η συνάρτηση αντιπροσωπεύει την κανονική κατανομή όταν έχουμε δείγματα, δηλαδή πρακτικώς σχεδόν πάντα.

Η σχέση υπολογισμού του t (t -statistic) θα δούμε παρακάτω χρησιμοποιείται για τη σύγκριση των μέσω τιμών διαφορετικών τεχνικών, ώστε να δούμε αν η διαφορά τους (ή η μη διαφορά τους) είναι στατιστικά σημαντική.

Δοκιμή: Μπείτε στην ιστοσελίδα:

<http://www-stat.stanford.edu/~naras/jsm/TDensity/TDensity.html> για να δείτε την μεταβολή της t ανάλογα με τους βαθμούς ελευθερίας. Όσο οι βαθμοί ελευθερίας αυξάνονται, τότε η κατανομή t προσεγγίζει την κανονική κατανομή, όπως φαίνεται και στο σχήμα στη συνέχεια.



Σχήμα 4.7. Μεταβολή του σχήματος της καμπύλης κατανομής t συναρτήσει των βαθμών ελευθερίας ν

4.3 Κατανομές χ^2 και F

Η κατανομή χ^2 είναι η κατανομή των διασπορών τυχαίων δειγμάτων που προήλθαν από κανονικούς πληθυσμούς. Επειδή η τιμή s^2 δεν μπορεί να είναι αρνητική, η κατανομή χ^2 δεν ομοιάζει με την κανονική κατανομή και ξεκινάει από την τιμή 0.

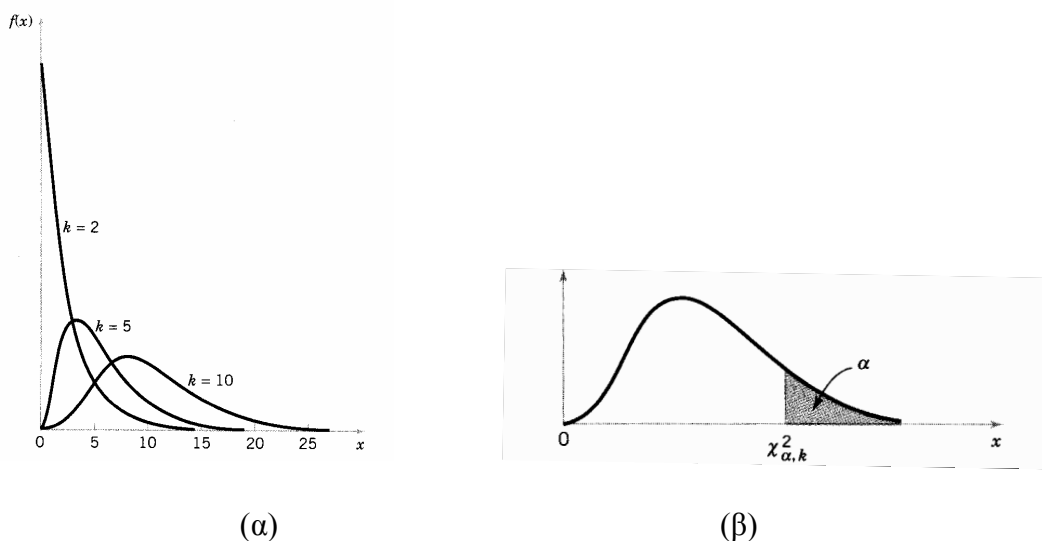
Αυτή η κατανομή ταιριάζει περισσότερο στα περιβαλλοντικά συστήματα. Δεν θα αναλυθεί ιδιαίτερα, διότι θεωρούμε πάντα ότι τα αρχικά δεδομένα μας κατανέμονται κανονικά. Η κατανομή χ^2 έχει μια κεντρική τιμή προς τα αριστερά και στη συνέχεια μια σχετικά μακριά «ουρά» προς τα δεξιά και τιμές πάντα θετικές. Η κατανομή χ^2 είναι ουσιαστικά η κατανομή του s^2 και είναι ανεξάρτητη του y . Η μορφή της κατανομής αυτής εξαρτάται από τους βαθμούς ελευθερίας του s . Δηλαδή, για ένα τυχαίο δείγμα μεγέθους n με διασπορά s^2 που προέρχεται από ένα πληθυσμό με διασπορά σ^2 , τότε η τιμή:

$$\chi^2 = \nu \frac{s^2}{\sigma^2}$$

, με ν τους βαθμούς ελευθερίας του s , ακολουθεί την κατανομή χ^2 .

Το άπλωμα της κατανομής αυτής εξαρτάται από τους βαθμούς ελευθερίας, ενώ το εμβαδόν στην άκρη της καμπύλης αυτής αποτελεί την πιθανότητα η τιμή χ^2 – όπως αυτή ορίστηκε παραπάνω - να ξεπερνάει μία συγκεκριμένη τιμή.

Τα γραφήματα στη συνέχεια δείχνουν την μορφή της χ^2 για διάφορους βαθμούς ελευθερίας, ενώ το σχήμα 4.8β δείχνει την επιφάνεια α , που ταυτίζεται με την πιθανότητα, δεξιά της τιμής $\chi^2_{\alpha,k}$ (ΠΡΟΣΟΧΗ: στο σχήμα 4.8 οι βαθμοί ελευθερίας συμβολίζονται με k αντί για ν). Όπως ισχύει και με τις παραπάνω κατανομές (κανονική, t), η καμπύλη κατανομής χ^2 χαρακτηρίζεται από τις τιμές α , $n-1$ ¹²



Σχήμα 4.8. (α) Μορφές της χ^2 ανάλογα τους βαθμούς ελευθερίας k . (β) Επιφάνεια α , που ταυτίζεται με την πιθανότητα, για γνωστή τιμή χ^2 , συναρτήσει των βαθμών ελευθερίας (k).

Η κατανομή χ^2 χρησιμοποιείται ευρέως και ομοιάζει αρκετά με την κατανομή F .

Η κατανομή F είναι η κατανομή του λόγου των διασπορών s_1^2 & s_2^2 δύο ανεξάρτητων τυχαίων δειγμάτων. Χρησιμοποιείται λοιπόν στην πράξη για να μελετήσουμε κατά πόσο δύο δείγματα προέρχονται από 2 πληθυσμούς που έχουν ίδιες διασπορές. Αν είναι όμοιες οι διασπορές των 2 δειγμάτων, τότε ο λόγος τους (F) πρέπει να είναι κοντά στο 1. Ο λόγος λοιπόν $F = s_1^2/s_2^2$ είναι μια τυχαία μεταβλητή που κατανέμεται με κατανομή F που εξαρτάται από τις παραμέτρους $\nu_1=n_1-1$ και $\nu_2=n_2-1$.

Σε διαφοροποίηση με την χ^2 , η κατανομή F εξαρτάται από 2 βαθμούς ελευθερίας ν_1, ν_2 και φυσικά από την επιθυμητό επίπεδο εμπιστοσύνης α . Στο παράρτημα περιέχονται τιμές της F για διάφορους συνδυασμούς ν_1, ν_2 για τιμές $\alpha = 0.05$ και $\alpha = 0.01$ (που έτσι και

¹² Οι βαθμοί ελευθερίας συμβολίζονται είτε με ν , είτε με k και σε κάθε περίπτωση ισούνται με $n-1$, όπου n ο αριθμός των μετρήσεων στο δείγμα. Στο παρόν σύγγραμμα θα χρησιμοποιείται τις περισσότερες φορές ο συμβολισμός ν , ενώ θα γίνεται κατάλληλη διευκρίνιση αν χρησιμοποιείται άλλος συμβολισμός.

αλλιώς χρησιμοποιούνται περισσότερο). Υπενθυμίζεται ότι στην καμπύλη F, η τιμή α αντιστοιχεί στο εμβαδόν στη δεξιά «ουρά» της καμπύλης.

Παράδειγμα κατανομής χ^2

Ένα μηχάνημα αυτόματου γεμίσματος χρησιμοποιείται για να γεμίσει μπουκάλια με ένα υγρό απορρυπαντικό. Ένα τυχαίο δείγμα 20 μπουκαλιών δίνει μία διασπορά του δείγματος ίση με $s^2 = 0.0153 \text{ lt}^2$. Αν η διασπορά του όγκου γεμίσματος υπερβαίνει τα 0.01 lt^2 , ένα μη αποδεκτό ποσοστό μπουκαλιών θα είναι λιγότερο ή περισσότερο γεμάτο του κανονικού. Υπάρχουν ενδείξεις από τα δεδομένα του δείγματος που να υποδεικνύουν ότι ο παρασκευαστής έχει πρόβλημα με υπο-γεμισμένα ή υπερ-γεμισμένα μπουκάλια? Κάνετε χρήση του $\alpha=0.05$ και θεωρήστε ότι οι όγκοι γεμίσματος κατανέμονται κανονικά.

Καταρχάς, η πληροφορία ότι οι όγκοι γεμίσματος κατανέμονται κανονικά είναι σημαντική γιατί μας βοηθάει να συμπεράνουμε ότι η διασπορά s^2 κατανέμεται ακολουθώντας την χ^2 κατανομή (Υπενθυμίζεται ότι η χ^2 είναι η κατανομή της διασποράς και όχι των αρχικών μετρήσεων).

Συνεπώς θέλουμε να δούμε αν ισχύει η μηδενική υπόθεση $H_0: \sigma^2=0.01$ ή η εναλλακτική υπόθεση $H_e: \sigma^2>0.01$ σε $\alpha=0.05$ ¹³.

Η τιμή χ^2 είναι: $\chi^2=(n-1) s^2 / \sigma^2$. Για $s^2 = 0.0153$, $\sigma^2=0.01$, και $n-1=19$, έχω $\chi^2=29.17$. Θέλω να συγκρίνω με την τιμή $\chi^2_{0.05,19}$, που βάσει πινακίων, είναι: 30.14. Άρα το $\chi^2_{0.05,19}$ βρίσκεται δεξιάτερα της τιμής 29.17, που σημαίνει ότι η πιθανότητα (εμβαδόν) που αντιστοιχεί στην δεύτερη περίπτωση είναι μεγαλύτερη από 5%.

Συνεπώς, δεν υπάρχουν σημαντικές ενδείξεις (το σημαντικό βέβαια εξαρτάται από το επίπεδο σημαντικότητας 95%, **που εμείς καθορίσαμε**) ότι η διασπορά των όγκων γεμίσματος θα υπερβαίνει το 0.01 lt^2 . Μάλιστα η πιθανότητα που αντιστοιχεί στο $\chi^2=29.17$ είναι (μπορεί να βρεθεί με παλινδρόμηση από σχετικό πινάκιο του παραρτήματος) 6.5% ήτοι μεγαλύτερη του 5%.

Παράδειγμα κατανομής F

Αν 2 ανεξάρτητα τυχαία δείγματα μεγέθους $n_1=7$ & $n_2=13$ λαμβάνονται από τον ίδιο κανονικό πληθυσμό, ποια είναι η πιθανότητα ότι η διασπορά του πρώτου δείγματος θα είναι τουλάχιστον 3 φορές μεγαλύτερη της διασποράς του δεύτερου δείγματος?

Από τον πίνακα της κατανομής F βρίσκουμε ότι $F_{6,12,0.05}$ είναι 3.00. Άρα η πιθανότητα να είναι ο λόγος s_1^2/s_2^2 ίσος με 3 είναι 5%.

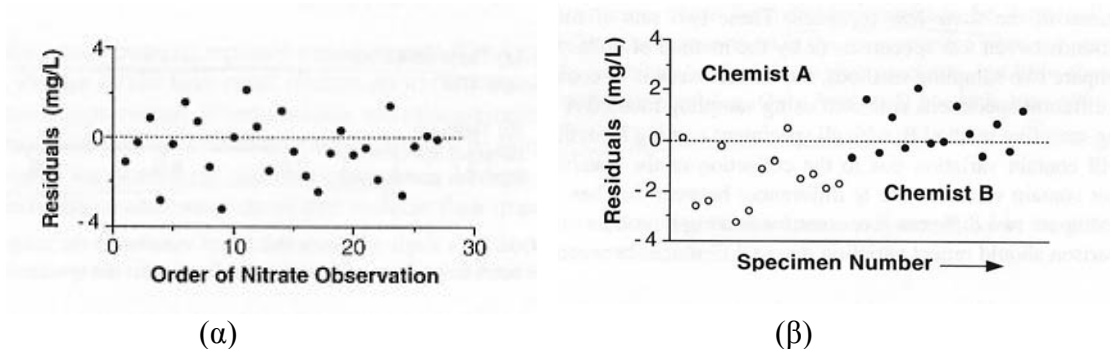
¹³ Η έννοια της μηδενικής και εναλλακτικής υπόθεσης θα εξηγηθεί σε επόμενο κεφάλαιο.

5. Αναγκαιότητα των γραφικών στην ανάλυση δεδομένων

Σύμφωνα με τον Tukey (1977), «Η μεγαλύτερη αξία των γραφημάτων είναι όταν μας αναγκάζουν να προσέξουμε κάτι που δεν αναμέναμε ποτέ να δούμε».

Πέρα από οποιονδήποτε υπολογισμό με τη χρήση μαθηματικών τύπων, είναι άκρως σημαντικό να κάνουμε γραφήματα των δεδομένων. Τα απλά γραφήματα, συνήθως γραφήματα χρονοσειρών στο χώρο της περιβαλλοντικής μηχανικής, μπορούν να αποκαλύψουν πολλά στοιχεία και τάσεις, που δεν είναι εύκολο να εξαχθούν άμεσα μέσω των στατιστικών τύπων ή μέσω πινάκων. Παραδειγματικά, στο σχήμα 5.1α στη συνέχεια φαίνονται τα πειραματικά σφάλματα από τις μετρήσεις των νιτρικών (27 τιμές με γνωστή τιμή 8.0 mg/L). Δηλαδή έχουμε κάνει γράφημα όλων των 27 διαφορών $y_i - \mu$, που αποτελούν τα πειραματικά σφάλματα. Τα (α) γράφημα που έγινε συναρτήσεως της σειράς λήψης των δειγμάτων δεν φανερώνει κάποια συγκεκριμένη τάση στα σφάλματα αυτά (ενώ φαίνεται ότι η μέση τιμή των σφαλμάτων είναι πολύ κοντά στο 0, αν και παρατηρείται ίσως ένας ελάχιστα μεγαλύτερος αριθμός αρνητικών σφαλμάτων). Αντίθετα αν τα δεδομένα αναπτυχθούν με τη σειρά του αριθμού του δείγματος (όπου ο αριθμός του δείγματος αντιστοιχεί σε διαφορετικό χημικό που έκανε τις μετρήσεις), όπως φαίνεται στο σχήμα 5.1β, τότε είναι ευκρινείς οι διαφοροποιήσεις μεταξύ των 2 χημικών. Συγκεκριμένα τα δείγματα με αριθμό 1-12 έγιναν από τον χημικό Α και τα υπόλοιπα από τον χημικό Β, κάτι το οποίο είναι εμφανές στο σχήμα 5.2β.

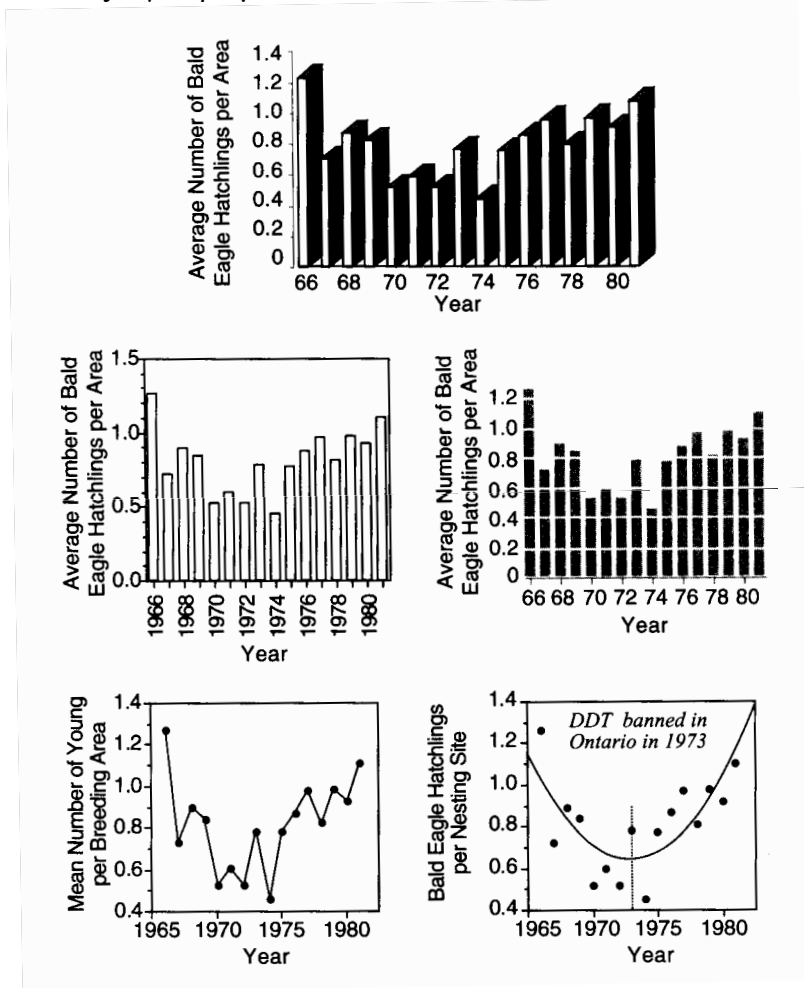
Γραφήματα σαν και αυτά, - δηλ. πειραματικά σφάλματα συναρτήσεως της σειράς της παρατήρησης - αποτελούν πρώτιστα διαγνωστικά γραφικά τεστ, ειδικά όταν ελέγχουμε την καταλληλότητα ενός εμπειρικού μοντέλου σε σχέση με τα πρωτογενή δεδομένα.



Σχήμα 5.1. Διαφορές ανάλογα με σειρά λήψης μετρήσεων και χημικό αναλυτή.

Σύμφωνα με τον Tufte, το άρτιο γραφικό σχήμα είναι αυτό που καταναλώνει το λιγότερο μελάνι στην εκτύπωση παρέχοντας την βασική πληροφορία. Με βάση το παραπάνω, που κατά τον γράφοντα είναι κανόνας, κάθε είδους γραφήματα με τρισδιάστατες μπάρες, σκιασμένες πίτες, κ.λ.π. είναι περιττά και ανούσια. Σαν παράδειγμα έχουμε τη σειρά των γραφημάτων στη συνέχεια. Το πρώτο είναι φανερά περιττό. Δεν απαιτούνται τρισδιάστατα σχέδια αφού δεν δίνουν κάποια επιπλέον πληροφορία με τις 3 διαστάσεις.

Το ίδιο ισχύει με το δεύτερο σχήμα. Το τρίτο δίνει την ίδια πληροφορία με το δεύτερο καταναλώνοντας λιγότερο μελάνι.



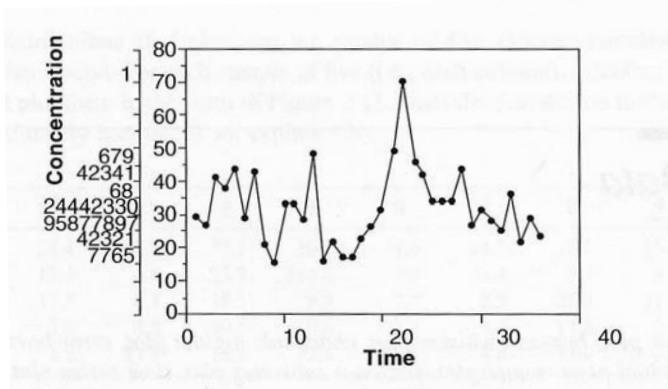
Σχήμα 5.2. Παράδειγμα ουσιαστικών και μη ουσιαστικών γραφημάτων (από Berthouex and Brown, 2002)

Το 4^ο σχήμα 5.2 δείχνει ότι κάποια μείωση παρατηρείται μεταξύ 1970 – 1975, αλλά οι γραμμές μεταξύ των σημείων μάλλον μπερδεύουν τον αναγνώστη. Σε αντίθεση, το 5^ο σχήμα είναι το ίδιο με το 4^ο, χωρίς ενωμένα σημεία (τα οποία δεν προσφέρουν παραπάνω πληροφορία, παρά μόνο παραπάνω μελάνι στο χαρτί) ενώ η απλή κατακόρυφη γραμμή στο 1973 δηλώνει ότι ήταν το έτος απαγόρευσης του DDT. Με τη πληροφόρηση αυτή, το σχήμα 5.2 (5^ο) γίνεται σαφές διατηρώντας τη λιτότητά του (δεν υπάρχουν γραμμές μεταξύ των σημείων).

Επίσης, υπάρχουν περιπτώσεις που όντως η χρήση γραφημάτων δεν έχει νόημα και είναι καλύτερη η χρήση ενός απλού πίνακα. Παράδειγμα, αν θέλουμε να αναφέρουμε τις τιμές 5 παραμέτρων των λυμάτων, τότε απλά μπορούμε να γράψουμε: pH = 5 mg/L, COD = 2300 mg/L, BOD = 1500 mg/L, TSS = 875 mg/L, TDS = 5700 mg/L και να μην κάνουμε απολύτως κανένα γράφημα (η χρήση τέτοιων γραφημάτων αποτελεί σύνηθες

περιεχόμενο μελετών, που το κριτήριο επιτυχίας τους – και πληρωμής τους - είναι το «πάχος» των σελίδων, παρά αυτά που περιέχονται).

Μερικά χρήσιμα γραφήματα φαίνονται στη συνέχεια. Το σχήμα 5.3 δείχνει, ταυτόχρονα, την κατανομή των δεδομένων, τη χρονοσειρά των δεδομένων και ένα πλήρες αριθμητικό αρχείο για μετέπειτα αριθμητική επεξεργασία 36 ωριαίων μετρήσεων (τιμές BOD σε mg/L), ήτοι των: 30, 27,41, 38, 44, 29, 43, 21, 15, 33, 33, 28, 49, 16, 22, 17, 17, 23, 27, 32, 47, 71, 46, 42, 34, 34, 34, 44, 27, 32, 28, 25, 36, 22, 29, 24.



Σχήμα 5.3. Χρησιμότερο γράφημα που παρέχει πολλαπλή πληροφορία (κατανομή, χρονοσειρά, τιμές). Το αριστερό «γράφημα» ονομάζεται «βλαστός και φύλλα» γράφημα¹⁴. Παραδειγματικά, μεταξύ του διαστήματος 40 και 45, φαίνεται ότι υπάρχουν 5 τιμές, που είναι οι 44,42,43,44,41 (με τη χρονική σειρά που εμφανίζονται). Οι

περισσότερες τιμές εμφανίζονται στο διάστημα 25-35, ενώ μόνο μία ακραία τιμή (71) εμφανίζεται. Η ακραία αυτή τιμή (outlier) δεν είναι απαραίτητα κάποια λάθος μέτρηση. Μπορεί να κρύβει κάποια σημαντική πληροφορία και χρήζει έρευνας.

Σημειώνεται ότι με βάση την αρχή του Tufte, στόχος είναι να κατασκευάζονται γραφήματα που να παρέχουν όσο το δυνατό περισσότερη πληροφορία εντός του ίδιου του γραφήματος, χωρίς να «ξοδεύεται» περιττό μελάνι.

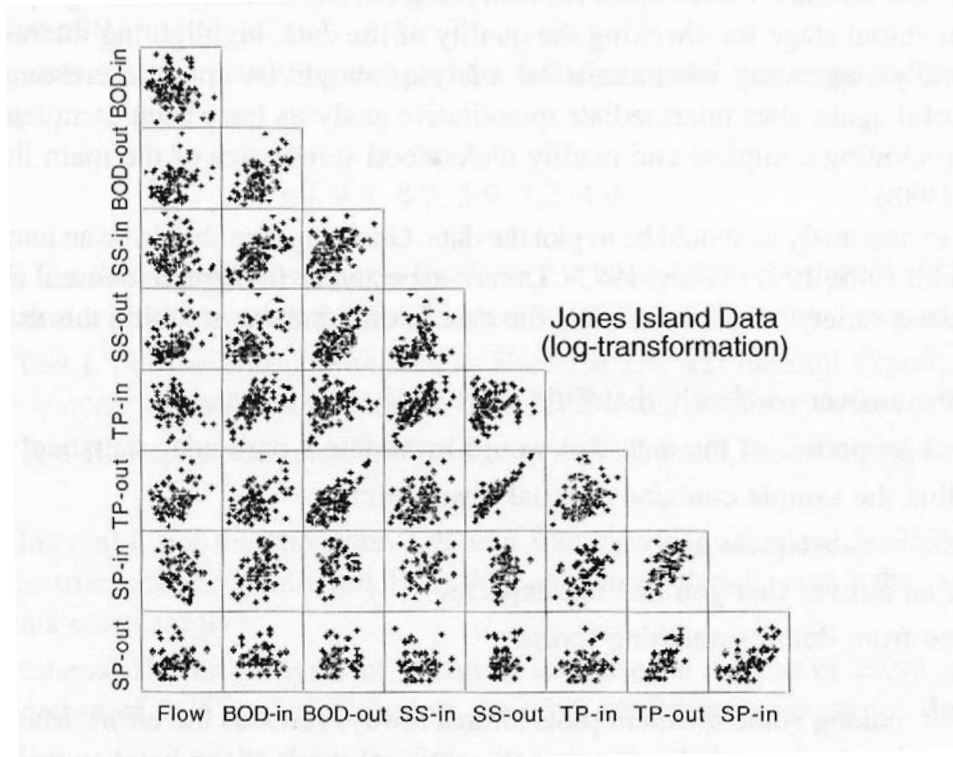
5.1 Γραφήματα διασποράς (scatter-plots)

Το σχήμα 5.4 αποτελεί ένα από τα πλέον κλασικά σχήματα που πρέπει να χρησιμοποιούν οι μηχανικοί περιβάλλοντος (και γενικότερα οι ερευνητές). Είναι πολλαπλά γραφήματα διασποράς (scatter-plots) που σαν στόχο έχουν να δείξουν (ή να μη δείξουν) την πιθανή συσχέτιση μεταξύ δύο οποιονδήποτε μεταβλητών, που έχουν συλλεχθεί ταυτόχρονα. Παράδειγμα, το σχήμα παρουσιάζει τις σχέσεις (όλους τους δυνατούς συνδυασμούς) μεταξύ ροής λυμάτων σε μία ΕΕΛ¹⁵, BOD εισόδου, BOD εξόδου, συνολικών στερεών εισόδου, συνολικών στερεών εξόδου, συνολικού φωσφόρου εισόδου, συνολικού φωσφόρου εξόδου και διαλυτού φωσφόρου εισόδου. Τα γραφήματα αυτά είναι εξαιρετικά στην παρατήρηση πιθανών συσχετίσεων και δεν έχουν διαβαθμίσεις στους άξονες, αφού ο στόχος είναι η παρατήρηση τάσεων (οι κλίμακες στους οριζόντιους και κατακόρυφους άξονες διαφέρουν σε κάθε γράφημα). Όπως δείχνει το σχήμα, ο συνολικός φώσφορος εξόδου συσχετίζεται αρκετά καλά (θετική συσχέτιση)

¹⁴ “stem & leaf”, στατιστικά πακέτα, όπως το STATISTICA, SYSTAT, SPSS και MINITAB μπορούν να παράγουν τέτοια γραφήματα, ενώ είναι εξαιρετα γενικότερα για στατιστικά γραφήματα και σχετικές αναλύσεις.

¹⁵ Εγκατάσταση Επεξεργασίας Λυμάτων

με τα συνολικά στερεά εξόδου και το BOD εξόδου και μετρώς με το BOD εισόδου. Είναι αναμενόμενες οι συσχετίσεις των BOD_{in} με SS_{in} και BOD_{out} με SS_{out} .



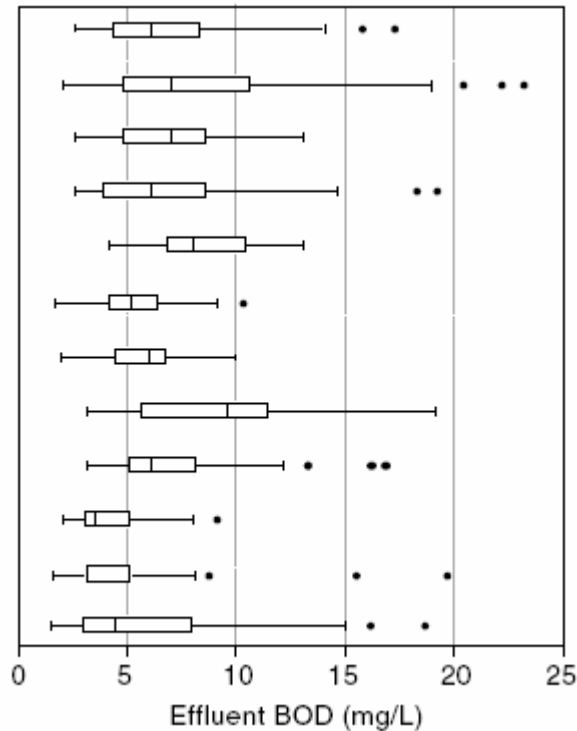
Σχήμα 5.4. Πολλαπλά διαγράμματα διασποράς (scatterplots). Οι άξονες είναι σε λογαριθμική κλίμακα

5.2 Box-whisker γράφημα

Το διάγραμμα κουτιού (διάγραμμα με μουστάκια!) ή αλλιώς box-whisker plot είναι αρκετά κατατοπιστικό αναφορικά με τις μέσες τιμές και τις ακραίες τιμές που εμφανίζονται σε ένα δείγμα. Δεν δείχνει παρόλα αυτά το μέγεθος του δείγματος, αν και αυτό είναι απλό να φανεί με μία σημείωση απλή πάνω στο γράφημα (π.χ. $n=27$). Για την πραγματοποίηση του γραφήματος αυτού τα δεδομένα των δειγμάτων πρέπει να ιεραρχηθούν από την μικρότερη τιμή προς την μεγαλύτερη. Παραδειγματικά έχουμε το παρακάτω γράφημα που δείχνει την ποιότητα της εξόδου από 12 διαφορετικές ΕΕΛ με χρήση των βιολογικών φίλτρων (trickling filters). Το γράφημα περιέχει:

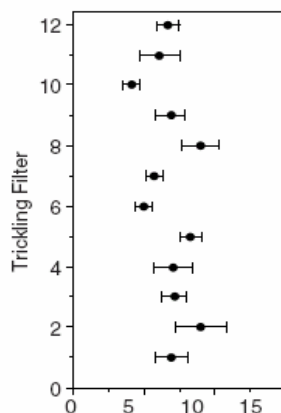
- Τον μέσο (median) (ΠΡΟΣΟΧΗ: δεν είναι ο μέσος όρος αλλά είναι απλά η τιμή που βρίσκεται στο μέσο των αριθμών, όταν αυτοί ιεραρχηθούν από τον μικρότερο προς τον μεγαλύτερο. Δηλαδή, ο μέσος (median) διαφέρει από τον μέσο όρο (average ή mean).
- Τα 25^α και 75^α εκατοστιαία, που είναι οι τιμές των δεδομένων, κάτω από την πρώτη, αντίστοιχα, περιέχεται το 25% των δεδομένων και άνω της δεύτερης περιέχεται το 75%.
- Οι ακραίες τιμές, που είναι αντίστοιχα η ελάχιστη και η μέγιστη τιμή των δεδομένων.

- Στο γράφημα μπορούν να αποτυπωθούν και οι πολύ ακραίες τιμές (outliers), ο ορισμός των οποίων δεν είναι πάντα σαφής (μεμονωμένα σημεία στο σχήμα). Μπορεί να είναι αποτέλεσμα λάθος μετρήσεων, αλλά μπορεί και να κρύβουν κάποια πολύ σημαντική πληροφορία και συνεπώς πρέπει να μελετώνται με προσοχή.



Σχήμα 5.5. Box-plot για 2 συγκεντρώσεις BOD σε εξόδους εργοστασίων επεξεργασίας λυμάτων με βιολογικά φίλτρα (Berthouex and Brown, 2002).

Το σχήμα 5.6 δείχνει τις τιμές των εν λόγω 12 ΕΕΛ με βιολογικά φίλτρα με μπάρες τυπικού σφάλματος της μέσης τιμής, συμμετρικά αριστερά και δεξιά της μέσης τιμής. Οι μπάρες είναι ίσες με το διπλάσιο του τυπικού σφάλματος της μέσης τιμής (standard error of the mean)¹⁶.



Σχήμα 5.6. Μέσος όρος ± 2 x τυπικό σφάλμα της μέσης τιμής (το σχήμα δείχνει ότι τα δεδομένα κατανέμονται κανονικά κάτι το οποίο δεν ισχύει τελικά).

¹⁶ Υπενθυμίζεται ότι το τυπικό σφάλμα είναι μία εκτίμηση της τυπικής απόκλισης της μέσης τιμής και ισούται με $S_{\bar{y}} = s/\sqrt{n}$

Γενικά το σχήμα 5.6 μπορεί να οδηγήσει τον αναγνώστη να πιστέψει ότι τα δεδομένα κατανέμονται συμμετρικά γύρω από τον μέσο όρο, κάτι που, όπως δείχνει το πρώτο σχήμα δεν είναι αληθές. Η χρήση της μέσης τιμής και της τυπικής απόκλισης, υποδεικνύει γενικά ότι τα δεδομένα κατανέμονται κανονικά (δηλ. με κανονική κατανομή). Αυτό δεν είναι πάντα αληθές για τα περιβαλλοντικά δεδομένα, που κυριαρχεί η χ^2 κατανομή συχνά. Το σχήμα 5.6 λοιπόν είναι καλό για να δείξει την τυπική απόκλιση πολλών επαναληπτικών τιμών μίας μέτρησης. Παρόλα αυτά, ΠΡΕΠΕΙ ΝΑ ΔΙΕΥΚΡΙΝΙΖΕΤΑΙ τι ακριβώς είναι η μπάρα γύρω από τον μέσο όρο, αφού αυτή μπορεί να είναι ίση με:

1. Μία τυπική απόκλιση του δείγματος (s)
2. Ένα (ή δύο) τυπικά σφάλματα της μέσης τιμής ($s_{\bar{y}}$ ή $2 \times s_{\bar{y}}$)
3. Το διάστημα εμπιστοσύνης της παραμέτρου (για ορισμένο επίπεδο εμπιστοσύνης, το οποίο πρέπει σαφώς και αυτό να καθορίζεται).

Είναι προφανές ότι τα παραπάνω γραφήματα απαιτούν έναν επαρκή αριθμό δεδομένων για να υλοποιηθούν. Συνήθως, δεν μπορεί να γίνουν με λιγότερες από δέκα (10) τιμές.

5.3 Διαγνωστικά τεστ των σφαλμάτων

Τα τρία κύρια χαρακτηριστικά των σφαλμάτων που συνοδεύουν κάποιες μετρήσεις είναι η κανονικότητα, η τυχαιότητα και η ανεξαρτησία.

- *Κανονικότητα* σημαίνει ότι τα πειραματικά σφάλματα που συνοδεύουν κάποιες μετρήσεις (δηλ. τα $\bar{y} - y_i$) κατανέμονται κανονικά.
- *Τυχαιότητα* σημαίνει ότι οι παρατηρήσεις προέρχονται από ένα πληθυσμό με κάθε παρατήρηση να έχει την ίδια πιθανότητα να «κληρωθεί» και να αποτελέσει μέρος του δείγματος.
- *Ανεξαρτησία* σημαίνει απλά ότι κάποια παρατήρηση δεν εξαρτάται από κάποια άλλη του ίδιου πληθυσμού. Δηλαδή η πιθανότητα του να συμβεί το Α με την πιθανότητα του να συμβεί το Β. Αν, π.χ σε μία σειρά παρατηρήσεων όταν το y_1 είναι μεγάλο τότε και η χρονικά κοντινή του τιμή y_2 είναι επίσης μεγάλη, τότε οι δύο μεταβλητές (παρατηρήσεις) δεν είναι στατιστικά ανεξάρτητες, δηλ. είναι χρονικά εξαρτημένες ή σειριακά συσχετισμένες.

Όταν φτιάχνουμε κάποιο μαθηματικό μοντέλο (εμπειρικό ή θεωρητικό), τελικώς θέλουμε να δούμε κατά πόσο αυτό το μοντέλο «ταιριάζει» στα πραγματικά δεδομένα. Ένα κύριο διαγνωστικό τεστ καταλληλότητας του μοντέλου είναι τα γραφήματα των σφαλμάτων. Τα σφάλματα, δηλαδή οι διαφορές των εκτιμήσεων του μοντέλου και των πραγματικών τιμών πρέπει:

- Να είναι ανεξάρτητα και να κατανέμονται κανονικά.
- Να έχουν μέση τιμή 0
- Να είναι τυχαία

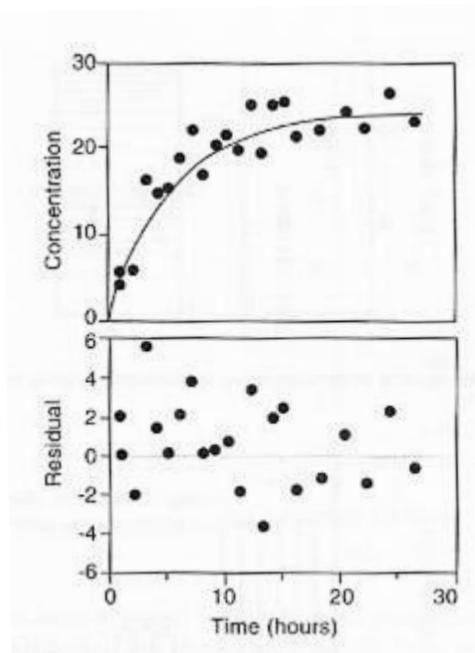
Τα παραπάνω είναι κατανοητά με το σχήμα 5.7. Το γράφημα στην κορυφή δείχνει το αποτελέσματα του μοντέλου και τις πραγματικές τιμές. Από το πρώτο γράφημα φαίνεται

οτι οι αποκλίσεις των υπολογισμένων τιμών είναι μάλλον μεγαλύτερες στις υψηλές τιμές (μεγαλύτερες ώρες). Στο δεύτερο γράφημα, που ακριβώς αποτελεί το διαγνωστικό γράφημα, δηλαδή το γράφημα των σφαλμάτων, παρατηρούμε ότι στις χαμηλές τιμές οι αποκλίσεις είναι τελικώς μεγαλύτερες. Γραφήματα σαν το δεύτερο πρέπει να γίνονται ώστε να εξετάζεται η ανεξαρτησία και τυχαιότητα των σφαλμάτων. Αν παρατηρούνται κάποιες τάσεις, τότε το μοντέλο δεν είναι επαρκές και πρέπει να δοκιμάζεται νέο.

Σε συνέχεια του παραπάνω, 4 επιπλέον παρεμφερή και σημαντικά διαγνωστικά τεστ σφαλμάτων είναι:

1. Κανονικότητα των σφαλμάτων, όπως μπορεί να φανεί με ένα απλό ιστόγραμμα (ή με γράφημα σε χαρτί πιθανοτήτων) των σφαλμάτων πάντα.
2. Γράφημα με τα σφάλματα σε σχέση με τον χρόνο λήψης των δειγμάτων (αυτό πάντα δεν είναι διαθέσιμο διότι πρέπει να έχουν καταγραφεί οι χρονικές στιγμές δειγματοληψίας).
3. Γράφημα συσχέτισης του σφάλματος με προηγούμενη τιμή (δηλαδή e_t vs e_{t+1}).
4. Γράφημα συσχέτισης του σφάλματος με την αντίστοιχη πραγματική τιμή παρατήρησης ή την τιμή που εξάγει το μοντέλο.

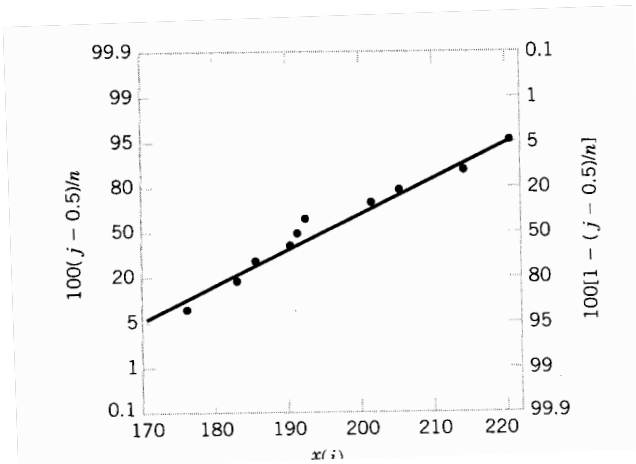
Καμία τάση δεν πρέπει να παρατηρείται στα γραφήματα των σημείων 2-4 αν το μοντέλο πρέπει να θεωρηθεί επαρκές. Σε περίπτωση που παρατηρείται τάση, τότε το μοντέλο πρέπει να ξαναδοκιμάζεται.



Σχήμα 5.7. Γράφημα σφαλμάτων (β γράφημα) ενός εμπειρικού μοντέλου που δοκιμάζεται για κάποια πρωτογενή δεδομένα, ώστε να ελεγχθεί η τυχαιότητά τους. Δεν πρέπει να παρουσιάζεται κάποια τάση και τα σημεία πρέπει να κατανέμονται τυχαία γύρω από το 0.

Ένας από τους πιο κοινά χρησιμοποιούμενους δείκτες κατά την μελέτη της καταλληλότητας ενός μοντέλου είναι ο συντελεστής απόφασης (ή συντελεστής παλινδρόμησης), R^2 , ο οποίος θα αναλυθεί στη συνέχεια. Ο συντελεστής αυτός, χρησιμοποιείται ευρέως ενώ έχει σαφώς παρεξηγηθεί, αφού μπορεί να καταλήξει, αν χρησιμοποιείται μόνος του χωρίς τα διαγνωστικά γραφήματα των σφαλμάτων, σε λάθος εκτιμήσεις.

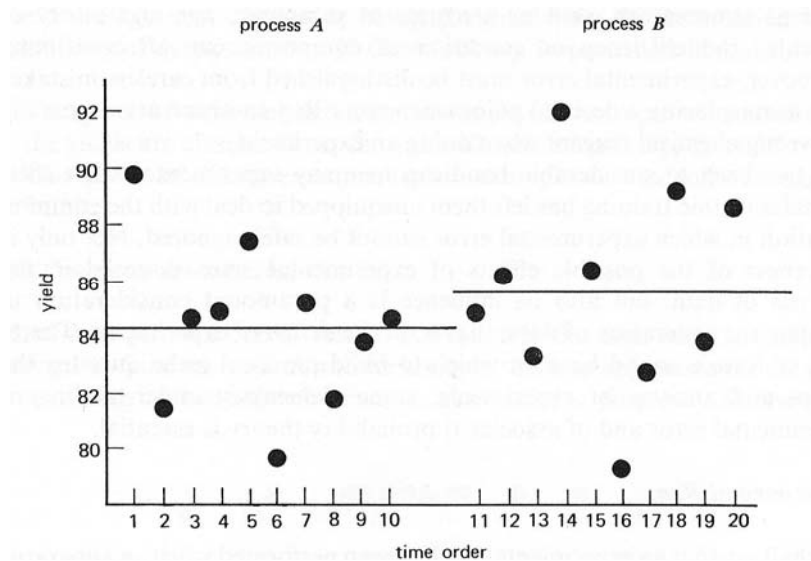
Η κανονικότητα των σφαλμάτων ελέγχεται με τα κλασσικά ιστογράμματα (όλα τα προγράμματα παρέχουν αυτή την δυνατότητα) αλλά η καλύτερη τεχνική είναι η χρήση γραφημάτων πιθανότητας. Τα γραφήματα αυτά έχουν στον κατακόρυφο άξονα μία χαρακτηριστική κλίμακα (κλίμακα πιθανότητας). Τα δεδομένα θέτονται στο γράφημα αφού πρώτα ιεραρχηθούν από το μικρότερο στο μεγαλύτερο. Στη συνέχεια, υπολογίζεται για κάθε μέτρηση, η τιμή $(j-0.5)/n$, όπου j η σειρά που έχει η μέτρηση στην ιεράρχηση και n είναι ο αριθμός των μετρήσεων. Στη συνέχεια σχεδιάζεται το γράφημα σε χαρτί πιθανοτήτων, όπου στον οριζόντιο άξονα βρίσκονται οι πρωτογενείς τιμές και στον κατακόρυφο άξονα, με ειδική κλίμακα πιθανότητας, οι τιμές $(j-0.5)/n$. Αν οι τιμές «πέφτουν» χοντρικά πάνω σε ευθεία γραμμή, τότε τα δεδομένα κατανέμονται κανονικά. Η μέση τιμή προφανώς είναι η τιμή που αντιστοιχεί στο 50 (δηλ. το 50%) του κατακόρυφου άξονα. Σημειώνεται ότι όλα τα στατιστικά προγράμματα παρέχουν τη δυνατότητα γραφημάτων πιθανότητας. Το EXCEL δεν έχει κλίμακα πιθανότητας.



Σχήμα 5.8. Γράφημα πιθανότητας για τον έλεγχο της κανονικότητας 10 μετρήσεων. Τα δεδομένα φαίνεται ότι κατανέμονται κανονικά σε αυτή την περίπτωση αφού μία ευθεία γραμμή περνάει σχετικά ομαλά ανάμεσά τους.

6. Τεστ σημαντικότητας και έννοια της υπόθεσης

Ένας από τους κύριους στόχους της διεξαγωγής πειραμάτων και στατιστικής ανάλυσης των αποτελεσμάτων είναι να μελετήσουμε και συμπεράνουμε για το αν υπάρχουν διαφορές μεταξύ 2 επεξεργασιών (treatments). Αν, παράδειγμα, κάνουμε τις μετρήσεις της μεθόδου A και B και αφού εξάγουμε τα βασικά στατιστικά αποτελέσματα (μέσος όρος, s) και κάνουμε τα σχετικά γραφήματα (δες σχήμα 6.1), στη συνέχεια πρέπει να αναρωτηθούμε αν όντως υπάρχουν διαφορές μεταξύ των τεχνικών. Βλέπουμε όντως διαφορετικούς μέσους όρους για τα δείγματα A & B - τα γραφήματα έγιναν συναρτήσει του χρόνου μέτρησης - αλλά υπάρχουν όντως διαφορές μεταξύ των δύο διαδικασιών? Πόσο στατιστικά σημαντικές είναι οι διαφορές αυτές συναρτήσει πάντα των μετρήσεων που κάναμε? Η απάντηση στο παραπάνω ερώτημα αποτελεί ένα από τα βασικά θέματα που αφορούν τους αναλυτές αποτελεσμάτων, τους ερευνητές και μηχανικούς.



Σχήμα 6.1. Αποτελέσματα από δύο διεργασίες A και B. Αν και έχουν διαφορετικούς μέσους όρους, είναι όντως διαφορετικές?

Το πως γίνεται η σύγκριση, δηλαδή το πως θα καταλήξουμε στο συμπέρασμα για το αν δύο δείγματα ανήκουν ή δεν ανήκουν στον ίδιο πληθυσμό, και συνεπώς για το αν είναι ίδιες οι διεργασίες ή διαφορετικές, βασίζονται στην επιβεβαίωση ή όχι κάποιας υπόθεσης. Η έννοια της υπόθεσης ταυτίζεται με την έννοια της δοκιμής σημαντικότητας.

Ένα από τα κύρια ερωτήματα λοιπόν που τίθενται είναι το κατά πόσο μία μέση τιμή από ένα δείγμα είναι όντως ίση με την πραγματική μέση τιμή του πληθυσμού ή κατά πόσο οι 2 διεργασίες που αναφέρονται παραπάνω είναι όμοιες ή όχι. Αυτό το ερώτημα δεν απαντάται με ένα απλό ναι ή όχι αλλά με μία πιθανότητα ναι ή μία πιθανότητα όχι. Σαν να λέμε ότι αύριο η πιθανότητα βροχής στο τάδε σημείο είναι 60%. Αυτό σημαίνει ότι

μπορεί να βρέξει αλλά και μπορεί να μην βρέξει αύριο, αλλά αν είχαμε 100.000 παρόμοιες καταστάσεις, στις 60.000 θα είχαμε βροχή.

Στόχος δηλαδή είναι να αποφανθούμε εάν:

- Το δείγμα A και B προέρχονται από τον ίδιο πληθυσμό (μηδενική υπόθεση H_0)
- Το δείγμα A και B δεν προέρχονται από τον ίδιο πληθυσμό (εναλλακτική υπόθεση H_e)

Όπως βλέπετε, χρησιμοποιούμε την έννοια της υπόθεσης – δηλαδή ότι υποθέτουμε ότι κάτι συμβαίνει ή δεν συμβαίνει – και θέλουμε να βρούμε ποιές από τις 2 υποθέσεις ισχύει. Η έννοια της υπόθεσης είναι σημαντική στη στατιστική ανάλυση και, ως έννοια μάλλον πρωτοπόρα. Πρακτικά τελικώς, ψάχνουμε για την πιθανότητα για το αν κάτι ισχύει ή όχι, δηλαδή στόχος μας είναι να υπολογίσουμε μία τιμή (x %) που να ταυτίζεται με την πιθανότητα να συμβεί η μηδενική υπόθεση H_0 , και αντίστροφα η πιθανότητα να συμβεί η εναλλακτική H_e που θα είναι $1-x\%$.

Δηλαδή, πρακτικά κάνουμε τις δύο υποθέσεις:

H_0 : Διεργασία A = Διεργασία B

H_e : Διεργασία A \neq Διεργασία B

Η ισχύς ή όχι των υποθέσεων ελέγχεται με το t τεστ ή με την έννοια του διαστήματος εμπιστοσύνης. Και οι δύο προσεγγίσεις είναι στατιστικά ισοδύναμες, όμως η δεύτερη έννοια είναι πιο κατανοητή διότι απλά δίνει το διάστημα εντός του οποίου αναμένεται (με πιθανότητα που εμείς ορίζουμε) να υφίσταται η παράμετρος του πληθυσμού υπό μελέτη. Και τα δύο παρέχουν την ίδια πληροφορία, απλά – όπως θα δούμε – το δεύτερο (κατά το γράφοντα) συνίσταται λόγω πιά εύκολων υπολογισμών.

6.1 Τεστ σημαντικότητας με δοκιμή t

Με βάση το παράδειγμα μας έως τώρα, «Πόσο πιθανό είναι να έχουμε μία μέση τιμή δείγματος ίση με 7.51 mg/L από ένα πληθυσμό με γνωστή μέση τιμή πληθυσμού ίση με 8.0 mg/L». Ανάλογα με την απάντησή μας, μπορούμε να δεχθούμε ότι το δείγμα όντως αντιπροσωπεύει τον πληθυσμό ή όχι.

Αντιστρόφως το ερώτημα που μπορεί να τεθεί είναι – σε περίπτωση μη γνώσης της πραγματικής μέσης τιμής του πληθυσμού – κάτι σύνηθες – «Αν η μέση τιμή του δείγματος μας με 27 παρατηρήσεις είναι 7.51 mg/L, τότε ποία η πιθανότητα η πραγματική μέση τιμή του πληθυσμού να είναι ίση με 8.0 mg/L». Για να απαντηθεί το παραπάνω ερώτημα, πραγματοποιούμε ένα τεστ σημαντικότητας ορίζοντας τις παρακάτω υποθέσεις:

- τη μηδενική υπόθεση, δηλ. H_0 ότι $\mu = 8.0$ και
- την εναλλακτική υπόθεση, δηλ. H_e ότι $\mu < 8.0$. Προσοχή η υπόθεση $\mu < 8.0$ διαφέρει από την υπόθεση $\mu \neq 8.0$. Στη δεύτερη περίπτωση, ουσιαστικά ελέγχουμε τις υποθέσεις $\mu < 8.0$ και $\mu > 8.0$ ταυτόχρονα.

Κάνουμε χρήση του t-τεστ και ορίζουμε ένα ποσοστό α , που είναι η πιθανότητα να απορρίψουμε την μηδενική υπόθεση κατά λάθος. Π.χ. $\alpha=0.05$ σημαίνει 5% ή αντίστροφα ($1-\alpha=95\%$) είναι η πιθανότητα να ισχύει η υπόθεση μηδέν.

Η σχετική συνάρτηση t για τον υπολογισμό του τεστ σημαντικότητας είναι η γνωστή:

$$t = \frac{\bar{y} - \mu}{s_y}$$

ΠΡΟΣΟΧΗ: μιλάμε για έλεγχο μέσων τιμών τελικά, οπότε στον παρονομαστή εισέρχεται η τυπική απόκλιση της μέσης τιμής (ή αλλιώς *το τυπικό σφάλμα της μέσης τιμής*, όπως σωστότερα ονομάζεται). Δεν χρησιμοποιείται δηλαδή η τυπική απόκλιση των παρατηρήσεων, s . Επίσης υπενθυμίζεται ότι η κατανομή t προϋποθέτει ότι τα αρχικά δεδομένα κατανέμονται κανονικά, είναι ανεξάρτητα μεταξύ τους, ενώ η διασπορά s^2 κατανέμεται ανεξάρτητα από την κατανομή της μέσης τιμής.

Μετά τον υπολογισμό του t κάνουμε χρήση των πινακίων που έχουν καταχωρημένα τις τιμές t (δες παράρτημα), πάντα συναρτήσει του επιθυμητού α (που αποτελεί το εμβαδό κάτω από την καμπύλη και συγκεκριμένα κάτω από την «ουρά» της καμπύλης στα δεξιά ή αριστερά) και των βαθμών ελευθερίας, αφού είπαμε ότι το σχήμα της καμπύλης t εξαρτάται από το πλήθος του δείγματος¹⁷. Άρα συγκρίνουμε την υπολογιζόμενη τιμή t με την τιμή t των πινάκων συναρτήσει ενός α και ενός ν , που συμβολίζεται ως $t_{\nu,\alpha}$.

Παράδειγμα

Για το δείγμα μας με μέση τιμή $\bar{y}=7.51$, θέλουμε να δούμε αν ισχύει η υπόθεση $H_0: \mu=8.0$ σε $\alpha=5\%$ (ή 0.05) με 27 δείγματα (δηλ. 26 β.ε.). Η εναλλακτική υπόθεση H_e είναι $\mu < 8.0$. Το t είναι: $t = (7.51 - 8.0) / 0.266 = -1.842$

(Το 0.266 έχει υπολογιστεί με τη σχέση που αναφέρεται παραπάνω, αφού έχουμε γνωστή και την τυπική απόκλιση s του δείγματος. Σημειώνεται επίσης ότι το πρόσημο (-) της τιμής 1.842 δεν μας αφορά άμεσα. Μπορούμε να δουλέψουμε με την απόλυτη τιμή. Το (-) απλά υποδεικνύει ότι η μέση τιμή του δείγματος είναι μικρότερη της μέσης τιμής του πληθυσμού και ότι βρισκόμαστε στα αριστερά της καμπύλης t)

Κοιτώντας το σχετικό πινάκιο, η τιμή $t_{26,0.05}$, δηλαδή για 26=27-1 βαθμούς ελευθερίας και σε ρίσκο 5% (δες πίνακα σχετικό) είναι -1.706. Η τιμή t που υπολογίσαμε είναι μικρότερη (βρίσκεται πιο αριστερά του -1.706), άρα δεν βρισκόμαστε στην επιφάνεια που αντιστοιχεί στο 95%, που σημαίνει ότι **δεν ισχύει** η μηδενική υπόθεση και **ισχύει** η εναλλακτική. Πρακτικά αυτό σημαίνει ότι από τις 100 δειγματοληψίες διαφόρων δειγμάτων από τον πληθυσμό με πραγματική μέση τιμή 8.0, οι 95 θα δίνουν μέση τιμή μεγαλύτερη του 7.51.

¹⁷ Υπενθυμίζεται ότι όσο μεγαλύτερο το πλήθος των δειγμάτων, τόσο περισσότερο η καμπύλη t προσεγγίζει την κανονική καμπύλη κατανομής.

Αν επιλέγαμε $\alpha = 99\%$, τότε το $t_{26,0.05}$ θα ήταν 2.479 (ή -2.479) και τώρα το t που υπολογίσαμε είναι μεγαλύτερο της παραπάνω τιμής. Σε αυτή την περίπτωση, η μηδενική υπόθεση ισχύει. Άρα η ισχύς ή όχι της μηδενικής ή εναλλακτικής υπόθεσης εξαρτάται από το ποσοστό α , που αλλιώς ορίζεται ως **επίπεδο εμπιστοσύνης ή σημαντικότητας**.

Όπως βλέπουμε λοιπόν, πολλά στηρίζονται στην επιλογή του α , που σημαίνει την πιθανότητα που εμείς επιθυμούμε, δηλ. το ρίσκο που εμείς δίνουμε στο να απορρίψουμε την υπόθεση. Σαφώς πρέπει λοιπόν να δηλώνουμε πάντα το α που χρησιμοποιούμε. Γενικά όσο μικρότερο το α , πιο «ασφαλή» τα αποτελέσματα. Συνηθίζονται τιμές όπως 90% ή το πολύ 95%.

Σημειώνεται ότι το παραπάνω παράδειγμα είναι τεστ σημαντικότητας t μίας πλευράς, αφού ελέγξαμε μόνο το $H_e: \mu < 8.0$. Είναι σύνηθες να κάνουμε εναλλακτικές υποθέσεις όπως $\mu \neq 8.0$, δηλαδή $\mu < 8.0$ ή $\mu > 8.0$ ταυτόχρονα, οπότε μιλάμε για τεστ και από τις δύο πλευρές. Η διαφορά εδώ είναι ότι κοιτάμε και την αριστερή και την δεξιά «ουρά» της καμπύλης t . Οι επιφάνειες κάτω από την «ουρά» συνολικά πρέπει να αθροίζονται, και αριστερά και δεξιά στις άκρες της καμπύλης, σε α συνολικά. Άρα, η επιφάνεια της κάθε «ουράς» είναι $\alpha/2$. Στην περίπτωση του $\alpha=0.05$, θα πρέπει λοιπόν να βρούμε την τιμή $t_{26,0.025}$, που προφανώς θα έχει ένα πρόσημο \pm , αφού δηλώνει ότι ελέγχουμε και στην αριστερή και δεξιά άκρη της καμπύλης.

Για το παράδειγμα και για έλεγχο των υποθέσεων:

$$H_0: \mu = 8.0$$

$$H_e: \mu \neq 8.0$$

Το $t_{26,0.025}$ είναι ίσο με ± 2.056 . Επειδή το t που υπολογίσαμε είναι ίσο με -1.842, είναι μεγαλύτερο του -2.056 και άρα δεν πέφτει εκτός των κρίσιμων τιμών t . Συνεπώς, δεν υπάρχουν ενδείξεις να απορρίψουμε την μηδενική υπόθεση στο συγκεκριμένο επίπεδο σημαντικότητας. Δηλαδή, ισχύει η μηδενική υπόθεση και το δείγμα όντως προέρχεται από πληθυσμό με πραγματική μέση τιμή ίση με $\mu=8.0$.

Παρατηρούμε ότι διαφορετικά αποτελέσματα έχουμε ανάλογα με το αν υπολογίζουμε t τεστ μίας πλευράς ή δύο πλευρών. Το ποιο τεστ θα χρησιμοποιηθεί εξαρτάται από την περίπτωση. Παράδειγμα, αν σε ένα πείραμα θέλουμε να δούμε η προσθήκη του χημικού Α αλλάζει την απόδοση ενός μείγματος (είτε για καλύτερο είτε για χειρότερα), τότε τεστ και από τις πλευρές πρέπει να χρησιμοποιηθεί.

6.2 Τεστ σημαντικότητας βάσει διαστήματος εμπιστοσύνης (confidence interval)

Πολλές φορές είναι πιο πρακτικό να ορίσουμε ένα διάστημα μέσα στο οποίο η τιμή κάποιου παραμέτρου αναμένεται (ή δεν αναμένεται) να υπάρχει. Το διάστημα εμπιστοσύνης ορίζεται (σε επίπεδο $1-\alpha$) και υπολογίζεται ως εξής:

$$\bar{y} - s_{\bar{y}} t_{\alpha/2} < \mu < \bar{y} + s_{\bar{y}} t_{\alpha/2}$$

Προσέξτε ότι το διάστημα εμπιστοσύνης προκύπτει από τη σχέση που ορίζει το t . Δηλαδή, πολλαπλασιάζω τον παρονομαστή με την τιμή t και αυτό μου δίνει τη διαφορά $y-\mu$. Η διαφορά αυτή «κυμαίνεται» αριστερά και δεξιά του μ , και έτσι ορίζεται το διάστημα εμπιστοσύνης.

Η έννοια του διαστήματος εμπιστοσύνης είναι «Αν μία σειρά τυχαίων δειγμάτων με n παρατηρήσεις ληφθούν από μία κανονική κατανομή με μέση τιμή μ και γνωστή τυπική απόκλιση σ , και ένα $1-\alpha$ διάστημα εμπιστοσύνης $\bar{y} \pm s_{\bar{y}} t_{\alpha/2}$ κατασκευαστεί για την κάθε σειρά δειγμάτων, τότε ένα ποσοστό $1-\alpha$ των συνολικών διαστημάτων εμπιστοσύνης θα περιέχει την πραγματική μέση τιμή μ και ένα ποσοστό α δεν θα την περιέχει. Ένας άλλος ορισμός είναι ότι υπάρχει $1-\alpha$ πιθανότητα ότι η πραγματική μέση τιμή θα εμπεριέχεται στο διάστημα εμπιστοσύνης).

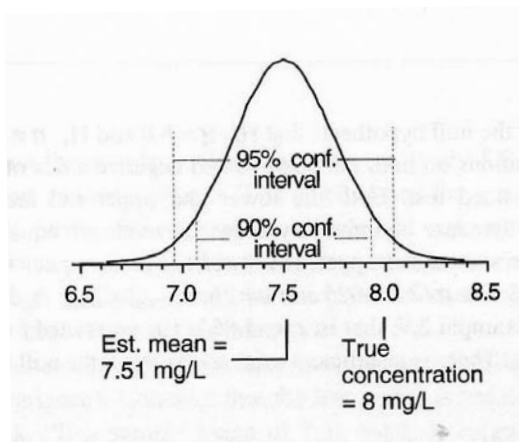
Με βάση το παράδειγμά μας, το ανώτερο και κατώτερο 95% όριο εμπιστοσύνης είναι (για μέση τιμή δειγμάτων 7.51, $s_{\bar{y}}=0.266$, και 26 βαθμούς ελευθερίας το $t_{26,0.025}$ είναι ίσο με 2.056):

$$7.51 - 2.056(0.266) < \mu < 7.51 + 2.056(0.266)$$

$$6.96 < \mu < 8.05$$

Το διάστημα περιέχει την (πραγματική) μέση τιμή 8.0 και συνεπώς μπορούμε να πούμε ότι η διαφορά y και μ δεν είναι τόσο μεγάλη (σε επίπεδο εμπιστοσύνης 95%).

Το γράφημα στη συνέχεια δείχνει τα διαφορετικά διαστήματα εμπιστοσύνης για 90% και 95%. Όπως φαίνεται, αν είχαμε επιλέξει το 90% ως επίπεδο εμπιστοσύνης, τότε η τιμή 8.0 δεν θα περιλαμβανόταν στο αντίστοιχο διάστημα.



Σχήμα 6.2. Διαφορετικά διαστήματα εμπιστοσύνης ανάλογα με επίπεδο εμπιστοσύνης

Το παραπάνω είναι t-τεστ και από τις δύο πλευρές. Η έκφραση «δύο πλευρές» σημαίνει απλά ότι η εναλλακτική υπόθεση $H_1 A \neq B$ μπορεί να σημαίνει ότι $A < B$ ή $A > B$. Υπάρχει και το t τεστ της μίας πλευράς, όπου η εναλλακτική υπόθεση μπορεί να είναι μόνο $A < B$ ή μόνο $A > B$. Στην πρώτη περίπτωση, το χαρακτηριστικό είναι ότι το επίπεδο σημαντικότητας α διαιρείται με 2 και κοιτάμε τη σχετική τιμή $t_{\alpha/2}$. Αυτό οφείλεται στο ότι το εμβαδό που αντιστοιχεί στο α (π.χ. 5%) είναι το άθροισμα των «εμβαδών» και από τις δύο πλευρές της t καμπύλης. Συνεπώς, το εμβαδόν από την κάθε πλευρά είναι ίσο με $\alpha/2$.

Γενικά η χρήση του τεστ του διαστήματος εμπιστοσύνης βολεύει στη σύγκριση δύο διεργασιών διότι πρακτικά γίνεται σύγκριση μέσων τιμών και ελέγχεται αν υπάρχει ή όχι το 0 στο αντίστοιχο διάστημα εμπιστοσύνης. Συγκεκριμένα:

- Στην περίπτωση που το διάστημα εμπιστοσύνης (για καθορισμένο α) της διαφοράς των μέσων τιμών περιέχει το 0, τότε οι πραγματικές μέσες τιμές και των δύο πληθυσμών είναι όμοιες και συνεπώς πρόκειται για τον ίδιο πληθυσμό
- Στην περίπτωση που το διάστημα εμπιστοσύνης (για καθορισμένο α) της διαφοράς των μέσων τιμών δεν περιέχει το 0, τότε οι πραγματικές μέσες τιμές και των δύο πληθυσμών είναι ανόμοιες και συνεπώς πρόκειται για δυο διαφορετικούς πληθυσμούς. Άρα υπάρχουν επαρκείς ενδείξεις ώστε να θεωρήσουμε ότι οι διεργασίες A και B είναι διαφορετικές. Ανάλογα μάλιστα με το αν το διάστημα είναι αρνητικό ή θετικό, μπορούμε να δούμε αν ο ένας πληθυσμός είναι μεγαλύτερος ή μικρότερος από τον άλλον.

Γενικά ο στόχος του t-test για δύο πληθυσμούς είναι να δούμε αν οι πληθυσμοί είναι όμοιοι ή ανόμοιοι. Παρόμοιο τεστ, που στηρίζεται στις ίδιες αρχές υπάρχει και πολλά δείγματα (δηλαδή για πληθυσμούς άνω των δύο) και ονομάζεται ανάλυση της διασποράς.

6.3 Δοκιμή t σε ζεύγη

Το εν λόγω τεστ βασίζεται στη **σύγκριση των διαφορών** των αποτελεσμάτων 2 τεχνικών. Οι τεχνικές εφαρμόστηκαν συγχρόνως (π.χ. δύο δείγματα λήφθηκαν συγχρόνως και το ένα αναλύθηκε με τη μέθοδο A, ενώ το δεύτερο αναλύθηκε με τη μέθοδο την ίδια στιγμή). Συνεπώς, δημιουργούνται ζεύγη τιμών ή παρατηρήσεων. Στο τεστ αυτό υπολογίζεται η μέση τιμή των διαφορών από όλα τα ζεύγη αυτά. Η θεωρία – που θα γίνει κατανοητή στη συνέχεια με ένα παράδειγμα – είναι η εξής:

Αν ορίσουμε ως δ την πραγματική μέση τιμή της διαφοράς μεταξύ των μεταβλητών y_1 & y_2 , που αποτελούν τα ζευγάρια των τιμών που αντιστοιχούν σε δύο διαφορετικές τεχνικές που θέλουμε να συγκρίνουμε. Το δ θα είναι ίσο με 0 αν τα y_1, y_2 προέρχονται από τον ίδιο πληθυσμό. Η στατιστική εκτίμηση του δ (\bar{d}) για n ζεύγη παρατηρήσεων είναι:

$$\bar{d} = \frac{\sum d_i}{n} = \frac{1}{n} \sum (y_{1,i} - y_{2,i})$$

Προφανώς, λόγω σφαλμάτων κατά την μέτρηση, η εκτίμηση του δ , αν όντως οι πληθυσμοί είναι όμοιοι, δεν θα είναι ίση με 0. Θα είναι κάπου κοντά στο 0. Η διασπορά (variance) των διαφορών d_i είναι:

$$s_d^2 = \frac{\sum (d_i - \bar{d})^2}{n-1}$$

Και, τέλος, το τυπικό σφάλμα της μέσης τιμής των διαφορών των μετρήσεων δίνεται από την κλασσική σχέση:

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}}$$

Το διάστημα εμπιστοσύνης (συνίσταται ως ο καλύτερος τρόπος καθορισμού της υπόθεσης μας, καλύτερος από το τεστ σημαντικότητας με χρήση των υποθέσεων) είναι $\bar{d} \pm s_{\bar{d}} \cdot t_{n-1, \alpha/2}$. Στη συνέχεια, με βάση τους παραπάνω κανόνες, κοιτάμε αν περιέχεται ή όχι η τιμή 0 εντός του υπολογιζόμενου διαστήματος εμπιστοσύνης (π.χ. το διάστημα -0.1 έως 0.05 περιέχει σαφώς την τιμή 0). Αν περιέχεται 0, τότε η πραγματική μέση τιμή της διαφοράς των παραμέτρων που εξετάζουμε περιέχει το μηδέν και συνεπώς οι παράμετροι προέρχονται από τον ίδιο πληθυσμό.

Το t-test σε ζευγάρια είναι προτιμητέο σε σχέση με το ανεξάρτητο t-test (το οποίο θα περιγραφεί στη συνέχεια). Αυτό ισχύει διότι μπορεί να υπάρχουν διακυμάνσεις (εποχικές, λόγω εργαστηρίου κ.λ.π.) μεταξύ των ζευγαριών που συγκρίνονται. Παρόλα αυτά, το t-test σε ζεύγη προϋποθέτει την διεξαγωγή των μετρήσεων με τις διαφορετικές τεχνικές σε ζεύγη και περίπου κατά την ίδια χρονική στιγμή. Ας δούμε τα παραπάνω με ένα παράδειγμα.

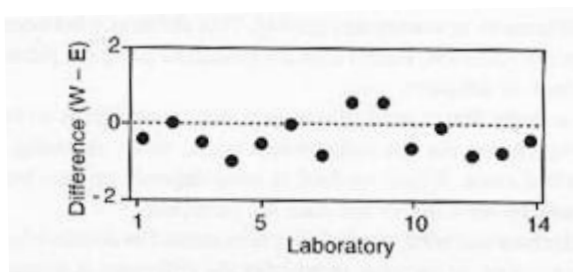
Παράδειγμα t-test

14 εργαστήρια κάνουν συγκρίσεις δύο τεχνικών μέτρησης διαλυτού οξυγόνου σε δείγμα γνωστής συγκέντρωσης διαλυτού οξυγόνου (1.2 mg/L). Οι τεχνικές είναι οι Winkler και η τεχνική του ηλεκτροδίου. Ο στόχος μας είναι να αποφανθούμε αν υπάρχουν διαφορές μεταξύ των δύο μεθόδων ή αν όντως και οι δύο εξάγουν την ίδια γνωστή τιμή (άρα αν είναι και οι δύο αξιόπιστες). Θεωρούμε ότι οι μετρήσεις που έχουν γίνει είναι ανεξάρτητες και τυχαίες (βασική προϋπόθεση για κάθε στατιστική ανάλυση είναι η τυχαιότητα και ανεξαρτησία των μετρήσεων).

Τα δεδομένα καταχωρούνται στον παρακάτω πίνακα και εξάγονται οι μέσες τιμές για κάθε μέθοδο. Επίσης γίνεται ένα γράφημα των διαφορών των αποτελεσμάτων των δύο τεχνικών για τα δύο εργαστήρια. Όπως αναλύθηκε στο κεφ. 5, κάνουμε πάντα γραφήματα ώστε να παρατηρήσουμε οπτικά πιθανές τάσεις ή διαφορές πριν προχωρήσουμε σε οποιαδήποτε στατιστική ανάλυση.

Πίνακας διαλυμένων οξυγόνων (όλες οι τιμές σε mg/L)

Εργαστήριο	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Μέση τιμή
Winkler (W)	1.2	1.4	1.4	1.3	1.2	1.3	1.4	2.0	1.9	1.1	1.8	1.0	1.1	1.4	1.4
Ηλεκτρόδιο (E)	1.6	1.4	1.9	2.3	1.7	1.3	2.2	1.4	1.3	1.7	1.9	1.8	1.8	1.8	1.7
Διαφορά (W-E)	-0.4	0	-0.5	-1	-0.5	0	-0.8	0.6	0.6	-0.6	-0.1	-0.8	-0.7	-0.4	-0.33



Σχήμα 6.3. Διαφορές των αποτελεσμάτων των τεχνικών Winkler και ηλεκτροδίου ανά εργαστήριο

Με μία πρώτη ματιά βλέπουμε ότι η μέση τιμή της τεχνικής Winkler είναι μικρότερη της τιμής του ηλεκτροδίου. Αυτό φαίνεται και από την μέση τιμή της διαφοράς W-E (-0.33) αλλά και από το παραπάνω γράφημα. Όμως πόσο ασφαλές είναι να πούμε ότι υπάρχουν διαφορές μεταξύ των τεχνικών. Οι μετρήσεις προέρχονται από 14 διαφορετικά εργαστήρια τα οποία μπορούν να εξάγουν διαφορετικές τιμές για την ίδια μέθοδο (λόγω διαφορετικών αντιδραστηρίων, συστημικών σφαλμάτων κ.λ.π). Συνεπώς είναι πιο ασφαλές να κάνουμε τις συγκρίσεις σε ζευγάρια, και στη συνέχεια να συγκρίνουμε τις διαφορές των μεθόδων μεταξύ τους.

Χρησιμοποιώντας τη θεωρία που αναπτύχθηκε παραπάνω, πραγματοποιούμε ένα t – τεστ για τις διαφορές των μεθόδων ανά εργαστήριο και παίρνουμε τη μέση τιμή των διαφορών, ήτοι την -0.33. Υπολογίζουμε επίσης το $s_d^2 = 0.24$ mg/L και το τυπικό σφάλμα της μέσης τιμής (των διαφορών), $s_{\bar{d}}$, είναι: 0.13 mg/L. Για 13 βαθμούς ελευθερίας (14-1) και $1-\alpha=95\%$, επιλέγω το $t_{13,0.025}$ το οποίο είναι 2.160. Τελικά το διάστημα εμπιστοσύνης, που περιέχει με βεβαιότητα 95% την πραγματική μέση τιμή της διαφοράς των 2 μεθόδων, είναι:

$$\bar{d} - s_{\bar{d}} \cdot t_{13,0.025} < \delta < \bar{d} + s_{\bar{d}} \cdot t_{13,0.025}$$

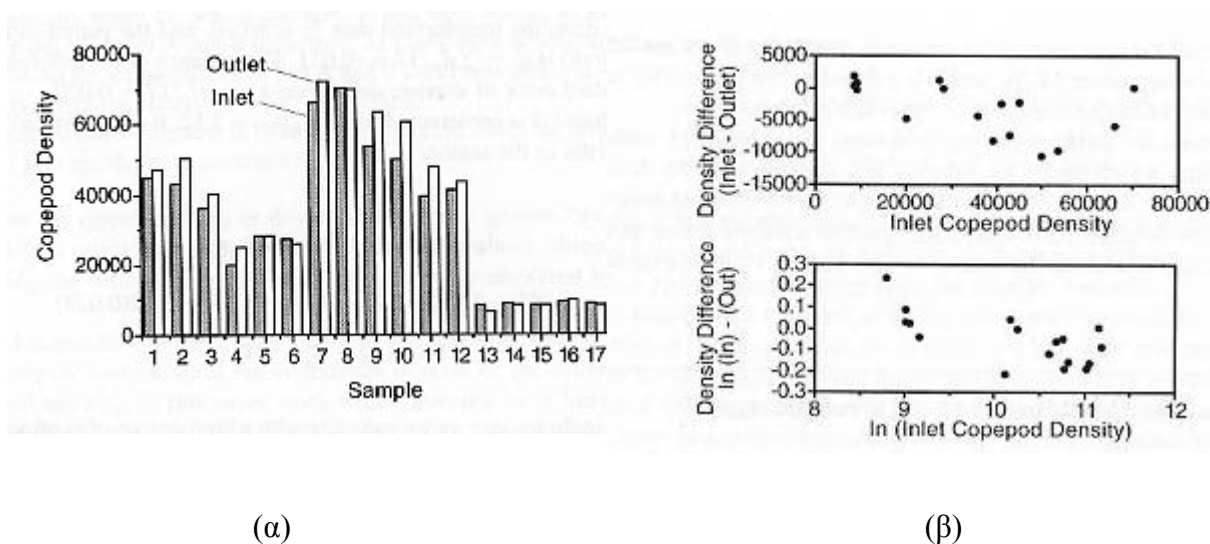
$$-0.61 \text{ mg/L} < \delta < -0.04 \text{ mg/L}$$

Είμαστε λοιπόν αρκετά σίγουροι ότι το δ δεν είναι 0 και συνεπώς οι δύο τεχνικές είναι διαφορετικές. Εξάγουν λοιπόν διαφορετικά αποτελέσματα κατά την μέτρηση του διαλυμένου οξυγόνου. Μάλιστα το γεγονός ότι το δ έχει αρνητικές τιμές, επιβεβαιώνει αυτό που αρχικά είδαμε, δηλ. ότι η τιμή του ηλεκτροδίου εξάγει μεγαλύτερες τιμές από τις τιμές Winkler.

Το θετικό της παραπάνω τεχνικής είναι ότι απαλείφει πιθανές διαφοροποιήσεις μεταξύ των εργαστηρίων, αφού στηρίζεται μόνο στις διαφορές των μεθόδων και δεν στηρίζεται στην σύγκριση των Winkler μεταξύ των διαφορετικών εργαστηρίων.

Με το παράδειγμα στη συνέχεια μπορούμε να δούμε πως μπορούν να απαλειφθούν εποχικές διακυμάνσεις όταν χρησιμοποιούμε το t-test σε ζεύγη.

Ας θεωρήσουμε ότι έχουμε τον πληθυσμό κάποιων υδρόβιων οργανισμών (ΥΟ) σε ποτάμι κοντά σε εργοστάσιο ηλεκτροπαραγωγής. Θέλουμε να δούμε τις επιπτώσεις του συστήματος ψύξης στους ΥΟ πριν και μετά την ψύξη (δηλ. σε κρύο και θερμό νερό αντίστοιχα). Λαμβάνουμε 17 δείγματα σε ζευγάρια (δηλ. κάθε δειγματοληψία αποτελείται από ένα δείγμα εισόδου και ένα δείγμα εξόδου και μέτρηση της πυκνότητας των ΥΟ (σε οργανισμούς/m³). Τα δείγματα 1-6 λαμβάνονται Νοέμβριο, τα 7-12 Φεβρουάριο, και τα 13-17 τον επόμενο Αύγουστο. Αναμένονται εποχικές διαφορές, που όντως φαίνονται στο παρακάτω σχήμα. Είναι λοιπόν μάλλον προφανές ότι η επεξεργασία δεν θα πρέπει να γίνει λαμβάνοντας την μέση τιμή των εισροών και την μέση τιμή των εκροών, αφού – λόγω των εποχικών διακυμάνσεων – θα οδηγούσε σε μεγάλες διασπορές. Η σύγκριση και ανάλυση πρέπει να γίνει σε ζεύγη, δηλ. υπολογίζοντας τη μέση τιμή των διαφορών της εισροής και εκροής για το κάθε δείγμα. Έτσι οι εποχικές διαφορές δεν επηρεάζουν άμεσα τις αναλύσεις.



Σχήμα 6.4. Εποχικές διακυμάνσεις μεταξύ συγκεντρώσεων υδατικών οργανισμών σε ποτάμι κοντά σε σταθμό ηλεκτροπαραγωγής (α) και διαφορές (απόλυτες και μετά από τροποποίηση των τιμών με χρήση λογαρίθμων) (β)

Σημειώνεται, ότι οι τροποποιήσεις απαιτούνται περιστασιακά στην στατιστική ανάλυση για εξομάλυνση των τιμών και ειδικά όταν έχουμε τιμές που διαφέρουν μεταξύ τους κατά τάξεις μεγέθους. Π.χ. Εδώ βλέπουμε ότι στο δείγμα 1 έχουμε πυκνότητα 44909 και στο δείγμα 13 8424 (στην είσοδο). Οι τροποποιημένες τιμές φαίνονται και στον παρακάτω πίνακα. Η τροποποίηση σε μεγάλες τιμές μπορεί να γίνει λαμβάνοντας και την τετραγωνική ρίζα. Γενικά, τροποποίηση τιμών γίνεται ώστε να επιτύχουμε ομοιόμορφη διασπορά των τιμών. π.χ. Οι τιμές στο κάτω γράφημα του 23 (β) είναι περισσότερο ομοιόμορφες σε σχέση με το πάνω γράφημα του 23 (α). Στο πρώτο γράφημα δηλ.

κυριαρχούν οι αρνητικές τιμές και φαίνεται ότι όσο αυξάνεται η πυκνότητα αυξάνεται και το μέγεθος της διαφοράς.

Πίνακας απόλυτων διαφορών και διαφορών τροποποιημένων τιμών

Sample	Original Counts (no./m ³)			Transformed Data, $z = \ln(y)$		
	y_{in}	y_{out}	$d = y_{in} - y_{out}$	z_{in}	z_{out}	$d_{in} = z_{in} - z_{out}$
1	44909	47069	-2160	10.712	10.759	-0.047
2	42858	50301	-7443	10.666	10.826	-0.160
3	35976	40431	-4455	10.491	10.607	-0.117
4	20048	24887	-4839	9.906	10.122	-0.216
5	28273	28385	-112	10.250	10.254	-0.004
6	27261	26122	1139	10.213	10.171	0.043
7	66149	72039	-5890	11.100	11.185	-0.085
8	70190	70039	151	11.159	11.157	0.002
9	53611	63228	-9617	10.890	11.055	-0.165
10	49978	60585	-10607	10.819	11.012	-0.192
11	39186	47455	-8269	10.576	10.768	-0.191
12	41074	43584	-2510	10.623	10.682	-0.059
13	8424	6640	1784	9.039	8.801	0.238
14	8995	8244	751	9.104	9.017	0.087
15	8436	8204	232	9.040	9.012	0.028
16	9195	9579	-384	9.126	9.167	-0.041
17	8729	8547	182	9.074	9.053	0.021
Average	33135	36196	-3062	10.164	10.215	-0.051
Std. deviation	20476	23013	4059	0.785	0.861	0.119
Std. error	4967	5582	984	0.190	0.209	0.029

Κοιτώντας τον παραπάνω πίνακα, βλέπουμε ότι αν στηριχτούμε απευθείας στις μέσες τιμές δεν θα μπορούμε να βγάλουμε άμεσα συμπέρασμα λόγω απλά της μεγάλης τυπικής απόκλισης των τιμών (αποτέλεσμα της εποχικής διακύμανσης). Π.χ. Η μέση τιμή στην εκροή είναι 36196 (> της μέσης τιμής της εισροής, 33135) αλλά οι τυπικές αποκλίσεις αντίστοιχα 23013 & 20476 είναι μεγάλες (αντίστοιχα οι τιμές συντελεστών απόκλισης είναι 64% & 62%).

Η σωστή προσέγγιση λοιπόν είναι καταρχάς:

- Η τροποποίηση των τιμών με χρήση λογαρίθμου (ή κάποια άλλη τροποποίηση) εφόσον δούμε ότι όντως η τροποποίηση αυτή κάνει περισσότερο ομοιόμορφα τα δεδομένα.
- Η σύγκριση και ανάλυση σε ζεύγη (των τροποποιημένων τιμών) ώστε να αποφευχθούν οι εποχικές διακυμάνσεις

6.4 t-test ανεξάρτητο

Σε περιπτώσεις που δεν είναι δυνατή η δειγματοληψία σε ζεύγη (π.χ. διαφορετικός αριθμός δειγμάτων ανά μέθοδο ή ακόμα και δειγματοληψία υπό διαφορετικές συνθήκες), τότε χρησιμοποιείται το ανεξάρτητο t-test. Συνήθως, αυτό είναι το t-test, που είναι συχνά περισσότερο γνωστό, όμως – για τους λόγους που προαναφέρθηκαν - το τεστ σε ζεύγη είναι αρτιότερο, όταν οι τιμές μπορούν να μετρηθούν σε ζεύγη φυσικά.

Η θεωρία πίσω από το ανεξάρτητο t-test είναι η εξής:

Έστω ότι έχουμε δύο πληθυσμούς με πραγματικές μέσες τιμές μ_1 και μ_2 και διασπορές σ_1^2 και σ_2^2 αντίστοιχα. Μπορούμε είτε να ελέγξουμε την μηδενική υπόθεση $\mu_1 - \mu_2 = 0$ είτε να ελέγξουμε το διάστημα εμπιστοσύνης της διαφοράς των μέσων τιμών. Το δεύτερο είναι προτιμότερο. Η αναμενόμενη τιμή (ΑΤ) της διαφοράς των πραγματικών μέσων τιμών των πληθυσμών είναι:

$$AT(\bar{y}_1 - \bar{y}_2) = \mu_1 - \mu_2$$

(Σημείωση: το ΑΤ σημαίνει αναμενόμενη τιμή).

Οι διασπορές των μέσων τιμών μ_1 και μ_2 είναι:

$V(\bar{y}_1) = \sigma_1^2/n_1$ και $V(\bar{y}_2) = \sigma_2^2/n_2$, με n_1 και n_2 ο αριθμός των μετρήσεων στο δείγμα 1 και 2 αντίστοιχα.

Η διαφορά των διασπορών είναι:

$$V(\bar{y}_1 - \bar{y}_2) = \sigma_1^2/n_1 + \sigma_2^2/n_2$$

Η τεχνική εδώ είναι ότι οι διασπορές μπορούν να ομογενοποιηθούν σε μία κοινή διασπορά, εφόσον οι επιμέρους διασπορές είναι σχεδόν παρόμοιες (ομοίου μεγέθους). Αν δεν ισχύει αυτό, τότε η διαφορά των διασπορών υπολογίζεται ως το απλό άθροισμα των επιμέρους διασπορών των 2 υπό μελέτη δειγμάτων.

(Υπενθυμίζεται ότι όλες οι προσεγγίσεις μας με τη χρήση της t κατανομής – που πρακτικά είναι μία κανονική κατανομή – προϋποθέτουν ακριβώς ότι τα δεδομένα μας κατανέμονται κανονικά, είναι τυχαία και ανεξάρτητα. Οι προϋποθέσεις αυτές βέβαια δεν ισχύουν πάντα στα περιβαλλοντικά συστήματα που τα δεδομένα συχνά κατανέμονται βάσει της κατανομής τύπου χ^2).

Παράδειγμα 1

Σε μία πόλη γίνονται 13 μετρήσεις σε διάφορα σημεία του δικτύου ύδρευσης ώστε να ελεγχθεί η ποιότητα του δημοτικού νερού, που προέρχεται από το δημοτικό δίκτυο ύδρευσης. Στα νερά μιας άλλης περιοχής, που υδρεύεται από ιδιωτικές γεωτρήσεις, γίνονται 10 μετρήσεις στα νερά. Παρατηρούνται συγκεντρώσεις υδραργύρου και στις δύο περιπτώσεις και το ερώτημα που τίθεται είναι αν οι συγκεντρώσεις υδραργύρου είναι διαφορετικές στις δύο περιοχές (πόλη, περιοχή με ιδιωτικές γεωτρήσεις) ή αν όντως τα νερά είναι όμοια (δηλ. προέρχονται από τον ίδιο υδροφόρο).

Οι τιμές του υδραργύρου είναι οι εξής ανά πηγή είναι:

Πηγή	Συγκεντρώσεις υδραργύρου (ppb)													
	0.34	0.18	0.13	0.09	0.16	0.09	0.16	0.10	0.14	0.26	0.06	0.26	0.07	
Πόλη ($n_{\pi}=13$)														
Ιδιωτικές πηγές ($n_{i\delta}=10$)	0.26	0.06	0.16	0.19	0.32	0.16	0.08	0.05	0.10	0.13				

Με βάση τις σχέσεις που αναφέρθηκαν παραπάνω, έχουμε ότι:

$$\bar{y}_{\pi} = 0.157 \text{ ppb}, s_{\pi}^2 = 0.0071 \text{ ppb}^2, s_{\pi} = 0.084 \text{ ppb}$$

$$\bar{y}_{i\delta} = 0.151 \text{ ppb}, s_{i\delta}^2 = 0.0076 \text{ ppb}^2, s_{i\delta} = 0.087 \text{ ppb}$$

Θεωρώντας ότι είναι ίδιου μεγέθους οι διασπορές¹⁸, οι παραπάνω διασπορές μπορούν να συνδυαστούν και να εξαχθεί μία κοινή διασπορά.. Η σχέση της κοινής (ή ομογενοποιημένης) διασποράς δίνεται από την παρακάτω σχέση:

$$s^2_{pool} = \frac{(n_{\pi} - 1)s_{\pi}^2 + (n_{i\delta} - 1)s_{i\delta}^2}{n_{\pi} + n_{i\delta} - 2} = 0.00734 \text{ ppb}^2$$

Η διασπορά της διαφοράς των μέσων τιμών είναι:

$$\begin{aligned} V(\bar{y}_{\pi} - \bar{y}_{i\delta}) &= \\ &= \frac{s^2_{pool}}{n_{\pi}} + \frac{s^2_{pool}}{n_{i\delta}} = s^2_{pool} \left(\frac{1}{n_{\pi}} + \frac{1}{n_{i\delta}} \right) = s^2_{y_{\pi}} + s^2_{y_{i\delta}} = 0.00734 \cdot \left(\frac{1}{13} + \frac{1}{10} \right) = 0.0013 \text{ ppb}^2 \end{aligned}$$

Η τυπική απόκλιση της διαφοράς των μέσων τιμών είναι προφανώς:

$$s_{\bar{y}_{\pi} - \bar{y}_{i\delta}} = \sqrt{s^2_{pool} \left(\frac{1}{n_{\pi}} + \frac{1}{n_{i\delta}} \right)} = s_{pool} \sqrt{\frac{1}{n_{\pi}} + \frac{1}{n_{i\delta}}} = \sqrt{0.0013} = 0.036 \text{ ppb}$$

Σημειώνεται ότι οι παραπάνω πράξεις θα μπορούσαν να γίνουν και με χρήση των διασπορών των 13 τιμών της πόλης και των 10 τιμών των ιδιωτικών πηγαδιών, ξεχωριστά (μπορεί να ελεγχθεί αυτό).

¹⁸ Ιδίου μεγέθους σημαίνει ότι ο λόγος s_1^2/s_2^2 , τιμή που κατανομείται βάσει της κατανομής χ^2 , που αλλιώς ονομάζεται κατανομή F, είναι μικρότερος από την τιμή της F που αντιστοιχεί σε v_1, v_2 βαθμούς ελευθερίας για επίπεδο σημαντικότητας α . Γενικά πρέπει να είναι μικρότερος από ένα εύρος 2.5, που σημαίνει ότι οι διασπορές είναι όμοιες για τα δύο υπό μελέτη δείγματα. Στην προκειμένη περίπτωση παρατηρούμε ότι οι τιμές s_{π}^2 & $s_{i\delta}^2$ βρίσκονται πολύ κοντά και συγκεκριμένα $s_{\pi}^2/s_{i\delta}^2 = 0.9$ και άρα οι διασπορές θεωρούνται ομοίου μεγέθους, επιτρέποντας το υπολογισμό στη συνέχεια μίας κοινής διασποράς s^2_{pool} . Αν δεν ήταν όμοιες, τότε ο υπολογισμός στη συνέχεια θα γινόταν με τις 2 διαφορετικές αυτές διασπορές.

Με βάση τα παραπάνω, και για $12+9=21$ βαθμούς ελευθερίας και $1-\alpha=95\%$ (δηλ. διάστημα εμπιστοσύνης 95%) βρίσκω το $t_{21,0.025}$, που είναι ίσο με 2.080.

Το διάστημα εμπιστοσύνης που αναμένεται να περιέχει την πραγματική διαφορά των τιμών υδραργύρου στα νερά της πόλης και τα ιδιωτικά νερά είναι:

$$(\bar{y}_\pi - \bar{y}_{id}) \pm t_{21,0.025} \cdot s_{\bar{y}_\pi - \bar{y}_{id}} = (0.157 - 0.151) \pm 2.080 \cdot (0.036) = 0.006 \pm 0.075 \text{ ppb}$$

Συνεπώς, παρατηρούμε ότι το παραπάνω διάστημα εμπιστοσύνης περιέχει σαφώς το 0. (κυμαίνεται από -0.069 έως 0.081 ppb). Άρα η πραγματική τιμή της διαφοράς των συγκεντρώσεων υδραργύρου στα νερά της πόλης και των γεωτρήσεων, αναμένεται με βεβαιότητα 95%, να ανήκει στο εν λόγω διάστημα εμπιστοσύνης. Άρα, με βεβαιότητα 95%, η διαφορά αυτή περιέχει το 0, και συνεπώς μπορούμε να συμπεράνουμε ότι δεν υπάρχουν ενδείξεις για το ότι οι συγκεντρώσεις υδραργύρου στα δύο νερά διαφέρουν. Άρα τα νερά, αναφορικά με τον υδράργυρο, είναι όμοια, δηλ. προέρχονται από τον ίδιο υδροφόρο. Μελλοντική δειγματοληψία για υδράργυρο μπορεί να γίνει σε οποιαδήποτε από τις δύο πηγές (δίκτυο πόλης ή ιδιωτικά πηγάδια) χωρίς να νοιαζόμαστε για τις διαφορετικές τοποθεσίες.

Παράδειγμα 2

Θέλουμε να μετρήσουμε το pH μίας μάζας στερεών αποβλήτων. Μία τεχνική που έχει προταθεί για τη μέτρηση αυτή είναι η χρήση λόγου 1:10 (υγρό βάρος στερεού / βάρος απεσταγμένου νερού), ενώ μία άλλη τεχνική αναφέρει απλά τη προσθήκη νερού ώστε να υγρανθεί το στερεό δείγμα και στη συνέχεια μέτρηση του pH. Αποφασίζετε να κάνετε πολλές μετρήσεις σε διάφορα μίγματα γρασιδιού και χαρτιού (συστατικά του οργανικού κλάσματος των αστικών στερεών αποβλήτων), και με τις δύο τεχνικές, ώστε να συγκρίνετε τις δοκιμές. Δηλαδή, ετοιμάζετε 7 δείγματα με λόγο 1:10 και 7 δείγματα με απλή προσθήκη νερού και μετράτε τις τιμές pH με το ίδιο pHμετρο. Δεν υπάρχει απαραίτητα αντιστοιχία του 1^{ου} δείγματος της μίας τεχνικής με το 1^ο δείγμα της άλλης τεχνικής. Τα αποτελέσματα είναι τα εξής:

Δείγματα με λόγο 1:10	Δείγματα με απλή προσθήκη νερού
5,6	6,1
5,8	6,2
5,4	5,9
5,5	6,0
5,6	6,0
5,7	6,3
5,9	6,1

Αναλύστε τα δεδομένα (t-test) ώστε να δείτε αν υπάρχουν διαφορές μεταξύ των τεχνικών τελικά? Τι συμπεραίνετε?

Θα ήταν καταρχάς λάθος να κάναμε χρήση του τεστ t με τις διαφορές, αφού δεν υπάρχει απαραίτητα αντιστοιχία μεταξύ της κάθε μέτρησης του ενός δείγματος με την αντίστοιχη

μέτρηση του άλλου δείγματος. Συνεπώς, θα γίνει χρήση του ανεξάρτητου t-test. Βάσει των σχέσεων που αναφέρονται παραπάνω, έχουμε:

Για δείγμα 1:10 έχω: $n=7, v=6, \bar{y}_{1_10} = 5,64, s_{1_10}^2 = 0,0295, s_{1_10} = 0,172$

Για δείγμα με νερό έχω: $n=7, v=6, \bar{y}_{\text{νερό}} = 6,09, s_{\text{νερό}}^2 = 0,0181, s_{\text{νερό}} = 0,135$

Θεωρώντας ότι είναι ίδιου μεγέθους οι διασπορές, οι παραπάνω διασπορές μπορούν να συνδυαστούν και να εξαχθεί μία κοινή/ομογενοποιημένη διασπορά s_{pool}^2 . Η σχέση της κοινής (ή ομογενοποιημένης) διασποράς δίνεται από την παρακάτω σχέση:

$$s_{\text{pool}}^2 = \frac{(n_{1_10} - 1) \cdot s_{1_10}^2 + (n_{\text{νερό}} - 1) \cdot s_{\text{νερό}}^2}{n_{1_10} + n_{\text{νερό}} - 2} = 0,0238$$

Η διασπορά της διαφοράς των μέσων τιμών είναι:

$$\begin{aligned} V(\bar{y}_{1_10} - \bar{y}_{\text{νερό}}) &= \\ &= \frac{s_{\text{pool}}^2}{n_{1_10}} + \frac{s_{\text{pool}}^2}{n_{\text{νερό}}} = s_{\text{pool}}^2 \left(\frac{1}{n_{1_10}} + \frac{1}{n_{\text{νερό}}} \right) = s_{\bar{y}_{1_10}}^2 + s_{\bar{y}_{\text{νερό}}}^2 = 0,0034 + 0,0034 = 0,0068 \end{aligned}$$

Η τυπική απόκλιση της διαφοράς των μέσων τιμών είναι προφανώς:

$$s_{\bar{y}_{1_10} - \bar{y}_{\text{νερό}}} = \sqrt{0,0068} = 0,0825$$

Με βάση τα παραπάνω, και για $6+6=12$ βαθμούς ελευθερίας και $1-\alpha=95\%$ (δηλ. διάστημα εμπιστοσύνης 95%) βρίσκω το $t_{12,0,025}$, που είναι ίσο με 2.179.

Το διάστημα εμπιστοσύνης είναι τελικά:

$$(\bar{y}_{1_10} - \bar{y}_{\text{νερό}}) \pm t_{12,0,025} \cdot s_{\bar{y}_{1_10} - \bar{y}_{\text{νερό}}} = (5,643 - 6,086) \pm 2,179 \cdot (0,0825) = (-0,62 \text{ έως } -0,26)$$

Επειδή το Δ.Ε. δεν περιέχει την τιμή 0, συμπεραίνουμε ότι τα διαφορετικά τεστ μέτρησης pH όντως παράγουν διαφορετικά αποτελέσματα. Συνεπώς, πρέπει σαφώς να διευκρινίζουμε τον τρόπο μέτρησης pH στα στερεά απόβλητα. Συγκεκριμένα, η μέτρηση pH με διάλυση 1:10 παράγει – με βεβαιότητα 95% - χαμηλότερες τιμές από την τεχνική μέτρησης pH με απευθείας διάλυσης με νερό.

7. Σύγκριση πολλαπλών μέσων τιμών (k)

Η σύγκριση πολλών μέσων όρων έχει προφανώς ομοιότητες με τη σύγκριση 2 μέσων όρων με τη χρήση t τεστ. Δηλαδή, το απλούστερο που θα μπορούσε κάποιος/α να κάνει όταν έχει άνω των 2 διεργασιών για σύγκριση, είναι πολλαπλά t-tests μεταξύ όλων των δυνατών συνδυασμών ζευγαριών. Για k μέσους όρους, παραδειγματικά, μπορούν να γίνουν $k(k-1)/2$ συγκρίσεις ζευγών. Π.χ. με 3 μέσες τιμές, έχω $3*2/2=3$ ζεύγη να συγκρίνω. Για 15 μέσες τιμές, υπάρχουν 105 συνδυασμοί! Κάτι τέτοιο βέβαια θα μπορούσε να γίνει, αλλά συνήθως όταν υλοποιούμε εργαστηριακά πειράματα, οι πόροι (οικονομικοί και χρονικοί) είναι συχνά περιορισμένοι. Συνεπώς, δεν είναι πάντα εφικτή η πραγματοποίηση μεγάλου αριθμού πειραμάτων.

Υπάρχει και το εξής αρνητικό με τη χρήση του t-test με πολλαπλά ζεύγη. Παράδειγμα, αν έχουμε 3 μεθόδους για σύγκριση, την A, B και Γ και αν η μέση τιμή του A είναι μεγαλύτερη του μέσου όρου του B και του Γ είναι ελάχιστα μεγαλύτερο του A (δηλ. $\bar{y}_\Gamma > \bar{y}_A > \bar{y}_B$), είναι πιθανό τα 3 ανεξάρτητα αυτά t τεστ μεταξύ όλων των μεθόδων (δηλ. A με B, B με Γ και A με Γ) να δείξουν ότι $\mu_A > \mu_B$, $\mu_A = \mu_\Gamma$ και $\mu_B = \mu_\Gamma$. Αυτή η αντίφαση μπορεί να συμβεί γιατί χρησιμοποιούνται διαφορετικές διασπορές σε κάθε σύγκριση. Για να αποφευχθεί αυτή η αντίφαση, θεωρείται ότι οι διασπορές όλων των διεργασιών είναι ιδίου μεγέθους και χρησιμοποιείται μία κοινή διασπορά (ομοιάζει με την προσέγγιση στην περίπτωση του t-τεστ για ανεξάρτητα δείγματα, όπου ορίσαμε την S^2_{pool}).

Παράδειγμα

Έστω ότι 5 διαφορετικά εργαστήρια κάνουν 10 επαναληπτικές μετρήσεις του ιδίου δείγματος. Τα δεδομένα, οι μέσοι όροι και οι διασπορές ανά εργαστήριο φαίνονται στον παρακάτω πίνακα.

	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5
	3.4	4.5	5.3	3.2	3.3
	3.0	3.7	4.7	3.4	2.4
	3.4	3.8	3.6	3.1	2.7
	5.0	3.9	5.0	3.0	3.2
	5.1	4.3	3.6	3.9	3.3
	5.5	3.9	4.5	2.0	2.9
	5.4	4.1	4.6	1.9	4.4
	4.2	4.0	5.3	2.7	3.4
	3.8	3.0	3.9	3.8	4.8
	4.2	4.5	4.1	4.2	3.0
Mean	$\bar{y} = 4.30$	3.97	4.46	3.12	3.34
Variance	$s^2 = 0.82$	0.19	0.41	0.58	0.54

Το ερώτημα που τίθεται πάλι είναι αν τα εργαστήρια παράγουν όμοιες μετρήσεις, ή κάποιο/α διαφέρουν σημαντικά μεταξύ τους (υπερεκτιμούν ή υποεκτιμούν την πραγματική τιμή)¹⁹. Δηλαδή, θα ήταν σωστό να πούμε ότι το εργαστήριο 3 παράγει

¹⁹ Είναι προφανές, και σε κάθε περίπτωση υπενθυμίζεται, ότι τέτοιες συγκρίσεις μπορούν να γίνουν εφόσον γίνονται επαναλήψεις για την κάθε μέτρηση. Δηλαδή αν είχαμε μόνο μία τιμή (μέτρηση) ανά

πάντα μεγαλύτερες τιμές από τα άλλα? Ή μήπως η τιμή 4.46 είναι απλά αποτέλεσμα πειραματικού σφάλματος και δεν υπάρχουν πραγματικές διαφορές μεταξύ των εργαστηρίων. Τα τεστ που θα εφαρμοστούν στη συνέχεια είναι:

- Το πολλαπλό t-test του Tukey (σύγκριση σε ζεύγη).
- Το πολλαπλό t-test του Dunnett (σύγκριση με control).

7.1 Δοκιμή Tukey για σύγκριση πολλαπλών διεργασιών

Το $(1-\alpha)\%$ διάστημα εμπιστοσύνης²⁰ της πραγματικής διαφοράς μεταξύ των μέσων τιμών δύο εργαστηρίων (ή 2 διεργασιών γενικότερα) είναι:

$$(\bar{y}_i - \bar{y}_j) \pm t_{v, \alpha/2} \cdot s_{pool} \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

ΥΠΕΝΘΥΜΙΣΗ: Η παράσταση $s_{\frac{y\pi - y\delta}{\gamma\pi - \gamma\delta}}$ είναι ίση με την $s_{pool} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$

Το κρίσιμο σημείο εδώ είναι ότι χρησιμοποιούμε μία κοινή τυπική απόκλιση για την κάθε διεργασία (εργαστήριο). Θεωρούμε δηλαδή ότι έχουν κοινό s. Το κοινό αυτό s υπολογίζεται από τη σχέση, που παρουσιάστηκε και παραπάνω, και που είναι μία σχέση που υπολογίζει ένα είδος μέσης τιμής διασποράς δίνοντας μεγαλύτερο βάρος στην διεργασία με τη μεγαλύτερη διασπορά, λαμβάνοντας υπόψη τους βαθμούς ελευθερίας:

$$s_{pool}^2 = \frac{(n_i - 1)s_i^2 + (n_j - 1)s_j^2}{n_i + n_j - 2}$$

Η πιθανότητα ότι το διάστημα εμπιστοσύνης περιέχει την πραγματική τιμή της διαφοράς για κάθε μεμονωμένη σύγκριση 2 διεργασιών είναι $1-\alpha$. Αλλά η πιθανότητα ότι όλα τα $k(k-1)/2$ διαστήματα εμπιστοσύνης θα περιέχουν ταυτόχρονα την πραγματική τιμή είναι μικρότερη του $1-\alpha$. Το διάστημα εμπιστοσύνης της διαφοράς δύο πραγματικών μέσων τιμών (μ_i, μ_j), λαμβάνοντας υπόψη ότι όλες οι δυνατές συγκρίσεις των k διεργασιών (5 εργαστηρίων στην προκειμένη περίπτωση) μπορούν να γίνουν, υπολογίζεται από τη σχέση²¹:

εργαστήριο (δηλ. 5 τιμές), τότε δεν μπορούν να βγουν συμπεράσματα αφού δεν μπορεί να εξαχθεί κάποια διασπορά ανά εργαστήριο.

²⁰ Όπως βλέπετε, προτείνεται η χρήση της έννοιας του διαστήματος εμπιστοσύνης, το οποίο είναι καλύτερο σε σχέση με το κλασικό τεστ σημαντικότητας με τη χρήση μηδενικής και εναλλακτικής υπόθεσης. Προφανώς, και οι 2 προσεγγίσεις παράγουν τα ίδια αποτελέσματα αφού στηρίζονται στις ίδιες σχέσεις (δηλ. τη συνάρτηση t).

²¹ Υπενθυμίζεται ότι k είναι ο αριθμός των διεργασιών που συγκρίνουμε, n ο αριθμός των μετρήσεων ανά διεργασία, v είναι οι συνολικοί βαθμοί ελευθερίας και $\alpha/2$ το επίπεδο εμπιστοσύνης διαιρούμενο με 2.

$$(\bar{y}_i - \bar{y}_j) \pm \frac{q_{k,v,\alpha/2}}{\sqrt{2}} \cdot s_{pool} \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

Παρατηρούμε ότι στην παραπάνω σχέση δεν εμφανίζεται η συνάρτηση t. Εμφανίζεται κάποιο q, το οποίο από τους δείκτες που το συνοδεύουν, εκτιμούμε ότι θα πρέπει να υπολογίζεται συναρτήσει των παραμέτρων k, v & α. Συγκεκριμένα, το $q_{k,v,\alpha/2}$ είναι το άνω επίπεδο σημαντικότητας του εύρους για k διεργασίες και v βαθμούς ελευθερίας κατά την εκτίμηση του s^2_{pool} . Η κοινή τιμή του s^2_{pool} , για όλες τις k διεργασίες, υπολογίζεται από τη σχέση:

$$s^2_{pool} = \frac{(n_1 - 1)s_1^2 + \dots + (n_k - 1)s_k^2}{n_1 + \dots + n_k - k}$$

Γενικά η εκτίμηση του διαστήματος εμπιστοσύνης όταν χρησιμοποιείται η συνάρτηση q είναι μεγαλύτερη σε σύγκριση με όταν χρησιμοποιούμε τη συνάρτηση t. Αυτό γίνεται γιατί η συνάρτηση q (τιμές της οποίας υπάρχουν σε πίνακες, όπως και με όλες τις κατανομές) θεωρεί ότι η εκτίμηση του διαστήματος εμπιστοσύνης μπορεί να γίνει από ένα οποιοδήποτε εκ των $k(k-1)/2$ ζευγαριών σύγκρισης. Είναι λοιπόν τιμή περισσότερο ασφαλής για την εκτίμηση του διαστήματος εμπιστοσύνης.

Ο πίνακας στη συνέχεια δίνει κάποιες τιμές της συνάρτησης $q_{k,v,\alpha/2}$ για 95% κοινό (ή «οικογενειακό» διάστημα εμπιστοσύνης), όπου k είναι ο αριθμός των διεργασιών και v οι συνολικοί βαθμοί ελευθερίας (δηλ. $v=n_i-1+n_j-1+\dots+n_k-1$).

v	k						
	2	3	4	5	6	8	10
5	4.47	5.56	6.26	6.78	7.19	7.82	8.29
10	3.73	4.47	4.94	5.29	5.56	5.97	6.29
15	3.52	4.18	4.59	4.89	5.12	5.47	5.74
20	3.43	4.05	4.43	4.70	4.91	5.24	5.48
30	3.34	3.92	4.27	4.52	4.72	5.02	5.24
60	3.25	3.80	4.12	4.36	4.54	4.81	5.01
∞	3.17	3.68	3.98	4.20	4.36	4.61	4.78

Note: Family error rate = 5%; $\alpha/2 = 0.05/2 = 0.025$.

Με $k=5$, $v=(10-1)+(10-1)+(10-1)+(10-1)+(10-1)=50-5=45$ (συνολικά για όλα τα k), και $s_{pool} = 0.71$, τότε $q_{5,45,0.025} = 4.49$. Έτσι το διάστημα εμπιστοσύνης για την κάθε διαφορά $\bar{y}_i - \bar{y}_j$ είναι:

$$(\bar{y}_i - \bar{y}_j) \pm \frac{4.49}{\sqrt{2}} \cdot 0.71 \cdot \sqrt{\frac{1}{10} + \frac{1}{10}}$$

Τελικά, το διάστημα εμπιστοσύνης είναι: $(\bar{y}_i - \bar{y}_j) \pm 1.01$

Στη συνέχεια το διάστημα που εξάγεται, εφαρμόζεται χωριστά για κάθε διαφορά διεργασιών, όπως εμείς την επιλέγουμε. Το διάστημα αυτό υπολογίζεται διαφορετικά,

ανάλογα με τις διεργασίες που θέλουμε να συγκρίνουμε. Συγκεκριμένα, τα ζεύγη των διαφορών για σύγκριση είναι $5(5-1)/2 = 10$. Τα διαστήματα εμπιστοσύνης για την κάθε διαφορά φαίνονται λοιπόν στον παρακάτω πίνακα:

	<i>Εργ.1</i>	<i>Εργ.2</i>	<i>Εργ.3</i>	<i>Εργ.4</i>	<i>Εργ.5</i>
<i>Εργ. 1:</i>	-				
<i>Εργ. 2:</i>	(-0.68, 1.34)	-			
<i>Εργ. 3:</i>	(-1.17, 0.85)	(-1.5, 0.52)	-		
<i>Εργ. 4:</i>	(0.17, 2.19)	(-0.16, 1.86)	(0.33, 2.35)	-	
<i>Εργ. 5:</i>	(-0.05, 1.97)	(-0.38, 1.64)	(0.11, 2.13)	(-1.23, 0.79)	-

Με βάση τα παραπάνω, για όλα τα διαστήματα που περιέχουν 0 μπορούμε να πούμε (με 95% βεβαιότητα) ότι δεν υπάρχουν σημαντικές διαφορές μεταξύ των εν λόγω διεργασιών. Από την άλλη μπορούμε να πούμε (με 95% βεβαιότητα) ότι το εργαστήριο 3 παράγει σταθερά μεγαλύτερες τιμές από το εργαστήριο 4 και 5, ενώ το εργαστήριο 1 παράγει επίσης σταθερά μεγαλύτερες τιμές από το εργαστήριο 4 (έχουν όλα διαστήματα εμπιστοσύνης στη θετική πλευρά). Δεν μπορούμε βέβαια να αποφανθούμε για το ποιο εργαστήριο είναι σωστό ή όχι, αφού δεν γνωρίζουμε την πραγματική τιμή του δείγματος. Το MINITAB έχει το εν λόγω τεστ με τον τίτλο «Pairwise Comparison».

7.2 Δοκιμή Dunnett για σύγκριση πολλαπλών διεργασιών

Η δοκιμή Dunnett στηρίζεται στο ότι μία εκ των διεργασιών μπορεί να είναι κάποιο standard ή κάποια διεργασία control. Π.χ. όταν κάνουμε δειγματοληψίες στον αέρα μίας ρυπασμένης περιοχής, λαμβάνουμε δείγματα κάπου μακριά, εκτός της περιοχής, ώστε να χρησιμοποιηθεί αυτή η μέτρηση για σύγκριση με τις υπόλοιπες. Αυτή είναι η μέτρηση control (τυφλό δείγμα). Αν λοιπόν έχω k διεργασίες, μία εκ των οποίων είναι η διεργασία control, τότε αναμένω να κάνω k-1 συγκρίσεις των διεργασιών ξεχωριστά κάθε φορά με το control. Δηλαδή δεν συγκρίνω όλες τις διεργασίες μεταξύ τους, αλλά την κάθε μία με το τυφλό. Για αυτό το λόγο έχω σχετικά λιγότερες συγκρίσεις. Στην περίπτωση αυτή, ο Dunnett όρισε το διάστημα εμπιστοσύνης της διαφοράς της μέσης τιμής της κάθε διεργασίας από την μέση τιμή του (καθορισμένου) control ως εξής:

$$(\bar{y}_i - \bar{y}_c) \pm t_{k-1, \nu, \alpha/2} \cdot s_{pool} \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_c}}$$

Εδώ δεν εμφανίζεται η συνάρτηση q του Tukey. Εμφανίζεται η συνάρτηση t, η οποία όμως εδώ ορίζεται από 3 παραμέτρους. Το k, το ν και το α/2. Έως τώρα ξέραμε ότι το t συνδέεται μόνο με το ν και το α. Ο πίνακας του Dunnett για τις τιμές t συναρτήσει των k, ν, α καθορίζεται από την τιμή των διεργασιών k, των βαθμών ελευθερίας ν και το επίπεδο εμπιστοσύνης. Αντί για k το t υπολογίζεται βάσει του k-1, γιατί έχουμε συγκρίσεις των k-1 διεργασιών κάθε φορά ξεχωριστά με το control. Και εδώ, όπως και στην περίπτωση του τεστ Tukey, οι υπολογισμοί γίνονται βάσει της κοινής διασποράς ($s_{\frac{y_i - y_{i0}}{y_i - y_{i0}}}$) της διαφοράς της διεργασίας i με το control. Το MINITAB έχει το παραπάνω τεστ με τον τίτλο «Comparisons with a control».

Ο πίνακας στη συνέχεια δίνει τις τιμές του t βάσει των k-1, v, α/2 για το τεστ Dunnett.

Πίνακας τιμών $t_{k-1,v,0.025}$ για κοινό επίπεδο εμπιστοσύνης 1-α=95% για σύγκριση k-1 διεργασιών με ένα control

v	k - 1 = Number of Treatments Excluding the Control						
	2	3	4	5	6	8	10
5	3.03	3.29	3.48	3.62	3.73	3.90	4.03
10	2.57	2.76	2.89	2.99	3.07	3.19	3.29
15	2.44	2.61	2.73	2.82	2.89	3.00	3.08
20	2.38	2.54	2.65	2.73	2.80	2.90	2.98
30	2.32	2.47	2.58	2.66	2.72	2.82	2.89
60	2.27	2.41	2.51	2.58	2.64	2.73	2.80
∞	2.21	2.35	2.44	2.51	2.57	2.65	2.72

Παράδειγμα

Λαμβάνοντας υπόψη τα δεδομένα παραπάνω, θεωρούμε ότι το εργαστήριο 2 είναι το control. Συνεπώς, αναπτύσσουμε τη σχέση του διαστήματος εμπιστοσύνης για τις υπόλοιπες 4 διεργασίες σε σχέση με το control. **ΠΡΟΣΟΧΗ:** Η κοινή διασπορά s_{pool} εξάγεται και από τις 5 διεργασίες (5 εργαστήρια στην περίπτωση αυτή). Η τιμή της λοιπόν παραμένει ίση με 0.71. Για 4 διεργασίες (k-1) η τιμή $t_{4,45,0.025}$ είναι 2.55 (αν και δεν φαίνεται με ακρίβεια από τον παραπάνω πίνακα). Έχω λοιπόν, ότι το διάστημα εμπιστοσύνης για τη διαφορά του εργαστηρίου i από το control είναι τελικά:

$$(\bar{y}_i - \bar{y}_c) \pm 2.55 \cdot (0.71) \cdot \sqrt{\frac{1}{10} + \frac{1}{10}}$$

Και στη συνέχεια υπολογίζω το διάστημα για την κάθε διεργασία, όπως φαίνεται στη συνέχεια:

Εργαστήριο	Control	1	3	4	5
Μέση τιμή	3.97	4.30	4.46	3.12	3.34
Διάστημα εμπιστοσύνης	-	(-0.48,1.14)	(-0.32,1.3)	(-1.66, -0.04)	(-1.44,0.18)

Με βάση τον παραπάνω πίνακα, μόνο το εργαστήριο 4 φαίνεται ότι διαφέρει από το control. Τα υπόλοιπα 3 εργαστήρια είναι όμοια με το control.

Ο Box επίσης πρότεινε την τεχνική της μελέτης της κατανομής των k μέσων τιμών που προέρχονται από τις k διεργασίες. Η ιδέα στηρίζεται στο ότι αν k διεργασίες είναι όμοιες, δηλ. προέρχονται από τον ίδιο πληθυσμό, θα έχουν την ίδια μέση τιμή και άρα οι μέσες τιμές των διεργασιών θα κατανέμονται κανονικά γύρω από την κοινή μέση τιμή με τυπική απόκλιση ίση με σ/\sqrt{n} . Αν λοιπόν κατασκευαστεί σε ένα γράφημα μία καμπύλη κανονικής κατανομής και «περάσει» πάνω από τις k μέσες τιμές, τότε οπτικά μπορεί να δεί κάποιος/α για το αν οι τιμές αυτές κατανέμονται κανονικά. Αν όχι, τότε υπάρχουν διαφορές μεταξύ των εν λόγω k τιμών. Γραφικά, μπορούμε να δούμε ποιά διεργασία απέχει αρκετά από μία πιθανή κοινή μέση τιμή.

8. Ανάλυση διασποράς για k δείγματα (ANOVA)

Η ανάλυση διασποράς έχει θεωρηθεί ως ένα από τα σημαντικότερα εργαλεία στατιστικής ανάλυσης. Στόχος της είναι να ελέγξει αν k (τουλάχιστον 3) δείγματα προέρχονται από τον ίδιο πληθυσμό ή όχι και κατά συνέπεια αν έχουν τον ίδιο μέσο όρο μ ή όχι. Παρόλα αυτά, η αρχή της ANOVA – σε αντίθεση με τα τεστ των Tukey, Dunnett – είναι στο ότι συγκρίνει τις διασπορές των δειγμάτων και όχι τους μέσους όρους. Συγκεκριμένα, συγκρίνει τη μέση διασπορά που παρατηρείται εντός των διαφορετικών διεργασιών με τη μέση διασπορά μεταξύ των διεργασιών αυτών. Αν αυτές οι διασπορές είναι περίπου όμοιες τότε δεν υπάρχει και διαφορά μεταξύ των διεργασιών. Συνεπώς, όλες προέρχονται από τον ίδιο πληθυσμό. Αν η διασπορά, όμως, μεταξύ των διεργασιών είναι αρκετά μεγαλύτερη (για επίπεδο εμπιστοσύνης που εμείς καθορίζουμε) από τη διασπορά εντός των διεργασιών, τότε υπάρχουν και διαφοροποιήσεις μεταξύ των δειγμάτων.

Η διασπορά (s_w^2) εντός των διεργασιών αποτελεί το πειραματικό σφάλμα, που φυσιολογικά υπάρχει μεταξύ τυχαίων επαναληπτικών μετρήσεων. Δηλαδή, η εντός των διεργασιών διασπορά – εφόσον οι επαναληπτικές μετρήσεις έγιναν τυχαία – είναι, κατά πάσα πιθανότητα, μία καλή εκτίμηση του φυσιολογικού και τυχαίου πειραματικού σφάλματος που πάντα υπάρχει για επαναληπτικές μετρήσεις. Αν λοιπόν οι διεργασίες διαφέρουν όντως μεταξύ τους, τότε η *μεταξύ των διεργασιών* διασπορά (s_b^2) αναμένεται να είναι μεγαλύτερη από την *εντός των διεργασιών* διασπορά, δηλ. από το τυχαίο πειραματικό σφάλμα. Με βάση τα παραπάνω, η ANOVA στηρίζεται στη σύγκριση του πηλίκου s_b^2/s_w^2 , καταρχάς με τη μονάδα. Πέρα από αυτό, επειδή το παραπάνω πηλίκο – όπως και η διασπορά η ίδια – κατανέμεται βάσει της κατανομής F , γίνονται συγκρίσεις του πηλίκου αυτού με την καθορισμένη τιμή $F_{v1,v2, \alpha}$, όπου $v1$ και $v2$ είναι βαθμοί ελευθερίας και $1-\alpha$ είναι το επίπεδο σημαντικότητας το οποίο μας ενδιαφέρει. Σχεδόν παρόμοια με τις κατανομές t , το α αποτελεί το εμβαδό στα δεξιά της αντίστοιχης τιμής $F_{v1,v2,\alpha}$ και κάτω από την «ουρά» της εν λόγω καμπύλης της κατανομής F .

Υπάρχουν αναλύσεις διασποράς ενός παράγοντα (one-way ή one-factor) και άνω του ενός παράγοντα (two-way / factor ή multi-factor). Το παράδειγμα στη συνέχεια αφορά ένα παράγοντα, δηλ. γίνεται διερεύνηση μεταξύ διαφορετικών «διεργασιών» ενός μόνο παράγοντα (εργαστήριο).

Το αρνητικό της ανάλυσης διασποράς είναι ότι δεν μας λέει ποιά διεργασία είναι διαφορετική από την άλλη. Δηλαδή, αν συγκρίνουμε 3 διεργασίες A, B, Γ , που είναι διαφορετικές, η ANOVA μας λέει απλά ότι δεν ισχύει η ισότητα $A=B=\Gamma$. Αλλά δεν μας λέει – τουλάχιστον άμεσα – αν $A>B$ ή αν $B<\Gamma=A$ ή $\Gamma<A<B$ κ.λ.π. Για να το μελετήσουμε αυτό, τότε πρέπει να κάνουμε τα τεστ συγκρίσεων των μέσων όρων με την κατανομή t , ανά ζεύγη ή με ένα καθορισμένο control, δηλ. τα τεστ Tukey και Dunnett αντίστοιχα, όπως αυτά περιγράφηκαν παραπάνω.

Οι σχέσεις που χρησιμοποιούμε κατά την ανάλυση διασποράς αναλύονται στη συνέχεια. Θεωρούμε ότι k είναι ο μέγιστος αριθμός των διεργασιών, n_i ο αριθμός των μετρήσεων ανά διεργασία, $i=1,2,3,\dots,n_i$, ο αριθμός των μετρήσεων ανά διεργασία και $t=1,2,3,\dots,k$, οι διαφορετικές διεργασίες. Η διασπορά λοιπόν εντός μίας διεργασίας είναι:

$$s_t^2 = \frac{\sum_{i=1}^{n_t} (y_{ti} - \bar{y}_t)^2}{n_t - 1}$$

Η ομογενοποιημένη διασπορά s_w^2 εντός όλων των διεργασιών λαμβάνει υπόψη της τους βαθμούς ελευθερίας ανά διεργασία και υπολογίζεται ως εξής:

$$s_w^2 = \frac{\sum_{t=1}^k (n_t - 1) s_t^2}{\sum_{t=1}^k (n_t - 1)}$$

Η ομογενοποιημένη διασπορά μεταξύ των διεργασιών (s_b^2) εισάγει τον υπολογισμό του «μεγάλου» μέσου όρου, \bar{y} , που υπολογίζεται από το άθροισμα όλων των μεμονωμένων μετρήσεων των διεργασιών διαιρούμενου με το συνολικό αριθμό των μετρήσεων αυτών,

$$N. \text{ Δηλαδή, } \bar{y} = \frac{\sum_{i=1}^N n_{ti}}{N}$$

$$\text{Τελικά, } s_b^2 = \frac{\sum_{t=1}^k n_t (\bar{y}_t - \bar{y})^2}{k - 1}$$

Είναι σημαντικό λοιπόν να τονιστεί ότι στη μεταξύ των διεργασιών διασπορά, το μεγαλύτερο βάρος έχει η διεργασία με το μεγαλύτερο αριθμό μετρήσεων.

Ο λόγος $F = \frac{s_b^2}{s_w^2}$ κατανέμεται βάσει της κατανομής F (δες προηγούμενο κεφάλαιο).

Συγκρίνω τελικά το λόγο F με την σταθερή τιμή $F_{vb,vw,\alpha}$, όπως αυτή προκύπτει από σχετικούς πίνακες ή σχετικά στατιστικά προγράμματα. Οι τιμές v_b , v_w είναι αντίστοιχα οι βαθμοί ελευθερίας των διεργασιών, δηλ. $k-1$, και οι βαθμοί ελευθερίας συνολικά των εντός των διεργασιών μετρήσεων, δηλ. $N-k$. Η τιμή α εκφράζει το εμβαδόν κάτω από την ουρά της καμπύλης F και στα δεξιά της εν λόγω τιμής. Συνεπώς όταν $F > F_{vb,vw,\alpha}$, τότε με βεβαιότητα $1-\alpha$ (συνήθως 95%), η τιμή s_b^2 είναι μεγαλύτερη του s_w^2 και συνεπώς οι διεργασίες είναι ανόμοιες. Στην αντίθετη περίπτωση, τα δείγματα θεωρείται ότι όλα προέρχονται από τον ίδιο πληθυσμό.

Παράδειγμα

Έστω ότι 5 εργαστήρια μετράνε υδάτινο διάλυμα με γνωστή συγκέντρωση μολύβδου σε ppb. Δεν πραγματοποιείται ο ίδιος αριθμός επαναληπτικών μετρήσεων ανά εργαστήριο. Οι μετρήσεις των εργαστηρίων φαίνονται στη συνέχεια.

Εργ. 1	Εργ. 2	Εργ. 3	Εργ. 4	Εργ.5
3.4	4.5	5.3	3.2	3.3
3.0	3.7	4.7	3.4	2.4
3.4	3.8	3.6	3.1	2.7
5.0	3.9	5.0	3.0	3.2
5.1	4.3	3.6	3.9	3.3
5.5	3.9	4.5	2.0	2.9
5.4	4.1	4.6	1.9	4.4
4.2	4.0	5.3	2.7	
3.8	3.0	3.9		
4.2		4.1		
$n_t = 10$	9	10	8	7
$v_t = 9$	8	9	7	6
$\bar{y}_t = 4.30$	3.91	4.46	2.90	3.17
$s_t^2 = 0.82$	0.179	0.411	0.463	0.406
$\bar{y} = 3.823$				
$N = 44$				
$k = 5$				

Με βάση τις παραπάνω σχέσεις, υπολογίζεται η «μέση» διασπορά εντός των διεργασιών, που είναι:

$$s_w^2 = \frac{9 \cdot 0.82 + 8 \cdot 0.179 + \dots + 6 \cdot 0.406}{44 - 5} = \frac{18.17}{39} = 0.466$$

Υπενθυμίζεται ότι οι μονάδες της διασποράς είναι οι μονάδες της μέσης τιμής υψωμένες στο τετράγωνο.

Στη συνέχεια υπολογίζω τη διασπορά μεταξύ των διεργασιών, όπου λαμβάνεται υπόψη ο «μεγάλος» μέσος όρος \bar{y} , ήτοι:

$$s_b^2 = \frac{10 \cdot (4.30 - 3.823)^2 + 9 \cdot (3.91 - 3.823)^2 + \dots + 7 \cdot (3.17 - 3.823)^2}{5 - 1} = \frac{16.190}{4} = 4.047$$

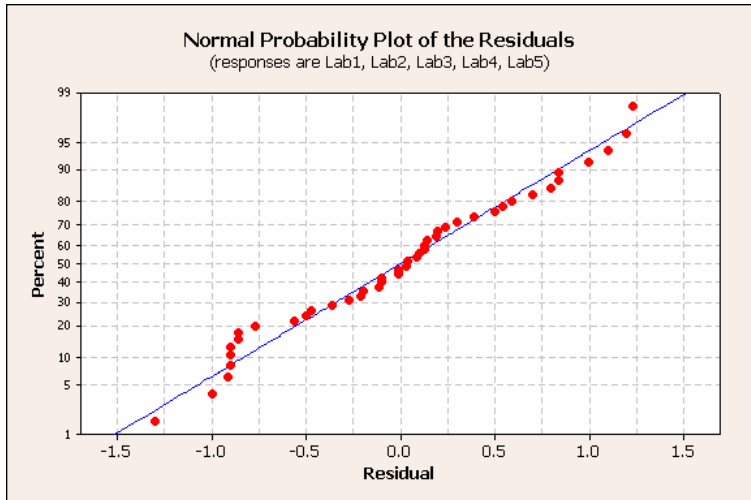
Ο λόγος F είναι τελικά $s_b^2/s_w^2 = 4.047/0.466 = 8.69$

Ο λόγος F αυτός θα πρέπει να συγκριθεί με την τιμή $F_{v_b, v_w, \alpha}$ για $\alpha=0.05$ και με $v_b=5-1$, $v_w=N-k$. Άρα $F_{4,39,0.05} = 2.62$. Προφανώς, ισχύει ότι $F > F_{4,39,0.05}$, άρα – με βεβαιότητα 95% - απορρίπτεται η μηδενική υπόθεση ότι οι διεργασίες είναι όμοιες. Τα εργαστήρια εξάγουν διαφορετικά αποτελέσματα, αφού τα δείγματα των διαφορετικών εργαστηρίων – που δεν γνωρίζουμε όμως ποιών – ανήκουν σε διαφορετικούς πληθυσμούς. Επίσης, από τους σχετικούς πίνακες, μπορεί κάποιος/α να υπολογίσει την τιμή P που αντιστοιχεί στο

υπολογισμένο λόγο F, δηλ. στο 8.69. Αν η τιμή αυτή είναι μικρότερη του 5%, τότε ομοίως μπορεί να εξαχθεί το συμπέρασμα ότι οι μετρήσεις δεν προέρχονται από τον ίδιο πληθυσμό και συνεπώς οι διεργασίες είναι ανόμοιες μεταξύ τους.

Η παραπάνω προσέγγιση δεν μας λέει αν το εργαστήριο Γ διαφέρει από το Β ή/και το Α. Πιθανά έχουμε ενδείξεις από τη σύγκριση των μέσων τιμών, όπως αναγράφονται στον πίνακα παραπάνω. Δηλαδή, μπορούμε να πούμε ότι το εργαστήριο 4 διαφέρει από το 3 και εξάγει μικρότερες τιμές. Δεν μπορούμε να πούμε, όμως, αν οι μετρήσεις των εργαστηρίων 1 και 3 προέρχονται από τον ίδιο πληθυσμό. Επίσης, δεν γνωρίζουμε την αξιοπιστία των εργαστηρίων αφού – παρόλο που όλες οι μετρήσεις έγιναν σε δείγματα με γνωστή συγκέντρωση μολύβδου – δεν γνωρίζουμε τη συγκέντρωση αυτή.

Απαραίτητα διαγνωστικά γραφήματα στην περίπτωση της ανάλυσης διασποράς είναι τα γραφήματα που αφορούν στα σφάλματα ή υπολειμματικά σφάλματα (errors). Εδώ προσοχή θέλει η διευκρίνιση ότι ως σφάλματα εννοούνται οι διαφορές των μεμονωμένων μετρήσεων των εργαστηρίων από αντίστοιχο μέσο όρο του κάθε εργαστηρίου (ή της κάθε διεργασίας γενικότερα). Τα σφάλματα λοιπόν δεν αντιστοιχούν με τη διαφορά από τον «**μεγάλο**» μέσο όρο. Τα N σφάλματα αυτά πρέπει να κατανέμονται κανονικά με βάση την μέση τιμή 0 και διασπορά σ^2 , να είναι ανεξάρτητα μεταξύ τους και να μην παρατηρούνται κάποιες τάσεις όταν τα σφάλματα αυτά τίθενται σε γράφημα συναρτήσεως των μεμονωμένων μετρήσεων ή συναρτήσεως του χρόνου της δειγματοληψίας. Στην περίπτωση των παραπάνω μετρήσεων, η κανονικότητα των σφαλμάτων επιβεβαιώνεται από το παρακάτω γράφημα πιθανότητας (όπως εξάγεται από το MINITAB 14), αφού όλα



τα υπολειμματικά σφάλματα φαίνεται ότι «πέφτουν» πάνω σε μία ευθεία. Επίσης περίπου στο 50% των τιμών (τιμή του κατακόρυφου άξονα) αντιστοιχεί η τιμή 0, δηλ. η μέση τιμή των σφαλμάτων αυτών είναι το 0). Αν και κάποια αρνητικά σφάλματα φαίνεται ότι απέχουν λίγο από την ευθεία, συνολικά δεν επηρεάζεται η κανονικότητά τους.

Το MINITAB εξάγει – πέραν των υπολειμματικών σφαλμάτων – και τις τιμές που προβλέπονται από το μοντέλο (fits) για κάθε μέτρηση του εργαστηρίου. Εδώ, αν και καταρχάς δημιουργείται το ερώτημα για ποιές προβλεπόμενες τιμές μιλάμε, πολύ απλά εννοείται ότι μιλάμε για τις ίδιες τις μέσες τιμές ανά εργαστήριο. Άρα η προβλεπόμενη (fit) τιμή της 1^{ης} μέτρησης των εργαστηρίων 1 και 2, αντίστοιχα, είναι οι 4.30 και 3.91. Η προβλεπόμενη τιμή της 2^{ης} μέτρησης των εργαστηρίων 1 & 2 είναι όμοια οι 4.30 και 3.91.

Σε κάθε περίπτωση, η σχέση μεταξύ των εργαστηρίων μπορεί να εξαχθεί μόνο με τεστ t σε ζεύγη, όπως είναι το τεστ Tukey.

Σε κλασικά στατιστικά προγράμματα, τα αποτελέσματα της ανάλυσης διασποράς με έναν παράγοντα παρουσιάζονται στον κλασικό πλέον πίνακα ANOVA, που αναλύεται παρακάτω. Τα αποτελέσματα παραπάνω αναλύθηκαν με το πρόγραμμα MINITAB 14 που τα παρουσιάζει στον εξής πίνακα ANOVA.

Πίνακας εξαγωγής αποτελεσμάτων ανάλυσης διασποράς (MINITAB 14)

Source	DF	SS	MS	F	P
Factor	4	16.190	4.048	8.69	0.000
Error	39	18.167	0.466		
Total	43	34.357			

S = 0.6825 R-Sq = 47.12% R-Sq(adj) = 41.70%

Ο όρος SS σημαίνει το άθροισμα των τετραγώνων και ισούται με:

- τον αριθμητή του κλάσματος υπολογισμού της διασποράς μεταξύ των διεργασιών (s^2_b) για τον υπό μελέτη παράγοντα (factor), που εδώ είναι το εργαστήριο.
- τον αριθμητή του κλάσματος υπολογισμού της διασποράς εντός των διεργασιών (s^2_w), που ταυτίζεται με τα τυχαία πειραματικά σφάλματα (errors).
- το άθροισμα των παραπάνω, που ουσιαστικά ταυτίζεται με το άθροισμα των τετραγώνων των διαφορών των μεμονωμένων τιμών από τον μεγάλο μέσο όρο (total).

Ο όρος MS σημαίνει το μέσο άθροισμα των τετραγώνων και ισούται με:

- Το πηλίκο του SS με τους αντίστοιχους βαθμούς ελευθερίας (DF), ξεχωριστά για μεταξύ των διεργασιών (4), εντός των διεργασιών (39) και το σύνολο των μετρήσεων (43=44-1). Τα MS δηλαδή είναι τιμές s^2_b , s^2_w .

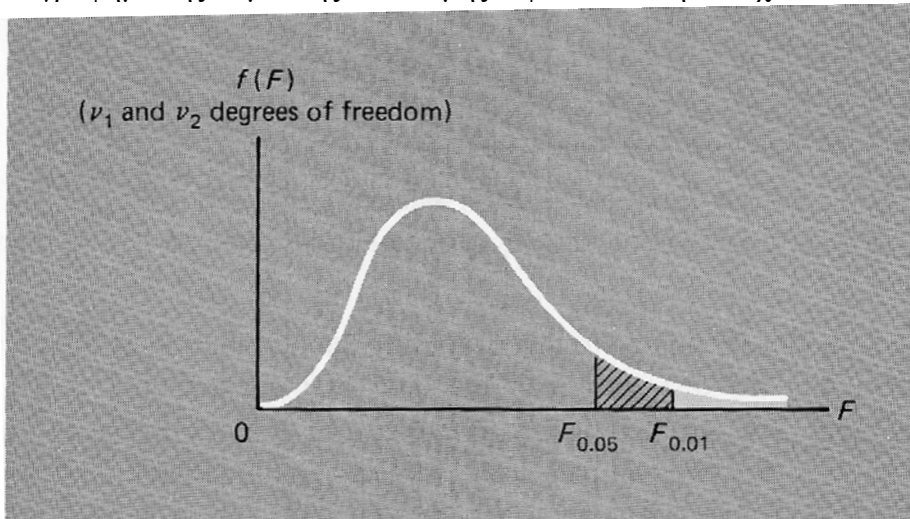
Ο λόγος F είναι το πηλίκο s^2_b/s^2_w , όπως ορίστηκε και παραπάνω, ενώ η τιμή P αντιστοιχεί στην πιθανότητα εμφάνισης τιμών ίσων ή μεγαλύτερων της τιμής 8.69 σε μία κατανομή $F_{4,39}$. Αν η τιμή αυτή είναι μικρότερη του 5% (0.05), αφού 95% είναι το επίπεδο εμπιστοσύνης που εμείς καθορίσαμε, τότε άμεσα μπορούμε να πούμε ότι ο λόγος F είναι «πολύ» μεγάλος και συνεπώς η διασπορά μεταξύ των διεργασιών πολύ μεγαλύτερη της διασποράς εντός των διεργασιών, που είναι αποτέλεσμα πειραματικού σφάλματος.

Η τιμή S του πίνακα του MINITAB είναι η κοινή (ομογενοποιημένη) τυπική απόκλιση, όπως είχε υπολογιστεί και σε προηγούμενο κεφάλαιο. Η δε τιμή R^2 και adj. R^2 είναι ο συντελεστής παλινδρόμησης και διορθωμένος συντ. παλινδρόμησης²², που θα αναλυθούν σε επόμενο κεφάλαιο²³.

²² Αρτιότερος όρος είναι συντελεστής απόφασης

²³ Το R^2 είναι το ποσοστό της συνολικής διασποράς που δεν ανήκει στα σφάλματα (errors) και προφανώς είναι το ποσοστό της διασποράς που προέρχεται από το μοντέλο. Δηλ. $R^2 = SS_{\text{factor}} / \text{Συνολικό SS} = 16.190 / 34.357 = 0.4712$. Το $1-R^2$ είναι συνεπώς το ποσοστό της συνολικής διασποράς που ανήκει στα τυχαία σφάλματα. Σε ένα «καλό» μοντέλο, γενικά επιθυμείται τα σφάλματα να είναι όσο το δυνατό μικρότερα και κατά συνέπεια το $1-R^2$ όσο το δυνατό μικρότερο. Συνεπώς, σε ένα «καλό» μοντέλο, το R^2 επιθυμείται να

Ένα τυπικό γράφημα της καμπύλης κατανομής F φαίνεται στη συνέχεια.



Σχήμα 8.1. Κατανομή F για ν_1, ν_2 βαθμούς ελευθερίας

Σημειώνεται ότι το MINITAB – όπως και τα περισσότερα στατιστικά πακέτα – εξάγουν και τα διαστήματα εμπιστοσύνης ανά διεργασία (στην περίπτωση μας εργαστήριο). Από τα διαστήματα αυτά, μπορεί να εξαχθεί και το συμπέρασμα για το ποιά διεργασία μετράει μεγαλύτερες ή μικρότερες τιμές σε σχέση με κάποια άλλη.

Το αποτελέσματα του MINITAB για το παραπάνω παράδειγμα είναι ως εξής:

Level	N	Mean	StDev
Lab1	10	4.3000	0.9043
Lab2	9	3.9111	0.4226
Lab3	10	4.4600	0.6415
Lab4	8	2.9000	0.6803
Lab5	7	3.1714	0.6370

Individual 95% CIs For Mean Based on Pooled StDev			
Lab1	(-----*-----)		
Lab2	(-----*-----)		
Lab3	(-----*-----)		
Lab4	(-----*-----)		
Lab5	(-----*-----)		

2.80 3.50 4.20 4.90

Pooled StDev = 0.6825

Από το παραπάνω, φαίνεται ότι υπάρχει σαφής διαφορά μεταξύ των εργαστηρίων 4 & 5 σε σχέση με τα 1,2 & 3. Σημειώνεται ότι τα διαστήματα εμπιστοσύνης εξήχθησαν συναρτήσει μίας κοινής (ομογενοποιημένης) διασποράς s^2_{pool} , που υπολογίστηκε με τη σχέση που παρουσιάστηκε σε προηγούμενο κεφάλαιο. Η κοινή τυπική απόκλιση και διασπορά ισούνται, αντίστοιχα, με 0.6825 και 0.466. Υπενθυμίζεται ότι βασική προϋπόθεση για τον υπολογισμό μιας κοινής διασποράς είναι να είναι οι επιμέρους διασπορές περίπου του ίδιου μεγέθους.

βρίσκεται όσο κοντύτερα στο 1. Παρόλα αυτά, απαιτείται προσοχή διότι μόνο η τιμή R^2 δεν αποτελεί ένδειξη ενός «καλού μοντέλου» (δες κεφ. 13 & 14).

Παράδειγμα

Ένα εργοστάσιο σε μονάδα επεξεργασίας λυμάτων λαμβάνει δείγματα από 3 παρακείμενες μονάδες επεξεργασίας λυμάτων που δεν διαθέτουν τα δικά τους εργοστάσια. Οι 3 μονάδες έχουν διαφορετικές τεχνικές επεξεργασίας των λυμάτων και συγκεκριμένα: Μονάδα 1. Παρατεταμένος αερισμός, Μονάδα 2. Βιολογικό φίλτρο, Μονάδα 3. Τυπικό σύστημα ενεργής ύλης χωρίς παρατεταμένο αερισμό. Οι μετρήσεις σε τυχαία δείγματα που έλαβε το εργοστάσιο σε διάστημα του Δεκεμβρίου 2005 (σε BOD₅ σε mg/L) στην έξοδο της κάθε μονάδας αναφέρονται στη συνέχεια. Μπορείτε να συμπεράνετε αν και τα 3 εργοστάσια επιτυγχάνουν την ίδια επεξεργασία, τουλάχιστον για το συγκεκριμένο μήνα, με χρήση ανάλυσης διασποράς? Παρουσιάστε όλα τα αποτελέσματα στον κλασικό πίνακα ANOVA.

Μονάδα 1	Μονάδα 2	Μονάδα 3
25,0	26,0	15,0
23,0	20,0	16,0
18,0	30,0	25,0
12,0		28,0
22,0		

Οι μέσες τιμές και διασπορές ανά μονάδα είναι οι εξής:

$$\bar{y}_1 = 20,0 \text{ mg/L}, s_1^2 = 26,5 \text{ mg/L}, n_1 = 5, v_1 = 4$$

$$\bar{y}_2 = 25,3 \text{ mg/L}, s_2^2 = 25,3 \text{ mg/L}, n_2 = 3, v_2 = 2$$

$$\bar{y}_3 = 21,0 \text{ mg/L}, s_3^2 = 42,0 \text{ mg/L}, n_3 = 4, v_3 = 3$$

Ο μεγάλος μέσος όρος είναι ίσος με: $\bar{y} = 21,7 \text{ mg/L}$, ενώ το $N = 5 + 3 + 4 = 12$, $k = 3$

Με βάση τις προαναφερόμενες σχέσεις, η διασπορά εντός των μονάδων είναι:

$$s_w^2 = \frac{4 \cdot 26,5 + 2 \cdot 25,3 + 3 \cdot 42,0}{4 + 2 + 3} = \frac{282,7}{9} = 31,4 \text{ (mg/L)}^2$$

Υπενθυμίζεται ότι οι μονάδες της διασποράς είναι οι μονάδες της μέσης τιμής υψωμένες στο τετράγωνο.

Στη συνέχεια υπολογίζω τη διασπορά μεταξύ των διεργασιών, όπου λαμβάνεται υπόψη ο «μεγάλος» μέσος όρος \bar{y} , ήτοι:

$$s_b^2 = \frac{5 \cdot (20,0 - 21,7)^2 + 3 \cdot (25,3 - 21,7)^2 + \dots + 4 \cdot (21,0 - 21,7)^2}{3 - 1} = \frac{56,0}{2} = 28,0$$

Ο λόγος F είναι: $F = \frac{s_b^2}{s_w^2} = \frac{28,0}{31,4} = 0,892$. Η τιμή βρίσκεται πολύ κοντά στη μονάδα, που

άμεσα οδηγεί στο συμπέρασμα ότι οι 2 διασπορές (μεταξύ των ΜΕΛ και εντός των μονάδων) είναι σχεδόν ίσες. Σε κάθε περίπτωση, η σύγκριση πρέπει να γίνει με την τιμή $F_{2,9,5\%}$

Συγκρίνω την παραπάνω τιμή με την τιμή $F_{2,9,5\%} = 4,26$

Προφανώς, $F \ll F_{2,9,5\%}$ και μπορούμε να πούμε με βεβαιότητα 95% ότι δεν διαφέρουν μεταξύ τους οι αποδόσεις επεξεργασίας των μονάδων. Δηλαδή, και οι τρεις μονάδες επιτυγχάνουν την ίδια συγκέντρωση των επεξεργασμένων λυμάτων.

Στη συνέχεια φαίνεται ο κλασσικός πίνακας ANOVA, όπως εξάγεται από τα κλασσικά στατιστικά λογισμικά (εδώ από το MINITAB 14).

Πίνακας ANOVA

Source	DF	SS	MS	F	P
Factor	2	56,0	28,0	0,89	0,443
Error	9	282,7	31,4		
Total	11	338,7			

Όπως φαίνεται, οι τιμές SS (Sum of Squares) των παραγόντων και των σφαλμάτων είναι αντίστοιχα οι αριθμητές των κλασμάτων που ορίζουν τα s_b^2 και s_w^2 , αντίστοιχα. Το άθροισμά τους είναι το συνολικό άθροισμα των τετραγώνων των σφαλμάτων των μεμονωμένων μετρήσεων (δηλαδή το άθροισμα των διαφορών – υψωμένων στο τετράγωνο - των μεμονωμένων 12 μετρήσεων από τον μεγάλο μέσο όρο).

Οι τιμές MS (Mean of Squares) είναι πρακτικά οι μέσες τιμές των αθροισμάτων των τετραγώνων των σφαλμάτων, δηλαδή απευθείας οι τιμές s_b^2 (28,0) και s_w^2 (31,4). Η τιμή

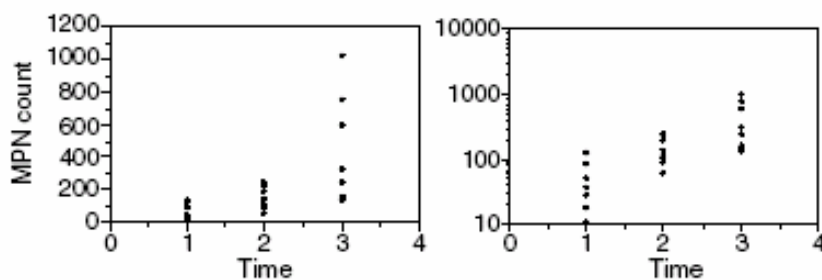
F είναι προφανώς το κλάσμα $\frac{s_b^2}{s_w^2}$, όπως υπολογίστηκε παραπάνω. Η τιμή P είναι η τιμή

πιθανότητας που αντιστοιχεί στην υπολογισμένη τιμή F (ήτοι στην 0,89). Δηλαδή, πρακτικά, είναι το εμβαδόν κάτω από την καμπύλη της κατανομής F (για 2 και 9 βαθμούς ελευθερίας) και δεξιά της τιμής F 0,89. Αν η τιμή αυτή F ήταν *μικρότερη* από 0,05 (5%), που είναι το όριο που εμείς θέτουμε, τότε θα συμπεραίναμε ότι οι ΜΕΛ επιτυγχάνουν διαφορετική επεξεργασία.

9. Μετατροπή δεδομένων

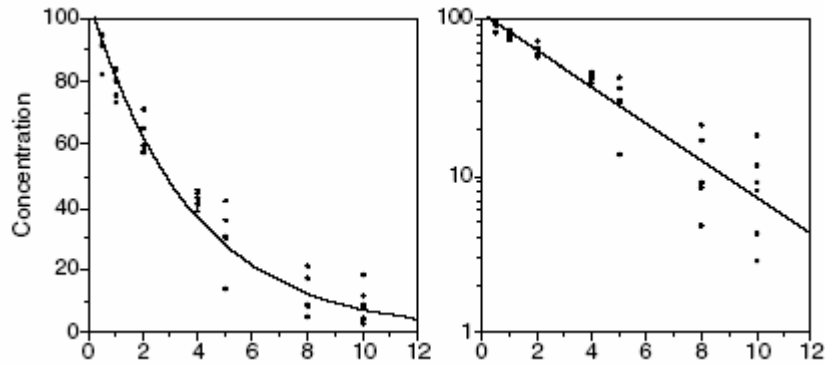
Πολλές φορές, όταν μας δοθούν δεδομένα ενός πειράματος ή απλών μετρήσεων είναι αδύνατο ή πολύ δύσκολο να τα αναλύσουμε στατιστικά στην κλίμακα στην οποία δίνονται. Δηλαδή, για δεδομένα y , έχει αποδειχθεί ότι – υπό συγκεκριμένες προϋποθέσεις – είναι βολικότερο και περισσότερο κατανοητό αν κάνουμε τις στατιστικές μας αναλύσεις, αφού μετατρέψουμε μαθηματικά τις τιμές y σε άλλες τιμές όπως π.χ. y^2 , $y^{1/2}$, $\log(y)$, $\ln(y)$, $1/y$ κ.λ.π. Τα τελικά αποτελέσματα της ανάλυσης μας (π.χ. μέσες τιμές, διασπορές, διάστημα εμπιστοσύνης) μπορούμε να τα μετατρέψουμε πίσω στην αρχική κλίμακα, ώστε να έχουμε τα τελικά αποτελέσματα της ανάλυσης μας στις «κατανοητές» αρχικές τιμές y .

Ένας από τους πιο συνήθεις λόγους να κάνουμε μετατροπές των αρχικών δεδομένων είναι όταν παρατηρείται μεταβολή της διασποράς των δεδομένων κατ' αναλογία με τον μέσο όρο. Τι σημαίνει αυτό? Αν δούμε το παρακάτω σχήμα, βλέπουμε στο αριστερό διάγραμμα ότι μεταβάλλεται η διασπορά (πρακτικά το «άπλωμα») των επαναληπτικών τιμών με το πέρασμα του χρόνου. Οι τιμές είναι μετρήσεις αποικιών, που διαφέρουν μεταξύ τους μάλιστα κατά τάξεις μεγέθους. Δηλαδή στο χρόνο 1 έχουμε τιμές έως και 100 MPN, ενώ στο χρόνο 3 έχουμε τιμές έως και 1000 MPN. Οι μεγάλες διαφοροποιήσεις τιμών προς στατιστική ανάλυση είναι ένας ακόμα λόγος για μετατροπή των δεδομένων. Η μετατροπή που προτιμάται στην περίπτωση που μόλις περιγράφηκε είναι η μετατροπή των τιμών σε λογαριθμική κλίμακα. Αυτό φαίνεται στο δεξιό γράφημα, όπου το πρώτο που παρατηρούμε είναι ότι η διασπορά στις 3 χρονικές στιγμές έχει λίγο-πολύ σταθεροποιηθεί, παρόλο που ο μέσος όρος συνεχίζει να αυξάνεται από την 1 έως την 3^η χρονική στιγμή. Δηλαδή, μπορεί να διαφέρει ο μέσος όρος μεταξύ των 3 χρονικών στιγμών, αλλά η διασπορά είναι μάλλον ομοιόμορφη πλέον μεταξύ των στιγμών αυτών. Άρα η μετατροπή των δεδομένων πέτυχε, όσον αφορά τουλάχιστον την επίτευξη ομοιόμορφης διασποράς.



Σχήμα 9.1. Μετατροπή σε λογαριθμική κλίμακα δημιουργεί σταθερή διασπορά

Η μετατροπή των δεδομένων δεν δημιουργεί πάντα θετικό αποτέλεσμα. Υπάρχουν πολλές περιπτώσεις που η μετατροπή των δεδομένων μπορεί να φέρει «χειρότερο» αποτέλεσμα, όπως φαίνεται στο παρακάτω σχήμα.



Σχήμα 9.2. Η μετατροπή των αρχικών δεδομένων σε λογαριθμική κλίμακα δημιουργεί μη σταθερή διασπορά μεταξύ των χρονικών στιγμών που έγιναν οι μετρήσεις. Η μετατροπή δημιούργησε «χειρότερα» αποτελέσματα, αναφορικά τουλάχιστον με τη διασπορά.

Γενικά, τροποποιήσεις απαιτούνται όταν η διασπορά δεν είναι σταθερή για όλα τα y , όπως στις παρακάτω περιπτώσεις:

- ο Μετρήσεις στις οποίες επικρατούν συνεχείς διαλύσεις, και συνεπώς εισέρχεται πολλαπλασιαστικό λάθος
- ο Μετρήσεις με όργανα που μετράνε σε λογαριθμικές κλίμακες (οπότε οι χαμηλές τιμές είναι πιο ακριβείς από τις υψηλές)
- ο Μετρήσεις βιολογικών αποικιών, όπου οι τιμές μπορούν να διαφοροποιούνται και κατά τάξεις μεγέθους όπως παρουσιάστηκε στο σχήμα 9.1.

Γενικά τροποποιήσεις δεδομένων κάνουμε όταν θέλουμε να έχουμε:

1. τις διασπορές διαφόρων τιμών y περίπου ίσου μεγέθους (δηλαδή το «άπλωμα» των διαφορετικών τιμών y να γίνει περίπου όμοιο για όλα τα y).
2. την κατανομή των σφαλμάτων κανονική (η κανονικότητα αφορά στα σφάλματα από κάποιο εμπειρικό μοντέλο που περιγράφει τα δεδομένα, και όχι στα ίδια τα αρχικά δεδομένα)
3. τις επιδράσεις κάποιων διεργασιών προσθετικές, δηλ. να απαλείψουμε τις τυχόν αλληλοεπιδράσεις (η έννοια των αλληλοεπιδράσεων συζητείται σε επόμενο κεφάλαιο). Ειδικότερα, αυτό γίνεται όταν σε παραγοντικές αναλύσεις εξάγονται σημαντικές αλληλεπιδράσεις, χωρίς οι επιπτώσεις των κυρίων παραγόντων, που συνιστούν τις αλληλοεπιδράσεις αυτές να είναι στατιστικά σημαντικές.

Συνεπώς, κάνουμε τροποποιήσεις των αρχικών δεδομένων ώστε να πληρούνται κατά το δυνατό οι παραπάνω προϋποθέσεις. Οι τροποποιήσεις που γίνονται στα δεδομένα ξεκινούν από τον έλεγχο της σχέσης μεταξύ της μέσης τιμής και της διασποράς. Η διασπορά πρέπει γενικώς να παραμένει σταθερή και να μην μεταβάλλεται συναρτήσει των τιμών των δεδομένων, δηλαδή να μην εξαρτάται από τον μέσο όρο, όπως αυτός προκύπτει από τα διαφορετικά πειράματα. Εφόσον αυτό δεν γίνεται, και π.χ. παρατηρείται αύξηση της διασποράς με την αύξηση του μέσου όρου ή κάποια άλλη τάση της διασποράς, τότε προσπαθούμε να κάνουμε τροποποιήσεις ώστε να εξαλείψουμε τις παραπάνω «ανωμαλίες».

Οι κλασσικές οδηγίες που προτάθηκαν από τους Box et al., 1978 για τις τροποποιήσεις περιλαμβάνονται στον πίνακα στη συνέχεια.

Πίνακας 9.1. Προτεινόμενες τροποποιήσεις δεδομένων, που βασίζονται στη σχέση της τιμής σ, σ^2 με το μέσο όρο των δεδομένων, με στόχο να επιτευχθεί σταθερή διασπορά για όλες τις τιμές y

Condition		Replace y by
σ uniform over range of y		no transformation needed
$\sigma \propto \bar{y}^2$		$x = 1/y$
$\sigma \propto \bar{y}^{3/2}$		$x = 1/\sqrt{\bar{y}}$
$\sigma \propto \bar{y}$ ($\sigma^2 > \bar{y}$)	all $y > 0$	$x = \log(y)$
	some $y = 0$	$x = \log(y + c)$
$\sigma \propto \bar{y}^{1/2}$	all $y > 0$	$x = \sqrt{\bar{y}}$
	some $y < 0$	$x = \sqrt{\bar{y} + c}$
$\sigma^2 > \bar{y}$	$p =$ ratio or percentage	$x = \arcsin \sqrt{p}$

Source: Box, G. E. P., W. G. Hunter, and J. S. Hunter (1978). *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, New York, Wiley Interscience.

Σύμφωνα με τον παραπάνω πίνακα, όταν η διασπορά είναι σταθερή για όλες τις τιμές των δεδομένων y , τότε γενικά δεν απαιτείται μετατροπή. Μία χαρακτηριστική μετατροπή είναι όταν η διασπορά είναι ανάλογη του μέσου όρου (και ειδικότερα όταν η διασπορά είναι αρκετά μεγαλύτερη του μέσου όρου), όπου τότε λαμβάνεται ο λογάριθμος των δεδομένων. Η λογική των τροποποιήσεων είναι ότι μετά την μετατροπή γίνεται ανάλυση των δεδομένων και όλα τα αποτελέσματα που εξάγονται βρίσκονται στην τροποποιημένη κλίμακα (μέσοι όροι, διασπορές, διαστ. εμπιστοσύνης κ.λ.π.). Συνεπώς, στη συνέχεια τα δεδομένα επανατροποποιούνται στην αρχική τους κλίμακα.

Στον πίνακα 9.2, παρουσιάζονται 8 επαναληπτικές μετρήσεις (y) βακτηρίων σε 3 δειγματοληπτικούς σταθμούς. Παρατηρούμε 2 χαρακτηριστικά. Καταρχάς, η διασπορά s^2 είναι αρκετά μεγαλύτερη του μέσου όρου για κάθε σταθμό και το s αυξάνει αναλογικά με το \bar{y} . Σύμφωνα με τον πίνακα 9.1, απαιτείται λογαριθμική μετατροπή των δεδομένων, κάτι που γίνεται στις 3 δεξιές στήλες του ίδιου πίνακα.

Πίνακας 9.2. Παράδειγμα αναγκαιότητας λογαριθμικής μετατροπής δεδομένων

$y = \text{Bacteria}/100 \text{ mL}$			$x = \log_{10}(\text{Bacteria}/100 \text{ mL})$		
1	2	3	1	2	3
27	225	1020	1.431	2.352	3.009
11	99	136	1.041	1.996	2.134
48	41	317	1.681	1.613	2.501
36	60	161	1.556	1.778	2.207
120	190	130	2.079	2.279	2.114
85	240	601	1.929	2.380	2.779
18	90	760	1.255	1.954	2.889
130	112	240	2.144	2.049	2.380
$\bar{y} = 59.4$	132	420.6	$\bar{x} = 1.636$	2.050	2.502
$s_y^2 = 2156$	5771	111,886	$s_x^2 = 0.151$	0.076	0.124

Στο δεξιό τμήμα του πίνακα 9.2 φαίνονται τα τροποποιημένα δεδομένα, δηλ. $x = \log(y)$. Παρατηρούμε ότι η διασπορά είναι – λίγο, πολύ – σταθερή και δεν μεταβάλλεται με την αύξηση του μέσου όρου από το 1^ο στο 3^ο δείγμα.

Όμοια και το επόμενο παράδειγμα, όπου υπάρχουν 20 επαναλήψεις πλαγκτόν από 5 θαλάσσιους σταθμούς μέτρησης, παρατηρείται αύξηση των μέσων τιμών από το σταθμό 1 στο σταθμό 5 (β πίνακιο του 9.3). Συγχρόνως όμως παρατηρείται και αύξηση των αντίστοιχων διασπορών. Συνεπώς, δεν ισχύει η αρχή της ομοιόμορφης διασποράς και τα δεδομένα χρήζουν μετατροπής.

Πίνακας 9.3. Παράδειγμα αναγκαιότητας μετατροπής δεδομένων με χρήση της τετραγωνικής ρίζας

Station 1	0	2	1	0	0	1	1	0	1	1	0	2	1	0	0	2	3	0	1	1
Station 2	3	1	1	1	4	0	1	4	3	3	5	3	2	2	1	1	2	2	2	0
Station 3	6	1	5	7	4	1	6	5	3	3	5	3	4	3	8	4	2	2	4	2
Station 4	7	2	6	9	5	2	7	6	4	3	5	3	6	4	8	5	2	3	4	1
Station 5	12	7	10	15	9	6	13	11	8	7	10	8	11	8	14	9	6	7	9	5

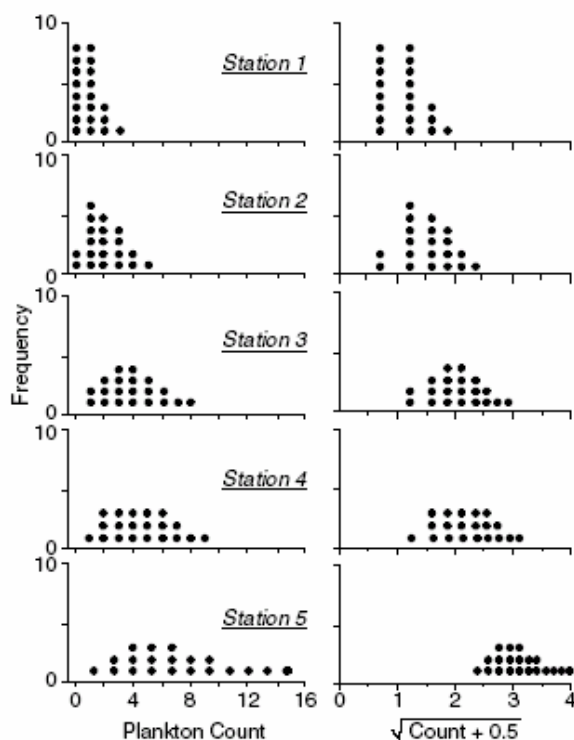
Source: Elliot, J. (1977). *Some Methods for the Statistical Analysis of Samples of Benthic Invertebrates*, 2nd ed., Ambleside, England, Freshwater Biological Association.

Station	1	2	4	5	6
Untransformed data	$\bar{y} = 0.85$	2.05	3.90	4.60	9.25
	$s_y^2 = 0.77$	1.84	3.67	4.78	7.57
Transformed $x = \sqrt{y+c}$	$\bar{x} = 1.10$	1.54	2.05	2.20	3.09
	$s_x^2 = 0.14$	0.20	0.22	0.22	0.19

Θα δοκιμαστεί η μετατροπή με την τετραγωνική ρίζα, ενώ θα μπορούσε να είχε δοκιμαστεί και η λογαριθμική μετατροπή όπως και παραπάνω. Λόγω των τιμών 0, θα πρέπει να προστεθεί και μία σταθερά c , αφού αποφεύγεται η χρήση της τετραγωνικής ρίζας για την τιμή 0. **Προσοχή:** Η πρόσθεση με κάποια σταθερά απαιτείται και σε αρνητικές ή μηδενικές τιμές, όταν δοκιμάζεται η λογαριθμική μετατροπή, αφού δεν ορίζεται ο λογάριθμος του 0 και αρνητικής τιμής.

Οι μέσες τιμές και οι διασπορές των τροποποιημένων αποτελεσμάτων φαίνονται στο κάτω μέρος του πίνακα 9.3 με χρήση της σταθεράς $c=0,5$. Είναι ξεκάθαρο ότι η διασπορά πλέον σταθεροποιείται, παρόλο που ο μέσος όρος αυξάνεται από το σταθμό 1 στον 5, όπως και πριν. Οι επιδράσεις της τροποποίησης φαίνονται επίσης στο Σχήμα 9.1.

Οι διασπορές τελικά είναι ομοιόμορφες και τα τροποποιημένα δεδομένα κατανέμονται «περισσότερα» κανονικά σε σχέση με τα αρχικά.



Σχήμα 9.3. Κατανομές συχνότητας ανά σταθμό για τα δεδομένα πλαγκτόν του πίνακα 9.3. Παρατηρείται ομοιόμορφη διασπορά στα τροποποιημένα δεδομένα στη δεξιά στήλη σε αντίθεση με τη μη σταθερή διασπορά στην αριστερή στήλη.

Το παράδειγμα στη συνέχεια δείχνει την επαναφορά των τροποποιημένων δεδομένων στην αρχική κλίμακα μετά το πέρας της στατιστικής ανάλυσης.

Ένα δείγμα $n=5$ παρατηρήσεων $[90, 18,73,190,85]$ δίνει μέση τιμή $\bar{y}=91,2$ και διασπορά $s^2 = 3873$. Ξεκάθαρα, $s^2 \gg \bar{y}$ και απαιτείται μετατροπή των πρωτογενών τιμών.

Δοκιμάζεται η λογαριθμική μετατροπή, που τελικώς δίνει $x=\log(y)$, ήτοι $[1.95, 1.26, 1.86, 2.28, 1.93]$ με μέση τιμή $\bar{x} = 1.86$ και διασπορά $s^2_x=0.14$. Η τιμή t για 4 βαθμούς ελευθερίας και $\alpha/2=0.025$ είναι 2.776. Τελικώς, το 95% διάστημα εμπιστοσύνης για την τροποποιημένη στη λογαριθμική κλίμακα μέση τιμή είναι:

$$\eta_x = 1.86 \pm t_{4,0.025} \cdot \sqrt{0.14/5} = 1.86 \pm 2.776 \cdot 0.17 = 1.86 \pm 0.46 \text{ ή } (1.39 \text{ έως } 2.32)$$

Το τελικό διάστημα εμπιστοσύνης μπορεί να «επανέρθει» στην αρχική κλίμακα παίρνοντας απλά τον αντίστροφο λογάριθμο των τιμών του διαστήματος εμπιστοσύνης. Δηλαδή, $\text{antilog}(1.39) < \bar{y} < \text{antilog}(2.32)$ ή τελικά **24.78** < \bar{y} < **208.2**. Προφανώς το τελικό διάστημα εμπιστοσύνης – στις πραγματικές τιμές – δεν είναι συμμετρικό. Αυτό είναι χαρακτηριστικό της επαναφοράς τροποποιημένων τιμών στις πρωτότυπες τιμές.

10. Πλήρη παραγοντικά πειράματα

Τα πλήρη παραγοντικά πειράματα αποτελούν τη βάση της σχεδίασης πειραμάτων σε όλες τις επιστημονικές κατευθύνσεις. Έχουν μεγάλη αξία διότι:

- Εξάγονται συμπεράσματα για το μέγεθος των επιπτώσεων (effect) των κυρίων παραγόντων (κύριες επιπτώσεις) και των αλληλοεπιδράσεων μεταξύ των παραγόντων
- Επικουρούν στη δημιουργία ενός εμπειρικού μοντέλου που περιγράφει την επιρροή της απόκρισης (response) από τους παράγοντες.
- Επιτρέπουν την μελέτη μεγάλου αριθμού μεταβλητών (η παραγόντων) με διεξαγωγή σχετικά μικρού αριθμού πειραμάτων.
- Δεν απαιτούν πολύπλοκους υπολογισμούς κατά την ανάλυση των δεδομένων.

Οι k ανεξάρτητες μεταβλητές, η επίδραση των οποίων σε κάποια εξαρτημένη μεταβλητή (απόκριση) πρόκειται να μελετηθεί, ονομάζονται **παράγοντες** (factors). Ένα πείραμα με k παράγοντες, με τον κάθε παράγοντα σε 2 επίπεδα, ονομάζεται παραγοντικό πείραμα 2 επιπέδων και συμβολίζεται ως 2^k . Ένα πλήρες παραγοντικό πείραμα 2^k απαιτεί να υλοποιηθούν 2^k πειράματα (χωρίς επανάληψη), που αποτελούν όλους τους συνδυασμούς των k παραγόντων μεταξύ τους σε 2 επίπεδα. Τα υψηλά και χαμηλά επίπεδα των παραγόντων συνηθίζεται να συμβολίζονται ως +1 και -1, αντίστοιχα. Επίσης, έχει προταθεί και ο συμβολισμός 1 & 0 αντίστοιχα. Οι παράγοντες (δηλ. οι ανεξάρτητες μεταβλητές) μπορεί να είναι συνεχείς (π.χ. πίεση, θερμοκρασία κ.λ.π.) ή ασυνεχείς (π.χ. παρουσία ή όχι κάποιου υλικού, χρήση ή όχι κάποιας συνθήκης κ.λ.π.). Η μεταβλητή απόκριση (response variable), που αποτελεί την εξαρτημένη μεταβλητή που μετράμε το πείραμα, και που θα ονομάζεται πλέον ως **απόκριση**, συνήθως συμβολίζεται με y .

Υπάρχουν και παραγοντικά πειράματα που απαιτούν λιγότερο των 2^k πειραμάτων ώστε να μελετηθούν k παράγοντες. Αυτά ονομάζονται κλασματικά (ή μερικά) παραγοντικά πειράματα και θα αναλυθούν στη συνέχεια. Είναι μάλλον περισσότερο χρήσιμα από τα πλήρη παραγοντικά πειράματα, αφού απαιτούν ακόμα μικρότερο αριθμό πειραμάτων για την μελέτη k παραγόντων.

Στα παραγοντικά πειράματα μιλάμε για τη *μήτρα σχεδίασης* και τη *μήτρα ανάλυσης* (ή μοντελοποίησης). Η μήτρα σχεδίασης ουσιαστικά ορίζει τις θέσεις που θα θέσουμε τον κάθε παράγοντα σε κάθε πείραμα. Η τυπική σειρά ενός 2^3 και ενός 2^4 πλήρους παραγοντικού πειράματος φαίνεται στον πίνακα στη συνέχεια.

Ο τρόπος που αναπτύσσεται η μήτρα σχεδίασης είναι σχετικά απλός. Αφού θέσουμε τη σειρά των πειραμάτων από 1 έως 2^k τότε ορίζουμε οριζόντια τους παράγοντες 1 έως k . Στην τελευταία στήλη, που αντιστοιχεί στον παράγοντα k , θέτουμε τα πρώτα μισά πειράματα $2^{k-1}/2$ σε - επίπεδο και τα υπόλοιπα $2^{k-1}/2$ πειράματα σε + επίπεδο. Στον προηγούμενο παράγοντα, $k-1$, θέτουμε το πρώτο ένα τέταρτο των πειραμάτων $2^{k-2}/4$ σε - επίπεδο, το επόμενο τέταρτο σε + επίπεδο, το επόμενο τέταρτο σε - κ.ο.κ. Συνεχίζουμε έτσι, δηλ. στο προηγούμενο παράγοντα λαμβάνουμε το $1/8$ των πειραμάτων όπου διαδοχικά αλλάζουμε πρόσημα, ενώ στον 1 παράγοντα αλλάζουμε πρόσημο από - σε +

ανα πείραμα. Η παραπάνω λογική δημιουργεί την τυποποιημένη σειρά των πειραμάτων, όπως φαίνεται στον πίνακα 10.1 παραδειγματικά για 2^3 και 2^4 σχεδιασμούς.

Πίνακας 10.1. Μήτρες σχεδίασης 2^3 και 2^4 πλήρων παραγοντικών πειραμάτων

Αριθμός πειραμάτων 2^3	Θέση παράγοντα			Αριθμός πειραμάτων 2^4	Θέση παράγοντα			
	1	2	3		1	2	3	4
1	-	-	-	1	-	-	-	-
2	+	-	-	2	+	-	-	-
3	-	+	-	3	-	+	-	-
4	+	+	-	4	+	+	-	-
5	-	-	+	5	-	-	+	-
6	+	-	+	6	+	-	+	-
7	-	+	+	7	-	+	+	-
8	+	+	+	8	+	+	+	+
				9	-	-	-	+
				10	+	-	-	+
				11	-	+	-	+
				12	+	+	-	+
				13	-	-	+	+
				14	+	-	+	+
				15	-	+	+	+
				16	+	+	+	+

Συνεπώς, βάσει του πίνακα σχεδίασης, στη 2^3 πειραματική σχεδίαση, το πείραμα με τυποποιημένη σειρά 7 θα πρέπει να υλοποιηθεί με τον παράγοντα 1 στο χαμηλό επίπεδο, και τους 2 και 3 σε υψηλό επίπεδο. Σε περίπτωση που τα πειράματα δεν υλοποιηθούν ταυτόχρονα, τότε πρέπει να μην γίνουν με τη σειρά που αναφέρονται στον πίνακα παραπάνω, αλλά με τυχαία σειρά. Αυτό γίνεται ώστε να μην υπάρξει κάποιο λαθεμένο συμπέρασμα για την επίδραση ενός παράγοντα. Παράδειγμα, αν υπάρχουν μεταβολές με τον χρόνο στις πειραματικές συνθήκες, τότε οι μεταβολές αυτές μπορούν να ενσωματωθούν και αποδοθούν στον παράγοντα 3, ο οποίος ξεκινάει με χαμηλές συνθήκες και τελειώνει με υψηλές συνθήκες στην τυποποιημένη σειρά.

Ο πίνακας μοντελοποίησης του 2^3 παραγοντικού πειράματος βοηθάει στον υπολογισμό των επιδράσεων για κάθε κύριο παράγοντα αλλά και για κάθε αλληλοεπίδραση. Ουσιαστικά ο πίνακας μοντελοποίησης παρουσιάζει τα πρόσημα που θα χρησιμοποιηθούν στον υπολογισμό των επιδράσεων των κυρίων παραγόντων και αλληλοεπιδράσεων.

Ο πίνακας παρουσιάζεται στη συνέχεια για ένα 2^3 πλήρες παραγοντικό σχεδιασμό.

Πίνακας 10.2. Μήτρα μοντελοποίησης για πείραμα σε 2 επίπεδα με 3 παράγοντες

Πείραμα	X0	X1	X2	X3	X12	X13	X23	X123	y
1	+	-	-	-	+	+	+	-	Y ₁
2	+	+	-	-	-	-	+	+	Y ₂
3	+	-	+	-	-	+	-	+	Y ₃
4	+	+	+	-	+	-	-	-	Y ₄
5	+	-	-	+	+	-	-	+	Y ₅
6	+	+	-	+	-	+	-	-	Y ₆
7	+	-	+	+	-	-	+	-	Y ₇
8	+	+	+	+	+	+	+	+	Y ₈

Ο πίνακας 10.2 μας δίνει τα πρόσημα για τον υπολογισμό της επίδρασης για κάθε παράγοντα ή αλληλοεπίδραση. Τα πρόσημα θέτονται μπροστά στην τιμή y, αθροίζονται και λαμβάνεται η μέση τιμή.

Βάσει του παραπάνω, η επίδραση του κάθε κύριου παράγοντα υπολογίζεται ως η διαφορά της μέση τιμής της απόκρισης με τον παράγοντα στο χαμηλό επίπεδο από την μέση τιμή της απόκρισης με τον παράγοντα στο υψηλό επίπεδο. Ας δούμε παραδειγματικά την επίδραση του παράγοντα 1 (όταν λέμε επίδραση του κυρίου παράγοντα εννοούμε την επίδραση στην απόκριση). Θα πρέπει να υπολογίσουμε τη μέση τιμή της απόκρισης y στα υψηλά επίπεδα του παράγοντα 1, που προέρχεται από 4 τιμές, και να υπολογίσουμε τη διαφορά της από την μέση τιμή της απόκρισης όταν ο παράγοντας 1 είναι στο χαμηλό επίπεδο (επίσης 4 τιμές).

Οι τιμές της απόκρισης με τον παράγοντα 1 σε υψηλό επίπεδο είναι οι: y₂, y₄, y₆, y₈

Οι τιμές της απόκρισης με τον παράγοντα 1 σε χαμηλό επίπεδο είναι οι: y₁, y₃, y₅, y₇

Συνεπώς, η μέση μεταβολή της απόκρισης όταν ο παράγοντας 1 μεταβάλλεται από το χαμηλό στο υψηλό επίπεδο είναι:

$$(y_2+y_4+y_6+y_8)/4 - (y_1+y_3+y_5+y_7)/4 \text{ ή}$$

$$\text{Επίδραση } 1 : \frac{-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8}{4}$$

Η παραπάνω σχέση υπολογίζει την επίδραση και του κάθε παράγοντα (κυρία επίδραση) και των αλληλοεπιδράσεων ξεχωριστά. Με βάση την τελευταία σχέση, βλέπουμε ότι είναι δυνατή η εξαγωγή της επίδρασης του κάθε παράγοντα «τοποθετώντας» τα πρόσημα + και -, όπως περιλαμβάνονται σε κάθε στήλη για την κάθε τιμή X, μπροστά στις αντίστοιχες τιμές της απόκρισης, και διαιρώντας στη συνέχεια με την τιμή 4. Η τιμή 4 βέβαια έχει να κάνει με το 2³ πείραμα. Διαιρώ δηλαδή πάντα με την τιμή 2^{k-1}, όπου k είναι οι παράγοντες που θέλω να μελετήσω. Αν είχα πείραμα 2⁴, τότε ο παρονομαστής στην παραπάνω εξίσωση θα ήταν ο 2⁴⁻¹=2³=8.

Μόνο στην περίπτωση της συνολικής μέσης τιμής (μεγάλος μέσος όρος), που συμβολίζεται με X_0 , διαιρώ με 8 (δηλ. 2^k) αντί για 4, που είναι ουσιαστικά η μέση τιμή των 8 αποκρίσεων των 8 πειραμάτων. Δηλαδή:

$$X_0 = \frac{+y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8}{8}$$

Μετά τον υπολογισμό του μεγάλου μέσου (X_0) και των κυρίων επιδράσεων (X_1, X_2, X_3), πρέπει να υπολογιστούν οι αλληλοεπιδράσεις, ήτοι οι $X_{12}, X_{13}, X_{23}, X_{123}$. Στο πίνακα μοντελοποίησης φαίνονται και οι αλληλοεπιδράσεις (συνέργιες ή ανταγωνισμοί) μεταξύ των παραγόντων. Τα πρόσημα των αλληλοεπιδράσεων εξάγονται εύκολα διότι απλά προκύπτουν από το γινόμενο των πρόσημων των αντίστοιχων κυρίων παραγόντων. Στην περίπτωση που έχουμε θετικό πρόσημο μιλάμε για συνέργια, ενώ στην αντίθετη περίπτωση μιλάμε για ανταγωνισμό. Παραδειγματικά, το πρόσημο του X_{12} προκύπτει από το γινόμενο του πρόσημου – για το X_1 και του πρόσημου – για το X_2 στο πείραμα 1, δηλ. τελικώς +. Το πρόσημο (δηλ. το επίπεδο) του X_{123} προκύπτει από το γινόμενο των πρόσημων των X_1, X_2, X_3 στο κάθε πείραμα (π.χ. – επί + επί + ήτοι - στο πείραμα 7). Με την ίδια λογική εξάγονται τα πρόσημα για όλες τις αλληλοεπιδράσεις ανά πείραμα. Τα πρόσημα αυτά, όμοια με τους κύριους παράγοντες, τίθενται μπροστά από την αντίστοιχη τιμή της απόκρισης y , αθροίζονται οι τιμές y και διαιρούνται με το 4.

Με βάση τον πίνακα μοντελοποίησης λοιπόν, μπορούμε να υπολογίσουμε το X_0 (μεγάλος μέσος όρος) και τα (επτά) $X_1, X_2, X_3, X_{12}, X_{23}, X_{13}, X_{123}$. Έχουμε δηλαδή 8 πειράματα και 8 αγνώστους, που υπολογίζονται σε ένα πείραμα 2^3 . Οι 7 επιδράσεις είναι αναλυτικά οι εξής:

$$\text{Επίδραση } X_1: \frac{-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8}{4}$$

$$\text{Επίδραση } X_2: \frac{-y_1 - y_2 + y_3 + y_4 - y_5 - y_6 + y_7 + y_8}{4}$$

$$\text{Επίδραση } X_3: \frac{-y_1 - y_2 - y_3 - y_4 + y_5 + y_6 + y_7 + y_8}{4}$$

$$\text{Επίδραση } X_{12}: \frac{+y_1 - y_2 - y_3 + y_4 + y_5 - y_6 - y_7 + y_8}{4}$$

$$\text{Επίδραση } X_{13}: \frac{+y_1 - y_2 + y_3 - y_4 - y_5 + y_6 - y_7 + y_8}{4}$$

$$\text{Επίδραση } X_{23}: \frac{+y_1 + y_2 - y_3 - y_4 - y_5 - y_6 + y_7 + y_8}{4}$$

$$\text{Επίδραση } X_{123}: \frac{-y_1 + y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8}{4}$$

Για ένα πείραμα 2^4 , θα είχαμε να υπολογίσουμε 16 αγνώστους, ήτοι:

- το μεγάλο μέσο όρο
- 4 κύριες επιδράσεις (X_1, X_2, X_3, X_4)
- 6 επιδράσεις μεταξύ δύο παραγόντων ($X_{12}, X_{13}, X_{14}, X_{23}, X_{24}, X_{34}$)
- 4 επιδράσεις μεταξύ τριών παραγόντων ($X_{123}, X_{234}, X_{134}, X_{124}$)
- 1 αλληλεπίδραση μεταξύ 4 παραγόντων (X_{1234}).

Το πείραμα 2^3 μπορεί να παρουσιαστεί και κατανοηθεί καλύτερα γραφικά. Ας υποθέσουμε ότι έχουμε μία χημική αντίδραση, η απόδοση της οποίας επηρεάζεται από 3 παράγοντες τους οποίους μελετούμε. Οι παράγοντες είναι η θερμοκρασία (T), η συγκέντρωση (C) και ο καταλύτης (K). Δύο παράγοντες είναι ποσοτικοί (θερμοκρασία, συγκέντρωση), δηλαδή μεταβάλλονται μεταξύ δύο τιμών και ένας παράγων είναι ποιοτικός, δηλαδή παρουσία ή απουσία καταλύτη στην αντίδραση.

Τα επίπεδα των παραγόντων είναι λοιπόν:

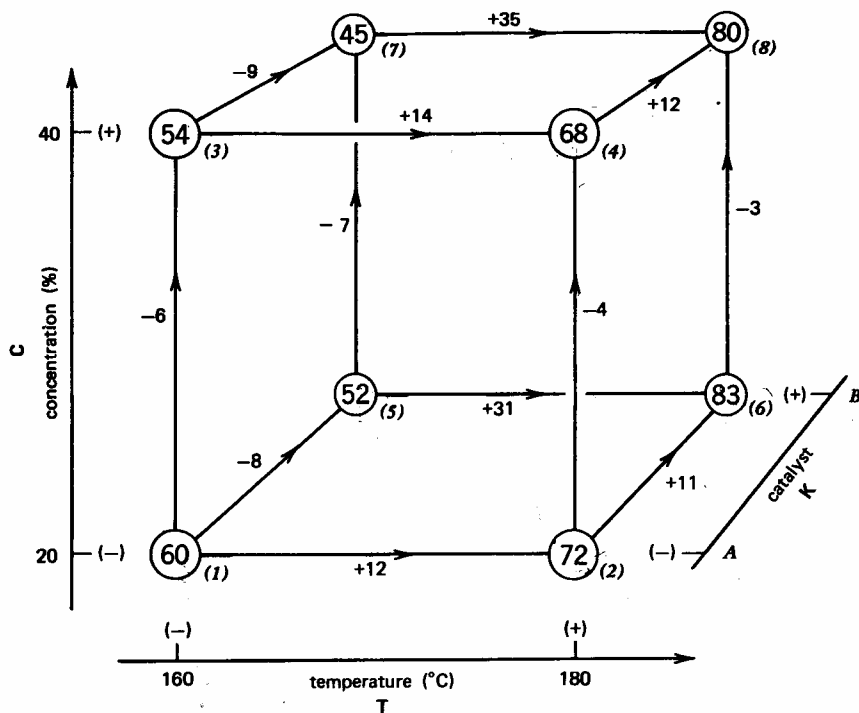
Παράγοντας	Χαμηλό επίπεδο (-)	Υψηλό επίπεδο (+)
Θερμοκρασία	160°C	180°C
Συγκέντρωση	20%	40%
Καταλύτης	Όχι	Ναι

Οι τιμές της απόδοσης της χημικής αντίδρασης που έχουν καταχωρηθεί, βάσει του πίνακα σχεδίασης περιλαμβάνονται στον πίνακα 10.3.

Πίνακας 10.3. Μήτρα σχεδίασης για 2^3 πείραμα που μετράει την απόδοση χημικής αντίδρασης

Πείραμα	Θερμοκρασία (°C)	Συγκέντρωση (%)	Καταλύτης	Απόδοση (gr) – y
1	-	-	-	60
2	+	-	-	72
3	-	+	-	54
4	+	+	-	68
5	-	-	+	52
6	+	-	+	83
7	-	+	+	45
8	+	+	+	80

Η σχεδίαση και οι αποδόσεις των 8 πειραμάτων φαίνονται στο σχήμα 10.1.



Σχήμα 10.1. Πείραμα με 3 παράγοντες σε 2 επίπεδα. Οι τιμές στο μέσο της κάθε γραμμής δείχνουν τη μεταβολή της απόκρισης μεταξύ της μίας θέσης του πειράματος σε άλλη (π.χ. μεταβολή της απόκρισης όταν μεταβάλλεται ο καταλύτης από το χαμηλό στο υψηλό επίπεδο, με τη θερμοκρασία και τη συγκέντρωση να παραμένουν στο χαμηλό επίπεδο είναι -8). Οι τιμές στην παρένθεση δεξιά των τιμών των αποκρίσεων είναι απλά η «κωδικοποίηση» των πειραμάτων βάσει του πίνακα σχεδίασης 10.3.

Με βάση τη μήτρα μοντελοποίησης όπως αναγράφεται παραπάνω, η επίδραση των κυρίων παραγόντων είναι:

$$\text{Επίδραση θερμοκρασίας (T)} = (72+68+83+80)/4 - (60+54+52+45)/4 = 23$$

$$\text{Επίδραση συγκέντρωσης (C)} = (54+68+45+80)/4 - (60+72+52+83)/4 = -5$$

$$\text{Επίδραση καταλύτη (K)} = (52+83+45+80)/4 - (60+72+54+68)/4 = 1,5$$

Παρατηρούμε ότι οι επιδράσεις υπολογίζονται από τη διαφορά της απόκρισης της μέσης τιμής στα 4 υψηλά επίπεδα του παράγοντα (\bar{y}_+) μείον τη μέση τιμή της απόκρισης στα 4 χαμηλά επίπεδα του ίδιου παράγοντα (\bar{y}_-). Δηλαδή:

$$\text{Κυρία επίδραση} = \bar{y}_+ - \bar{y}_-$$

Είναι φανερό εξάλλου ότι η επίδραση της θερμοκρασίας είναι πολύ μεγαλύτερη από τις επιδράσεις των 2 άλλων κυρίων παραγόντων.

Παρατηρούμε εξάλλου από το σχήμα 10.1 ότι η επίδραση της θερμοκρασίας είναι πολύ μεγαλύτερη όταν υπάρχει καταλύτης σε σχέση με την απουσία αυτού. Αυτό παρατηρείται και στα δύο επίπεδα της θερμοκρασίας, αφού με απουσία καταλύτη η αύξηση είναι +12 και με παρουσία καταλύτη είναι +31, στη χαμηλή θερμοκρασία, και με απουσία καταλύτη η αύξηση είναι +14 και με παρουσία καταλύτη είναι +35, στην υψηλή θερμοκρασία. Αυτό σημαίνει ότι η επίδραση της θερμοκρασίας δεν είναι προσθετική (αν ήταν θα αναμέναμε αύξηση όμοια στην απόκριση σε όλες τις περιπτώσεις), αλλά υπάρχει όπως φαίνεται συνέργια με τον καταλύτη. Η αλληλεπίδραση αυτή υπολογίζεται λαμβάνοντας τη μέση επίδραση της θερμοκρασίας σε παρουσία καταλύτη μείον τη μέση επίδραση της θερμοκρασίας σε απουσία καταλύτη, και ονομάζεται αλληλεπίδραση T, K.

Καταλύτης	Μέση επίδραση θερμοκρασίας
(+)	$(31+35)/2 = 33$
(-)	$(14+12)/2 = 13$
Διαφορά	20

Η αλληλεπίδραση είναι το μισό της παραπάνω διαφοράς, ήτοι $20/2=10$

Όμοια η αλληλεπίδραση μεταξύ θερμοκρασίας και καταλύτη μπορεί να υπολογιστεί λαμβάνοντας τη μέση επίδραση του καταλύτη σε υψηλή θερμοκρασία μείον τη μέση επίδραση του καταλύτη σε χαμηλή θερμοκρασία, δηλαδή:

Θερμοκρασία	Μέση επίδραση καταλύτη
(+)	$(12+11)/2 = 11.5$
(-)	$(-9-8)/2 = - 8.5$
Διαφορά	20

Όμοια, η αλληλεπίδραση είναι το μισό της παραπάνω διαφοράς, ήτοι $20/2=10$.

Βλέπουμε λοιπόν ότι η αλληλεπίδραση T x K είναι όμοια όπως και να υπολογιστεί.

Τελικά, η αλληλεπίδραση T x K υπολογίζεται ως εξής:

$$33 = (31+35)/2 = \frac{1}{2} (y_6-y_5+y_8-y_7)$$

$$13 = (12+14)/2 = \frac{1}{2} (y_2-y_1+y_4-y_3)$$

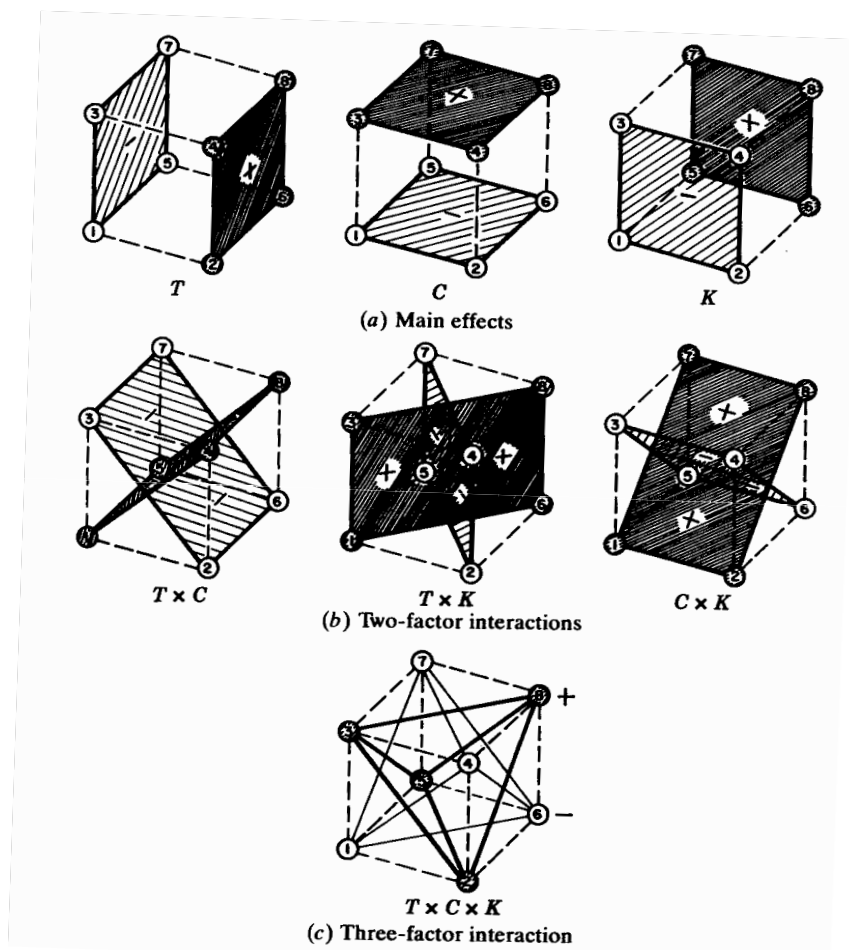
$$T \times K \text{ αλληλεπίδραση} = \frac{1}{2} \text{ διαφοράς} = \frac{1}{2} (33-13) = 10 =$$

$$= \frac{1}{4} (y_1 - y_2 + y_3 - y_4 - y_5 + y_6 - y_7 + y_8) =$$

$$\frac{(y_1 + y_3 + y_6 + y_8)}{4} - \frac{(y_2 + y_4 + y_5 + y_7)}{4}$$

Παρατηρώντας τη δεύτερη παραπάνω γραμμή, θα δούμε ότι τα πρόσημα που έχουν τεθεί μπροστά από τις τιμές y αντιστοιχούν με πρόσημα της αλληλεπίδρασης X13 στον πίνακα μοντελοποίησης, ενώ ο αριθμητής διαιρείται πάλι με 4, αφού έχουμε πείραμα 2^3 . Συνεπώς, όπως και στην περίπτωση των κυρίων επιδράσεων, οι αλληλοεπιδράσεις

υπολογίζονται από τη μήτρα μοντελοποίησης με τη χρήση των σχετικών προσήμων. Οι αλληλεπιδράσεις βασίζονται και υπολογίζονται στις διαγώνιους του «κύβου» του σχήματος 10.1, όπως φαίνεται στο σχήμα 10.2.



Σχήμα 10.2. Γραφική απεικόνιση των υπολογισμών των κυρίων επιδράσεων (T,C,K) και των αλληλεπιδράσεων (TxC, TxK, CxK, TxCxK)

Συνδυαστικά με την τεχνική των πρόσημων και του πίνακα μοντελοποίησης, έχει προταθεί και ο αλγόριθμος Yates για τον υπολογισμό των επιδράσεων και αλληλοεπιδράσεων. Ο αλγόριθμος αυτός είναι αρκετά δημοφιλής, κατά τον γράφοντα όμως δεν διαφέρει ούτε απλοποιεί ιδιαίτερα την τεχνική των πρόσημων και συνεπώς δεν θα παρουσιαστεί εδώ. Ο αλγόριθμος περιγράφεται αναλυτικά στον Box et al. (1978, σελ. 323). Ο αλγόριθμος Yates βασίζεται απλά στο ότι δεν χρειάζεται να αναπτυχθεί η μήτρα μοντελοποίησης αλλά μπορούν να υπολογιστούν οι επιδράσεις μόνο βάσει της μήτρας σχεδίασης χρησιμοποιώντας μία συγκεκριμένη διαδοχή προσθέσεων και αφαιρέσεων²⁴.

²⁴ Εξάλλου, με τη χρήση στατιστικών προγραμμάτων, δεν απαιτείται στην πράξη ούτε η χρήση του αλγόριθμου Yates, ούτε της μήτρας μοντελοποίησης.

Αφού υπολογιστούν οι επιδράσεις, βάσει των εξισώσεων που αναφέρονται παραπάνω, το επόμενο βήμα είναι η εκτίμηση της σημαντικότητας των επιδράσεων αυτών και τελικά η εξαγωγή του μοντέλου που περιγράφει και υπολογίζει την απόκριση συναρτήσει των επιπέδων των παραγόντων.

Εκτίμηση ακρίβειας επιδράσεων

Σε περίπτωση που υπάρχουν επαναλήψεις, τότε η ακρίβεια των επιδράσεων υπολογίζεται από τον υπολογισμό του διαστήματος εμπιστοσύνης αυτών. Για τον υπολογισμό του διαστήματος εμπιστοσύνης καταρχάς υπολογίζεται η διασπορά της κάθε επίδρασης, που στην περίπτωση ενός 2^3 πειράματος, βασίζεται στις παρακάτω σχέσεις.

$$\begin{aligned} \text{Var}(\text{επίδραση}) &= \text{Var}\left[\frac{\pm \bar{y}_1 \pm \bar{y}_2 \pm \bar{y}_3 \pm \bar{y}_4 \pm \bar{y}_5 \pm \bar{y}_6 \pm \bar{y}_7 \pm \bar{y}_8}{4}\right] = \\ &= \left(\frac{1}{4}\right)^2 \cdot [\text{Var}(\bar{y}_1) + \text{Var}(\bar{y}_2) + \dots + \text{Var}(\bar{y}_8)]^{25} \end{aligned}$$

Οι τιμές \pm δηλώνουν απλά ότι η παραπάνω σχέση υπολογίζει την (όμοια) διασπορά για όλες τις επτά επιδράσεις και αλληλεπιδράσεις ενός πειράματος 2^3 , αφού τελικά οι επιμέρους διασπορές των τιμών y αθροίζονται πάντα, ανεξάρτητα από το αν υπάρχει αρνητικό ή θετικό πρόσημο μπροστά από τις τιμές y .

Η διασπορά του μεγάλου μέσου όρου υπολογίζεται σχεδόν παρόμοια, δηλαδή:

$$\begin{aligned} \text{Var}(\text{μέσου όρου}) &= \text{Var}\left[\frac{+ \bar{y}_1 + \bar{y}_2 + \bar{y}_3 + \bar{y}_4 + \bar{y}_5 + \bar{y}_6 + \bar{y}_7 + \bar{y}_8}{8}\right] = \\ &= \left(\frac{1}{8}\right)^2 \cdot [\text{Var}(\bar{y}_1) + \text{Var}(\bar{y}_2) + \dots + \text{Var}(\bar{y}_8)] \end{aligned}$$

Όπως, φαίνεται, ανεξάρτητα από τα πρόσημα, στην τελική διασπορά όλες οι επιμέρους διασπορές αθροίζονται, σύμφωνα με τους κανονισμούς εκτίμησης των διασπορών αθροισμάτων ή διαφορών. Μόνο ο παρονομαστής διαφέρει, όπως φαίνεται, αφού στη μία περίπτωση είναι 4^2 και στην περίπτωση του μέσου όρου είναι 8^2 ²⁶.

Συνεπώς, με τον υπολογισμό των διασπορών, που είναι εφικτός αν υπάρχει τουλάχιστον μία επανάληψη ανά μέτρηση y , είναι δυνατή η εκτίμηση του διαστήματος εμπιστοσύνης, όπως έχει παρουσιαστεί σε προηγούμενο κεφάλαιο. Αν το διάστημα εμπιστοσύνης περιέχει το 0, για το επιθυμητό επίπεδο σημαντικότητας, τότε η συγκεκριμένη επίδραση θεωρείται μηδενική.

²⁵ Βασίζεται στη σχέση: $y = k \pm k_a a \pm k_b b \pm \dots \pm k_n n$, τότε $\sigma_y^2 = k_a^2 \sigma_a^2 + k_b^2 \sigma_b^2 + \dots + k_n^2 \sigma_n^2$

²⁶ Προφανώς αυτό ισχύει για το 2^3 μοντέλο. Για μοντέλο 2^4 , οι τιμές θα ήταν αντίστοιχα 8^2 και 16^2 .

Το τελικό βήμα της παραπάνω διαδικασίας είναι να αναπτυχθεί μία εμπειρική εξίσωση που να περιγράφει την απόκριση συναρτήσει των επιπέδων των παραγόντων. Η εξίσωση αυτή, για ένα 2^3 πείραμα, έχει τη μορφή:

$$\text{Επίδραση } y = X_0 + C_1 X_1 + C_2 X_2 + C_3 X_3 + C_{12} X_1 X_2 + C_{13} X_1 X_3 + C_{23} X_2 X_3 + C_{123} X_1 X_2 X_3$$

Η εξίσωση αυτή, έχει ως ανεξάρτητες μεταβλητές τις τιμές X , που είναι τα επίπεδα των παραγόντων, και τα οποία μπορούν να πάρουν τιμές -1 ή $+1$ μόνο. Η εξαρτημένη μεταβλητή είναι προφανώς η μετρηθείσα απόκριση y , ενώ οι συντελεστές C είναι οι ζητούμενοι συντελεστές του εμπειρικού μοντέλου. Η τιμή X_0 είναι ο μεγάλος μέσος όρος, όπως έχει ήδη υπολογιστεί. Οι τιμές C είναι ίσες με το $\frac{1}{2}$ των τιμών των επιδράσεων. Ο λόγος που ο συντελεστής είναι το $\frac{1}{2}$ της επίδρασης έχει να κάνει με τις τιμές X , που είναι είτε -1 είτε $+1$. Συνεπώς, μεταβολή από το -1 στο $+1$ είναι μεταβολή ίση με 2 και συνεπώς ο συντελεστής του X πρέπει να λάβει τιμή ίση με επίδραση/ 2 . Αν ο συντελεστής ήταν ίσος με την επίδραση, τότε μετακίνηση από το χαμηλό στο υψηλό επίπεδο θα είχε σαν αποτέλεσμα διπλασιασμό (ή υποδιπλασιασμό) της επίδρασης στην απόκριση. Παραδειγματικά, για την περίπτωση του μοντέλου με τους παράγοντες θερμοκρασία (παράγοντας 1), συγκέντρωση (παράγοντας 2), καταλύτης (παράγοντας 3), οι τιμές των συντελεστών είναι:

$$\begin{aligned} X_0 &= 64,25 \\ C_1 &= \text{Επίδραση } 1 : 2 = 23 : 2 = 11,5 \\ C_2 &= \text{Επίδραση } 2 : 2 = -5 : 2 = -2,5 \\ C_3 &= \text{Επίδραση } 3 : 2 = 1,5 : 2 = 0,75 \\ C_{12} &= \text{Αλληλοεπίδραση } 12 : 2 = 1,5 : 2 = 0,75 \\ C_{13} &= \text{Αλληλοεπίδραση } 13 : 2 = 10 : 2 = 5,0 \\ C_{23} &= \text{Αλληλοεπίδραση } 23 : 2 = 0,0 : 2 = 0,0 \\ C_{123} &= \text{Αλληλοεπίδραση } 123 : 2 = 0,5 : 2 = 0,25 \end{aligned}$$

Συνεπώς, το πλήρες εμπειρικό μοντέλο που προκύπτει από το παραπάνω 2^3 πείραμα είναι:

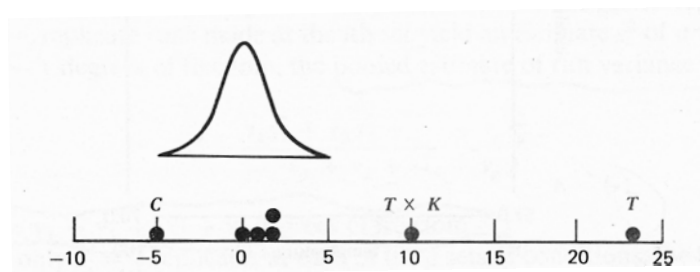
$$y = 64,25 + 11,5 X_1 - 2,5 X_2 + 0,75 X_3 + 0,75 X_1 X_2 + 5,0 X_1 X_3 + 0,0 X_2 X_3 + 0,25 X_1 X_2 X_3$$

Με την παραπάνω λοιπόν εξίσωση, μπορεί να περιγραφεί η απόκριση σε οποιοδήποτε συνδυασμό επιπέδων των 3 παραγόντων. Δηλαδή, αν μας ζητήσουν να εκτιμήσουμε την απόκριση στο χαμηλό επίπεδο θερμοκρασίας και στα δύο υψηλά επίπεδα συγκέντρωσης και καταλύτη, τότε θα θέσουμε το X_1 στο -1 και τα X_2 και X_3 στα $+1$. Φυσικά, τα γινόμενα των επιπέδων των παραγόντων (αλληλεπιδράσεις) μπορούν εύκολα να υπολογιστούν στη συνέχεια και να εισαχθούν στην εξίσωση. Οπότε, π.χ. $X_1 X_2 = (-1)*(+1) = -1$, $X_1 X_2 X_3 = (-1)*(+1)*(+1) = -1$ κ.λ.π.

Τελικά όμως είναι στατιστικά σημαντικές όλες οι επιδράσεις και αλληλοεπιδράσεις? Είναι προφανές παραδειγματικά ότι η αλληλοεπίδραση των παραγόντων 2×3 είναι

μηδενική, αφού έχει υπολογιστεί τιμή σχεδόν ίση με 0 για την αλληλοεπίδραση αυτή. Όμως τι ισχύει για την κύρια επίδραση του παράγοντα 2 (που είναι -5) ή του παράγοντα 3 (που είναι 1,5)? Είναι στατιστικά σημαντικές οι επιπτώσεις αυτές ή μήπως θα μπορούσαμε να αναπτύξουμε το εμπειρικό μας μοντέλο χωρίς να εισάγουμε τις συγκεκριμένες επιπτώσεις?

Η κανονική κατανομή των επτά (7) επιδράσεων και αλληλοεπιδράσεων φαίνεται στο σχήμα 10.3. Η κατανομή – που φαίνεται στο πάνω μέρος του σχήματος - βασίστηκε σε κατανομή t με 8 βαθμούς ελευθερίας και μέση τιμή 0 και κλίμακα ίση με 1,4. Η τιμή 1,4 αποτελεί το τυπικό σφάλμα της κάθε επίδρασης, δηλ. θεωρήθηκε ότι έγινε μία επανάληψη ανά πείραμα και τελικά το τυπικό σφάλμα είχε την τιμή 1,4 για κάθε επίδραση ή αλληλοεπίδραση.



Σχήμα 10.3. Κύριες επιδράσεις και αλληλεπιδράσεις για πείραμα 2^3 συγκρινόμενες με κατανομή αναφοράς t με 8 βαθμούς ελευθερίας και κλίμακα μετατροπής 1,4²⁷

Από το σχήμα 10.3 μπορούμε να δούμε ότι 3 από τις επιδράσεις / αλληλοεπιδράσεις (T, C, T x K) απέχουν αρκετά από τις υπόλοιπες τιμές και τη μέση τιμή της κατανομής. Αυτές οι επιδράσεις είναι που θεωρούνται στατιστικά σημαντικές αφού έχουν φαινομενικά πιθανότητα εμφάνισης αρκετά μικρή. Ανάλογα με την τιμή P που συνοδεύει τις επιδράσεις αυτές και εφόσον η τιμή P είναι μικρότερη από το επίπεδο σημαντικότητας (π.χ. $5\%/2=2,5\%$) τότε οι επιδράσεις / αλληλοεπιδράσεις αυτές κρίνονται ως στατιστικά σημαντικές. Το βέλτιστο μειωμένο εμπειρικό μοντέλο (best reduced model) θα έχει τη μορφή:

$$y = 64,25 + 11,5 X1 - 2,5 X2 + 5,0 X1 X3$$

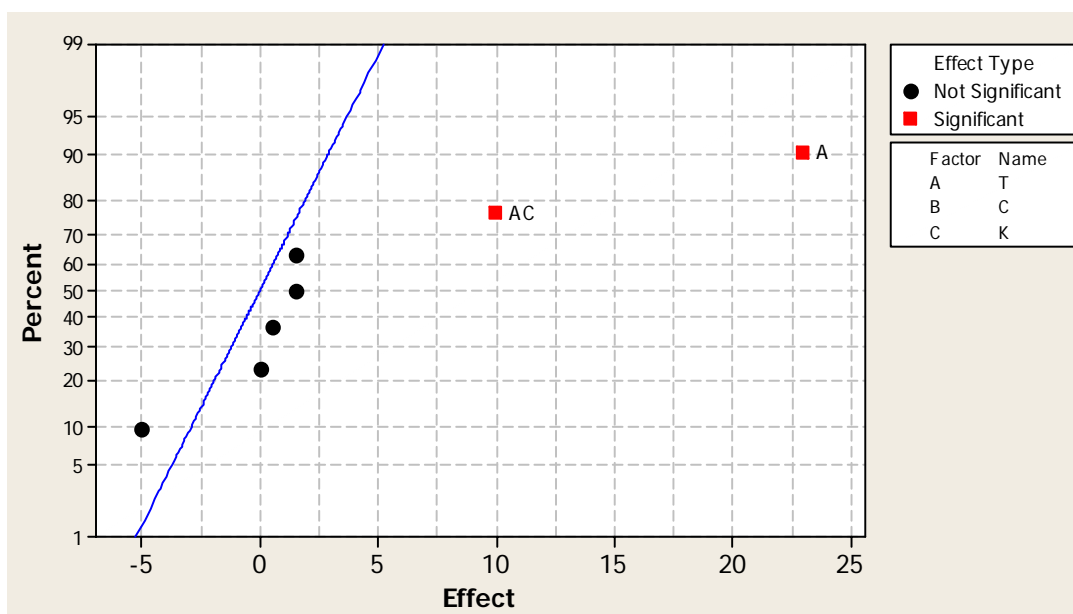
Κοιτώντας τα μεγέθη των επιδράσεων μπορούμε να συμπεράνουμε τα εξής:

²⁷ Η κλίμακα μετατροπής σημαίνει απλά ότι αν για την κατανομή t με 8 βαθμούς ελευθερίας έχουμε για κάθε τιμή t συγκεκριμένη τιμή πιθανότητας (δηλαδή τιμή του εμβαδόν δεξιά ή αριστερά του t) τότε η μετατροπή σε κλίμακα με τυπικό σφάλμα 1,4 γίνεται πολλαπλασιάζοντας την τιμή t με το τυπικό σφάλμα (οριζόντιος άξονας), ενώ την αντίστοιχη τιμή στο κατακόρυφο άξονα πρέπει να την διαιρέσουμε με το τυπικό σφάλμα. Η χρήση της κλίμακας μετατροπής έχει σαν αποτέλεσμα ότι το εμβαδό κάτω από την νέα καμπύλη t να είναι ίσο με 1.00.

1. Η επίδραση του παράγοντα 2 (συγκέντρωση) είναι η μείωση της απόδοσης κατά 5 μονάδες και αυτό – όπως φαίνεται από τις αλληλοεπιδράσεις – δεν επηρεάζεται από τα επίπεδα των άλλων παραγόντων.
2. Οι (κύριες) επιδράσεις των παραγόντων θερμοκρασία (T) και καταλύτη (K) δεν μπορούν από μόνες τους να εξηγηθούν. Υπάρχει σαφώς αλληλοεπίδραση μεταξύ τους, αφού η επίδραση της θερμοκρασίας απουσία καταλύτη είναι 13 μονάδες και η επίδραση της θερμοκρασίας παρουσία καταλύτη είναι 33 μονάδες. Αυτό εξάλλου φαίνεται και από την αλληλοεπίδραση T x K, που υπολογίστηκε και βρέθηκε ίση με 10,0 και είναι η δεύτερη μεγαλύτερη επίδραση εκ των 7 που υπολογίστηκαν, μετά την κυρία επίδραση της θερμοκρασίας.
3. Στο σημείο αυτό σημειώνεται ότι το γεγονός ότι η αλληλοεπίδραση T x K είναι μεγαλύτερη από την κυρία επίδραση του ιδίου του καταλύτη (1,5) είναι γενικά περίεργο. Γενικά δεν αναμένεται (και δεν έχει φυσική έννοια) μία αλληλοεπίδραση να είναι μεγαλύτερη από την επίδραση ενός εκ των δύο παραγόντων από τους οποίους προέρχεται. Συνεπώς το μοντέλο πρέπει να έχει τη μορφή:

$$y = 64,25 + 11,5 X_1 - 2,5 X_2 + 0,75 X_3 + 5,0 X_1 X_3$$

Η παραπάνω κατανομή των επιδράσεων μπορεί να εξαχθεί και σε γράφημα πιθανοτήτων, όπως παρακάτω (Minitab 14®).



Σχήμα 10.4. Διάγραμμα πιθανοτήτων για τις 7 επιδράσεις / αλληλοεπιδράσεις

Αν όλες οι κύριες επιδράσεις και αλληλοεπιδράσεις αντιπροσωπεύουν τυχαίες αποκλίσεις, τότε όλες θα βρίσκονταν περίπου επί της μπλε ευθείας, δηλ. θα κατανέμονταν κανονικά. Θα βρίσκονταν δηλαδή εντός ενός διαστήματος που θα όριζε το διάστημα εμπιστοσύνης στο επιθυμητό επίπεδο (π.χ. 95%). Οι δύο παράλληλες γραμμές (δεν αποτυπώνονται στο παραπάνω γράφημα) θα περιείχαν μεταξύ τους τις επιδράσεις

που κατανέμονται κανονικά και βρίσκονται κοντά στο μέσο όρο. Αν κάποιος παράγων έχει προκαλέσει μία επίδραση που να είναι μεγαλύτερη από την αναμενόμενη λόγω τυχαίου σφάλματος μόνο, τότε αυτή η επίδραση είναι στατιστικά σημαντική. Τότε αυτή η επίδραση δεν θα πέφτει πάνω στην ευθεία γραμμή και η επίδραση αυτή θα έχει πιθανότητα εμφάνισης μικρότερη του 2,5%. Αυτές είναι και οι στατιστικά σημαντικές επιπτώσεις.

Από το σχήμα 10.4, φαίνεται ότι τουλάχιστον 2 σημεία, η θερμοκρασία (συμβολίζεται με A) και η αλληλεπίδραση (AC) απέχουν αρκετά από την ευθεία. Το Minitab 14 υπολογίζει ποιες επιδράσεις απέχουν σημαντικά από τη μέση τιμή (δηλ. έχουν P μικρότερο του καθορισμένου από εμάς 2,5%) και τις παρουσιάζει στο σχήμα 10.4 σαν σημεία με κόκκινο χρώμα. Όπως φαίνεται, η κυρία επίδραση του παράγοντα «συγκέντρωση» δεν έχει κόκκινο χρώμα, δηλαδή θεωρείται μη στατιστικά σημαντικός στο καθορισμένο επίπεδο σημαντικότητας. Συνεπώς, γραφήματα όπως τα παραπάνω μπορούν να βοηθήσουν στην εκτίμηση της σημαντικότητας ή όχι κάποιων παραγόντων, ανεξάρτητα από την ύπαρξη επαναλήψεων.

Στην περίπτωση που υπάρχουν επαναλήψεις, τότε η εκτίμηση των διαστημάτων εμπιστοσύνης χρησιμοποιείται για την εκτίμηση της σημαντικότητας. Το τυπικό σφάλμα (S.E.) της επίδρασης, που υπενθυμίζεται είναι το ίδιο για όλες τις 7 επιδράσεις και αλληλοεπιδράσεις, δίνεται από τη σχέση:

$$\text{S.E.}(επίδρασης) = \sqrt{\frac{\text{Var}(effect)}{n}} = \frac{s(effect)}{\sqrt{n}}$$

Το τυπικό σφάλμα (S.E.aver) της μέσης τιμής δίνεται από τη σχέση:

$$\text{S.E.}(μέσης τιμής) = \sqrt{\frac{\text{Var}(aver)}{n}} = \frac{s(aver)}{\sqrt{n}}$$

, όπου s είναι η τυπική απόκλιση της επίδρασης. Η τιμή n απαιτεί μία προσοχή. Είναι ο αριθμός των δειγμάτων (ή επαναλήψεων) από τον οποίο προκύπτει η απόκριση σε κάθε θέση του πειράματος. Δηλαδή, σε αυτή την περίπτωση, το n δεν είναι η τιμή 2^k . Προσέξτε, όμως ότι όσον αφορά το μεγάλο μέσο όρο, το τυπικό σφάλμα υπολογίζεται αφού διαιρέσουμε τη διασπορά του μεγάλου μέσου όρου με τον αριθμό των πειραμάτων χωρίς τις επαναλήψεις, δηλαδή 2^k .

Βάσει των παραπάνω, το διάστημα εμπιστοσύνης για μία επίδραση/αλληλοεπίδραση ή/και το μέσο όρο και σε επίπεδο σημαντικότητας 95% είναι:

$$\begin{aligned} \mu \pm t_{v, 0.025} \text{ S.E.}(effect) \text{ ή} \\ \mu \pm t_{v, 0.025} \text{ S.E.}(aver) \end{aligned}$$

με v τους βαθμούς ελευθερίας που είναι ίσοι με $n \cdot 2^k - 1$ (n είναι ο αριθμός των επαναλήψεων. Χωρίς επανάληψη $n=1$).

Προσοχή: Το MINITAB υπολογίζει και παρουσιάζει τα τυπικά σφάλματα των συντελεστών του εμπειρικού μοντέλου. Συνεπώς, το διάστημα εμπιστοσύνης που υπολογίζεται αφορά στο συντελεστή και όχι στην επίδραση. Βέβαια, το MINITAB απευθείας υπολογίζει τις τιμές p για κάθε επίδραση και αλληλεπίδραση. Τιμές p μικρότερες του α που χρησιμοποιούμε (συνήθως το 0.05) δηλώνουν τη στατιστική σημαντικότητα της επίδρασης ή αλληλεπίδρασης.

Είναι προφανές ότι αν δεν υπάρχουν επαναλήψεις, τότε δεν μπορούν να υπολογιστούν διαστήματα εμπιστοσύνης. Επίσης, όπως θα φανεί και στο παράδειγμα στη συνέχεια, το MINITAB υπολογίζει σε γράφημα πιθανοτήτων τις τυποποιημένες επιδράσεις (standardized effects), οι οποίες δεν ταυτίζονται με τις επιδράσεις. Συγκεκριμένα, οι τυποποιημένες επιδράσεις, οι οποίες υπολογίζονται μόνο εφόσον έχουν πραγματοποιηθεί επαναληπτικά πειράματα, ορίζονται ως εξής:

$$\text{stand effect} = \frac{\text{effect}}{SE(\text{effect})}$$

Παράδειγμα

Ας δούμε την περίπτωση που έχουμε επαναληπτικές μετρήσεις (διπλές μετρήσεις) για ένα πείραμα 2^3 με την παρακάτω διάταξη. Ας υποθέσουμε ότι η απόκριση είναι συγκεντρώσεις σιδήρου σε ένα υδατικό διάλυμα, στο οποίο εξετάζουμε την επίδραση 3 κροκιδωτικών μέσων. Τα επίπεδα των παραγόντων είναι διαφορετικές συγκεντρώσεις των 3 κροκιδωτικών, σε πείραμα που πραγματοποιείται μέσω jar tests.

Πίνακας 10.4. Αποκρίσεις 2^3 πειράματος

	Τιμή δείγμα 1	Τιμή δείγμα 2	Μέση τιμή (mg/L)	Διασπορά (mg/L) ²	Παράγων A	Παράγων B	Παράγων C
1	38,9	41,5	40,20	3,38	-	-	-
2	45,7	45,4	45,55	0,05	+	-	-
3	47,8	48,8	48,30	0,50	-	+	-
4	45,8	43,8	44,80	2,00	+	+	-
5	45,2	47,6	46,40	2,88	-	-	+
6	46,9	48,3	47,60	0,98	+	-	+
7	41	45,8	43,40	11,52	-	+	+
8	53,5	52,4	52,95	0,60	+	+	+

Σημειώνεται ότι παρόλο που τα πειράματα παραπάνω έχουν γραφεί με την κανονική σειρά (standard order), διεξήχθησαν στην πράξη με τυχαία σειρά, όπως εξάλλου επιβάλλεται.

Πίνακας 10.5. Επιδράσεις και συντελεστές πειράματος 2^3

Επίδραση (effect)	Συντελεστής (coeff)	Διασπορά επίδρασης	Τυπικό σφάλμα επίδρασης	Τυπικό σφάλμα συντελεστή	Δ.Ε. συντ.*	Χαμηλή τιμή συντελ.	Υψηλή τιμή συντελ.
Μέσος όρος	46,15	0,34	0,21	0,21	0,44	45,71	46,59
A	3,15	1,37	0,83	0,41	0,88	0,69	2,46
B	2,43	1,37	0,83	0,41	0,88	1,54	3,31
C	2,87	1,37	0,83	0,41	0,88	1,99	3,76
AB	-0,12	1,37	0,83	0,41	0,88	-1,01	0,76
BC	-1,25	1,37	0,83	0,41	0,88	-2,13	-0,37
AC	2,23	1,37	0,83	0,41	0,88	1,34	3,11
ABC	4,30	1,37	0,83	0,41	0,88	3,42	5,18

*: Διάστημα εμπιστοσύνης σε επίπεδο εμπιστοσύνης 95%.

Το διάστημα εμπιστοσύνης υπολογίστηκε για επίπεδο σημαντικότητας 95% και για 15 βαθμούς ελευθερίας, που δίνει τελικά $t_{15, 0,025} = 2,131$. Από τον παραπάνω πίνακα εξάλλου βλέπουμε ότι ο μέσος όρος είναι σε μονάδες απόκρισης και συνεπώς δεν είναι κάποια επίδραση. Βέβαια και οι επιδράσεις είναι σε μονάδες της απόκρισης, αφού ουσιαστικά εκφράζουν τη μεταβολή της απόκρισης που παρατηρείται όταν οι παράγοντες αλλάζουν από το χαμηλό στο υψηλό επίπεδο. Οι επιδράσεις στο 2^3 πείραμα είναι συνεπώς πάντα 7 και οι συντελεστές του μοντέλου είναι ίσοι με το $\frac{1}{2}$ της τιμής των επιδράσεων. Από τα διαστήματα εμπιστοσύνης μπορεί να εξαχθεί το συμπέρασμα ότι η αλληλοεπίδραση AxB και οριακά η αλληλοεπίδραση BxC δεν είναι στατιστικά σημαντικές, ενώ όλες οι άλλες αλληλεπιδράσεις είναι στατιστικά σημαντικές (δηλ. οι 3 κύριες επιδράσεις των παραγόντων A,B,C και η αλληλοεπίδραση A x C, ενώ φαίνεται ότι και η τριπλή αλληλοεπίδραση A x B x C είναι επίσης στατιστικά σημαντική και περιέργως περισσότερο σημαντική από τις 3 κύριες επιδράσεις)

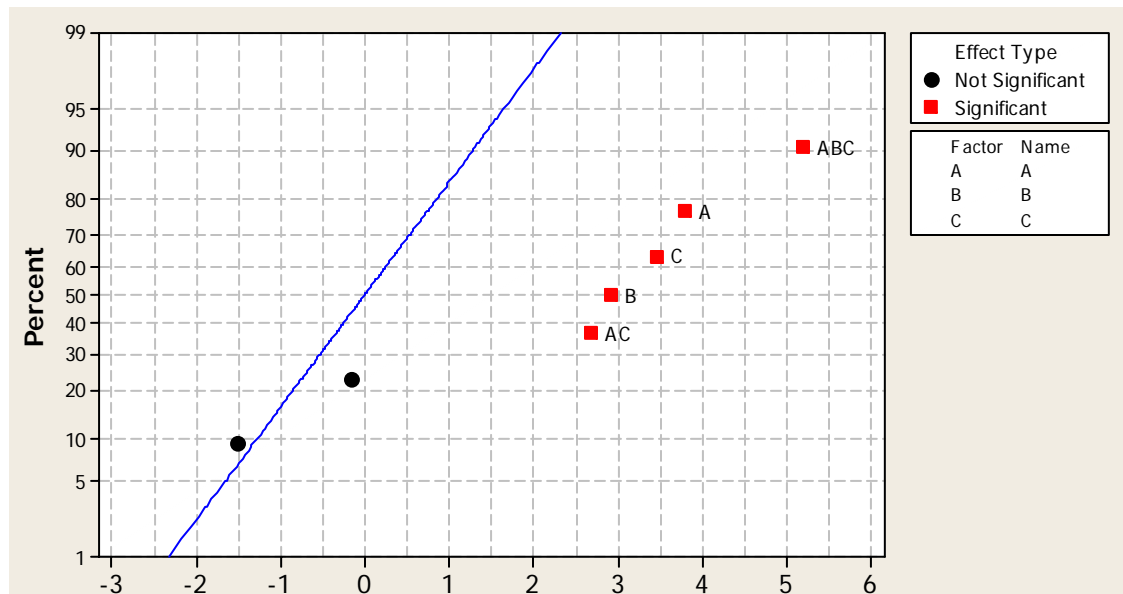
Το MINITAB14 εξάγει και τις τιμές t και p για κάθε επίδραση, οι οποίες περιλαμβάνονται στον πίνακα 10.6.

Πίνακας 10.6. Τιμές t και p κυρίων επιδράσεων και αλληλεπιδράσεων

	Τιμή t	Τιμή p	Κατάσταση
Μέσος όρος	111,55	0,000	
X1	3,81	0,005	<0,05 Σημαντικός
X2	2,93	0,019	<0,05 Σημαντικός
X3	3,47	0,008	<0,05 Σημαντικός
X12	-0,15	0,884	> 0,05 Μη σημαντικός
X23	2,69	0,028	<0,05 Σημαντικός
X13	-1,51	0,169	> 0,05 Μη σημαντικός
X123	5,20	0,001	<0,05 Σημαντικός

Στατιστικά σημαντικές είναι επιδράσεις με τιμή P (εκφράζει πιθανότητα προφανώς) μικρότερη από το καθορισμένο από εμάς όριο του 2,5%.

Το σχετικό γράφημα στην κλίμακα πιθανότητας είναι το παρακάτω:



Σχήμα 10.5. Γράφημα πιθανότητας για πείραμα 2^3 . Πέντε εκ των επτά επιδράσεων είναι στατιστικά σημαντικές. Το γράφημα περιέχει τις τυποποιημένες επιδράσεις και όχι τις επιδράσεις.

Το βέλτιστο μειωμένο εμπειρικό μοντέλο που περιγράφει το σύστημα είναι τελικά:

$$Y = 46,15 + 1,58 \cdot A + 1,21 \cdot B + 1,44 \cdot C + 1,11 \cdot A \cdot C + 2,15 \cdot A \cdot B \cdot C$$

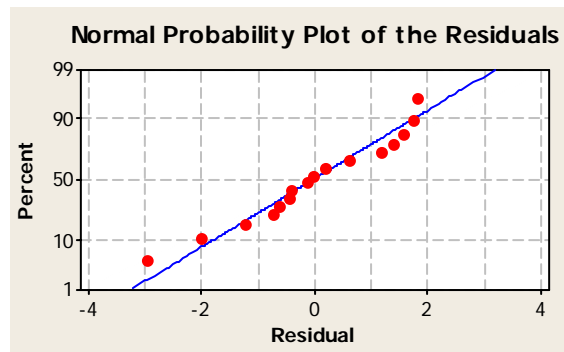
με τις τιμές A, B, C να παίρνουν τιμές -1 ή +1, ανάλογα με τον αν ο παράγοντας βρίσκεται στο χαμηλό ή υψηλό επίπεδο, αντίστοιχα, ενώ προφανώς Y είναι η απόκριση.

Αν θέλαμε να υπολογίσουμε τα τυπικά σφάλματα των συντελεστών στην παραπάνω εξίσωση, τότε απλά θα διαιρούσαμε τα τυπικά σφάλματα των επιδράσεων – όπως αυτά φαίνονται στον πίνακα 10.5 - με 2. Δηλαδή, η εξίσωση θα γραφόταν ως εξής:

$$Y = 46,15 + 1,58 \cdot A + 1,21 \cdot B + 1,4 \cdot C + 1,11 \cdot A \cdot C + 2,15 \cdot A \cdot B \cdot C$$

Είναι σημαντικό να τονιστεί ότι η τροποποιημένη εξίσωση, που δεν περιέχει τις μη στατιστικά σημαντικές επιδράσεις, θα έχει διαφορετικά σφάλματα σε σχέση με την πρώτη (πλήρη) εξίσωση. Και στις δύο περιπτώσεις, αλλά ειδικά στην δεύτερη, τα σφάλματα πρέπει να ελέγχονται με διαγνωστικά γραφήματα.

Το MINITAB δημιουργεί και τα κλασσικά διαγνωστικά γραφήματα για να ελεγχθούν τα σφάλματα του μοντέλου που εξάγεται. Τα διαγνωστικά γραφήματα περιλαμβάνουν τον έλεγχο της κανονικότητας των (16) σφαλμάτων (τα σφάλματα υπολογίζονται ως οι διαφορές των 16 αποκρίσεων από το μεγάλο μέσο όρο, ήτοι 46,15). Παραδειγματικά, τα σφάλματα για το παραπάνω παράδειγμα κατανέμονται κανονικά, όπως φαίνεται από το σχήμα 10.6.



Σχήμα 10.6. Διαγνωστικό γράφημα που ελέγχει κανονικότητα σφαλμάτων του βελτιωμένου μειωμένου μοντέλου. Τα συγκεκριμένα 16 σφάλματα κατανέμονται κανονικά και συνεπώς το μοντέλο κρίνεται ως άρτιο.

Τα διαγνωστικά γραφήματα ελέγχουν και την ανεξαρτησία των σφαλμάτων αλλά και την πιθανή τάση των σφαλμάτων σε σχέση με τις παρατηρούμενες τιμές, με τις τιμές που εξάγει το μοντέλο και σε σχέση με το χρόνο. Γενικά ελέγχεται η συνθήκη IIDN(0,σ), δηλαδή τα σφάλματα να είναι ανεξάρτητα, κανονικά, με μέση τιμή 0 και σταθερή διασπορά.

Τα πειράματα στα οποία γίνεται μεταβολή μόνος ενός παράγοντα κάθε φορά (One Factor At a Time experiments ή OFAT πειράματα), δεν έχουν τη δυνατότητα μελέτης αλληλοεπιδράσεων, απαιτούν περισσότερα πειράματα για να εξάγουν την ίδια πληροφορία με τα παραγοντικά ενώ και η πληροφορία που εξάγουν για τις επιδράσεις των κύριων παραγόντων δεν είναι τόσο πλήρης όσο με ένα παραγοντικό πείραμα.

Παράδειγμα, στην περίπτωση του 2^3 πειράματος για τη μελέτη της θερμοκρασίας, συγκέντρωσης και καταλύτη, αν θέλαμε να κάναμε χρήση των OFAT πειραμάτων θα απαιτούντο συνολικά 2^4 πειράματα για να πάρουμε την ίδια πληροφορία με τα $2^3=8$ πειράματα το:

- 8 πειράματα για τη μελέτη της επίδρασης της θερμοκρασίας, ήτοι 4 πειράματα σε κάθε επίπεδο θερμοκρασίας.
- 8 πειράματα για τη μελέτη της επίδρασης της συγκέντρωσης, ήτοι 4 πειράματα σε κάθε επίπεδο συγκέντρωσης.
- 8 πειράματα για τη μελέτη της επίδρασης του καταλύτη, ήτοι 4 πειράματα σε κάθε επίπεδο καταλύτη.

Γενικά, οι OFAT σχεδιασμοί απαιτούν $(k+1)/2 \times 2^k$ παραπάνω πειράματα από ότι τα πλήρη παραγοντικά. Συνεπώς για τρεις παράγοντες, απαιτούνται $2 \times 8=16$ παραπάνω πειράματα σε σχέση με την πλήρη παραγοντική σχεδίαση 2^3 .

Παράδειγμα

Την κροκιδώση των λυμάτων, με τη χρήση του κροκιδωτικού $FeCl_3$, την επηρεάζει, όπως έχουν δείξει προκαταρκτικά πειράματα, η συγκέντρωση του ίδιου του κροκιδωτικού μέσου, η συγκέντρωση H_2SO_4 και ο ρυθμός ανάμειξης του διαλύματος (εκφράζεται σε μονάδες ισχύος W). Η απόδοση της κροκιδώσης εκφράζεται με μετρήσεις συγκέντρωσης συνολικών στερεών (TSS σε mg/L) μετά την κροκιδώση. Προφανώς, ο στόχος μας είναι να μειωθούν τα TSS όσο περισσότερο αυτό είναι δυνατόν. Τα επίπεδα στα οποία θέτουμε τους 3 παράγοντες του πειράματος φαίνονται στον παρακάτω πίνακα.

Παράγοντας	Χαμηλό επίπεδο (-1)	Υψηλό επίπεδο (+1)
1. H_2SO_4 (mg/L)	10	35
2. Συγκέντρωση $FeCl_3$ (mg/L)	500	1000
3. Ρυθμός ανάδευσης (W)	100	500

Τα αποτελέσματα των μετρήσεων των TSS είναι (με βάση την κανονική σειρά, δηλ το y1 αντιστοιχεί στο -, -, - και το y8 στο +, +, +):

y1 = 28, y2=22, y3=10, y4=5, y5=42, y6=43, y7=45, y8=50 (όλα σε mg/L).

Γράψτε τη μήτρα σχεδίασης και μοντελοποίησης του παραπάνω 2^3 πειράματος. Αναλύστε τα δεδομένα και γράψτε το πλήρες και, κατά τη γνώμη σας, το βέλτιστο εμπειρικό μοντέλο που περιγράφει τα δεδομένα. Ποιοί/ποιός παράγοντες/ας φαίνεται (κατ'εκτίμηση) ότι είναι οι/ο στατιστικά σημαντικοί/ός?

Όπως φαίνεται από τα δεδομένα εκ πρώτης όψεως, χαμηλές συγκεντρώσεις TSS (άρα καλή απόδοση των κροκιδωτικών) επιτυγχάνεται στα πειράματα 3 & 4. Και στα δύο αυτά πειράματα, ο παράγοντας 2 (κροκιδωτικό) και ο παράγοντας 3 (ισχύς) βρίσκονται στο υψηλό και στο χαμηλό επίπεδο, αντίστοιχα. Αντίθετα, σχετικά υψηλές τιμές TSS παρατηρούνται όταν ο παράγοντας 3 βρίσκεται στο υψηλό επίπεδο (δηλ. ισχυρή ανάδευση).

Πίνακας 10.7. Μήτρα μοντελοποίησης του πειράματος με $FeCl_3$

Πείραμα	X0	X1	X2	X3	X12	X13	X23	X123	Y (mg/L)
1	+	-	-	-	+	+	+	-	28
2	+	+	-	-	-	-	+	+	22
3	+	-	+	-	-	+	-	+	10
4	+	+	+	-	+	-	-	-	5
5	+	-	-	+	+	-	-	+	42
6	+	+	-	+	-	+	-	-	43
7	+	-	+	+	-	-	+	-	45
8	+	+	+	+	+	+	+	+	50

X1: οξύ, X2: $FeCl_3$, X3: ρυθμός ανάδευσης (W)

Τα σκιασμένα κελιά του πίνακα 10.7 παρουσιάζουν τη μήτρα σχεδίασης του πειράματος ανεπτυγμένη σε κανονική σειρά. Σημειώνεται ότι στην πράξη τα πειράματα δεν πρέπει να διεξαχθούν ακολουθώντας τη κανονική σειρά. Δηλαδή, όταν δεν υπάρχει η

δυνατότητα ταυτόχρονης υλοποίησης και των 8 (ή 16) πειραμάτων, τότε πρέπει να γίνει τυχαία η επιλογή της σειράς υλοποίησης. Σε περίπτωση που υπάρχει η αναγκαιότητα ομαδοποίησης κάποιων πειραμάτων, τότε πρέπει να διερευνηθεί και η πιθανή επίδραση των ομάδων (blocks). Αυτό θα περιγραφεί στη συνέχεια (κεφ. 12).

Ακολουθώντας τις εξισώσεις που αναφέρονται παραπάνω, και βάσει του πίνακα μοντελοποίησης, υπολογίζονται οι επιδράσεις των κυρίων παραγόντων και των αλληλεπιδράσεων. Οι επιδράσεις είναι:

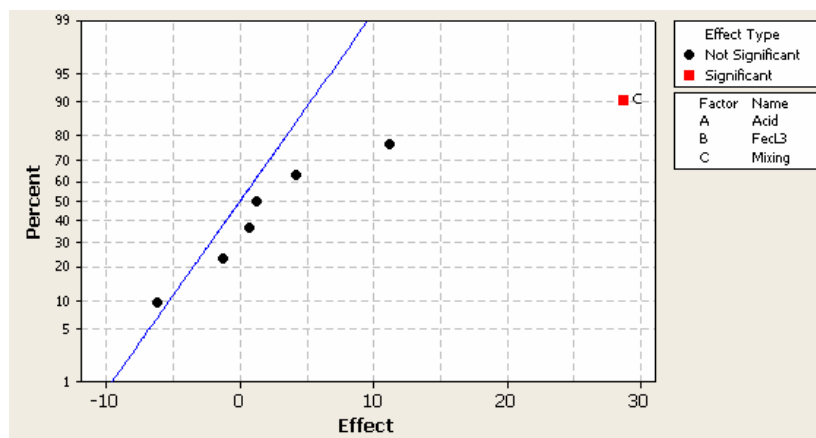
Πίνακας 10.8. Αποτελέσματα πλήρους παραγοντικής ανάλυσης

Παράγοντας	Επίδραση	Συντελεστής
Μέσος όρος		30,625
Acid	-1,25	-0,625
FecL3	-6,25	-3,125
Mixing	28,75	14,375
Acid*FecL3	1,25	0,625
Acid*Mixing	4,25	2,125
FecL3*Mixing	11,25	5,625
Acid*FecL3*Mixing	0,75	0,375

Αφού δεν υπήρξαν επαναλήψεις ανά πείραμα, δεν είναι δυνατό να εξαχθούν διασπορές και τυπικά σφάλματα για κάθε παράμετρο. Από το μέγεθος της τιμής της κάθε επίδρασης (ή προφανώς του συντελεστή) μπορούμε λοιπόν να συμπεράνουμε τη στατιστική σημαντικότητα του κάθε κυρίου παράγοντα ή αλληλεπίδρασης. Βλέπουμε λοιπόν ότι η ανάμειξη είναι ο σημαντικότερος κύριος παράγοντας. Η αλληλεπίδραση του κροκιδωτικού με την ανάμειξη ακολουθεί ως η σημαντικότερη αλληλεπίδραση αντίστοιχα. Το θετικό πρόσημο στο συντελεστή της ανάμειξης φανερώνει απλά ότι η μεταβολή της από το χαμηλό στο υψηλό επίπεδο προκαλεί αύξηση των στερεών, δηλαδή υψηλοί ρυθμοί ανάδευσης δεν ευνοούν την καθίζηση των στερεών. Ο δεύτερος πιο σημαντικός κύριος παράγοντας είναι το κροκιδωτικό άλας του σιδήρου, του οποίου η επίδραση, όπως αναμενόταν, είναι αρνητική. Δηλαδή, μεταβολή της συγκέντρωσης του σιδήρου από το χαμηλό στο υψηλό επίπεδο προκαλεί μείωση των συγκεντρώσεων των στερεών.

Γενικά τα αποτελέσματα της παραπάνω ανάλυσης δεν είναι και τόσο στέρεα. Παραδειγματικά, η αλληλεπίδραση της ανάμειξης με το άλας του σιδήρου αποδεικνύεται περισσότερο σημαντική από την επίδραση του άλατος μόνο. Αυτό γενικά δεν είναι σύνηθες και μάλλον υπονοεί ότι έπρεπε τα πρωτογενή δεδομένα να τροποποιηθούν. Δηλαδή, στατιστικά σημαντικές αλληλεπιδράσεις χωρίς να είναι ιδιαίτερα σημαντικές οι κύριες επιδράσεις από τις οποίες αυτές προέρχονται δεν έχουν βάση.

Το γράφημα πιθανότητας, όπως εξάγεται από το MINITAB, για $\alpha=5\%$ είναι το παρακάτω:



Σχήμα 10.7. Γράφημα πιθανοτήτων 7 επιδράσεων των κυρίων παραγόντων και των αλληλοεπιδράσεων. Στατιστικά σημαντική επίπτωση – σε επίπεδο σημαντικότητας 95% - είναι μόνο ο παράγοντας «Ανάμειξη».

Παρατηρήστε στο παραπάνω γράφημα, ότι οι επιδράσεις έχουν ιεραρχηθεί από την μικρότερη στη μεγαλύτερη και «μεσαία» τιμή, ήτοι η 4^η επίδραση από την αρχή ή το τέλος, έχει τοποθετηθεί στο 50% του άξονα y. Αντιστοιχεί δηλαδή στον μέσο όρο των επτά τιμών και είναι η τιμή 1,25. Από τις παραπάνω τιμές, η επίδραση με $P < 0,025$ είναι τελικά μόνο η επίδραση του παράγοντα 3, δηλ. της *ανάμειξης*. Η επίδραση αυτή, όπως φαίνεται και από το γράφημα, απέχει αρκετά από τις υπόλοιπες τιμές. Συνεπώς, απέχει αρκετά από την καθορισμένη μέση τιμή 1,25 και συνεπώς αν τα δεδομένα τοποθετηθούν στο κλασικό γράφημα της κανονικής κατανομής, όπως στο σχήμα 10.3, τότε το σημείο C θα βρίσκεται αρκετά «δεξιά» στην ουρά της καμπύλης με εμβαδό στο δεξιό τμήμα της καμπύλης που να αντιστοιχεί σε $P < 0.025$.

Είναι ενδιαφέρον να σημειωθεί, ότι σε επίπεδο σημαντικότητας 90% ($\alpha=10\%$), και η αλληλεπίδραση του κροκιδωτικού (εκτός της ανάμειξης) κρίνεται και αυτή ως στατιστικά σημαντική, αφού έχει $P < 0,05$. Το γράφημα των πιθανοτήτων δεν αλλάζει. Δηλαδή, οι επτά επιδράσεις έχουν τις ίδιες τιμές πιθανοτήτων στον άξονα y – και προφανώς στον άξονα x.

Το συνολικό μοντέλο έχει την μορφή:

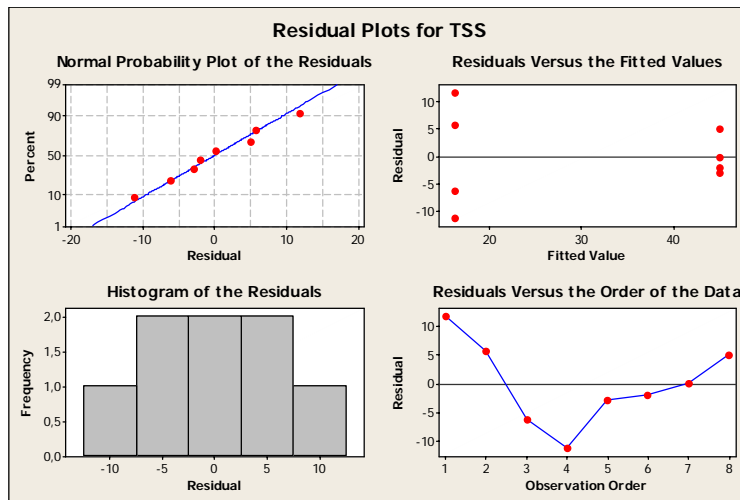
$$Y = 30,626 - 0,625 \cdot X_1 - 3,125 \cdot X_2 + 14,375 \cdot X_3 + 0,625 \cdot X_1 \cdot X_2 + 2,125 \cdot X_1 \cdot X_3 + 5,625 \cdot X_2 \cdot X_3 + 0,375 \cdot X_1 \cdot X_2 \cdot X_3$$

Επειδή είδαμε τελικά ότι μόνο μία κύρια επίδραση είναι στατιστικά σημαντική σε επίπεδο σημαντικότητας 95%, το βέλτιστο μειωμένο μοντέλο (best reduced model) είναι τελικά το:

$$Y = 30,63 + 14,37 \cdot X_3$$

Το παραπάνω προκύπτει αν επιλέξουμε μόνο τον παράγοντα «Ανάμειξη» στο εμπειρικό μοντέλο και στη συνέχεια προχωρήσουμε στη προσαρμογή (fitting) του μοντέλου στα δεδομένα. Το πόσο κατάλληλο είναι το μοντέλο κρίνεται εξάλλου και από τα κλασικά

διαγνωστικά τεστ (γραφήματα των σφαλμάτων) τα οποία οφείλουμε να δούμε. Συγκεκριμένα, τα 4 κλασικά διαγνωστικά τεστ σφαλμάτων – όπως εξάγονται και από το MINITAB – περιλαμβάνονται στο σχήμα 10.8.



Σχήμα 10.8. Διαγνωστικά τεστ σφαλμάτων για το βέλτιστο μειωμένο εμπειρικό μοντέλο στο τεστ κροκίδωσης. Η κανονική κατανομή των σφαλμάτων, με μέση τιμή 0, φαίνεται από τα 2 αριστερά γραφήματα. Μία ελαφριά μείωση των σφαλμάτων όσο μεγαλώνουν οι προβλεπόμενες τιμές φαίνεται στο άνω δεξιό γράφημα, χωρίς όμως η τάση αυτή να χαρακτηρίζει το μοντέλο ως ακατάλληλο. Το δεξί κάτω γράφημα δείχνει ότι τα θετικά και αρνητικά σφάλματα κατανέμονται σχεδόν όμοια. Το βέλτιστο μειωμένο μοντέλο είναι το κατάλληλο

11. Κλασματικά παραγοντικά πειράματα

Τα κλασματικά παραγοντικά πειράματα, από μία άποψη, είναι σημαντικότερα των πλήρων παραγοντικών πειραμάτων, αφού επιτρέπουν με μικρότερο αριθμό πειραμάτων να μελετηθούν οι ίδιοι παράγοντες ενός πλήρους παραγοντικού πειράματος. Παραδειγματικά, αν απαιτούνται $2^4=16$ πειράματα για την πλήρη μελέτη 4 παραγόντων, με ένα κλασματικό πείραμα μπορούμε με 8 μόνο πειράματα να μελετήσουμε τους ίδιους 4 αυτούς παράγοντες. Βέβαια, είναι προφανές, ότι κάποια πληροφορία «θυσιάζεται» με τα κλασματικά πειράματα και όπως θα δούμε στη συνέχεια δεν μπορούμε να ερμηνεύσουμε κάποιες αλληλοεπιδράσεις όταν κάνουμε κλασματικά παραγοντικά πειράματα.

Κλασματικά σημαίνει ότι υλοποιούμε ένα κλάσμα ή μέρος του πλήρους παραγοντικού σχεδιασμού. Συνεπώς, μπορούμε να κάνουμε το $\frac{1}{2}$ ή $\frac{1}{3}$ ή $\frac{1}{4}$ των πειραμάτων. Το $\frac{1}{2}$ σημαίνει ότι μπορούμε να κάνουμε $(1/2)*2^4=16/2=8$ πειράματα για να διερευνήσουμε 4 παράγοντες, ή το $\frac{1}{4}$ σημαίνει ότι μπορούμε να κάνουμε $(1/4)*2^5=8$ πειράματα για να μελετήσουμε 5 παράγοντες.

Ας πάρουμε την περίπτωση που θέλουμε να διερευνήσουμε 4 παράγοντες, αλλά ο προϋπολογισμός μας δεν επιτρέπει τη διεξαγωγή άνω των 8 πειραμάτων, αν και ιδανικά βέβαια θα θέλαμε να κάνουμε $2^4=16$ πειράματα. Το κλασματικό πείραμα με τα μισά πειράματα συμβολίζεται ως 2^{4-1} που ισούται άμεσα με 8.

Γενικά, ο συμβολισμός των κλασματικών πειραματικών σχεδιασμών είναι 2^{k-p} , όπου k είναι ο αριθμός των παραγόντων που θα μπορούσαν να μελετηθούν σε έναν πλήρη παραγοντικό σχεδιασμό και p είναι ο αριθμός των επιπλέον παραγόντων που προσθέτουμε στο πείραμα. Συνεπώς, αν σε ένα 2^3 σχεδιασμό με 8 πειράματα προσθέσουμε 1 επιπλέον παράγοντα διατηρώντας τα ίδια πειράματα, τότε μιλάμε για ένα 2^{4-1} παραγοντικό σχεδιασμό.

Ας πάρουμε το παράδειγμα ενός πλήρους παραγοντικού πειράματος 2^3 . Όταν σχεδιάζουμε ένα κλασματικό πείραμα 2^{4-1} , τότε θα πρέπει ο τέταρτος παράγοντας που θα εισάγουμε στο μοντέλο να εμφανιστεί 4 φορές στο χαμηλό επίπεδο και 4 φορές στο υψηλό επίπεδο, όπως ακριβώς και οι υπόλοιποι 3 παράγοντες. Για να γίνει αυτό, συνηθίζεται ο 4 παράγοντας να «τρέχει» σε ένα από τους συνδυασμούς των αλληλοεπιδράσεων μεταξύ των 3 παραγόντων. Δηλαδή, όλοι οι δυνατοί συνδυασμοί μεταξύ των επιπέδων των 3 παραγόντων δημιουργούν τις 4 αλληλοεπιδράσεις του 2^3 πειράματος. Συνεπώς, ο 4^{ος} παράγοντας που θα εισάγουμε στο μοντέλο θα πρέπει να ταυτιστεί με τις θέσεις των επιπέδων ενός εκ των τεσσάρων αλληλοεπιδράσεων. Συνηθίζεται ο 4^{ος} παράγοντας να ταυτίζεται με τις θέσεις των επιπέδων της τριπλής αλληλοεπίδρασης. Αυτό γίνεται γιατί συνήθως, η 3πλή αλληλοεπίδραση αναμένεται να είναι – και συνήθως είναι – αμελητέα. Το αποτέλεσμα της επίδρασης που θα εξαχθεί για τον 4^ο παράγοντα στην περίπτωση αυτή είναι πρακτικά ένα *άθροισμα* της επίδρασης του 4^{ου} παράγοντας και της μάλλον μικρής έως μηδενικής 3πλής αλληλοεπίδρασης. Δηλαδή, με την παραδοχή που γίνεται δεν είναι ξεκάθαρο ότι η επίδραση του 4^{ου} παράγοντα που εκτιμάται οφείλεται μόνο στον παράγοντα. Αυτό είναι εξάλλου και το τίμημα που

πληρώνουμε αφού θέλουμε με λιγότερα πειράματα να εξάγουμε την ίδια πληροφορία που θα μας έδινε ένα πλήρες 2^3 πείραμα.

Η μήτρα σχεδίασης λοιπόν για ένα 2^{4-1} πείραμα είναι η εξής:

Πίνακας 11.1. Μήτρα σχεδίασης για ένα 2^{4-1} πείραμα

Πείραμα	Παράγοντας 1	Παράγοντας 2	Παράγοντας 3	Παράγοντας 4=1x2x3
1	-	-	-	-
2	+	-	-	+
3	-	+	-	+
4	+	+	-	-
5	-	-	+	+
6	+	-	+	-
7	-	+	+	-
8	+	+	+	+

Συνεπώς, για να καταλήξουμε στον παραπάνω πίνακα, φτιάχνουμε τη σχεδίαση κανονικά για ένα πλήρες 2^3 πείραμα με τους πρώτους 3 παράγοντες και στη συνέχεια υπολογίζουμε τα επίπεδα του 4^{ου} παράγοντα σαν να είναι τα επίπεδα της 3πλης αλληλοεπίδρασης. Δηλαδή, υπολογίζονται τα επίπεδα του 4^{ου} παράγοντα ως το γινόμενο των επιπέδων των 3 παραγόντων ανά πείραμα.

Οι υπόλοιπες διπλές αλληλοεπιδράσεις υπολογίζονται κανονικά, όπως και στο πλήρες 2^3 πείραμα. Προσοχή όμως εδώ. Αφού υπάρχουν 4 παράγοντες, οι διπλές αλληλοεπιδράσεις θα είναι πλέον όχι μόνο 3 αλλά 6. Τι γίνεται λοιπόν με τις επιπλέον 3 διπλές αλληλοεπιδράσεις? Η απάντηση είναι ότι εκτιμώνται τελικά ζευγάρια 2πλών αλληλοεπιδράσεων και η συνολική αλληλοεπίδραση που εκτιμάται αποτελεί τελικά άθροισμα των 2 διπλών αλληλοεπιδράσεων.

Όμως και για την κύρια επίδραση του κάθε παράγοντα δεν έχουμε κάποια ξεκάθαρη εκτίμηση. Αν δούμε τις θέσεις των επιπέδων του παράγοντα 1, δηλ. σε σειρά -,+,-,+,-,+,-,+ θα δούμε ότι είναι ίδιες με τις θέσεις των επιπέδων της 3πλης αλληλοεπίδρασης $2 \times 3 \times 4$. Ομοίως, οι θέσεις των επιπέδων του 2 είναι όμοιες με τη 3πλή αλληλοεπίδραση $1 \times 3 \times 4$. Γενικά όλες οι κύριες επιδράσεις προστίθενται με τις 3πλες αλληλοεπιδράσεις, βάσει της παρακάτω σχέσης:

$$\begin{aligned}
 & \mathbf{1 + 234} \\
 & \mathbf{2 + 134} \\
 & \mathbf{3 + 124} \\
 & \mathbf{4 + 123}
 \end{aligned}$$

Συνεπώς, η επίδραση που υπολογίζουμε για κάποιον κύριο παράγοντα προστίθεται στην (ή μερδεύεται με την) τριπλή αλληλοεπίδραση των υπολοίπων τριών παραγόντων. Το μέγεθος της τριπλής αυτής αλληλοεπίδρασης – δηλαδή το πόσο «συνεισφέρει» στην επίδραση του κυρίου παράγοντα - δεν θα το μάθουμε ποτέ. Υποθέτουμε – και ελπίζουμε

– ότι η τριπλή αυτή αλληλοεπίδραση είναι μικρή έως μηδαμινή και συνεπώς η επίδραση που υπολογίζεται είναι μόνο η κύρια επίδραση. Στην πράξη αλληλοεπιδράσεις άνω των 2 παραγόντων είναι γενικά πολύ μικρές και μπορούν συνεπώς να μην λαμβάνονται υπόψη. Σε αυτό εξάλλου στηρίζεται και η λογική των κλασματικών παραγοντικών σχεδιασμών. Δηλαδή στο ότι θυσιάζεται κάποια πολλαπλή αλληλοεπίδραση για να υπολογιστεί ένας επιπλέον κύριος παράγοντας.

Γενικά, με βάση τα παραπάνω, ορίζεται μία σχέση ορισμού που βοηθάει στην εύρεση των συνδυασμένων επιδράσεων και αλληλοεπιδράσεων. Για ένα 2^{4-1} πείραμα, η σχέση αυτή είναι η $I=1234$. Η σχέση αυτή σημαίνει ότι αν πολλαπλασιαστούν τα επίπεδα (ή πρόσημα) των παραγόντων 1,2,3,4 για κάθε πείραμα, το αποτέλεσμα θα ισούται με +1. Επίσης, αν πολλαπλασιαστούν τα πρόσημα των 1,2,3 για κάθε πείραμα, το αποτέλεσμα θα δίνει τα επίπεδα του παράγοντα 4. Ή, αν πολλαπλασιαστούν τα πρόσημα των 2,3,4, το αποτέλεσμα αντιπροσωπεύει το αντίστοιχο επίπεδο του παράγοντα 1 για το κάθε πείραμα. Το παραπάνω σημαίνει αυτό ακριβώς που έχουμε πει, δηλαδή ότι η επίδραση του κάθε κύριου παράγοντα προστίθεται (ή από μία άποψη μπερδεύεται) με μία τριπλή αλληλοεπίδραση.

Ομοίως έχουμε και συνδυασμένες διπλές αλληλοεπιδράσεις. Βάσει της σχέσης ορισμού I , η διπλή αλληλοεπίδραση των 1,2 προστίθεται στην διπλή αλληλοεπίδραση των 3,4. Αυτό είναι ένα ακόμα αρνητικό της κλασματικής παραγοντικής ανάλυσης, δηλαδή η μη ξεκάθαρη εκτίμηση των διπλών αλληλοεπιδράσεων. Συνεπώς, όταν λαμβάνουμε μία τιμή για την αλληλοεπίδραση των 1,3 δεν γνωρίζουμε σε τι βαθμό «συμμετέχει» στην επίδραση αυτή και η διπλή επίδραση 2,4. Έτσι για τις διπλές αλληλοεπιδράσεις ισχύει:

$$\begin{aligned} 12 + 34 \\ 13 + 24 \\ 14 + 23 \end{aligned}$$

Γενικά, στις κλασματικές παραγοντικές σχεδιάσεις ελπίζουμε ότι οι τριπλές αλληλοεπιδράσεις είναι μηδαμινές, διότι μόνο έτσι μπορούμε να εκτιμήσουμε σωστά τις κύριες επιδράσεις. Στην περίπτωση των διπλών αλληλοεπιδράσεων η κατάσταση είναι ακόμα «χειρότερη» (!), αφού δεν μπορούμε να ξεχωρίσουμε το μέγεθος των διπλών αλληλοεπιδράσεων στα διάφορα ζεύγη. Συνεπώς, μία πιθανά μεγάλη τιμή για την αλληλοεπίδραση 12 δεν γνωρίζουμε αν οφείλεται μερικώς (ή πλήρως) στην διπλή αλληλοεπίδραση 34.

Η μήτρα μοντελοποίησης τελικά για ένα 2^{4-1} πείραμα είναι η εξής:

Πίνακας 11.2. Μήτρα μοντελοποίησης για ένα 2^{4-1} πείραμα

Πείραμα	Μέση τιμή	Παράγοντας 1=234	Παράγοντας 2=134	Παράγοντας 3=124	Παράγοντας 4 = 123	12=34	13=24	14=23
1	+	-	-	-	-	+	+	+
2	+	+	-	-	+	-	-	+
3	+	-	+	-	+	-	+	-
4	+	+	+	-	-	+	-	-
5	+	-	-	+	+	+	-	-
6	+	+	-	+	-	-	+	-
7	+	-	+	+	-	-	-	+
8	+	+	+	+	+	+	+	+

Θα μπορούσαμε επίσης με 8 πειράματα να σχεδιάσουμε μία 2^{5-2} πειραματική σχεδίαση. Εδώ μιλάμε για σχεδίαση του $\frac{1}{4}$ αφού θα προσπαθούσαμε να διερευνήσουμε 5 παράγοντες με μόνο $\frac{1}{4} \times 2^5 = \frac{1}{4} \times 32 = 8$ πειράματα. Στη περίπτωση αυτή θα «θυσιάζαμε» ακόμα περισσότερες αλληλοεπιδράσεις σε σχέση με τη 2^{4-1} σχεδίαση. Η μήτρα σχεδίασης του 2^{5-2} πειράματος φαίνεται στη συνέχεια. Εδώ θα πρέπει να ταυτιστούν και ο $4^{ος}$ και ο $5^{ος}$ παράγοντας με μία διπλή αλληλοεπίδραση ξεχωριστά. Συγκεκριμένα, ο $4^{ος}$ παράγοντας θα ταυτιστεί με την αλληλοεπίδραση 12 και ο $5^{ος}$ παράγοντας με την αλληλοεπίδραση 13. Δηλαδή, $4=12$ και συνεπώς $4 \times 4 = 12 \times 4$ ή $42=124$ ή $I=124$. Όμοια, $5=13$ ή $5 \times 5 = 13 \times 5$ ή $52=135$ ή $I=135$. Δηλαδή, η σχέση ορισμού I στην περίπτωση αυτή είναι η **$I=124=135=2345$** .

Πίνακας 11.3 Μήτρα σχεδίασης για ένα 2^{5-2} πείραμα (με $I=124=135$)

Πείραμα	Παράγοντας 1	Παράγοντας 2	Παράγοντας 3	Παράγοντας 4=1x2	Παράγοντας 5=1x3
1	-	-	-	+	+
2	+	-	-	-	-
3	-	+	-	-	+
4	+	+	-	+	-
5	-	-	+	+	-
6	+	-	+	-	+
7	-	+	+	-	-
8	+	+	+	+	+

Η πληροφορία που παίρνουμε τώρα έχει ακόμα μεγαλύτερη αβεβαιότητα αφού θυσιάζονται και διπλές αλληλοεπιδράσεις. Η έννοια «θυσία» έχει να κάνει φυσικά με το ότι θα πρέπει να αγνοηθούν οι συγκεκριμένες διπλές αλληλοεπιδράσεις ώστε να μπορούμε να πούμε ότι η υπολογισμένη επίδραση του παράγοντα 4 και 5 οφείλεται **αποκλειστικά** στους εν λόγω κύριους παράγοντες. Η μήτρα μοντελοποίησης θα είχε τελικά την μορφή του πίνακα 11.4. Οι ταυτίσεις των επιδράσεων / αλληλοεπιδράσεων στην περίπτωση αυτή είναι:

$$\begin{aligned}
 & \mathbf{1+24+35+12345}^{28} \\
 & \mathbf{2+14+345+1235} \\
 & \mathbf{3+15+245+1234} \\
 & \mathbf{4+12+235+1345} \\
 & \mathbf{5+13+234+1245} \\
 & \mathbf{23+45+125+134} \\
 & \mathbf{25+34+123+145}
 \end{aligned}$$

Μπορούμε να δούμε ότι ισχύει η σχέση $1 \times 1 = 1^2 = I$, $2 \times 2 = 2^2 = I$, $3 \times 3 = 3^2 = I$, $4 \times 4 = 4^2 = I$, $5 \times 5 = 5^2 = I$. Δηλαδή, $2 \times 4 = 2 \times (1 \times 2) = 1 \times I = 1$ ή $3 \times 5 = 3 \times 1 \times 3 = 1 \times I = 1$. Δηλαδή, η σχέση ορισμού I στην περίπτωση εδώ είναι η $I = 124 = 135 = 2345$.

Στην περίπτωση ενός κλασματικού πειράματος της τάξης του $\frac{1}{2}$ με 5 παράγοντες ή 2^{5-1} πείραμα, μιλάμε προφανώς για τη διερεύνηση 5 παραγόντων με $2^{5-1} = 2^4 = 16$ πειράματα αντί για 32 όπως θα απαιτείτο σε ένα πλήρες παραγοντικό πείραμα των 5 παραγόντων σε 2 επίπεδα. Η τιμή I εδώ είναι η 12345 και συνεπώς ταυτίζεται ο παράγοντας 5 με την τετραπλή αλληλοεπίδραση $1 \times 2 \times 3 \times 4$. Θυσιάζουμε δηλαδή την πληροφορία για την τετραπλή αλληλοεπίδραση (και καλά κάνουμε για αλληλοεπίδραση τέτοιου μεγάλου βαθμού είναι μάλλον μηδαμινή) με τον 5^ο παράγοντα. Συνεπώς, $5 = 1234$ ή $5 \times 5 = 1234 \times 5$ ή $5^2 = 12345$ ή $I = 12345$.

Με βάση τη σχέση ορισμού $I = 12345$ μπορούμε να δούμε όλους τους συνδυασμούς μεταξύ κυρίων επιδράσεων και αλληλοεπιδράσεων, ήτοι:

$$\begin{aligned}
 & \mathbf{1+2345} \\
 & \mathbf{2+1345} \\
 & \mathbf{3+1245} \\
 & \mathbf{4+1235} \\
 & \mathbf{5+1234} \\
 & \mathbf{12+345} \\
 & \mathbf{13+245} \\
 & \mathbf{14+235} \\
 & \mathbf{15+234} \\
 & \mathbf{23+145} \\
 & \mathbf{24+135} \\
 & \mathbf{25+134} \\
 & \mathbf{34+125} \\
 & \mathbf{35+124} \\
 & \mathbf{45+123}
 \end{aligned}$$

Οι παραπάνω συνδυασμοί εξάγονται βάσει της ιδιότητας ότι 1^2 ή 2^2 ή 3^2 κ.ο.κ. = I . Άρα, $I = 12345$ ή $I \times 1 = 12345 \times 1 = (1 \times 1) \times 2345$ ή $I \times 1 = I \times 2345$ ή $1 = 2345$. Συνεπώς με τον τρόπο αυτό μπορούμε να βρούμε τις επιδράσεις και αλληλοεπιδράσεις που ταυτίζονται. Άρα, παραδειγματικά, η αλληλοεπίδραση 2×5 επηρεάζεται από (ή εμπεριέχει την) τριπλή αλληλοεπίδραση $1 \times 3 \times 4$. Αν η τελευταία είναι αμελητέα – και μάλλον είναι – τότε η τιμή της επίδρασης που υπολογίζουμε οφείλεται αποκλειστικά στην διπλή αλληλοεπίδραση 25.

²⁸ Αυτό σημαίνει ότι η κύρια επίδραση του παράγοντα 1 εμπεριέχει τις αλληλοεπιδράσεις 2×4 , 3×5 , $1 \times 2 \times 3 \times 4 \times 5$. Δεν μπορούν να υπολογιστούν ξεχωριστά οι παραπάνω επιδράσεις και αλληλοεπιδράσεις.

Πίνακας 11.4 Μήτρα μοντελοποίησης για ένα 2^{5-2} πείραμα (οι 4πλές αλληλοεπιδράσεις δεν περιλαμβάνονται)

Πείραμα	Μέση τιμή	Παράγοντας 1=2x4=3x5	Παράγοντας 2=1x4=3x4x5	Παράγοντας 3=1x5=2x4x5	Παράγοντας 4 = 1x2=2x3x5	Παράγοντας 5=1x3=2x3x4	2x3=4x5= 1x2x5=1x3x4	2x5=3x4=1x2x 3=1x4x5
1	+	-	-	-	-	+	+	+
2	+	+	-	-	+	-	-	+
3	+	-	+	-	+	-	+	-
4	+	+	+	-	-	+	-	-
5	+	-	-	+	+	+	-	-
6	+	+	-	+	-	-	+	-
7	+	-	+	+	-	-	-	+
8	+	+	+	+	+	+	+	+

Η μήτρα σχεδιασμού για το 2^{5-1} πείραμα βασίζεται πάλι στο ότι αναπτύσσουμε ένα πλήρη σχεδιασμό για τους 4 πρώτους παράγοντες και τον 5^ο παράγοντα τον ταυτίζουμε με την 4πλη αλληλοεπίδραση 1234 (πίνακας 11.5).

Πίνακας 11.5. Μήτρα σχεδίασης για ένα 2^{5-1} πείραμα (I=12345)

Πείραμα	Παράγοντας 1	Παράγοντας 2	Παράγοντας 3	Παράγοντας 4	Παράγοντας 5 =1x2x3x4
1	-	-	-	-	+
2	+	-	-	-	-
3	-	+	-	-	-
4	+	+	-	-	+
5	-	-	+	-	-
6	+	-	+	-	+
7	-	+	+	-	+
8	+	+	+	-	-
9	-	-	-	+	-
10	+	-	-	+	+
11	-	+	-	+	+
12	+	+	-	+	-
13	-	-	+	+	+
14	+	-	+	+	-
15	-	+	+	+	-
16	+	+	+	+	+

Είναι προφανές, κοιτώντας τον πίνακα 11.5, ότι πρέπει να υπάρχουν 8 υψηλά και 8 χαμηλά επίπεδα του 5 παράγοντα, όπως φυσικά και με όλους τους παράγοντες, ώστε να έχουμε «δίκαια» αποτελέσματα για όλους τους 5 παράγοντες.

Ο πίνακας 11.6 δείχνει την αντίστοιχη μήτρα μοντελοποίησης για ένα 2^{5-1} πείραμα.

Πίνακας 11.6 Μήτρα μοντελοποίησης για ένα 2^{5-1} πείραμα

Πείραμα	Μέση τιμή	1 =2345	2 =1345	3 =1245	4 =1235	5 =1234	12 =345	13 =245	14 =235	15 =234	23 =145	24 =135	25 =134	34 =125	35 =124	45 =123
1	+	-	-	-	-	+	+	+	+	-	+	+	-	+	-	-
2	+	+	-	-	-	-	-	-	-	-	+	+	+	+	+	+
3	+	-	+	-	-	-	-	+	+	+	-	-	-	+	+	+
4	+	+	+	-	-	+	+	-	-	+	-	-	+	+	-	-
5	+	-	-	+	-	-	+	-	+	+	-	+	+	-	-	+
6	+	+	-	+	-	+	-	+	-	+	-	+	-	-	+	-
7	+	-	+	+	-	+	-	-	+	-	+	-	+	-	+	-
8	+	+	+	+	-	-	+	+	-	-	+	-	-	-	-	+
9	+	-	-	-	+	-	+	+	-	+	+	-	+	-	+	-
10	+	+	-	-	+	+	-	-	+	+	+	-	-	-	-	+
11	+	-	+	-	+	+	-	+	-	-	-	+	+	-	-	+
12	+	+	+	-	+	-	+	-	+	-	-	+	-	-	+	-
13	+	-	-	+	+	+	+	-	-	-	-	-	-	+	+	+
14	+	+	-	+	+	-	-	+	+	-	-	-	+	+	-	-
15	+	-	+	+	+	-	-	-	-	+	+	+	-	+	-	-
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Στον πίνακα αυτόν αναγράφονται όλες οι κύριες επιδράσεις και διπλές αλληλοεπιδράσεις καθώς και οι πολλαπλές αλληλοεπιδράσεις από τις οποίες επηρεάζονται οι παραπάνω επιδράσεις.

Παράδειγμα

Ας μελετήσουμε την περίπτωση ενός πλήρους παραγοντικού πειράματος με 5 παράγοντες (και συνεπώς 32 πειράματα) και ενός κλασματικού παραγοντικού πειράματος που μελετάει τους ίδιους 5 παράγοντες αλλά με το μισό αριθμό πειραμάτων, ήτοι 16. Στη συγκεκριμένη περίπτωση μελετάμε την επίδραση 5 παραγόντων στην απόδοση μίας χημικής αντίδρασης. Η απόκριση που μετράμε είναι το ποσοστό αντιδρώντων που μετατρέπονται στην αντίδραση (% αντιδρώντων).

Ας υποθέσουμε ότι οι 5 παράγοντες που μελετάμε είναι οι εξής:

	Χαμηλό επίπεδο (-)	Υψηλό επίπεδο (+)
Ρυθμός προσθήκης χημικών (L/min)	10	15
Καταλύτης (%)	1	2
Ρυθμός ανάδευσης (rpm)	100	120
Θερμοκρασία (°C)	140	180
Συγκέντρωση (%)	3	6

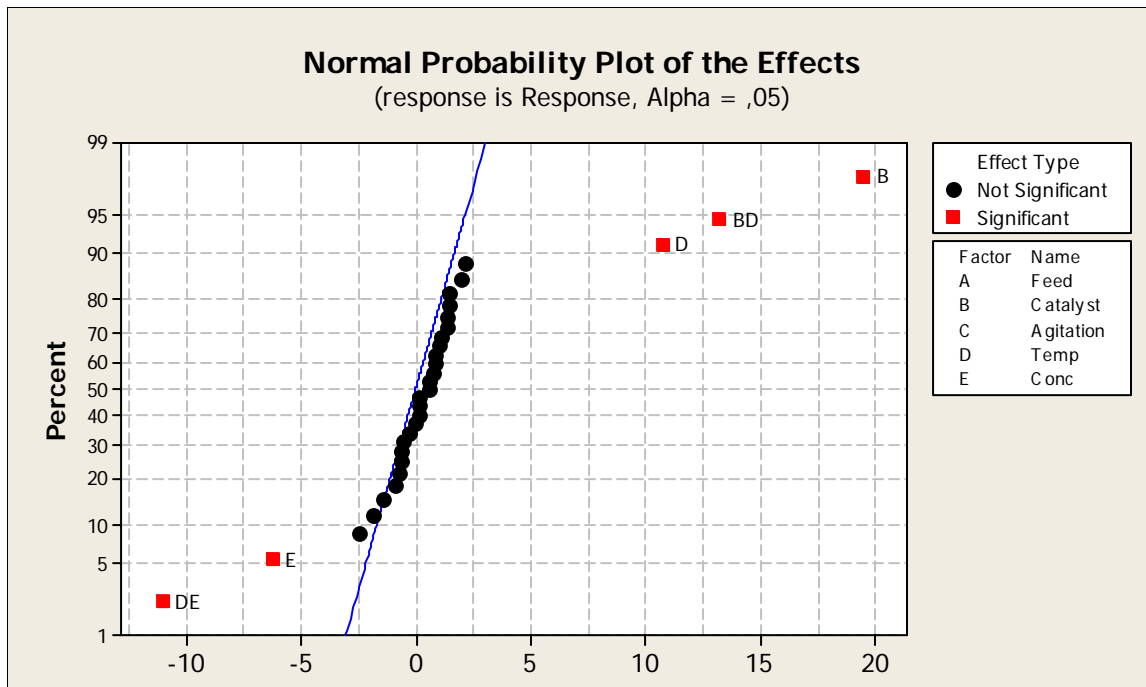
Τα αποτελέσματα του πλήρους παραγοντικού σχεδιασμού περιλαμβάνονται στον πίνακα 11.7.

Πίνακας 11.7 Αποτελέσματα πλήρους παραγοντικού σχεδιασμού 2⁵

Πείραμα	Παράγοντας 1	Παράγοντας 2	Παράγοντας 3	Παράγοντας 4	Παράγοντας 5	Απόκριση (% χημικών που αντιδρούν)
1	-	-	-	-	-	61
2	+	-	-	-	-	53
3	-	+	-	-	-	63
4	+	+	-	-	-	61
5	-	-	+	-	-	53
6	+	-	+	-	-	56
7	-	+	+	-	-	54
8	+	+	+	-	-	61
9	-	-	-	+	-	69
10	+	-	-	+	-	61
11	-	+	-	+	-	94
12	+	+	-	+	-	93
13	-	-	+	+	-	66
14	+	-	+	+	-	60
15	-	+	+	+	-	95
16	+	+	+	+	-	98
17	-	-	-	-	+	56
18	+	-	-	-	+	63
19	-	+	-	-	+	70
20	+	+	-	-	+	65
21	-	-	+	-	+	59
22	+	-	+	-	+	55
23	-	+	+	-	+	67
24	+	+	+	-	+	65
25	-	-	-	+	+	44
26	+	-	-	+	+	45
27	-	+	-	+	+	78
28	+	+	-	+	+	77
29	-	-	+	+	+	49
30	+	-	+	+	+	42
31	-	+	+	+	+	81
32	+	+	+	+	+	82

Τα πειράματα θεωρείται ότι έχουν γίνει με τυχαία σειρά και όχι με τη σειρά που αναφέρεται στη στήλη 1, που είναι και η κανονική σειρά (standard order).

Το χαρακτηριστικό γράφημα πιθανοτήτων των επιδράσεων του παραπάνω πειράματος φαίνεται στο σχήμα 11.1



Σχήμα 11.1 Γράφημα ελέγχου κανονικότητας των επιδράσεων του πειράματος 2^5 όπως εξάγεται από το MINITAB 14®

Βάσει του σχήματος 11.1 βλέπουμε ότι 3 κύριοι παράγοντες (καταλύτης, θερμοκρασία, Συγκέντρωση) και οι διπλές αλληλοεπιδράσεις (θερμοκρασία x συγκέντρωση) και (καταλύτης x θερμοκρασία) είναι οι στατιστικά σημαντικές επιδράσεις. Όλες οι τριπλές και τετραπλές αλληλοεπιδράσεις είναι στατιστικά μη σημαντικές, όπως εξάλλου αναμενόταν, αφού όπως είπαμε σπάνια οι πολλαπλές αλληλοεπιδράσεις είναι σημαντικές. Οι τιμές των επιδράσεων φαίνονται παρακάτω:

Μέση τιμή = 65,5

1=	-1,375	123=	1,50
2=	19,5	124=	1,375
3=	-0,625	125=	-1,875
4=	10,75	134=	-0,75
5=	-6,25	135=	-2,50
		145=	0,625
12=	1,375	235=	0,125
13=	0,75	234=	1,125
14=	0,875	245=	-0,250
15=	0,125	345=	0,125
23=	0,875	1234=	0,0
24=	13,25	1245=	0,625
25=	2,0	2345=	-0,625
34=	2,125	1235=	1,5
35=	0,875	1345=	1,0
45=	-11,0	12345=	-0,50

Ας υποθέσουμε λοιπόν ότι έχουμε μόνο 16 πειράματα που μπορούμε να υλοποιήσουμε για να μελετήσουμε τους ίδιους 5 παράγοντες. Θεωρώντας I=12345 και χρησιμοποιώντας τη μήτρα σχεδιασμού που περιγράφηκε παραπάνω (δηλ. θέτοντας τον παράγοντα 5 στα ίδια επίπεδα με τον 4πλη αλληλοεπίδραση 1234), τα αποτελέσματα είναι ως εξής:

Πίνακας 11.8 Αποτελέσματα κλασματικού σχεδιασμού 2^{5-1}

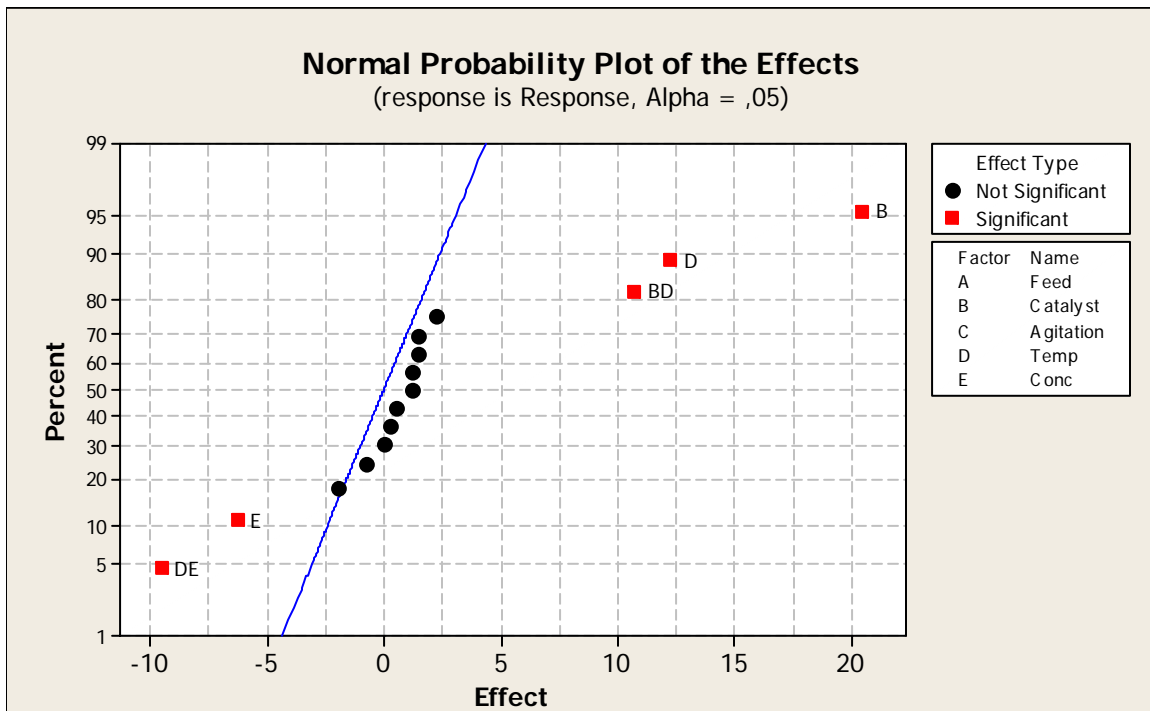
Πείραμα	Παράγοντας 1	Παράγοντας 2	Παράγοντας 3	Παράγοντας 4	Παράγοντας 5=1234	Απόκριση (% χημικών που αντιδρούν)
17	-	-	-	-	+	56
2	+	-	-	-	-	53
3	-	+	-	-	-	63
20	+	+	-	-	+	65
5	-	-	+	-	-	53
22	+	-	+	-	+	55
23	-	+	+	-	+	67
8	+	+	+	-	-	61
9	-	-	-	+	-	69
26	+	-	-	+	+	45
27	-	+	-	+	+	78
12	+	+	-	+	-	93
29	-	-	+	+	+	49
14	+	-	+	+	-	60
15	-	+	+	+	-	95
32	+	+	+	+	+	82

Τα αποτελέσματα των επιδράσεων είναι ως εξής (θεωρούμε ότι οι 3πλές και ανώτερες αλληλοεπιδράσεις είναι μηδαμινές):

Μέση τιμή=65,25

1=-2,0	12= 1,5
2=20,5	13= 0,5
3=0,0	14= -0,75
4=12,25	15= 1,25
5=-6,25	23= 1,50
	24= 10,75
	25= 1,25
	34= 0,25
	35= 2,25
	45= -9,50

Το σχετικό γράφημα σε κλίμακα πιθανοτήτων είναι το παρακάτω:



Σχήμα 11.2 Γράφημα ελέγχου κανονικότητας των επιδράσεων του πειράματος 2^{5-1} όπως εξάγεται από το MINITAB 14®

Από το σχήμα 11.2 μπορούμε να δούμε ότι εξάγονται τα ίδια ακριβώς συμπεράσματα με το πλήρες 2^5 παραγοντικό πείραμα. Δηλαδή, οι ίδιες 3 κύριες επιδράσεις και οι ίδιες 2 διπλές αλληλοεπιδράσεις είναι στατιστικά σημαντικές. Συνεπώς, η ίδια πληροφορία μπορεί να εξαχθεί με το μισό των συνολικών πειραμάτων, ήτοι 16 αντί για 32. Κάτι τέτοιο σίγουρα μειώνει χρόνο και το συνολικό κόστος του πειράματος.

Συμπερασματικά μπορούμε να πούμε ότι τα κλασματικά παραγοντικά πειράματα είναι περισσότερο χρήσιμα από τα πλήρη, αφού σπάνια οι πολλαπλές αλληλοεπιδράσεις είναι σημαντικές. Συνεπώς, αξίζει να «θυσιάσουμε» τις αλληλοεπιδράσεις αυτές για να εισάγουμε και μελετήσουμε επιπλέον παράγοντες εξάγοντας την ίδια πληροφορία με λιγότερο αριθμό πειραμάτων σε σχέση με τα αντίστοιχα πλήρη παραγοντικά πειράματα.

12. Πειράματα με μίγματα

Τα πειράματα με μίγματα βασίζονται στη χρήση μιγμάτων από συγκεκριμένα συστατικά που αναμιγνύονται σε γνωστά ποσοστά στο μίγμα. Παράδειγμα, τα καύσιμα αποτελούνται από μίγματα πετρελαίου και άλλων πρόσθετων ή τα αστικά στερεά απόβλητα είναι τυπικό παράδειγμα μίγματος συστατικών αποβλήτων, όπως χαρτί, υπολείμματα κήπου, υπολείμματα φαγητών κ.λ.π.

Κάθε πείραμα διεξάγεται με ένα μίγμα, που εμείς ορίζουμε από πριν. Π.χ. γνωρίζουμε ότι ένα μίγμα μας 3 συστατικών αστικών στερεών αποβλήτων περιέχει 30% χαρτί, 50% φαγητά, 20% πλαστικά. Τα ποσοστά των 3 συστατικών πρέπει να αθροίζουν απαραίτητα στο 100% (δηλ. στο 1). Ένα μίγμα μπορεί να ταυτιστεί και με ένα συστατικό των αποβλήτων, δηλαδή ένα «μίγμα» να περιέχει 100% χαρτί.

Η γραμμική εξίσωση πρώτου βαθμού που περιγράφει ένα πείραμα με μίγματα είναι:

$$\eta = \beta_0 + \beta_1 \cdot F_1 + \beta_2 \cdot F_2 + \dots + \beta_n \cdot F_n$$

$$Y = \beta_0 + \beta_1 \cdot F_1 + \beta_2 \cdot F_2 + \dots + \beta_n \cdot F_n + \varepsilon$$

η : πραγματική τιμή της απόκρισης (π.χ. σε μονάδες gr / kg)

Y : η τιμή που μετράμε στο πείραμά μας (ίδιες μονάδες, δηλ. gr/kg)

$\beta_0, \beta_1, \dots, \beta_n$: πραγματικές τιμές των συντελεστών των μεταβλητών του μοντέλου

F_1, F_2, \dots, F_n : τα ποσοστά των συστατικών 1,2,...,n, αντίστοιχα στο μίγμα. Το n είναι ο αριθμός των συστατικών στο μίγμα. Προφανώς, οι τιμές των F_1, F_2, \dots, F_n κυμαίνονται από 0 έως 1. Δηλαδή, $0 \leq F_1, F_2, \dots, F_n \leq 1$. Το χαρακτηριστικό όμως (αλλά και η απαίτηση) της παραπάνω εξίσωσης μίγματος – που αποτελεί τη βασική προϋπόθεση κατά την ανάπτυξη μοντέλων σε πειράματα με μίγματα – είναι ότι τα ποσοστά F πρέπει να αθροίζουν σε 1 (δηλ. $F_1 + F_2 + \dots + F_n = 1$).

ε : τυχαίο πειραματικό σφάλμα. Τα σφάλματα αυτά φυσιολογικώς πρέπει να κατανέμονται κανονικά, να είναι ανεξάρτητα και να έχουν σταθερή διασπορά.

Το μοντέλο μας τελικά θα έχει τη μορφή:

$$\hat{Y} = b_0 + b_1 \cdot F_1 + b_2 \cdot F_2 + \dots + b_n \cdot F_n$$

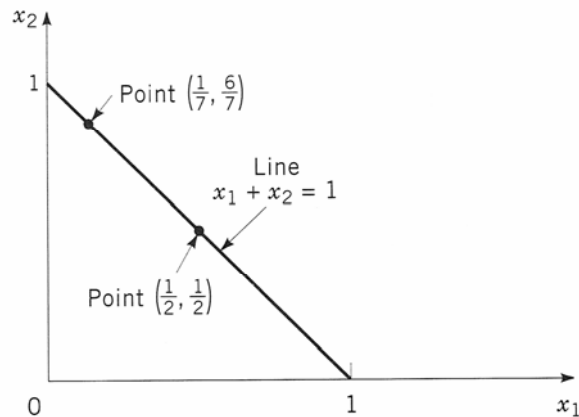
\hat{Y} : είναι η τιμή που εξάγει (ή προβλέπει) το συγκεκριμένο μοντέλο μας (π.χ. σε μονάδες gr / kg)

$b_0, b_1, b_2, \dots, b_n$: συντελεστές της εξίσωσης (εκτιμήσεις των πραγματικών τιμών)

12.1. Μίγμα με δύο συστατικά

Παραδειγματικά, στην περίπτωση των 2 συστατικών, $n=2$ και $F_1+F_2=1$. Αυτό ορίζεται από μία ευθεία γραμμή στο διάστημα (x_1, x_2) . Όλα τα σημεία επί της γραμμής που φαίνεται στο σχήμα 12.1 ορίζουν το μονοδιάστατο χώρο (ευθεία) του μίγματος. Τα σημεία συμπεριλαμβάνουν και τα ακραία σημεία (0,1) και (1,0). Όλα τα σημεία επί της γραμμής παίρνουν τη μορφή $(x_1, 1-x_1)$ ή $(1-x_2, x_2)$. Παράδειγμα, αν x_1 είναι το ποσοστό του ούισκι σε ένα κοκτέιλ και x_2 είναι το ποσοστό της σόδας στο κοκτέιλ, όλοι οι δυνατοί συνδυασμοί θα πέφτουν πάνω στη γραμμή που περιγράφεται. Είναι σημαντικό

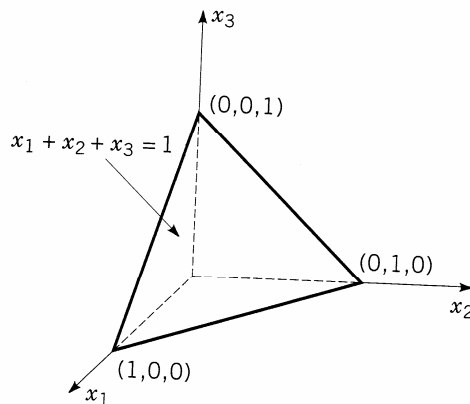
να τονιστεί ότι και τα ποτά που είναι μόνο ούισκι ή μόνο σόδα, δηλ. τα σημεία 1,0 και 0,1, θεωρούνται επίσης ως «μίγματα» και απλώς ονομάζονται «καθαρά μίγματα».



Σχήμα 12.1. Απεικόνιση πειράματος μιγμάτων με 2 συστατικά σε μία γραμμή (από Draper & Smith, 1998)

12.2. Μίγμα με τρία συστατικά

Όταν $x_1+x_2+x_3=1$, το διάστημα του μίγματος καθορίζεται από μία τριγωνική επιφάνεια, οι κορυφές της οποίας περιέχουν τα 3 ακραία σημεία των μιγμάτων, που είναι $(x_1,x_2,x_3) = (1,0,0)$, $(0,1,0)$ και $(0,0,1)$ και $0 \leq x_i \leq 1$. Η επιφάνεια ορίζεται λοιπόν από ένα ισόπλευρο τρίγωνο, στο οποίο η κάθε πλευρά είναι ένα υπο-διάστημα 2 συστατικών, και το τρίτο συστατικό είναι μηδέν.



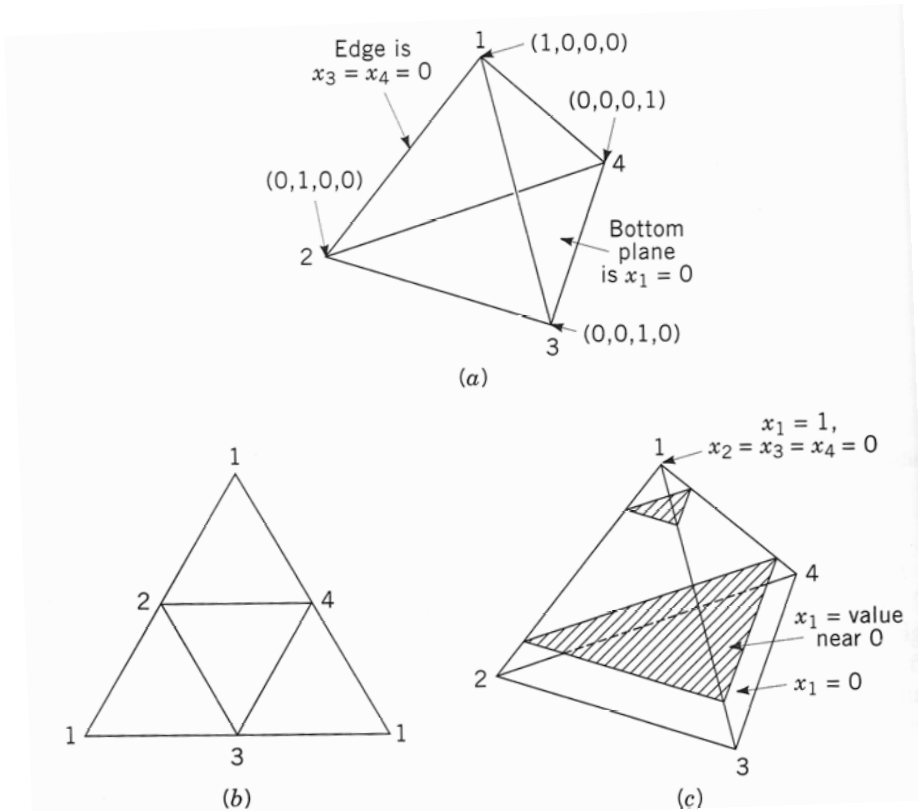
Σχήμα 12.2. Απεικόνιση πειράματος μιγμάτων με 3 συστατικά σε επιφάνεια (από Draper & Smith, 1998)

Το σχήμα 12.2. δείχνει πως μπορούμε να εντοπίσουμε ένα σημείο εντός του τριγώνου. Το σημείο εντός του τριγώνου στο παρακάτω σχήμα είναι το σημείο $(0,35, 0,45, 0,20)$, που σημαίνει ότι το x_1 είναι σε ποσοστό 35%, x_2 σε ποσοστό 45%, και x_3 σε ποσοστό 20% στο μίγμα. Ο άξονας που είναι κάθετος στην οριζόντια γραμμή του τριγώνου ορίζει τις συντεταγμένες του x_1 (με τιμή 0 στο κατώτατο σημείο και 1 στο ανώτατο σημείο). Ό

άξονας που είναι κάθετος στην δεξιά πλευρά του τριγώνου ορίζει τις συντεταγμένες του x_2 (με τιμή 0 στο κατώτατο-δεξιό σημείο και 1 στο ανώτατο-αριστερό σημείο). Ο άξονας που είναι κάθετος στην αριστερά πλευρά του τριγώνου ορίζει τις συντεταγμένες του x_3 (με τιμή 0 στο κατώτατο-αριστερό σημείο και 1 στο ανώτατο-δεξιό σημείο).

12.3. Μίγμα με τέσσερα συστατικά

Τα 4 συστατικά είναι και αυτά δυνατό να τα απεικονίσουμε γραφικά. Από τα 5 συστατικά και άνω αυτό δεν είναι δυνατό. Τα σημεία που ορίζουν το μίγμα των τεσσάρων συστατικών βρίσκεται εντός ή στις επιφάνειες του ισόπλευρου τετραέδρου του σχήματος 12.3.



Σχήμα 12.3. Απεικόνιση πειράματος μιγμάτων με 4 συστατικά σε τρισδιάστατο χώρο (από Draper & Smith, 1998)

Τα παραδείγματα στη συνέχεια θα βασιστούν σε μίγματα από 3 συστατικά, τα οποία παρουσιάζονται σε γράφημα σχετικά εύκολα. Το χαρακτηριστικό (και αξιόλογο) με τα μίγματα 3 συστατικών είναι ότι εντός του τριγώνου μπορούν να σχεδιαστούν καμπύλες ίδιας τιμής (ισοπληθείς καμπύλες), που τελικά απεικονίζουν την επιφάνεια απόκρισης (response surface). Συνεπώς ο τρισδιάστατος χώρος που απαιτείται για την απεικόνιση των μιγμάτων με 3 συστατικά θα απεικονιστεί σε ένα δισδιάστατο χώρο με χρήση ισοπληθών καμπύλων.

12.4 Μοντέλα πειραμάτων μιγμάτων με 3 συστατικά

Το βασικό μοντέλο γραμμικό (πρώτου βαθμού) με μίγματα έχει τη μορφή:

$$\eta = \beta_0 + \beta_1 \cdot F_1 + \beta_2 \cdot F_2 + \beta_3 \cdot F_3$$

Το οποίο προσεγγίζεται από την γραμμική (σε αυτή την περίπτωση) εξίσωση:

$$Y = \beta_0 + \beta_1 \cdot F_1 + \beta_2 \cdot F_2 + \beta_3 \cdot F_3 + \varepsilon \quad (1)$$

Όπου β είναι οι πραγματικές των παραμέτρων στο μοντέλο, ενώ Y είναι η τιμή της παρατήρησης μας από το πείραμα. Το η εξάλλου είναι η πραγματική τιμή της απόκρισης, που πιθανά να μην μάθουμε ποτέ. Προφανώς, ε είναι το τυχαίο πειραματικό σφάλμα.

Βάσει του μοντέλου, η τιμή που εξάγει το μοντέλο (\hat{Y}) βάσει των εκτιμώμενων παραμέτρων b συμβολίζεται με:

$$\hat{Y} = b_0 + b_1 \cdot F_1 + b_2 \cdot F_2 + \dots + b_n \cdot F_n$$

με b να είναι οι εκτιμήσεις των πραγματικών τιμών των παραμέτρων β .

Παρόλα αυτά, όπως και στην περίπτωση των παραγοντικών πειραμάτων, η παραπάνω εξίσωση μπορεί να αναπτυχθεί έτσι ώστε να ποσοτικοποιηθούν και οι πιθανές αλληλοεπιδράσεις. Οι αλληλοεπιδράσεις, στην περίπτωση των πειραμάτων με μίγματα, ονομάζονται συνέργιες ή ανταγωνισμοί, αντίστοιχα, αν η αλληλοεπίδραση είναι θετική (δηλ. αυξάνει την απόκριση) ή αρνητική (μειώνει την απόκριση). Η πλήρης γραμμική πολυωνυμική εξίσωση (δευτέρου βαθμού) μίγματος με 3 συστατικά που περιλαμβάνει και τις διπλές αλληλοεπιδράσεις είναι:

$$Y = \beta_0 + \beta_1 \cdot F_1 + \beta_2 \cdot F_2 + \beta_3 \cdot F_3 + \beta_{12} \cdot F_1 \cdot F_2 + \beta_{13} \cdot F_1 \cdot F_3 + \beta_{23} \cdot F_2 \cdot F_3 + \varepsilon \quad (2)$$

ενώ η γραμμική πλήρης κυβική (τρίτου βαθμού) εξίσωση μίγματος με 3 συστατικά που περιλαμβάνει και τη μοναδική τριπλή αλληλοεπίδραση είναι:

$$Y = \beta_0 + \beta_1 \cdot F_1 + \beta_2 \cdot F_2 + \beta_3 \cdot F_3 + \beta_{12} \cdot F_1 \cdot F_2 + \beta_{13} \cdot F_1 \cdot F_3 + \beta_{23} \cdot F_2 \cdot F_3 + \beta_{123} \cdot F_1 \cdot F_2 \cdot F_3 + \varepsilon \quad (3)$$

Εφόσον υπάρχουν πραγματικά δεδομένα, η παραπάνω εξίσωση χρησιμοποιείται για να περιγράψει τα δεδομένα αυτά και συνεπώς να εξαχθούν οι συντελεστές $\beta_0, \beta_1, \beta_2, \beta_3, \beta_{12}, \beta_{13}, \beta_{23}, \beta_{123}$. Οι συντελεστές $\beta_{12}, \beta_{23}, \beta_{13}$ εκφράζουν το μέγεθος των διπλών συνεργιών ή ανταγωνισμών, ενώ ο συντελεστής β_{123} εκφράζει το μέγεθος της τριπλής συνέργιας ή ανταγωνισμού. Συγκεκριμένα μιλάμε για:

- ο Συνέργια, εφόσον οι συντελεστές β είναι θετικοί αριθμοί
- ο Ανταγωνισμό, εφόσον οι συντελεστές β είναι αρνητικοί αριθμοί

Οι δε τιμές των σφαλμάτων ε στις εξισώσεις (2) και (3) είναι τα πειραματικά σφάλματα.

Προφανώς, με τη χρήση του μοντέλου θα εκτιμήσουμε τις παραμέτρους β μέσω των συντελεστών b . Δηλαδή, αυτό που τελικώς υπολογίζουμε είναι οι τιμές: $b_1, b_2, b_3, b_{12}, \dots$ κ.λ.π., που αποτελούν τις εκτιμήσεις των πραγματικών τιμών των παραμέτρων β . Συνεπώς, οι τιμές b δεν θα συνοδεύονται από μία τιμή μόνο αλλά και από ένα τυπικό σφάλμα ή από το διάστημα εμπιστοσύνης, το οποίο πρακτικώς θα μας λέει αν η πραγματική τιμή της παραμέτρου β πιθανά να περιέχει το 0 ή όχι. Στην πρώτη περίπτωση, δεν θα εισαχθεί προφανώς η αντίστοιχη μεταβλητή στο τελικό μας μοντέλο. Βάσει των αρχών των εξισώσεων των μιγμάτων, $F_1 + F_2 + F_3 = 1$ και $0 \leq F_1, F_2, F_3 \leq 1$. Το παράδειγμα στη συνέχεια θα βοηθήσει στην κατανόηση των παραπάνω.

12.4.1 Παράδειγμα πειράματος με μίγμα αστικών στερεών αποβλήτων με 3 συστατικά

Τα πειράματα είχαν σαν στόχο την μέτρηση της παραγωγής του CO₂ από διάφορα μίγματα αστικών στερεών αποβλήτων (ΑΣΑ) από 3 συστατικά. Τα 3 συστατικά ήταν τα:

- ο υπολείμματα χαρτιού (YX)
- ο υπολείμματα κήπου (YK)
- ο υπολείμματα φαγητού (YΦ).

Τα μίγματα που έγιναν ήταν 10 (συμπεριλαμβανομένων 2 επαναλήψεων τα μεμονωμένων συστατικά YK και των YΦ) και καταγράφηκαν στο πείραμα οι παραγωγές CO₂ (σε gr C-CO₂/dry kg μίγματος) και οι παραγωγές N-NH₃, που φαίνονται στον πίνακα 12.1. Οι παραγωγές αυτές θα συμβολίζονται αντίστοιχα Y_C & Y_N. Οι τιμές στις στήλες 3 έως 5 είναι οι τιμές των F (που είναι τα ποσοστά του κάθε συστατικού στο μίγμα). Παραδειγματικά, στο τελευταίο μίγμα, το χαρτί (YX) αποτελούσε το 34% του ξηρού βάρους του μίγματος, τα YK το 37% και τα YΦ το 30%. Οι τιμές της παραγωγής είναι σε g αερίου / ξηρό kg του συνολικού μίγματος.

Πίνακας 12.1. Ποσοστά 3 συστατικών σε πειράματα με μίγματα ΑΣΑ

A/A	Συστατικά μίγματος	YX (F ₁)	YK (F ₂)	YΦ (F ₃)	CO ₂ (g C / dry kg)	NH ₃ (g N / dry kg)
1	Υπολ. χαρτιού (YX)	1	0	0	150	2.0
2	Υπολ. κήπ. (YK)	0	1	0	217	4.4
3	Υπολ. κήπ. (YK)	0	1	0	220	4.6
4	Υπολ. φαγ. (YΦ)	0	0	1	360	34
5	Υπολ. φαγ. (YΦ)	0	0	1	369	41
6	YX / YK	0.83	0.17	0	246	0.6
7	YX / YΦ	0.96	0	0.04	237	1.1
8	YΦ/YK	0	0.78	0.22	302	14.5
9	YΦ/YX/YK	0.80	0.155	0.045	265	0.5
10	YΦ/YX/YK	0.34	0.37	0.30	266	6.5

*: Τα πειράματα 2 και 3 είναι επαναλήψεις των YK, ενώ τα πειράματα 4 & 5 είναι επαναλήψεις των YΦ.

Με χρήση της μεθόδου των ελαχίστων τετραγώνων, και μέσω στατιστικών προγραμμάτων όπως το MINITAB[®] 14, γίνεται υπολογισμός των συντελεστών της εξίσωσης (2). Μία προτεινόμενη τεχνική για τον εντοπισμό των συντελεστών της εξίσωσης είναι με χρήση του Solver από το Excel του Microsoft Office. Στην περίπτωση αυτή, υπολογίζεται η τιμή του μοντέλου για κάθε μίγμα – όπως έγινε στα πρωτογενή πειράματα – και υπολογίζεται το τετράγωνο των σφαλμάτων, δηλ. της διαφοράς της τιμής του μοντέλου από την πραγματική τιμή, για το κάθε πείραμα. Με τη χρήση του Solver μπορεί να ελαχιστοποιηθεί το **άθροισμα των τετραγώνων των σφαλμάτων**, μεταβάλλοντας τις τιμές των συντελεστών της εξίσωσης (2). Με αυτόν τον τρόπο, υπολογίζεται ο βέλτιστος «συνδυασμός» των συντελεστών της πολυωνυμικής εξίσωσης.

Ο διαμορφωμένος πίνακας που χρησιμοποιείται για τη βελτιστοποίηση είναι ο πίνακας 12.1. Για κάθε ένα από τα 10 πειράματα υπολογίζεται το τετράγωνο των σφαλμάτων, δηλ. το $(CO_{2real} - CO_{2model})^2$. Τα σφάλματα αυτά αθροίζονται $[\Sigma(CO_{2real} - CO_{2model})^2]$ και το άθροισμα αυτό συνήθως ορίζεται ως Residual Sum of Squares (RSS). Το μοντέλο με το οποίο θέλουμε να περιγράψουμε τα δεδομένα δεν περιλαμβάνει το συντελεστή b₀.

Θεωρείται δηλαδή ότι αν δεν υπάρχουν συστατικά στο μίγμα, η παραγωγή διοξειδίου του άνθρακα θα είναι ίση με 0. Το μοντέλο έχει τελικά τη μορφή:

$$\widehat{Y}_c = b_1 \cdot F_1 + b_2 \cdot F_2 + b_3 \cdot F_3 + b_{12} \cdot F_1 \cdot F_2 + b_{13} \cdot F_2 \cdot F_3 + b_{23} \cdot F_2 \cdot F_3 + b_{123} \cdot F_1 \cdot F_2 \cdot F_3 \quad (3)$$

όπου \widehat{Y}_c είναι η πρόβλεψη της εκπομπής του CO₂ από το συγκεκριμένο μοντέλο.

Στη συνέχεια γίνεται ελαχιστοποίηση του αθροίσματος αυτού μεταβάλλοντας τις τιμές των συντελεστών b_1, b_2, \dots, b_{123} . Συνεπώς, ζητούνται 7 συντελεστές ($b_1, b_2, b_3, \dots, b_{123}$), ενώ υπάρχουν 10 εξισώσεις. Συνεπώς, είναι δυνατός ο υπολογισμός των τιμών αυτών, αφού έχουμε ένα αλγεβρικό σύστημα 7 αγνώστων με 10 εξισώσεις. Οι εξισώσεις που χρησιμοποιούνται είναι οι εξής, ακολουθώντας τη σειρά των πειραμάτων από το 1^ο στο 10^ο, όπως αυτή αναγράφεται στον πίνακα 12.1.

Πείραμα 1

$$b_1 \cdot 1 + b_2 \cdot 0 + b_3 \cdot 0 + b_{12} \cdot 0 + b_{13} \cdot 0 + b_{23} \cdot 0 + b_{123} \cdot 0 = 150$$

Πείραμα 2

$$b_1 \cdot 0 + b_2 \cdot 1 + b_3 \cdot 0 + b_{12} \cdot 0 + b_{13} \cdot 0 + b_{23} \cdot 0 + b_{123} \cdot 0 = 217$$

Πείραμα 3

$$b_1 \cdot 0 + b_2 \cdot 1 + b_3 \cdot 0 + b_{12} \cdot 0 + b_{13} \cdot 0 + b_{23} \cdot 0 + b_{123} \cdot 0 = 220$$

Πείραμα 4

$$b_1 \cdot 0 + b_2 \cdot 0 + b_3 \cdot 1 + b_{12} \cdot 0 + b_{13} \cdot 0 + b_{23} \cdot 0 + b_{123} \cdot 0 = 360$$

Πείραμα 5

$$b_1 \cdot 0 + b_2 \cdot 0 + b_3 \cdot 1 + b_{12} \cdot 0 + b_{13} \cdot 0 + b_{23} \cdot 0 + b_{123} \cdot 0 = 369$$

Πείραμα 6

$$b_1 \cdot 0,83 + b_2 \cdot 0,17 + b_3 \cdot 0 + b_{12} \cdot 0,141 + b_{13} \cdot 0 + b_{23} \cdot 0 + b_{123} \cdot 0 = 246$$

Πείραμα 7

$$b_1 \cdot 0,96 + b_2 \cdot 0 + b_3 \cdot 0,04 + b_{12} \cdot 0 + b_{13} \cdot 0,0384 + b_{23} \cdot 0 + b_{123} \cdot 0 = 237$$

Πείραμα 8

$$b_1 \cdot 0 + b_2 \cdot 0,78 + b_3 \cdot 0,22 + b_{12} \cdot 0 + b_{13} \cdot 0 + b_{23} \cdot 0,172 + b_{123} \cdot 0 = 302$$

Πείραμα 9

$$b_1 \cdot 0,80 + b_2 \cdot 0,16 + b_3 \cdot 0,045 + b_{12} \cdot 0,124 + b_{13} \cdot 0,036 + b_{23} \cdot 0,007 + b_{123} \cdot 0,00558 = 265$$

Πείραμα 10

$$b_1 \cdot 0,34 + b_2 \cdot 0,36 + b_3 \cdot 0,30 + b_{12} \cdot 0,122 + b_{13} \cdot 0,102 + b_{23} \cdot 0,108 + b_{123} \cdot 0,0367 = 266$$

Η μέθοδος των ελαχίστων τετραγώνων χρησιμοποιείται για τον υπολογισμό των 7 συντελεστών b_1, \dots, b_{123} . Σημειώνεται εξάλλου ότι είναι δυνατός ο υπολογισμός των 7 αυτών παραμέτρων, αφού τα πειράματά μας είναι $10 > 7$. Υπάρχουν δηλαδή 3 βαθμοί ελευθερίας σε σχέση με την διαδικασία της παλινδρόμησης. Αν παραδειγματικά είχαμε μόνο 5 πειράματα, δεν θα μπορούσαμε να υπολογίσουμε και τις 7 παραμέτρους του πλήρους μοντέλου. Θα μπορούσαμε να υπολογίσουμε έως και 4 παραμέτρους σε ένα παρόμοιο πολυωνυμικό μοντέλο.

Τα αποτελέσματα της μεθόδου ελαχίστων τετραγώνων φαίνονται συνολικά στον πίνακα 12.1. Σύμφωνα με τον πίνακα αυτόν, το πλήρες μοντέλο που θα περιέγραφε τα αποτελέσματα των πειραμάτων για ένα μίγμα 3 συστατικών, θα ήταν το:

$$\begin{aligned} \hat{Y}_c = & 152 \cdot F_1 + 219 \cdot F_2 + 365 \cdot F_3 + 555 \cdot F_1 \cdot F_2 + 1909 \cdot F_2 \cdot F_3 + 301 \cdot F_2 \cdot F_3 \\ & - 7348 \cdot F_1 \cdot F_2 \cdot F_3 \end{aligned} \quad (4)$$

Πίνακας 12.1. Χρήση λογιστικού φύλλου για υπολογισμό συντελεστών στο μοντέλο μιγμάτων με 3 συστατικά (χρήση 7 παραμέτρων στο μοντέλο) με απόκριση το CO₂

YX (F ₁)	YK (F ₂)	YΦ (F ₃)	YX / YK (F ₁ , F ₂)	YX / YΦ (F ₁ , F ₃)	YK / YΦ (F ₂ , F ₃)	YX / YK / YΦ (F ₁ , F ₂ , F ₃)	CO ₂ μετρ.	CO ₂ μοντ.	b ₁	b ₂	b ₃	b ₁₂	b ₁₃	b ₂₃	b ₁₂₃	Res ²
1	0	0	0	0	0	0	150	152,2	152	219	365	555	1909	301	-7348	
0	1*	0	0	0	0	0	217	218,5	152,2	0,0	0,0	0,0	0,0	0,0	0,0	4,7
0	1*	0	0	0	0	0	220	218,5	0,0	218,5	0,0	0,0	0,0	0,0	0,0	2,1
0	0	1 ⁺	0	0	0	0	360	364,5	0,0	218,5	0,0	0,0	0,0	0,0	0,0	2,4
0	0	1 ⁺	0	0	0	0	369	364,5	0,0	0,0	364,5	0,0	0,0	0,0	0,0	20,5
0,83	0,17	0	0,14	0	0	0	246	241,8	0,0	0,0	364,5	0,0	0,0	0,0	0,0	20,0
0,96	0	0,04	0	0,038	0	0	237	234,0	126,3	37,1	0,0	78,3	0,0	0,0	0,0	17,9
0	0,78	0,22	0	0	0,17	0	302	302,3	146,1	0,0	14,6	0,0	73,3	0,0	0,0	9,1
0,8	0,16	0,045	0,12	0,036	0,007	0,0056	265	270,7	0,0	170,4	80,2	0,0	0,0	51,7	0,0	0,1
0,34	0,36	0,3	0,12	0,10	0,11	0,037	266	265,1	121,7	33,9	16,4	68,8	68,7	2,1	-41,0	32,1
									51,7	78,6	109,4	67,9	194,7	32,6	-269,8	0,7
															SS _R =>	109,7

* : επαναλήψεις του πειράματος με YK μόνο

+ : επαναλήψεις του πειράματος με YΦ μόνο

SS_R: Sum of Squared Residuals (Άθροισμα των τετραγώνων των σφαλμάτων)

Πόσο σωστό όμως είναι αυτό το μοντέλο? Δεν μπορούμε να συμπεράνουμε τίποτε ακόμα αν δεν υπολογίσουμε πρώτα την ακρίβεια των συντελεστών b, όπως αυτή καθορίζεται από το μέγεθος του τυπικού σφάλματος ή του διαστήματος εμπιστοσύνης για τον κάθε συντελεστή στο μοντέλο ξεχωριστά. Αυτά θα αναλυθούν εκτενέστερα στο κεφάλαιο 13.

Ας αναλύσουμε λίγο τον πίνακα διασποράς του παραπάνω μοντέλου. Οι βασικές παράμετροι που υπολογίζονται σε ένα πίνακα διασποράς (ANOVA table) είναι οι:

$$SS_{\text{regr}} = \sum (\hat{y}_i - \bar{\hat{y}})^2$$

$$SS_{\text{res}} = \sum (\hat{y}_i - y_i)^2 \text{ (ή RSS)}$$

$$SS_{\text{tot}} = SS_{\text{regr}} + SS_{\text{res}} = \sum (y_i - \bar{y})^2$$

Με:

$\bar{\hat{y}}$ την μέση τιμή των υπολογισμένων από το μοντέλο τιμών \hat{y}_i (απόκρισης)

y_i τις πραγματικές τιμές που μετρώνται κατά το πείραμα και,

\hat{y}_i τις υπολογισμένες από το μοντέλο τιμές της απόκρισης για το κάθε πείραμα i .

Σε ένα πίνακα ANOVA λοιπόν, θα έχουμε την παρακάτω εμφάνιση τιμών:

Πίνακας 12.2. Πίνακας ANOVA εμπειρικού μοντέλου μιγμάτων του πίνακα 12.1

Διασπορά λόγω	B.E.	SS	MS	F
Παλινδρόμησης (Regression)	6 (p-1)	39767,9	6628	181,3
Σφαλμάτων (Residuals)	3 (n-p)	109,7	36,6	
Συνόλου (Total)	9 (n-1)	39877,6		

B.E.: βαθμοί ελευθερίας; n: συνολικός αριθμός πειραμάτων, που εδώ είναι 10; p: ο αριθμός των παραμέτρων στο μοντέλο που θέλουμε να υπολογιστούν, που εδώ είναι 7; SS: Sum of Squares; MS: Mean Sum of Squares, που είναι το πηλίκο του SS με τους αντίστοιχους B.E.

Οι βαθμοί ελευθερίας για τα **σφάλματα** είναι n-p, όπου n ο αριθμός των πειραμάτων (ή δειγμάτων) και p ο αριθμός των παραμέτρων στο μοντέλο που επεξεργαζόμαστε. Για τα σφάλματα λοιπόν, B.E. = 10-7=3. Πως υπολογίζονται οι βαθμοί ελευθερίας όμως για τα υπόλοιπα? Οι **συνολικοί** βαθμοί ελευθερίας εξαρτώνται από το συνολικό αριθμό των δειγμάτων ή πειραμάτων και είναι απλά n - 1, ήτοι 10-1 = 9. Οι βαθμοί ελευθερίας για την παλινδρόμηση είναι p-1, ήτοι 7-1=6, εφόσον 7 είναι ο αριθμός των παραμέτρων στο συγκεκριμένο μοντέλο που θέλουμε να αναπτύξουμε. Το p μεταβάλλεται ανάλογα με το μοντέλο, ενώ το n παραμένει σταθερό για ένα συγκεκριμένο πείραμα. Οι παραπάνω πράξεις μπορούν σχετικά εύκολα να εισαχθούν σε ένα λογιστικό φύλλο (Excel) και να υπολογιστούν οι τιμές. Αν και υπάρχουν πολλά στατιστικά πακέτα που άμεσα εξάγουν τις τιμές αυτές, συνίσταται πάντα η κατανόηση των παραπάνω παραμέτρων.

Για την περίπτωση των MS της παλινδρόμησης, έχω:

$$MS_{\text{Regr}} = \frac{SS_{\text{Regr}}}{p-1} = \frac{39767,9}{6} = 6628$$

Για την περίπτωση των MS των σφαλμάτων, έχω:

$$MS_{Res} = s^2 = \frac{SS_{Res}}{n - p} = \frac{109,7}{3} = 36,6 \quad 29$$

Η παραπάνω σχέση μας λέει ότι όσο μεγαλύτερος ο αριθμός των πειραμάτων n, τόσο μικρότερη η διασπορά. Αντίθετα, όσο μεγαλώνει ο αριθμός των εκτιμώμενων παραμέτρων p τόσο ελαττώνεται ο παρονομαστής και συνεπώς τόσο αυξάνει η συνολική διασπορά s^2 . Κατά συνέπεια τόσο θα αυξάνει και η διασπορά του κάθε συντελεστή b στο πολυωνυμικό μας μοντέλο.

Σε αυτό το σημείο αξίζει να ορίσουμε και τον περίφημο συντελεστή παλινδρόμησης (ή καλύτερα συντελεστής απόφασης) R^2 . Ο συντελεστής αυτός ορίζεται ως εξής:

$$R^2 = \frac{SS_{Regr}}{SS_{tot}} = \frac{39767,9}{39877,6} = 99.72\%$$

Πρακτικά ο συντελεστής αυτός είναι το ποσοστό της συνολικής διασποράς που οφείλεται στην (ή που εξηγείται από την) **παλινδρόμηση**. Αν και συνήθως επιθυμείται αυτή η τιμή να βρίσκεται όσο γίνεται κοντύτερα στο 1 (100%), δεν σημαίνει ότι ένα καλό και επαρκές μοντέλο κρίνεται μόνο από την υψηλή τιμή R^2 . Αντίθετα, υπάρχουν πολλά παραδείγματα μη επαρκών μοντέλων με υψηλό R^2 . Δεν είναι σωστό δηλαδή να φορτώνουμε το μοντέλο μας με όσο περισσότερες παραμέτρους μπορούμε ώστε να αυξήσουμε το R^2 . Ο στόχος μας είναι να βρούμε το απλούστερο μοντέλο που βέλτιστα χαρακτηρίζει και περιγράφει επαρκώς τα δεδομένα μας (best reduced model).

Σε άλλο κεφάλαιο θα δούμε την καταλληλότητα του R^2 ως τιμή για να κρίνουμε την επάρκεια ενός μοντέλου. Σε κάθε περίπτωση πάντως, είναι ασφαλές να πούμε ότι γενικά επιθυμούμε σχετικά υψηλές τιμές του SS_{Regr} για να έχουμε ένα καλό μοντέλο.

Η τιμή F ορίζεται προφανώς ως MS_{Regr} / MS_R , ήτοι $6628 / 36,6 = 181,3$. Τι σημαίνει η τιμή αυτή? Οτι η διασπορά λόγω της παλινδρόμησης είναι αρκετά μεγαλύτερη από τη διασπορά λόγω του τυχαίου πειραματικού σφάλματος. Η τιμή πρέπει να συγκριθεί με την τιμή F για το άνω 5% (σύνηθες επίπεδο εμπιστοσύνης) που αντιστοιχεί σε βαθμούς ελευθερίας v_1, v_2 , που αντίστοιχα είναι οι βαθμοί ελευθερίας του αριθμητή και του παρονομαστή του λόγου F. Στην περίπτωσή μας $v_1=6$ και $v_2=3$, άρα $F_{6,3,0.05} = 8.94$. Επειδή, $181.3 \gg 8.94$, μπορούμε να συμπεράνουμε ότι η διασπορά λόγω της παλινδρόμησης είναι όντως μεγαλύτερη από τη διασπορά λόγω του πειραματικού σφάλματος. Συνεπώς, η παλινδρόμησης μας έχει κάποια στατιστικά σημαντική βάση και το μοντέλο μας έχει δυνατότητες πρόβλεψης. Ο πρακτικός κανόνας λέει ότι ο λόγος F πρέπει να είναι 4 φορές μεγαλύτερος του $F_{v_1, v_2, 0.05}$, ώστε το εμπειρικό μας μοντέλο να έχει δυνατότητες πρόβλεψης. Στην παραπάνω περίπτωση, $181.3 \gg 4 \cdot 8.94$ και συνεπώς το εμπειρικό μοντέλο που έχουμε αναπτύξει έχει σαφώς δυνατότητες πρόβλεψης. Δεν αποτελεί δηλαδή μόνο μοντέλο «περιγραφής» των δεδομένων.

²⁹ Υπάρχουν και εκφράσεις του MS_R όπου ο παρονομαστής είναι $n-p-1$ αντί για $n-p$.

Που καταλήγουμε όμως? Έχει άραγε νόημα να διατηρήσουμε όλους τους συντελεστές $b_1 \dots b_{123}$ στο τελικό μοντέλο. Ποια είναι ακρίβεια των συντελεστών αυτών? Με μια πρώτη ματιά στον πίνακα βλέπουμε ότι ο συντελεστής του συστατικού 1 είναι μικρότερος από το συντελεστή του συστατικού. Άρα το φαγητό συνεισφέρει περισσότερο στην παραγωγή CO₂ σε ένα μίγμα σε σχέση με το χαρτί στο μίγμα. Επίσης αυτό που μας λέει ο πίνακας 12.1 είναι ότι υπάρχουν συνέργιες (θετικές αλληλεπιδράσεις) 2^{ου} βαθμού. Δηλαδή, η μίξη χαρτιού και φαγητού παράγει περισσότερο CO₂ από ότι το άθροισμα των επιμέρους παραγωγών CO₂ των μεμονωμένων συστατικών, χαρτιού και Υ.Κ. Προτού όμως βγάλουμε αυτό το συμπέρασμα θα πρέπει να κοιτάξουμε την ακρίβεια των παραμέτρων b_i , δηλαδή τα τυπικά σφάλματα των εν λόγω παραμέτρων ή τα διαστήματα εμπιστοσύνης τους. Αυτό θα το δούμε σε επόμενο κεφάλαιο, αφού αφορά συνολικά στην ακρίβεια παραμέτρων εμπειρικών μοντέλων και συνεπώς εμπίπτει στην έννοια της ανάπτυξης εμπειρικών μοντέλων από πειραματικά δεδομένα μέσω γραμμικής παλινδρόμησης.

Αυτό που μας λένε οι συντελεστές του μοντέλου μας στον πίνακα 12.1, είναι ότι αν έχουμε ένα μίγμα με 30% χαρτί και 70% υπολείμματα κήπου (δηλ. με 0% στο 3^ο συστατικό), τότε η παραγωγή διοξειδίου του άνθρακα θα εκτιμάται βάσει του (πλήρους) μοντέλου από τη σχέση:

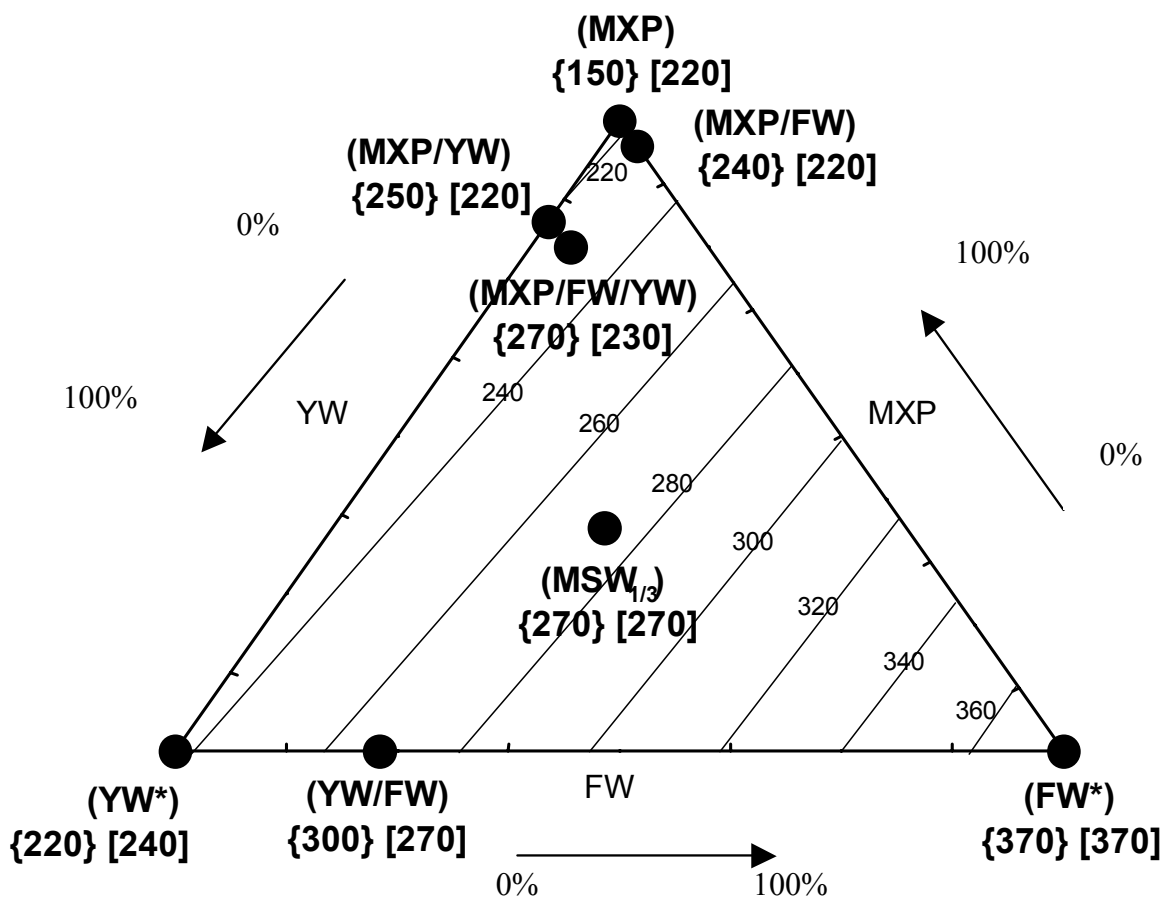
$$\begin{aligned} \hat{Y}_c &= 152 * F_1 + 219 * F_2 + 555 * F_1 * F_2 = \\ &= 152 * 0,3 + 219 * 0,7 + 555 * 0,3 * 0,7 = 315 \text{ g αερίου / kg μίγματος αποβλήτου} \end{aligned}$$

Το F₃ δεν εισάγεται καθόλου στη σχέση, αφού ισούται με 0. Ομοίως και τα γινόμενα F₁*F₃, F₂*F₃, F₁*F₂*F₃ ισούνται με το 0, αρά δεν εμφανίζονται στο τελικό μοντέλο.

Αν υποθέσουμε λοιπόν ότι το βέλτιστο μοντέλο μας τελικά δεν περιλαμβάνει συνέργιες, ούτε δευτέρου και ούτε τρίτου βαθμού. Αυτό θα εξαρτηθεί από τον υπολογισμό των διαστημάτων εμπιστοσύνης του κάθε συντελεστή b_i . Πολύ απλά σημειώνεται ότι αν το διάστημα εμπιστοσύνης ενός συντελεστή b_i περιλαμβάνει το 0, τότε ο συντελεστής αυτός δεν κρίνεται σημαντικός και το μοντέλο επαναδοκιμάζεται χωρίς το συγκεκριμένο όρο του μοντέλου. Μετά από αυτή τη διαδικασία, η οποία θα περιγραφεί στο επόμενο κεφάλαιο, εξάγουμε για το μοντέλο της εξίσωσης (3), βάσει των πρωτογενών δεδομένων που εμφανίζονται στον πίνακα 12.1 το παρακάτω βέλτιστο μειωμένο μοντέλο (best reduced model):

$$\hat{Y}_c = 220 F_p + 240 F_Y + 370 F_F \quad (4)$$

Η γραφική απεικόνιση (επιφάνεια απόκρισης) της παραπάνω εξίσωσης που περιγράφει ένα μίγμα με 3 συστατικά φαίνεται στο σχήμα 12.4.



Σχήμα 12.4 Επιφάνεια απόκρισης για το CO₂ βάσει της εξίσωσης (4), σημείωση: MXP=YX, FW=YΦ, YW=YK ((Komilis & Ham, 2006).

[Οι μαύρες τελείες δείχνουν τις περιοχές που πραγματοποιήθηκαν τα 10 πειράματα. Τα επαναληπτικά πειράματα για τα YK και τα YΦ πέφτουν στο ίδιο σημείο και συνεπώς αντιστοιχεί στο κάθε ζεύγος ένα σημείο. Οι ισοπληθείς καμπύλες, που εδώ είναι ευθείες γραμμές, αφού δεν υπάρχουν συνέργειες ή ανταγωνισμοί, αποτελούνται από σημεία ίδιας συνολικής εκπομπής CO₂ σε g C-CO₂ / ξηρό kg μίγματος].
{: πειραματικά δεδομένα, [: αποτελέσματα εξίσωσης (4)

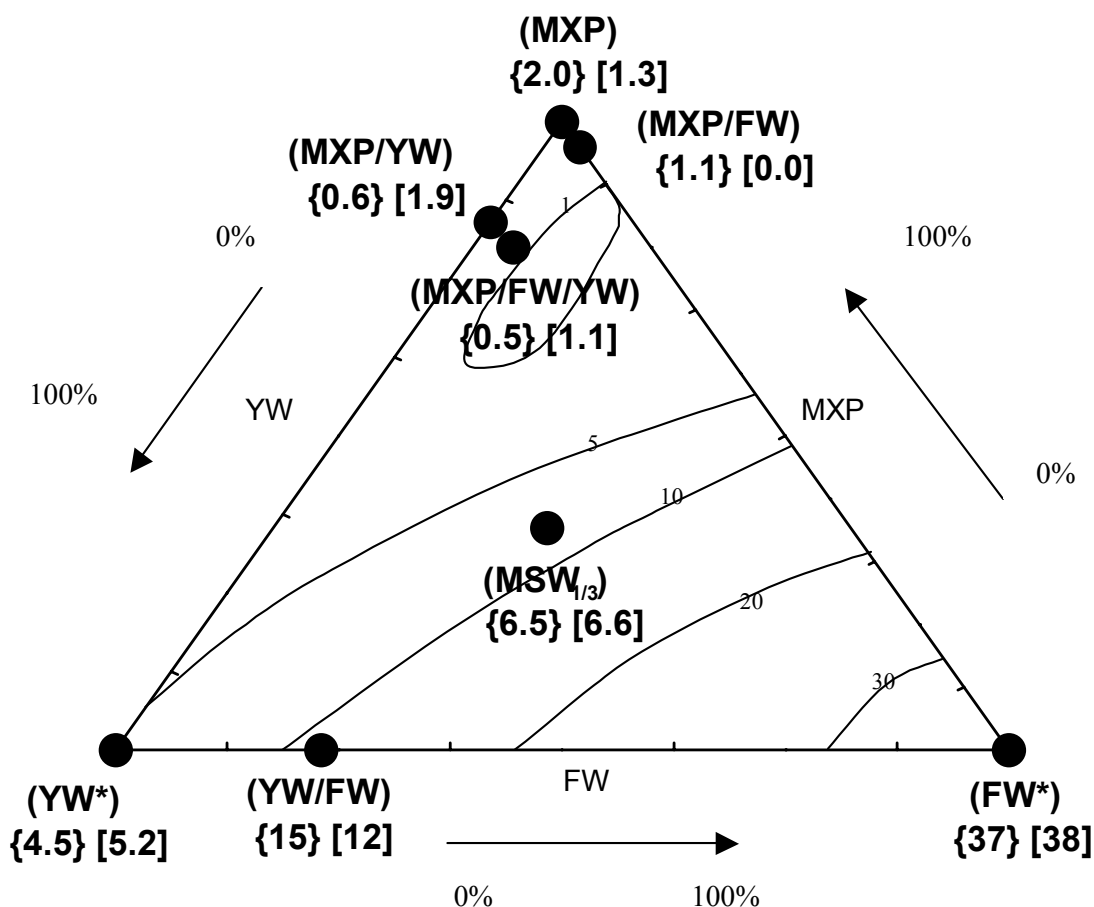
Το σχήμα 12.4 εξηγείται ως εξής. Εντός του τριγώνου, που περικλείεται από τα 3 ακραία σημεία που υποδηλώνουν 3 ξεχωριστά πειράματα με ένα συστατικό ανά μίγμα, η επιφάνεια είναι επίπεδη με κλίση ανοδική προς τα υπολείμματα φαγητού (FW). Συνεπώς, όσο αυξάνεται το ποσοστό φαγητού σε ένα μίγμα στερεών αποβλήτων, αυξάνεται η μοναδιαία παραγωγή διοξειδίου του άνθρακα του μίγματος. Πάνω στις πλευρές του τριγώνου, υπάρχουν μίγματα δύο συστατικών. Παραδειγματικά, στην αριστερή πλευρά του τριγώνου είναι οι «θέσεις» όλων των μιγμάτων με υπολείμματα χαρτιού (YX) και υπολείμματα κήπου (YK). Στην κορυφή του τριγώνου το μίγμα περιλαμβάνει 100% YX και στο κάτω αριστερά σημείο το μίγμα περιλαμβάνει 100% YK.

Η επιφάνεια απόκρισης δεν είναι πάντα επίπεδη. Παραδειγματικά, στην περίπτωση που η απόκριση είναι η εκπομπή αμμωνίας, το μειωμένο βέλτιστο μοντέλο που περιγράφει την εκπομπή της αμμωνίας από το μίγμα των ΑΣΑ βρέθηκε ότι είναι το:

$$\hat{Y}_N = 1.3 (\pm 1.4) \cdot F_1 + 5.2 (\pm 1.4) \cdot F_2 + 38 (\pm 1.6) \cdot F_3 - 69 (\pm 23) \cdot F_1 \cdot F_3 \quad (5)$$

με \hat{Y}_N να είναι η συνολική εκπομπή αμμωνίας (NH₃-N) σε gr NH₃-N / ξηρό kg μίγματος των 3 συστατικών των ΑΣΑ. Οι τιμές εντός των παρενθέσεων είναι τα τυπικά σφάλματα των αντίστοιχων συντελεστών (η παραμέτρων του μοντέλου). Το πώς εξάγονται τα σφάλματα αυτά, θα παρουσιαστεί σε επόμενο κεφάλαιο. Η σημασία πάντως είναι ότι όσο μεγαλύτερο το τυπικό σφάλμα ενός συντελεστή τόσο λιγότερο ακριβής είναι η αντίστοιχη μεταβλητή και δεν αξίζει να τη διατηρούμε στο μοντέλο. Το διάστημα εμπιστοσύνης του κάθε συντελεστή b εκτιμάται (χοντρικά) πολλαπλασιάζοντας με 2.5 το τυπικό σφάλμα. Συνεπώς το διάστημα εμπιστοσύνης για το συντελεστή 5.2 είναι (περίπου) 5.2-2.5*1.4 έως 5.2+2.5*1.4, δηλαδή κυμαίνεται από 1,7 έως 8.7. Επειδή το διάστημα αυτό δεν περιέχει το 0, η αντίστοιχη μεταβλητή του μοντέλου (δηλ. η F2) μπορεί να διατηρηθεί στο μοντέλο, αφού η πραγματική της τιμή είναι κατά πάσα πιθανότητα εντός του παραπάνω εύρους. Αν περιείχε το 0, τότε θα έπρεπε να κάναμε παλινδρόμηση του μοντέλου μέσα από τα δεδομένα μας χωρίς την μεταβλητή F2. Σε αυτό το σημείο αξίζει να τονίσουμε ένα περίεργο χαρακτηριστικό του μοντέλου (5). Βλέπουμε ότι η μεταβλητή F1 συνοδεύεται από ένα συντελεστή (τον 1.3), που σίγουρα το διάστημα εμπιστοσύνης του περιέχει το 0. Άρα γιατί διατηρούμε τη μεταβλητή F1 στο μοντέλο μας? Ο λόγος που διατηρείται η μεταβλητή έχει να κάνει με το ότι η μία διπλή αλληλεπίδραση που περιέχει τη μεταβλητή F1 (δηλ. η F1*F3) είναι στατιστικά σημαντική (αφού το διάστημα 64-2.5*23 έως 64+2.5*23 δεν περιέχει σαφώς το 0). Δεν είναι λοιπόν δυνατό ένα μοντέλο με μία στατιστικά σημαντική αλληλεπίδραση να μην περιλαμβάνει στις παραμέτρους του και τις αντίστοιχες μεμονωμένες μεταβλητές από τις οποίες αποτελείται η αλληλεπίδραση. Για αυτό το λόγο διατηρείται η μη στατιστικά σημαντική μεταβλητή F₁ στην εξίσωση (5).

Επειδή το μοντέλο που εκφράζεται από την εξίσωση (5) περιέχει και τον δεύτερου βαθμού όρο F1*F3, η επιφάνεια απόκρισης δεν μπορεί να είναι επίπεδη. Η επιφάνεια αυτή απεικονίζεται στο σχήμα 12.5.



Σχήμα 12.5. Επιφάνεια απόκρισης για την αμμωνία ($N-NH_3$) βάσει της εξίσωσης (5), σημείωση: $MXP=YX$, $FW=Y\Phi$, $YW=YK$ (Komilis & Ham, 2006).

[Οι μαύρες τελείες δείχνουν τις περιοχές που πραγματοποιήθηκαν τα 10 πειράματα. Τα επαναληπτικά πειράματα για τα YK και τα $Y\Phi$ πέφτουν στο ίδιο σημείο και συνεπώς αντιστοιχεί στο κάθε ζεύγος ένα σημείο. Οι ισοπληθείς καμπύλες, που εδώ δεν είναι ευθείες γραμμές, αφού υπάρχει μία αρνητική αλληλεπίδραση (δηλ. ανταγωνισμός), αποτελούνται από σημεία ίδιας συνολικής εκπομπής NH_3 σε $g N-NH_3 / \xi\eta\rho\acute{o} kg$ μίγματος]. {}: πειραματικά δεδομένα, []: αποτελέσματα εξίσωσης (5)

Οι παραπάνω επιφάνειες απόκρισης (response surfaces), όπως στο σχήμα 12.4 και 12.5, είναι ο κλασσικός τρόπος απεικόνισης απόκρισης από πειράματα μιγμάτων με 3 συστατικά. Υπάρχουν και στατιστικά πακέτα που μπορούν να απεικονίσουν το κλασσικό αυτό τρίγωνο των 3 συστατικών και να σημειώσουν και τις θέσεις των πρωτογενών αποτελεσμάτων των πειραμάτων ακόμα και να εξάγουν τις ισοπληθείς καμπύλες³⁰. Η απεικόνιση της ίδιας της επιφάνειας μπορεί να γίνει και «χειρωνακτικά» υπολογίζοντας τις τιμές εντός του τριγώνου, όπως αυτές υπολογίζονται από το μοντέλο. Φυσικά, υπάρχουν πλέον και προγράμματα που μπορούν να απεικονίσουν και σε ψευδο-3 διαστάσεις τις επιφάνειες απόκρισης (από 3 και 4 συστατικά μιγμάτων). Για μίγματα άνω

³⁰ Είναι σαν την απεικόνιση ισοϋψών σε κάποιο χάρτη.

των 4 συστατικών, μιλάμε για πολυδιάστατους χώρους και συνεπώς δεν είναι δυνατή η γραφική απεικόνιση. Τα μοντέλα παρουσιάζονται απλώς με την μαθηματική τους φόρμα.

12.4.1 Σχεδιασμοί Simplex σε πειράματα με μίγματα

Στο παραπάνω παράδειγμα μίγματος με 3 συστατικά, τα σημεία (πειράματα) δεν είχαν τοποθετηθεί σε ιδανικά σημεία. Παρατηρούμε συγκεκριμένα ότι 4 από τα 10 σημεία έχουν συσσωρευθεί στο απάνω τμήμα του τριγώνου, που υποδηλώνει μίγματα με υψηλό σχετικά ποσοστό χαρτιού. Γενικά στα πειράματα με μίγματα χρησιμοποιείται η αρχή των σχεδιασμών simplex, που βασίζεται στη λίγο πολύ συμμετρική τοποθέτηση των πειραμάτων εντός του τριγώνου (εφόσον μιλάμε για 3 συστατικά) ή εντός του οποιοδήποτε χώρου η διαστάσεων, εφόσον μιλάμε για άνω των 3 συστατικών στο μίγμα. Οι δύο κλασσικές μορφές σχεδιασμού simplex είναι η τύπου lattice και η τύπου centroid.

Σχεδιασμός τύπου simplex lattice (σχεδιασμός πίνακα)

Μία $[p,m]$ σχεδίαση πίνακα simplex lattice για p συστατικά σε ένα μίγμα αποτελείται από θέσεις συντεταγμένων, που επιτρέπουν ανάμειξη μόνο δύο εκ των τριών συστατικών. Οι αναλογίες που χρησιμοποιούνται για κάθε συστατικό λαμβάνουν θέσεις $m+1$, που απέχουν ίσα διαστήματα από το 0 έως το 1. Δηλαδή, αν $m=1$ τότε πρέπει το κάθε συστατικό να συμπεριληφθεί σε $1+1=2$ πειράματα. Οι συντεταγμένες των θέσεων των πειραμάτων είναι:

$$x_i = 0, \frac{1}{m}, \frac{2}{m}, \dots, 1 \quad i = 1, 2, \dots, p$$

Παραδειγματικά, αν $p=3$ (δηλαδή αν έχω 3 συστατικά) και $m=2$, τότε για κάθε συστατικό πρέπει να κάνω πείραμα στις θέσεις:

$$x_i = 0, \frac{1}{2}, \frac{2}{2} \text{ δηλαδή στις 3 θέσεις} \\ x_i = 0, \frac{1}{2}, 1 \text{ με } i=1, 2, 3.$$

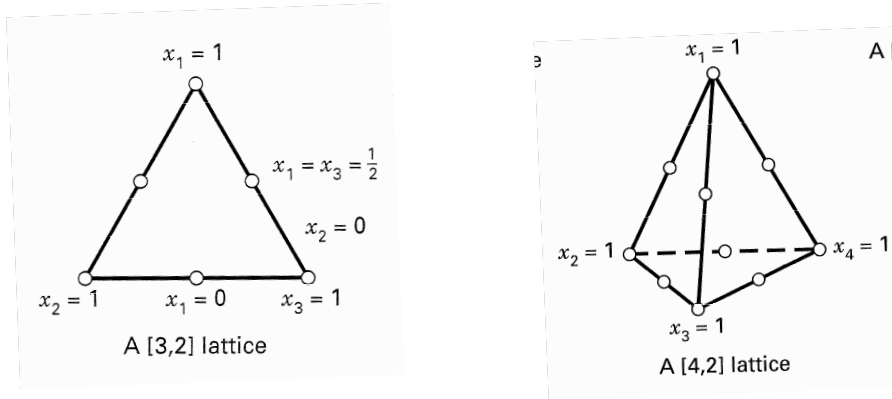
Ο σχεδιασμός simplex τύπου lattice περιλαμβάνει τα παρακάτω 6 πειράματα, που είναι:

$$(x_1, x_2, x_3) = (1,0,0), (0,1,0), (0,0,1), (\frac{1}{2}, \frac{1}{2}, 0), (\frac{1}{2}, 0, \frac{1}{2}), (0, \frac{1}{2}, \frac{1}{2})$$

Οι θέσεις $(1,0,0)$, $(0,1,0)$ & $(0,0,1)$ είναι προφανώς τα καθαρά μίγματα, που αποτελούνται από ένα συστατικό. Δηλαδή, στην πρώτη περίπτωση, το μίγμα περιλαμβάνει 100% το συστατικό x_1 . Τα σημεία $(\frac{1}{2}, \frac{1}{2}, 0)$, $(\frac{1}{2}, 0, \frac{1}{2})$, $(0, \frac{1}{2}, \frac{1}{2})$, είναι τα μίγματα μεταξύ δύο συστατικών. Τα σημεία αυτά βρίσκονται στο ενδιάμεσο σημείο μεταξύ των δύο ακραίων μεμονωμένων συστατικών. Παρατηρούμε, ότι στη συγκεκριμένη περίπτωση σχεδίασης simplex δεν έχουμε μίξη τριών συστατικών. Γενικά, ο συνολικός αριθμός των απαιτούμενων σημείων (ή τελικά πειραμάτων που πρέπει να διεξάγουμε) υπολογίζεται από την εξίσωση:

$$N = \frac{(p+m-1)!}{m!(p-1)!}$$

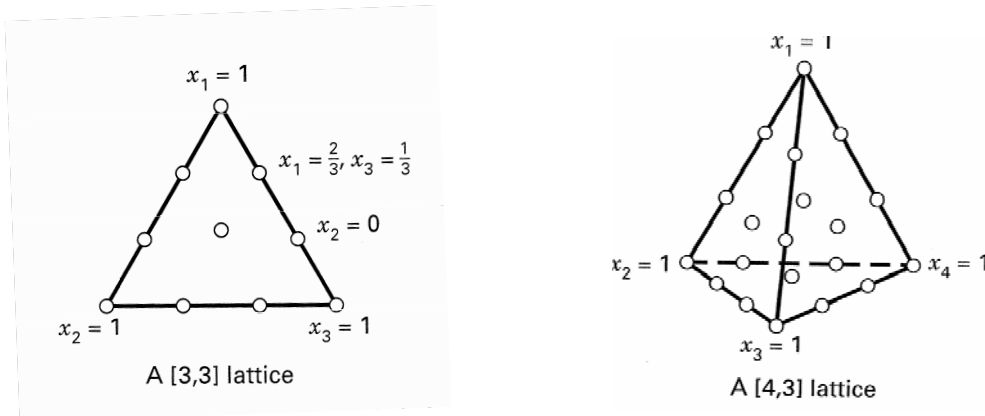
Το σχήμα 12.5 δείχνει το σχεδιασμό τύπου lattice για $p,m=3,2$ & $p,m=4,2$



Σχήμα 12.6. Γραφική απεικόνιση σχεδίασης μίγματος $(p,m)=(3,2)$ και $(p,m)=(4,2)$. Το πρώτο περιλαμβάνει 3 συστατικά και το δεύτερο 4 συστατικά. Αμφότερα βασίζονται στη σχεδίαση τύπου simplex lattice (από Montgomery, 2005)

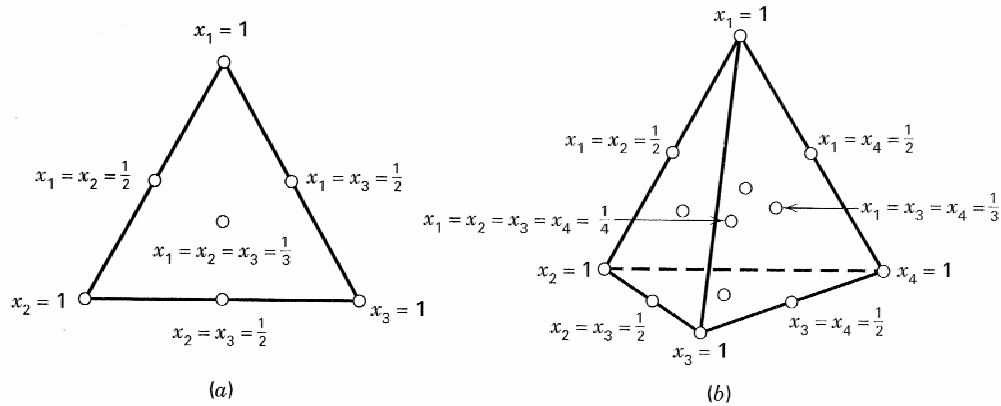
Σχεδιασμός τύπου simplex centroid (κεντρικός σχεδιασμός)

Στην περίπτωση περιλαμβάνεται και κεντρικό μίγμα με όλα τα συστατικά που συμμετέχουν στο μίγμα. Για μία σχεδίαση simplex centroid με p συστατικά απαιτούνται $2^p - 1$ σημεία (θέσεις πειραμάτων). Στα πειράματα αυτά, οι θέσεις των πειραμάτων αντιστοιχούν στους p συνδυασμούς του $(1,0,0,\dots,0)$, στους $p/2$ συνδυασμούς του $(\frac{1}{2}, \frac{1}{2}, 0, \dots, 0)$, στους $p/3$ συνδυασμούς του $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, \dots, 0)$, ..., και στο συνολικό κεντρικό συνδυασμό $(\frac{1}{p}, \frac{1}{p}, \frac{1}{p}, \dots, \frac{1}{p})$. Το σχήμα 12.6 δείχνει τις γραφικές απεικονίσεις centroid σχεδιασμών με μίγματα.



Σχήμα 12.7. Γραφική απεικόνιση σχεδίασης μίγματος $(p,m)=(3,3)$ και $(p,m)=(4,3)$, δηλ. που περιλαμβάνει ένα κεντρικό πείραμα. Το πρώτο περιλαμβάνει 3 συστατικά και το δεύτερο 4 συστατικά. Αμφότερα βασίζονται στη σχεδίαση τύπου simplex centroid (από Montgomery, 2005)

Σχετικά απλούστερες μορφές centroid πειραμάτων φαίνονται στο παρακάτω σχήμα.



Σχήμα 12.8. Σχεδιασμοί simplex centroid με (α) 3 συστατικά, (β) 4 συστατικά ((από Montgomery, 2005)

Γενικά μία κριτική των παραπάνω σχεδιασμών τύπου simplex, είναι ότι τα περισσότερα πειράματα λαμβάνουν χώρα στα όρια της περιοχής (πλευρές των τριγώνων ή ακμές των τετράπλευρων σχημάτων, ανάλογα με το αν έχουμε 3 ή 4 συστατικά αντίστοιχα) και πολύ μικρότερος αριθμός στο κέντρο. Ανεξάρτητα λοιπόν με τους παραπάνω προτεινόμενους σχεδιασμούς, ο πειραματιστής δύναται να τοποθετήσει τα σημεία όπου αυτός/αυτή επιθυμεί – κατά προτίμηση με μία συμμετρική λογική. Ακόμα και αν δεν είναι συμμετρικές οι τοποθετήσεις των σημείων (δηλ. των πειραμάτων), πάλι μπορούν να εξαχθούν αξιόπιστα μοντέλα, όπως εξάλλου παρουσιάστηκε και με τα πρωτογενή δεδομένα του πίνακα 12.1 (Komilis & Ham, 2006), όπου τα σημεία (πειράματα) δεν ήταν σαφώς τοποθετημένα συμμετρικά (βλέπε σχήματα 12.4, 12.5).

Κάποιες τυπικές μορφές των μοντέλων σε πειράματα με μίγματα, που κοινώς χρησιμοποιούνται, είναι οι παρακάτω:

Γραμμικά μοντέλα (δεν περιλαμβάνονται αλληλεπιδράσεις)

$$Y = \sum_{i=1}^p \beta_i \cdot x_i$$

Δυωνυμικά μοντέλα

$$Y = \sum_{i=1}^p \beta_i \cdot x_i + \sum_{i<j} \beta_{ij} x_i x_j$$

Πλήρη κυβικά μοντέλα

$$Y = \sum_{i=1}^p \beta_i \cdot x_i + \sum_{i<j} \beta_{ij} x_i x_j + \sum_{i<j} \delta_{ij} x_i x_j (x_i - x_j) + \sum_{i<j<k} \beta_{ijk} x_i x_j x_k$$

Ειδικά κυβικά μοντέλα

$$Y = \sum_{i=1}^p \beta_i \cdot x_i + \sum_{i<j} \beta_{ij} x_i x_j + \sum_{i<j<k} \beta_{ijk} x_i x_j x_k$$

Το τμήμα $\sum_{i=1}^p \beta_i \cdot x_i$ του μοντέλου ονομάζεται το γραμμικό τμήμα του μίγματος.

Όταν αρχίζει να εμφανίζεται καμπύλωση, τότε έχουμε την εμφάνιση των συντελεστών β_{ij} , που όπως ορίστηκε και παραπάνω εκφράζουν συνεργιστική ή ανταγωνιστική αλληλεπίδραση.

13. Ανάπτυξη εμπειρικών μοντέλων μέσω γραμμικής παλινδρόμησης

Ο στόχος του κεφαλαίου αυτού είναι να παρουσιάσει τη μεθοδολογία ανάπτυξης εμπειρικών μοντέλων, μέσω γραμμικής παλινδρόμησης. Τα μοντέλα προφανώς θα στηρίζονται σε πρωτογενή πειραματικά δεδομένα. Έμφαση δίνεται στον υπολογισμό της ακρίβειας των συντελεστών του μοντέλου αλλά και στην μέθοδο που ακολουθούμε ώστε να αποφασίζουμε τελικά για το ποιές μεταβλητές είναι σημαντικές. Ποιες μεταβλητές λοιπόν αξίζει να παραμείνουν στο μοντέλο μας. Κατ' επέκταση, ποιες μεταβλητές δεν αξίζει να παραμείνουν στο μοντέλο. Έχει νόημα να προσπαθήσουμε να υπολογίσουμε πολύπλοκα μοντέλα με στόχο να αυξήσουμε όσο περισσότερο τον περίφημο συντελεστή R^2 ? Ο υπολογισμός των τιμών των συντελεστών b σε εμπειρικές εξισώσεις, όπως είναι αυτές που εμφανίζονται στο κεφάλαιο 12, είναι λοιπόν μόνο ένα μέρος της δουλειάς που απαιτείται για να γίνει. Πρέπει να υπολογιστεί και η ακρίβεια του κάθε συντελεστή. Η ακρίβεια αυτή θα εκτιμηθεί μέσω του τυπικού σφάλματος και κατά συνέπεια μέσω του διαστήματος εμπιστοσύνης για το κάθε συντελεστή ξεχωριστά, όπως θα δειχθεί στη συνέχεια.

Θα βασιστούμε στο παράδειγμα του μοντέλου μίγμάτων με 3 συστατικά που παρουσιάστηκε στο προηγούμενο κεφάλαιο, με τα πρωτογενή δεδομένα, όπως αυτά εμφανίζονται στον πίνακα 12.1. Θα προσπαθήσουμε να αναπτύξουμε εμπειρικά μοντέλα τόσο για το CO_2 όσο και για την NH_3 .

Ο στόχος μας είναι λοιπόν να περιγράψουμε τα δεδομένα με κάποιο μοντέλο. Αφού μιλάμε για παραδείγματα με μίγματα 3 συστατικών, το απλούστερο (εμπειρικό) μοντέλο είναι το απλό γραμμικό³¹ μοντέλο:

$$Y = \beta_1 \cdot F_1 + \beta_2 \cdot F_2 + \beta_3 \cdot F_3 + \varepsilon \quad (1)$$

Παρατηρούμε ότι εδώ δεν έχουμε εισάγει καθόλου την παράμετρο β_0 (τεταγμένη), αφού έτσι απλοποιείται ακόμα περισσότερο το μοντέλο. Είναι σαν να λέμε ότι όταν οι τιμές των μεταβλητών F είναι όλες ίσες με 0, τότε η απόκριση αναμένεται να είναι ίση με 0.

Το πιο πολύπλοκο γραμμικό μοντέλο θα ήταν το παρακάτω 3^{ου} βαθμού μοντέλο:

$$Y = \beta_1 \cdot F_1 + \beta_2 \cdot F_2 + \beta_3 \cdot F_3 + \beta_{12} \cdot F_1 \cdot F_2 + \beta_{13} \cdot F_1 \cdot F_3 + \beta_{23} \cdot F_2 \cdot F_3 + \beta_{123} \cdot F_1 \cdot F_2 \cdot F_3 + \varepsilon \quad (2)$$

και τελικώς το μοντέλο που θα υπολογίζουμε θα έχει τη μορφή:

³¹ Υπενθυμίζεται ότι ο όρος γραμμικό αφορά στους συντελεστές (ή παραμέτρους) του μοντέλου β . Δηλαδή ένα μοντέλο τύπου $Y = \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2$, είναι γραμμικό ως προς τις παραμέτρους β και 2^{ου} βαθμού ως προς τις ανεξάρτητες μεταβλητές X . Ένα μοντέλο της μορφής $Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{123} X_1 X_2 X_3$ είναι ομοίως γραμμικό ως προς τις παραμέτρους β και 3^{ου} βαθμού ως προς τις ανεξάρτητες μεταβλητές X . Γενικώς, στα εμπειρικά μοντέλα που θίγονται στο κεφάλαιο αυτό, θα εννοείται ότι αυτά είναι γραμμικά.

$$\hat{Y} = b_1 \cdot F_1 + b_2 \cdot F_2 + b_3 \cdot F_3 + b_{12} \cdot F_1 \cdot F_2 + b_{13} \cdot F_1 \cdot F_3 + b_{23} \cdot F_2 \cdot F_3 + b_{123} \cdot F_1 \cdot F_2 \cdot F_3 \quad (3)$$

με παραμέτρους όπως προσδιορίστηκαν στο κεφάλαιο 12.

Στο κεφάλαιο 12 έγινε χρήση της μεθοδολογίας των ελαχίστων τετραγώνων για τον προσδιορισμό των συντελεστών b . Συγκεκριμένα, έγινε χρήση της διαδικασίας ελαχιστοποίησης του SS_R μεταβάλλοντας τις σταθερές b . Αυτό πλέον μπορεί να γίνει εύκολα με τη χρήση προγραμμάτων όπως το Excel και τη ρουτίνα Solver, που επιτρέπει την ελαχιστοποίηση μίας εξαρτημένης μεταβλητής μεταβάλλοντας συγκεκριμένες μεταβλητές και θέτοντας – αν θέλουμε – περιορισμούς (π.χ. οι μεταβλητές που μεταβάλλονται να είναι θετικοί αριθμοί). Με αυτόν τον τρόπο υπολογίστηκαν οι μεταβλητές b στο κεφάλαιο 12. Παρόλα αυτά, για να εκτιμήσουμε την ακρίβεια των συντελεστών πρέπει να έχουμε, όπως είπαμε, και τα αντίστοιχα τυπικά σφάλματα (ή τις διασπορές) των συντελεστών αυτών. Με τη μεθοδολογία της ελαχιστοποίησης μέσω λογιστικού φύλλου, δεν είναι εύκολη η εκτίμηση των σφαλμάτων αυτών. Συνεπώς, στη συνέχεια παρουσιάζεται η μεθοδολογία της εκτίμησης ελαχίστων τετραγώνων με χρήση αλγεβρικών πινάκων, μέσω της οποίας υπολογίζονται με πολλαπλασιασμούς απλών πινάκων τόσο οι συντελεστές b όσο και οι αντίστοιχες διασπορές (ή τα τυπικά σφάλματα) των εν λόγω συντελεστών b του μοντέλου που εμείς επιθυμούμε.

13.1 Μεθοδολογία ελαχίστων τετραγώνων με χρήση αλγεβρικών πινάκων

Το παραπάνω μοντέλο γραμμένο σε μορφή πίνακα είναι:

$$Y = F \beta + \varepsilon$$

Όπου Y είναι το διάνυσμα των παρατηρούμενων τιμών (τιμών που εξάγονται από το πείραμα), F είναι ο πίνακας των ανεξάρτητων μεταβλητών F και β είναι το διάνυσμα των συντελεστών του μοντέλου (τιμές που ψάχνουμε). Βάσει των δεδομένων του πίνακα 12.1, τα αντίστοιχα διανύσματα και πίνακες για το μοντέλο που εκφράζεται από την εξίσωση (2) είναι:

Διάνυσμα παρατηρούμενων τιμών

$$Y = \begin{bmatrix} 150 \\ 217 \\ 220 \\ 360 \\ 369 \\ 246 \\ 237 \\ 302 \\ 265 \\ 266 \end{bmatrix}$$

Πίνακας ανεξάρτητων μεταβλητών (τιμές των μεταβλητών F)

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0,83 & 0,17 & 0 & 0,14 & 0 & 0 & 0 \\ 0,96 & 0 & 0,04 & 0 & 0,038 & 0 & 0 \\ 0 & 0,78 & 0,22 & 0 & 0 & 0,17 & 0 \\ 0,8 & 0,16 & 0,045 & 0,12 & 0,036 & 0,007 & 0,0056 \\ 0,34 & 0,36 & 0,30 & 0,12 & 0,10 & 0,11 & 0,037 \end{bmatrix}$$

Βλέπουμε ότι ο πίνακας F έχει διαστάσεις 10X7, όπου η κάθε στήλη από αριστερά προς τα δεξιά αντιστοιχεί στις τιμές της μεταβλητής F1, F2, F3, F1*F2, F1*F3, F2*F3, F1*F2*F3, αντίστοιχα. Το δε διάνυσμα Y είναι οι 10 παρατηρούμενες πειραματικές τιμές που αντιστοιχούν στα 10 πειράματα, όπως αντίστοιχα έχουν εισαχθεί και στον πίνακα F.

Με βάση τα παραπάνω, οι εκτιμήσεις των συντελεστών b βάσει ελαχίστων τετραγώνων δίνεται από την εξίσωση:

$$b = (F^T \cdot F)^{-1} \cdot F^T \cdot Y$$

Η διασπορά των συντελεστών δίνεται από την εξίσωση:

$$\text{Var}(b) = (F^T \cdot F)^{-1} \cdot \sigma^2$$

Ιδανικά, οι επαναληπτικές μετρήσεις είναι αυτές που χρησιμοποιούνται για τον υπολογισμό του σ για κάθε παράμετρο ξεχωριστά. Επειδή στην περίπτωση μας δεν είχαμε επαναληπτικές μετρήσεις για όλα τα 10 πειράματα, τότε το σ^2 θα υπολογιστεί μέσω της τιμής SS_R διαιρούμενης με τους αντίστοιχους βαθμούς ελευθερίας (όπως θα δειχθεί στη συνέχεια).

Το παραπάνω είναι η μέθοδος των ελαχίστων τετραγώνων βασισμένη σε απλές πράξεις πινάκων, που μπορούν εύκολα να γίνουν από διάφορα υπολογιστικά πακέτα ή αν αυτά δεν είναι διαθέσιμα, βάσει των γνώσεων των πράξεων μεταξύ πινάκων. Το τελευταίο μπορεί κάποιος/α να βρει εύκολα σε οποιοδήποτε εισαγωγικό βιβλίο άλγεβρας και συνεπώς δεν θα αφιερωθεί ιδιαίτερος χρόνος στην εξήγηση των πράξεων μεταξύ πινάκων.

Σημειώνεται ότι αν το μοντέλο περιείχε και την τεταγμένη β_0 , δηλαδή ήταν της μορφής:

$$Y = \beta_0 + \beta_1 \cdot F_1 + \beta_2 \cdot F_2 + \beta_3 \cdot F_3 + \beta_{12} \cdot F_1 \cdot F_2 + \beta_{13} \cdot F_1 \cdot F_3 + \beta_{23} \cdot F_2 \cdot F_3 + \beta_{123} \cdot F_1 \cdot F_2 \cdot F_3 + \varepsilon \quad (4)$$

τότε ο πίνακας F θα είχε την εξής διαμόρφωση.

$$F = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0,83 & 0,17 & 0 & 0,14 & 0 & 0 & 0 \\ 1 & 0,96 & 0 & 0,04 & 0 & 0,038 & 0 & 0 \\ 1 & 0 & 0,78 & 0,22 & 0 & 0 & 0,17 & 0 \\ 1 & 0,8 & 0,16 & 0,045 & 0,12 & 0,036 & 0,007 & 0,0056 \\ 1 & 0,34 & 0,36 & 0,30 & 0,12 & 0,10 & 0,11 & 0,037 \end{bmatrix}$$

Δηλαδή, θα είχαμε έναν πίνακα 10X8, αφού πλέον 8 είναι οι παράμετροι που πρέπει να υπολογιστούν, και η πρώτη στήλη, που αντιστοιχεί στην παράμετρο β_0 , θα έχει τιμές ίσες με 1.

Ο πίνακας F^T είναι ο ανάστροφος του F, δηλαδή έχει ως στήλες τις γραμμές του F και ως γραμμές τις στήλες του F. Συνεπώς, ο ανάστροφος του F είναι:

$$F^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0,83 & 0,96 & 0 & 0,8 & 0,34 \\ 0 & 1 & 1 & 0 & 0 & 0,17 & 0 & 0,78 & 0,16 & 0,36 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0,04 & 0,22 & 0,045 & 0,3 \\ 0 & 0 & 0 & 0 & 0 & 0,14 & 0 & 0 & 0,12 & 0,12 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,038 & 0 & 0,036 & 0,10 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,17 & 0,007 & 0,11 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,0056 & 0,037 \end{bmatrix}$$

Ο πολλαπλασιασμός του πίνακα F^T (mXn) με τον ανάστροφό F (nXm) του δίνει έναν συμμετρικό πίνακα διαστάσεων (mXm), όπου $b_{ij}=b_{ji}$. Πολλαπλασιάζοντας λοιπόν τους πίνακες $F^T \cdot F$, έχω τον παρακάτω συμμετρικό πίνακα διαστάσεων 7X7:

$$F^T * F = \begin{bmatrix} 3,37 & 0,39 & 0,18 & 0,26 & 0,10 & 0,042 & 0,017 \\ 0,39 & 2,79 & 0,29 & 0,087 & 0,042 & 0,17 & 0,014 \\ 0,18 & 0,29 & 2,14 & 0,042 & 0,034 & 0,070 & 0,011 \\ 0,26 & 0,087 & 0,042 & 0,050 & 0,017 & 0,014 & 0,0052 \\ 0,10 & 0,042 & 0,034 & 0,017 & 0,013 & 0,011 & 0,0039 \\ 0,042 & 0,17 & 0,070 & 0,014 & 0,011 & 0,041 & 0,0040 \\ 0,017 & 0,014 & 0,011 & 0,0052 & 0,0039 & 0,0040 & 0,0014 \end{bmatrix}$$

Ο αντίστροφος του $F^T * F$ είναι ο παρακάτω πίνακας:

$$(F^T * F)^{-1} = \begin{bmatrix} 0,93 & 0,0015 & -0,00094 & -4,45 & -20,5 & -0,067 & 64,2 \\ 0,0015 & 0,49 & 0,00002 & -0,63 & -0,09 & -2,27 & 4,1 \\ -0,0009 & 0,00002 & 0,49 & 0,018 & -0,46 & -0,64 & -0,95 \\ -4,4 & -0,63 & 0,018 & 57,3 & 59,1 & 4,05 & -336 \\ -20,5 & -0,094 & -0,46 & 59,1 & 1030 & 4,8 & -2924 \\ -0,067 & -2,27 & -0,64 & 4,05 & 4,8 & 45,1 & -130 \\ 64,2 & 4,12 & -0,95 & -336 & -2924 & -130 & 9908 \end{bmatrix}$$

Τελικά:

$$b = (F^T * F)^{-1} * F^T * Y = [152,2 \quad 218,5 \quad 364,5 \quad 555,1 \quad 1909,0 \quad 301,4 \quad -7348]$$

Το παραπάνω λοιπόν διάνυσμα (αφού προκύπτει από πολλαπλασιασμό πινάκων με διάνυσμα) εξάγει τους συντελεστές, που προφανώς είναι οι ίδιοι με αυτούς που εξήγησαν στον πίνακα 12.1 μέσω της ελαχιστοποίησης του SS_R . Συνεπώς και οι δύο μεθοδολογίες εξήγαγαν τα ίδια αποτελέσματα, αφού και οι δύο αποτελούν εναλλακτικές μορφές της παλινδρόμησης μέσω ελαχίστων τετραγώνων.

13.2 Ακρίβεια συντελεστών

Για τον υπολογισμό της ακρίβειας των συντελεστών χρησιμοποιείται η σχέση:

$$\text{Var}(b) = (F^T \cdot F)^{-1} \cdot \sigma^2$$

Εδώ θέλει όμως προσοχή στα παρακάτω σημεία:

- ο Πολλαπλασιάζεται ένας πίνακας διαστάσεων $m \times m$ (δηλ. ο $(F^T * F)^{-1}$, που στην περίπτωση του παραδείγματος μας είναι 7×7) με ένα σταθερό αριθμό. Συνεπώς, πρέπει όλα τα στοιχεία του πίνακα να πολλαπλασιαστούν με τον αριθμό αυτό και ο πίνακας που εξάγεται είναι πάλι 7×7 .
- ο Αφού δεν υπάρχουν επαναλήψεις για κάθε πείραμα, ως διασπορά θα ληφθεί η συνολική διασπορά SS_R διαιρούμενη όμως με τους αντίστοιχους βαθμούς ελευθερίας, που εδώ είναι $10 - 7 = 3$, αφού μελετούμε 7 παραμέτρους. Άρα το σ^2 ταυτίζεται με την τιμή MS_R του πίνακα ANOVA (βλέπε κεφ. 12, πιν. 12.2). Η τιμή MS_R για 7 παραμέτρους και 10 πειράματα είναι 36,6, που προκύπτει από τη

διαίρεση του SS_R με $10-7=3$.

- ο Τελικώς λοιπόν έχουμε ότι: $Var(b) = (F^T \cdot F)^{-1} \cdot \frac{s^2}{n-p}$, όπου στην περίπτωση μας $n=10$ και $p=7$, ενώ $s^2 = SS_R = 109$.

Ο πίνακας 7×7 που εξάγεται μας δίνει τις διασπορές των 7 συντελεστών του μοντέλου, που φαίνονται στη **διαγώνιο του πίνακα**.

$$(F^T \cdot F)^{-1} \cdot \frac{s^2}{n-p} =$$

$$\begin{bmatrix} 0,93 & 0,0015 & -0,00094 & -4,45 & -20,5 & -0,067 & 64,2 \\ 0,0015 & 0,49 & 0,00002 & -0,63 & -0,09 & -2,27 & 4,1 \\ -0,0009 & 0,00002 & 0,49 & 0,018 & -0,46 & -0,64 & -0,95 \\ -4,4 & -0,63 & 0,018 & 57,3 & 59,1 & 4,05 & -336 \\ -20,5 & -0,094 & -0,46 & 59,1 & 1030 & 4,8 & -2924 \\ -0,067 & -2,27 & -0,64 & 4,05 & 4,8 & 45,1 & -130 \\ 64,2 & 4,12 & -0,95 & -336 & -2924 & -130 & 9908 \end{bmatrix} * \frac{109,7}{10-7} =$$

$$= \begin{bmatrix} 33,9 & 0,06 & -0,03 & -162,7 & -751 & -2,45 & 2350 \\ 0,06 & 18,3 & 0,001 & -23,2 & -3 & -83,1 & 151 \\ -0,03 & 0,001 & 18,3 & 0,7 & -17 & -23,5 & -35 \\ -162,7 & -23,2 & 0,7 & 2098,5 & 2164 & 148,3 & -12280 \\ -751 & -3,46 & -16,9 & 2164 & 37684 & 174 & -107041 \\ -2,45 & -83,1 & -23,5 & 148,3 & 174 & 1649 & -4772 \\ 2350 & 151 & -35 & -12280 & -107041 & -4772 & 362632 \end{bmatrix}$$

Στον παραπάνω πίνακα παρατηρώ ότι $b_{ij} = b_{ji}$, άρα ο πίνακας είναι συμμετρικός. Ο πίνακας αυτός ονομάζεται πίνακας διασποράς (covariance matrix). Οι τιμές που μας ενδιαφέρουν είναι οι τιμές της διαγώνιου, δηλαδή οι τιμές b_{ii} , όπου το i κυμαίνεται από 1 έως 7. Οι 7 αυτές τιμές είναι οι διασπορές των συντελεστών b . Τα τυπικά σφάλματα είναι προφανώς οι τετραγωνικές ρίζες των διασπορών αυτών. Δηλαδή:

$$\begin{array}{ll} Var(b_1)=33,9 & SE(b_1)=5,8 \\ Var(b_2)=18,3 & SE(b_2)=4,3 \\ Var(b_3)=18,3 & SE(b_3)=4,3 \\ Var(b_{12})=2099 & SE(b_{12})=45,8 \\ Var(b_{13})=37684 & SE(b_{13})=194 \\ Var(b_{23})=1649 & SE(b_{23})=40,6 \\ Var(b_{123})=362632 & SE(b_{123})=602 \end{array}$$

Με βάση τα παραπάνω τυπικά σφάλματα, μπορούν πλέον να υπολογιστούν τα διαστήματα εμπιστοσύνης για την κάθε παράμετρο b . Πρέπει φυσικά να διευκρινιστεί το επίπεδο σημαντικότητας (συνήθως λαμβάνουμε το 95%) και να υπολογιστεί το $t_{10-7,0.025}$

που εδώ είναι ίσο με 3,18. Παρατηρούμε ότι το στατιστικό t βασίζεται σε $n-p=10-7=3$ βαθμούς ελευθερίας, αφού εδώ έχουμε 7 μεταβλητές (ή καλύτερα 7 παραμέτρους) που δοκιμάζουμε στο μοντέλο μας, το οποίο βασίστηκε σε 10 πρωτογενείς τιμές. Αν δοκιμάσαμε διαφορετικό αριθμό μεταβλητών (παραμέτρων), τότε το στατιστικό t προφανώς θα άλλαζε. Τα αντίστοιχα διαστήματα εμπιστοσύνης των 7 συντελεστών b φαίνονται στον πίνακα 13.1:

Πίνακας 13.1. Διαστήματα εμπιστοσύνης 7 συντελεστών εμπειρικού μοντέλου

Συντελεστής	Τιμή	t*SE	Διάστημα εμπιστοσύνης
b ₁	152,2	±3.18*5,8	(133.73, 170.6)
b ₂	218,5	±3.18*4,3	(204.8, 232.1)
b ₃	364,5	±3.18*4,3	(350.9, 378.2)
b ₁₂	555,1	±3.18*45,8	(409.5, 700.8)
b ₁₃	1909	±3.18*194	(1292.1, 2525.9)
b ₂₃	301,4	±3.18*40,6	(172.3, 430.5)
b ₁₂₃	-7348	±3.18*602	(-5433.7, -9262.4)

Τα διαστήματα εμπιστοσύνης είναι τελικά αυτά που μας δίνουν την εγκυρότερη πληροφορία για τη στατιστική σημαντικότητα των συντελεστών. Εφόσον δεν περιέχουν την τιμή 0, ο συντελεστής είναι στατιστικά σημαντικός και η αντίστοιχη μεταβλητή μπορεί να παραμείνει στο μοντέλο.

Πλέον με τα στατιστικά πακέτα που κυκλοφορούν, δεν είναι απαραίτητα η χρήση των παραπάνω διαδικασιών. Άμεσα, εξάγονται οι εκτιμήσεις των παραμέτρων b καθώς και τα αντίστοιχα τυπικά σφάλματα (ή διαστήματα εμπιστοσύνης) χωρίς να φαίνονται οι ενδιάμεσες πράξεις μέσω πινάκων. Παραδειγματικά, εμφανίζεται στη συνέχεια ο τυπικός πίνακας εξόδου από το πρόγραμμα MINITAB για το εμπειρικό μοντέλο μιγμάτων στο οποίο διατηρήσαμε και τις 7 παραμέτρους b, όπως αυτές φαίνονται στην εξίσωση (3).

Regression for Mixtures: CO2 meas versus MXP; YW; FW

Term	Coef	SE Coef	T	P	VIF
MXP	152	5,821	*	*	3,120
YW	218	4,276	*	*	1,395
FW	365	4,276	*	*	1,071
MXP*YW	555	45,788	12,12	0,001	2,882
MXP*FW	1909	194,032	9,84	0,002	13,565
YW*FW	301	40,587	7,43	0,005	1,854
MXP*YW*FW	-7348	601,903	-12,21	0,001	13,668

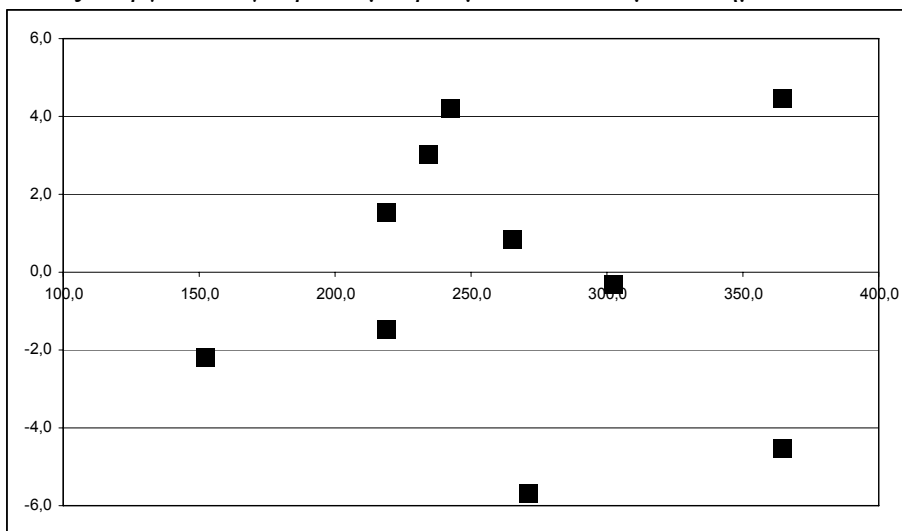
S = 6,04692 PRESS = 50682,9
 R-Sq = 99,72% R-Sq(pred) = 0,00% R-Sq(adj) = 99,17%

Όπως βλέπουμε στον παραπάνω πίνακα εξόδου του MINITAB, οι τιμές των παραμέτρων b και των αντίστοιχων τυπικών σφαλμάτων είναι οι ίδιες με αυτές που έχουμε βρει έως τώρα – τόσο με την τεχνική της ελαχιστοποίησης όσο και με την τεχνική των πινάκων. Αναφορικά με την τιμή t (συμβολίζεται ως T στον παραπάνω πίνακα), ο πρακτικός κανόνας λέει ότι αν είναι μεγαλύτερο από 2,5 τότε η αντίστοιχη παράμετρος είναι

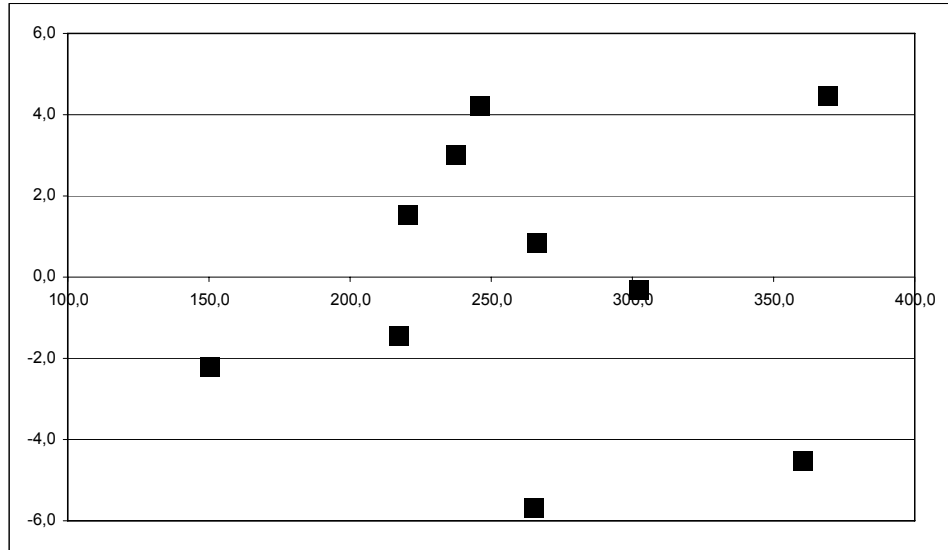
στατιστικά σημαντική. Τι σημαίνει αυτό? Σχετικά μεγάλες τιμές t σημαίνουν ότι η διαφορά της πραγματικής τιμής του συντελεστή από την εκτιμώμενη τιμή απέχει αρκετά από την τιμή 0. Αν βρίσκεται κοντά στο 0, τότε η διαφορά θεωρούμε ότι περιλαμβάνει το 0 και συνεπώς ο συντελεστής είναι στατιστικά μη σημαντικός. Όσον αφορά τις τιμές P , αν οι τιμές αυτές είναι μικρότερες από το επίπεδο σημαντικότητας που εμείς ορίζουμε (π.χ. συνήθως 0,025), τότε η παράμετρος είναι στατιστικά σημαντική. Παρατηρώ επίσης, ότι το R^2 είναι προφανώς το ίδιο με την τιμή που έχω εξάγει στο κεφάλαιο 12.

Συμπερασματικά

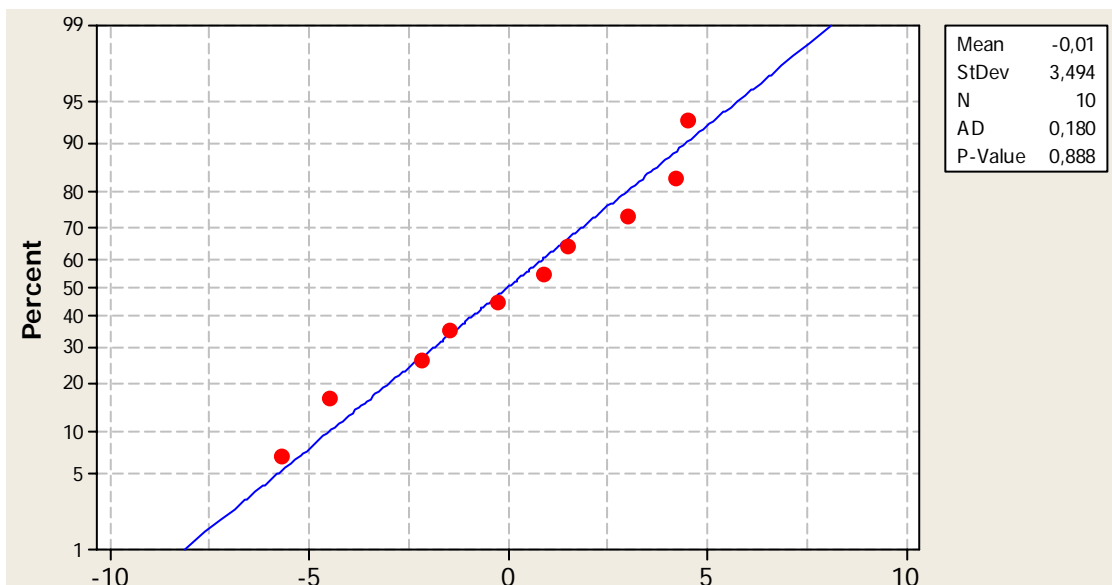
Με βάση τους παραπάνω ορισμούς που βασίζονται στην άλγεβρα πινάκων, οι διασπορές των συντελεστών b προσδιορίζονται λαμβάνοντας υπόψη τη διασπορά σ^2 . Εφόσον δεν υπάρχουν επαναλήψεις για κάθε πείραμα, ως σ^2 θα χρησιμοποιήσουμε την τιμή SS_R που είναι ίση με 109,7 διαιρούμενη με τους βαθμούς ελευθερίας που μεταβάλλονται ανάλογα με τον αριθμό των πειραμάτων και ανάλογα με τις παραμέτρους που θέλω να διατηρήσω στο μοντέλο μου. Τελικώς δηλαδή χρησιμοποιώ την τιμή MS_R , που θεωρούμε ότι εκφράζει το τυχαίο πειραματικό σφάλμα. Σε κάθε περίπτωση, αφού καταλήξουμε στο μοντέλο που επιθυμούμε και υπολογίσουμε την ακρίβεια των συντελεστών, δεν σταματάμε εκεί. Το επόμενο, ίσως πλέον σημαντικό βήμα, είναι τα ελέγχουμε την επάρκεια των μοντέλων μέσω των κλασικών διαγνωστικών γραφημάτων, όπως είναι γράφημα των σφαλμάτων (διαφορές προβλέψεων μοντέλου από πειραματικές τιμές) συναρτήσει των προβλεπόμενων τιμών, της σειράς των παρατηρήσεων, να ελέγχουμε την κανονικότητα των σφαλμάτων αυτών καθώς και την ανεξαρτησία τους. Τα σφάλματα αυτά πρέπει επίσης να έχουν μέση τιμή ίση με 0 και σταθερή διασπορά. Στη συνέχεια φαίνεται το γράφημα με τα σφάλματα του μοντέλου των 7 παραμέτρων. Δεν παρατηρείται κάποια συγκεκριμένη τάση στα 10 αυτά σφάλματα, αν και σίγουρα δεν μπορεί κάποιος να βγάλει σίγουρα συμπεράσματα από 10 μόνο σημεία.



Σχήμα 13.1. Διαγνωστικό γράφημα επάρκειας του μοντέλου. Εδώ έχω τα σφάλματα (residuals) συναρτήσει των προβλεπόμενων από το μοντέλο τιμών. Δεν παρατηρείται κάποια συγκεκριμένη τάση που αποτελεί ένδειξη για σταθερή διασπορά και συνεπώς ένδειξη επάρκειας του μοντέλου.



Σχήμα 13.2. Διαγνωστικό γράφημα επάρκειας του μοντέλου. Εδώ έχω τα σφάλματα (residuals) συναρτήσει των πραγματικών (παρατηρούμενων) τιμών. Ομοίως δεν παρατηρείται κάποια συγκεκριμένη τάση που αποτελεί ένδειξη για σταθερή διασπορά και συνεπώς ένδειξη επάρκειας του μοντέλου.



Σχήμα 13.3. Γράφημα πιθανοτήτων για τον έλεγχο της κανονικότητας των σφαλμάτων. Εφόσον τα σφάλματα βρίσκονται λίγο – πολύ πάνω σε μία ευθεία – όπως συμβαίνει εδώ - τότε μπορεί να θεωρηθεί ότι κατανέμονται κανονικότητα. Η κανονικότητα των σφαλμάτων είναι απαραίτητη συνθήκη για την επάρκεια ενός οποιουδήποτε μοντέλου και συνεπώς από τα σημαντικότερα διαγνωστικά τεστ.

13.3. Δοκιμές με διάφορα μοντέλα

Εδώ δοκιμάσαμε ένα γραμμικό πολυωνμικό μοντέλο τρίτου βαθμού. Δηλαδή δοκιμάσαμε 7 παραμέτρους που αντιστοιχούσαν στις μεταβλητές $F_1, F_2, F_3, F_1 * F_2, F_1 * F_3, F_2 * F_3, F_1 * F_2 * F_3$. Είναι αυτό το σωστό μοντέλο? Τι γίνεται αν δοκιμάσω μόνος τις 3 πρώτες μεταβλητές? Ή αν διατηρήσω μόνο τις διπλές αλληλοεπιδράσεις και δεν λάβω

υπόψη μου την τριπλή? Τι γίνεται αν (ξεχνώντας ότι αυτό είναι πείραμα μιγμάτων) εισάγω όρους όπως $F1^2$, $F2^2$ κ.λ.π. Στη συνέχεια θα δούμε πως μεταβάλλονται οι τιμές MS_R (που όσο μικρότερο είναι τόσο κοντύτερα το μοντέλο στα πραγματικά δεδομένα) καθώς και η τιμή του R^2 . Σε κάθε περίπτωση, ο στόχος μας είναι:

«Να βρούμε το απλούστερο μοντέλο, δηλαδή αυτό με τους λιγότερους όρους, που περιγράφει επαρκώς τα δεδομένα μας».

Το πώς αφαιρούμε (ή προσθέτουμε) όρους σε ένα μοντέλο φαίνεται στο παρακάτω παράδειγμα.

Παράδειγμα δημιουργίας εμπειρικού μοντέλου από πρωτογενή πειραματικά δεδομένα (βασισμένο σε Berthouex and Brown, 2002)

Έστω ότι κάνουμε ένα πείραμα στο οποίο μετράμε τη συγκέντρωση των αιωρούμενων στερεών σε μία στήλη σε διαφορετικά βάθη και σε διαφορετικούς χρόνους. Η αρχική συγκέντρωση των στερεών είναι 560 mg/L. Τα πρωτογενή μας δεδομένα είναι τα παρακάτω:

Πίνακας 13.2. Πρωτογενή δεδομένα για ανάπτυξη εμπειρικού μοντέλου

Βάθος (D) σε ft	Συγκέντρωση στερεών σε χρόνο T (min)			
	20	40	60	120
2	135	90	75	48
4	170	110	90	53
6	180	126	96	60

Βλέπουμε λοιπόν ότι όσο αυξάνεται ο χρόνος, μειώνεται η συγκέντρωση στο ίδιο βάθος δειγματοληψίας. Επίσης, σε όσο μικρότερο βάθος γίνεται η δειγματοληψία, τόσο μικρότερη η συγκέντρωση των στερεών, αφού τα τελευταία καθιζάνουν. Η εξαρτημένη μας μεταβλητή εδώ είναι η συγκέντρωση των στερεών (SC) και οι ανεξάρτητες μεταβλητές μας είναι το βάθος (D) και ο χρόνος (T). Θέλουμε λοιπόν να φτιάξουμε ένα μοντέλο της μορφής: $SC = f(D,T)$

Το μοντέλο αυτό θα είναι γραμμικό (όσον αφορά τους συντελεστές) αλλά μπορεί να έχει όρους δευτέρου ή τρίτου βαθμού όσον αφορά τις ανεξάρτητες μεταβλητές.

Παραδειγματικά, το πιο πολύπλοκο μοντέλο που μπορούμε να δοκιμάσουμε μπορεί να είναι της μορφής:

$$SC = b_0 + b_1 D + b_2 T + b_3 D^2 + b_4 T^2 + b_5 DT$$

ενώ σταδιακά μπορούμε να δοκιμάσουμε και μοντέλα που δεν έχουν ή έχουν κάποιους από τους παραπάνω όρους. Π.χ. $SC = b_0 + b_1 D + b_2 T + b_5 DT$ ή $SC = b_0 + b_1 D + b_2 T + b_3 D^2$ ή $SC = b_0 + b_1 D + b_2 T + b_4 T^2 + b_5 DT$ κ.λ.π. Ο στόχος μας είναι να βρούμε το απλούστερο μοντέλο που περιγράφει επαρκώς τα δεδομένα.

Χωρίς να περιγράψουμε τις παραπάνω μεθοδολογίες ξανά, χρησιμοποιούμε απευθείας ένα στατιστικό πακέτο (π.χ. MINITAB) και μέσω της ρουτίνας «Regression» δοκιμάζουμε τα διάφορα μοντέλα. Καταγράφουμε στην κάθε περίπτωση τα: SS_R , MS_R , R^2 και φυσικά τους συντελεστές b με τα αντίστοιχα τυπικά σφάλματά τους και τις τιμές t . Υπενθυμίζεται ότι τιμές t μεγαλύτερες του 2,5 περίπου δηλώνουν τη

στατιστική σημαντικότητα του συντελεστή (σε επίπεδο σημαντικότητας περίπου 95%) και συνεπώς το ότι ο αντίστοιχος όρος μπορεί να παραμείνει στο μοντέλο.

Τα δεδομένα στο στατιστικό μας πακέτο εισάγονται (συνήθως) εντός 3 στηλών, όπως παραδειγματικά φαίνεται παρακάτω. Εισάγονται τα δεδομένα της εξαρτημένες μεταβλητής σε μία στήλη (εδώ είναι η SC). Σε διπλανές στήλες εισάγονται τα δεδομένα των ανεξάρτητων μεταβλητών που αντιστοιχούν στην κάθε εξαρτημένη μεταβλητή. Στην περίπτωση του παραδείγματος, τα δεδομένα εισάγονται σε 3 στήλες με τη μορφή. Σε περίπτωση που καταγράφαμε και άλλες ανεξάρτητες μεταβλητές (π.χ. θερμοκρασία) τότε θα εισάγαμε μία επιπλέον στήλη.

SC (εξαρτημένη)	D (ανεξάρτητη)	T (ανεξάρτητη)
135	2	20
90	2	40
75	2	60
48	2	120
170	4	20
110	4	40
90	4	60
53	4	120
180	6	20
126	6	40
96	6	60
60	6	120

Επειδή εμείς θέλουμε να δοκιμάσουμε και άλλους όρους (D^2 , T^2 , $D*T$)³² δημιουργούμε επιπλέον στήλες που προκύπτουν από το τετράγωνο των τιμών στις στήλες D & T αντίστοιχα και από το γινόμενο των στηλών D, T. Κάνουμε δηλαδή οι ίδιοι τους πολλαπλασιασμούς και εισάγουμε τα δεδομένα, ως νέες μεταβλητές, στους χώρους εισαγωγής του στατιστικού μας προγράμματος σε ξεχωριστές στήλες. Έτσι τελικώς, τα δεδομένα μας θα φαίνονται ως εξής:

SC	D	T	D*T	D ²	T ²
135	2	20	40	4	400
90	2	40	80	4	1600
75	2	60	120	4	3600
48	2	120	240	4	14400
170	4	20	80	16	400
110	4	40	160	16	1600
90	4	60	240	16	3600
53	4	120	480	16	14400

³² Πιθανά να θέλαμε να δοκιμάσουμε και την 3^{ου} βαθμού μεταβλητή D^3 ή T^3 , αλλά γενικά δεν συνιστάται να φορτώνουμε εξ αρχής το μοντέλο μας με υψηλού βαθμού όρους.

180	6	20	120	36	400
126	6	40	240	36	1600
96	6	60	360	36	3600
60	6	120	720	36	14400

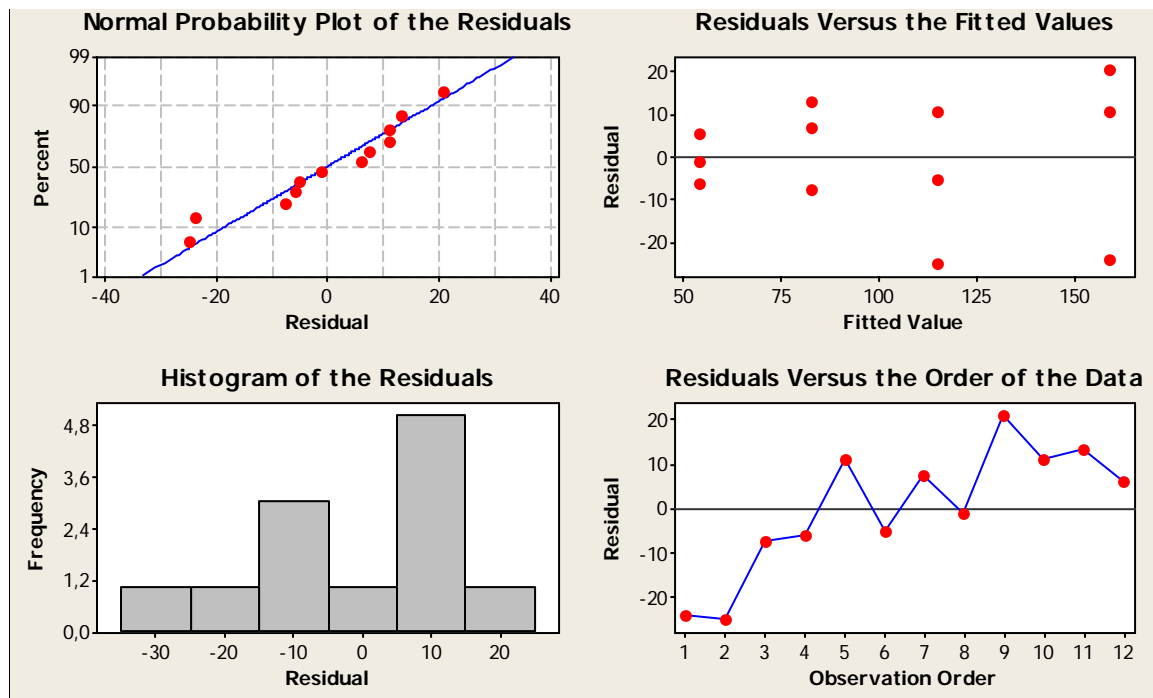
Ο πίνακας παραπάνω περιέχει όλες τις πληροφορίες που απαιτούνται για να δοκιμάσουμε τα διάφορα μοντέλα μας. Στη συνέχεια κάνουμε χρήση της ρουτίνας Regression (σε MINITAB) και προσθέτουμε ή αφαιρούμε τους όρους που θέλουμε. Οι ανεξάρτητες μεταβλητές ονομάζονται «predictors» και η εξαρτημένη μεταβλητή ονομάζεται “response”. Στο παράδειγμά μας ο μέγιστος αριθμός των predictors είναι 5. Ανάλογα επιλέγουμε αν θέλουμε σταθερά (constant, b_0) στο μοντέλο μας επιλέγοντας ή όχι ανάλογα του κουτί Fit Intercept.

Η μεθοδολογία εύρεσης του βέλτιστου μοντέλου θα βασιστεί στα ελάχιστα τετράγωνα. Ο πίνακας στη συνέχεια περιέχει τους όρους που έχουμε διατηρήσει στο μοντέλο, την αντίστοιχη τιμή της παραμέτρου του κάθε όρου και τις τιμές των R^2 και MS_R . Η παράμετρος κλειδί για να μας βοηθήσει να διατηρήσουμε ή όχι κάποιον όρο στο τελικό μας μοντέλο είναι η MS_R . Αν και είναι καταρχάς λογικό να μας εντυπωσιάζουν οι υψηλές τιμές R^2 , δεν σημαίνει απαραίτητα ότι το μοντέλο με την υψηλότερη τιμή είναι και το καλύτερο. Δηλαδή δεν είναι απαραίτητα το μοντέλο A ή το μοντέλο D τα καλύτερα μοντέλα επειδή έχουν τις υψηλότερες τιμές R^2 . Στόχος μας είπαμε να βρούμε το απλούστερο μοντέλο με τους λιγότερους όρους, που επαρκώς περιγράφει τα δεδομένα μας. Πως γίνεται αυτό?

Ένας τρόπος είναι να κοιτάξουμε τα τυπικά σφάλματα των συντελεστών b και να υπολογίσουμε τα διαστήματα εμπιστοσύνης με τον τρόπο που γνωρίζουμε (δηλ. $b \pm SE * t_{n-p, 0,025}$). Γενικά, αν οι βαθμοί ελευθερίας του μοντέλου ($n-p$)³³ είναι άνω των 20, τότε το τυπικό σφάλμα πολλαπλασιάζεται με 2 ώστε να βρεθεί το διάστημα εμπιστοσύνης. Αν $n-p < 10$, τότε το τυπικό σφάλμα πολλαπλασιάζεται με τιμές που κυμαίνονται από 2.3 έως 3.0. Κοιτώντας το μοντέλο A (που είναι το πλέον πολύπλοκο ή «φορτωμένο») βλέπουμε ότι τα όντως κάποιοι συντελεστές έχουν τιμές t μικρότερες του 2.5. Παράδειγμα, ο συντελεστής b_1 έχει τυπικό σφάλμα 9.1. Το διάστημα εμπιστοσύνης για το συντελεστή αυτό σε ε.σ. 95% είναι: $\pm t_{12-6, 0,025} * SE = \pm 2.447 * 9.1 = \pm 22.3$ άρα (20.9-22.3 έως 20.9+22.3) ή (-1,4 έως 43,2). Το Δ.Ε. αυτό περιλαμβάνει το 0 άρα η πραγματική τιμή του μάλλον περιλαμβάνει το 0. Συνεπώς ο όρος αυτός πρέπει να αφαιρεθεί από το μοντέλο. Το ίδιο κάνουμε και με τους υπόλοιπους όρους στους οποίους ο αντίστοιχος συντελεστής έχει Δ.Ε. που περιέχει το 0. Με αυτόν τον τρόπο μπορεί να απλοποιηθεί το μοντέλο. Κοιτώντας λοιπόν το πλέον πολύπλοκο μοντέλο A, μπορώ με το παραπάνω σκεπτικό να αφαιρέσω τους όρους D, D^2 , $D * T$. Καταλήγω έτσι στο μοντέλο I. Ξαναπροσαρμόζουμε λοιπόν το μοντέλο μας στα πραγματικά δεδομένα έχοντας αφαιρέσει τους όρους που δεν επιθυμούμε (ή αντίστροφα που επιθυμούμε να παραμείνουν στο μοντέλο). Στο μοντέλο I βλέπουμε ότι όλοι οι όροι είναι σημαντικοί στο Ε.Σ. 95%. Πρακτικώς το μοντέλο I μας λέει ότι η συγκέντρωση των στερεών είναι ανεξάρτητη του βάθους και εξαρτάται μόνο από το χρόνο. Αυτό είναι ένα απλούστερο

³³ n είναι ο αριθμός των πειραμάτων (πραγματικών τιμών) και p ο αριθμός των παραμέτρων (συντελεστών) που εισάγουμε στο μοντέλο μας.

μοντέλο που περιγράφει επαρκώς τα δεδομένα. Δεν μένει παρά να κοιτάξουμε την κατανομή των σφαλμάτων, καθώς και τα υπόλοιπα διαγνωστικά τεστ των σφαλμάτων ώστε να πειστούμε για την επάρκεια του μοντέλου.



Σχήμα 13.4. Διαγνωστικά τεστ σφαλμάτων μοντέλου I (βλέπε πίνακα 13.3) όπως εξάγονται από το στατιστικό πρόγραμμα MINITAB 14.

Το σχήμα 13.4. δείχνει ότι τα σφάλματα δεν κατανέμονται κανονικά, αν βασιστούμε στο ιστόγραμμα μόνο, αν και το γράφημα πιθανοτήτων μάλλον θα μας οδηγούσε σε διαφορετικό συμπέρασμα. Επίσης παρατηρείται αύξηση του μεγέθους των σφαλμάτων (residuals) με τη σειρά παρατήρησης, όπως η σειρά παρατήρησης έχει τεθεί σε παραπάνω πίνακα. Το γράφημα των σφαλμάτων συναρτήσει των προβλεπόμενων τιμών δείχνει μία τάση «χωνιού» (funneling). Δηλαδή ότι τα σφάλματα μεγαλώνουν όσο μεγαλώνουν οι προβλεπόμενες τιμές. Γενικά τα διαγνωστικά τεστ δεν είναι τόσο καλά ώστε να μπορέσουμε να χαρακτηρίσουμε το μοντέλο μας επαρκές. Βλέπουμε λοιπόν, ότι πέρα από τα διαστήματα εμπιστοσύνης των συντελεστών, τον τελευταίο λόγο τον έχουν τα διαγνωστικά τεστ των σφαλμάτων στην αποδοχή ή όχι ενός μοντέλου.

13.4 Μεθοδολογία ανάπτυξης μοντέλων βασισμένη σε F τεστ

Ο άλλος τρόπος να προχωρήσουμε στην ανάπτυξη του ορθού μοντέλου είναι να ξεκινήσουμε από το πλέον πολύπλοκο μοντέλο. Στη συνέχεια προχωρούμε αφαιρώντας όρους και συγκρίνουμε τη μεταβολή του αθροίσματος των τετραγώνων λόγω παλινδρόμησης (SS_{reg}), από την αφαίρεση του όρου, με το άθροισμα των τετραγώνων λόγω του πειραματικού σφάλματος του πολύπλοκου μοντέλου (SS_{res}). Υπενθυμίζεται ότι όσο μεγαλύτερο το SS_{reg} , τόσο η παλινδρόμηση θεωρείται καλύτερη. Στην ιδανική περίπτωση που το SS_{reg} είναι ίσο με το SS_{tot} τότε το μοντέλο περνάει ακριβώς «πάνω» από όλα τα πειραματικά δεδομένα (και το R^2 είναι ίσο με 100%). Η περίπτωση αυτή

είναι σπάνια και μάλλον σημαίνει ότι το μοντέλο είναι υπερφορτωμένο με πολλούς υψηλού βαθμού όρους.

Ας προχωρήσουμε σε ένα παράδειγμα. Ας πάρουμε λοιπόν το πλέον πολύπλοκο μοντέλο A (που περιέχει και τη σταθερά b_0). Το αντίστοιχο SS_{regr} για το A είναι η τιμή 20256. Αν αφαιρέσω έναν όρο (π.χ. τον όρο D^*T) τότε πηγαίνω στο μοντέλο C με $SS_{\text{regr}} = 19966$. Η μεταβολή λοιπόν του SS_{regr} από το μοντέλο A στο C είναι 290 και επειδή μεταβλήθηκε ένας όρος από το A στο C, οι βαθμοί ελευθερίας που αντιστοιχούν στη διαφορά 290 είναι $k=1$. Αν θεωρήσω ότι αυτή είναι μία διασπορά που σχετίζεται με τον όρο D^*T , τότε θα πρέπει να τη συγκρίνω με το απλό πειραματικό σφάλμα για να δω αν όντως ο όρος D^*T παρέχει κάποια ουσιαστική πληροφορία στο μοντέλο ή αν όχι, τότε θα πρέπει να αφαιρεθεί. Αν δηλαδή η διασπορά που σχετίζεται με τον όρο D^*T είναι αρκετά μεγάλη σε σχέση με το τυχαίο πειραματικό σφάλμα, τότε θα πρέπει να διατηρήσω τον όρο αυτό στο μοντέλο. Για το λόγο αυτό κάνω ένα F test. Χρησιμοποιώ δηλαδή το λόγο:

$$F = \frac{\Delta SS_{\text{regr}}/k}{s^2} = \frac{\Delta SS_{\text{regr}}/k}{SS_{\text{res}}/n-p} = \frac{\Delta SS_{\text{regr}}/k}{MS_{\text{res}}}$$

, όπου k είναι οι βαθμοί ελευθερίας της μεταβολής του SS_{regr} , που εδώ είναι ίσο με 1. Η διαφορά $n-p$ αφορά στο μοντέλο A, οπότε έχουμε n πειραματικά δεδομένα (εδώ είναι 12) και p παραμέτρους (στο A είναι 6). Άρα $n-p=6$ πάντα.

Επειδή δεν έχω επαναλήψεις των πειραμάτων ώστε να εξάγω ένα πραγματικό πειραματικό σφάλμα, θα χρησιμοποιήσω ως εκτίμηση του πειραματικού σφάλματος (όπως έχω κάνει και παλαιότερα) τη μεταβλητή MS_{res} (δες πίνακα 13.3). Η μεταβλητή αυτή είναι ίση με 52 για το A. Ο λόγος της διασποράς λόγω του όρου D^*T προς τη διασπορά των πειραματικών σφαλμάτων είναι λοιπόν $F = (290:1) / 52 = 5,57$. Αυτή η τιμή συγκρίνεται με το άνω 5% της κατανομής F με (1,6) βαθμούς ελευθερίας, δηλαδή με την τιμή $F_{1,6,0.05}$. Οι βαθμοί ελευθερίας είναι 1 για τον αριθμητή (αφού αφαιρώ 1 όρο για να πάω από το μοντέλο A στο C) και 6 για τον παρονομαστή, αφού έχω $n-p=12-6=6$. Η τιμή $F_{1,6,0.05}$ είναι 5.99. Επειδή $5,57 < 5.99$, τότε αφαιρώντας τον όρο D^*T από το μοντέλο δεν οδηγεί σε στατιστικά σημαντική μείωση του αθροίσματος των τετραγώνων λόγω παλινδρόμησης. Συνεπώς, ο όρος αυτός μπορεί να αφαιρεθεί και δεν χρειάζεται στο τελικό μοντέλο.

Με τον ίδιο τρόπο προχωρούμε ώστε να ελέγξουμε όλα τα μοντέλα που έχουν (καταρχάς) ένα λιγότερο όρο από το μοντέλο A. Δηλαδή τα μοντέλα D και E. Στη συνέχεια προχωρούμε στο να ελέγξουμε και μοντέλα που έχουν λιγότερες δύο παραμέτρους, όπως το μοντέλο F & K, κ.ο.κ. Εδώ προσοχή θέλει στην εκτίμηση του F που πλέον βασίζεται σε 2,6 βαθμούς ελευθερίας. Το $F_{2,6,0.05}$ είναι πλέον 5,14 και όχι 5.99. Ομοίως προχωρώ αν αφαιρέσω 3 όρους κ.ο.κ. Ο παρακάτω πίνακας περιλαμβάνει συγκρίσεις μεταξύ των μοντέλων που έχουν την σταθερά b_0 .

Πίνακας 13.3. Όροι και συντελεστές εμπειρικών μοντέλων που βασίζονται στα πειραματικά δεδομένα του πίνακα 13.2*

Μοντέλο	b ₀	b ₁ (D)	b ₂ (T)	b ₃ (D ²)	b ₄ (T ²)	b ₅ (D*T)	SS _{Regr}	MS _{Res}	R ²
A	152 (19,5) (7,8)	20,9 (9,1) (2,3)	-2,7 (0,33) (-8,27)	-1,1 (1,1) (-1,02)	0,014 (0,002) (7,04)	-0,08 (0,03) (-2,4)	20256	52	98,5%
B		81,5 (14,6) (5,6)	-1,3 (0,85) (-1,55)	-7,5 (2,2) (-3,38)	0,0081 (0,0058) (1,41)	-0,19 (0,095) (-2,01)	143828	490	97,7%
C	171 (22,8) (7,5)	16,1 (11,4) (1,41)	-3,1 (0,4) (-7,9)	-1,1 (1,4) (-0,80)	0,014 (0,003) (5,5)		19966	85,4	97,1%
D	167 (12,9) (12,9)	11,9 (2,4) (5,0)	-2,7 (0,33) (-8,2)		0,014 (0,002) (7,02)	-0,08 (0,03) (-2,4)	20202	52	98,2%
E	98 (50,5) (1,9)	21 (26) (0,4)	-0,65 (0,4) (-1,6)	-1,1 (3,1) (-0,36)		-0,08 (0,095) (-0,84)	17705	408	86,1%
F	113 (27,5) (4,1)	11,9 (6,4) (1,9)	-0,65 (0,39) (-1,66)			-0,08 (0,09) (-0,89)	17651	364	85,8%
G	132 (16,85) (7,85)	7,1 (3,33) (2,14)	-0,97 (0,15) (-6,7)				17362	355,8	84,4%
H		27,9 (5,4) (5,14)	-0,38 (0,33) (-1,13)				122101	2515	83%
I	214,5 (19) (11,3)		-3,1 (0,67) (-4,6)		0,014 (0,0045) (3,2)		18288	253	88,9%
K	186 (12,5) (14,8)	7,1 (1,6) (4,5)	-3,1 (0,4) (-8,05)		0,014 (0,0026) (5,6)		19912	81,5	96,8%

*: η πρώτη τιμή στην παρένθεση είναι το τυπικό σφάλμα και η δεύτερη τιμή είναι η τιμή t του κάθε συντελεστή; SS_{Regr} =

$$\sum(\hat{y}_i - \bar{\hat{y}})^2, SS_{res} = \sum(\hat{y}_i - y_i)^2, SS_{tot} = \sum(y_i - \bar{y})^2, MS_{Res} = SS_{res} / df, R^2 = \frac{SS_{Regr}}{SS_{tot}};$$

Ο παρακάτω πίνακας δείχνει τη μείωση των SS_{regr} για το κάθε μοντέλο και τις αντίστοιχες τιμές F.

Πίνακας 13.4. Σύγκριση διαφορετικών εμπειρικών μοντέλων βασιζόμενες στο λόγο F

Μοντέλο	Διαφορά σε SS_{regr} από μοντέλο A	Αριθμός όρων που αφαιρώ από A	Λόγος F	Σημαντικός όρος
C	290	1 (όρος D*T)	290/52=5,57	5.57 < 5.99 OXI
D	54	1 (όρος D ²)	54/52 = 1.05	1.05 < 5.99 OXI
E	2551	1 (όρος T ²)	2551/52=49.1	49.1 >> 5.99 ΝΑΙ
K	344	2 (όροι D*T, D ²)	(344:2)/52=3.3	3.3 < 5.14 OXI
G	2894	3 (όροι D*T, D ² , T ²)	(2894:3)/52=18.6	18.6 > 4.76 ΝΑΙ

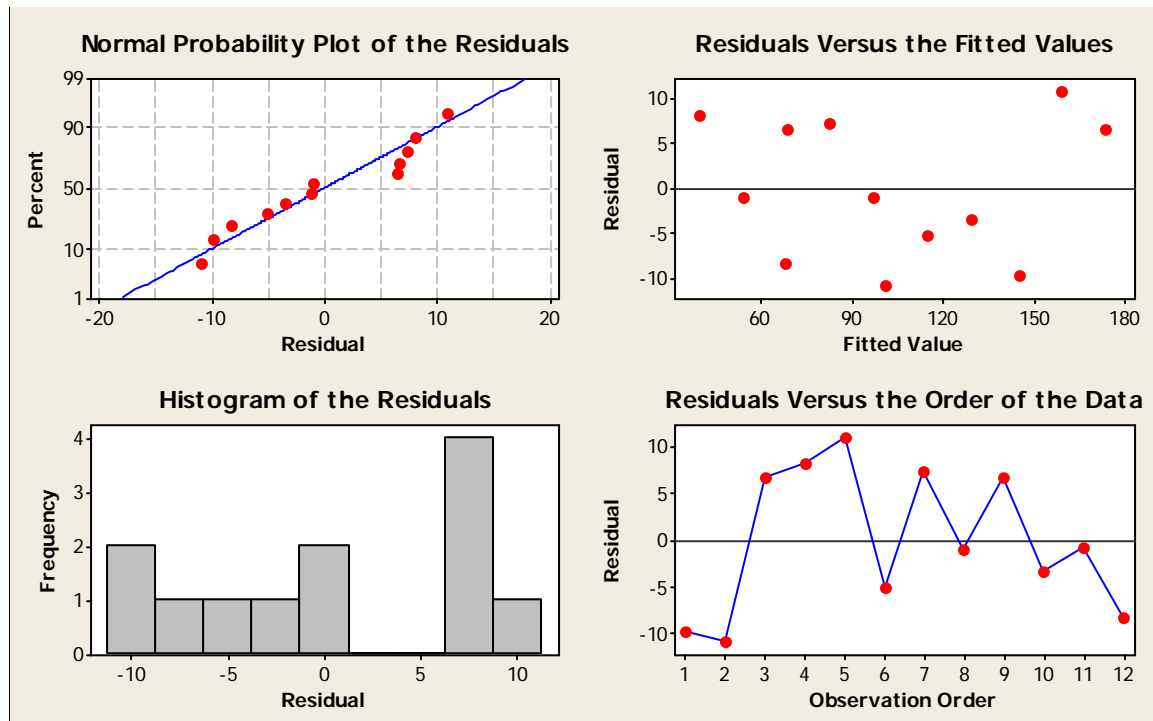
Από τα παραπάνω βλέπω ότι αφαιρώντας τον όρο D² & D*T (μοντέλα D & C αντίστοιχα) δεν επηρεάζεται σημαντικά το μοντέλο. Δηλαδή μπορώ να προχωρήσω στην αφαίρεση των όρων αυτών έχοντας ένα μοντέλο απλούστερο, το οποίο μπορεί να προβλέψει (ή περιγράψει) με απλούστερη μορφή τα πειραματικά μου δεδομένα. Με βάση την παραπάνω πληροφορία, αναπτύσσω ένα μοντέλο χωρίς τους όρους D² & D*T (μοντέλο K) και το συγκρίνω με την ίδια λογική με το μοντέλο A. ΠΡΟΣΟΧΗ: Στην περίπτωση του K έχω αφαιρέσει δύο όρους. Παρόλο που η συνολική διαφορά στο SS_{regr} είναι 344, η διαφορά ανά βαθμό ελευθερίας είναι 344/2=172, αφού 2 βαθμοί ελευθερίας σχετίζονται με το ΔSS_{regr} . Την τιμή 172 χρησιμοποιώ τελικά στον αριθμητή του λόγου F ώστε να τον συγκρίνω με το $F_{2,6,0.05}$. Ας συγκρίνω και το μοντέλο G (που έχει 3 λιγότερους όρους) με το A. Παρατηρώ ότι αν αφαιρέσω 3 όρους (δηλαδή έναν όρο επιπλέον σε σχέση με το K, ήτοι τον όρο T², στο μοντέλο G), καταλήγω σε ένα μοντέλο που διαφέρει σημαντικά από το αρχικό A, αφού ο λόγος F που υπολογίζεται είναι 18.6. Το μοντέλο G δεν μπορεί να προβλέψει λοιπόν επαρκώς τα δεδομένα μας και δεν συνίσταται, κάτι που επιβεβαιώνει το ότι ο όρος T² πρέπει να παραμείνει στο τελικό μοντέλο.

Βάσει των παραπάνω, καταλήγω ότι το μοντέλο K (στο οποίο έχουν αφαιρεθεί οι όροι D² & D*T) είναι το πλέον κατάλληλο και απλούστερο μοντέλο που περιγράφει επαρκώς τα δεδομένα μας. Το μοντέλο K λοιπόν είναι:

$$SC = 186 + 7.1 * D - 3.1 * T + 0.014 * T^2$$

Φυσικά δεν πρέπει να σταματήσουμε εδώ. Πρέπει να ελέγξουμε τα διαστήματα εμπιστοσύνης των 4 παραμέτρων και στη συνέχεια να κοιτάξουμε τα διαγνωστικά γραφήματα των σφαλμάτων. Όσον αφορά τα Δ.Ε. είδαμε ότι όλες οι τιμές t των εν λόγω συντελεστών είναι άνω του 2.5 (δες πίνακα 13.3 για το μοντέλο K) και συνεπώς κρίνονται ως στατιστικά σημαντικοί. Στο ίδιο συμπέρασμα θα κατέλγα αν υπολόγιζα τα Δ.Ε. μέσω των τυπικών σφαλμάτων (που φαίνονται στον πίνακα) και των τιμών t, που διαφοροποιείται ανάλογα με το μοντέλο. ΠΡΟΣΟΧΗ: Αυτό δεν είναι το ίδιο t με αυτό που φαίνεται στον πίνακα 13.3.

Τα διαγνωστικά γραφήματα των σφαλμάτων του μοντέλου K φαίνονται στη συνέχεια:



Σχήμα 13.5. Διαγνωστικά γραφήματα σφαλμάτων μοντέλου K.

Από το σχήμα 13.5 βλέπω ότι δεν παρατηρείται κάποια τάση των τιμών των σφαλμάτων σε σχέση με τις προβλεπόμενες τιμές (σταθερή διασπορά). Ομοίως δεν παρατηρείται κάποια τάση με τη σειρά της παρατήρησης. Το γράφημα πιθανοτήτων υποδεικνύει την κανονικότητα των σφαλμάτων, αν και το ιστόγραμμα δεν υποδεικνύει το στηρίζει αυτό. Σε κάθε περίπτωση, τα διαγνωστικά γραφήματα των σφαλμάτων θεωρούνται κατάλληλα και υποδεικνύουν την καταλληλότητα του μοντέλου K στο να περιγράψει επαρκώς τα δεδομένα.

Συμπέρασμα

Ο στόχος κατά την ανάπτυξη εμπειρικών μοντέλων δεν πρέπει να είναι να αυξήσουμε το R^2 όσο το δυνατό περισσότερο. Μια τέτοια προσέγγιση έχει ως αποτέλεσμα την προσθήκη μη απαιτούμενων υψηλόβαθμων όρων στο μοντέλο. Το καλύτερο μοντέλο θα πρέπει να έχει τις λιγότερες δυνατές παραμέτρους διότι έτσι ελαχιστοποιείται το σφάλμα πρόβλεψης του μοντέλου.

Μία προσέγγιση για να βρεθεί το απλούστερο επαρκές μοντέλο είναι να ξεκινήσει κάποιος με ένα απλό μοντέλο βασισμένο σε διαστήματα εμπιστοσύνης των συντελεστών και να χρησιμοποιήσει τα διαγνωστικά γραφήματα των σφαλμάτων ώστε να ελέγξει την επάρκεια του μοντέλου. Η εναλλακτική τεχνική είναι να ξεκινήσει κάποιος με το πλέον πολύπλοκο μοντέλο και να προχωρήσει σε απλοποιήσεις σταδιακά αφαιρώντας όρους. Κάθε φορά που κάποιος όρος προστίθεται ή αφαιρείται, έλεγχος πρέπει να γίνεται αν η διαφορά του αθροίσματος των τετραγώνων από την παλινδρόμηση του νεότερου μοντέλου από το προηγούμενο αντίστοιχο άθροισμα είναι αρκετά μεγάλη ώστε να δικαιολογεί την τροποποίηση.

14. Ο συντελεστής απόφασης R^2

Ο συντελεστής R^2 ορίστηκε στο κεφάλαιο 12 και αποτελεί το λόγο των αθροισμάτων των τετραγώνων των σφαλμάτων λόγω της παλινδρόμησης προς το άθροισμα των τετραγώνων των συνολικών σφαλμάτων. Είναι δηλαδή:

$$R^2 = \frac{SS_{Regr}}{SS_{tot}}$$

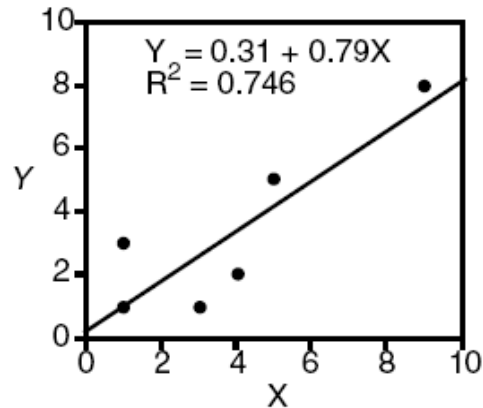
με όρους όπως έχουν οριστεί στο κεφ. 12. Ο R^2 είναι ευρέως χρησιμοποιούμενος κατά την παλινδρόμηση δεδομένων και την ανάπτυξη απλών εμπειρικών μοντέλων. Γενικά υπάρχει η λάθος άποψη, ότι κάποιο εμπειρικό μοντέλο γίνεται καλύτερο, όσο αυξάνεται ο συντελεστής R^2 . Υπάρχει όμως και μία στατιστική σημαντικότητα με τον R^2 . Δηλαδή, υπάρχει ένα ποσοστό πιθανότητας ώστε να συμπεράνουμε λάθος ότι υπάρχει σημαντικότητα όταν δεν υπάρχει πραγματική συσχέτιση. Ο πίνακας στη συνέχεια δείχνει ακριβώς τις ελάχιστες απαιτούμενες τιμές R^2 για να καταλήξουμε σε επίπεδα στατιστικής σημαντικότητας συναρτήσει του μεγέθους του δείγματος για μία απλή γραμμική παλινδρόμηση. Τι σημαίνει αυτό? Παράδειγμα, όταν έχω 10 πειραματικές τιμές ($n=10$), τότε αν το απλό γραμμικό μου μοντέλο πετύχει ένα ελάχιστο $R^2 = 59\%$, τότε έχω πιθανότητα 1% να κάνω λάθος και να συμπεράνω ότι δεν υπάρχει συσχέτιση μεταξύ των μεταβλητών της παλινδρόμησης. Για τα ίδια πειραματικά δεδομένα, απαιτείται ένα ελάχιστο $R^2=30\%$ ώστε να μπορώ να πω με 90% βεβαιότητα ότι υπάρχει συσχέτιση μεταξύ των μεταβλητών μου.

Πίνακας 14.1. Απαιτούμενες τιμές R^2 για να επιτευχθεί στατιστική σημαντικότητα στη συσχέτιση για το απλό γραμμικό μοντέλο τύπου $y = \beta_0 + \beta_1 x + \varepsilon$

Sample Size <i>n</i>	Statistical Significance Level		
	10%	5%	1%
3	0.98	0.99	0.99
4	0.81	0.90	0.98
5	0.65	0.77	0.92
6	0.53	0.66	0.84
8	0.39	0.50	0.70
10	0.30	0.40	0.59
12	0.25	0.33	0.50
15	0.19	0.26	0.41
20	0.14	0.20	0.31
25	0.11	0.16	0.26
30	0.09	0.13	0.22
40	0.07	0.10	0.16
50	0.05	0.08	0.13
100	0.03	0.04	0.07

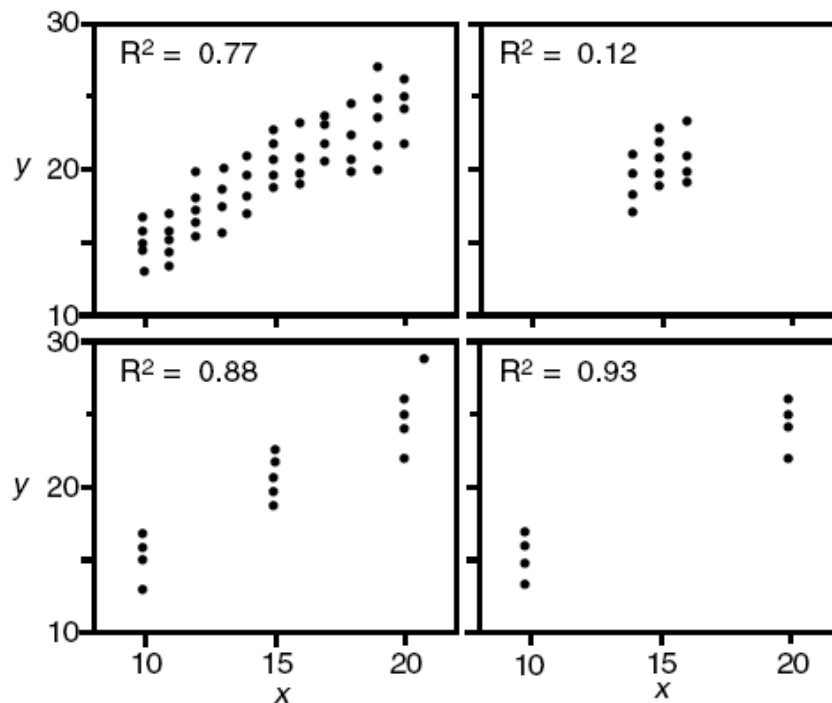
Το βασικό που εξάγεται από τον πίνακα 14.1 είναι ότι η σημασία του R^2 εξαρτάται από το μέγεθος του δείγματος. Μπορούμε να έχουμε σχετικά μικρές τιμές R^2 που να υποδηλώνουν στατιστική σημαντικότητα, εφόσον το δείγμα είναι σχετικά μεγάλο. Παραδειγματικά, για δείγμα 100 μετρήσεων ($n=100$), απαιτείται ένα ελάχιστο R^2 ίσο με 7% ώστε να είμαστε σίγουροι με πιθανότητα 99% ότι υπάρχει στατιστική σημαντικότητα της συσχέτισης μεταξύ x και y για το απλό γραμμικό μοντέλο που προαναφέραμε.

Το βασικό στη συσχέτιση κάποιων μεταβλητών x, y , είναι η σχέση αίτιου με αιτιατό. Δεν έχει κανένα νόημα να συσχετίσουμε μεταβλητές που δεν έχουν καμία σχέση μεταξύ τους, απλά επειδή τυγχάνει να υπολογίζονται υψηλές τιμές R^2 , όπως φαίνεται στο παρακάτω γράφημα. Το σχήμα 14.1 είναι κλασική περίπτωση κατάχρησης του συντελεστή R^2 .



Σχήμα 14.1. Συσχέτιση μεταξύ των 6 αρχικών αριθμών του π και των 6 αρχικών αριθμών της σειράς Fibonacci. Δεν υπάρχει καμία απολύτως σχέση μεταξύ αυτών των αριθμών, παρόλα αυτά επιτυγχάνεται μία γραμμική συσχέτιση με συντελεστή $R^2 = 75\%$. Αυτό είναι ένα τυπικό παράδειγμα κατάχρησης του R^2 , αφού ο έβδομος αριθμός του π δεν προβλέπει τον έβδομο αριθμό Fibonacci. Βασική προϋπόθεση στη χρήση του συντελεστή R^2 λοιπόν είναι η σχέση αίτιου – αιτιατού.

Το μέγεθος του R^2 εξαρτάται εξάλλου από το εύρος των δεδομένων. Η τιμή του R^2 μειώνεται με μείωση του εύρους των ανεξάρτητων μεταβλητών και αντίστροφα, εφόσον το σωστό μοντέλο χρησιμοποιείται για να περιγράψει τα δεδομένα. Στο σχήμα 14.2, βλέπω τις μεταβολές του R^2 για το ίδιο μοντέλο, όταν αλλάζει το εύρος των δεδομένων x (ανεξάρτητων μεταβλητών). Παραδειγματικά, αν αντί για 50 δεδομένα με εύρος από 10 έως 20 (η συσχέτιση του γραμμικού μου μοντέλου δίνει $R^2 = 77\%$) χρησιμοποιήσω τα δεδομένα που βρίσκονται στο εύρος 14 με 16, βλέπω ότι για το ίδιο γραμμικό μοντέλο μειώνεται αρκετά το R^2 (στο 12%). Ο λόγος της μείωσης αυτής είναι επειδή μειώνεται το εύρος του x και όχι επειδή μειώνεται ο αριθμός των δεδομένων (από 50 σε 13). Παράδειγμα, αν χρησιμοποιήσω τον ίδιο αριθμό δεδομένων (δηλαδή περίπου 13) στο εύρος 10 με 20 με θέσεις x στο 10, 15, 20 επιτυγχάνω $R^2=88\%$. Με ακόμα λιγότερες τιμές (10) αλλά στις θέσεις x 10 και 20, επιτυγχάνω $R^2=93\%$! Τα παραδείγματα αυτά (που γραφικά απεικονίζονται στο σχήμα 14.2) δείχνουν ότι μπορεί να επιτευχθεί υψηλό R^2 μόνο και μόνο επειδή τα δεδομένα x ανήκουν σε ένα μεγάλο εύρος. Αντίστροφα, χαμηλά R^2 μπορούμε να έχουμε όταν οι μετρήσεις της ανεξάρτητης μεταβλητής γίνονται σε ένα μικρό σχετικά εύρος. Το χαμηλό R^2 στη δεύτερη περίπτωση δεν σημαίνει απαραίτητα ότι το μοντέλο δεν είναι επαρκές.



Σχήμα 14.2. Γραφήματα που δείχνουν την εξάρτηση του R^2 (για το ίδιο γραμμικό μοντέλο) από το εύρος της ανεξάρτητης μεταβλητής x .

Με βάση τα παραπάνω μπορούμε να πούμε ότι ένα χαμηλό σχετικά R^2 μπορεί να είναι στατιστικά σημαντικό. Αντίστροφα, μπορούμε να έχουμε υψηλή σχετικά τιμή του R^2 αλλά να μην υπάρχει στατιστική σημαντικότητα. Αυτό εξαρτάται από το μέγεθος του δείγματος αλλά και από την εξίσωση με την οποία περιγράφουμε τα δεδομένα. Ποιά είναι η ικανότητα της εξίσωσης όμως να μπορεί να προβλέψει τιμές της εξαρτημένης μεταβλητής. Η «ικανότητα πρόβλεψης» ενός μοντέλου δεν εξαρτάται από το συντελεστή R^2 . Εξαρτάται από το λόγο $F = MS_{\text{regr}} / MS_{\text{res}}$, όπου το MS_{regr} είναι ο λόγος SS_{regr} προς τους αντίστοιχους βαθμούς ελευθερίας $(n-1)$ και MS_{res} είναι ο λόγος SS_{res} προς τους αντίστοιχους βαθμούς ελευθερίας $(n-p)$. Και οι δύο αυτές τιμές εμφανίζονται στον πίνακα της ανάλυσης διασποράς (ANOVA table) που εξάγεται κατά τη διαδικασία γραμμικής παλινδρόμησης σε ένα οποιοδήποτε στατιστικό πακέτο. Ο πρακτικός κανόνας λέει ότι για να έχει δυνατότητες πρόβλεψης ένα μοντέλο πρέπει αυτός ο λόγος F να είναι 4 φορές τουλάχιστον μεγαλύτερος του επιλεγμένου σημείου F που αντιστοιχεί στο επίπεδο εμπιστοσύνης που επιλέγουμε (π.χ. 5%).

Συμπερασματικά μπορούμε να πούμε ότι το R^2 από μόνο του δεν μπορεί να μας υποδείξει αν κάποιο εμπειρικό μοντέλο είναι επαρκές ή όχι. Πρέπει, εκτός του παραπάνω συντελεστή, να μελετάται με χρήση γραφικών η συσχέτιση. Αυτό ίσως είναι το πλέον σημαντικό, αφού τα γραφήματα αποκαλύπτουν πληροφορίες που δεν φαίνονται με απλούς υπολογισμούς. Είναι λάθος λοιπόν να επιδιώκουμε υψηλά R^2 χωρίς να βλέπουμε γραφική συσχέτιση και χωρίς να υπάρχει σχέση αίτιου-αιτιατού.

Μία επιπλέον σημαντική ιδιότητα που κρίνει την επάρκεια ενός μοντέλου είναι η ακρίβεια της εκτίμησης (δηλ. της εξαρτημένης μεταβλητής). Η ακρίβεια της εξαρτημένης μεταβλητής κρίνεται από το *διάστημα εμπιστοσύνης της εξαρτημένης μεταβλητής*. Το διάστημα εμπιστοσύνης αναμένεται να περιέχει την πραγματική τιμή της εξαρτημένης μεταβλητής – βάσει κάποιας βεβαιότητας σε % που καθορίζουμε – και σε καθορισμένες τιμές των ανεξάρτητων μεταβλητών. Το διάστημα εμπιστοσύνης αυτό διαφοροποιείται από το *διάστημα πρόβλεψης της εξαρτημένης μεταβλητής*. Το *διάστημα πρόβλεψης* είναι το εύρος των τιμών εντός του οποίου αναμένεται – πάλι με κάποια βεβαιότητα σε % - μία μελλοντική τιμή της εξαρτημένης μεταβλητής για ένα συγκεκριμένο σετ τιμών των ανεξάρτητων μεταβλητών. Τέλος, υπάρχουν και τα *διαστήματα εμπιστοσύνης των παραμέτρων* (συντελεστών b) του μοντέλου, που έχουμε δει ήδη πως υπολογίζονται. Τα διαστήματα αυτά είναι το εύρος τιμών εντός των οποίων αναμένεται να υπάρχει η πραγματική τιμή των συντελεστών b του μοντέλου, βάσει πάλι κάποιου επιπέδου εμπιστοσύνης (ή βεβαιότητας), που εκφράζεται σε %.

Συμπληρωματικά με τα παραπάνω, το πλέον σημαντικό τεστ ελέγχου εμπειρικών μοντέλων είναι ο έλεγχος των απαραίτητων συνθηκών που πρέπει να πληρούν τα σφάλματα (δηλ. οι διαφορές πρόβλεψης του μοντέλου από τα πραγματικά δεδομένα). Τα σφάλματα πρέπει να κατανέμονται κανονικά, να είναι ανεξάρτητα, να μην παρατηρείται κάποια τάση σε σχέση με τις πραγματικές ή προβλεπόμενες τιμές ή σε σχέση με το χρόνο και να έχουν μέση τιμή ίση με 0 και σταθερή διασπορά.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Berthouex, P.M., L.C.Brown (2002), “Statistics for Environmental Engineers”, 2nd ed., CRC Press LLC, Boca Raton, FL, USA.

Box, G, W.Hunter, J.Hunter (1978), “Statistics for Experimenters: An introduction to design, data analysis and model building”, J. Wiley & Sons Inc., NJ, USA.

Cornell, J.A., (1990) “Experiments with mixtures: designs, models, and the analysis of mixture data”, 2nd ed., John Wiley and Sons.

Draper, N, H. Smith (1998), “Applied Regression Analysis”, 3rd ed., J. Wiley & Sons Inc., NJ, USA.

Johnson, R., (1994) “Miller and Friends’s Probability and Statistics for Engineers”, Prentice-Hall Inc., NJ, USA.

Komilis, D.P., R.K.Ham. (2006) “Carbon dioxide and ammonia emissions during composting of mixed paper, yard waste and food waste” *Waste Management*, 26, 62-70.

McCuen, R. H. (1985) “Statistical Methods for Engineers”, Prentice-Hall Inc., NJ, USA

Minitab, Release 14.1, Statistical Software by Minitab Inc.

Montgomery, D., G. Runger, N. Hubele (1998) “Engineering Statistics”, 1st ed., John Wiley & Sons, NJ, USA

Montgomery, D. (2005), “Design and Analysis of Experiments”, 6th ed., J. Wiley & Sons Inc., NJ, USA.

ΠΑΡΑΡΤΗΜΑ ΠΙΝΑΚΕΣ ΚΑΤΑΝΟΜΩΝ

Κανονική κατανομή (z)

Table of Normal Distribution Function

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998									
4.0	0.99997									
5.0	0.9999997									
6.0	0.999999999									

Κατανομή t (ν = βαθμοί ελευθερίας = $n - 1$)

ν	Tail Area Probability						
	0.4	0.25	0.1	0.05	0.025	0.01	0.005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576

Κατανομή χ^2 (ν = βαθμοί ελευθερίας)

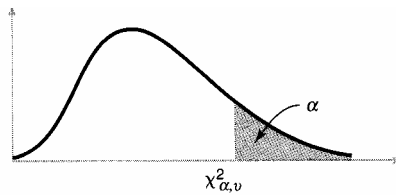


Table III Percentage Points $\chi^2_{\alpha, \nu}$ of the Chi-Square Distribution

ν \ α	.995	.990	.975	.950	.900	.500	.100	.050	.025	.010	.005
1	.00+	.00+	.00+	.00+	.02	.45	2.71	3.84	5.02	6.63	7.88
2	.01	.02	.05	.10	.21	1.39	4.61	5.99	7.38	9.21	10.60
3	.07	.11	.22	.35	.58	2.37	6.25	7.81	9.35	11.34	12.84
4	.21	.30	.48	.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	.41	.55	.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	.68	.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	17.34	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.34	34.28	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.65
28	12.46	13.57	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	69.33	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	79.33	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	89.33	107.57	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	99.33	118.50	124.34	129.56	135.81	140.17

Κατανομή F ($\alpha=5\%$)

ν_2/ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

