

**ΣΗΜΕΙΩΣΕΙΣ ΓΙΑ ΤΟ ΜΑΘΗΜΑ
ΣΤΑΤΙΣΤΙΚΗ ΓΙΑ ΠΟΛΙΤΙΚΟΥΣ ΜΗΧΑΝΙΚΟΥΣ**

ΜΕΡΟΣ Β

Δημήτρης Κουγιουμτζής

<http://users.auth.gr/dkugiu/Teach/CivilEngineer>

E-mail: dkugiu@gen.auth.gr

1/11/2009

Περιεχόμενα

1	ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ	7
1.1	Περιγραφή Στατιστικών Δεδομένων	7
1.2	Περιγραφικά Μέτρα Στατιστικών Δεδομένων	13
1.2.1	Μέτρα θέσης	13
1.2.2	Μέτρα μεταβλητότητας	14
2	ΕΚΤΙΜΗΣΗ ΠΑΡΑΜΕΤΡΩΝ	23
2.1	Σημειακή Εκτίμηση	23
2.1.1	Κριτήρια καλών εκτιμητριών	24
2.1.2	Μέθοδος υπολογισμού της σημειακής εκτίμησης	26
2.2	Εκτίμηση Διαστήματος Εμπιστοσύνης	29
2.2.1	Διάστημα εμπιστοσύνης της μέσης τιμής μ	29
2.2.2	Διάστημα εμπιστοσύνης της διασποράς σ^2	36
2.2.3	Διάστημα εμπιστοσύνης της αναλογίας p	38
2.2.4	Διάστημα εμπιστοσύνης της διαφοράς δύο μέσων τιμών $\mu_1 - \mu_2$	40
2.2.5	Διάστημα της διαφοράς δύο αναλογιών $p_1 - p_2$	44
3	ΕΛΕΓΧΟΣ ΣΤΑΤΙΣΤΙΚΩΝ ΥΠΟΘΕΣΕΩΝ	47
3.1	Διαδικασία Ελέγχου Στατιστικής Υπόθεσης	47
3.1.1	Στατιστική υπόθεση	47
3.1.2	Στατιστική ελέγχου και περιοχή απόρριψης	48
3.1.3	Απόφαση ελέγχου	49
3.2	Έλεγχος Παραμέτρων Πληθυσμού	50
3.2.1	Έλεγχος για τη μέση τιμή μ	50
3.2.2	Έλεγχος για τη διασπορά σ^2	52
3.2.3	Έλεγχος για την αναλογία p	54
3.2.4	Έλεγχος για τη διαφορά μέσων τιμών $\mu_1 - \mu_2$	55
3.2.5	Έλεγχος για τη διαφορά δύο αναλογιών $p_1 - p_2$	58
3.3	Χαρακτηριστικά Ελέγχου	59
3.3.1	Αποφάσεις ελέγχου και πραγματική κατάσταση	60
3.3.2	Δύναμη ελέγχου	61
3.3.3	Δίπλευρος έλεγχος και διάστημα εμπιστοσύνης	64
4	ΣΥΣΧΕΤΙΣΗ ΚΑΙ ΠΑΛΙΝΔΡΟΜΗΣΗ	67
4.1	Συσχέτιση δύο τ.μ.	67
4.1.1	Ο συντελεστής συσχέτισης ρ	68

4.1.2	Σημειακή εκτίμηση του συντελεστή συσχέτισης	68
4.2	Απλή Γραμμική Παλινδρόμηση	71
4.2.1	Το πρόβλημα της γραμμικής παλινδρόμησης	72
4.2.2	Σημειακή εκτίμηση των παραμέτρων της γραμμικής παλινδρόμησης . . .	74
4.2.3	Σχέση του συντελεστή συσχέτισης και παλινδρόμησης	77

Οι στατιστικές έννοιες και μέθοδοι είναι εργαλεία όχι απλά χρήσιμα αλλά κι απαραίτητα για να καταλάβουμε τον κόσμο γύρω μας. Αν για ένα φαινόμενο που μελετάμε δεν υπάρχει καθόλου αβεβαιότητα ή τυχαιότητα δε χρειάζεται η στατιστική προσέγγιση αφού **καθοριστικά μοντέλα** (deterministic models) μπορούν να περιγράψουν το φαινόμενο επακριβώς. Στα πραγματικά φαινόμενα όμως σχεδόν πάντα υπάρχει ο παράγοντας της αβεβαιότητας ή τυχαιότητας κι αυτό κάνει σημαντική τη γνώση των **στατιστικών μεθόδων και μοντέλων** για την επίλυση πραγματικών προβλημάτων. Για παράδειγμα αν θέλουμε να μελετήσουμε την απόσταση από το σπίτι του φοιτητή στο Πανεπιστήμιο, τότε δε μπορούμε να μιλήσουμε με βεβαιότητα γι αυτό το μέγεθος όσους φοιτητές κι αν συμπεριλάβουμε για να συλλέξουμε στοιχεία για την απόσταση. Μπορούμε όμως να υπολογίσουμε τη μέση απόσταση από ένα δείγμα φοιτητών. Ένα άλλο παράδειγμα είναι η σχέση της κατανάλωσης ηλεκτρικής ενέργειας ή νερού με τη διάσταση του χώρου ενός κτιρίου. Η εκτίμηση μιας τέτοιας σχέσης μπορεί να γίνει μόνο με στατιστικές μεθόδους αφού οι μετρήσεις ηλεκτρικής ενέργειας και διάστασης χώρου για κάθε κτίριο επηρεάζονται από πολλούς άλλους παράγοντες κι εμπεριέχουν έτσι τυχαιότητα. Συγκεκριμένα για κλάδους της μηχανικής όπως οι πολιτικοί μηχανικοί, η στατιστική μπορεί να βοηθήσει στην κατανόηση κάποιων φαινομένων και διαδικασιών σχετικών με τη μηχανική που περιέχουν αβεβαιότητα ή μεταβλητότητα, όπως η μελέτη της αντοχής θραύσης σκυροδέματος, ή η επιμήκυνση αλουμινίου για διάφορες τάσεις.

Για να μάθουμε για ένα φαινόμενο συνήθως πρώτα συλλέγουμε δεδομένα. **Στατιστική** (statistics) είναι η επιστήμη ή 'τέχνη' του να μαθαίνουμε από τα δεδομένα. Η στατιστική συνίσταται στη συλλογή δεδομένων, στην περιγραφή τους και κυρίως στην ανάλυση τους που οδηγεί και στην απόκτηση συμπερασμάτων. Η συλλογή των δεδομένων αναφέρεται κι ως **δειγματοληψία** (sampling) κι αποτελεί ξεχωριστό πεδίο της στατιστικής που δεν θα μας απασχολήσει εδώ. Η περιγραφή των δεδομένων και η παρουσίαση συνοπτικών στοιχείων και γραφημάτων αναφέρεται ως **περιγραφική στατιστική** (descriptive statistics). Η ανάλυση στατιστικών δεδομένων και η ερμηνεία των αποτελεσμάτων αναφέρεται ως **στατιστική συμπερασματολογία** (statistical inference) και αποτελεί την βάση της στατιστικής και γι αυτό συχνά αναφέρεται απλά και ως **στατιστική**.

Παραθέτουμε κάποιους βασικούς ορισμούς που χρησιμοποιούνται στην στατιστική ανάλυση :

τυχαία μεταβλητή (τ.μ.) (random variable): οποιοδήποτε χαρακτηριστικό του οποίου η τιμή αλλάζει στα διάφορα στοιχεία του πληθυσμού, π.χ. η αντοχή θραύσης σκυροδέματος.

δεδομένα (data): ένα σύνολο τιμών μιας τ.μ. που έχουμε στη διάθεση μας, π.χ. μετρήσεις αντοχής θραύσης σκυροδέματος.

πληθυσμός (population): μια ομάδα ή μια κατηγορία στην οποία αναφέρεται η τ.μ., π.χ. σκυρόδεμα γενικά, ή ένας τύπος σκυροδέματος.

δείγμα (sample): ένα υποσύνολο του πληθυσμού που μελετάμε, π.χ. 25 δοκίμια σκυροδέματος που χρησιμοποιήσαμε για να κάνουμε μετρήσεις αντοχής θραύσης.

παράμετρος (parameter): ένα μέγεθος που συνοψίζει με κάποιο τρόπο τις τιμές της τ.μ. στον πληθυσμό, π.χ. η μέση τιμή αντοχής θραύσης σκυροδέματος.

στατιστική (statistic): ένα μέγεθος που συνοψίζει με κάποιο τρόπο τις τιμές της τ.μ. στο δείγμα, π.χ. ο μέσος όρος αντοχής θραύσης από τα 25 δοκίμια σκυροδέματος που μετρήσαμε.

Στο Μέρος Α ασχοληθήκαμε με την τυχαία μεταβλητή (τ.μ.) X , μελετήσαμε τη συνάρτηση πυκνότητας ή μάζας πιθανότητας $f_X(x)$ και τη συνάρτηση κατανομής $F_X(x)$ καθώς και κύριες παραμέτρους της κατανομής, όπως τη μέση (ή προσδοκώμενη) τιμή $E(X) \equiv \mu$ και τη διασπορά (ή διακύμανση) $\text{Var}(X) \equiv \sigma^2$.

Η μελέτη μιας τ.μ. με τη βοήθεια της πιθανοθεωρίας προϋποθέτει ότι γνωρίζουμε (ή υποθέτουμε) την κατανομή της τ.μ., κι άρα και τον πληθυσμό στον οποίο παίρνει τιμές, καθώς και τις παραμέτρους της. Στην πράξη, οι παράμετροι πληθυσμού είναι *σταθερές* αλλά *άγνωστες* ενώ οι στατιστικές δείγματος είναι *γνωστές* αλλά *μεταβλητές*. Θα ασχοληθούμε μόνο με τις πιο γνωστές από την πιθανοθεωρία παραμέτρους κατανομής μιας τ.μ., δηλαδή τη μέση τιμή μ , τη διασπορά σ^2 και την αναλογία p .

Γενικά δε γνωρίζουμε τον πληθυσμό αλλά συλλέγουμε δεδομένα για την τ.μ. που μας ενδιαφέρει από ένα αντιπροσωπευτικό δείγμα του πληθυσμού. Πρώτα παρουσιάζουμε τα δεδομένα μέσα από πίνακες και διαγράμματα και συνοψίζουμε τα δεδομένα υπολογίζοντας συνοπτικά μέτρα (αυτό είναι το θέμα της *περιγραφικής στατιστικής*). Στη συνέχεια προσπαθούμε να βγάλουμε συμπεράσματα για την τ.μ. και τις παραμέτρους του πληθυσμού που μελετάμε από τα δεδομένα του δείγματος και τις στατιστικές που υπολογίζουμε από αυτά τα δεδομένα (αυτό είναι το θέμα της *στατιστικής συμπερασματολογίας*).

Για παράδειγμα, για τη μελέτη της αντοχής θραύσης κάποιου τύπου σκυροδέματος και συγκεκριμένα για τη μέση τιμή της δεν είναι δυνατόν να γνωρίζουμε την αντοχή θραύσης για κάθε δοκίμιο που παράγουμε αλλά μπορούμε να μετρήσουμε την αντοχή θραύσης σε κάποια δοκίμια που επιλέγουμε τυχαία. Πρώτα παρουσιάζουμε τις μετρήσεις αντοχής θραύσης που συλλέξαμε χρησιμοποιώντας τις τεχνικές της περιγραφικής στατιστικής για να σχηματίσουμε μια πρώτη εντύπωση για την κατανομή της αντοχής θραύσης. Μετά χρησιμοποιώντας μεθόδους της στατιστικής συμπερασματολογίας προχωράμε στην εκτίμηση της μέσης τιμής της και στον έλεγχο κάποιας υπόθεσης σχετικά με τιμές που μπορεί να πάρει η μέση τιμή.

Ένα άλλο παράδειγμα είναι η μελέτη της μεταβολής της επιμήκυνσης αλουμινίου με την τάση. Το δείγμα περιλαμβάνει τις μετρήσεις της επιμήκυνσης δοκιμίων αλουμινίου σε διάφορες τάσεις. Με την περιγραφική στατιστική μπορούμε να απεικονίσουμε και να περιγράψουμε την εξάρτηση της επιμήκυνσης του αλουμινίου από την τάση και με μεθόδους της στατιστικής (συμπερασματολογίας) μπορούμε να εκτιμήσουμε ποσοτικά τη σχέση αυτής της εξάρτησης.

Στο Μέρος Β θα ασχοληθούμε με κάποια κύρια θέματα στατιστικής που αναφέραμε παραπάνω. Στο Κεφάλαιο 1 θα παρουσιάσουμε συνοπτικά κύριες τεχνικές της περιγραφικής στατιστικής. Στο Κεφάλαιο 2 θα μελετήσουμε την εκτίμηση κυρίων παραμέτρων κατανομής μιας τ.μ. ενός πληθυσμού και στο Κεφάλαιο 3 θα περιγράψουμε στατιστικούς ελέγχους για συγκεκριμένες τιμές αυτών των παραμέτρων. Στο Κεφάλαιο 4 θα μελετήσουμε τη συσχέτιση δύο τ.μ. καθώς και την εξάρτηση (παλινδρόμηση) μιας τ.μ. από μια άλλη.

Κεφάλαιο 1

ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Σ' αυτό το κεφάλαιο θα δούμε πρώτα τρόπους να παρουσιάσουμε τα δεδομένα με στατιστικούς πίνακες και διαγράμματα και μετά να συνοψίσουμε τα δεδομένα υπολογίζοντας συνοπτικά μέτρα.

1.1 Περιγραφή Στατιστικών Δεδομένων

Οι στατιστικοί πίνακες και γραφικές παραστάσεις αποτελούν χρήσιμα μέσα για να παρουσιάσουμε τα δεδομένα καθαρά, σύντομα και με σαφήνεια. Επίσης μπορούν να αποκαλύψουν σημαντικά χαρακτηριστικά των δεδομένων, όπως το εύρος τους, τη συμμετρικότητα τους ή την ύπαρξη ακραίων τιμών.

Πίνακας συχνοτήτων Τα δεδομένα x_1, x_2, \dots, x_n ενός δείγματος για μια τ.μ. X που παίρνει τιμές σ' ένα σχετικά μικρό σύνολο διακεκριμένων τιμών (κατηγορίες ή αριθμητικές τιμές), a_1, a_2, \dots, a_m , μπορούν εύκολα να παρουσιαστούν σ' ένα **πίνακα συχνοτήτων** (frequency table). Ο πίνακας συχνοτήτων παρουσιάζει για κάθε τιμή a_i της X τη συχνότητα εμφάνισής της f_i , δηλαδή πόσες φορές εμφανίζεται η κάθε διακεκριμένη τιμή στο δείγμα. Εύκολα μπορούμε επίσης να υπολογίσουμε και τη **σχετική συχνότητα** (relative frequency) εμφάνισης ή αλλιώς το **ποσοστό** (percent) p_i που ορίζεται από το λόγο της συχνότητας εμφάνισης f_i μιας τιμής a_i προς το σύνολο των παρατηρήσεων n του δείγματος

$$p_i = \frac{f_i}{n}. \quad (1.1)$$

Όταν οι διακεκριμένες τιμές a_1, a_2, \dots, a_m μπορούν να διαταχθούν σε μια σειρά, δηλαδή όταν ισχύει $a_1 < a_2 < \dots < a_m$, ορίζεται η **αθροιστική συχνότητα** (cumulative frequency) F_i μιας τιμής a_i ως το άθροισμα των συχνοτήτων όλων των τιμών που είναι μικρότερες ή ίσες της a_i ,

$$F_i = \sum_{j=1}^i f_j.$$

Με τον ίδιο τρόπο ορίζεται και η αθροιστική σχετική συχνότητα P_i

$$P_i = \sum_{j=1}^i p_j.$$

Ο πίνακας της σχετικής συχνότητας αντιστοιχεί στην εμπειρική ή δειγματική συνάρτηση μάζας πιθανότητας της διακριτής τ.μ. X , δηλαδή στην εκτίμηση της $f_X(x)$ βασισμένη στο δείγμα. Όμοια ο πίνακας της αθροιστικής σχετικής συχνότητας μας δίνει μια εκτίμηση της αθροιστικής συνάρτησης κατανομής $F_X(x)$ από το δείγμα.

Ραβδόγραμμα Τα δεδομένα του πίνακα συχνοτήτων εύκολα μπορούν να παρασταθούν γραφικά σ' ένα **ραβδόγραμμα** (bar chart), όπου η κάθε ράβδος παρουσιάζει τη συχνότητα (ή τη σχετική συχνότητα ή την αθροιστική συχνότητα, ή ακόμα την αθροιστική σχετική συχνότητα) για κάθε τιμή x_i . Το ραβδόγραμμα σχετικής συχνότητας είναι το γράφημα της εμπειρικής ή δειγματικής συνάρτησης μάζας πιθανότητας (δηλαδή της εκτίμησης της $f_X(x)$ από το δείγμα). Αντίστοιχα το ραβδόγραμμα της αθροιστικής σχετικής συχνότητας απεικονίζει τη δειγματική αθροιστική συνάρτηση κατανομής. Η ίδια πληροφορία μπορεί να δοθεί και με άλλου είδους γραφήματα, όπως μ' ένα **κυκλικό διάγραμμα** ή **διάγραμμα πίτας** (pie chart) όπου το κάθε κομμάτι της επιφάνειας του κύκλου ('πίτα') παρουσιάζει τη συχνότητα της αντίστοιχης τιμής.

Παράδειγμα 1.1 Για τη μελέτη του αριθμού των διακεκριμένων δωματίων ανά διαμέρισμα σε μια περιοχή της πόλης συλλέχτηκαν στοιχεία από 120 διαμερίσματα πολυκατοικιών όπως δείχνει ο Πίνακας 1.1. Η τ.μ. X που μετρήσαμε είναι ο αριθμός δωματίων διαμερίσματος. Τα δεδομένα

1	4	2	2	2	3	4	3	1	1	3	3
1	2	1	2	1	1	2	3	3	5	1	2
2	3	2	1	1	4	3	4	1	1	6	2
1	3	2	1	2	2	3	2	4	3	3	5
1	3	5	3	1	2	2	3	1	2	6	4
1	2	5	4	3	1	2	4	2	1	3	4
2	2	2	3	2	1	3	3	4	2	1	5
2	2	3	3	2	4	6	3	2	3	1	3
2	1	5	1	1	4	4	2	5	4	2	2
4	2	1	2	2	2	3	2	3	2	1	4

Πίνακας 1.1: Αριθμός δωματίων σε δείγμα 120 διαμερισμάτων.

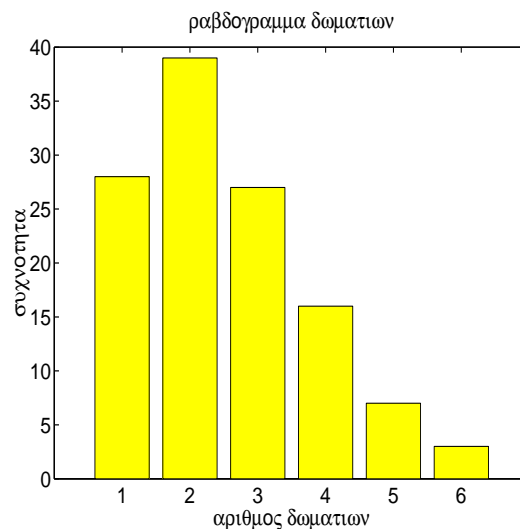
του Πίνακα 1.1 μπορούν να χαρακτηριστούν είτε σαν διακριτά αριθμητικά (ως αριθμοί από το 1 ως το ανώτατο αριθμό δωματίων) ή σαν διατακτικά κατηγορικά (κατηγορίες διαμερισμάτων με βάση τον αριθμό δωματίων, γκαρσονιέρα, δυάρι, κτλ.). Με βάση τον Πίνακα 1.1 δεν είναι εύκολο να μελετήσουμε την συχνότητα εμφάνισης των διαφόρων αριθμών δωματίων. Ποιος αριθμός δωματίων εμφανίζεται συχνότερα; Είναι περισσότερα διαμερίσματα 2 δωματίων ή 3 δωματίων; Για να απαντήσουμε σε τέτοια ερωτήματα μπορούμε να μετρήσουμε πόσες φορές εμφανίζεται ο κάθε αριθμός δωματίων και να φτιάξουμε έτσι τον πίνακα συχνοτήτων. Αυτούς τους απλούς υπολογισμούς μπορούμε εύκολα να τους κάνουμε μόνοι μας αλλιώς όταν το δείγμα είναι μεγάλο θα χρειαστεί να χρησιμοποιήσουμε κάποιο υπολογιστικό πρόγραμμα (όπως το στατιστικό πακέτο SPSS).

Ο Πίνακας 1.2 παρουσιάζει τις τιμές της X (αριθμός δωματίων x_i για $i = 1, \dots, 6$) στην πρώτη στήλη, τη συχνότητα f_i της κάθε τιμής x_i στη δεύτερη στήλη, τη σχετική συχνότητα (ποσοστό) p_i στην τρίτη στήλη, την αθροιστική συχνότητα F_i στην τέταρτη στήλη και τη σχετική αθροιστική συχνότητα P_i στην πέμπτη στήλη.

x_i	f_i	p_i	F_i	P_i
1	28	0.23	28	0.23
2	39	0.33	67	0.56
3	27	0.23	94	0.78
4	16	0.13	110	0.92
5	7	0.06	117	0.97
6	3	0.03	120	1.00
Άθροισμα	120	1.00		

Πίνακας 1.2: Πίνακας συχνοτήτων για τον αριθμό δωματίων 120 διαμερισμάτων του Πίνακα 1.1 που περιλαμβάνει τη συχνότητα f_i , τη σχετική συχνότητα p_i , τη αθροιστική συχνότητα F_i και τη σχετική αθροιστική συχνότητα P_i).

Στο Σχήμα 1.1 παρουσιάζεται το ραβδόγραμμα που προκύπτει από τις συχνότητες εμφάνισης του κάθε αριθμού δωματίων. Από τον πίνακα συχνοτήτων και το ραβδόγραμμα είναι φανερό πως



Σχήμα 1.1: Ραβδόγραμμα του αριθμού δωματίων των 120 διαμερισμάτων του Πίνακα 1.1.

τα περισσότερα διαμερίσματα στο δείγμα μας έχουν 2 δωμάτια, λιγότερα διαμερίσματα έχουν 1 δωμάτιο ή 3 δωμάτια και η συχνότητα φθίνει καθώς αυξάνει ο αριθμός δωματίων, δηλαδή για διαμερίσματα με 4, 5 και 6 δωμάτια.

Ομαδοποίηση Όταν τα δεδομένα είναι αριθμητικά και είτε ο αριθμός των διακεκριμένων τιμών που παίρνει η τ.μ. X είναι μεγάλος, ή η X είναι συνεχής (δηλαδή παίρνει τιμές σ' ένα διάστημα τιμών), είναι προφανές ότι οι πίνακες και τα γραφήματα συχνοτήτων των τιμών δεν προσφέρονται για την απεικόνιση των δεδομένων. Σε τέτοιες περιπτώσεις πρώτα χωρίζουμε τα δεδομένα σε k **ομάδες** (groups), ή κλάσεις διαστημάτων, και μετά παρουσιάζουμε σε πίνακα ή σε γράφημα τη συχνότητα της κάθε ομάδας, δηλαδή τον αριθμό των παρατηρήσεων που ανήκουν σε κάθε ομάδα.

Το εύρος τιμών r_i της κάθε i ομάδας είναι συνήθως το ίδιο ($r_i = r$). Δεν υπάρχει συγκεκριμένος τρόπος να το καθορίσουμε και το διαλέγουμε ανάλογα με την κλίμακα τιμών για

την οποία μας ενδιαφέρει να δούμε διαφορές. Γενικά φροντίζουμε να είναι τέτοιο ώστε να μην προκύπτουν πολλές ομάδες με αποτέλεσμα να έχουμε μικρές συχνότητες γιατί τότε δε μπορούμε να διακρίνουμε κάποιο σχηματισμό στα δεδομένα. Από την άλλη δε θα πρέπει οι ομάδες να είναι πολύ λίγες γιατί τότε δε θα μπορούμε να διακρίνουμε διαφορές παρά μόνο για μεγάλες κλίμακες τιμών.

Για το χωρισμό των δεδομένων σε ομάδες βρίσκουμε πρώτα τη μικρότερη (ελάχιστη) τιμή x_{\min} και μεγαλύτερη (μέγιστη) τιμή x_{\max} και υπολογίζουμε το εύρος των δεδομένων

$$R = x_{\max} - x_{\min}.$$

Διαιρώντας το R με τον αριθμό των ομάδων k που επιλέγουμε έχουμε το εύρος τιμών r της κάθε ομάδας το οποίο συνήθως στρογγυλοποιούμε για να έχουμε εύχρηστα νούμερα. Η πρώτη ομάδα έχει σαν κάτω άκρο του διαστήματος κάποιον κατάλληλα στρογγυλοποιημένο αριθμό μικρότερου ή ίσου του x_{\min} , τα διαστήματα των ομάδων είναι ισομήκη και το διάστημα της τελευταίας ομάδας περιλαμβάνει το x_{\max} . Για τα διαστήματα διαλέγουμε συμβατικά να είναι κλειστά από αριστερά (να περιέχουν την ακραία μικρότερη τιμή) κι ανοιχτά από δεξιά (να μην περιέχουν την ακραία μεγαλύτερη τιμή).

Ιστογράμμα Έχοντας ομαδοποιήσει τα αριθμητικά δεδομένα μπορούμε να κάνουμε τον πίνακα και τα γραφήματα συχνοτήτων όπως και πριν. Ειδικότερα για το ραβδόγραμμα για την κάθε ράβδο παίρνουμε το κέντρο του διαστήματος που αντιστοιχεί σε κάθε ομάδα. Επίσης δεν υπάρχει κενό διάστημα μεταξύ των ράβδων και το γράφημα αυτό λέγεται **ιστόγραμμα** (histogram). Στον κάθετο άξονα του ιστογράμματος μπορεί να είναι η συχνότητα f_i , η σχετική συχνότητα (ποσοστό) p_i , η αθροιστική συχνότητα F_i , ή ακόμα η σχετική αθροιστική συχνότητα P_i για την κάθε i ομάδα. Σε πλήρη αντιστοιχία με το ραβδόγραμμα σχετικής συχνότητας το ιστόγραμμα σχετικής συχνότητας (ή απλά το ιστόγραμμα συχνότητας) δίνει μια εκτίμηση του γραφήματος της συνάρτησης πυκνότητας πιθανότητας $f_X(x)$ της συνεχούς τ.μ. X , βασισμένη στο δείγμα και στην επιλεγμένη ομαδοποίηση των παρατηρήσεων.

Το ιστόγραμμα είναι χρήσιμο για να κρίνουμε αν μπορούμε να δεχθούμε ότι η τ.μ. X , όπως την παρατηρήσαμε από το δείγμα, ακολουθεί κάποια γνωστή κατανομή. Εδώ κυρίως ενδιαφερόμαστε για την **κανονική κατανομή** γιατί τότε μπορούμε να χρησιμοποιήσουμε μια ολόκληρη μεθοδολογία εκτιμητικής που βασίζεται στην κανονική κατανομή της X . Για μικρά δείγματα (με λιγότερες από 30 παρατηρήσεις) δεν περιμένουμε το ιστόγραμμα να δίνει το σχήμα καμπάνας της κανονικής κατανομής (ακόμα και για δεδομένα που προέρχονται από κανονική κατανομή είναι δυνατόν να παρατηρήσουμε σημαντικές αποκλίσεις). Συνήθως δεχόμαστε ότι η τ.μ. X ακολουθεί κανονική κατανομή όταν το ιστόγραμμα φαίνεται να διατηρεί κάποια σχετική συμμετρία με τις υψηλότερες συχνότητες να παρουσιάζονται στα κεντρικά διαστήματα τιμών.

Υπάρχουν κι άλλα γραφήματα που απεικονίζουν την εμπειρική κατανομή της τ.μ. X με διαφορετικό τρόπο. Αναφέρουμε ενδεικτικά χωρίς να τα περιγράψουμε το *φυλλογράφημα* (stem and leaf plot) και το *σημειογράφημα* (dotplot).

Παράδειγμα 1.2 Στον Πίνακα 1.3 δίνονται οι μετρήσεις αντοχής θραύσης σκυροδέματος από 25 σκυροδέματα ενός τύπου A που διαλέχτηκαν τυχαία (και από 20 σκυροδέματα ενός άλλου τύπου B που θα μελετήσουμε αργότερα).

Ονομάζουμε X την τυχαία μεταβλητή της αντοχής θραύσης του σκυροδέματος τύπου A . Η μικρότερη αντοχή θραύσης είναι $x_{\min} = 4.5$ ksi, η μεγαλύτερη αντοχή θραύσης είναι $x_{\max} =$

A/A	τύπος A	τύπος B
1	5.3	5.0
2	4.5	4.2
3	5.7	5.4
4	5.8	5.5
5	4.8	4.6
6	6.4	6.1
7	6.4	6.1
8	5.6	5.3
9	5.8	5.5
10	5.7	5.4
11	5.5	5.2
12	6.1	5.8
13	5.2	4.9
14	7.0	6.7
15	5.5	5.2
16	5.7	5.4
17	6.3	6.0
18	5.6	5.3
19	5.5	5.2
20	5.0	4.8
21	5.8	
22	4.7	
23	6.1	
24	6.7	
25	5.1	
Σύνολο	141.8	107.6

Πίνακας 1.3: Δεδομένα αντοχής θραύσης σκυροδέματος (σε μονάδα μέτρησης ksi) δύο τύπων A και B.

7.0 ksi και το εύρος των δεδομένων του δείγματος είναι

$$R = x_{\max} - x_{\min} = 7.0 - 4.5 = 2.5 \text{ ksi.}$$

Διαλέγουμε να χωρίσουμε τα δεδομένα σε 10 ομάδες ($k = 10$) κι άρα η κάθε ομάδα καλύπτει εύρος τιμών

$$r = \frac{R}{k} = \frac{2.5}{10} = 0.25 \text{ ksi,}$$

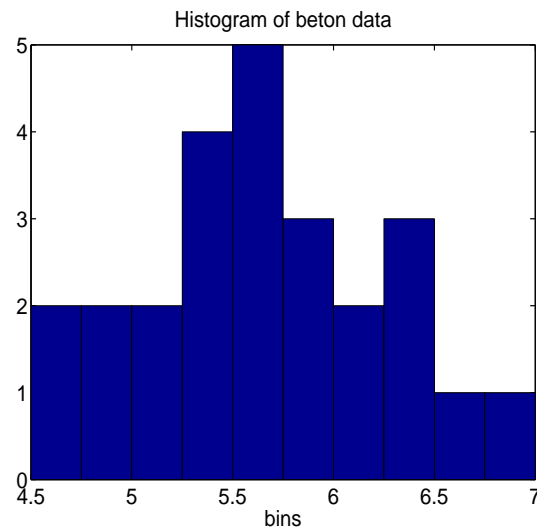
αρχίζοντας από την τιμή 4.5 ksi. Στη συνέχεια μπορούμε να υπολογίσουμε τον πίνακα συχνοτήτων μετρώντας τον αριθμό των δεδομένων σε κάθε ομάδα.

Στον Πίνακα 1.4 δίνεται ο πίνακας συχνοτήτων, που περιλαμβάνει τη συχνότητα, τη σχετική συχνότητα, την αθροιστική συχνότητα και την σχετική αθροιστική συχνότητα. Παρατηρούμε ότι οι τιμές της αντοχής θραύσης του δείγματος συγκεντρώνονται σε ένα κεντρικό διάστημα τιμών (περίπου οι μισές παρατηρήσεις εμφανίζονται στις τρεις ομάδες που καλύπτουν το διάστημα [5.25, 6.0]). Μπορούμε να σχηματίσουμε καλύτερη εντύπωση για την κατανομή της τυχαίας

Διάστημα τιμών	f_i	p_i	F_i	P_i
4.50 - 4.75	2	0.08	2	0.08
4.75 - 5.00	2	0.08	4	0.16
5.00 - 5.25	2	0.08	6	0.24
5.25 - 5.50	4	0.16	10	0.40
5.50 - 5.75	5	0.20	15	0.60
5.75 - 6.00	3	0.12	18	0.72
6.00 - 6.25	2	0.08	20	0.80
6.25 - 6.50	3	0.12	23	0.92
6.50 - 6.75	1	0.04	24	0.96
6.75 - 7.00	1	0.04	25	1.00
Άθροισμα	25	1.00		

Πίνακας 1.4: Πίνακας συχνοτήτων για τα δεδομένα αντοχής θραύσης σκυροδέματος τύπου Α που περιλαμβάνει τη συχνότητα f_i , τη σχετική συχνότητα p_i , την αθροιστική συχνότητα F_i και τη σχετική αθροιστική συχνότητα P_i).

μεταβλητής X της αντοχής θραύσης από το ιστόγραμμα συχνοτήτων, που παρουσιάζεται στο Σχήμα 1.2. Πράγματι από το ιστόγραμμα φαίνεται ότι οι παρατηρήσεις συγκεντρώνονται σ' ένα



Σχήμα 1.2: Ιστόγραμμα των δεδομένων αντοχής θραύσης σκυροδέματος τύπου Α του Πίνακα 1.3.

κεντρικό διάστημα τιμών κι απλώνονται με κάποια συμμετρία γύρω από αυτό. Η εικόνα που δίνει το ιστόγραμμα είναι συνεπής με την γραφική παράσταση κανονικής κατανομής (σχήμα καμπάνας) και μας επιτρέπει να δεχτούμε ότι η κατανομή της αντοχής θραύσης του σκυροδέματος τύπου Α είναι κανονική.

1.2 Περιγραφικά Μέτρα Στατιστικών Δεδομένων

Ο πίνακας συχνοτήτων και το ραβδόγραμμα ή ιστόγραμμα δίνουν μια συνοπτική παρουσίαση των δεδομένων και μας επιτρέπουν να μελετήσουμε ποιοτικά την κατανομή της τυχαίας μεταβλητής που παρατηρήσαμε. Στη συνέχεια θα ορίσουμε ποσοτικά μεγέθη που περιγράφουν περιληπτικά τα βασικά χαρακτηριστικά της κατανομής της τ.μ. X και λέγονται **συνοπτικά ή περιγραφικά μέτρα** (summarizing or descriptive statistics). Κάθε τέτοιο μέτρο υπολογίζεται από τις παρατηρήσεις του δείγματος κι όπως θα δούμε στα επόμενα κεφάλαια αποτελεί εκτίμηση κάποιας παραμέτρου της κατανομής της τ.μ. που μελετάμε.

Θα ασχοληθούμε με δύο τύπους περιγραφικών μέτρων:

- τα **μέτρα θέσης** (measures of location) που προσδιορίζουν χαρακτηριστικές θέσεις μέσα στο εύρος των δεδομένων και
- τα **μέτρα μεταβλητότητας** (variability measures) που δίνουν περιληπτικά τη διασκόπηση και μεταβλητότητα των δεδομένων.

1.2.1 Μέτρα θέσης

Ως μέτρα θέσης εννοούμε κυρίως τα μέτρα κεντρικής τάσης που προσδιορίζουν ένα κεντρικό σημείο γύρω από το οποίο τείνουν να συγκεντρώνονται τα δεδομένα. Τα κυριότερα μέτρα κεντρικής τάσης είναι:

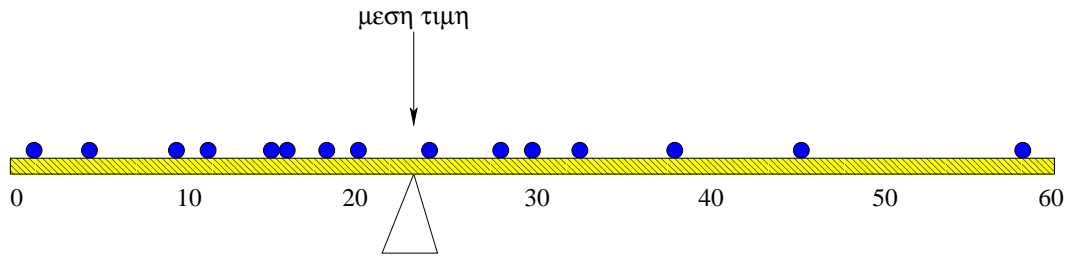
- η **δειγματική μέση τιμή** (sample mean value) ή **αριθμητικός μέσος** (arithmetic mean), ή **μέσος όρος** (average),
- η **δειγματική διάμεσος** (sample median),
- η **δειγματική επικρατούσα τιμή** (sample mode).

Μέση τιμή Η δειγματική μέση τιμή είναι το πιο γνωστό και χρήσιμο μέτρο του κέντρου των δεδομένων. Έστω x_1, x_2, \dots, x_n , οι τιμές των παρατηρήσεων του δείγματος για μια τ.μ. X που μελετάμε. Η δειγματική μέση τιμή συμβολίζεται \bar{x} κι ορίζεται ως

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (1.2)$$

Η μέση τιμή είναι το 'κέντρο ισορροπίας' των δεδομένων. Για να καταλάβουμε τη φυσική της σημασία φανταστούμε μία σανίδα πάνω στην οποία σκορπίζουμε ένα αριθμό n ίδιων βαριδιών. Το σημείο στήριξης της σανίδας (ώστε να ισορροπεί σε οριζόντια θέση) είναι η μέση τιμή της θέσης των βαριδιών πάνω στη σανίδα, όπως φαίνεται και στο Σχήμα 1.3.

Διάμεσος Η δειγματική διάμεσος είναι ένα άλλο μέτρο του κέντρου των δεδομένων και ορίζεται ως η κεντρική τιμή όταν διατάξουμε τα δεδομένα σε αύξουσα σειρά. Θα τη συμβολίζουμε ως \tilde{x} . Αν ο αριθμός n των δεδομένων είναι περιττός τότε η διάμεσος είναι η τιμή στη θέση



Σχήμα 1.3: Σχηματική παρουσίαση της μέσης τιμής.

$(n + 1)/2$, ενώ αν το n είναι άρτιος τότε είναι το ημίαθροισμα των τιμών στις θέσεις $n/2$ και $n/2 + 1$, δηλαδή

$$\tilde{x} = \begin{cases} x_{(n+1)/2} & n = 2k + 1 \\ \frac{x_{n/2} + x_{n/2+1}}{2} & n = 2k. \end{cases} \quad (1.3)$$

Για παράδειγμα σε δείγμα τριών τιμών η διάμεσος είναι η δεύτερη μικρότερη τιμή και σε δείγμα τεσσάρων τιμών η διάμεσος είναι ο μέσος όρος της δεύτερης και τρίτης μικρότερης τιμής.

Επικρατούσα τιμή Η δειγματική επικρατούσα τιμή χρησιμοποιείται επίσης για να δηλώσει την κεντρική τάση των δεδομένων κι ορίζεται ως η τιμή που εμφανίζεται με τη μεγαλύτερη συχνότητα. Αν υπάρχουν πάνω από μία τέτοιες τιμές, τότε όλες αυτές θεωρούνται επικρατούσες τιμές. Είναι φανερό πως η επικρατούσα τιμή δεν έχει νόημα όταν το δείγμα δεν αποτελείται από διακεκριμένες επαναλαμβανόμενες τιμές.

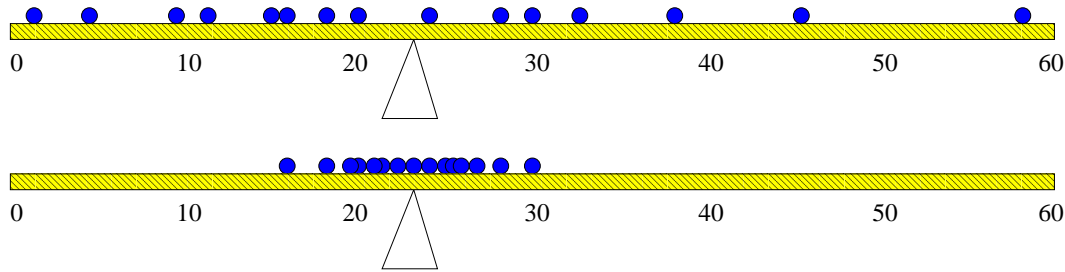
Η δειγματική μέση τιμή είναι το πιο σημαντικό από τα τρία μέτρα κεντρικής τάσης και θα μας απασχολήσει ιδιαίτερα καθώς θα τη χρησιμοποιήσουμε στη στατιστική συμπερασματολογία (στα επόμενα κεφάλαια) για να βγάλουμε συμπεράσματα για τη μέση τιμή μ του πληθυσμού. Για τον υπολογισμό της μέσης τιμής χρησιμοποιούνται όλες οι τιμές του δείγματος, ενώ για τη διάμεσο μόνο η τάξη τους. Γι αυτό και η μέση τιμή επηρεάζεται από μακρινές τιμές αλλά η διάμεσος όχι. Όταν η κατανομή των αριθμητικών δεδομένων είναι μονοκόρυφη και συμμετρική, τότε και τα τρία μέτρα κεντρικής τάσης συμπίπτουν.

Η ύπαρξη μακρινών παρατηρήσεων στο δείγμα δυσκολεύει τη στατιστική περιγραφή κι ανάλυση. Γι αυτό πριν προχωρήσουμε θα πρέπει να αποφασίσουμε αν θα συμπεριλάβουμε τη μακρινή παρατήρηση (αν πιστεύουμε ότι είναι σωστή) ή αν θα την αγνοήσουμε (αν έχουμε λόγους να πιστεύουμε ότι δεν είναι ακριβής).

1.2.2 Μέτρα μεταβλητότητας

Εκτός από την κεντρική τάση μας ενδιαφέρει επίσης και η μεταβλητότητα ή διασπορά των παρατηρήσεων. Όταν τα δεδομένα είναι συγκεντρωμένα γύρω από μια κεντρική τιμή, δηλαδή η διασπορά των δεδομένων είναι μικρή, τότε η κεντρική τιμή αντιπροσωπεύει ικανοποιητικά τα δεδομένα. Από την άλλη, όταν τα δεδομένα είναι πολύ σκορπισμένα τα μέτρα κεντρικής τιμής δε δίνουν καλή περιληπτική περιγραφή των δεδομένων. Επίσης, διαφορετικά δείγματα από τον ίδιο πληθυσμό μπορεί να έχουν το ίδιο μέτρο κεντρικής τάσης αλλά να διαφέρουν κατά κάποιο σημαντικό τρόπο ως προς τη διασπορά των παρατηρήσεων. Χρησιμοποιώντας το

παράδειγμα με τα βαρίδια και τη σανίδα βλέπουμε στο Σχήμα 1.4 πώς δύο δείγματα που έχουν την ίδια μέση τιμή μπορεί να διαφέρουν σημαντικά και κατά χαρακτηριστικό τρόπο ως προς τη διασπορά τους.



Σχήμα 1.4: Σχηματική παρουσίαση δύο δειγμάτων ίσου πλήθους με ίδια μέση τιμή και διαφορετική μεταβλητότητα.

Τα κυριότερα μέτρα διασποράς είναι:

- το **δειγματικό εύρος** (sample range) R ,
- η **δειγματική διακύμανση** ή **δειγματική διασπορά** (sample variance) s^2 και η **δειγματική τυπική απόκλιση** (standard deviation) s .
- τα **εκατοστιαία σημεία** (percentiles) και το **ενδοτεταρτομοριακό εύρος** (interquartile range).

Εύρος Όπως αναφέρθηκε παραπάνω το εύρος των δεδομένων $R = x_{\max} - x_{\min}$ είναι η διαφορά της ελάχιστης από τη μέγιστη τιμή του δείγματος. Το εύρος υπολογίζεται εύκολα αλλά δεν είναι ανθεκτικό μέτρο μεταβλητότητας. Εξαρτάται μόνο από τις δύο ακραίες παρατηρήσεις x_{\min} και x_{\max} και αγνοεί τις υπόλοιπες παρατηρήσεις. Γι αυτό μπορεί να αλλάζει σημαντικά από δείγμα σε δείγμα (ίδιου πλήθους κι από τον ίδιο πληθυσμό). Γενικά το εύρος αυξάνει όταν μεγαλώνει το δείγμα καθώς αναμένεται να συμπεριληφθούν πιο ακραίες τιμές.

Διασπορά Η διασπορά ή διακύμανση μετράει τη μεταβλητότητα των παρατηρήσεων γύρω από τη μέση τιμή. Αν ορίσουμε την απόκλιση μιας παρατήρησης x_i από τη μέση τιμή ως $x_i - \bar{x}$, είναι φανερό πως το άθροισμα όλων αυτών των αποκλίσεων είναι 0 γιατί χρησιμοποιώντας τον ορισμό της δειγματικής μέσης τιμής (1.2) έχουμε

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} = n\bar{x} - n\bar{x} = 0.$$

Η δειγματική μέση τιμή \bar{x} έχει οριστεί έτσι ώστε οι θετικές αποκλίσεις για τιμές μεγαλύτερες του \bar{x} να είναι αθροιστικά ίδιες με τις αρνητικές αποκλίσεις για τιμές μικρότερες του \bar{x} . Για να μετρήσουμε λοιπόν τη μεταβλητότητα των παρατηρήσεων γύρω από τη μέση τιμή διαλέγουμε να αθροίσουμε όχι τις ίδιες τις αποκλίσεις αλλά τα τετράγωνα των αποκλίσεων. Επίσης για να πάρουμε ένα μέτρο της μέσης απόκλισης που δεν εξαρτάται από το πλήθος των παρατηρήσεων θα πρέπει να διαιρέσουμε με το πλήθος n των παρατηρήσεων. Όμως για τεχνικούς λόγους που

θα εξηγήσουμε παρακάτω διαιρούμε με $n - 1$ αντί για n και η δειγματική διασπορά s^2 ορίζεται ως

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (1.4)$$

Αναπτύσσοντας τα τετράγωνα της (1.4) έχουμε τον ισοδύναμο τύπο

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right), \quad (1.5)$$

που είναι πιο εύχρηστος και χρησιμοποιείται στους υπολογισμούς.

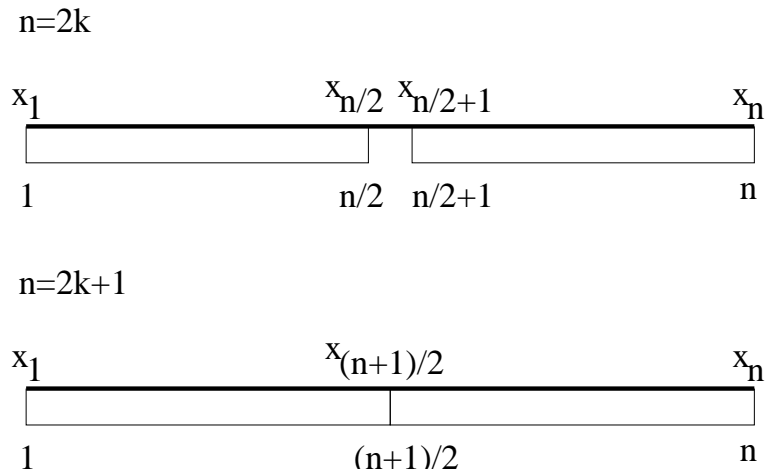
Η διασπορά s^2 προκύπτει από τα τετράγωνα των παρατηρήσεων και συχνά είναι δύσκολο να την ερμηνεύσουμε ως πραγματικό φυσικό μέγεθος. Γι αυτό ορίζουμε τη δειγματική τυπική απόκλιση s , που είναι απλά η θετική ρίζα της δειγματικής διασποράς s^2 . Η τυπική απόκλιση s μετριέται με τη μονάδα μέτρησης της τ.μ. X κι εκφράζει (όπως δηλώνει η ονομασία της) την τυπική απόκλιση των δεδομένων από τη δειγματική μέση τιμή, δηλαδή μέχρι πόσο περίπου περιμένουμε μια τυπική τιμή της X να απέχει από τη μέση τιμή.

Σημείωση: Χρήση του $n - 1$ αντί του n

Όπως η δειγματική μέση τιμή εκτιμά τη μέση τιμή του πληθυσμού μ , έτσι και η δειγματική διασπορά s^2 εκτιμά τη διασπορά του πληθυσμού σ^2 . Αν γνωρίζαμε τη μ τότε θα τη χρησιμοποιούσαμε στον τύπο για τον υπολογισμό του s^2 , αλλά συνήθως η μ είναι άγνωστη. Οι παρατηρήσεις x_i , τείνουν να είναι πιο κοντά στη \bar{x} παρά στη μ κι άρα οι υπολογισμοί με βάση τις αποκλίσεις $x_i - \bar{x}$ δίνουν μικρότερες τιμές απ' ότι αν χρησιμοποιούσαμε τις αποκλίσεις $x_i - \mu$. Για να αντισταθμίσουμε αυτήν την τάση για υποεκτίμηση της διασποράς του πληθυσμού σ^2 διαιρούμε με $n - 1$ αντί με n .

Μία άλλη πιο τεχνική εξήγηση βασίζεται στους *βαθμούς ελευθερίας* (degrees of freedom). Οι n 'ελεύθερες' παρατηρήσεις αποτελούν τους n βαθμούς ελευθερίας. Για τον υπολογισμό της \bar{x} σχηματίζουμε το μέσο όρο διαιρώντας το άθροισμα των παρατηρήσεων με τους βαθμούς ελευθερίας n αφού δεν έχουμε καμιά συνθήκη για τις n παρατηρήσεις που χρησιμοποιούμε. Για τον υπολογισμό της s^2 όμως έχουμε της συνθήκη $\sum_{i=1}^n (x_i - \bar{x}) = 0$, δηλαδή αν ξέρουμε $n - 1$ από τις αποκλίσεις μπορούμε να βρούμε αυτήν που απομένει. Άρα για τον υπολογισμό της s^2 οι βαθμοί ελευθερίας είναι $n - 1$ και γι αυτό διαιρούμε με $n - 1$.

Εκατοστιαία σημεία – ενδοτεταρτομοριακό εύρος – θηκόγραμμα Η διάμεσος χωρίζει τα δεδομένα στα δύο. Μπορούμε να ορίσουμε άλλα σημεία χωρισμού του διατεταγμένου συνόλου τιμών που παίρνουμε από το δείγμα. Τέτοια σημεία είναι τα εκατοστιαία σημεία. Μια παρατήρηση καλείται το *p-εκατοστιαίο σημείο* (*p*-percentile) όταν ποσοστό παρατηρήσεων το πολύ $p\%$ είναι μικρότερες απ' αυτήν την παρατήρηση ($0 \leq p < 1$). Η διάμεσος είναι το 50-εκατοστιαίο σημείο. Αλλά χαρακτηριστικά εκατοστιαία σημεία είναι αυτά που ορίζουν τέταρτα ή *τεταρτομόρια* (quartiles). Το 25-εκατοστιαίο σημείο είναι το **πρώτο** ή **κατώτερο τεταρτομόριο** (first or lower quartile) και το συμβολίζουμε Q_1 , ενώ το 75-εκατοστιαίο σημείο είναι το **τρίτο** ή **ανώτερο τεταρτομόριο** (third or upper quartile) και το συμβολίζουμε Q_3 . Το πρώτο και τρίτο τεταρτομόριο ορίζονται όπως η διάμεσος αλλά περιορίζοντας το σύνολο των δεδομένων στα αντίστοιχα υποσύνολα (κατώτερο ή ανώτερο μισό). Ειδικότερα, έχοντας πρώτα διατάξει τις παρατηρήσεις σε αύξουσα σειρά, αν το σύνολο των παρατηρήσεων n είναι άρτιος αριθμός τότε το κατώτερο υποσύνολο περιέχει τις παρατηρήσεις από 1 ως $n/2$ και το ανώτερο

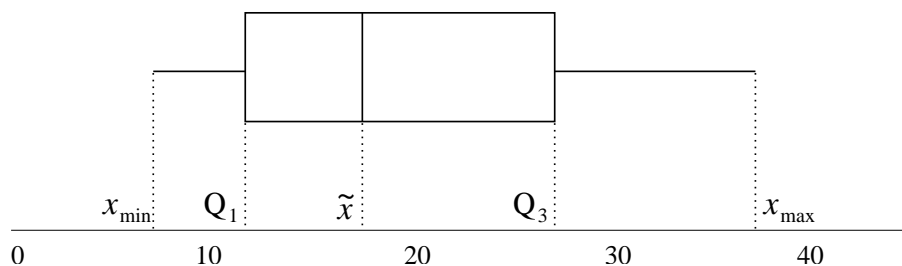


Σχήμα 1.5: Σχηματική παρουσίαση των δύο υποσυνόλων του δείγματος που σχηματίζονται για άρτιο και περιττό πλήθος παρατηρήσεων. Το πρώτο και τρίτο τεταρτομόριο υπολογίζονται ως οι διάμεσοι του πρώτου και δεύτερου υποσύνολου αντίστοιχα.

από $n/2 + 1$ ως n , ενώ αν το n είναι περιττός το κατώτερο υποσύνολο περιέχει τις παρατηρήσεις από 1 ως $(n + 1)/2$ και το ανώτερο από $(n + 1)/2$ ως n (δες Σχήμα 1.5).

Η διαφορά $I = Q_3 - Q_1$ λέγεται **ενδοτεταρτομοριακό εύρος** (interquartile range) και δίνει το εύρος που καλύπτουν τα μισά από τα δεδομένα που είναι πιο κοντά στην κεντρική τιμή (διάμεσο). Το I είναι ένα άλλο μέτρο διασποράς των δεδομένων (πιο ανθεκτικό από το εύρος R) που δεν ορίζεται ως προς τη δειγματική μέση τιμή κι άρα δεν επηρεάζεται απ' αυτήν.

Η διάμεσος \tilde{x} , το πρώτο και τρίτο τεταρτομόριο Q_1 και Q_3 αντίστοιχα, καθώς και η ελάχιστη και μέγιστη τιμή των δεδομένων x_{\min} και x_{\max} , αποτελούν τη **σύνοψη των 5 αριθμών** (five number summary). Γραφικά η παρουσίαση της σύνοψης των 5 αριθμών γίνεται με το **θηκόγραμμα** (box plot) σε οριζόντια ή κάθετη θέση όπως δείχνει το Σχήμα 1.6.



Σχήμα 1.6: Σχηματική παρουσίαση οριζόντιου θηκογράμματος.

Σημείωση: Θηκόγραμμα στον υπολογιστή

Το θηκόγραμμα που παράγει κάποιο στατιστικό πρόγραμμα, όπως το SPSS, μπορεί να διακόπτει τις γραμμές που ενώνουν τα άκρα του κουτιού με την ελάχιστη και μέγιστη τιμή (που λέγονται *μύστακες* (whiskers)) σε κάποια άλλα σημεία και να παρουσιάζει με ειδικά σύμβολα τις υπόλοιπες μακρινές τιμές. Για το σκοπό αυτό χρησιμοποιείται κάποιο κριτήριο για να χαρακτηριστεί μια μακρινή τιμή x_i σαν *ύποπτη ακραία τιμή* (outlier), ή *ακραία τιμή* (extreme). Το κριτήριο είναι η απόσταση της τιμής από τα άκρα του κουτιού (δηλαδή από το Q_1 ή Q_3). Τυπικά αν η απόσταση αυτή είναι μεγαλύτερη από $1.5I$

(όπου τελειώνει ο μύστακας) τότε η τιμή χαρακτηρίζεται *ύποπτη ακραία*, ενώ αν είναι μεγαλύτερη από $3I$ χαρακτηρίζεται *ακραία*.

Θηκόγραμμα και κανονική κατανομή Το θηκόγραμμα, όπως και το ιστόγραμμα, μας επιτρέπει να κρίνουμε αν μπορούμε να δεχτούμε ότι η κατανομή της συνεχούς τυχαίας μεταβλητής που παρατηρήσαμε είναι κανονική. Για να κάνουμε αυτήν την παραδοχή θα πρέπει:

- η διάμεσος να μην αποκλίνει σημαντικά προς το πρώτο ή το τρίτο τεταρτομόριο, δηλαδή η γραμμή που αντιστοιχεί στη διάμεσο να μην πλησιάζει σε κάποιο από τα δύο άκρα του κουτιού (γιατί αλλιώς αυτό θα σήμαινε πως η κατανομή δεν είναι συμμετρική και δείχνει λοξότητα),
- το εύρος των τιμών στα δύο ακραία τεταρτομόρια να μη διαφέρει σημαντικά, δηλαδή τα μήκη των δύο μυστάκων να είναι συγκρίσιμα (για τη διατήρηση της συμμετρίας).
- να μην υπάρχουν ακραίες τιμές, δηλαδή να μην υπάρχουν σημεία μακριά από τους δύο μυστακές (η ύπαρξη ακραίων σημείων δηλώνει πως οι ουρές της κατανομής είναι 'παχιές' που δε συμφωνεί με την κανονική κατανομή).

Είναι σημαντικό να αναλογιστούμε ότι με λίγα δεδομένα δεν είναι δυνατόν να τηρούνται αυστηρά οι παραπάνω προϋποθέσεις για το θηκόγραμμα ακόμα κι αν τα δεδομένα προέρχονται πράγματι από κανονική κατανομή. Αν όμως το θηκόγραμμα (όπως και το ιστόγραμμα, ή οποιοδήποτε άλλο γράφημα της κατανομής των δεδομένων) δίνει ενδείξεις σημαντικής απόκλισης από συμμετρική κατανομή τότε δεν θα πρέπει να θεωρήσουμε πως η κατανομή είναι κανονική στην στατιστική ανάλυση που θα κάνουμε στη συνέχεια.

Παράδειγμα 1.3 *Θέλουμε να εκτιμήσουμε την περιεκτικότητα σε ραδιενέργεια του χάλυβα που παράγεται από ένα εργοστάσιο Α. Γι αυτό έγιναν μετρήσεις της ραδιενέργειας (σε Bq/g) σε 10 δοκίμια από το εργοστάσιο Α. Τα αποτελέσματα δίνονται στον Πίνακα 1.5.*

A/A	εργοστάσιο Α	εργοστάσιο Β
1	0.40	0.11
2	0.51	0.13
3	0.51	0.26
4	0.54	0.27
5	0.55	0.33
6	0.59	0.37
7	0.63	0.52
8	0.67	0.65
9	0.75	
10	2.10	
Σύνολο	7.25	2.64

Πίνακας 1.5: Δεδομένα περιεκτικότητας ραδιενέργειας (σε μονάδα μέτρησης Bq/g) σε δοκίμια από χάλυβα κατασκευασμένα από δύο εργοστάσια Α και Β.

Θα ασχοληθούμε με τις παρατηρήσεις της περιεκτικότητας ραδιενέργειας στο χάλυβα από το εργοστάσιο A κι ας ονομάσουμε αυτήν την τυχαία μεταβλητή X . Η δειγματική μέση τιμή υπολογίζεται από το άθροισμα των 10 παρατηρήσεων που δίνονται στη δεύτερη στήλη του Πίνακα 1.5 ως

$$\bar{x} = \frac{1}{10} \sum_{i=1}^{10} x_i = \frac{7.25}{10} = 0.725.$$

Η δειγματική διάμεσος εύκολα μπορεί να βρεθεί μια κι οι παρατηρήσεις δίνονται σε αύξουσα σειρά. Αφού το n είναι άρτιο η διάμεσος δίνεται ως

$$\tilde{x} = \frac{x_{n/2} + x_{n/2+1}}{2} = \frac{x_5 + x_6}{2} = \frac{0.55 + 0.59}{2} = 0.57.$$

Παρατηρούμε ότι η δειγματική μέση τιμή δίνει μεγαλύτερη τιμή στην εκτίμηση της κεντρικής τάσης των δεδομένων από τη δειγματική διάμεσο. Αυτό συμβαίνει γιατί η μέση τιμή επηρεάζεται από την ακραία τιμή της δεκάτης παρατήρησης που είναι πολύ μεγαλύτερη από όλες τις άλλες. Θα πρέπει να γνωρίζουμε αν αυτή η ακραία τιμή είναι πραγματική ή οφείλεται σε κάποιο σφάλμα της παρατήρησης (σφάλμα του μηχανήματος μέτρησης, σφάλμα στην καταγραφή κτλ).

Η μικρότερη περιεκτικότητα ραδιενέργειας στο δείγμα είναι $x_{\min} = 0.40$ Bq/g κι η μεγαλύτερη είναι $x_{\max} = 2.10$ Bq/g. Άρα το εύρος των δεδομένων είναι $R = 1.70$ Bq/g, που είναι μεγάλο εξαιτίας της ακραίας τιμής που συμπεριλάβαμε στο δείγμα.

Για να βρούμε τη διασπορά s^2 της περιεκτικότητας της ραδιενέργειας στο χάλυβα στο δείγμα μας υπολογίζουμε πρώτα το άθροισμα των τετραγώνων των παρατηρήσεων

$$\sum_{i=1}^{10} x_i^2 = 0.40^2 + 0.51^2 + \dots + 0.75^2 + 2.10^2 = 7.44$$

κι αντικαθιστώντας το στον τύπο της δειγματικής διασποράς (1.5) βρίσκουμε

$$s^2 = \frac{1}{9} \left(\sum_{i=1}^{10} x_i^2 - 10\bar{x}^2 \right) = \frac{1}{9} (7.44 - 10 \cdot 0.725^2) = 0.243.$$

Η δειγματική τυπική απόκλιση είναι

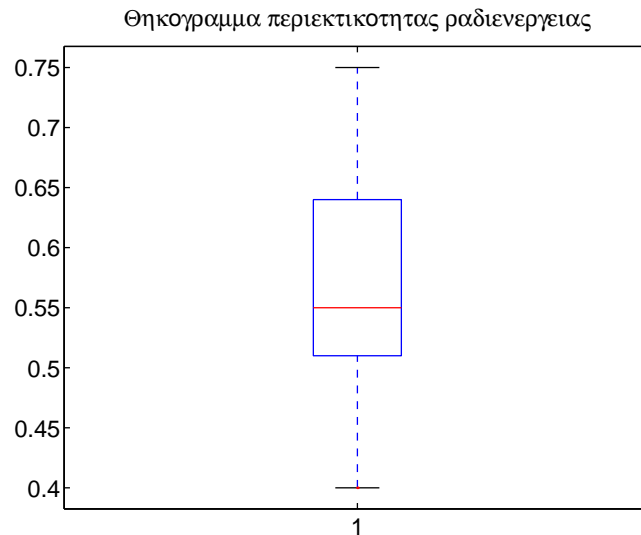
$$s = \sqrt{s^2} = \sqrt{0.243} = 0.493,$$

δηλαδή η αντιπροσωπευτική τυπική απόκλιση από τη μέση περιεκτικότητα ραδιενέργειας είναι περίπου 0.5 Bq/g, που είναι πολύ μεγάλη (όση περίπου και η διάμεσος).

Η σύνοψη των 5 αριθμών δίνεται στο Σχήμα 1.7 όπου παρουσιάζεται σχηματικά και η εύρεση του πρώτου και τρίτου τεταρτομορίου. Από το πρώτο και τρίτο τεταρτομόριο υπολογίζουμε το ενδοτεταρτομοριακό εύρος, $I = Q_3 - Q_1 = 0.67 - 0.51 = 0.16$ Bq/g. Έχοντας βρεί τη σύνοψη των 5 αριθμών μπορούμε εύκολα να παραστήσουμε το θηκόγραμμα. Στο Σχήμα 1.8 δίνεται το θηκόγραμμα σε κατακόρυφη θέση. Παρατηρούμε ότι ο άνω μύστακας δεν προεκτείνεται ως τη μέγιστη τιμή των δεδομένων που είναι 2.10. Αυτή η τιμή δηλώνεται ως ακραία τιμή με ιδιαίτερο σύμβολο, αφού η απόσταση του αντίστοιχου σημείου από το πάνω μέρος του κουτιού, $Q_3 = 0.67$, είναι μεγαλύτερη από $3I = 3 \cdot 0.16 = 0.48$.

Διασπορά	$s^2 = \frac{1}{8} (3.03 - 9 \cdot 0.572^2) = 0.010.$
Τυπική απόκλιση	$s = \sqrt{0.010} = 0.10$
Ελάχιστη τιμή	$x_{\min} = 0.40$
Μέγιστη τιμή	$x_{\max} = 0.75$
Εύρος	$R = 0.75 - 0.40 = 0.35$
Πρώτο τεταρτομόριο	(διάμεσος των $\{x_1, \dots, x_5\}$) $Q_1 = x_3 = 0.51$
Τρίτο τεταρτομόριο	(διάμεσος των $\{x_5, \dots, x_9\}$) $Q_3 = x_7 = 0.63$
Ενδοτεταρτομοριακό εύρος	$I = 0.63 - 0.51 = 0.12$

Το αντίστοιχο θηκόγραμμα δίνεται στο Σχήμα 1.9. Από τα μέτρα θέσης και μεταβλητότητας



Σχήμα 1.9: Θηκόγραμμα των 9 παρατηρήσεων περιεκτικότητας ραδιενέργειας σε χάλυβα του εργοστασίου Α.

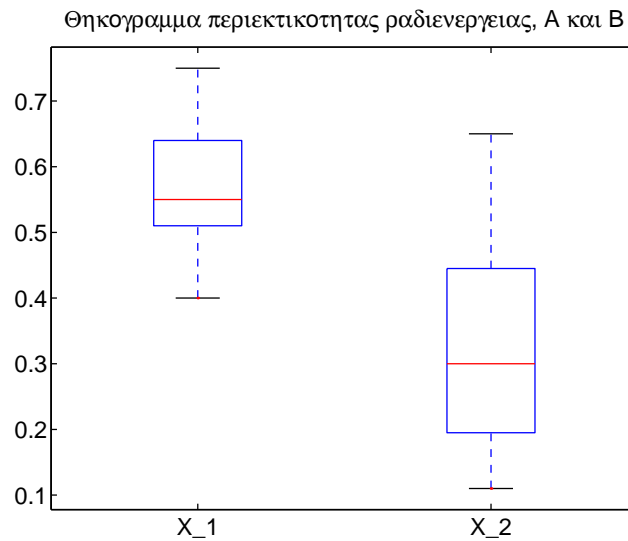
καθώς κι από το θηκόγραμμα παρατηρούμε ότι με την απαλοιφή της ακραίας τιμής η κατανομή της περιεκτικότητας ραδιενέργειας στο χάλυβα του εργοστασίου Α φαίνεται να είναι συμμετρική και μπορούμε να κάνουμε τώρα την παραδοχή ότι η κατανομή είναι κανονική με βάση το δείγμα.

Παράδειγμα 1.5 Στον Πίνακα 1.5 παρουσιάζονται επίσης οι μετρήσεις της περιεκτικότητας ραδιενέργειας σε 8 δοκίμια χάλυβα που κατασκευάστηκαν από ένα άλλο εργοστάσιο Β. Θέλουμε να συγκρίνουμε την περιεκτικότητα ραδιενέργειας στο χάλυβα από τις δύο μονάδες παραγωγής με βάση τα δείγματα των 9 και 8 παρατηρήσεων που έχουμε από το εργοστάσιο Α και από το εργοστάσιο Β αντίστοιχα. Έστω X_1 η τ.μ. της περιεκτικότητας ραδιενέργειας στο χάλυβα για το εργοστάσιο Α και X_2 η αντίστοιχη τ.μ. για το εργοστάσιο Β. Στο προηγούμενο παράδειγμα υπολογίσαμε τα μέτρα θέσης και μεταβλητότητας για τη X_1 . Κάνουμε το ίδιο για τη X_2 . Τα συγκεντρωτικά αποτελέσματα δίνονται στον Πίνακα 1.6. Επίσης στο Σχήμα 1.10 δίνεται το συνδυασμένο θηκόγραμμα για το δείγμα περιεκτικότητας ραδιενέργειας σε χάλυβα κι από τα δύο εργοστάσια.

Από τα μέτρα θέσης (μέση τιμή και διάμεσο) φαίνεται ότι η κεντρική τάση της περιεκτικότητας ραδιενέργειας είναι μεγαλύτερη για το χάλυβα του εργοστασίου Α. Από τα μέτρα μεταβλητότητας (τυπική απόκλιση, εύρος δεδομένων και ενδοτεταρτομοριακό εύρος) φαίνεται πως η περιεκτικότητα ραδιενέργειας μεταβάλλεται λιγότερο στο χάλυβα του εργοστασίου Α.

Μέτρο	X_1	X_2
Μέση τιμή	$\bar{x}_1 = 0.572$	$\bar{x}_2 = 0.33$
Διάμεσος	$\tilde{x}_1 = 0.55$	$\tilde{x}_2 = 0.30$
Διασπορά	$s_1^2 = 0.010$	$s_2^2 = 0.034$
Τυπική απόκλιση	$s_1 = 0.10$	$s_2 = 0.18$
Ελάχιστη τιμή	$x_{1,\min} = 0.40$	$x_{2,\min} = 0.11$
Μέγιστη τιμή	$x_{1,\max} = 0.75$	$x_{1,\max} = 0.65$
Εύρος	$R_1 = 0.35$	$R_2 = 0.54$
Πρώτο τεταρτομόριο	$Q_{1,1} = 0.51$	$Q_{2,1} = 0.195$
Τρίτο τεταρτομόριο	$Q_{1,3} = 0.63$	$Q_{2,3} = 0.445$
Ενδοτεταρτομοριακό εύρος	$I_1 = 0.12$	$I_2 = 0.250$

Πίνακας 1.6: Μέτρα θέσης και μεταβλητότητας για τα δεδομένα περιεκτικότητας ραδιενέργειας σε 9 και 8 δοκίμια χάλυβα κατασκευασμένα από τα εργοστάσια Α και Β στη στήλη 2 και 3 αντίστοιχα.



Σχήμα 1.10: Θηκόγραμμα των 9 παρατηρήσεων περιεκτικότητας ραδιενέργειας σε χάλυβα του εργοστασίου Α και των 8 παρατηρήσεων περιεκτικότητας ραδιενέργειας σε χάλυβα του εργοστασίου Β.

Κεφάλαιο 2

ΕΚΤΙΜΗΣΗ ΠΑΡΑΜΕΤΡΩΝ

Οι στατιστικές δείγματος που υπολογίζονται από τα δεδομένα που έχουν συλλεχθεί, όπως η δειγματική μέση τιμή \bar{x} και η δειγματική διασπορά s^2 , χρησιμοποιούνται για την *εκτίμηση* των σχετικών παραμέτρων πληθυσμού (μ και σ^2 αντίστοιχα).

Γενικά από τις παρατηρήσεις του δείγματος μπορούμε να υπολογίσουμε τη **σημειακή εκτίμηση** (point estimation) και την **εκτίμηση διαστήματος** (interval estimation) της παραμέτρου μιας τ.μ..

2.1 Σημειακή Εκτίμηση

Η *σημειακή εκτίμηση* μιας παραμέτρου είναι η στατιστική που υπολογίζουμε από το δείγμα, δηλαδή είναι μια τιμή, που υπολογίζεται με βάση τα δεδομένα του δείγματος και αντιπροσωπεύει την πραγματική τιμή της σχετικής παράμετρου του πληθυσμού. Για παράδειγμα, ο μέσος όρος των 25 τιμών αντοχής θραύσης που μετρήθηκαν σε δείγμα 25 σκυροδεμάτων αποτελεί μια σημειακή εκτίμηση της μέσης αντοχής θραύσης σκυροδέματος (δες Παράδειγμα 1.2).

Έστω X μια τ.μ. με αθροιστική συνάρτηση κατανομής $F_X(x; \theta)$ ή απλά $F(x; \theta)$ που εξαρτάται από την παράμετρο θ την οποία θέλουμε να εκτιμήσουμε. Έστω ακόμα ότι έχουμε παρατηρήσεις $\{x_1, \dots, x_n\}$ της X από ένα δείγμα μεγέθους n . Τότε η σημειακή εκτίμηση της θ δίνεται από τη συνάρτηση $g(x_1, \dots, x_n)$ των τιμών του δείγματος που λέγεται **εκτιμητήρια συνάρτηση**. Η **εκτιμητήρια** (estimator) της θ από το δείγμα είναι $\hat{\theta} = g(x_1, \dots, x_n)$.

Επειδή οι παρατηρήσεις $\{x_1, \dots, x_n\}$ αλλάζουν κάθε φορά που μελετάμε διαφορετικό δείγμα μεγέθους n , μπορούμε να υποθέσουμε ότι οι παρατηρήσεις $\{x_1, \dots, x_n\}$ είναι τιμές των τ.μ. $\{X_1, \dots, X_n\}$, που είναι ανεξάρτητες μεταξύ τους κι ακολουθούν την ίδια κατανομή $F(x; \theta)$. Άρα η παράμετρος $\hat{\theta}$ είναι συνάρτηση αυτών των τ.μ.. Για ευκολία θα χρησιμοποιούμε το συμβολισμό $\{x_1, \dots, x_n\}$ και θεωρητικά (εννοώντας τις τ.μ. $\{X_1, \dots, X_n\}$) και πρακτικά (εννοώντας τις παρατηρούμενες αριθμητικές τιμές αυτών των τ.μ.).

Είναι φανερό ότι για διαφορετικά δείγματα (διαφορετικές τιμές $\{x_1, \dots, x_n\}$) η εκτιμητήρια συνάρτηση της παραμέτρου $\hat{\theta}$ παίρνει διαφορετικές τιμές, δηλαδή η $\hat{\theta}$ είναι η ίδια τ.μ. με κάποια κατανομή κι έχει μέση τιμή $\mu_{\hat{\theta}} = E(\hat{\theta})$ και διασπορά $\sigma_{\hat{\theta}}^2 = \text{Var}(\hat{\theta})$.

Δύο σημαντικές παράμετροι μιας τ.μ. X που θέλουμε να εκτιμήσουμε είναι η μέση τιμή μ κι η διασπορά σ^2 .

Εκτίμηση μέσης τιμής Είναι φυσικό ως εκτιμήτρια της μ να ορίσουμε τη *δειγματική μέση τιμή* που ορίσαμε στην Παράγραφο 1.2.1

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (2.1)$$

Εκτίμηση διασποράς Όπως για τη μέση τιμή έτσι και για τη διασπορά σ^2 η εκτιμήτρια είναι η *δειγματική διασπορά* που ορίσαμε στην Παράγραφο 1.2.2 ως

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (2.2)$$

Βέβαια μπορεί κάποιος να ορίσει την εκτιμήτρια της σ^2 ως

$$\tilde{s}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (2.3)$$

Οι εκτιμήτριες s^2 και \tilde{s}^2 διαφέρουν μόνο ως προς το συντελεστή του αθροίσματος ($\frac{1}{n-1}$ και $\frac{1}{n}$ αντίστοιχα). Για μεγάλο n οι δύο εκτιμήτριες συγκλίνουν στην ίδια τιμή. Αναπτύσσοντας τα τετράγωνα της (2.2) έχουμε τον ισοδύναμο τύπο (όμοια για την (2.3), δες επίσης την (1.5) στην Παράγραφο 1.2.2)

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right), \quad (2.4)$$

που συνήθως χρησιμοποιούμε στους υπολογισμούς.

2.1.1 Κριτήρια καλών εκτιμητριών

Παραπάνω ορίσαμε κάπως αυθαίρετα εκτιμήτριες της μέσης τιμής μ και της διασποράς σ^2 χωρίς να γνωρίζουμε αν είναι 'καλές' εκτιμήτριες ή όχι. Γενικά όταν ορίζουμε μια εκτιμήτρια $\hat{\theta}$ κάποιας παραμέτρου θ θέλουμε να ελέγξουμε αν είναι κατάλληλη και γι αυτό θέτουμε κάποια κριτήρια ή ιδιότητες που πρέπει να πληρεί μια 'καλή' εκτιμήτρια. Παρακάτω περιγράφονται ορισμένες επιθυμητές ιδιότητες μιας εκτιμήτριας $\hat{\theta}$.

Αμεροληψία Η $\hat{\theta}$ είναι αμερόληπτη (unbiased) αν η μέση τιμή της είναι ίση με την παράμετρο θ , δηλαδή αν ισχύει

$$E(\hat{\theta}) = \theta.$$

Αλλιώς λέγεται μεροληπτική με μεροληψία

$$b(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

Παράδειγμα 2.1 Η *δειγματική μέση τιμή* \bar{x} είναι αμερόληπτη εκτιμήτρια της μέσης τιμής μ μιας τ.μ. X ενός πληθυσμού.

Έχουμε $\theta = \mu$ και $\hat{\theta} = \bar{x}$ και θέλουμε να δείξουμε ότι $E(\bar{x}) = \mu$. Αυτό προκύπτει ως

$$E(\bar{x}) = E\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n} \sum_{i=1}^n E(x_i) = \frac{1}{n} \sum_{i=1}^n \mu = \mu.$$

Παράδειγμα 2.2 Για την εκτίμηση της διασποράς σ^2 μιας τ.μ. X ενός πληθυσμού μπορεί ναδειχθεί ότι:

1. Η δειγματική διασπορά s^2 είναι αμερόληπτη εκτιμήτρια της σ^2 , δηλαδή ισχύει

$$E(s^2) = \sigma^2.$$

2. Η δειγματική διασπορά \tilde{s}^2 που ορίστηκε στη (2.3) είναι μεροληπτική εκτιμήτρια της σ^2 με μεροληψία

$$b(\tilde{s}^2) = -\frac{\sigma^2}{n}.$$

Συνέπεια Η ιδιότητα αυτή ορίζει πως όσο αυξάνει το μέγεθος του δείγματος τόσο μεγαλώνει η πιθανότητα η εκτίμηση να είναι 'κοντά' στην πραγματική τιμή της παραμέτρου, όπου το 'κοντά' σημαίνει ότι η διαφορά της εκτιμώμενης από την πραγματική τιμή της παραμέτρου είναι μικρότερη από κάποια αυθαίρετα μικρή απόσταση ϵ . Δηλαδή η $\hat{\theta}$ είναι συνεπής (consistent) αν ισχύει

$$P(|\theta - \hat{\theta}| \leq \epsilon) \rightarrow 1 \quad \text{όταν} \quad n \rightarrow \infty,$$

όπου ϵ είναι αυθαίρετα μικρός θετικός αριθμός.

Παράδειγμα 2.3 Η εκτιμήτρια \bar{x} της μέσης τιμής μ μιας τ.μ. X είναι συνεπής. Αν όμως αντί της δειγματικής μέσης τιμής διαλέξουμε σαν εκτιμήτρια της μ τον αριθμητικό μέσο της μικρότερης και μεγαλύτερης τιμής του δείγματος

$$x_d = \frac{x_{\min} + x_{\max}}{2},$$

τότε μπορεί ναδειχθεί ότι η εκτιμήτρια x_d δεν είναι συνεπής εκτιμήτρια της μ .

Αποτελεσματικότητα Η αποτελεσματικότητα αναφέρεται στη διασπορά της εκτιμήτριας και δίνεται συγκριτικά. Μια εκτιμήτρια $\hat{\theta}_1$ της θ είναι πιο αποτελεσματική (effective) από μια άλλη εκτιμήτρια $\hat{\theta}_2$ αν έχει μικρότερη διασπορά, $\sigma_{\hat{\theta}_1}^2 < \sigma_{\hat{\theta}_2}^2$.

Παράδειγμα 2.4 Η δειγματική μέση τιμή \bar{x} και η x_d είναι δύο εκτιμήτριες της μέσης τιμής μ κι έχουν διασπορές $\sigma_{\bar{x}}^2$ και $\sigma_{x_d}^2$ αντίστοιχα. Μπορεί ναδειχθεί ότι $\sigma_{\bar{x}}^2 < \sigma_{x_d}^2$ κι άρα η εκτιμήτρια \bar{x} είναι πιο αποτελεσματική από τη x_d .

Επάρκεια Μια εκτιμήτρια της παραμέτρου θ είναι επαρκής (adequate) όταν χρησιμοποιεί όλη την πληροφορία από το δείγμα που σχετίζεται με τη θ .

Παράδειγμα 2.5 Η δειγματική μέση τιμή \bar{x} , εκτιμήτρια της μέσης τιμής μ μιας τ.μ. X , είναι επαρκής γιατί χρησιμοποιεί όλες τις παρατηρήσεις που μετρήθηκαν στο δείγμα, ενώ η x_d δεν είναι επαρκής γιατί χρησιμοποιεί μόνο δύο τιμές των παρατηρήσεων του δείγματος (x_{\min} και x_{\max}).

Παρατηρήσεις

Από τα παραπάνω παραδείγματα βλέπουμε πως οι εκτιμήτριες, \bar{x} για την παραμέτρο μ και s^2 για την παράμετρο σ^2 , που ορίσαμε αυθαίρετα, πληρούν όλες τις τέσσερις ιδιότητες και είναι καλές εκτιμήτριες.

Στον ορισμό των εκτιμητριών \bar{x} και s^2 δεν κάναμε κάποια υπόθεση για την κατανομή της τ.μ. X κι άρα μπορούμε να τις χρησιμοποιήσουμε για οποιαδήποτε τ.μ. X που παρατηρούμε.

Στη συνέχεια θα υποθέσουμε πως η κατανομή της X είναι γνωστή (ως προς τη γενική μορφή της) αλλά δεν είναι γνωστή κάποια παράμετρο θ της κατανομής και θα δούμε πως μπορούμε γενικά να υπολογίσουμε την εκτιμήτρια της θ .

2.1.2 Μέθοδος υπολογισμού της σημειακής εκτίμησης

Υποθέτουμε ότι η τ.μ. X έχει κάποια γνωστή κατανομή, δηλαδή γνωρίζουμε τη γενική μορφή της αθροιστικής συνάρτησης κατανομής $F(x; \theta)$ και της $f(x; \theta)$, που είναι η συνάρτηση πυκνότητας πιθανότητας αν η X είναι συνεχής και η συνάρτηση μάζας πιθανότητας αν η X είναι διακριτή. Η παράμετρος θ της κατανομής είναι άγνωστη και θέλουμε να την εκτιμήσουμε από το δείγμα των παρατηρήσεων $\{x_1, \dots, x_n\}$. Θα θεωρήσουμε επίσης το πρόβλημα να έχουμε περισσότερες από μία άγνωστες παραμέτρους.

Μέθοδος των Ροπών Για συνήθεις κατανομές, μια παράμετρος θ της κατανομής $F(x; \theta)$ σχετίζεται με τις δύο κύριες παραμέτρους μ και σ^2 . Για παράδειγμα, για την κανονική κατανομή οι μ και σ^2 είναι οι μόνες δύο παράμετροι που καθορίζουν πλήρως της συνάρτηση της κατανομής. Για την ομοιόμορφη κατανομή σε διάστημα $[a, b]$, η σχέση των παραμέτρων της κατανομής a και b με τις μ και σ^2 δίνεται ως $\mu = \frac{a+b}{2}$ και $\sigma^2 = \frac{(b-a)^2}{12}$.

Γενικά όταν υπάρχει κάποια σχέση που μας επιτρέπει να υπολογίσουμε την παράμετρο θ (ή τις παραμέτρους θ_1, θ_2) από τις μ και σ^2 , τότε βρίσκουμε την εκτιμήτρια $\hat{\theta}$ (ή τις εκτιμήτριες $\hat{\theta}_1, \hat{\theta}_2$) ως εξής:

1. υπολογίζουμε τις εκτιμήσεις \bar{x} και s^2 των μ και σ^2 αντίστοιχα,
2. αντικαθιστούμε τις εκτιμήσεις \bar{x} και s^2 στην έκφραση της θ (ή των θ_1, θ_2) ως προς μ και σ^2 .

Αυτή είναι η μέθοδος των ροπών (method of moments). Η ονομασία προκύπτει από τη χρήση των ροπών στην εκτίμηση των παραμέτρων: τη μέση τιμή μ που είναι η πρώτη ροπή και τη διασπορά σ^2 που είναι η δεύτερη κεντρική ροπή. Αν οι δύο αυτές ροπές δεν επαρκούν, δηλαδή έχουμε να εκτιμήσουμε περισσότερες από δύο παραμέτρους ή οι σχέσεις δε δίνουν μοναδικότητα λύσης για τις παραμέτρους, χρησιμοποιούμε και ροπές μεγαλύτερου βαθμού, αλλά δε θα ασχοληθούμε με τέτοια προβλήματα.

Παράδειγμα 2.6 Στο Κεφάλαιο 1 μελετήσαμε ένα δείγμα από μετρήσεις της αντοχής θραύσης 25 σκυροδεμάτων κάποιου τύπου A. Τα δεδομένα αυτά (καθώς και τα δεδομένα από 20 σκυροδέματα ενός άλλου τύπου B που θα μελετήσουμε αργότερα) δίνονται στον Πίνακα 1.3. Για την αντοχή θραύσης σκυροδέματος τύπου A είχαμε βρει πως η δειγματική μέση τιμή \bar{x} είναι

$$\bar{x} = \frac{1}{25} \sum_{i=1}^{25} x_i = \frac{1}{25} 141.8 = 5.67$$

και η δειγματική διασπορά s^2 είναι (όπου έχουμε πρώτα υπολογίσει πως $\sum_{i=1}^{25} x_i^2 = 813.3$)

$$s^2 = \frac{1}{24} \left(\sum_{i=1}^{25} x_i^2 - 25\bar{x}^2 \right) = \frac{1}{24} (813.3 - 25 \cdot 5.67^2) = 0.375.$$

Με βάση αυτό το δείγμα η εκτίμηση της μέση τιμής μ είναι 5.67 ksi και της διασποράς σ^2 είναι 0.375 (ksi)².

Αν η τ.μ. X (αντοχή θραύσης σκυροδέματος) ακολουθεί κανονική κατανομή $N(\mu, \sigma^2)$, τότε είναι φανερό πως αυτές οι εκτιμήσεις περιγράφουν πλήρως την κανονική κατανομή της X με βάση αυτό το δείγμα.

Αν η τ.μ. X ακολουθεί ομοιόμορφη κατανομή σε κάποιο διάστημα $[a, b]$, τότε μπορούμε να εκτιμήσουμε τις παραμέτρους a και b από τις σχέσεις $\mu = \frac{a+b}{2}$ και $\sigma^2 = \frac{(b-a)^2}{12}$. Αντικαθιστούμε τις μ και σ^2 με τις εκτιμήσεις \bar{x} και s^2 και λύνουμε το παρακάτω σύστημα εξισώσεων

$$\begin{aligned} \bar{x} &= \frac{a+b}{2} & \Rightarrow & \hat{a} = \bar{x} - \sqrt{3}s & \Rightarrow & \hat{a} = 4.61 \\ s^2 &= \frac{(b-a)^2}{12} & \Rightarrow & \hat{b} = \bar{x} + \sqrt{3}s & \Rightarrow & \hat{b} = 6.73 \end{aligned}$$

Μέθοδος της Μεγίστης Πιθανοφάνειας Η μέθοδος αυτή δίνει την εκτίμηση που έχει τη μέγιστη πιθανοφάνεια, δηλαδή δίνει την τιμή της παραμέτρου η οποία, μεταξύ όλων των δυνατών τιμών της παραμέτρου, είναι η πιο πιθανή με βάση το δείγμα.

Η συνάρτηση πυκνότητας πιθανότητας για κάποια τιμή $X = x_i$ που εξαρτάται από κάποια παράμετρο θ παίρνει την τιμή $f(x_i; \theta)$ (αν η X είναι διακριτή τότε αυτή η τιμή εκφράζει την πιθανότητα $P(X = x_i)$). Επειδή $\{x_1, \dots, x_n\}$ είναι ανεξάρτητες, η πιθανότητα να τις παρατηρήσουμε σ' ένα τυχαίο δείγμα μεγέθους n δίνεται από τη **συνάρτηση πιθανοφάνειας** (likelihood function) ως προς θ

$$L(x_1, \dots, x_n; \theta) = f(x_1; \theta) \cdots f(x_n; \theta).$$

Στα προβλήματα εκτίμησης, θεωρούμε τα $\{x_1, \dots, x_n\}$ δεδομένα και ενδιαφερόμαστε για τη θ . Αν λοιπόν $L(x_1, \dots, x_n; \theta_1) > L(x_1, \dots, x_n; \theta_2)$ για δύο τιμές θ_1 και θ_2 της θ , τότε η τιμή θ_1 είναι πιο αληθοφανής από τη θ_2 γιατί δίνει μεγαλύτερη πιθανότητα να παρατηρήσουμε τα $\{x_1, \dots, x_n\}$. Θέλουμε λοιπόν να βρούμε την 'πιο αληθοφανή' τιμή της θ , δηλαδή την τιμή $\hat{\theta}$ που μεγιστοποιεί τη $L(x_1, \dots, x_n; \theta)$ ή καλύτερα (για ευκολότερους υπολογισμούς) τη $\log L(x_1, \dots, x_n; \theta)$. Άρα η **εκτιμήτρια μεγίστης πιθανοφάνειας** (maximum likelihood estimator) $\hat{\theta}$ βρίσκεται από τη σχέση

$$\frac{\partial \log L(x_1, \dots, x_n; \theta)}{\partial \theta} = 0. \quad (2.5)$$

Αν θέλουμε να εκτιμήσουμε δύο ή περισσότερες παραμέτρους $\theta_1, \dots, \theta_m$, η συνάρτηση πιθανοφάνειας είναι $L(x_1, \dots, x_n; \theta_1, \dots, \theta_m)$ και οι εκτιμήτριες $\hat{\theta}_1, \dots, \hat{\theta}_m$ βρίσκονται λύνοντας το σύστημα των m εξισώσεων

$$\frac{\partial \log L(x_1, \dots, x_n; \theta_1, \dots, \theta_m)}{\partial \theta_j} = 0 \quad \text{για } j = 1, \dots, m. \quad (2.6)$$

Παράδειγμα 2.7 Έχουμε ένα τυχαίο δείγμα $\{x_1, \dots, x_n\}$ από κανονική κατανομή $N(\mu, \sigma^2)$ και θέλουμε να εκτιμήσουμε τη μέση τιμή μ θεωρώντας τη σ^2 γνωστή. Η συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής είναι

$$f_X(x; \mu) \equiv f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Η συνάρτηση πιθανότητας (για την οποία μόνο η παράμετρος μ είναι άγνωστη) είναι

$$L(x_1, \dots, x_n; \mu) = \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right],$$

όπου $\exp(x) \equiv e^x$. Ο λογάριθμος της συνάρτησης πιθανότητας είναι

$$\log L(x_1, \dots, x_n; \mu) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

Η εκτιμήτρια μεγίστης πιθανότητας $\hat{\mu}$ βρίσκεται μηδενίζοντας την παράγωγο της $\log L$ (σχέση (2.5))

$$\frac{\partial \log L}{\partial \mu} = 0 \quad \Rightarrow \quad \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \quad (2.7)$$

που δίνει τη λύση

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x},$$

δηλαδή είναι ίδια με την εκτιμήτρια \bar{x} της μέσης τιμής μ που ορίσαμε για οποιαδήποτε κατανομή της τ.μ. X .

Παράδειγμα 2.8 Άς υποθέσουμε στο προηγούμενο παράδειγμα πως κι η διασπορά σ^2 είναι άγνωστη. Τότε στην παραπάνω εξίσωση (2.7) προστίθεται κι η εξίσωση

$$\frac{\partial \log L}{\partial \sigma^2} = 0 \quad \Rightarrow \quad -\frac{n}{2\sigma^2} + \frac{1}{\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0. \quad (2.8)$$

Η επίλυση του συστήματος των εξισώσεων (2.7) και (2.8) δίνει την ίδια λύση για τη μ και για τη σ^2 είναι

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Οι εκτιμήτριες μεγίστης πιθανότητας λοιπόν για τη μέση τιμή μ και τη διασπορά σ^2 μιας τ.μ. που ακολουθεί κανονική κατανομή είναι απλά η δειγματική μέση τιμή και διασπορά αντίστοιχα, απλά για τη διασπορά έχουμε τη 'μεροληπτική' δειγματική διασπορά \hat{s}^2 (σχέση (2.3)). Ασυμπτωτικά όμως (για μεγάλο n) η εκτιμήτρια $\hat{\sigma}^2 = \hat{s}^2$ είναι αμερόληπτη.

Η μέθοδος μεγίστης πιθανότητας είναι η καλύτερη μέθοδος εκτίμησης αν γνωρίζουμε την κατανομή της τ.μ. X και μπορεί να εφαρμοσθεί σε οποιοδήποτε πρόβλημα εκτίμησης παραμέτρων.

Μέθοδος των Ελαχίστων Τετραγώνων Αυτή η μέθοδος εφαρμόζεται μόνο στην περίπτωση που οι άγνωστες παράμετροι εμφανίζονται σε σχέσεις τυχαίων μεταβλητών και οι σχέσεις αυτές είναι γραμμικές ως προς τις παραμέτρους που θέλουμε να εκτιμήσουμε. Μια απλή περίπτωση είναι να έχουμε μια τ.μ. Y και η κάθε τιμή της y να δίνεται από τη σχέση

$$y = \theta_1 x_1 + \dots + \theta_m x_m + \epsilon,$$

όπου οι τιμές x_1, \dots, x_m είναι γνωστές, $\theta_1, \dots, \theta_m$ είναι οι άγνωστες παράμετροι και ϵ είναι μια άλλη τ.μ. με $E(\epsilon) = 0$. Θα ασχοληθούμε με τη μέθοδο αυτή στο Κεφάλαιο 4.

Παρατηρήσεις

Η μέθοδος μεγίστης πιθανοφάνειας μπορεί να εφαρμοσθεί για οποιοδήποτε θ αν γνωρίζουμε την κατανομή $F_X(x; \theta)$ ενώ η μέθοδος των ροπών δεν εφαρμόζεται αν η θ δε μπορεί να υπολογισθεί από τις ροπές.

Η εκτίμηση μεγίστης πιθανοφάνειας έχει όλες τις ιδιότητες καλής εκτιμήτριας, δηλαδή είναι αμερόληπτη (ασυμπτωτικά, δηλαδή η μεροληψία $b(\theta)$ τείνει στο μηδέν για μεγάλα n), συνεπής, αποτελεσματική κι επαρκής. Γι αυτό κι αυτή η μέθοδος είναι η καλύτερη μέθοδος εκτίμησης παραμέτρων αν γνωρίζουμε την κατανομή $F_X(x; \theta)$.

2.2 Εκτίμηση Διαστήματος Εμπιστοσύνης

Η σημειακή εκτίμηση $\hat{\theta}$ από κάποιο δείγμα δεν περιέχει καμιά πληροφορία για την ακρίβεια της εκτίμησης της θ . Η εκτίμηση $\hat{\theta}$ είναι μια τιμή που δε γνωρίζουμε πόσο κοντά είναι στην πραγματική τιμή της θ κι επίσης η τιμή αυτή αλλάζει με το δείγμα. Για παράδειγμα, υπολογίζουμε τη δειγματική μέση τιμή \bar{x} από ένα τυχαίο δείγμα μεγέθους n . Αν πάρουμε ένα άλλο τυχαίο δείγμα ίδιου μεγέθους, η τιμή της \bar{x} θα είναι διαφορετική. Μπορεί να είναι πιο κοντά ή πιο μακριά στην πραγματική τιμή της μ απ' ότι αυτή από το προηγούμενο δείγμα. Γενικά λοιπόν η εκτιμήτρια $\hat{\theta}$ εξαρτάται από το δείγμα και είναι λοιπόν τ.μ. με κάποια κατανομή. Γι αυτό στην εκτίμηση της θ είναι σημαντικό εκτός από τη σημειακή εκτίμηση $\hat{\theta}$ να υπολογίσουμε και διάστημα $[\theta_1, \theta_2]$ που να μπορούμε να πούμε με μεγάλη εμπιστοσύνη ότι θα περιέχει την πραγματική τιμή της παραμέτρου θ . Στη συνέχεια θα δούμε τέτοια διαστήματα εμπιστοσύνης για διάφορες παραμέτρους, αρχίζοντας από τη μέση τιμή.

2.2.1 Διάστημα εμπιστοσύνης της μέσης τιμής μ

Η σημειακή εκτίμηση της μέσης τιμής μ μιας τ.μ. X είναι η δειγματική μέση τιμή \bar{x} (σχέση (2.1)) που είναι κι αμερόληπτη εκτιμήτρια της μ , δηλαδή η μέση τιμή της \bar{x} είναι η πραγματική τιμή μ που είναι άγνωστη όμως σε μας, $\mu_{\bar{x}} = E(\bar{x}) = \mu$. Παρ' όλο που η εκτιμήτρια \bar{x} είναι διαφορετική από δείγμα σε δείγμα, επειδή η \bar{x} είναι συνεπής εκτιμήτρια, όταν αυξάνεται το μέγεθος n του δείγματος η \bar{x} πλησιάζει τη μέση τιμή μ . Η διασπορά λοιπόν της \bar{x} θα πρέπει να εξαρτάται από το n έτσι ώστε καθώς ο αριθμός των παρατηρήσεων μεγαλώνει η διασπορά να μικραίνει. Πράγματι για τη διασπορά $\sigma_{\bar{x}}^2$ έχουμε

$$\sigma_{\bar{x}}^2 = \text{Var}(\bar{x}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(x_i) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}, \quad (2.9)$$

δηλαδή η διασπορά της εκτιμήτριας \bar{x} είναι ανάλογη της διασποράς σ^2 της X κι αντιστρόφως ανάλογη του αριθμού των παρατηρήσεων n . Στην παραπάνω σχέση υποθέσαμε πως οι παρατηρήσεις x_1, \dots, x_n είναι ανεξάρτητες (εννοώντας και πάλι τις τ.μ. X_1, \dots, X_n). Την τυπική απόκλιση (τετραγωνική ρίζα της διασποράς) $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ της \bar{x} θα την ονομάζουμε **σταθερό σφάλμα** (standard error) της εκτιμήτριας \bar{x} .

Η τ.μ. \bar{x} λοιπόν έχει κάποια κατανομή με μέση τιμή $\mu_{\bar{x}} = \mu$ και διασπορά $\sigma_{\bar{x}}^2 = \sigma^2/n$. Στη συνέχεια, για να ορίσουμε διάστημα εμπιστοσύνης της μ θα βασιστούμε στην κατανομή της \bar{x} . Για να υποθέσουμε κάποια γνωστή κατανομή για τη \bar{x} χρειάζεται να ελέγξουμε το μέγεθος του δείγματος, αν η κατανομή της X είναι κανονική κι αν γνωρίζουμε τη διασπορά της.

Γνωστή διασπορά Θεωρούμε εδώ ότι γνωρίζουμε τη διασπορά σ^2 της τ.μ. X στον πληθυσμό. Για την κατανομή της \bar{x} διακρίνουμε δύο περιπτώσεις:

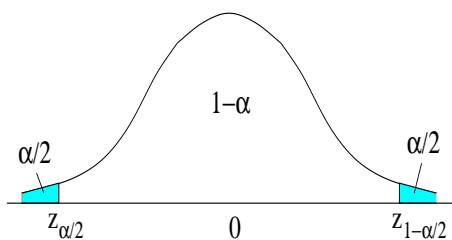
1. Αν η τ.μ. X ακολουθεί κανονική κατανομή $N(\mu, \sigma^2)$ ή το δείγμα είναι μεγάλο ($n > 30$) τότε και η τ.μ. \bar{x} ακολουθεί κανονική κατανομή $N(\mu, \sigma^2/n)$.
2. Αν δε συμβαίνει το παραπάνω, δηλαδή αν η τ.μ. X δεν ακολουθεί κανονική κατανομή **και** το δείγμα είναι μικρό τότε γενικά δε γνωρίζουμε την κατανομή της \bar{x} .

Αν το δείγμα είναι μεγάλο η κανονική κατανομή της \bar{x} δίνεται από το *Κεντρικό Οριακό Θεώρημα*, ενώ αν η τ.μ. X ακολουθεί κανονική κατανομή τότε και το άθροισμα τέτοιων τ.μ. X_1, \dots, X_n ακολουθεί κανονική κατανομή κι έτσι προκύπτει πως και η \bar{x} ακολουθεί κανονική κατανομή.

Υποθέτοντας ότι η \bar{x} ακολουθεί κανονική κατανομή $N(\mu, \sigma^2/n)$, η τ.μ. z που προκύπτει από τον απλό μετασχηματισμό $z \equiv \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ ακολουθεί την τυπική κανονική κατανομή

$$\bar{x} \sim N(\mu, \sigma^2/n) \Rightarrow z \equiv \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Για την τυπική κανονική κατανομή μπορούμε εύκολα να ορίσουμε ένα διάστημα $[z_{\alpha/2}, z_{1-\alpha/2}]$, στο οποίο θα ανήκει η z με κάποια δοθείσα πιθανότητα $1 - \alpha$, όπως φαίνεται στο Σχήμα 2.1. Ο λόγος που χρησιμοποιούμε το δείκτη $\alpha/2$ για την κρίσιμη τιμή $z_{\alpha/2}$ φαίνεται απο το Σχήμα 2.1. Τα άκρα του διαστήματος, $z_{\alpha/2}$ και $z_{1-\alpha/2}$, λέγονται *κρίσιμες τιμές*. Οι δείκτες $\alpha/2$ και $1 - \alpha/2$



Σχήμα 2.1: Συνάρτηση πυκνότητας πιθανότητας της τυπικής κανονικής κατανομής και οι ουρές της για κάποιο α .

δηλώνουν τις τιμές της αθροιστικής συνάρτησης για $z_{\alpha/2}$ και $z_{1-\alpha/2}$ αντίστοιχα, δηλαδή ισχύει

$$\begin{aligned} \Phi(z_{\alpha/2}) &= P(z < z_{\alpha/2}) = \alpha/2 \\ \Phi(z_{1-\alpha/2}) &= P(z < z_{1-\alpha/2}) = 1 - \alpha/2 \end{aligned}$$

όπου $\Phi(z)$ είναι η αθροιστική συνάρτηση της τυπικής κανονικής κατανομής. Άρα η πιθανότητα να είναι $z < z_{\alpha/2}$ ή $z > z_{1-\alpha/2}$ είναι α . Οι δύο σκιασμένες περιοχές στο Σχήμα 2.1 κατέχουν μαζί ποσοστό $\alpha\%$ του συνολικού εμβαδού του ολοκληρώματος της συνάρτησης πυκνότητας

πιθανότητας. Αντίστοιχα η πιθανότητα να συμβαίνει $z \in [z_{\alpha/2}, z_{1-\alpha/2}]$ είναι $1 - \alpha$. Γενικά λοιπόν ισχύει

$$P(z_{\alpha/2} < z \leq z_{1-\alpha/2}) = \Phi(z_{1-\alpha/2}) - \Phi(z_{\alpha/2}) = 1 - \alpha.$$

Επειδή η συνάρτηση πυκνότητας πιθανότητας της τυπικής κανονικής κατανομής είναι συμμετρική ως προς το 0 ισχύει

$$z_{\alpha/2} = -z_{1-\alpha/2}.$$

Άρα στην ουσία για να ορίσουμε το διάστημα $[z_{\alpha/2}, z_{1-\alpha/2}]$ χρειαζόμαστε μία μόνο κρίσιμη τιμή.

Ανακεφαλαιώνοντας, βρήκαμε ότι σε κάθε δοθείσα πιθανότητα $1 - \alpha$ αντιστοιχεί ένα διάστημα τιμών της τ.μ. z που ορίζεται από την κρίσιμη τιμή $z_{1-\alpha/2}$ που την υπολογίζουμε ως $z_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2)$ από τον στατιστικό πίνακα της τυπικής κανονικής κατανομής. Για παράδειγμα για $\alpha = 0.05$ το διάστημα $[-1.96, 1.96]$ περιέχει την τ.μ. z με πιθανότητα 0.95, όπου $z_{0.975} = \Phi^{-1}(0.975) = 1.96$. [Ο συμβολισμός που χρησιμοποιούμε εδώ, όπως και για άλλες κατανομές που θα δούμε παρακάτω, βασίζεται στην αρχή της αντιστοίχισης του δείκτη της κρίσιμης τιμής στην τιμή της αντίστοιχης αθροιστικής συνάρτησης. Σε κάποια βιβλία η θετική κρίσιμη τιμή $z_{1-\alpha/2}$ συμβολίζεται ως $z_{\alpha/2}$ και η αρνητική κρίσιμη τιμή $z_{\alpha/2}$ συμβολίζεται ως $-z_{\alpha/2}$].

Θέλουμε να μετασχηματίσουμε το διάστημα $[z_{\alpha/2}, z_{1-\alpha/2}]$ για πιθανότητα $1 - \alpha$ στο αντίστοιχο διάστημα που περιέχει την παραμέτρο μ . Γι αυτό λύνουμε τις σχέσεις

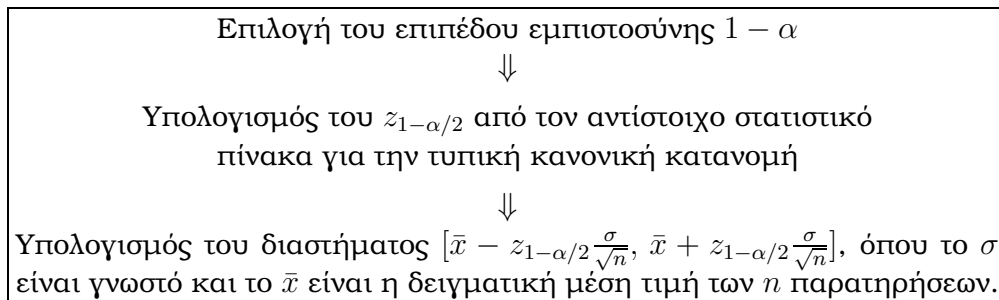
$$z_{\alpha/2} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad z_{1-\alpha/2} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

ως προς μ και βρίσκουμε τα άκρα του διαστήματος για τη μέση τιμή μ

$$\bar{x} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \rightarrow \left[\bar{x} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right]. \quad (2.10)$$

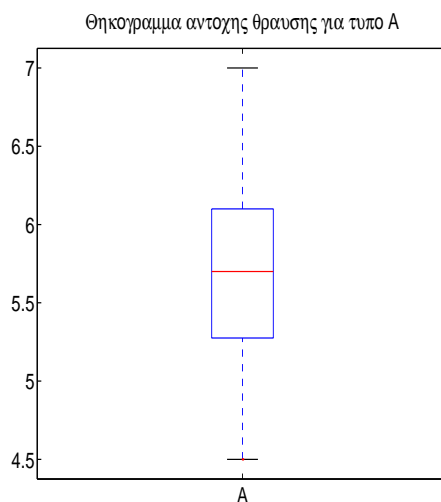
Το διάστημα αυτό υπολογίστηκε για κάποια δοθείσα πιθανότητα $1 - \alpha$ που είναι το προκαθορισμένο **επίπεδο εμπιστοσύνης** (confidence level) και λέγεται **διάστημα εμπιστοσύνης** (confidence interval) της μ σε επίπεδο $1 - \alpha$. Η ερμηνεία αυτού του διαστήματος θέλει προσοχή. Σε μια πρώτη προσέγγιση θα λέγαμε ότι σημαίνει 'με πιθανότητα (εμπιστοσύνη) $1 - \alpha$ η μέση τιμή μ βρίσκεται μέσα σ' αυτό το διάστημα', το οποίο δεν είναι ορθό αφού η μ είναι σταθερό μέγεθος κι όχι τ.μ. για να μιλάμε για την τιμή της με πιθανότητες. Το μέγεθος που αλλάζει (ανάλογα με το δείγμα) είναι το διάστημα και σ' αυτό πρέπει να αναφέρεται η πιθανότητα ή εμπιστοσύνη. Σωστότερη λοιπόν ερμηνεία είναι ότι 'αν χρησιμοποιούσαμε πολλά τέτοια διαστήματα από διαφορετικά δείγματα, ποσοστό $(1 - \alpha)\%$ απ' αυτά θα περιείχαν τη μ ' ή 'με $1 - \alpha$ πιθανότητα (εμπιστοσύνη) το διάστημα αυτό θα περιέχει την πραγματική μ '.

Συνοπτικά η διαδικασία για τον προσδιορισμό του διαστήματος εμπιστοσύνης της μ είναι



Παράδειγμα 2.9 Θέλουμε να εκτιμήσουμε διάστημα εμπιστοσύνης σε επίπεδο 95% για τη μέση αντοχή θραύσης σκυροδέματος τύπου A από τα δεδομένα του Πίνακα 1.3. Υποθέτουμε ότι από παλιότερες μετρήσεις γνωρίζουμε ότι η διασπορά για την αντοχή θραύσης σκυροδέματος είναι $\sigma^2 = 0.38 \text{ (ksi)}^2$. Εξετάζοντας τη δειγματική κατανομή αντοχής θραύσης από τα δεδομένα μας, π.χ. σχεδιάζοντας το ιστόγραμμα των δεδομένων του Πίνακα 1.3 (όπως κάναμε στο Σχήμα 1.2) ή το θηκόγραμμα που παρουσιάζεται στο Σχήμα 2.2, βλέπουμε ότι φαίνεται να είναι κανονική (είναι συμμετρική και δεν έχει μακριές ουρές). Ειδικά για το θηκόγραμμα φαίνεται να τηρούνται τα τρία χαρακτηριστικά που συνιστούν κανονική κατανομή όπως τα θέσαμε στην Παράγραφο 1.2:

- Η διάμεσος δεν τείνει προς το Q_1 ή το Q_3 .
- Οι μύστακες έχουν περίπου το ίδιο μήκος.
- Δεν υπάρχουν ακραίες τιμές ή ύποπτες ακραίες τιμές.



Σχήμα 2.2: Θηκόγραμμα των δεδομένων αντοχής θραύσης σκυροδέματος τύπου A του Πίνακα 1.3.

Άρα μπορούμε να υποθέσουμε ότι η αντοχή θραύσης σκυροδέματος X ακολουθεί κανονική κατανομή $N(\mu, 0.38)$ και τότε η δειγματική μέση τιμή \bar{x} της αντοχής θραύσης σκυροδέματος ακολουθεί επίσης κανονική κατανομή $N(\mu, 0.38/25)$, όπου $n = 25$ είναι το μέγεθος του δείγματος.

Για την εκτίμηση του διαστήματος εμπιστοσύνης ακολουθούμε τα βήματα:

1. Το επίπεδο εμπιστοσύνης είναι $1 - \alpha = 0.95$ ($\alpha = 0.05$).
2. Από τον στατιστικό πίνακα έχουμε $z_{0.975} = \Phi^{-1}(0.975) = 1.96$.
3. Το διάστημα για τη μέση τιμή μ είναι (έχουμε υπολογίσει ότι $\bar{x} = 5.67 \text{ ksi}$, δεξ Παράδειγμα 2.6)

$$5.67 \pm 1.96 \frac{\sqrt{0.38}}{5} \rightarrow [5.43, 5.91].$$

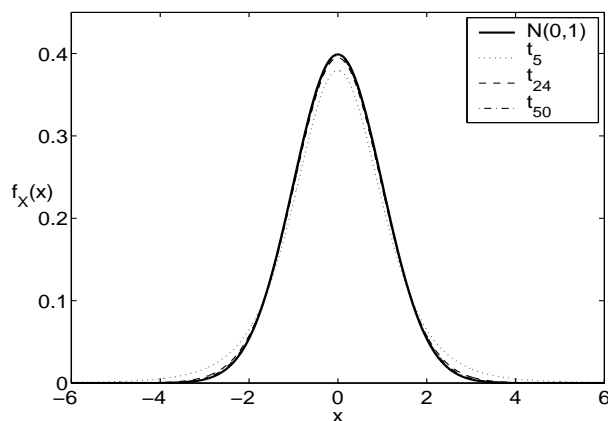
Άρα το 95% διάστημα εμπιστοσύνης της μέσης αντοχής θραύσης σκυροδέματος με βάση αυτό το δείγμα είναι $[5.43, 5.91]$. Μπορούμε να πούμε ότι η σημειακή εκτίμηση $\bar{x} = 5.67$ είναι αρκετά ακριβής αφού το αντίστοιχο 95% διάστημα εμπιστοσύνης είναι αρκετά μικρό.

Άγνωστη διασπορά Γενικά η διασπορά σ^2 της τ.μ. X είναι άγνωστη και την εκτιμούμε από το δείγμα με την s^2 (π.χ. δεξ την (2.4) που χρησιμοποιούμε για τους υπολογισμούς). Αν το δείγμα είναι μεγάλο ($n > 30$) η εκτιμήτρια s^2 είναι αρκετά ακριβής κι απλά μπορούμε να αντικαταστήσουμε τη διασπορά σ^2 με τη δειγματική διασπορά s^2 στην παραπάνω διαδικασία για να πάρουμε το διάστημα εμπιστοσύνης για τη μ . Αν όμως το δείγμα είναι μικρό τότε η προσέγγιση δεν είναι καλή και το διάστημα μπορεί να είναι αρκετά ανακριβές.

Για μικρό n και υποθέτοντας πως η X ακολουθεί κανονική κατανομή, η τ.μ. t που ορίζεται ως $t \equiv \frac{\bar{x} - \mu}{s/\sqrt{n}}$ ακολουθεί την κατανομή student ή την t -κατανομή με $n - 1$ βαθμούς ελευθερίας

$$t \equiv \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}.$$

Η κατανομή αυτή μοιάζει με την τυπική κανονική κατανομή και την προσεγγίζει καθώς αυξάνει ο αριθμός των βαθμών ελευθερίας όπως φαίνεται στο Σχήμα 2.3. Για μεγάλο n η προσέγγιση



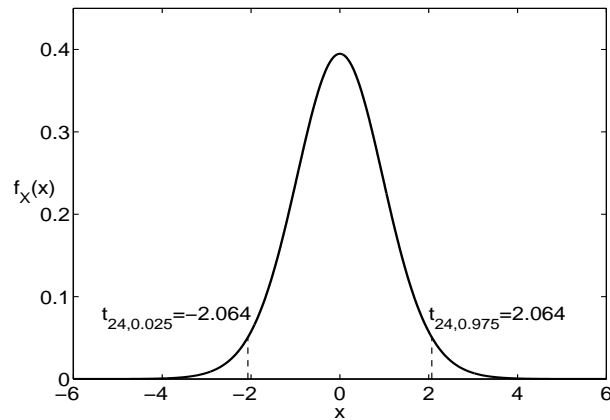
Σχήμα 2.3: Συνάρτηση πυκνότητας πιθανότητας για την τυπική κανονική κατανομή και την κατανομή student με 5, 24 και 50 βαθμούς ελευθερίας, όπως δείχνει το επεξηγημα των γραμμών.

είναι πολύ καλή κι οι τιμές των t και z είναι πρακτικά ίδιες.

Η διαδικασία για την εκτίμηση του διαστήματος εμπιστοσύνης είναι ίδια όπως στην περίπτωση της γνωστής διασποράς, αλλά η άγνωστη διασπορά σ^2 αντικαθίσταται από τη δειγματική διασπορά s^2 που υπολογίζεται από το δείγμα κι η κρίσιμη τιμή είναι $t_{n-1, 1-\alpha/2}$ αντί για $z_{1-\alpha/2}$ και τη βρίσκουμε από το στατιστικό πίνακα για την κατανομή student. [Η κρίσιμη τιμή της t ορίζεται από το $1 - \alpha/2$ αλλά και από τους βαθμούς ελευθερίας $n - 1$. Στο Σχήμα 2.4 παρουσιάζεται η κατανομή student για $n - 1 = 24$ βαθμούς ελευθερίας κι οι κρίσιμες τιμές της για $1 - \alpha = 0.95$.] Το διάστημα εμπιστοσύνης σε επίπεδο $1 - \alpha$ είναι

$$\bar{x} \pm t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}}. \quad (2.11)$$

Παράδειγμα 2.10 Για το προηγούμενο παράδειγμα υποθέτουμε πως η διασπορά είναι άγνωστη (που είναι και η πιό πιθανή περίπτωση για το πραγματικό πρόβλημα). Το δείγμα είναι αρκετά μεγάλο κι ίσως θα μπορούσαμε να υποθέσουμε πως η \bar{x} σαν τ.μ. ακολουθεί κανονική κατανομή



Σχήμα 2.4: Συνάρτηση πυκνότητας πιθανότητας για την κατανομή student με 24 βαθμούς ελευθερίας κι η δεξιά κι αριστερή κρίσιμη τιμή της για $1 - \alpha = 0.95$.

και να χρησιμοποιήσουμε στους υπολογισμούς του διαστήματος εμπιστοσύνης την κρίσιμη τιμή $z_{1-\alpha/2}$. Για να εκτιμήσουμε όμως το διάστημα εμπιστοσύνης για τη μέση αντοχή θραύσης με την καλύτερη δυνατή ακρίβεια χρησιμοποιούμε την κρίσιμη τιμή $t_{n-1,1-\alpha/2}$ από την κατανομή student (δες Σχήμα 2.4). Η δειγματική διασπορά έχει βρεθεί να είναι $s^2 = 0.375 \text{ (ksi)}^2$ (δες Παράδειγμα 2.6). Από το στατιστικό πίνακα για την κατανομή student, για $1 - \alpha/2 = 0.975$ και $n - 1 = 24$, βρίσκουμε $t_{24,0.975} = 2.064$ και το διάστημα εμπιστοσύνης για τη μ είναι

$$5.67 \pm 2.064 \frac{\sqrt{0.375}}{5} \rightarrow [5.42, 5.92].$$

Η χρήση της κρίσιμης τιμής $t_{24,0.975} = 2.064$ υπαγορεύεται από τη χρήση της εκτίμησης s^2 αντί της πραγματικής διασποράς σ^2 που δεν τη γνωρίζουμε. Αν αποφασίζαμε εσφαλμένα να χρησιμοποιήσουμε στους παραπάνω υπολογισμούς τη $z_{0.975} = 1.96$ της κανονικής κατανομής θα βρίσκαμε το διάστημα $[5.52, 5.82]$ που είναι φυσικά πιο μικρό αφού $z_{0.975} < t_{24,0.975}$. Το διάστημα όμως αυτό δεν είναι ακριβές γιατί κάναμε την υπόθεση για κανονική κατανομή της \bar{x} που δεν εφαρμόζεται στην περίπτωση που η διασπορά είναι άγνωστη και το δείγμα είναι μικρό.

Γενικά όταν τον n δεν είναι μεγάλο το διάστημα εμπιστοσύνης της \bar{x} που παίρνουμε υποθέτοντας ότι η \bar{x} ακολουθεί κατανομή student είναι πιο μεγάλο από αυτό που παίρνουμε υποθέτοντας ότι η \bar{x} ακολουθεί κανονική κατανομή και η διασπορά παραμένει η ίδια.

Μη κανονική κατανομή και μικρά δείγματα Σε κάποιες περιπτώσεις το δείγμα μπορεί να είναι μικρό **και** η κατανομή της τ.μ. X που παρατηρούμε να μην είναι κανονική. [Όταν δε ξέρουμε τίποτε για την κατανομή της X αυτό το εκτιμούμε από τα δεδομένα του δείγματος, π.χ. από τη μορφή του ιστογράμματος ή του θηκογράμματος των δεδομένων.] Σε μια τέτοια περίπτωση (κι ανεξάρτητα από το αν η διασπορά είναι γνωστή ή όχι) δε μπορούμε να υποθέσουμε πως η \bar{x} ακολουθεί κάποια συγκεκριμένη κατανομή και στη συνέχεια να εκτιμήσουμε διάστημα εμπιστοσύνης για τη μέση τιμή μ .

Σε τέτοιες περιπτώσεις ενδιαφερόμαστε για τη διάμεσο $\delta = \delta_X$ της τ.μ. X αντί της $\mu = \mu_X$. Η σημειακή εκτίμηση \tilde{x} της δ είναι απλά η κεντρική τιμή των παρατηρήσεων διαταγμένες

σε αύξουσα σειρά (δες Παράγραφο 1.2.1). Το διάστημα εμπιστοσύνης για τη δ βρίσκεται με τη μέθοδο Wilcoxon που βασίζεται στην τάξη των παρατηρήσεων παρά στις τιμές τους. Αυτή η εκτίμηση διαστήματος εμπιστοσύνης λέγεται **μη παραμετρική** (επειδή δεν υποθέτουμε κάποια κατανομή και τις παραμέτρους της για την εκτιμήτρια).

Παρατηρήσεις

Γενικά θα θέλαμε το διάστημα εμπιστοσύνης να είναι όσο το δυνατό μικρότερο για να έχουμε όσο το δυνατόν πιο ακριβή σημειακή εκτίμηση. Το διάστημα εμπιστοσύνης εξαρτάται από:

- την κατανομή και τη διασπορά σ^2 της τ.μ. X [αυτά δε μπορούμε να τα αλλάξουμε, είναι χαρακτηριστικά της τ.μ. που μελετάμε].
- το μέγεθος n του δείγματος [αύξηση του n έχει σαν αποτέλεσμα τη μείωση του εύρους του διαστήματος, που είναι φυσικά επιθυμητό αλλά όχι πάντοτε εφικτό, αφού η απόκτηση πολλών παρατηρήσεων μπορεί να είναι εργασία επίπονη και πολυέξοδη].
- το επίπεδο εμπιστοσύνης $1 - \alpha$ [αυτό το καθορίζουμε εμείς, αλλά δε θα θέλαμε να μικρύνουμε το διάστημα μειώνοντας την εμπιστοσύνη μας σ' αυτό γιατί τότε τα αποτελέσματά μας δε θα έχουν την επιθυμητή στατιστική σημαντικότητα]. Το επίπεδο εμπιστοσύνης που συνήθως χρησιμοποιείται στην πράξη είναι 95%.

Στον Πίνακα 2.1 συνοψίζεται η εκτίμηση του διαστήματος εμπιστοσύνης της μ στις διάφορες περιπτώσεις.

διασπορά	κατανομή της X	n	κατανομή της \bar{x}	διάστημα εμπιστοσύνης
γνωστή	κανονική		$z \equiv \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$	$\bar{x} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$
γνωστή	μη κανονική	μεγάλο	$z \equiv \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$	$\bar{x} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$
γνωστή	μη κανονική	μικρό	—	—
άγνωστη		μεγάλο	$z \equiv \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim N(0, 1)$	$\bar{x} \pm z_{1-\alpha/2} \frac{s}{\sqrt{n}}$
άγνωστη	κανονική	μικρό	$t \equiv \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$	$\bar{x} \pm t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}}$
άγνωστη	μη κανονική	μικρό	—	—

Πίνακας 2.1: Εκτίμηση του διαστήματος εμπιστοσύνης της μ ανάλογα με τη γνώση της διασποράς και κατανομής της τ.μ. X και το μέγεθος n του δείγματος.

Εύρος διαστήματος εμπιστοσύνης Πολλές φορές πριν να κάνουμε το πείραμα και συλλέξουμε τις μετρήσεις προκαθορίζουμε ένα συγκεκριμένο εύρος για το δ.ε. ή ζητάμε το εύρος του δ.ε. να μην ξεπερνάει κάποιο ανώτατο όριο για να έχουν νόημα τα αποτελέσματα. Για να το πετύχουμε αυτό χωρίς να αλλάξουμε τη σημαντικότητα των στατιστικών αποτελεσμάτων, βρίσκουμε το μέγεθος n του δείγματος που μας δίνει αυτό το εύρος του δ.ε. Αυτό υπολογίζεται θέτοντας το εύρος του δ.ε. ίσο με την τιμή που ζητάμε και λύνοντας την εξίσωση ως προς

το n . Για παράδειγμα, ας υποθέσουμε πως το δείγμα είναι μικρό και η τ.μ. X ακολουθεί κανονική κατανομή με άγνωστη διασπορά, χρησιμοποιούμε δηλαδή τη σχέση (2.11) για να υπολογίσουμε το δ.ε. της μέσης τιμής μ . Το εύρος του δ.ε. είναι

$$w = 2 t_{n-1, 1-\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad (2.12)$$

και λύνοντας ως προς n βρίσκουμε ότι για να είναι το εύρος του δ.ε. ίσο με w πρέπει το δείγμα να έχει μέγεθος

$$n = \left(2 t_{n-1, 1-\alpha/2} \cdot \frac{s}{w} \right)^2. \quad (2.13)$$

Στην παραπάνω σχέση η τιμή $t_{n-1, 1-\alpha/2}$ δεν είναι γνωστή αφού το n είναι άγνωστο και ζητούμενο. Θα πρέπει λοιπόν να χρησιμοποιήσουμε την τιμή $t_{n-1, 1-\alpha/2}$ που συμφωνεί καλύτερα με το αποτέλεσμα για το n από την (2.13). Αν το n παίρνει μεγάλες τιμές το παραπάνω πρόβλημα δεν υφίσταται αφού για μεγάλα n η κρίσιμη τιμή $t_{n-1, 1-\alpha/2}$ είναι πρακτικά ίδια με την αντίστοιχη κρίσιμη τιμή της τυπικής κανονικής κατανομής $z_{1-\alpha/2}$. Άρα για μεγάλο n η σχέση (2.13) δίνεται ως

$$n = \left(2 z_{1-\alpha/2} \cdot \frac{s}{w} \right)^2. \quad (2.14)$$

Παράδειγμα 2.11 Στο προηγούμενο παράδειγμα για την αντοχή θραύσης σκυροδέματος υπολογίσαμε το 95% δ.ε. της μέσης αντοχής θραύσης κάνοντας χρήση της κατανομής student και βρήκαμε ότι είναι $[5.42, 5.92]$ ksi. Το εύρος του δ.ε. είναι $w = 5.92 - 5.42 = 0.50$ ksi. Αν θέλουμε να το μειώσουμε στο μισό (δηλαδή σε $w = 0.25$ ksi) τότε αντί για 25 δοκίμια πρέπει να χρησιμοποιήσουμε

$$n = \left(2 \cdot 1.96 \cdot \frac{\sqrt{0.375}}{0.25} \right)^2 = 92.2 \simeq 93,$$

δηλαδή πρέπει να αυξήσουμε το δείγμα σε 93 δοκίμια.

2.2.2 Διάστημα εμπιστοσύνης της διασποράς σ^2

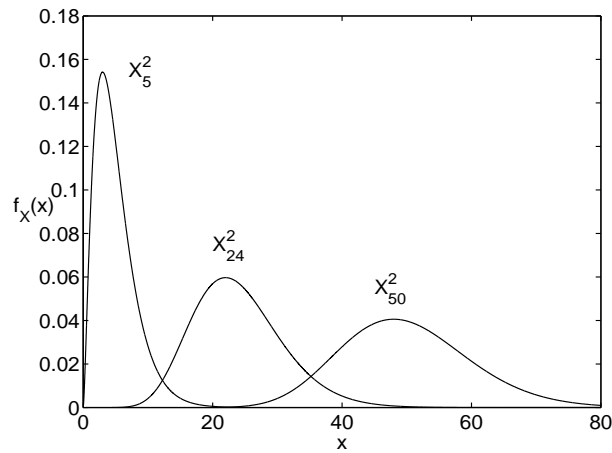
Όπως για να βρούμε διάστημα εμπιστοσύνης για τη μέση τιμή μ ορίσαμε πρώτα την κατανομή της εκτιμήτριας \bar{x} έτσι και για να βρούμε διάστημα εμπιστοσύνης για τη διασπορά σ^2 ορίζουμε πρώτα την κατανομή της αμερόληπτης εκτιμήτριας s^2 της σ^2 (π.χ. από τη σχέση (2.4)). Γνωρίζουμε ότι η $\chi^2 \equiv \frac{(n-1)s^2}{\sigma^2}$ ακολουθεί την κατανομή χ^2 με $n-1$ βαθμούς ελευθερίας

$$\chi^2 \equiv \frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2.$$

Στο Σχήμα 2.5 παρουσιάζεται η μορφή της κατανομής χ^2 για χαρακτηριστικούς βαθμούς ελευθερίας. Για λίγους βαθμούς ελευθερίας η κατανομή χ^2 είναι αρκετά λοξή και γίνεται πιο συμμετρική καθώς αυξάνουν οι βαθμοί ελευθερίας.

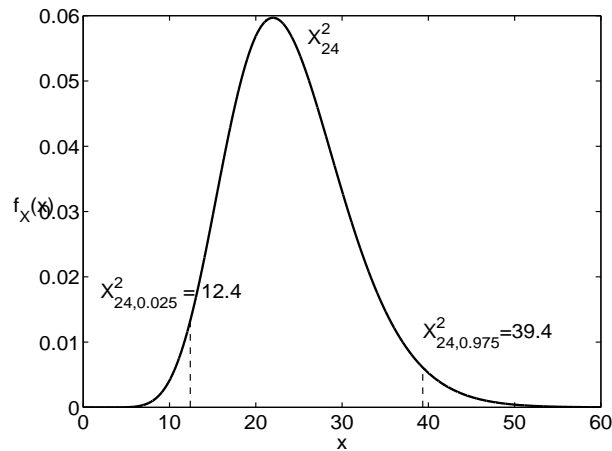
Για δοθείσα πιθανότητα $1 - \alpha$ μπορούμε να βρούμε από τον στατιστικό πίνακα για τη κατανομή χ^2 τις δύο κρίσιμες τιμές $\chi_{n-1, \alpha/2}^2$ και $\chi_{n-1, 1-\alpha/2}^2$ για τις οποίες ισχύει

$$P(\chi_{n-1, \alpha/2}^2 < \chi^2 < \chi_{n-1, 1-\alpha/2}^2) = 1 - \alpha. \quad (2.15)$$



Σχήμα 2.5: Συνάρτηση πυκνότητας πιθανότητας για την κατανομή χ^2 με 5, 24 και 50 βαθμούς ελευθερίας.

Επειδή η χ^2 δεν είναι συμμετρική έχουμε δύο κρίσιμες τιμές: την αριστερή κρίσιμη τιμή $\chi_{n-1,\alpha/2}^2$ που είναι τέτοια ώστε $P(\chi^2 < \chi_{n-1,\alpha/2}^2) = \alpha/2$ και τη δεξιά κρίσιμη τιμή $\chi_{n-1,1-\alpha/2}^2$ που είναι τέτοια ώστε $P(\chi^2 < \chi_{n-1,1-\alpha/2}^2) = 1 - \alpha/2$. [Οι δείκτες $\alpha/2$ και $1 - \alpha/2$ των κρίσιμων τιμών είναι κι οι τιμές της αντίστοιχης αθροιστικής συνάρτησης κατανομής.] Στο Σχήμα 2.6 παρουσιάζεται η κατανομή χ^2 για $n-1 = 24$ βαθμούς ελευθερίας καθώς κι οι κρίσιμες τιμές της για $1 - \alpha = 0.95$.



Σχήμα 2.6: Συνάρτηση πυκνότητας πιθανότητας για την κατανομή χ^2 με 24 βαθμούς ελευθερίας κι η δεξιά κι αριστερή κρίσιμη τιμή της για $1 - \alpha = 0.95$.

Στη σχέση (2.15), λύνοντας τις δύο ανισότητες

$$\chi_{n-1,\alpha/2}^2 < \frac{(n-1)s^2}{\sigma^2} \leq \chi_{n-1,1-\alpha/2}^2$$

ως προς σ^2 βρίσκουμε το $(1 - \alpha)\%$ διάστημα εμπιστοσύνης για τη σ^2

$$\left[\frac{(n-1)s^2}{\chi_{n-1,1-\alpha/2}^2}, \frac{(n-1)s^2}{\chi_{n-1,\alpha/2}^2} \right], \quad (2.16)$$

όπου η δειγματική διασπορά s^2 υπολογίζεται από το δείγμα των n παρατηρήσεων.

Το 95% δ.ε. για την τυπική απόκλιση σ έχει σαν άκρα τις τετραγωνικές ρίζες των αντίστοιχων άκρων του 95% δ.ε. για τη διασπορά σ^2 .

Παράδειγμα 2.12 Από τα δεδομένα για την αντοχή θραύσης σκυροδέματος τύπου A του Πίνακα 1.3 θέλουμε να εκτιμήσουμε τη διασπορά σ^2 της αντοχής θραύσης σκυροδέματος τύπου A. Η σημειακή εκτίμηση βρέθηκε να είναι $s^2 = 0.375$ (ksi)². Για $n - 1 = 24$ και $\alpha = 0.05$ από τον στατιστικό πίνακα για τη χ^2 βρίσκουμε $\chi_{24,0.025}^2 = 12.4$ και $\chi_{24,0.975}^2 = 39.4$ (δες Σχήμα 2.6). Το 95% διάστημα εμπιστοσύνης για τη σ^2 είναι

$$\left[\frac{24 \cdot 0.375}{39.4}, \frac{24 \cdot 0.375}{12.4} \right] = [0.228, 0.726].$$

Το 95% δ.ε. για την τυπική απόκλιση σ της αντοχής θραύσης σκυροδέματος είναι

$$[\sqrt{0.228}, \sqrt{0.726}] = [0.478, 0.852].$$

Στο παραπάνω παράδειγμα παρατηρούμε ότι η εκτίμηση $s^2 = 0.375$ (ksi)² είναι πιά κοντά στο αριστερό άκρο του διαστήματος εμπιστοσύνης. Γενικά το διάστημα εμπιστοσύνης για τη διασπορά σ^2 δεν είναι συμμετρικό ως προς τη σημειακή εκτίμηση s^2 κι αυτό γιατί η κατανομή χ^2 δεν είναι συμμετρική όπως είναι η κανονική κατανομή κι η κατανομή student. Όσο μεγαλώνουν όμως οι βαθμοί ελευθερίας (δηλαδή το μέγεθος δείγματος) η χ^2 κατανομή προσεγγίζει την κανονική κατανομή (δες Σχήμα 2.5). Γι αυτό για πολύ μεγάλα δείγματα το διάστημα εμπιστοσύνης μπορεί να υπολογισθεί κι από άλλο τύπο που περιέχει κρίσιμες τιμές της τυπικής κανονικής κατανομής.

2.2.3 Διάστημα εμπιστοσύνης της αναλογίας p

Σε αρκετά προβλήματα τα δεδομένα δεν είναι αριθμητικές τιμές μιας τ.μ. του πληθυσμού αλλά δυαδικές τιμές, δηλαδή κάποιο στοιχείο του πληθυσμού έχει μια ιδιότητα (επιτυχία ή 1) ή δεν την έχει (αποτυχία ή 0). Ο λόγος των στοιχείων του πληθυσμού που πληρούν την ιδιότητα προς το σύνολο όλων των στοιχείων του πληθυσμού λέγεται **αναλογία** p . [Η αναλογία p είναι η πιθανότητα 'επιτυχίας σε μια δοκιμή' όταν αναφερόμαστε σε ακολουθίες Bernoulli.] Ένα παράδειγμα είναι η αναλογία κοριτσιών στις Πολυτεχνικές Σχολές της Ελλάδας ή η αναλογία σκουριασμένων ράβδων χάλυβα σε μια αποθήκη. Σε πολλές περιπτώσεις θέλουμε να εκτιμήσουμε την αναλογία p από ένα δείγμα μεγέθους n . Η σημειακή εκτίμηση της p είναι απλά $\hat{p} = \frac{m}{n}$, ο λόγος των 'επιτυχίων' m στο δείγμα προς το πλήθος n των στοιχείων του δείγματος.

Γνωρίζουμε ότι για μεγάλο n η κατανομή της εκτιμήτριας \hat{p} είναι κανονική με μέση τιμή $E(\hat{p}) = p$ και διασπορά $\text{Var}(\hat{p}) = \frac{p(1-p)}{n}$, δηλαδή ισχύει

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right).$$

Ακολουθώντας την ίδια διαδικασία όπως για τη μέση τιμή με γνωστή διασπορά, με τη βοήθεια του μετασχηματισμού

$$z \equiv \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$

καταλήγουμε στο διάστημα εμπιστοσύνης για την αναλογία p

$$\hat{p} \pm z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}}.$$

Αντικαθιστώντας στο τυπικό σφάλμα την αναλογία p με την δειγματική αναλογία \hat{p} έχουμε το διάστημα εμπιστοσύνης για την p από το δείγμα

$$\hat{p} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}. \quad (2.17)$$

Το εύρος του δ.ε. της αναλογίας p είναι

$$w = 2 z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}. \quad (2.18)$$

Λύνοντας την παραπάνω σχέση ως προς n βρίσκουμε

$$n = \left(\frac{2 z_{1-\alpha/2}}{w} \right)^2 \hat{p}(1-\hat{p}).$$

Πριν από τη μελέτη του δείγματος δε γνωρίζουμε την εκτίμηση της αναλογίας p . Άρα από την παραπάνω σχέση δε μπορούμε να βρούμε τον αριθμό των παρατηρήσεων που θα χρειαστούμε για να υπολογίσουμε δ.ε. της p εύρους w . Μπορεί να δειχθεί όμως πως η μέγιστη τιμή που μπορεί να πάρει η έκφραση $\hat{p}(1-\hat{p})$ είναι 0.25. Άρα το μέγεθος δείγματος που χρειαζόμαστε για να εξασφαλίσουμε μέγιστο εύρος w του δ.ε. είναι

$$n = \left(\frac{2 z_{1-\alpha/2}}{w} \right)^2 0.25 = \left(\frac{z_{1-\alpha/2}}{w} \right)^2. \quad (2.19)$$

Παράδειγμα 2.13 Για να εκτιμήσουμε την αναλογία σκουριασμένων ραβδών χάλυβα μιας αποθήκης πήραμε ένα δείγμα από $n = 100$ ράβδους και βρήκαμε $m = 12$ σκουριασμένες.

Η σημειακή εκτίμηση για την αναλογία p των σκουριασμένων ραβδών είναι $\hat{p} = \frac{12}{100} = 0.12$. Το 95% διάστημα εμπιστοσύνης είναι ($z_{1-\alpha/2} = z_{0.975} = 1.96$)

$$0.12 \pm 1.96 \sqrt{\frac{0.12 \cdot (1-0.12)}{100}} \rightarrow [0.056, 0.184].$$

Άρα σε 95% επίπεδο εμπιστοσύνης και με βάση το δείγμα μας μπορούμε να πούμε ότι το διάστημα $[0.056, 0.184]$ θα περιέχει την πραγματική αναλογία σκουριασμένων ραβδών χάλυβα, την οποία θα μπορούσαμε να τη βρούμε αν καταμετρούσαμε όλες τις ράβδους στην αποθήκη.

Έστω ότι θέλουμε να εκτιμήσουμε το διάστημα εμπιστοσύνης της αναλογίας p των σκουριασμένων ραβδών στο ίδιο επίπεδο εμπιστοσύνης 95% αλλά με μικρότερο εύρος $w = 0.05$ (ή αν αναφερόμαστε σε ποσοστά το εύρος να είναι 5 εκατοστιαίες μονάδες). Από την σχέση (2.19) βρίσκουμε ότι το μέγιστο μέγεθος n του δείγματος που θα χρειαστούμε είναι (αντικαθιστώντας τις τιμές $z_{1-\alpha/2} = z_{0.975} = 1.96$ και $w = 0.05$)

$$n = \left(\frac{1.96}{0.05} \right)^2 = 1536.6 \simeq 1537,$$

δηλαδή για να μειώσουμε το εύρος του 95% διαστήματος εμπιστοσύνης της αναλογίας σκουριασμένων ραβδών περίπου στο 1/3 πρέπει να αυξήσουμε το μέγεθος του δείγματος πάνω από 15 φορές.

2.2.4 Διάστημα εμπιστοσύνης της διαφοράς δύο μέσων τιμών $\mu_1 - \mu_2$

Θέλουμε να εκτιμήσουμε τη διαφορά των μέσων τιμών μ_1 και μ_2 δύο ανεξάρτητων τ.μ. X_1 και X_2 έχοντας δύο δείγματα μεγέθους n_1 και n_2 από τον πληθυσμό της X_1 και τον πληθυσμό της X_2 αντίστοιχα.

Η σημειακή εκτίμηση της διαφοράς $\mu_1 - \mu_2$ είναι απλά η διαφορά των δειγματικών μέσων τιμών $\bar{x}_1 - \bar{x}_2$. Για το διάστημα εμπιστοσύνης της διαφοράς $\mu_1 - \mu_2$ πρέπει να ελέγξουμε την κατανομή της εκτιμήτριας $\bar{x}_1 - \bar{x}_2$, όπως κάναμε στην εκτίμηση διαστήματος εμπιστοσύνης για τη μέση τιμή μ .

Γνωστές διασπορές Θεωρούμε ότι γνωρίζουμε τις διασπορές σ_1^2 και σ_2^2 των X_1 και X_2 . Υποθέτουμε επίσης ότι το δείγμα είναι αρκετά μεγάλο ή η κατανομή των X_1 και X_2 είναι κανονική. Τότε η εκτιμήτρια $\bar{x}_1 - \bar{x}_2$ ακολουθεί κανονική κατανομή με μέση τιμή $\mu_1 - \mu_2$ και διασπορά $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$. Αν οι X_1 και X_2 είναι ομοσκεδαστικές, δηλαδή $\sigma_1^2 = \sigma_2^2 = \sigma^2$, τότε η διασπορά είναι $\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$.

Η διαδικασία είναι ίδια όπως για την εύρεση του διαστήματος εμπιστοσύνης για τη μέση τιμή μ αν κάνουμε τις παρακάτω αντικαταστάσεις:

$$\begin{array}{lll} \text{εκτιμήτρια} & \bar{x} & \longrightarrow \bar{x}_1 - \bar{x}_2 \\ \text{μέση τιμή εκτιμήτριας} & \mu & \longrightarrow \mu_1 - \mu_2 \\ \text{διασπορά εκτιμήτριας} & \frac{\sigma^2}{n} & \longrightarrow \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}. \end{array}$$

Επομένως το διάστημα εμπιστοσύνης της διαφοράς $\mu_1 - \mu_2$ είναι

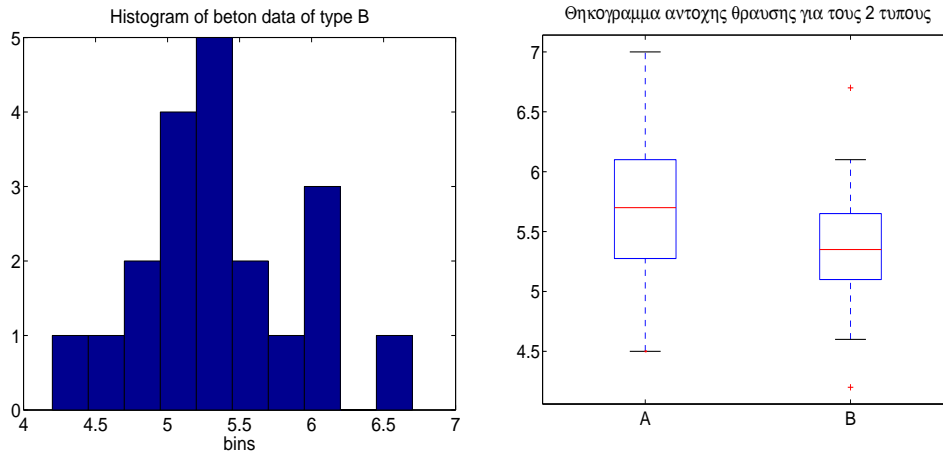
$$(\bar{x}_1 - \bar{x}_2) \pm z_{1-\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}. \quad (2.20)$$

Στην πράξη, όταν εκτιμάμε το διάστημα εμπιστοσύνης της διαφοράς $\mu_1 - \mu_2$ είναι γιατί θέλουμε να διαπιστώσουμε αν κατά μέσο όρο η μια τ.μ. είναι διαφορετική (μεγαλύτερη ή μικρότερη) από την άλλη και αν ναι να εκτιμήσουμε το μέγεθος αυτής της διαφοράς. Το διάστημα εμπιστοσύνης λοιπόν το ερμηνεύουμε ως εξής:

- Αν περιέχει το μηδέν τότε δε μπορούμε να πούμε ότι οι μέσες τιμές των τ.μ. X_1 και X_2 διαφέρουν για το επίπεδο εμπιστοσύνης που χρησιμοποιήσαμε και με βάση τα συγκεκριμένα δεδομένα.
- Αν είναι θετικό τότε μπορούμε να πούμε πως για το επίπεδο εμπιστοσύνης που χρησιμοποιήσαμε η τ.μ. X_1 είναι κατά μέσο όρο μεγαλύτερη από τη X_2 κατά ένα ποσό που κυμαίνεται στα όρια του διαστήματος που εκτιμήσαμε. Ανάλογα ερμηνεύουμε το διάστημα εμπιστοσύνης όταν είναι αρνητικό.

Παράδειγμα 2.14 Ενδιαφερόμαστε να διαπιστώσουμε αν η αντοχή θραύσης σκυροδέματος ενός τύπου Α είναι κατά μέσο όρο διαφορετική από αυτή ενός άλλου τύπου Β. Γι αυτό θέλουμε να εκτιμήσουμε το 95% διάστημα εμπιστοσύνης για τη διαφορά των μέσων τιμών μ_1 και μ_2 της αντοχής θραύσης σκυροδέματος του τύπου Α και Β αντίστοιχα. Δίνεται ότι η διασπορά είναι κοινή και γνωστή και είναι $\sigma^2 = 0.38 \text{ (ksi)}^2$.

Στον Πίνακα 1.3 έχουμε 25 μετρήσεις αντοχής θραύσης σκυροδέματος από τον τύπο A και 20 από τον τύπο B. Στο Παράδειγμα 2.9 είδαμε με τη βοήθεια ιστογράμματος (Σχήμα 1.2) και θηκογράμματος (Σχήμα 2.2) πως η αντοχή θραύσης του σκυροδέματος τύπου A θα μπορούσε να ακολουθεί κανονική κατανομή. Από το ιστόγραμμα και το θηκόγραμμα του Σχήματος 2.7 μπορούμε να πούμε το ίδιο και για τον τύπο B. Άρα μπορούμε να υποθέσουμε πως οι τ.μ. αντοχής



Σχήμα 2.7: Ιστόγραμμα των δεδομένων αντοχής θραύσης σκυροδέματος του τύπου B του Πίνακα 1.3 και θηκογράμματα των δεδομένων αντοχής θραύσης σκυροδέματος των δύο τύπων.

θραύσης X_1 και X_2 και για τους δύο πληθυσμούς, δηλαδή τους δύο τύπους σκυροδέματος, ακολουθούν κανονική κατανομή. Συγκρίνοντας τα θηκογράμματα για τον τύπο A και B (δες Σχήμα 2.2) φαίνεται ότι η κεντρική τάση (εδώ διάμεσος) της αντοχής θραύσης σκυροδέματος τύπου B να είναι χαμηλότερη από αυτή του τύπου A, αλλά ίσως όχι σημαντικά αφού οι δύο θήκες (τα διαστήματα των 50% κεντρικών τιμών του κάθε δείγματος) επικαλύπτονται κατά μεγάλο ποσοστό. Στη συνέχεια θα υπολογίσουμε αν αυτή η διαφορά είναι στατιστικά σημαντική κάνοντας χρήση του διαστήματος εμπιστοσύνης της $\mu_1 - \mu_2$.

Οι δειγματικές μέσες τιμές υπολογίζονται σε $\bar{x}_1 = 5.67$ ksi και $\bar{x}_2 = 5.38$ ksi. Η διαφορά τους είναι $\bar{x}_1 - \bar{x}_2 = 0.29$ ksi. Από τη σχέση (2.20), όπου $z_{1-\alpha/2} = z_{0.975} = 1.96$ και $\sigma_1^2 = \sigma_2^2 = \sigma^2 = 0.38$ βρίσκουμε ότι το 95% διάστημα εμπιστοσύνης είναι

$$0.29 \pm 1.96 \sqrt{0.38 \cdot \left(\frac{1}{25} + \frac{1}{20} \right)} \rightarrow [-0.073, 0.653].$$

Συμπεραίνουμε λοιπόν πως σε επίπεδο εμπιστοσύνης 95% δε μπορούμε να πούμε πως οι δύο τύποι σκυροδέματος διαφέρουν σημαντικά ως προς τη μέση αντοχή θραύσης. Το διάστημα $[-0.073, 0.653]$ είναι σχεδόν θετικό αλλά δε μας επιτρέπει να συμπεράνουμε στατιστικά σημαντική διαφορά. Για τέτοια οριακά αποτελέσματα η μόνη λύση είναι να αυξήσουμε τα δείγματα και να κάνουμε πάλι την εκτίμηση.

Άγνωστες διασπορές Συνήθως όταν δε γνωρίζουμε τις μ_1, μ_2 αγνοούμε και τις σ_1^2, σ_2^2 . Όταν το δείγμα είναι μεγάλο ($n > 30$) μπορούμε να αντικαταστήσουμε στη σχέση (2.20) τις σ_1^2, σ_2^2 με τις δειγματικές διασπορές s_1^2, s_2^2 και να εκτιμήσουμε έτσι το διάστημα εμπιστοσύνης για τη διαφορά $\mu_1 - \mu_2$.

Όταν όμως το μέγεθος του ενός ή και των δύο δειγμάτων είναι μικρό η εκτίμηση του διαστήματος εμπιστοσύνης είναι πιο περίπλοκη. Στην περίπτωση που οι κατανομές των X_1 και X_2 δε δίνονται (ή φαίνονται από τα δεδομένα) να είναι κανονικές, δε μπορούμε γενικά να προσδιορίσουμε την κατανομή της διαφοράς $\bar{x}_1 - \bar{x}_2$ για να εκτιμήσουμε έτσι το διάστημα εμπιστοσύνης. Σε μια τέτοια περίπτωση πρέπει να καταφύγουμε σε μή-παραμετρική εκτίμηση.

Στη συνέχεια υποθέτουμε πως οι κατανομές των X_1 και X_2 είναι κανονικές κι επιπλέον ομοσκεδαστικές ($\sigma_1^2 = \sigma_2^2 = \sigma^2$). Σ' αυτήν την περίπτωση ορίζουμε πρώτα τη *δειγματική κοινή διασπορά* s^2 ως συνάρτηση των s_1^2 και s_2^2

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}. \quad (2.21)$$

Μπορεί να δειχθεί ότι η s^2 είναι αμερόληπτη εκτιμήτρια της κοινής διασποράς σ^2 . Με τη βοήθεια της s^2 η εκτίμηση της διασποράς της διαφοράς $\bar{x}_1 - \bar{x}_2$ είναι $s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$. Για την εκτιμήτρια $\bar{x}_1 - \bar{x}_2$ μπορούμε να ορίσουμε τον παρακάτω μετασχηματισμό που ακολουθεί κατανομή student με $n_1 + n_2 - 2$ βαθμούς ελευθερίας

$$t \equiv \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

από το οποίο προκύπτει πως το $(1 - \alpha)\%$ διάστημα εμπιστοσύνης της διαφοράς $\mu_1 - \mu_2$ είναι

$$(\bar{x}_1 - \bar{x}_2) \pm t_{n_1+n_2-2, 1-\alpha/2} s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}. \quad (2.22)$$

Παράδειγμα 2.15 Για το προηγούμενο παράδειγμα ας υποθέσουμε πως η διασπορά αντοχής θραύσης για τους δύο τύπους σκυροδέματος είναι άγνωστη. Από τα δύο ιστογράμματα στα Σχήματα 1.2 και 2.7 καθώς και τα θηκογράμματα στο Σχήμα 2.7 μπορούμε να δεχτούμε ότι οι τ.μ. αντοχής θραύσης και για τους δύο τύπους ακολουθούν κανονική κατανομή και μάλιστα έχουν την ίδια διασπορά (το εύρος των τιμών των δύο δειγμάτων είναι περίπου το ίδιο). [Τα θηκογράμματα δείχνουν κάποια διαφορά στη διασπορά των τιμών μεταξύ των δύο τύπων αλλά για να αποκλείσουμε ότι οι διασπορές μπορεί να είναι ίσες θα πρέπει η μια θήκη να είναι τουλάχιστον διπλάσια της άλλης.] Δεχόμαστε λοιπόν ότι η κατανομή της αντοχής θραύσης και για τα δύο σκυροδέματα είναι κανονική, οι διασπορές είναι ίδιες κι επειδή επιπλέον τα δείγματα είναι σχετικά μικρά συμπεραίνουμε ότι η διαφορά $\bar{x}_1 - \bar{x}_2$ ακολουθεί κατανομή student.

Η δειγματική διασπορά για τον τύπο A είναι $s_1^2 = 0.375 \text{ (ksi)}^2$ και για τον τύπο B είναι $s_2^2 = 0.326 \text{ (ksi)}^2$ (σχετικά κοντά). Εφαρμόζοντας τη σχέση (2.21) βρίσκουμε ότι η δειγματική κοινή διασπορά είναι $s^2 = 0.353 \text{ (ksi)}^2$ κι επομένως $s = 0.594 \text{ ksi}$ (όπου $n_1 = 25$ και $n_2 = 20$). Οι βαθμοί ελευθερίας είναι $n_1 + n_2 - 2 = 43$ και για επίπεδο εμπιστοσύνης 95% βρίσκουμε από τον στατιστικό πίνακα για την κατανομή student την κρίσιμη τιμή $t_{43, 0.975} = 2.02$ (πολύ κοντά στην αντίστοιχη κρίσιμη τιμή $z_{0.975} = 1.96$ της τυπικής κανονικής κατανομής γιατί οι βαθμοί ελευθερίας είναι πολλοί). Το 95% διάστημα εμπιστοσύνης είναι $(\bar{x}_1 - \bar{x}_2 = 0.29)$

$$0.29 \pm 2.02 \cdot 0.594 \sqrt{\frac{1}{25} + \frac{1}{20}} \rightarrow [-0.07, 0.65].$$

Όπως και πριν που η διασπορά ήταν γνωστή δε μπορούμε να πούμε πως οι μέσες τιμές αντοχής θραύσης των δύο τύπων σκυροδέματος διαφέρουν με στατιστική σημαντικότητα.

Παρατηρήσεις

Οι παράγοντες που επηρεάζουν τον υπολογισμό του διαστήματος εμπιστοσύνης της $\mu_1 - \mu_2$ είναι οι ίδιες όπως για τη μ . Σ' αυτές προστίθεται και ο παράγοντας της ισότητας των διασπορών των X_1 και X_2 . Στον Πίνακα 2.2 συνοψίζεται η εκτίμηση του διαστήματος εμπιστοσύνης της $\mu_1 - \mu_2$ στις διάφορες περιπτώσεις. Υπάρχει φανερή αντιστοιχία των περιπτώσεων για

διασπορές των X_1, X_2	κατανομή των X_1, X_2	n_1, n_2	κατανομή της $\bar{x}_1 - \bar{x}_2$	διάστημα εμπιστοσύνης
γνωστές	κανονική		$z \equiv \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$	$(\bar{x}_1 - \bar{x}_2) \pm z_{1-\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
γνωστές	μη κανονική	μεγάλα	$z \equiv \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$	$(\bar{x}_1 - \bar{x}_2) \pm z_{1-\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
γνωστές	μη κανονική	μικρά	—	—
άγνωστες άνισες/ίσες		μεγάλα	$z \equiv \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0, 1)$	$(\bar{x}_1 - \bar{x}_2) \pm z_{1-\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
άγνωστες ίσες	κανονική	μικρά	$t \equiv \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$	$(\bar{x}_1 - \bar{x}_2) \pm t_{n_1+n_2-2, 1-\alpha/2} s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
άγνωστες ίσες	μη κανονική	μικρά	—	—
άγνωστες άνισες		μικρά	—	—

Πίνακας 2.2: Εκτίμηση του διαστήματος εμπιστοσύνης της διαφοράς $\mu_1 - \mu_2$ ανάλογα με τη γνώση των διασπορών και κατανομών των τ.μ. X_1 και X_2 καθώς και των μεγεθών n_1 και n_2 των αντιστοιχών δειγμάτων.

το διάστημα εμπιστοσύνης της μέσης τιμής (Πίνακας 2.1) και για το διάστημα εμπιστοσύνης της διαφοράς δύο μέσων τιμών (Πίνακας 2.2, οι έξι πρώτες σειρές). Για τη διαφορά μέσω τιμών υπάρχει ακόμα η περίπτωση των άνισων κι αγνώστων διασπορών σε συνδιασμό με μικρά δείγματα (τελευταία σειρά του Πίνακα 2.2) για την οποία δεν μπορούμε να καθορίσουμε την κατανομή της εκτιμήτριας $\bar{x}_1 - \bar{x}_2$ και απ' αυτήν να βρούμε το διάστημα εμπιστοσύνης. Σ' αυτήν την περίπτωση ούτε η μη παραμετρική εκτίμηση μπορεί να δώσει διάστημα εμπιστοσύνης (για τη διάμεσο).

Για το διάστημα εμπιστοσύνης της διαφοράς μέσω τιμών υποθέσαμε ότι οι τ.μ. X_1 και X_2 είναι ανεξάρτητες. Δε θα ασχοληθούμε με την περίπτωση που οι X_1 και X_2 είναι εξαρτημένες (όταν δηλαδή έχουμε *ζευγαρωτές παρατηρήσεις*) γιατί τέτοια προβλήματα δεν παρουσιάζονται συχνά στη μηχανική.

2.2.5 Διάστημα της διαφοράς δύο αναλογιών $p_1 - p_2$

Η εκτίμηση διαστήματος εμπιστοσύνης της διαφοράς δύο αναλογιών $p_1 - p_2$ παρουσιάζεται όταν θέλουμε να συγκρίνουμε δύο πληθυσμούς ως προς μια ιδιότητα, δηλαδή αν η αναλογία p_1 των στοιχείων που πληρούν μια ιδιότητα στον ένα πληθυσμό είναι διαφορετική, κι αν ναι κατά πόσο, από την αντίστοιχη αναλογία p_2 του άλλου πληθυσμού. Για παράδειγμα, θέλουμε να δούμε κατά πόσο διαφέρει η αναλογία κοριτσιών στην Πολυτεχνική Σχολή και στη Φυσικομαθηματική ή θέλουμε να διερευνήσουμε αν οι αναλογίες σκουριασμένων ραβδών χάλυβα σε δύο αποθήκες διαφέρουν και κατά πόσο.

Η σημειακή εκτίμηση της διαφοράς $p_1 - p_2$ δίνεται από τη διαφορά των $\hat{p}_1 = \frac{m_1}{n_1}$ και $\hat{p}_2 = \frac{m_2}{n_2}$, όπου m_1 είναι το πλήθος 'επιτυχιών' στο δείγμα μεγέθους n_1 από τον πρώτο πληθυσμό και m_2 είναι το πλήθος 'επιτυχιών' στο δείγμα μεγέθους n_2 από το δεύτερο πληθυσμό. Γνωρίζουμε ότι για μεγάλα n_1 και n_2 η κατανομή της εκτιμήτριας $\hat{p}_1 - \hat{p}_2$ είναι κανονική με μέση τιμή $E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2$ και διασπορά $\text{Var}(\hat{p}_1 - \hat{p}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$ κι άρα

$$z \equiv \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim N(0, 1).$$

Αντικαθιστώντας στον τύπο της διασποράς τις άγνωστες αναλογίες p_1 και p_2 με τις δειγματικές εκτιμήσεις \hat{p}_1 και \hat{p}_2 βρίσκουμε το $(1 - \alpha)\%$ διάστημα εμπιστοσύνης της διαφοράς $p_1 - p_2$

$$(\hat{p}_1 - \hat{p}_2) \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}. \quad (2.23)$$

Για την εκτίμηση της διασποράς της εκτιμήτριας $\hat{p}_1 - \hat{p}_2$, μπορούμε να υποθέσουμε ότι οι αναλογίες p_1 και p_2 δε διαφέρουν παρά πολύ έτσι ώστε να μπορούμε να χρησιμοποιήσουμε την εκτίμηση της κοινής αναλογίας \hat{p} που ορίζεται ως

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}.$$

Τότε η διασπορά της $\hat{p}_1 - \hat{p}_2$ δίνεται ως $\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})$ και το διάστημα εμπιστοσύνης της διαφοράς $p_1 - p_2$ δίνεται από τη σχέση

$$(\hat{p}_1 - \hat{p}_2) \pm z_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}. \quad (2.24)$$

Παράδειγμα 2.16 Για να εκτιμήσουμε αν υπάρχει διαφορά στο ποσοστό σκουριασμένων ραβδών χάλυβα δύο αποθηκών πήραμε ένα δείγμα 100 ραβδών από την πρώτη αποθήκη και βρήκαμε 12 σκουριασμένες κι ένα δείγμα 120 ραβδών από τη δεύτερη αποθήκη και βρήκαμε 26 σκουριασμένες.

Η εκτίμηση από τα δείγματα για τις αναλογίες σκουριασμένων ραβδών είναι $\hat{p}_1 = \frac{12}{100} = 0.12$ για την πρώτη αποθήκη και $\hat{p}_2 = \frac{26}{120} = 0.217$ για τη δεύτερη αποθήκη. Η διαφορά είναι $\hat{p}_1 - \hat{p}_2 = -0.097$ (σε ποσοστό 9.7%) κι από το 95% διάστημα εμπιστοσύνης μπορούμε να κρίνουμε αν είναι στατιστικά σημαντική. Από τη σχέση (2.23) και για $z_{1-\alpha/2} = 1.96$ βρίσκουμε

$$-0.097 \pm 1.96 \sqrt{\frac{0.12 \cdot 0.88}{100} + \frac{0.217 \cdot 0.783}{120}} \rightarrow [-0.198, 0.004].$$

Το 95% διάστημα εμπιστοσύνης περιέχει οριακά το μηδέν και γι αυτό συμπεραίνουμε πως με αυτά τα δείγματα και σ' αυτό το επίπεδο εμπιστοσύνης η δειγματική διαφορά των αναλογιών αν και είναι σε ποσοστό περίπου 10% δεν είναι στατιστικά σημαντική. Φαίνεται όμως το ποσοστό σκουριασμένων ραβδών στη πρώτη αποθήκη να είναι μικρότερο. Ίσως θα έπρεπε να αυξήσουμε το δείγμα μας και να πάρουμε έτσι στενότερο διάστημα εμπιστοσύνης για να διαπιστώσουμε αν πράγματι περιέχει το 0 ή όχι, αν δηλαδή η διαφορά των δύο αναλογιών είναι στατιστικά σημαντική.

Τα διαστήματα εμπιστοσύνης, πέρα από το ότι δίνουν ένα εύρος τιμών για την παράμετρο που μελετάμε, μας δίνουν τη δυνατότητα να ελέγξουμε αν μια συγκεκριμένη τιμή της παραμέτρου είναι πιθανή βλέποντας αν ανήκει στο διάστημα εμπιστοσύνης. Έτσι, στο τελευταίο παράδειγμα, ελέγχοντας αν η τιμή μηδέν ανήκει στο διάστημα εμπιστοσύνης μας επιτρέπει στην ουσία να ελέγξουμε αν οι δύο αναλογίες διαφέρουν σημαντικά ή όχι.

Στο επόμενο κεφάλαιο θα παρουσιάσουμε τέτοιους στατιστικούς ελέγχους για τις ίδιες παραμέτρους για τις οποίες εκτιμήσαμε διαστήματα εμπιστοσύνης.

Κεφάλαιο 3

ΕΛΕΓΧΟΣ ΣΤΑΤΙΣΤΙΚΩΝ ΥΠΟΘΕΣΕΩΝ

Σε πολλά προβλήματα της μηχανικής δεν ενδιαφερόμαστε να εκτιμήσουμε την τιμή της παραμέτρου αλλά να διαπιστώσουμε αν η παραμέτρος είναι μικρότερη ή μεγαλύτερη από μια δεδομένη τιμή που έχει φυσική σημασία για το πρόβλημα μας. Για παράδειγμα, μας ενδιαφέρει να δούμε αν κατά μέσο όρο η άνοδος (πάνω από το μέσο ετήσιο επίπεδο) της επιφάνειας του νερού ενός ποταμού κατά τη διάρκεια πλημμυρών είναι μεγαλύτερη από μια συγκεκριμένη τιμή, που μας ενδιαφέρει για το κτίσιμο μιας γέφυρας. Βέβαια μπορούμε να απαντήσουμε σ' ένα τέτοιο ερώτημα εκτιμώντας διάστημα εμπιστοσύνης κι ελέγχοντας αν η δεδομένη τιμή ανήκει σ' αυτό ή όχι, αλλά εδώ θα δούμε μια διαφορετική προσέγγιση, θέτοντας κατάλληλες στατιστικές υποθέσεις κι ελέγχοντας αν είναι αποδεκτές ή όχι.

3.1 Διαδικασία Ελέγχου Στατιστικής Υπόθεσης

Για τη διαδικασία ελέγχου μιας στατιστικής υπόθεσης πρώτα ορίζουμε τη στατιστική υπόθεση, μετά υπολογίζουμε τη στατιστική ελέγχου και την περιοχή απόρριψης και τέλος αποφασίζουμε για την υπόθεση με βάση την ένδειξη που έχουμε από το δείγμα.

3.1.1 Στατιστική υπόθεση

Η *στατιστική υπόθεση* (statistical hypothesis) μπορεί να είναι μια οποιαδήποτε 'στατιστική' δήλωση (για κατανομές πληθυσμών, στοχαστικές διαδικασίες κτλ) που θέτουμε υπό έλεγχο με βάση τις παρατηρήσεις, αλλά εδώ θα ασχοληθούμε μόνο με υποθέσεις που αναφέρονται σε κάποια από τις γνωστές παράμετρους που εκτιμήσαμε στο Κεφάλαιο 2.

Υποθέτουμε πως η παράμετρος θ που μελετάμε μπορεί να πάρει μια συγκεκριμένη τιμή θ_0 κι αυτήν την υπόθεση με την οποία ξεκινάμε θα τη λέμε **μηδενική υπόθεση** (null hypothesis) H_0 και θα την παρουσιάζουμε σύντομα ως $H_0 : \theta = \theta_0$. Η **εναλλακτική υπόθεση** (alternative hypothesis) H_1 την οποία δεχόμαστε αν απορρίψουμε τη H_0 μπορεί να είναι:

1. $H_1 : \theta \neq \theta_0$, δηλαδή η παράμετρος να είναι μικρότερη ή μεγαλύτερη από την τιμή θ_0 (**δίπλευρος έλεγχος**). Έτσι καλύπτουμε όλο το σύνολο των δυνατών τιμών της θ .
2. $H_1 : \theta < \theta_0$, δηλαδή η παράμετρος να είναι μικρότερη από την τιμή θ_0 (**μονόπλευρος έλεγχος**). Εδώ περιορίζουμε το σύνολο των δυνατών τιμών της θ στη θ_0 ($H_0 : \theta = \theta_0$) και στις μικρότερες από τη θ_0 ($H_1 : \theta < \theta_0$). Γι αυτό και είναι σωστότερο σ' αυτήν την

περίπτωση να θεωρούμε πως $H_0 : \theta \geq \theta_0$, έστω κι αν πρακτικά η θ δε μπορεί να πάρει τιμές μεγαλύτερες της θ_0 .

3. $H_1 : \theta > \theta_0$, παρόμοια με τη προηγούμενη αλλά για τιμές μεγαλύτερες της θ_0 .

Η επιλογή μονόπλευρου ή δίπλευρου ελέγχου εξαρτάται από την έρευνα που θέλουμε να κάνουμε και από το κατά πόσο μπορούμε να προβλέψουμε το αποτέλεσμα της έρευνας. Για παράδειγμα, αν θέλουμε να ελέγξουμε αν η μέση απόδοση μ_1 ενός μηχανήματος υπό μελέτη μπορεί να φθάσει τη μέση απόδοση μ_2 ενός άριστου μηχανήματος (μηχάνημα αναφοράς) τότε ο έλεγχος πρέπει να είναι μονόπλευρος ($H_0 : \mu_1 \geq \mu_2$ και $H_1 : \mu_1 < \mu_2$) γιατί *γνωρίζουμε* πως δε μπορεί $\mu_1 > \mu_2$ αφού η απόδοση του πρώτου μηχανήματος δε μπορεί να ξεπεράσει αυτή του μηχανήματος αναφοράς. Αν όμως συγκρίνουμε τη μέση απόδοση μ_1 του μηχανήματος που μας ενδιαφέρει με τη μέση απόδοση μ_3 ενός άλλου μηχανήματος της αγοράς κι *ελπίζουμε* το μηχάνημα μας να αποδίδει καλύτερα, ο έλεγχος δε θα πρέπει να είναι μονόπλευρος ($H_0 : \mu_1 \leq \mu_3$ και $H_1 : \mu_1 > \mu_3$) αλλά δίπλευρος ($H_0 : \mu_1 = \mu_3$ και $H_1 : \mu_1 \neq \mu_3$) γιατί από πριν δε μπορούμε να αποκλείσουμε το ενδεχόμενο $\mu_1 < \mu_3$, δηλαδή το μηχάνημα μας να αποδίδει χειρότερα από το άλλο μηχάνημα της αγοράς.

3.1.2 Στατιστική ελέγχου και περιοχή απόρριψης

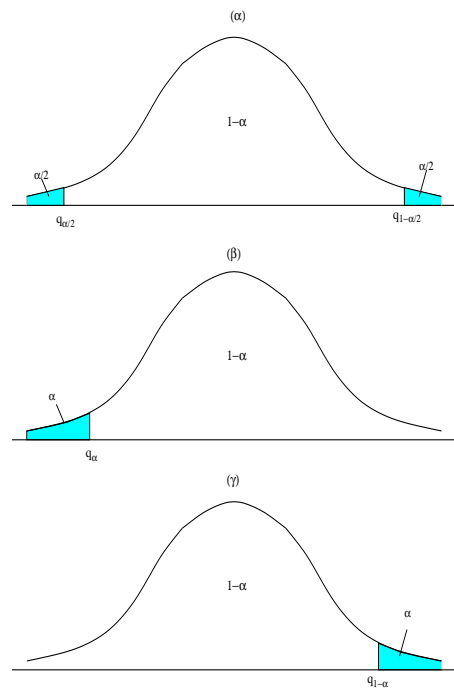
Η μηδενική υπόθεση $H_0 : \theta = \theta_0$ υποδηλώνει ότι η εκτίμηση θ_0 της θ είναι σωστή. Επίσης, τιμές της θ 'κοντά' στη θ_0 υποστηρίζουν την ορθότητα της H_0 ενώ τιμές της θ 'μακριά' από τη θ_0 δεν την υποστηρίζουν. Έτσι, χωρίζουμε τις δυνατές τιμές της παραμέτρου θ σε αυτές για τις οποίες αποδεχόμαστε την H_0 που αποτελούν την περιοχή αποδοχής και σ' αυτές για τις οποίες την απορρίπτουμε που αποτελούν την **περιοχή απόρριψης** (rejection region) που συμβολίζουμε R .

Η απόφαση για την αποδοχή ή απόρριψη της H_0 γίνεται βάση πιθανοτήτων κι όπως για τα διαστήματα εμπιστοσύνης έτσι κι εδώ ορίζουμε επίπεδο εμπιστοσύνης $(1 - \alpha)$ για την απόφαση ελέγχου. Το α λέγεται **επίπεδο σημαντικότητας** (significance level) και καθορίζει το 'κοντά' και 'μακριά' που αναφέραμε παραπάνω, δηλαδή καθορίζει το εύρος της περιοχής αποδοχής κι απόρριψης. Όταν μικραίνει το α μικραίνει κι η περιοχή απόρριψης R και μεγαλώνει η περιοχή αποδοχής. Η σχέση α και περιοχής αποδοχής είναι ίδια με τη σχέση α και διαστήματος εμπιστοσύνης της θ .

Επειδή είναι συνήθως υπολογιστικά δύσκολο να καθορίσουμε από την κατανομή της εκτιμήτριας θ τα κρίσιμα σημεία που θα μας δώσουν την R καταφεύγουμε σε κάποια άλλη τ.μ. που την ονομάζουμε **στατιστική ελέγχου** (test statistic) q με γνωστή κατανομή και για την οποία μπορούμε να ορίσουμε την R . Στο στατιστικό έλεγχο κυρίων παραμέτρων, για τις οποίες βρήκαμε διαστήματα εμπιστοσύνης στο Κεφάλαιο 2, η q είναι μια από τις z , t και χ^2 που ακολουθούν γνωστές κατανομές.

Ο έλεγχος ξεκινάει υποθέτοντας ότι η μηδενική υπόθεση H_0 είναι σωστή. Με βάση τη H_0 προσδιορίζουμε την στατιστική ελέγχου q και την κατανομή της. Όταν ο έλεγχος είναι *παραμετρικός* υποθέτουμε κάποια κατανομή για την παράμετρο θ κι η q σχετίζεται άμεσα με τη θ (η q προκύπτει από μετασχηματισμό της θ). Όταν ο έλεγχος είναι *μη-παραμετρικός* δε θεωρούμε κάποια κατανομή για τη θ κι η κατανομή της q βασίζεται σε άλλες ιδιότητες της θ . Στη συνέχεια θα ασχοληθούμε μόνο με παραμετρικούς ελέγχους.

Η κατανομή της στατιστικής ελέγχου q δίνει την πιθανότητα η q να πάρει κάποια τιμή (αν η q είναι διακριτή τ.μ.) ή να βρίσκεται σ' ένα διάστημα τιμών (αν η q είναι συνεχής τ.μ.) όταν η H_0 είναι αληθής. Αντίστροφα μπορούμε να πούμε πως με βάση αυτή την κατανομή, αν παρατηρήσουμε τιμές της q που αντιστοιχούν σε μεγάλες πιθανότητες αυτό δείχνει πως η H_0 είναι αληθής, ενώ αν παρατηρήσουμε τιμές της q που αντιστοιχούν σε μικρές πιθανότητες αυτό υποδηλώνει αμφιβολία για την ισχύ της H_0 . Άρα μη πιθανές τιμές της q συνιστούν την απόρριψη της H_0 . Η οριακή πιθανότητα για την αποδοχή ή απόρριψη της H_0 είναι το επίπεδο σημαντικότητας α κι αυτό καθορίζει την **κρίσιμη τιμή** (critical value), ή τις κρίσιμες τιμές, της q για τον προσδιορισμό της περιοχής αποδοχής και της περιοχής απόρριψης R της H_0 . Κατά κανόνα η περιοχή απόρριψης σχηματίζεται από τα άκρα της κατανομής της στατιστικής ελέγχου q όπως αυτά ορίζονται από τις κρίσιμες τιμές. Αν ο έλεγχος είναι δίπλευρος τότε οι κρίσιμες τιμές $q_{\alpha/2}$ και $q_{1-\alpha/2}$ ορίζουν την περιοχή απόρριψης της H_0 ως $R = \{q \mid q < q_{\alpha/2} \vee q > q_{1-\alpha/2}\}$, δηλαδή σχηματίζεται από τις δύο ουρές της κατανομής της q με συνολική πιθανότητα α . Αν ο έλεγχος είναι μονόπλευρος τότε υπάρχει μόνο μία κρίσιμη τιμή, q_α ή $q_{1-\alpha}$, που ορίζει την περιοχή απόρριψης της H_0 , $R = \{q \mid q < q_\alpha\}$ για την αριστερή πλευρά και $R = \{q \mid q > q_{1-\alpha}\}$ για τη δεξιά πλευρά. Στο Σχήμα 3.1 δίνονται σχηματικά οι περιοχές αποδοχής κι απόρριψης για δίπλευρο και μονόπλευρο έλεγχο.



Σχήμα 3.1: Σχηματικά παρουσιάζεται η συνάρτηση πυκνότητας πιθανότητας της στατιστικής ελέγχου q , η περιοχή αποδοχής και η περιοχή απόρριψης (σκιασμένη), για δίπλευρο έλεγχο στο διάγραμμα (α) και μονόπλευρο έλεγχο στο (β) και (γ).

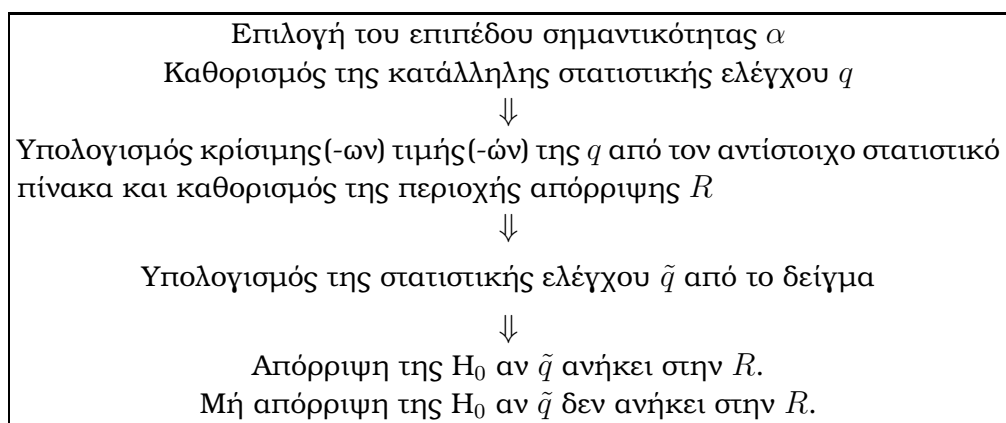
3.1.3 Απόφαση ελέγχου

Η απόφαση του έλεγχου γίνεται με βάση το δείγμα που δίνει και την εκτίμηση $\hat{\theta}$ για την παράμετρο θ . Βέβαια αφού έχουμε σχηματίσει την R ως προς την q , χρησιμοποιούμε αντί της

τιμής $\hat{\theta}$ την τιμή \tilde{q} (που προκύπτει από τον ίδιο μετασχηματισμό που κάναμε για να πάρουμε την q) και είναι η τιμή της στατιστικής ελέγχου από το δείγμα. Στη συνέχεια εξετάζουμε αν η \tilde{q} ανήκει στην περιοχή απόρριψης R που ορίσαμε κι ανάλογα απορρίπτουμε ή δεν απορρίπτουμε τη H_0 στο επίπεδο σημαντικότητας α . Συνήθως αποφεύγουμε να λέμε ότι δεχόμαστε την H_0 γιατί αυτό το ενδεχόμενο δεν έχει ελεγχθεί με την ίδια αυστηρότητα όπως το ενδεχόμενο της απόρριψης της H_0 .

Μπορούμε να δώσουμε ακόμα πιο αναλυτική απάντηση για το επίπεδο σημαντικότητας της απόρριψης ή μη της H_0 για το συγκεκριμένο δείγμα. Στην τιμή \tilde{q} της στατιστικής ελέγχου από το δείγμα αντιστοιχεί σύμφωνα με την κατανομή της q κάποια πιθανότητα p . Αυτή η p -**τιμή** (p -value) του ελέγχου είναι η πιθανότητα να παρατηρήσουμε για τη q μια τιμή τόσο ακραία όσο η \tilde{q} όταν ισχύει η H_0 . Δηλαδή η p -τιμή δηλώνει το μικρότερο επίπεδο σημαντικότητας για το οποίο μπορούμε να απορρίψουμε την H_0 με βάση το δείγμα. Γι αυτό το λόγο η p -τιμή είναι χρήσιμη στον έλεγχο υποθέσεων αλλά ο υπολογισμός της είναι πρακτικά πιο δύσκολος αφού πρέπει να υπολογίσουμε την αθροιστική συνάρτηση κατανομής για $q = \tilde{q}$.

Συνοπτικά η διαδικασία του στατιστικού ελέγχου για $H_0 : \theta = \theta_0$ είναι



3.2 Έλεγχος Παραμέτρων Πληθυσμού

Περνάμε τώρα να ορίσουμε ελέγχους υποθέσεων για τις παραμέτρους για τις οποίες ορίσαμε διαστήματα εμπιστοσύνης στο Κεφάλαιο 2.

3.2.1 Έλεγχος για τη μέση τιμή μ

Ο παραμετρικός έλεγχος για τη μέση τιμή μ μιας τ.μ. X γίνεται με βάση την κατανομή της εκτιμήτριας \bar{x} για τυχαίο δείγμα μεγέθους n . Ανάλογα με το αν η διασπορά της τ.μ. X είναι γνωστή ή όχι, αν η κατανομή της X είναι γνωστή ή όχι κι αν το n είναι μεγάλο ή μικρό, μπορούμε να υποθέσουμε μια συγκεκριμένη κατανομή (κανονική ή student) για τη στατιστική ελέγχου q που είναι κάποιος μετασχηματισμός της \bar{x} . Τις περιπτώσεις αυτές τις είχαμε συνοψίσει στον Πίνακα 2.1 και οι κατανομές της στατιστικής ελέγχου είναι ίδιες με τις κατανομές που χρησιμοποιήσαμε για να ορίσουμε τα αντίστοιχα διαστήματα εμπιστοσύνης. Γι αυτό εδώ θα δούμε περιληπτικά τους ελέγχους γι αυτές τις περιπτώσεις.

Γνωστή διασπορά Όταν γνωρίζουμε τη διασπορά σ^2 της τ.μ. X η εκτιμήτρια \bar{x} ακολουθεί κανονική κατανομή με μέση τιμή τη μέση τιμή μ του πληθυσμού και διασπορά σ^2/n . Άρα

σύμφωνα με τη μηδενική υπόθεση $H_0 : \mu = \mu_0$ η στατιστική ελέγχου είναι

$$q = z \equiv \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1), \quad (3.1)$$

όπου \bar{x} είναι η εκτιμήτρια συνάρτηση της μ . Η περιοχή απόρριψης της H_0 καθορίζεται από τις κρίσιμες τιμές της τυπικής κανονικής κατανομής, όπως και για το διάστημα εμπιστοσύνης της μ . [Η κρίσιμη τιμή της z για τον στατιστικό έλεγχο συμβολίζεται με τη βοήθεια ενός δείκτη που είναι η αντίστοιχη τιμή της αθροιστικής συνάρτησης κατανομής Φ , π.χ. η κρίσιμη τιμή z_α είναι τέτοια ώστε $P(z < z_\alpha) = \Phi(z_\alpha) = \alpha$. Λόγω συμμετρικότητας της τυπικής κανονικής κατανομής γύρω από το μηδέν, είναι $z_\alpha = -z_{1-\alpha}$ (δες Παράγραφο 2.2.1).] Η περιοχή απόρριψης για επίπεδο σημαντικότητας α είναι ανάλογα με την εναλλακτική υπόθεση

1. $H_1 : \mu \neq \mu_0$, δίπλευρος έλεγχος και $R = \{z \mid z < z_{\alpha/2} \vee z > z_{1-\alpha/2}\}$ ή $R = \{z \mid |z| > z_{1-\alpha/2}\}$.
2. $H_1 : \mu < \mu_0$, μονόπλευρος έλεγχος από αριστερά και $R = \{z \mid z < z_\alpha\} = \{z \mid z < -z_{1-\alpha}\}$.
3. $H_1 : \mu > \mu_0$, μονόπλευρος έλεγχος από δεξιά και $R = \{z \mid z > z_{1-\alpha}\}$.

Στη συνέχεια υπολογίζουμε από τη σχέση (3.1) την στατιστική ελέγχου \tilde{z} που παίρνουμε από το δείγμα μας (θέτοντας \bar{x} την τιμή της εκτίμησης από το δείγμα). Αν η τιμή \tilde{z} ανήκει στην R (ανάλογα με την περίπτωση ελέγχου), απορρίπτουμε την H_0 στο επίπεδο σημαντικότητας α , αλλιώς λέμε ότι δεν έχουμε αρκετές ενδείξεις για να την απορρίψουμε.

Η p -τιμή για τα τρία είδη ελέγχου είναι

1. $H_1 : \mu \neq \mu_0$, $p = 2P(z > |\tilde{z}|) = 2(1 - P(z < |\tilde{z}|))$.
2. $H_1 : \mu < \mu_0$, $p = P(z < \tilde{z}) = 1 - P(z < -\tilde{z})$.
3. $H_1 : \mu > \mu_0$, $p = P(z > \tilde{z}) = 1 - P(z < \tilde{z})$.

Παράδειγμα 3.1 Θέλουμε να ελέγξουμε αν η μέση αντοχή θραύσης σκυροδέματος μ είναι 6 ksi. Υποθέτουμε ότι γνωρίζουμε πως η διασπορά σ^2 είναι 0.4 ksi². Χρησιμοποιούμε τα δεδομένα του σκυροδέματος τύπου A του Πίνακα 1.3. Οι υποθέσεις είναι $H_0 : \mu = 6$ και $H_1 : \mu \neq 6$. Για το επίπεδο σημαντικότητας $\alpha = 0.05$ η κρίσιμη τιμή είναι $z_{0.975} = 1.96$ και η περιοχή απόρριψης της H_0 είναι $R = \{z \mid |z| > 1.96\}$. Η στατιστική ελέγχου \tilde{z} από το δείγμα είναι ($\bar{x} = 5.67$, $n = 25$)

$$\tilde{z} = \frac{5.67 - 6}{\sqrt{0.4/25}} = -2.61$$

κι ανήκει στην R . Άρα σε επίπεδο σημαντικότητας 0.05 (ή σε επίπεδο εμπιστοσύνης 95%) μπορούμε να απορρίψουμε τη H_0 και συμπεραίνουμε πως η μέση αντοχή θραύσης σκυροδέματος δε μπορεί να είναι 6 ksi. Από την αθροιστική τυπική κανονική κατανομή για $\tilde{z} = -2.61$ βρίσκουμε την p -τιμή

$$p = 2(1 - P(z < |\tilde{z}|)) = 2(1 - \Phi(|\tilde{z}|)) = 2(1 - \Phi(2.61)) = 2(1 - 0.9955) = 0.009$$

που δηλώνει πως μπορούμε με εμπιστοσύνη μέχρι και ποσοστού 99.1% να πούμε πως η μέση αντοχή θραύσης σκυροδέματος δε μπορεί να είναι 6 ksi.

Άγνωστη διασπορά Όταν η διασπορά της τ.μ. X είναι άγνωστη και το δείγμα μεγάλο τότε η στατιστική ελέγχου q έχει την ίδια κατανομή όπως παραπάνω αλλά στη σχέση (3.1) αντικαθιστούμε τη διασπορά σ^2 με τη δειγματική διασπορά s^2 .

Όταν το δείγμα είναι μικρό αλλά η X ακολουθεί κανονική κατανομή τότε η στατιστική ελέγχου είναι

$$q = t \equiv \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}. \quad (3.2)$$

Η περιοχή απόρριψης R ορίζεται με τη βοήθεια των κρίσιμων τιμών της κατανομής student (λόγω συμμετρικότητας της κατανομής είναι $t_{n-1,1-\alpha} = -t_{n-1,\alpha}$)

1. $H_1 : \mu \neq \mu_0, \quad R = \{t \mid |t| > t_{n-1,1-\alpha/2}\}.$
2. $H_1 : \mu < \mu_0, \quad R = \{t \mid t < -t_{n-1,1-\alpha}\}.$
3. $H_1 : \mu > \mu_0, \quad R = \{t \mid t > t_{n-1,1-\alpha}\}.$

Η p -τιμή για τα τρία είδη ελέγχου είναι όπως για γνωστή διασπορά αντικαθιστώντας το z με t .

Παράδειγμα 3.2 Στο παραπάνω παράδειγμα έστω ότι δε ξέρουμε τη διασπορά σ^2 . Υποθέτουμε ωστόσο πως η αντοχή θραύσης του σκυροδέματος τύπου A ακολουθεί κανονική κατανομή (δες Σχήμα 2.2).

Για το επίπεδο σημαντικότητας $\alpha = 0.05$ η κρίσιμη τιμή βρίσκεται από την κατανομή student και είναι $t_{24,0.975} = 2.06$. Άρα η περιοχή απόρριψης της H_0 είναι $R = \{t \mid |t| > 2.06\}$. Από προηγούμενους υπολογισμούς (δες Παράδειγμα 2.6) έχουμε $\bar{x} = 5.67$, $n = 25$, $\sigma^2 = 0.375$. Άρα η στατιστική ελέγχου \tilde{t} από το δείγμα είναι

$$\tilde{t} = \frac{5.67 - 6.0}{\sqrt{0.375/25}} = -2.69$$

που ανήκει και πάλι στην R και η απόφαση ελέγχου είναι η ίδια όπως και πριν. Από την αθροιστική συνάρτηση της t -κατανομής για $\tilde{z} = -2.69$ βρίσκουμε την p -τιμή

$$p = 2(1 - P(t \leq |\tilde{t}|)) = 2(1 - P(t \leq 2.69)) = 2(1 - 0.994) = 0.012$$

που δηλώνει ποσοστό εμπιστοσύνης της απόρριψης της H_0 στο ίδιο περίπου επίπεδο όπως και πριν.

Στον Πίνακα 3.1 συνοψίζεται ο έλεγχος υπόθεσης $H_0 : \mu = \mu_0$ στις διάφορες περιπτώσεις (φυσικά υπάρχει απόλυτη αντιστοίχιση των περιπτώσεων του έλεγχου υποθέσης στον Πίνακα 3.1 με τις περιπτώσεις του διαστήματος εμπιστοσύνης στον Πίνακα 2.1).

3.2.2 Έλεγχος για τη διασπορά σ^2

Η στατιστική υπόθεση για τη διασπορά σ^2 είναι όπως και για τη μέση τιμή, δηλαδή $H_0 : \sigma^2 = \sigma_0^2$ με κατάλληλη εναλλακτική υπόθεση H_1 ανάλογα αν ο έλεγχος είναι δίπλευρος ή μονόπλευρος. Η στατιστική q του παραμετρικού ελέγχου είναι όπως και για το διάστημα εμπιστοσύνης της σ^2 (Παράγραφος 2.2.2)

$$q = \chi^2 \equiv \frac{(n-1)s^2}{\sigma_0^2} \sim \chi_{n-1}^2, \quad (3.3)$$

διασπορά	κατανομή της X	n	κατανομή στατιστικής q
γνωστή	κανονική		$z \equiv \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$
γνωστή	μη κανονική	μεγάλο	$z \equiv \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$
γνωστή	μη κανονική	μικρό	—
άγνωστη		μεγάλο	$z \equiv \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim N(0, 1)$
άγνωστη	κανονική	μικρό	$t \equiv \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$
άγνωστη	μη κανονική	μικρό	—

Πίνακας 3.1: Στατιστική ελέγχου υπόθεσης $H_0 : \mu = \mu_0$ ανάλογα με τη γνώση της διασποράς και κατανομής της τ.μ. X καθώς και με το μέγεθος n του δείγματος.

όπου s^2 είναι η εκτιμήτρια συνάρτηση της διασποράς. Από την κατανομή χ^2 με $n - 1$ βαθμούς ελευθερίας και για επίπεδο σημαντικότητας α βρίσκουμε τις κρίσιμες τιμές κι η περιοχή απόρριψης είναι ανάλογα με τον τύπο ελέγχου (αυτή η κατανομή δεν είναι συμμετρική)

1. $H_1 : \sigma^2 \neq \sigma_0^2, \quad R = \{\chi^2 | \chi^2 < \chi_{n-1, \alpha/2}^2 \vee \chi^2 > \chi_{n-1, 1-\alpha/2}^2\}$.
2. $H_1 : \sigma^2 < \sigma_0^2, \quad R = \{\chi^2 | \chi^2 < \chi_{n-1, \alpha}^2\}$.
3. $H_1 : \sigma^2 > \sigma_0^2, \quad R = \{\chi^2 | \chi^2 > \chi_{n-1, 1-\alpha}^2\}$.

Αν η δειγματική στατιστική ελέγχου $\tilde{\chi}^2$ (υπολογίζεται θέτοντας στη σχέση (3.3) την εκτίμηση s^2 από το δείγμα) ανήκει στην R η H_0 απορρίπτεται. Η p -τιμή για τα τρία είδη ελέγχου είναι

1. $H_1 : \sigma^2 \neq \sigma_0^2, \quad p = P(\chi^2 < \tilde{\chi}^2 \vee \chi^2 > \tilde{\chi}^2)$.
2. $H_1 : \sigma^2 < \sigma_0^2, \quad p = P(\chi^2 < \tilde{\chi}^2)$.
3. $H_1 : \sigma^2 > \sigma_0^2, \quad p = P(\chi^2 > \tilde{\chi}^2)$.

Παράδειγμα 3.3 Θέλουμε να ελέγξουμε αν θα μπορούσαμε να πούμε με μεγάλη σιγουριά (επίπεδο εμπιστοσύνης 99%) πως η τυπική απόκλιση σ της αντοχής θραύσης σκυροδέματος μπορεί να ξεπεράσει το επίπεδο του 1 ksi. Γι αυτό εφαρμόζουμε μονόπλευρο έλεγχο για τη διασπορά: $H_0 : \sigma^2 \leq 1, H_1 : \sigma^2 > 1$, αφού δε μας ενδιαφέρει αν η διασπορά μπορεί να είναι πολύ μικρή, πράγμα που το αποκλείουμε, αλλά αν είναι στο επίπεδο της μονάδας (μηδενική υπόθεση) ή σημαντικά μεγαλύτερη του 1 (εναλλακτική υπόθεση). (Αν ο έλεγχος αφορούσε κάποια άλλη τιμή της τυπικής απόκλισης θα έπρεπε να πάρουμε το τετράγωνο της.)

Για τον έλεγχο χρησιμοποιούμε και πάλι το δείγμα σκυροδέματος τύπου A μεγέθους $n = 25$ (Πίνακας 1.3) και η δειγματική διασπορά είναι $s^2 = 0.375$ (ksi)². Η κρίσιμη τιμή της στατιστικής χ^2 για $\alpha = 0.01$ είναι $\chi_{24, 0.99}^2 = 42.98$ κι η περιοχή απόρριψης είναι $R = \{\chi^2 | \chi^2 > 42.98\}$. Η στατιστική ελέγχου που παίρνουμε από το δείγμα είναι

$$\tilde{\chi}^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{24 \cdot 0.375}{1} = 9.$$

Η $\tilde{\chi}^2$ δεν ανήκει στην περιοχή απόρριψης κι άρα δε μπορούμε να απορρίψουμε την H_0 . Είναι απίθανο λοιπόν η διασπορά της αυτοχθής θραύσης σκυροδέματος να ξεπεράσει το επίπεδο του 1 ksi^2 και το ίδιο ισχύει και για την τυπική απόκλιση.

Γενικά όταν θέλουμε να κάνουμε στατιστικό έλεγχο για την τυπική απόκλιση σ υψώνουμε στο τετράγωνο την τιμή της τυπικής απόκλισης που θέλουμε να ελέγξουμε και κάνουμε έλεγχο διασποράς γι αυτήν την τιμή.

3.2.3 Έλεγχος για την αναλογία p

Για να ελέγξουμε αν η αναλογία p ενός πληθυσμού παίρνει κάποια τιμή p_0 χρησιμοποιούμε τις ίδιες στατιστικές υποθέσεις και την ίδια κατανομή της στατιστικής ελέγχου όπως για τη μέση τιμή μ με τη διασπορά γνωστή. Έτσι η μηδενική υπόθεση είναι $H_0 : p = p_0$ και η στατιστική ελέγχου είναι (δες επίσης Παράγραφο 2.2.3 για το διάστημα εμπιστοσύνης της p)

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} \sim N(0, 1), \quad (3.4)$$

όπου \hat{p} είναι η εκτιμήτρια συνάρτηση της αναλογίας p κι έχουμε κάνει την υπόθεση πως το μέγεθος του δείγματος n είναι μεγάλο.

Η περιοχή απόρριψης είναι

1. $H_1 : p \neq p_0, R = \{z \mid |z| > z_{1-\alpha/2}\}.$
2. $H_1 : p < p_0, R = \{z \mid z < -z_{1-\alpha}\}.$
3. $H_1 : p > p_0, R = \{z \mid z > z_{1-\alpha}\}.$

Η δειγματική στατιστική ελέγχου \tilde{z} υπολογίζεται από τη σχέση (3.4) θέτοντας ως \hat{p} τη εκτίμηση της αναλογίας από το δείγμα. Αν η \tilde{z} ανήκει στην R απορρίπτουμε την H_0 στο επίπεδο σημαντικότητας α .

Η p -τιμή για τα τρία είδη ελέγχου είναι όπως και για τη μ με γνωστή διασπορά.

Παράδειγμα 3.4 Μια εταιρεία ενδιαφέρεται να ελέγξει αν το ποσοστό των σκουριασμένων ραβδών χάλυβα στην αποθήκη της είναι 19%. Ο έλεγχος είναι δίπλευρος κι η στατιστική υπόθεση είναι: $H_0 : p = 0.19$ και $H_1 : p \neq 0.19$.

Για τον έλεγχο αυτό χρησιμοποιήθηκε το ίδιο δείγμα όπως στο Παράδειγμα 2.13 (12 σκουριασμένες ραβδοί σε δείγμα 100 ραβδών). Η κρίσιμη τιμή για τη στατιστική έλεγχο σε επίπεδο σημαντικότητας $\alpha = 0.05$ είναι $z_{0.975} = 1.96$ κι η περιοχή απόρριψης είναι $R = \{z \mid |z| > 1.96\}$. Η εκτίμηση της αναλογίας από το δείγμα των 100 ραβδών είναι $\hat{p} = 0.12$ κι από τη σχέση (3.4) βρίσκουμε τη στατιστική έλεγχο από το δείγμα

$$\tilde{z} = \frac{0.12 - 0.19}{\sqrt{0.19 \cdot 0.81/100}} = -1.784.$$

Η τιμή της \tilde{z} οριακά δεν ανήκει στην R κι άρα δε μπορούμε να απορρίψουμε την H_0 και συμπεραίνουμε πως το ποσοστό των σκουριασμένων ραβδών μπορεί και να είναι 19%.

Αν τώρα η εταιρεία γνωρίζει από προηγούμενους ελέγχους ότι είναι απίθανο το ποσοστό των σκουριασμένων ραβδών στην αποθήκη της να κυμαίνεται σε επίπεδα μεγαλύτερα του 19% κι

αυτό το ποσοστό είναι το όριο της αγοράς που δε θα πρέπει να ξεπερνάει το κάθε φορτίο ραβδών χάλυβα, τότε η εταιρεία ενδιαφέρεται να ελέγξει αν πράγματι το ποσοστό των σκουριασμένων ραβδών στην αποθήκη της δεν φτάνει αυτό το όριο. Ο έλεγχος εδώ είναι μονόπλευρος κι η στατιστική υπόθεση είναι: $H_0 : p \geq 0.19$ και $H_1 : p < 0.19$.

Η δειγματική στατιστική ελέγχου είναι η ίδια, $\tilde{z} = -1.784$, αλλά η κρίσιμη τιμή για το ίδιο επίπεδο σημαντικότητας $\alpha = 0.05$ είναι $z_{0.95} = 1.65$ κι η περιοχή απόρριψης είναι $R = \{z \mid z < -1.65\}$. Άρα σ' αυτήν την περίπτωση μπορούμε (οριακά πάλι) να απορρίψουμε την H_0 και να συμπεράνουμε ότι το ποσοστό των σκουριασμένων ραβδών στην αποθήκη της εταιρείας δε ξεπερνάει το όριο της αγοράς 19%. Το ανώτατο επίπεδο εμπιστοσύνης που μπορούμε να απορρίψουμε την $H_0 : p \geq 0.19$ είναι περίπου 96% όπως προκύπτει από την p -τιμή που αντιστοιχεί στη δειγματική στατιστική του μονόπλευρου ελέγχου

$$P(z < \tilde{z}) = P(z < -1.784) = \Phi(-1.784) = 1 - \Phi(1.784) = 1 - 0.963 = 0.037.$$

Το παράδειγμα αυτό δείχνει πως γενικά ο δίπλευρος έλεγχος είναι πιο αυστηρός από τον μονόπλευρο γιατί στο δίπλευρο έλεγχο η ουρά της κατανομής (στα αριστερά και δεξιά) που αποτελεί την περιοχή απόρριψης είναι μικρότερη (μισή σε μέγεθος) από την αντίστοιχη ουρά για το μονόπλευρο έλεγχο (δες Σχήμα 3.1).

3.2.4 Έλεγχος για τη διαφορά μέσων τιμών $\mu_1 - \mu_2$

Ο έλεγχος της διαφοράς δύο μέσων τιμών μ_1 και μ_2 χρησιμοποιείται συχνά όταν θέλουμε να συγκρίνουμε δύο ανεξάρτητους πληθυσμούς ως προς μια μετρήσιμη ποσότητα, όπως για παράδειγμα τη βροχόπτωση σε δύο περιοχές. Αν ονομάσουμε X_1 και X_2 τις αντίστοιχες ανεξάρτητες τ.μ. θέλουμε να ελέγξουμε αν έχουν την ίδια κατανομή. Αυτό είναι αρκετά πολύπλοκο πρόβλημα γι αυτό και περιορίζουμε την έρευνα στη μέση τιμή και η μηδενική υπόθεση είναι $H_0 : \mu_1 = \mu_2$ ή ισοδύναμα $H_0 : \mu_1 - \mu_2 = 0$. Μπορούμε λοιπόν να θεωρήσουμε ότι το πρόβλημα μας είναι ο έλεγχος της μέσης τιμής $\mu_1 - \mu_2$ της τ.μ. $X_1 - X_2$ ως προς την τιμή 0.

Στην Παράγραφο 2.2.4 όπου μελετήσαμε το διάστημα εμπιστοσύνης της $\mu_1 - \mu_2$ είδαμε ότι αν οι τ.μ. X_1 και X_2 ακολουθούν κανονική κατανομή ή τα μεγέθη των δειγμάτων n_1 και n_2 είναι μεγάλα ($n_1, n_2 > 30$) τότε η εκτιμήτρια $\bar{x}_1 - \bar{x}_2$ ακολουθεί κανονική κατανομή με μέση τιμή $\mu_1 - \mu_2$ και διασπορά $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$. Εδώ και πάλι θα ξεχωρίσουμε τις περιπτώσεις που οι διασπορές σ_1^2 και σ_2^2 είναι γνωστές ή άγνωστες.

Γνωστές διασπορές Όταν γνωρίζουμε τις διασπορές σ_1^2 και σ_2^2 , τότε θεωρώντας τη μηδενική υπόθεση $H_0 : \mu_1 - \mu_2 = 0$, η στατιστική του παραμετρικού ελέγχου είναι

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1). \quad (3.5)$$

Οι κρίσιμες τιμές κι οι περιοχές απόρριψης για δίπλευρο και μονόπλευρο έλεγχο είναι όπως και για τον έλεγχο μέσης τιμής με γνωστή διασπορά (δες Παράγραφο 3.2.1). Η στατιστική ελέγχου \tilde{z} από το δείγμα υπολογίζεται από τη σχέση (3.5), όπου \bar{x}_1 και \bar{x}_2 είναι οι εκτιμήσεις των μέσων τιμών από το δείγμα.

Άγνωστες διασπορές Όταν οι διασπορές είναι άγνωστες και τα δείγματα είναι μεγάλα χρησιμοποιούμε την ίδια στατιστική ελέγχου z της σχέσης (3.5) όπου απλά αντικαθιστούμε τις άγνωστες διασπορές σ_1^2 και σ_2^2 με τις εκτιμήσεις s_1^2 και s_2^2 από τα δείγματα.

Όταν τα δείγματα είναι μικρά αλλά υποθέτουμε ότι οι τ.μ. X_1 και X_2 ακολουθούν κανονική κατανομή και οι διασπορές είναι ίδιες ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) τότε η στατιστική ελέγχου είναι

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2} \quad (3.6)$$

όπου $s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$ είναι η αμερόληπτη εκτιμήτρια της κοινής διασποράς σ^2 (δες Παράγραφο 2.2.4). Οι κρίσιμες τιμές κι οι περιοχές απόρριψης για δίπλευρο και μονόπλευρο έλεγχο είναι όπως και για τον έλεγχο μέσης τιμής με άγνωστη διασπορά (δες Παράγραφο 3.2.1). Η στατιστική ελέγχου \tilde{t} από το δείγμα υπολογίζεται από τη σχέση (3.6), όπου \bar{x}_1 , \bar{x}_2 και s^2 υπολογίζονται από τα δείγματα των δύο πληθυσμών.

Παράδειγμα 3.5 Θέλουμε να ερευνήσουμε αν δύο τύποι σκυροδέματος A και B είναι ίδιοι ως προς την αντοχή θραύσης. Γι αυτό κάνουμε δίπλευρο έλεγχο για τις μέσες αντοχές θραύσης μ_1 και μ_2 για τους δύο τύπους A και B αντίστοιχα και είναι $H_0 : \mu_1 = \mu_2$ και $H_1 : \mu_1 \neq \mu_2$. Για τον έλεγχο χρησιμοποιούμε τα δείγματα αντοχής θραύσης για τον τύπο σκυροδέματος A και B του Πίνακα 1.3. Υποθέτουμε πως η αντοχή θραύσης για κάθε τύπο σκυροδέματος ακολουθεί κανονική κατανομή το οποίο φαίνεται να ισχύει κι από τα ιστογράμματα και τα θηκογράμματα αντοχής θραύσης των δύο δειγμάτων (δες Σχήματα 1.2 και 2.7). Οι δειγματικές μέσες τιμές είναι $\bar{x}_1 = 5.67$ ksi και $\bar{x}_2 = 5.38$ ksi για τον τύπο A και B αντίστοιχα.

Πρώτα θεωρούμε πως από παλιότερες μετρήσεις γνωρίζουμε πως η διασπορά της αντοχής θραύσης είναι ίδια και για τους δύο τύπους και είναι $\sigma^2 = 0.4$ (ksi)². Άρα η στατιστική ελέγχου είναι η z όπως ορίσαμε στη σχέση (3.5). Για το τυπικό επίπεδο σημαντικότητας $\alpha = 0.05$ η κρίσιμη τιμή για το δίπλευρο έλεγχο είναι $z_{0.975} = 1.96$ κι η περιοχή απόρριψης είναι $R = \{z \mid |z| > 1.96\}$. Η στατιστική ελέγχου του δείγματος είναι

$$\tilde{z} = \frac{5.67 - 5.38}{\sqrt{\frac{0.4}{25} + \frac{0.4}{20}}} = 1.53$$

κι αφού δεν ανήκει στην R δεν έχουμε αρκετές ενδείξεις για να απορρίψουμε την $H_0 : \mu_1 = \mu_2$. Άρα όπως προκύπτει από το δείγμα μας, δε μπορούμε με στατιστική εμπιστοσύνη 95% να πούμε πως η μέση αντοχή θραύσης στα σκυροδέματα τύπου A και B είναι διαφορετική. Η τιμή $\tilde{z} = 1.53$ είναι αρκετά κοντά στην περιοχή απόρριψης κι ίσως ένα μεγαλύτερο δείγμα από τους δύο τύπους να μας επέτρεπε να απορρίψουμε την H_0 . Πράγματι η p -τιμή για αυτόν τον έλεγχο είναι

$$p = 2(1 - P(z < |\tilde{z}|)) = 2(1 - \Phi(|\tilde{z}|)) = 2(1 - \Phi(1.53)) = 2(1 - 0.937) = 0.126$$

δηλαδή η πιθανότητα να έχουμε αυτές τις παρατηρήσεις όταν ισχύει η $H_0 : \mu_1 = \mu_2$ είναι 0.126, ή αλλιώς μπορούμε να απορρίψουμε την H_0 σε μικρό επίπεδο εμπιστοσύνης 87% περίπου.

Θεωρούμε τώρα ότι η διασπορά της αντοχής θραύσης για τους δύο τύπους είναι κοινή αλλιώς δεν εμπιστευόμαστε την εμπειρική εκτίμηση $\sigma^2 = 0.4$ (ksi)². Άρα θα πρέπει να χρησιμοποιήσουμε στον έλεγχο τις εκτιμήσεις των διασπορών για τους δύο τύπους. Στο Παράδειγμα 2.15 βρήκαμε $s_1^2 = 0.375$ (ksi)² για τον τύπο A , $s_2^2 = 0.326$ (ksi)² για τον τύπο B και $s^2 = 0.353$ (ksi)² για την κοινή διασπορά σ^2 ($s = 0.594$ ksi).

Η στατιστική ελέγχου σ' αυτήν την περίπτωση είναι η t όπως ορίσαμε στη σχέση (3.6). Η κρίσιμη τιμή είναι $t_{43,0.975} = -t_{43,0.025} = 2.02$ κι η περιοχή απόρριψης είναι $R = \{t \mid |t| > 2.02\}$. Η στατιστική ελέγχου του δείγματος είναι

$$\tilde{t} = \frac{5.67 - 5.38}{0.594 \sqrt{\frac{1}{25} + \frac{1}{20}}} = 1.63$$

η οποία και πάλι δεν ανήκει στην R κι άρα δε μπορούμε να απορρίψουμε την $H_0 : \mu_1 = \mu_2$. Η p -τιμή για αυτόν τον έλεγχο είναι

$$p = 2(1 - P(t < |\tilde{t}|)) = 2(1 - P(t \leq 1.63)) = 2(1 - 0.945) = 0.11,$$

δηλαδή μπορούμε να απορρίψουμε την H_0 σε επίπεδο εμπιστοσύνης 0.89% που είναι κι αυτό μικρό. Το αποτέλεσμα του στατιστικού ελέγχου είναι σε συμφωνία με το διάστημα εμπιστοσύνης $[-0.07, 0.65]$ της $\mu_1 - \mu_2$ στο ίδιο επίπεδο εμπιστοσύνης που βρήκαμε στο Παράδειγμα 2.15, το οποίο οριακά περιέχει το μηδέν.

Στον Πίνακα 3.2 συνοψίζεται ο έλεγχος της $H_0 : \mu_1 - \mu_2 = 0$ για ανεξάρτητους πληθυσμούς στις διάφορες περιπτώσεις.

διασπορές των X_1, X_2	κατανομή των X_1, X_2	n_1, n_2	κατανομή στατιστικής q
γνωστή	κανονική		$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$
γνωστές	μη κανονική	μεγάλα	$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$
γνωστές	μη κανονική	μικρά	—
άγνωστες ίσες/άνισες		μεγάλα	$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0, 1)$
άγνωστες άνισες	κανονική	μικρά	$t = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$
άγνωστες άνισες	μη κανονική	μικρά	—
άγνωστες άνισες		μικρά	—

Πίνακας 3.2: Έλεγχος της $H_0 : \mu_1 - \mu_2 = 0$ ανάλογα με τη γνώση των διασπορών και κατανομών των ανεξάρτητων τ.μ. X_1 και X_2 καθώς και των μεγεθών n_1 και n_2 των αντιστοίχων δειγμάτων.

Ζευγαρωτές παρατηρήσεις Ο έλεγχος γίνεται διαφορετικά αν οι πληθυσμοί δεν είναι ανεξάρτητοι ή ισοδύναμα αν οι τ.μ. X_1 και X_2 δεν είναι ανεξάρτητες, όπως όταν θέλουμε να συγκρίνουμε ένα χαρακτηριστικό ενός πληθυσμού πριν και μετά τη μεταβολή ενός εξωτερικού παράγοντα.

Για παράδειγμα αν υπολογίζουμε την ανθεκτικότητα των δοκών ενός κτιρίου σε κάποια αριθμητική κλίμακα, μας ενδιαφέρει να συγκρίνουμε την ανθεκτικότητα πριν και μετά ένα αρκετά μεγάλο σεισμό. Βέβαια αυτό είναι δυνατόν αν πρώτα μετρήσαμε την ανθεκτικότητα των δοκών σ' ένα δείγμα κτιρίων πριν το σεισμό και μετά το σεισμό μετρήσαμε πάλι την ανθεκτικότητα των ίδιων δοκών. Είναι φανερό ότι οι τ.μ. X_1 και X_2 είναι εξαρτημένες (η ανθεκτικότητα των δοκών ενός κτιρίου μετά το σεισμό καθορίζεται ως ένα βαθμό από την ανθεκτικότητά τους πριν το σεισμό).

Η διαδικασία ελέγχου είναι αρκετά απλή σ' αυτήν την περίπτωση. Προϋποθέτει βέβαια ότι έχουμε παρατηρήσει τα αντίστοιχα στοιχεία των πληθυσμών (π.χ. τους δοκούς των κτιρίων πριν και μετά το σεισμό) γι αυτό κι αυτή η περίπτωση αναφέρεται συχνά ως **ζευγαρωτές παρατηρήσεις**. Πριν προχωρήσουμε στη διαδικασία ελέγχου διατάσσουμε τις n ζευγαρωτές παρατηρήσεις από του δύο πληθυσμούς σε ζευγάρια $(x_{11}, x_{21}), \dots, (x_{1n}, x_{2n})$ και υπολογίζουμε τη διαφορά $d_i = x_{1i} - x_{2i}$ για $i = 1, \dots, n$. Οι τιμές d_1, d_2, \dots, d_n αποτελούν το τυχαίο δείγμα των παρατηρήσεων της τ.μ. $D \equiv X_1 - X_2$ που έχει μέση τιμή $\mu_D = \mu_1 - \mu_2$. Στη συνέχεια εφαρμόζουμε τον έλεγχο μέσης τιμής με $H_0 : \mu_D = 0$ βασιζόμενη στο δείγμα d_1, d_2, \dots, d_n (η διαδικασία ελέγχου είναι όπως στην Παράγραφο 3.2.1).

3.2.5 Έλεγχος για τη διαφορά δύο αναλογιών $p_1 - p_2$

Όταν θέλουμε να συγκρίνουμε δύο πληθυσμούς ως προς μια ιδιότητα κάνουμε έλεγχο για την μηδενική υπόθεση ότι οι δύο αναλογίες p_1 και p_2 είναι ίσες, ή αντίστοιχα $H_0 : p_1 - p_2 = 0$, όπου η κάθε αναλογία εκφράζει το ποσοστό των στοιχείων που πληρούν την συγκεκριμένη ιδιότητα στον αντίστοιχο πληθυσμό.

Αν \hat{p}_1 είναι η αναλογία που μετρήσαμε σε τυχαίο δείγμα μεγέθους n_1 από τον πρώτο πληθυσμό και \hat{p}_2 είναι η αναλογία σε τυχαίο δείγμα μεγέθους n_2 από το δεύτερο πληθυσμό και τα δείγματα είναι μεγάλα τότε είδαμε στην Παράγραφο 2.2.5 πως

$$z \equiv \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim N(0, 1).$$

Σύμφωνα με τη μηδενική υπόθεση είναι $p_1 = p_2 = p$ και η στατιστική ελέγχου είναι

$$q = z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim N(0, 1). \quad (3.7)$$

Για να βρούμε τη στατιστική ελέγχου από το δείγμα εκτιμούμε την κοινή αναλογία p από τα δύο δείγματα

$$\hat{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$$

και βρίσκουμε

$$\tilde{z} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}. \quad (3.8)$$

Οι κρίσιμες τιμές κι οι περιοχές απόρριψης για δίπλευρο και μονόπλευρο έλεγχο είναι όπως και στις άλλες περιπτώσεις που η στατιστική ελέγχου ακολουθεί τυπική κανονική κατανομή.

Παράδειγμα 3.6 Στο Παράδειγμα 2.16 εκτιμήσαμε το διάστημα εμπιστοσύνης για τη διαφορά αναλογιών των σκουριασμένων ραβδών χάλυβα από δύο αποθήκες. Εδώ θέλουμε να ελέγξουμε αν οι αναλογίες σκουριασμένων ραβδών χάλυβα διαφέρουν στις δύο αποθήκες.

Τα δεδομένα είναι όπως στο Παράδειγμα 2.16, δηλαδή 12 σκουριασμένες ράβδοι σε δείγμα 100 ραβδών από την πρώτη αποθήκη και 26 σκουριασμένες ράβδοι σε δείγμα 120 ραβδών από τη δεύτερη αποθήκη. Οι δειγματικές αναλογίες είναι $\hat{p}_1 = 0.12$ και $\hat{p}_2 = 0.217$ αντίστοιχα.

Για δίπλευρο έλεγχο της $H_0 : p_1 - p_2 = 0$ η κρίσιμη τιμή σε επίπεδο σημαντικότητας $\alpha = 0.05$ είναι $z_{0.975} = 1.96$ και η περιοχή απόρριψης είναι $R = \{z \mid |z| > 1.96\}$. Η εκτίμηση της κοινής αναλογίας είναι

$$\hat{p} = \frac{100 \cdot 0.12 + 120 \cdot 0.217}{100 + 120} = 0.173$$

και η στατιστική ελέγχου \tilde{z} από το δείγμα είναι

$$\tilde{z} = \frac{0.12 - 0.217}{\sqrt{0.173 \cdot (1 - 0.173) \cdot \left(\frac{1}{100} + \frac{1}{120}\right)}} = -1.89.$$

Οριακά λοιπόν η \tilde{z} δεν ανήκει στην R και δε μπορούμε να απορρίψουμε σε επίπεδο εμπιστοσύνης 95% πως οι δύο αναλογίες είναι ίσες. Το συμπέρασμα αυτό είναι σε συμφωνία με το διάστημα εμπιστοσύνης $[-0.198, 0.004]$ για τη διαφορά $p_1 - p_2$ στο ίδιο επίπεδο εμπιστοσύνης (δες Παράδειγμα 2.16). Το διάστημα αυτό επίσης οριακά περιέχει το 0. Η δειγματική στατιστική ελέγχου $\tilde{z} = -1.89$ αντιστοιχεί σε p -τιμή

$$p = 2(1 - P(z < |\tilde{z}|)) = 2(1 - \Phi(|\tilde{z}|)) = 2(1 - \Phi(1.89)) = 2(1 - 0.97) = 0.06.$$

Αυτό σημαίνει πως θα μπορούσαμε να απορρίψουμε την H_0 αν είχαμε αυξήσει το επίπεδο σημαντικότητας κατά μία ποσοστιαία μονάδα (δηλαδή αν είχαμε θέσει σαν επίπεδο εμπιστοσύνης το 94%).

Αν τώρα πριν τον έλεγχο γνωρίζαμε πως η δεύτερη αποθήκη έχει πάντοτε πιο παλιές ράβδους από την πρώτη, που σημαίνει ότι η δεύτερη αποθήκη θα έχει το ίδιο ποσοστό σκουριασμένων ραβδών με την πρώτη ή μεγαλύτερο τότε ο έλεγχος θα ήταν μονόπλευρος ($H_0 : p_1 \geq p_2$, $H_1 : p_1 < p_2$). Σ' αυτήν την περίπτωση η κρίσιμη τιμή για επίπεδο σημαντικότητας $\alpha = 0.05$ είναι $z_{0.95} = 1.65$, η περιοχή απόρριψης είναι $R = \{z \mid z < -1.65\}$ κι άρα απορρίπτουμε την H_0 αφού η δειγματική στατιστική ελέγχου $\tilde{z} = -1.89$ ανήκει στην R .

Συμπερασματικά, φαίνεται πως η αναλογία σκουριασμένων ραβδών στην δεύτερη αποθήκη είναι μεγαλύτερη απλά με αυτό το δείγμα δε μπορούμε να το αποφασίσουμε με μεγάλη βεβαιότητα μετά από στατιστικό έλεγχο. Σε τέτοιες ακραίες περιπτώσεις θα πρέπει να επιδιώξουμε να αυξήσουμε το μέγεθος του δείγματος μας.

Είδαμε κάποιους ελέγχους υποθέσεων που αφορούν κύριες παραμέτρους κατανομής μιας ή δύο τ.μ.. Υπάρχουν κι άλλοι έλεγχοι παραμέτρων που δεν παρουσιάζονται εδώ γιατί έχουν λιγότερη εφαρμογή σε προβλήματα μηχανικής, όπως ο έλεγχος της ισότητας δύο διασπορών. Υπάρχουν επίσης έλεγχοι που δεν αναφέρονται σε παραμέτρους κατανομής αλλά στις ίδιες τις κατανομές, όπως ο έλεγχος προσαρμογής (αν η εμπειρική κατανομή από ένα δείγμα είναι η θεωρητική κατανομή που υποθέτουμε) κι ο έλεγχος ανεξαρτησίας για δύο τ.μ..

3.3 Χαρακτηριστικά Ελέγχου

Περνάμε τώρα να δούμε κάποια χαρακτηριστικά του ελέγχου στατιστικής υπόθεσης.

3.3.1 Αποφάσεις ελέγχου και πραγματική κατάσταση

Ο έλεγχος υπόθεσης καταλήγει στην απόφαση επιλογής της μίας από τις δύο υποθέσεις, της μηδενικής H_0 ή της εναλλακτικής H_1 . Κάθε μία από τις δύο υποθέσεις μπορεί να είναι σωστή ή λανθασμένη. Συνδυάζοντας τις δύο δυνατές αποφάσεις του ελέγχου με τις δύο δυνατές πραγματικές καταστάσεις για την στατιστική υπόθεση έχουμε τέσσερις περιπτώσεις:

1. **Σωστή απόφαση:** Αποδεχόμαστε την H_0 όταν η H_0 είναι σωστή. Η πιθανότητα αυτής της απόφασης είναι

$$P(\text{αποδοχή της } H_0 \mid H_0 \text{ σωστή}) = 1 - \alpha.$$

2. **Σφάλμα τύπου II** (type II error): Αποδεχόμαστε την H_0 όταν η H_0 είναι λανθασμένη. Η πιθανότητα αυτού του σφάλματος είναι

$$P(\text{αποδοχή της } H_0 \mid H_0 \text{ λανθασμένη}) = \beta.$$

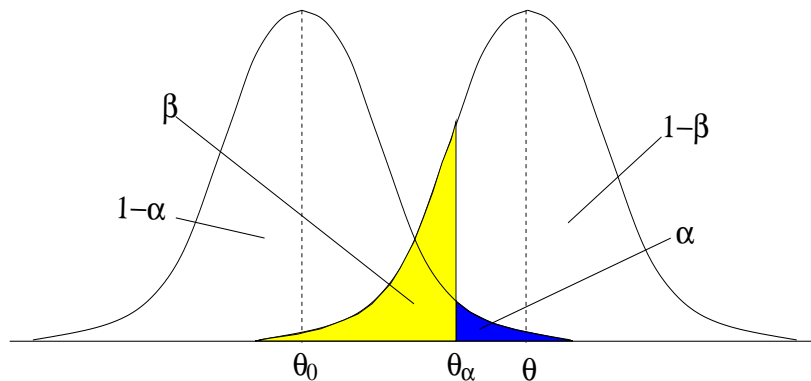
3. **Σφάλμα τύπου I** (type I error): Απορρίπτουμε την H_0 όταν η H_0 είναι σωστή. Η πιθανότητα αυτού του σφάλματος είναι το επίπεδο σημαντικότητας

$$P(\text{απόρριψη της } H_0 \mid H_0 \text{ σωστή}) = \alpha.$$

4. **Σωστή απόφαση:** Απορρίπτουμε την H_0 και η H_0 είναι λανθασμένη. Η πιθανότητα αυτής της απόφασης είναι

$$P(\text{απόρριψη της } H_0 \mid H_0 \text{ λανθασμένη}) = 1 - \beta$$

και δηλώνει τη **δύναμη ελέγχου** (power of the test).



Σχήμα 3.2: Σχηματικά δίνονται οι συναρτήσεις πυκνότητας πιθανότητας της εκτιμήτριας της παραμέτρου θ όταν η H_0 είναι σωστή (με κέντρο θ_0) και λανθασμένη (με κέντρο κάποια άλλη τιμή θ) κι οι πιθανότητες για κάθε περιοχή ανάλογα με την απόφαση ελέγχου.

Οι πιθανότητες των παραπάνω περιπτώσεων για μονόπλευρο έλεγχο κάποιας παραμέτρου θ διαγράφονται στο Σχήμα 3.2. Η μία κατανομή της εκτιμήτριας $\hat{\theta}$ της θ , αριστερά στο Σχήμα 3.2, είναι σύμφωνη με τη μηδενική υπόθεση $H_0 : \theta = \theta_0$ κι η άλλη κατανομή είναι σύμφωνη με την εναλλακτική υπόθεση. Η κρίσιμη τιμή της στατιστικής ελέγχου αντιστοιχεί στην κρίσιμη τιμή

θ_α της θ . Θεωρώντας την πρώτη κατανομή της $\hat{\theta}$ (η H_0 είναι σωστή), αριστερά από τη θ_α είναι η περιοχή σωστής αποδοχής της H_0 με πιθανότητα $1 - \alpha$, ενώ δεξιά της θ_α (η βαθεία σκούρη περιοχή) είναι η περιοχή λανθασμένης απόρριψης της H_0 με πιθανότητα α (σφάλμα τύπου I). Θεωρώντας τη δεύτερη κατανομή της $\hat{\theta}$ (η H_0 είναι λανθασμένη), αριστερά από τη θ_α είναι η περιοχή λανθασμένης αποδοχής της H_0 (η ελαφρά σκούρη περιοχή) με πιθανότητα β (σφάλμα τύπου II), ενώ δεξιά της θ_α είναι η περιοχή σωστής απόρριψης της H_0 με πιθανότητα $1 - \beta$.

Όταν σχεδιάζουμε να κάνουμε έναν στατιστικό έλεγχο έχοντας ένα τυχαίο δείγμα, πρέπει να διαλέξουμε πως θα αντισταθίσουμε τους δύο τύπους σφάλματος. Αν προσπαθήσουμε να αποτρέψουμε το σφάλμα τύπου I μειώνοντας το α αυξάνουμε τον κίνδυνο να διαπράξουμε μεγαλύτερο σφάλμα τύπου II κι αντίστροφα, όπως φαίνεται κι από το Σχήμα 3.2. Στην πραγματικότητα (όπου δε γνωρίζουμε την πραγματική τιμή της παραμέτρου) δε μπορούμε να διαπιστώσουμε τι σφάλμα μπορεί να κάνουμε γι αυτό σε κάθε έλεγχο πρέπει να λάβουμε υπόψη τις πρακτικές συνέπειες του κάθε σφάλματος.

Παράδειγμα 3.7 *Ας δεχτούμε πως ένα φορτίο ραβδών χάλυβα θεωρείται αποδεκτό αν το ποσοστό των σκουριασμένων ραβδών δε ξεπερνάει το 19%. Όπως είδαμε στο Παράδειγμα 3.4, μια εταιρεία ελέγχει (πριν να παραδώσει φορτία από την αποθήκη της) αν στην αποθήκη της το ποσοστό των σκουριασμένων ραβδών δε ξεπερνάει το 19% ($H_0 : p \geq 0.19$). Η απόφαση ελέγχου με βάση το δείγμα του Παραδείγματος 3.4 είναι η απόρριψη της H_0 σε επίπεδο σημαντικότητας $\alpha = 0.05$ (μάλιστα η απόρριψη μπορεί να γίνει ως το επίπεδο σημαντικότητας που δίνεται από την p -τιμή $p = 0.037$).*

Η εταιρεία μπορεί να κάνει το σφάλμα τύπου I, δηλαδή αποφασίζει ότι το ποσοστό των σκουριασμένων ραβδών δε ξεπερνάει το 19% ενώ στην πραγματικότητα το ξεπερνάει. Αν η εταιρεία θέλει να μειώσει αυτόν τον κίνδυνο θα πρέπει να ελαττώσει το επίπεδο σημαντικότητας α στον έλεγχο της υπόθεσης. Αυτό μάλλον είναι σημαντικό για την εταιρεία που θέλει να ελαχιστοποιήσει τον κίνδυνο να χάσει την παραγγελία επειδή μπορεί να βρεθεί σε κάποιο φορτίο ποσοστό σκουριασμένων ραβδών μεγαλύτερο του 19%. Αν απορρίψει την $H_0 : p \geq 0.19$ για μικρότερο α , π.χ. $\alpha = 0.01$, τότε είναι ικανοποιημένη πως ο κίνδυνος σφάλματος τύπου I που θα την έκανε να χάσει την παραγγελία είναι πολύ μικρός (μόνο 1%). Αν δεν την απορρίψει (όπως θα γινόταν για το δείγμα του παραδείγματος) τότε έχει αμφιβολία για το αν τηρεί το όριο αφού αποδέχεται ότι μπορεί το ποσοστό σκουριασμένων ραβδών να υπερβαίνει το 19%. Τότε ίσως επιλέξει να αφαιρέσει σκουριασμένες ραβδούς από την αποθήκη για να μειώσει το ποσοστό των σκουριασμένων ραβδών και να το ρίξει κάτω από το όριο. Είναι όμως πράγματι το ποσοστό πάνω από το όριο;

Για μικρότερο α η πιθανότητα β , δηλαδή να αποφασίσει ότι το ποσοστό των σκουριασμένων ραβδών ξεπερνάει το 19% ενώ στην πραγματικότητα είναι μικρότερο από 19% (σφάλμα τύπου II), είναι μεγαλύτερη. Αυτό σημαίνει ότι μεγαλώνει η πιθανότητα η εταιρεία να αποφασίσει λανθασμένα να ελαττώσει τον αριθμό των σκουριασμένων ραβδών από την αποθήκη της. Σε αυτήν την περίπτωση η εταιρεία θα μείωνε ακόμα περισσότερο το ήδη μικρό ποσοστό των σκουριασμένων ραβδών, δηλαδή θα μείωνε χωρίς λόγο τα έσοδα της. Γι αυτό και η εταιρεία πρέπει να αποφασίσει σε ποιά συνέπεια θα δώσει μεγαλύτερο βάρος, στην πιθανή απόρριψη της παραγγελίας ή στην πιθανή μείωση εσόδων της από την αφαίρεση σκουριασμένων ραβδών από την αποθήκη;

3.3.2 Δύναμη ελέγχου

Σε μια στατιστική υπόθεση συνήθως η εναλλακτική υπόθεση είναι αυτή που παρουσιάζει ενδιαφέρον για το πρόβλημα μας γι αυτό και θέλουμε με τον έλεγχο που κάνουμε να μπορούμε

με μεγάλη πιθανότητα να απορρίψουμε τη μηδενική υπόθεση όταν είναι λανθασμένη. Θέλουμε δηλαδή ο έλεγχος να έχει μεγάλη δύναμη $1 - \beta$. Για κάποια στατιστική υπόθεση μπορεί να προκύπτουν διαφορετικοί έλεγχοι, ένας για κάθε κατάλληλη στατιστική ελέγχου που έχουμε τη δυνατότητα να χρησιμοποιήσουμε. Για παράδειγμα, για τον έλεγχο της μέσης τιμής μ μπορεί να διαλέξουμε τον παραμετρικό έλεγχο με στατιστικές z ή t ανάλογα με την περίπτωση (δες Πίνακα 3.1) ή το μή-παραμετρικό έλεγχο με άλλη στατιστική που δε μελετήσαμε εδώ. Για να αξιολογήσουμε δύο τέτοιους ελέγχους για την ίδια στατιστική υπόθεση συγκρίνουμε τη δύναμη τους στο ίδιο φυσικά επίπεδο σημαντικότητας α και στο ίδιο δείγμα. Έτσι έχει βρεθεί ότι οι παραμετρικοί έλεγχοι έχουν μεγαλύτερη δύναμη από τους μη-παραμετρικούς για αυτό και προτιμούνται όταν είναι δυνατόν να γίνουν (στον Πίνακα 3.1 για παράδειγμα παραθέτονται περιπτώσεις όπου ο παραμετρικός έλεγχος δεν είναι δυνατός).

Όπως φαίνεται κι από το Σχήμα 3.2, η δύναμη ελέγχου μεγαλώνει όταν μικραίνει το σφάλμα τύπου II β , δηλαδή όταν μικραίνει η περιοχή που οι δύο κατανομές αλληλοκαλύπτονται. Τρεις παράγοντες που καθορίζουν αυτήν την αλληλοκάλυψη είναι:

1. Το επίπεδο σημαντικότητας α : Όταν αυξάνεται το α μειώνεται το β κι άρα αυξάνεται η δύναμη ελέγχου $1 - \beta$.
2. Η διασπορά της κατανομής της εκτιμήτριας $\hat{\theta}$: Όταν μικραίνει η διασπορά, 'στενεύει' η κατανομή, η περιοχή αλληλοκάλυψης μικραίνει κι άρα αυξάνει η δύναμη ελέγχου. Η διασπορά της κατανομής μειώνεται πάντα με την αύξηση του μεγέθους του δείγματος n . Η διασπορά της κατανομής της εκτιμήτριας εξαρτάται επίσης από τη διασπορά σ^2 της τ.μ. X .
3. Το μέγεθος της διαφοράς $\theta - \theta_0$: Φανερά η δύναμη ελέγχου μεγαλώνει καθώς η υποτιθέμενη τιμή της παραμέτρου θ_0 απομακρύνεται από την πραγματική τιμή της παραμέτρου θ .

Παράδειγμα 3.8 Στο Παράδειγμα 3.1 ελέγξαμε αν η μέση αντοχή θραύσης σκυροδέματος μ μπορεί να είναι 6 ksi ($H_0 : \mu = \mu_0 = 6$). Υποθέσαμε γνωστή διασπορά $\sigma^2 = 0.4$ (ksi)² και το δείγμα μεγέθους $n = 25$ έδωσε την εκτίμηση $\bar{x} = 5.67$.

Οι κρίσιμες τιμές της στατιστικής ελέγχου για το επίπεδο εμπιστοσύνης $\alpha = 0.05$ είναι $z_{0.025} = -1.96$ και $z_{0.975} = 1.96$ αφού ο έλεγχος είναι δίπλευρος. Οι αντίστοιχες κρίσιμες τιμές για την παράμετρο μ μπορούν να υπολογισθούν εύκολα από το μετασχηματισμό $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$, όπου θεωρούμε πως ισχύει η H_0 . Ο μετασχηματισμός αυτός δίνει για την αριστερή κρίσιμη τιμή

$$-1.96 = \frac{\bar{x}_{0.025} - 6}{\sqrt{0.4/25}} \Rightarrow \bar{x}_{0.025} = 5.752$$

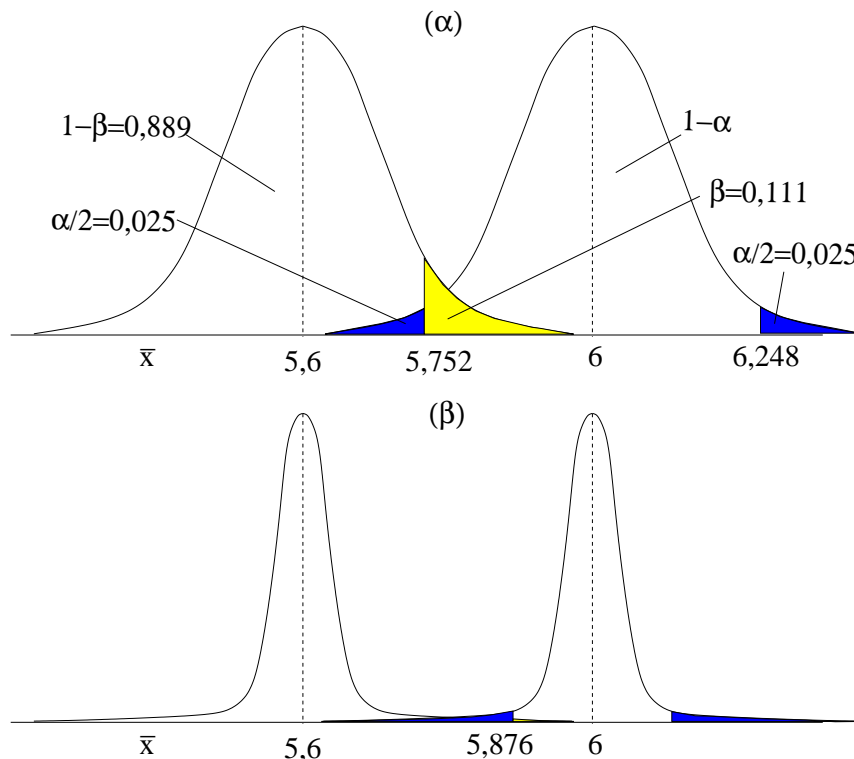
και για τη δεξιά κρίσιμη τιμή

$$1.96 = \frac{\bar{x}_{0.975} - 6}{\sqrt{0.4/25}} \Rightarrow \bar{x}_{0.975} = 6.248.$$

Άρα η πιθανότητα του σφάλματος τύπου I ορίζεται ως

$$P(\text{απόρριψη της } H_0 \mid H_0 \text{ σωστή}) = P(\bar{x} < 5.752 \vee \bar{x} > 6.248 \mid \mu = 6) = 0.05$$

όπως φαίνεται στο Σχήμα 3.3α από τις βαθεία σκούρες περιοχές.



Σχήμα 3.3: Σχηματικά οι συναρτήσεις πυκνότητας πιθανότητας της εκτιμήτριας \bar{x} της παραμέτρου μ όταν η H_0 είναι σωστή ($\mu = 6$) και λανθασμένη ($\mu = 5.6$) και οι πιθανότητες για κάθε περιοχή ανάλογα με την απόφαση ελέγχου. Στο (α) το μέγεθος του δείγματος είναι 25 και στο (β) είναι 100.

Ας υποθέσουμε τώρα ότι η πραγματική τιμή της μ είναι 5.6 ksi. Η δύναμη ελέγχου είναι

$$1 - \beta = P(\bar{x} < 5.752 \vee \bar{x} > 6.248 \mid \mu = 5.6).$$

Η πρώτη πιθανότητα είναι

$$P(\bar{x} < 5.752 \mid \mu = 5.6) = P\left(\frac{\bar{x} - 5.6}{\sqrt{0.4/25}} < \frac{5.752 - 5.6}{\sqrt{0.4/25}}\right) = P(z < 1.202) = 0.889,$$

όπου για το τελικό αποτέλεσμα χρησιμοποιήσαμε τον στατιστικό πίνακα για την τυπική κανονική κατανομή. Η δεύτερη πιθανότητα είναι αμελητέα κι άρα η δύναμη ελέγχου είναι 0.889, όπως φαίνεται στο Σχήμα 3.3α, όπου η συμπληρωματική πιθανότητα β του σφάλματος τύπου II δίνεται από τη ελαφρά σκούρα περιοχή.

Είναι φανερό πως η δύναμη ελέγχου θα αυξηθεί αν μετατοπίσουμε την κρίσιμη τιμή $\bar{x}_{0.025} = 5.752$ προς τα δεξιά, που μπορεί να γίνει με τους παρακάτω τρόπους:

- Αυξάνουμε το επίπεδο σημαντικότητας α . Έτσι διακινδυνεύουμε περισσότερο να απορρίψουμε την H_0 ενώ είναι σωστή. Σίγουρα δε θα θέλαμε να αυξήσουμε τη δύναμη ελέγχου μ' αυτόν τον τρόπο.
- Διαλέγουμε τιμή $\mu_0 > 6$ για τη μηδενική υπόθεση. Προφανώς έχουμε κάποιο λόγο να ελέγξουμε για την τιμή $\mu_0 = 6$ και όχι για κάποια άλλη τιμή (ίσως είναι η τυπική αντοχή θραύσης σκυροδέματος που θα πρέπει να έχει το σκυρόδεμα για να θεωρείται αποδεκτό).

- Υποθέτουμε ότι η διασπορά της αντοχής θραύσης είναι μικρότερη ($\sigma^2 < 0.4 \text{ ksi}^2$). Η μεταβλητότητα της αντοχής θραύσης όμως είναι φυσικό φαινόμενο και δε μπορούμε να την ορίσουμε όπως θέλουμε (μπορούμε βέβαια να την εκτιμήσουμε λιάδος!).
- Μένει ο τελευταίος παράγοντας που είναι το μέγεθος του δείγματος κι ας υποθέσουμε πως αυξάνουμε το δείγμα έτσι ώστε $n = 100$. Τότε η αριστερή κρίσιμη τιμή της εκτιμήτριας \bar{x} για το ίδιο επίπεδο σημαντικότητας είναι

$$-1.96 = \frac{\bar{x}_{0.025} - 6}{\sqrt{0.4/100}} \Rightarrow \bar{x}_{0.025} = 5.876.$$

Όπως φαίνεται κι από το Σχήμα 3.3β αυτό μειώνει σημαντικά την πιθανότητα β σφάλματος τύπου II και μεγαλώνει σημαντικά την περιοχή της δύναμης ελέγχου. Πράγματι όταν το n αυξάνει από 25 σε 100 η διασπορά $\frac{\sigma^2}{n}$ της εκτιμήτριας \bar{x} μειώνεται κατά τέσσερις φορές και η δύναμη ελέγχου $1 - \beta$ γίνεται

$$P(\bar{x} < 5.876 \mid \mu = 5.6) = P\left(\frac{\bar{x} - 5.6}{\sqrt{0.4/100}} < \frac{5.876 - 5.6}{\sqrt{0.4/100}}\right) = P(z < 4.36) \simeq 1.$$

Το παραπάνω παράδειγμα δείχνει πόσο σημαντικό είναι για τον έλεγχο να επιδιώκουμε να συλλέγουμε μεγάλα δείγματα. Η σχέση της δύναμης ελέγχου $1 - \beta$ με το μέγεθος του δείγματος n χρησιμοποιείται στην πράξη για να καθορίσουμε το απαραίτητο μέγεθος του δείγματος. Για συγκεκριμένη δύναμη ελέγχου $1 - \beta$ και διαφορά $\theta - \theta_0$ που δίνεται έτσι ώστε να έχει αξία ο έλεγχος που θέλουμε να κάνουμε μπορούμε να βρούμε το μέγεθος του δείγματος που θα χρειαστούμε. Γι αυτό υπάρχουν στατιστικοί πίνακες που μας δίνουν για δεδομένα α , $1 - \beta$ και διαφορά $\theta - \theta_0$ την τιμή του n για έλεγχο της κάθε παραμέτρου θ .

3.3.3 Δίπλευρος έλεγχος και διάστημα εμπιστοσύνης

Είδαμε στα παραδείγματα ότι το $(1 - \alpha)\%$ διάστημα εμπιστοσύνης για μια παράμετρο θ είναι σε συμφωνία με την απόφαση του αντίστοιχου δίπλευρου ελέγχου. Αν η μηδενική υπόθεση ενός δίπλευρου ελέγχου $H_0 : \theta = \theta_0$ απορρίπτεται τότε η τιμή θ_0 δεν ανήκει στο διάστημα εμπιστοσύνης της θ ενώ αν δεν απορρίπτεται ανήκει. Αυτό συμβαίνει γιατί και τα δύο αποτελέσματα βασίζονται στη γνωστή κατανομή κάποιας κατάλληλης τ.μ., στην οποία μετασχηματίζεται η εκτιμήτρια $\hat{\theta}$ (z ή t όταν η παράμετρος θ είναι μ , $\mu_1 - \mu_2$, p ή $p_1 - p_2$ και χ^2 όταν θ είναι σ^2).

Παράδειγμα 3.9 Στο Παράδειγμα 2.10 από ένα δείγμα 25 παρατηρήσεων αντοχής θραύσης σκυροδέματος με δειγματική μέση τιμή $\bar{x} = 5.67 \text{ ksi}$, βρήκαμε το 95% διάστημα εμπιστοσύνης $[5.42, 5.92]$ για τη μέση αντοχή θραύσης σκυροδέματος μ , όπου θεωρήσαμε άγνωστη διασπορά. Στο Παράδειγμα 3.2, για το ίδιο δείγμα κάναμε δίπλευρο έλεγχο για τη μηδενική υπόθεση $H_0 : \mu = \mu_0 = 6$ και την απορρίψαμε σε επίπεδο σημαντικότητας $\alpha = 0.05$. Μάλιστα θα απορρίπταμε την H_0 για οποιαδήποτε τιμή μ_0 μεγαλύτερη του 5.92 (και μικρότερη του 5.42 αν μας ενδιέφερε να ελέγξουμε για μικρές τιμές αντοχής θραύσης σκυροδέματος).

Το παραπάνω παράδειγμα φανερώνει ότι το διάστημα εμπιστοσύνης δίνει περισσότερη πληροφορία από τον έλεγχο. Ο έλεγχος μας δίνει μια απόφαση τύπου 'ναι / όχι' (απόρριψη ή μη απόρριψη) για το αν μια τιμή μπορεί να είναι πιθανή για μια παράμετρο θ , ενώ το διάστημα εμπιστοσύνης μας δίνει ένα σύνολο πιθανών τιμών για τη θ , και φυσικά το συμπληρωματικό σύνολο απίθανων τιμών. Για παράδειγμα για δύο σκυροδέματα τύπου Α και Β, μας ενδιαφέρει να δούμε αν έχουν την ίδια μέση αντοχή θραύσης κι αν όχι να εκτιμήσουμε πόση είναι η διαφορά τους. Ο έλεγχος της $H_0 : \mu_1 - \mu_2 = 0$ μπορεί να δώσει απάντηση στο πρώτο ερώτημα αλλά όχι στο δεύτερο (αν απορρίψουμε την H_0). Το διάστημα εμπιστοσύνης όμως απαντάει ταυτόχρονα και στα δύο ερωτήματα. Αν η διαφορά δεν είναι σημαντική, τότε το διάστημα εμπιστοσύνης της διαφοράς $\mu_1 - \mu_2$ θα περιέχει το 0. Αν δεν το περιέχει και το διάστημα είναι θετικό τότε εκτιμούμε ότι η μέση αντοχής θραύσης μ_1 του σκυροδέματος τύπου Α είναι μεγαλύτερη από τη μ_2 του σκυροδέματος τύπου Β, κατά ένα μέγεθος που κυμαίνεται μεταξύ των άκρων του διαστήματος εμπιστοσύνης. Αντίστοιχα αν το διάστημα εμπιστοσύνης είναι αρνητικό εκτιμούμε ότι $\mu_1 < \mu_2$ κατά ένα μέγεθος που κυμαίνεται μεταξύ των άκρων του διαστήματος εμπιστοσύνης.

Σημείωση Η συμφωνία διαστήματος εμπιστοσύνης και δίπλευρου ελέγχου δεν ισχύει στην περίπτωση που η παράμετρος είναι η αναλογία p για τον εξής λόγο. Η διασπορά της εκτιμήτριας \hat{p} για το διάστημα εμπιστοσύνης θεωρείται άγνωστη και εκτιμάται από το δείγμα ως $\frac{\hat{p}(1-\hat{p})}{n}$ (όπου \hat{p} είναι η εκτιμώμενη τιμή της αναλογίας από το δείγμα), ενώ για το δίπλευρο έλεγχο θεωρείται γνωστή κάτω από τη μηδενική υπόθεση $H_0 : p = p_0$ και είναι $\frac{p_0(1-p_0)}{n}$. Έτσι οι υπολογισμοί για το διάστημα εμπιστοσύνης και για το δίπλευρο έλεγχο γίνονται για διαφορετικές τιμές της διασποράς της εκτιμήτριας \hat{p} και γι αυτό τα αποτελέσματα δεν είναι ισοδύναμα.

Κεφάλαιο 4

ΣΥΣΧΕΤΙΣΗ ΚΑΙ ΠΑΛΙΝΔΡΟΜΗΣΗ

Στα προηγούμενα κεφάλαια ορίσαμε και μελετήσαμε την τ.μ. με τη βοήθεια της πιθανοθεωρίας (κατανομή, ροπές) και της στατιστικής (εκτίμηση, στατιστική υπόθεση). Σ' αυτό το κεφάλαιο θα μελετήσουμε πως μια τ.μ. μεταβάλλεται όταν αλλάζει μια άλλη μεταβλητή (τυχαία ή μη).

Πρώτα θα μελετήσουμε τη σχέση δύο τ.μ. X και Y . Συχνά στη μελέτη ενός τεχνικού συστήματος ή φυσικού φαινομένου ενδιαφερόμαστε να προσδιορίσουμε τη σχέση μεταξύ δύο μεταβλητών. Για παράδειγμα στην παρασκευή σκυροδέματος μας ενδιαφέρει η σχέση της αντοχής συμπίεσης και της αντοχής κάμψης. Θα προσδιορίσουμε και θα εκτιμήσουμε το συντελεστή συσχέτισης που μετράει τη γραμμική συσχέτιση δύο τ.μ..

Στη συνέχεια θα μελετήσουμε τη συναρτησιακή σχέση εξάρτησης μιας τ.μ. Y ως προς μια άλλη μεταβλητή X . Η σχέση αυτή είναι πιθανοκρατική κι ορίζεται με την κατανομή της Y για κάθε τιμή της X . Για παράδειγμα η διατμητική αντοχή του αργίλου είναι τ.μ. με κατανομή που μεταβάλλεται με το βάθος τους εδάφους. Συνήθως η μεταβολή αφορά μόνο τη μέση τιμή (και σπανιότερα και τη διασπορά), γι αυτό κι η περιγραφή της κατανομής της Y ως προς τη X περιορίζεται στη δεσμευμένη μέση τιμή $E(Y|X)$ και γίνεται με τη λεγόμενη ανάλυση παλινδρόμησης. Θα μελετήσουμε την απλή γραμμική παλινδρόμηση, δηλαδή θα περιοριστούμε να εκτιμήσουμε τη γραμμική σχέση για τη μέση τιμή $E(Y|X)$ ως προς μια τ.μ. X .

4.1 Συσχέτιση δύο τ.μ.

Δύο τ.μ. X και Y μπορεί να συσχετίζονται με κάποιο τρόπο. Αυτό συμβαίνει όταν επηρεάζει η μία την άλλη, ή αν δεν αλληλοεπηρεάζονται όταν επηρεάζονται και οι δύο από μια άλλη μεταβλητή. Για παράδειγμα η αντοχή συμπίεσης σκυροδέματος και η περιεκτικότητα του σε τσιμέντο μπορούν να θεωρηθούν σαν δύο τ.μ. που συσχετίζονται, όπου η αντοχή συμπίεσης εξαρτάται από την περιεκτικότητα σε τσιμέντο (το αντίθετο δεν έχει πρακτική σημασία). Μπορούμε επίσης να θεωρήσουμε τη συσχέτιση της αντοχής συμπίεσης και της αντοχής κάμψης σκυροδέματος, αλλά τώρα δεν εξαρτάται η μια από την άλλη παρά εξαρτιούνται κι οι δύο από τον τρόπο παρασκευής του σκυροδέματος, π.χ. από την περιεκτικότητα σε τσιμέντο.

Στη συνέχεια θα θεωρήσουμε ότι οι δύο τ.μ. X και Y είναι συνεχείς. Για διακριτές τ.μ. μπορούμε πάλι να ορίσουμε μέτρο συσχέτισης τους αλλά δε θα μας απασχολήσει εδώ.

4.1.1 Ο συντελεστής συσχέτισης ρ

Ο βαθμός της γραμμικής συσχέτισης δύο τ.μ. X και Y με διασπορά σ_X^2 και σ_Y^2 αντίστοιχα και συνδιασπορά $\sigma_{XY} = \text{Cov}(X, Y) = E(X, Y) - E(X)E(Y)$, μετριέται με τον **συντελεστή συσχέτισης** (correlation coefficient) ρ που ορίζεται ως

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}. \quad (4.1)$$

Ο συντελεστής συσχέτισης ρ , όπως και η συνδιασπορά σ_{XY} , εκφράζει το βαθμό και τον τρόπο που οι δύο μεταβλητές συσχετίζονται, δηλαδή πως η μία τ.μ. μεταβάλλεται ως προς την άλλη. Η σ_{XY} παίρνει τιμές που εξαρτώνται από το πεδίο τιμών των X και Y ενώ ο συντελεστής ρ παίρνει τιμές στο διάστημα $[-1, 1]$. Οι χαρακτηριστικές τιμές του ρ ερμηνεύονται ως εξής:

- $\rho = 1$: υπάρχει *τέλεια θετική* συσχέτιση μεταξύ των X και Y ,
- $\rho = 0$: δεν υπάρχει καμιά (γραμμική) συσχέτιση μεταξύ των X και Y ,
- $\rho = -1$: υπάρχει *τέλεια αρνητική* συσχέτιση μεταξύ των X και Y .

Όταν $\rho = \pm 1$ η σχέση είναι αιτιοκρατική κι όχι πιθανοκρατική γιατί γνωρίζοντας την τιμή της μιας τ.μ. γνωρίζουμε και την τιμή της άλλης τ.μ. ακριβώς. Όταν ο συντελεστής συσχέτισης είναι κοντά στο -1 ή 1 η γραμμική συσχέτιση των δύο τ.μ. είναι ισχυρή (συνήθως χαρακτηρίζουμε ισχυρές τις συσχετίσεις όταν $|\rho| > 0.9$) ενώ όταν είναι κοντά στο 0 οι τ.μ. είναι πρακτικά ασυσχέτιστες.

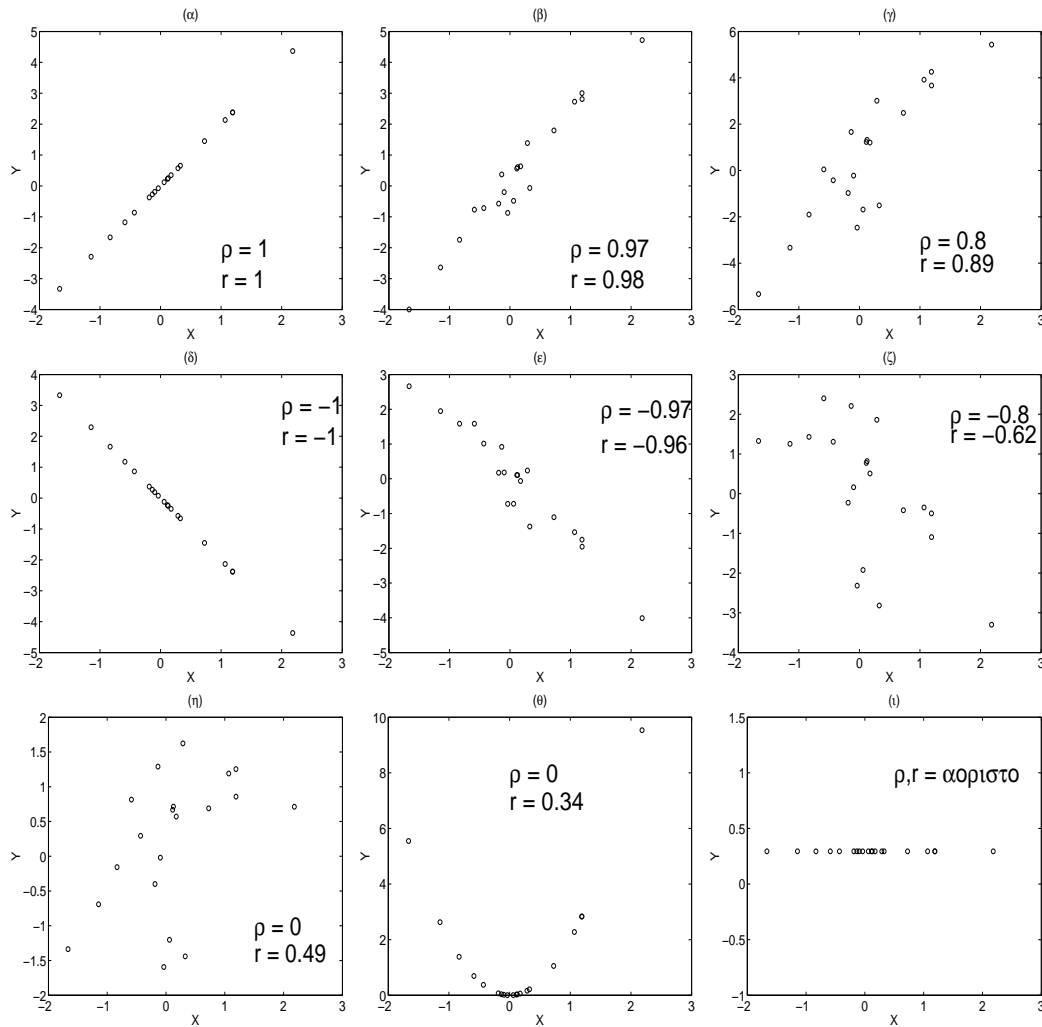
Όπως φαίνεται από τον ορισμό στη σχέση (4.1), ο συντελεστής συσχέτισης ρ δεν εξαρτάται από τη μονάδα μέτρησης των X και Y και είναι συμμετρικός ως προς τις X και Y .

4.1.2 Σημειακή εκτίμηση του συντελεστή συσχέτισης

Όταν έχουμε παρατηρήσεις των δύο τ.μ. X και Y κατά ζεύγη $(x_1, y_1), \dots, (x_n, y_n)$, μπορούμε να εκτιμήσουμε τη συσχέτιση τους ποιοτικά από το **διάγραμμα διασποράς** (scatter diagram), που είναι η απεικόνιση των σημείων (x_i, y_i) , $i = 1, \dots, n$, σε καρτεσιανό σύστημα συντεταγμένων. Στο Σχήμα 4.1 παρουσιάζονται τυπικά διαγράμματα διασποράς για ισχυρές κι ασθενείς συσχετίσεις δύο τ.μ. X και Y . Στα Σχήματα 4.1α και 4.1δ η σχέση είναι τέλεια ($\rho = 1$ και $\rho = -1$ αντίστοιχα), στα Σχήματα 4.1β και 4.1ε είναι ισχυρή (θετική με $\rho = 0.97$ κι αρνητική με $\rho = -0.97$ αντίστοιχα) και στα Σχήματα 4.1γ και 4.1ζ είναι λιγότερο ισχυρή (θετική με $\rho = 0.8$ κι αρνητική με $\rho = -0.8$ αντίστοιχα). Στο Σχήμα 4.1η είναι $\rho = 0$ γιατί οι τ.μ. X και Y είναι ανεξάρτητες ενώ στο Σχήμα 4.1θ είναι πάλι $\rho = 0$ αλλά οι X και Y δεν είναι ανεξάρτητες αλλά συσχετίζονται μόνο μη-γραμμικά. Τέλος για το Σχήμα 4.1ι ο συντελεστής συσχέτισης δεν ορίζεται γιατί η Y είναι σταθερή ($\sigma_Y = 0$ στον ορισμό του ρ στην (4.1)).

Η σημειακή εκτίμηση του συντελεστή συσχέτισης ρ του πληθυσμού από το δείγμα των n ζευγαρωτών παρατηρήσεων των X και Y γίνεται με την αντικατάσταση στη σχέση (4.1) της συνδιασποράς σ_{XY} και των διασπορών σ_X^2 και σ_Y^2 από τις αντίστοιχες εκτιμήσεις από το δείγμα

$$\hat{\rho} \equiv r = \frac{s_{XY}}{s_X s_Y}. \quad (4.2)$$



Σχήμα 4.1: Διάγραμμα διασποράς δύο τ.μ. X και Y από $n = 20$ παρατηρήσεις για θετική συσχέτιση στα σχήματα (α), (β) και (γ), για αρνητική συσχέτιση στα σχήματα (δ), (ε) και (ζ) και για ασυσχέτιστες τ.μ. στα σχήματα (η), (θ) και (ι). Σε κάθε σχήμα δίνεται η πραγματική τιμή του συντελεστή συσχέτισης ρ κι η δειγματική r . Στο (ι) ο συντελεστής συσχέτισης δεν ορίζεται.

Οι αμερόληπτες εκτιμήτριες s_{XY} , s_X^2 και s_Y^2 δίνονται ως

$$s_{XY} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \right) \quad (4.3)$$

$$s_X^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \quad (4.4)$$

όπου \bar{x} και \bar{y} είναι οι δειγματικές μέσες τιμές των X και Y . Από τα παραπάνω προκύπτει η

έκφραση της εκτιμήτρια r

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - n \bar{x}^2\right) \left(\sum_{i=1}^n y_i^2 - n \bar{y}^2\right)}}. \quad (4.5)$$

Είναι προφανές πως η παραπάνω σχέση για το r δεν αλλάζει αν θεωρήσουμε τις μεροληπτικές εκτιμήτριες των σ_{XY} , σ_X και σ_Y .

Στο Σχήμα 4.1 δίνεται ο δειγματικός συντελεστής συσχέτισης r για κάθε περίπτωση. Επειδή το δείγμα είναι μικρό ($n = 20$) η τιμή του r δεν είναι πάντα κοντά στην πραγματική τιμή ρ . Αυτό συμβαίνει γιατί η εκτιμήτρια r όπως δίνεται στη σχέση (4.5) είναι μια τ.μ. που εξαρτάται από τις τιμές και το πλήθος των ζευγών των παρατηρήσεων.

Για να εκφράσουμε τη συσχέτιση δύο τ.μ. χρησιμοποιούμε επίσης την ποσότητα r^2 που λέγεται και **συντελεστής προσδιορισμού** (coefficient of determination) (εκφράζεται συνήθως σε ποσοστό %, $100r^2$). Ο συντελεστής προσδιορισμού δίνει το ποσοστό μεταβλητότητας των τιμών της Y που υπολογίζεται από τη X (κι αντίστροφα) κι είναι ένας χρήσιμος τρόπος να συνοψίσουμε τη συσχέτιση δύο τ.μ..

Παράδειγμα 4.1 Θέλουμε να διερευνήσουμε τη συσχέτιση της αντοχής συμπίεσης κι αντοχής κάμψης σκυροδέματος κάποιας παρασκευής. Γι αυτό πήραμε τις αντίστοιχες μετρήσεις από δείγμα 20 δοκιμών σκυροδέματος που παρουσιάζονται στον Πίνακα 4.1. Για να βρούμε τον συντελεστή συσχέτισης r υπολογίζουμε πρώτα τα παρακάτω

$$\begin{aligned} \bar{x} &= 3092.9 & \bar{y} &= 428.5 \\ \sum_{i=1}^{20} x_i^2 &= 200680560 & \sum_{i=1}^{20} y_i^2 &= 4083287 & \sum_{i=1}^{20} x_i y_i &= 28088503. \end{aligned}$$

Αντικαθιστώντας τα παραπάνω αποτελέσματα στη σχέση (4.5) βρίσκουμε

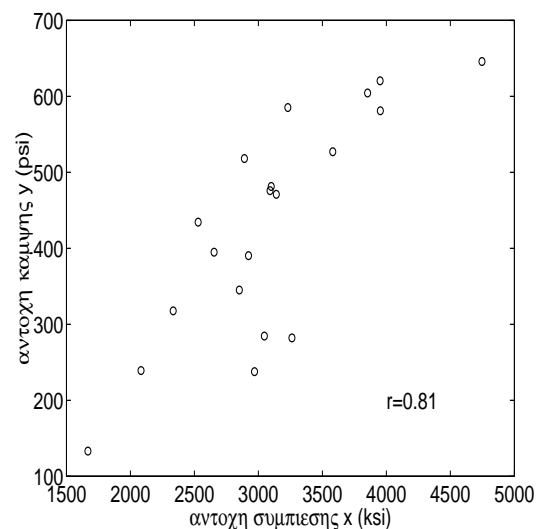
$$r = \frac{28088503 - 20 \cdot 3092.9 \cdot 428.5}{\sqrt{(200680560 - 20 \cdot 3092.9^2)(4083287 - 20 \cdot 428.5^2)}} = 0.807.$$

Η τιμή $r = 0.807$ υποδηλώνει ότι η αντοχή συμπίεσης κι η αντοχή κάμψης έχουν γραμμική θετική συσχέτιση αλλιά όχι πολύ ισχυρή. Αυτό φαίνεται κι από το διάγραμμα διασποράς στο Σχήμα 4.2. Η μεταβλητότητα της μιας τ.μ. (αντοχή συμπίεσης ή αντοχή κάμψης) μπορεί να εξηγηθεί από τη συσχέτιση της με την άλλη κατά ποσοστό που δίνεται από το συντελεστή προσδιορισμού, που είναι $r^2 \cdot 100 = 0.807^2 \cdot 100 = 65\%$. Συμπεραίνουμε λοιπόν πως η γνώση της μιας τ.μ. δε μας επιτρέπει να προσδιορίσουμε την άλλη με ακρίβεια. Πρέπει επίσης να σημειωθεί ότι η εκτίμηση r του συντελεστή συσχέτισης μπορεί να αλλιάξει σημαντικά με την πρόσθεση ή αφαίρεση λίγων ζευγών παρατηρήσεων γιατί το μέγεθος του δείγματος είναι μικρό.

Για τον συντελεστή συσχέτισης ρ μπορούμε να υπολογίσουμε διαστήματα εμπιστοσύνης αν θεωρήσουμε γνωστή την κατανομή της εκτιμήτριας r . Μπορούμε επίσης να κάνουμε στατιστικό έλεγχο για κάποια τιμή r_0 , αλλά εδώ δε θα ασχοληθούμε με αυτά τα θέματα.

A/A	Αντοχή συμπίεσης x_i (ksi)	Αντοχή κάμψης y_i (psi)
1	1668	133
2	2083	239
3	2334	318
4	2529	434
5	2654	395
6	2851	345
7	2891	518
8	2923	390
9	2970	238
10	3047	284
11	3091	476
12	3100	481
13	3140	471
14	3230	585
15	3262	282
16	3581	527
17	3853	604
18	3951	621
19	3953	581
20	4747	646

Πίνακας 4.1: Δεδομένα αντοχής συμπίεσης (x_i) κι αντοχής κάμψης (y_i) σκυροδέματος.



Σχήμα 4.2: Διάγραμμα διασποράς για το δείγμα παρατηρήσεων αντοχής συμπίεσης κι αντοχής κάμψης σκυροδέματος του Πίνακα 4.1.

4.2 Απλή Γραμμική Παλινδρόμηση

Στη *συσχέτιση* που μελετήσαμε παραπάνω μετρήσαμε με το συντελεστή συσχέτισης τη γραμμική συσχέτιση δύο τ.μ. X και Y . Στην *παλινδρόμηση* που θα μελετήσουμε τώρα σχεδιάζουμε

την εξάρτηση μιας τ.μ. Y , που την ονομάζουμε **εξαρτημένη μεταβλητή** (dependent variable), από κάποια άλλη μεταβλητή X που την ονομάζουμε **ανεξάρτητη μεταβλητή** (independent variable). Η ανεξάρτητη μεταβλητή δε θεωρείται τυχαία αλλά παίρνει καθορισμένες τιμές που διαλέγουμε εμείς ή δίνονται από το πρόβλημα που μελετάμε. Ενώ λοιπόν η συσχέτιση είναι συμμετρική ως προς τα X και Y , στην παλινδρόμηση η εξαρτημένη μεταβλητή Y 'καθοδηγείται' από την ανεξάρτητη μεταβλητή X . Γι αυτό και στην ανάλυση που κάνουμε παίζει ρόλο ποιόν από τους δύο παράγοντες που μετράμε ορίζουμε σαν ανεξάρτητη μεταβλητή και ποιόν σαν εξαρτημένη, όταν αυτές δεν ορίζονται ξεκάθαρα από το πρόβλημα. Για παράδειγμα, όταν μετράμε τη διατμητική αντοχή του αργίλου σε διάφορα βάθη, θέλουμε να μελετήσουμε την (γραμμική) εξάρτηση της διατμητικής αντοχής του αργίλου από το βάθος τους εδάφους και γι αυτό η διατμητική αντοχή του αργίλου είναι η εξαρτημένη μεταβλητή Y και το βάθος του εδάφους η ανεξάρτητη μεταβλητή X .

4.2.1 Το πρόβλημα της γραμμικής παλινδρόμησης

Η εξαρτημένη τ.μ. Y ακολουθεί κάποια κατανομή με αθροιστική συνάρτηση κατανομής $F_Y(y|X = x)$, δεσμευμένη για κάθε τιμή x της μεταβλητής X . Περιορίζουμε τη μελέτη του προβλήματος στη μέση τιμή και υποθέτουμε εδώ ότι η εξάρτηση εκφράζεται από μια γραμμική σχέση

$$E(Y|X = x) = \alpha + \beta x \quad (4.6)$$

και η σχέση αυτή λέγεται **γραμμική παλινδρόμηση της Y στη X** (linear regression) [οι τιμές της Y για κάθε τιμή $X = x$ παλινδρομούνται γύρω από το σημείο $y = E(Y|X = x)$ της ευθείας $y = \alpha + \beta x$, δηλαδή οι τιμές της Y για κάθε τιμή της X βρίσκονται πάνω και κάτω από αυτήν την ευθεία]. Το πρόβλημα της παλινδρόμησης είναι η εύρεση των παραμέτρων α και β που εκφράζουν καλύτερα τη γραμμική εξάρτηση της Y από τη X . Κάθε ζεύγος τιμών (α, β) καθορίζει μια διαφορετική γραμμική σχέση που εκφράζεται γεωμετρικά από ευθεία γραμμή και οι δύο παράμετροι ορίζονται ως:

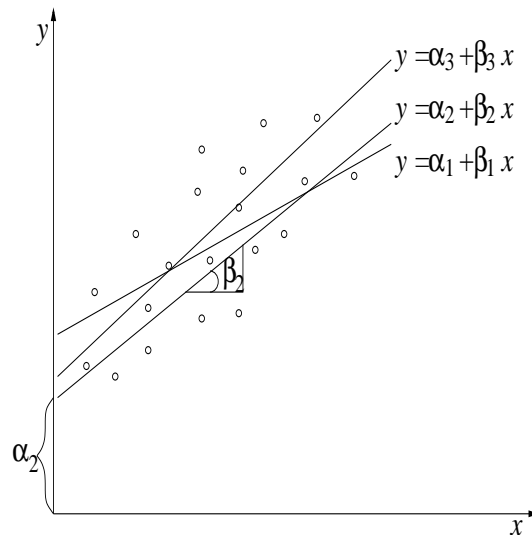
- Ο σταθερός όρος α είναι η τιμή του y για $x = 0$ (intercept).
- Ο συντελεστής β του x είναι η κλίση (slope) της ευθείας ή αλλιώς ο συντελεστής παλινδρόμησης (regression coefficient).

Αν θεωρήσουμε τις παρατηρήσεις $(x_1, y_1), \dots, (x_n, y_n)$ και το διάγραμμα διασποράς που τις απεικονίζει σαν σημεία, μπορούμε να σχηματίσουμε πολλές τέτοιες ευθείες που προσεγγίζουν την υποτιθέμενη γραμμική εξάρτηση της $E(Y|X = x)$ ως προς X , όπως φαίνεται στο Σχήμα 4.3.

Για κάποια τιμή x_i της X αντιστοιχούν διαφορετικές τιμές y_i της Y , σύμφωνα με κάποια κατανομή πιθανότητας $F_Y(y_i|X = x_i)$, δηλαδή μπορούμε να θεωρήσουμε την y_i σαν τ.μ. [θα ήταν σωστότερο να χρησιμοποιούσαμε το συμβολισμό Y_i αντί y_i , όπου ο δείκτης i ορίζει την εξάρτηση από το $X = x_i$, αλλά θα χρησιμοποιήσουμε εδώ τον ίδιο συμβολισμό y_i για την τ.μ. και την παρατήρηση]. Η τ.μ. y_i για κάποια τιμή x_i της X θα δίνεται κάτω από την υπόθεση της γραμμικής παλινδρόμησης ως

$$y_i = \alpha + \beta x_i + \epsilon_i, \quad (4.7)$$

όπου ϵ_i είναι κι αυτή τ.μ., λέγεται **σφάλμα παλινδρόμησης** (regression error) κι ορίζεται ως η διαφορά της y_i από τη δεσμευμένη μέση τιμή $E(Y|X = x_i)$ (δες σχέση (4.6)).



Σχήμα 4.3: Ευθείες γραμμικής παλινδρόμησης

Για την ανάλυση της γραμμικής παλινδρόμησης κάνουμε τις παρακάτω υποθέσεις:

- Η μεταβλητή X είναι *ελεγχόμενη* για το πρόβλημα που μελετάμε, δηλαδή γνωρίζουμε τις τιμές της χωρίς καμιά αμφιβολία.
- Η σχέση (4.6) ισχύει, δηλαδή η εξάρτηση της Y από τη X είναι γραμμική.
- $E(\epsilon_i) = 0$ και $\text{Var}(\epsilon_i) = \sigma_\epsilon^2$ για κάθε τιμή x_i της X , δηλαδή το σφάλμα παλινδρόμησης έχει μέση τιμή μηδέν για κάθε τιμή της X και η διασπορά του είναι σταθερή και δεν εξαρτάται από τη X .

Η τελευταία συνθήκη είναι ισοδύναμη με τη συνθήκη $\text{Var}(Y|X = x) = \sigma_{Y|X}^2$, δηλαδή ότι η διασπορά της εξαρτημένης μεταβλητής Y είναι η ίδια για κάθε τιμή της X και μάλιστα είναι $\sigma_{Y|X}^2 = \sigma_\epsilon^2 \equiv \sigma^2$, όπως προκύπτει από τη σχέση (4.7), αφού οι παράμετροι α και β είναι σταθερές και το x_i γνωστό. Η ιδιότητα αυτή λέγεται *ομοσκεδαστικότητα* (δες επίσης 2.2.4) και αντίθετα έχουμε *ετεροσκεδαστικότητα* όταν η διασπορά της Y (ή του σφάλματος ϵ) μεταβάλλεται με τη X .

Γενικά για να εκτιμήσουμε τις παραμέτρους της γραμμικής παλινδρόμησης με τη μέθοδο των ελαχίστων τετραγώνων, όπως θα δούμε παρακάτω, δεν είναι απαραίτητο να υποθέσουμε κάποια συγκεκριμένη δεσμευμένη κατανομή $F_Y(y_i|X = x_i)$ της Y ως προς τη X . Αν θέλουμε όμως να υπολογίσουμε διαστήματα εμπιστοσύνης για τις παραμέτρους θα χρειαστούμε να υποθέσουμε κανονική δεσμευμένη κατανομή για τη Y . Επίσης οι παραπάνω υποθέσεις για γραμμική σχέση και σταθερή διασπορά αποτελούν χαρακτηριστικά πληθυσμών με κανονική κατανομή. Συνήθως λοιπόν σε προβλήματα γραμμικής παλινδρόμησης υποθέτουμε ότι η δεσμευμένη κατανομή της Y είναι κανονική

$$Y|X = x \sim N(\alpha + \beta x, \sigma^2).$$

4.2.2 Σημειακή εκτίμηση των παραμέτρων της γραμμικής παλινδρόμησης

Λύση στο πρόβλημα της γραμμικής παλινδρόμησης με τις υποθέσεις που ορίστηκαν παραπάνω αποτελεί ο προσδιορισμός της σταθερού όρου της παλινδρόμησης α και του συντελεστή της παλινδρόμησης β για να γνωρίζουμε την ευθεία της παλινδρόμησης αλλά και της διασποράς σ^2 για να γνωρίζουμε το βαθμό μεταβλητότητας γύρω από την ευθεία.

Εκτίμηση των παραμέτρων της ευθείας παλινδρόμησης Η εκτίμηση των παραμέτρων α και β γίνεται με τη μέθοδο των **ελαχίστων τετραγώνων** (method of least squares). Η μέθοδος λέγεται έτσι γιατί βρίσκει την ευθεία παλινδρόμησης με παραμέτρους a και b έτσι ώστε το άθροισμα των τετραγώνων των κατακόρυφων αποστάσεων των σημείων από την ευθεία να είναι το ελάχιστο. Οι εκτιμήσεις των α και β δίνονται από την ελαχιστοποίηση του αθροίσματος των τετραγώνων των σφαλμάτων

$$\min_{\alpha, \beta} \sum_{i=1}^n \epsilon_i^2 \quad \text{ή} \quad \min_{\alpha, \beta} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2. \quad (4.8)$$

Για να λύσουμε αυτό το πρόβλημα θέτουμε τις μερικές παραγώγους ως προς τα α και β ίσες με το μηδέν και καταλήγουμε στο σύστημα δύο εξισώσεων με δύο αγνώστους

$$\left. \begin{aligned} \frac{\partial \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2}{\partial \alpha} = 0 \\ \frac{\partial \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2}{\partial \beta} = 0 \end{aligned} \right\} \begin{aligned} \sum_{i=1}^n y_i &= n\alpha + \beta \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i y_i &= \alpha \sum_{i=1}^n x_i + \beta \sum_{i=1}^n x_i^2 \end{aligned}$$

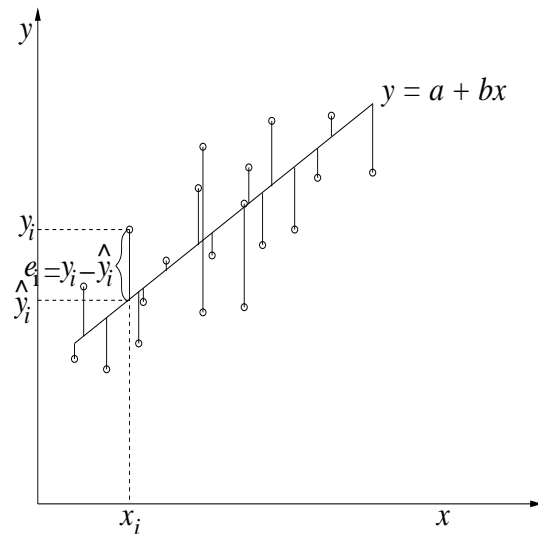
από το οποίο παίρνουμε τις εκτιμήσεις των α και β

$$b = \frac{s_{XY}}{s_X^2}, \quad a = \bar{y} - b\bar{x}, \quad (4.9)$$

όπου s_{XY} και s_X^2 είναι η δειγματική συνδιασπορά των X και Y και η δειγματική διασπορά της X που ορίστηκαν στις σχέσεις (4.3) και (4.4) αντίστοιχα. Τα a και b ορίζουν την ευθεία, $\hat{y} = a + bx$, που λέγεται κι **ευθεία ελαχίστων τετραγώνων**.

Εκτίμηση της διασποράς των σφαλμάτων παλινδρόμησης Για κάθε δοθείσα τιμή x_i με τη βοήθεια της ευθείας ελαχίστων τετραγώνων εκτιμούμε την τιμή \hat{y}_i που γενικά είναι διαφορετική από την πραγματική τιμή y_i . Η διαφορά $e_i = y_i - \hat{y}_i$ είναι η κατακόρυφη απόσταση της πραγματικής τιμής από την ευθεία ελαχίστων τετραγώνων και λέγεται σφάλμα ελαχίστων τετραγώνων ή απλά **υπόλοιπο** (residual). Στο Σχήμα 4.4 απεικονίζονται τα υπόλοιπα της παλινδρόμησης.

Το υπόλοιπο e_i είναι η εκτίμηση του σφάλματος παλινδρόμησης ϵ_i αντικαθιστώντας απλά τις παραμέτρους παλινδρόμησης α και β με τις εκτιμήσεις ελαχίστων τετραγώνων a και b στον ορισμό του σφάλματος $\epsilon_i = y_i - \alpha - \beta x_i$. Άρα η εκτίμηση της διασποράς σ^2 του σφάλματος



Σχήμα 4.4: Ευθεία ελαχίστων τετραγώνων και υπόλοιπα

(που είναι κι η δεσμευμένη διασπορά της Y ως προς X) δίνεται από τη δειγματική διασπορά s^2 των υπολοίπων e_i

$$s^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (4.10)$$

όπου διαιρούμε με $n-2$ γιατί από τους βαθμούς ελευθερίας n του μεγέθους του δείγματος αφαιρούμε δύο για τις δύο παραμέτρους που έχουν ήδη εκτιμηθεί. Η δειγματική διασπορά s^2 μπορεί να εκφραστεί ως προς τις δειγματικές διασπορές των X και Y και της συνδιασποράς τους, αν αντικαταστήσουμε τις εκφράσεις των a και b από την (4.9) στην παραπάνω σχέση (όπου θέτουμε $\hat{y}_i = a + bx_i$)

$$s^2 = \frac{n-1}{n-2} \left(s_Y^2 - \frac{s_{XY}^2}{s_X^2} \right) = \frac{n-1}{n-2} (s_Y^2 - b^2 s_X^2) \quad (4.11)$$

όπου και πάλι υποθέτουμε τις αμερόληπτες εκτιμήτριες για τις διασπορές.

Παρατηρήσεις

1. Η ευθεία ελαχίστων τετραγώνων περνάει από το σημείο (\bar{x}, \bar{y}) γιατί

$$a + b\bar{x} = \bar{y} - b\bar{x} + b\bar{x} = \bar{y}.$$

Άρα η ευθεία ελαχίστων τετραγώνων μπορεί επίσης να οριστεί ως

$$y_i - \bar{y} = b(x_i - \bar{x}).$$

2. Η εκτίμηση των α και β με τη μέθοδο των ελαχίστων τετραγώνων δεν προϋποθέτει σταθερή διασπορά και κανονική κατανομή της εξαρτημένης μεταβλητής Y για κάθε τιμή της ανεξάρτητης μεταβλητής X . Όταν όμως ισχύουν οι δύο αυτές συνθήκες οι εκτιμήτριες ελαχίστων τετραγώνων a και b είναι οι εκτιμήτριες μεγίστης πιθανοφάνειας (κι άρα έχουν και τις επιθυμητές ιδιότητες εκτιμητριών).

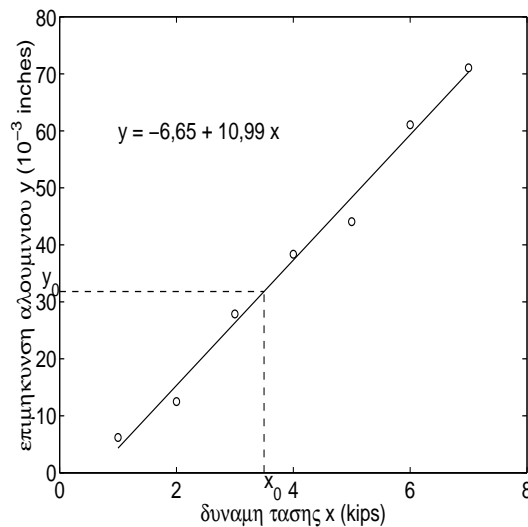
3. Αν η διασπορά της δεσμευμένης κατανομής της Y αλλάζει με τη X , τότε η διαδικασία των ελαχίστων τετραγώνων πρέπει να διορθωθεί έτσι ώστε να δίνει περισσότερο βάρος στις παρατηρήσεις που αντιστοιχούν σε μικρότερη διασπορά.
4. Για κάθε τιμή x_0 της X μπορούμε να προβλέψουμε την αντίστοιχη τιμή y_0 της Y από την ευθεία ελαχίστων τετραγώνων, $y_0 = a + bx_0$. Εδώ πρέπει να προσέξουμε ότι η τιμή x_0 πρέπει να ανήκει στο εύρος τιμών της X που έχουμε από το δείγμα. Για τιμές έξω από αυτό το διάστημα η πρόβλεψη είναι επικίνδυνη. Επίσης οι προβλέψεις έχουν νόημα όταν η διασπορά του σφάλματος είναι σχετικά μικρή.

Παράδειγμα 4.2 *Θέλουμε να μελετήσουμε την αντοχή αλουμινίου και γι αυτό κάναμε ένα πείραμα και μετρήσαμε την επιμήκυνση δοκιμίου αλουμινίου για διάφορες τάσεις. Τα αποτελέσματα δίνονται στον Πίνακα 4.2.*

Δύναμη τάσης x_i (kips)	1	2	3	4	5	6	7
Επιμήκυνση δοκιμίου y_i (10^{-3} inches)	6	13	28	38	44	61	71

Πίνακας 4.2: Δεδομένα αντοχής αλουμινίου (y_i) για διαφορετικές τάσεις (x_i).

Υποθέτουμε πως η παραμόρφωση του αλουμινίου εξαρτάται γραμμικά από την τάση και το διάγραμμα διασποράς στο Σχήμα 4.5 από το δείγμα επιβεβαιώνει αυτήν την υπόθεση. Για να



Σχήμα 4.5: Διάγραμμα διασποράς για τα δεδομένα του Πίνακα 4.2 κι ευθεία ελαχίστων τετραγώνων.

εκτιμήσουμε τις παραμέτρους a και b της ευθείας ελαχίστων τετραγώνων υπολογίζουμε πρώτα τα παρακάτω

$$\bar{x} = 4 \qquad \bar{y} = 37.31$$

$$\sum_{i=1}^7 x_i^2 = 140 \qquad \sum_{i=1}^7 y_i^2 = 13165 \qquad \sum_{i=1}^7 x_i y_i = 1352.5$$

και χρησιμοποιώντας τις σχέσεις (4.3) και (4.4) για τη δειγματική συνδιασπορά και διασπορά αντίστοιχα, βρίσκουμε

$$s_{XY} = 51.32 \quad s_X^2 = 4.67 \quad s_Y^2 = 570.50.$$

Οι εκτιμήσεις b και a είναι

$$b = \frac{51.32}{4.67} = 10.99$$

$$a = 37.31 - 10.99 \cdot 4 = -6.65.$$

Από τη σχέση (4.11) υπολογίζουμε την εκτίμηση διασποράς των σφαλμάτων παλινδρόμησης

$$s^2 = \frac{6}{5}(570.50 - 10.99^2 \cdot 4.67) = 7.75.$$

Τα αποτελέσματα αυτά ερμηνεύονται ως εξής:

1. $b = 10.99$: Για αύξηση της δύναμης τάσης κατά μία μονάδα μέτρησης (1 kips) το δοκίμιο αλουμινίου επιμηκύνεται κατά περίπου 0.011 ίντσες ($10.99 \cdot 10^{-3}$ inches).
2. $a = -6.65$: Όταν δεν ασκείται τάση στο δοκίμιο αλουμινίου ($x = 0$), η επιμήκυνση του δοκιμίου είναι $-6.65 \cdot 10^{-3}$ ίντσες, που φυσικά είναι αδύνατον. Αυτό συμβαίνει γιατί η τιμή $x = 0$ δεν ανήκει στο διάστημα τιμών δύναμης τάσης του πειράματος. Δε θα πρέπει λοιπόν να επιχειρήσουμε προβλέψεις για τιμές δύναμης τάσης μικρότερης του 1 kips και μεγαλύτερης των 7 kips.
3. $s^2 = 7.75$: Η διασπορά γύρω από την ευθεία παλινδρόμησης για κάθε τιμή του X (στο διάστημα τιμών του πειράματος) εκτιμάται να είναι 7.75, ή αλλιώς το τυπικό σφάλμα της εκτίμησης της παλινδρόμησης είναι $2.78 \cdot 10^{-3}$ ίντσες, που είναι μικρό σε σχέση με το επίπεδο τιμών της Y και άρα το μοντέλο της γραμμική παλινδρόμησης εξηγεί ικανοποιητικά τη σχέση επιμήκυνσης αλουμινίου και τάσης και επιτρέπει καλές προβλέψεις.

Με βάση το μοντέλο παλινδρόμησης που εκτιμήσαμε μπορούμε να προβλέψουμε την επιμήκυνση του αλουμινίου για κάθε δύναμη τάσης στο διάστημα [1, 7] kips. Στο Σχήμα 4.5 απεικονίζεται η πρόβλεψη της επιμήκυνσης για δύναμη τάσης $x_0 = 3.5$ kips και είναι

$$y_0 = -6.65 + 10.99 \cdot 3.5 = 31.82.$$

Για τις παραμέτρους α και β , καθώς και για τη διασπορά σ^2 , μπορούμε να εκτιμήσουμε διαστήματα εμπιστοσύνης και να κάνουμε στατιστικούς ελέγχους.

4.2.3 Σχέση του συντελεστή συσχέτισης και παλινδρόμησης

Η παλινδρόμηση ορίζεται θεωρώντας την ανεξάρτητη μεταβλητή X καθορισμένη και την εξαρτημένη μεταβλητή Y τυχαία, ενώ για τη συσχέτιση θεωρούμε και τις δύο μεταβλητές X και Y τυχαίες. Για τις μεταβλητές X και Y της παλινδρόμησης, μπορούμε να αγνοήσουμε ότι η X δεν είναι τ.μ. και να ορίσουμε το συντελεστή συσχέτισης ρ όπως και πριν. Η σχέση μεταξύ του r (της εκτιμήτριας του ρ από το δείγμα) και του συντελεστή της παλινδρόμησης b

(της εκτιμήτριας του β από το δείγμα) δίνεται ως εξής (συνδιάζοντας τις σχέσεις $r = \frac{s_{XY}}{s_X s_Y}$ και $b = \frac{s_{XY}}{s_X^2}$)

$$r = b \frac{s_X}{s_Y} \quad \text{ή} \quad b = r \frac{s_Y}{s_X}. \quad (4.12)$$

Και τα δύο μεγέθη, r και b , εκφράζουν ποιοτικά τη γραμμική συσχέτιση των μεταβλητών X και Y , αλλά το b εξαρτάται από τη μονάδα μέτρησης των X και Y ενώ το r παίρνει τιμές στο διάστημα $[-1, 1]$. Έτσι αν η συσχέτιση είναι θετική ($r > 0$) τότε η κλίση της ευθείας παλινδρόμησης b είναι επίσης θετική, αν η συσχέτιση είναι αρνητική ($r < 0$) τότε είναι $b < 0$ και αν οι μεταβλητές X και Y δε συσχετίζονται ($r = 0$) τότε η ευθεία παλινδρόμησης είναι οριζόντια ($b = 0$).

Επίσης μπορούμε να εκφράσουμε το συντελεστή προσδιορισμού r^2 ως προς τη δειγματική διασπορά του σφάλματος s^2 και αντίστροφα

$$s^2 = \frac{n-1}{n-2} s_Y^2 (1-r^2) \quad \text{ή} \quad r^2 = 1 - \frac{n-2}{n-1} \frac{s^2}{s_Y^2}. \quad (4.13)$$

Η παραπάνω σχέση δηλώνει πως όσο μεγαλύτερο είναι το r^2 (ή το $|r|$) τόσο μικρότερη είναι η διασπορά του σφάλματος της παλινδρόμησης, δηλαδή τόσο καλύτερη είναι η πρόβλεψη που βασίζεται στην ευθεία παλινδρόμησης.

Παράδειγμα 4.3 Στο παραπάνω παράδειγμα, ο συντελεστής συσχέτισης της επιμήκυνσης του αλουμινίου και της τάσης είναι

$$r = \frac{s_{XY}}{s_X s_Y} = \frac{41.32}{\sqrt{4.67 \cdot 570.50}} = 0.995$$

που θα μπορούσαμε να υπολογίσουμε κι από τις σχέσεις (4.12) ή (4.13). Ο συντελεστής συσχέτισης δηλώνει την πολύ ισχυρή θετική συσχέτιση της επιμήκυνσης του αλουμινίου και της τάσης, που δε μπορούμε όμως εύκολα να συμπεράνουμε από την τιμή του συντελεστή παλινδρόμησης $b = 10.99$ ή την τιμή της διασποράς των σφαλμάτων $s^2 = 7.75$ γιατί οι τιμές τους ορίζονται σε σχέση με το πεδίο τιμών της επιμήκυνσης του αλουμινίου και της τάσης.

Σ' αυτό το κεφάλαιο ασχοληθήκαμε μόνο με την απλή γραμμική παλινδρόμηση. Η μελέτη της παλινδρόμησης επεκτείνεται στη μή-γραμμική παλινδρόμηση, που αποτελεί γενικότερη και πιο ρεαλιστική (αλλά και πιο πολύπλοκη) προσέγγιση για τον προσδιορισμό της εξάρτησης της Y από τη X . Επίσης η τ.μ. Y μπορεί να εξαρτάται από περισσότερες από μια μεταβλητές που είναι το πρόβλημα της πολλαπλής παλινδρόμησης (γραμμική ή μή-γραμμική).

ΑΣΚΗΣΕΙΣ**Περιγραφική Στατιστική, Εκτίμηση και Έλεγχος Παραμέτρων**

1. Δείξτε ότι η εκτιμήτρια s^2 της διασποράς σ^2 είναι αμερόληπτη.
2. Δύο τυχαίες μεταβλητές X_1 και X_2 έχουν κοινή διασπορά σ^2 και s_1^2, s_2^2 είναι οι αμερόληπτες δειγματικές διασπορές των X_1 και X_2 , αντίστοιχα, από δείγματα μεγέθους n_1 και n_2 . Δείξτε ότι η εκτιμήτρια $s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$ της σ^2 είναι επίσης αμερόληπτη.
3. Υπολογίστε την εκτιμήτρια της μέγιστης πιθανοφάνειας της παραμέτρου λ της εκθετικής κατανομής $f_X(x) = \lambda e^{-\lambda x}$ από ένα δείγμα της X μεγέθους n . Είναι αυτή η εκτιμήτρια αμερόληπτη;
4. Μετρήθηκε η αντοχή διάτμησης σε κάθε ένα από 10 σημεία συγκόλλησης που χρησιμοποιήθηκαν σε ένα πείραμα και τα αποτελέσματα είναι (σε psi)

392	376	401	367	389	362	409	415	358	375
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

- (α) Υποθέτοντας πως η αντοχή διάτμησης ακολουθεί κανονική κατανομή εκτιμήστε την πραγματική μέση αντοχή διάτμησης και την τυπική απόκλιση της αντοχής διάτμησης με τη μέθοδο της μέγιστης πιθανοφάνειας.
 - (β) Υποθέτουμε πως θέλουμε να ελέγξουμε την αντοχή διάτμησης ενός άλλου σημείου συγκόλλησης. Χρησιμοποιώντας τα παραπάνω δεδομένα βρείτε την πιθανότητα η αντοχή διάτμησης να είναι μικρότερη από 400 psi.
5. Για τον έλεγχο της περιεκτικότητας του χάλυβα σε ραδιενέργεια σε δύο εργοστάσια παραγωγής χάλυβα Α και Β έγιναν οι παρακάτω μετρήσεις ραδιενέργειας σε τυχαία δοκίμια χάλυβα (οι μετρήσεις είναι σε Bq/g):

Δοκίμια	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A (Bq/g)	0.37	0.00	0.54	0.59	0.16	0.86	0.86	0.49	0.60	0.55					
B (Bq/g)	0.24	0.52	0.12	0.95	0.26	0.33	0.62	0.32	0.27	0.05	0.39	0.10	0.51	0.79	0.09

Θεωρείται ότι η διασπορά της ραδιενέργειας στο χάλυβα είναι ίδια για τα δύο εργοστάσια ($\sigma_1^2 = \sigma_2^2 = \sigma^2$).

- (α) Σχηματίστε το θηκόγραμμα για τα δεδομένα ραδιενέργειας στο χάλυβα των δύο δειγμάτων και σχολιάστε αν η ραδιενέργεια στους χάλυβες των δύο εργοστασίων φαίνεται να ακολουθούν κανονική κατανομή. Σχολιάστε επίσης αν φαίνεται να διαφέρουν αυτές οι δύο κατανομές.
- (β) Εκτιμήστε τη μέση ραδιενέργεια στο χάλυβα για το εργοστάσιο Α (σημειακή εκτίμηση και 95% διάστημα εμπιστοσύνης).
- (γ) Κάνετε την ίδια εκτίμηση για το εργοστάσιο Β.
- (δ) Έστω ότι το ανώτατο επιτρεπτό όριο για τη μέση ραδιενέργεια στο χάλυβα είναι 0.5 Bq/g. Με βάση τα παραπάνω διαστήματα εμπιστοσύνης σε επίπεδο εμπιστοσύνης 95% απαντήστε αν θα γινόταν αποδεκτός στην αγορά ο χάλυβας από το εργοστάσιο Α και από το εργοστάσιο Β.

- (ε) Για το παραπάνω ερώτημα κάνετε κατάλληλο στατιστικό έλεγχο στο ίδιο επίπεδο εμπιστοσύνης. Σχολιάστε τις απαντήσεις σας από τον έλεγχο και το διάστημα εμπιστοσύνης.
- (ϛ) Ελέγξτε χρησιμοποιώντας διάστημα εμπιστοσύνης και έλεγχο υπόθεσης σε επίπεδο εμπιστοσύνης 95% αν η μέση ραδιενέργεια στους χάλυβες των δύο εργοστασίων είναι ίδια.

6. Ένας πολιτικός μηχανικός επιβλέπει την ποιότητα του νερού μετρώντας την ποσότητα των αιωρούμενων στερεών (αναλογία σωματιδίων ανά εκατομμύριο) σε δείγμα νερού από ένα ποταμό. Τα αποτελέσματα των μετρήσεων για 10 μέρες είναι

14	25	21	28	30	63	29	63	55	19
----	----	----	----	----	----	----	----	----	----

Από προηγούμενες μετρήσεις σε άλλους ποταμούς είναι γνωστό ότι η τυπική απόκλιση της ποσότητας των αιωρούμενων σωματιδίων στο νερό είναι 20 σωματίδια ανά εκατομμύριο.

- (α) Μπορούμε να δεχτούμε την παραπάνω εμπειρική τιμή της τυπικής απόκλισης των αιωρούμενων στερεών γι αυτόν το ποταμό; Γιατί; Κάνετε τον κατάλληλο στατιστικό έλεγχο για να απαντήσετε αυτό το ερώτημα σε στατιστική σημαντικότητα $\alpha = 0.01$.
- (β) Υπολογίστε τα μέτρα θέσης και μεταβλητότητας για τα δεδομένα του δείγματος και σχηματίστε το κατάλληλο θηκόγραμμα. Σχολιάστε αν φαίνεται η περιεκτικότητα των αιωρούμενων σωματιδίων στο νερό του ποταμού να ακολουθεί κανονική κατανομή.
- (γ) Θεωρώντας γνωστή την τυπική απόκλιση των αιωρούμενων σωματιδίων και ίση με 20, εκτιμήστε τη μέση ποσότητα των αιωρούμενων στερεών στο νερό του ποταμού υπολογίζοντας τη σημειακή εκτίμηση και το 95% διάστημα εμπιστοσύνης.
- (δ) Το νερό του ποταμού θεωρείται κακής ποιότητας όταν η μέση ποσότητα των αιωρούμενων σωματιδίων στο νερό υπερβαίνει την ποσότητα των 50 σωματιδίων (ανά εκατομμύριο). Διατυπώστε την κατάλληλη στατιστική υπόθεση και κάνετε τον έλεγχο για να απαντήσετε σε στατιστική σημαντικότητα $\alpha = 0.05$ αν το νερό του ποταμού θα χαρακτηριζόταν ως κακής ποιότητας με βάση τό δείγμα. Ποιά είναι η απάντηση στο ίδιο ερώτημα σε επίπεδο σημαντικότητας $\alpha = 0.01$;
- (ε) Πόσες επιπλέον παρατηρήσεις πρέπει να κάνουμε για να έχει το 95% διάστημα εμπιστοσύνης της μέσης ποσότητας αιωρούμενων στερεών εύρος 10 (δηλαδή η ακρίβεια της μέσης ποσότητας να δίνεται με ± 5);
7. Έχει μελετηθεί η ρωγματογενής διάβρωση σιδήρου κι ατσάλιού από καυστική τάση γιατί προκαλεί θραύσεις του μετάλλου, για παράδειγμα παρουσιάζονται θραύσεις γύρω από πιρτσίνια (καρφιά) σε χαλύβδινους λέβητες. Σε δοκιμές που έγιναν μετρήθηκε το μήκος ρωγμής (σε μm) ως αποτέλεσμα της σταθερής τάσης διάβρωσης που ασκήθηκε σε λείες ράβδους από κάποιο κράμα χάλυβα για ορισμένο χρονικό διάστημα. Έγιναν οι παρακάτω μετρήσεις μήκους ρωγμής (σε μm) σε 21 δοκίμια :

16.1	9.6	24.9	20.4	12.7	21.2	30.2	25.8	18.5	10.3	25.3
14.0	27.1	45.0	23.3	24.2	14.6	8.9	32.4	11.8	28.5	

- (α) Σχηματίστε το ιστόγραμμα των δεδομένων του μήκους ρωγμής και σχολιάστε αν το μήκος ρωγμής του μετάλλου φαίνεται να ακολουθεί κανονική κατανομή.

- (β) Υπολογίστε το 95% διάστημα εμπιστοσύνης καθώς και το 99% διάστημα εμπιστοσύνης για το μέσο μήκος ρωγμής του μετάλλου με βάση το δείγμα.
- (γ) Το μέταλλο θεωρείται ακατάλληλο όταν το μήκος ρωγμής υπερβαίνει τα 30 μm (επειδή πιστεύεται ότι τότε μπορεί να προκληθεί θραύση από τη ρωγματογενή διάβρωση). Με βάση στατιστικό έλεγχο υπόθεσης σε επίπεδο εμπιστοσύνης 95% μπορούμε να χαρακτηρίσουμε αυτό το κράμα χάλυβα ακατάλληλο; Θα άλλαζε η απάντηση αν κάναμε τον ίδιο έλεγχο σε επίπεδο εμπιστοσύνης 99%; Συγκρίνετε τις απαντήσεις με τα διαστήματα εμπιστοσύνης από το ερώτημα 7β'.
- (δ) Θέλουμε να εκτιμήσουμε το μέσο μήκος ρωγμής με μεγάλη ακρίβεια, $\pm 2\mu m$, δηλαδή το 95% διάστημα εμπιστοσύνης του μέσου μήκους ρωγμής να έχει εύρος $4\mu m$. Πόσες ακόμα μετρήσεις πρέπει να κάνουμε;
8. Η φθορά στους αγωγούς ενός αστικού δικτύου αγωγών έχει προκαλέσει σοβαρές ανησυχίες. Η αποκατάσταση ενός αγωγού με φθορές μπορεί να επιτευχθεί εφαρμόζοντας μια ευπροσάρμοστη επένδυση στον αγωγό. Για να δοκιμασθεί μια νέα τεχνική συγκόλλησης της επένδυσης στον αγωγό έγιναν μετρήσεις της αντοχής εφελκυσμού (σε μονάδα psi) σε 18 δοκίμια με επένδυση. Σε 10 από αυτά χρησιμοποιήθηκε η νέα τεχνική συγκόλλησης και στα άλλα 8 δε χρησιμοποιήθηκε. Τα αποτελέσματα είναι:

Με συγκόλληση (μέση τιμή $\bar{x}_1 = 2917.5$, τυπική απόκλιση $s_1 = 272.3$)

2748	2700	2655	2822	2511	3149	3257	3213	3220	2900
------	------	------	------	------	------	------	------	------	------

Χωρίς συγκόλληση (μέση τιμή $\bar{x}_2 = 3082.9$, τυπική απόκλιση $s_2 = 243.7$)

3027	3356	3359	3297	3125	2910	2889	2700
------	------	------	------	------	------	------	------

- (α) Σχηματίζοντας κατάλληλα γραφήματα σχολιάστε αν φαίνεται η αντοχή εφελκυσμού στους αγωγούς με την επένδυση να ακολουθεί κανονική κατανομή (με ή χωρίς συγκόλληση). Από τα γραφήματα φαίνεται να διαφέρουν οι δύο κατανομές;
- (β) Υπολογίστε το 99% διάστημα εμπιστοσύνης για τη μέση αντοχή εφελκυσμού σε κάθε μια από τις δύο περιπτώσεις επένδυσης με βάση το δείγμα.
- (γ) Αν θέλουμε να εκτιμήσουμε τη μέση αντοχή εφελκυσμού στην επένδυση χωρίς συγκόλληση με ακρίβεια ± 80 psi, δηλαδή το 99% διάστημα εμπιστοσύνης του μέσου μήκους ρωγμής να έχει εύρος 160 psi, πόσες ακόμα μετρήσεις πρέπει να κάνουμε;
- (δ) Χρησιμοποιώντας είτε διάστημα εμπιστοσύνης ή έλεγχο υπόθεσης, εξετάστε σε επίπεδο εμπιστοσύνης 95% αν η μέση αντοχή εφελκυσμού διαφέρει με στατιστική σημαντικότητα σε επενδύσεις χωρίς συγκόλληση και σε επενδύσεις με συγκόλληση.
9. Στα πλαίσια μελέτης για τη μόλυνση του αέρα που έγινε σε έναν πειραματικό σταθμό μετρήθηκε η ποσότητα αιωρούμενης οργανικής ύλης διαλυτής στο βενζόλιο (σε μικρογραμμάρια ανά κυβικό μέτρο, $\mu g/m^3$) σε 12 δείγματα αέρα.

1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.9	2.4	2.6	3.0	3.4
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

- (α) Υπολογίστε τη δειγματική διάμεσο, το πρώτο και τρίτο τεταρτομόριο και σχηματίστε το θηκόγραμμα για τα δεδομένα της ποσότητας μολυσματικής ύλης. Με βάση

το θηκόγραμμα σχολιάστε αν η ποσότητα αυτή φαίνεται να ακολουθεί κανονική κατανομή.

Στη συνέχεια θεωρούμε πως η ποσότητα μολυσματικής ύλης στον αέρα ακολουθεί κανονική κατανομή.

- (β) Υπολογίστε το 95% διάστημα εμπιστοσύνης για τη μέση ποσότητα μολυσματικής ύλης στον αέρα της περιοχής του σταθμού με βάση το δείγμα των παραπάνω μετρήσεων.
- (γ) Χρησιμοποιώντας είτε διάστημα εμπιστοσύνης ή έλεγχο υπόθεσης, εξετάστε σε επίπεδο εμπιστοσύνης 99% αν η μέση ποσότητα μολυσματικής ύλης στον αέρα της περιοχής του σταθμού δε ξεπερνάει τα $3 \mu g/m^3$.
- (δ) Χωρίς να κάνετε υπολογισμούς απαντήστε στο ερώτημα θγ' για επίπεδο εμπιστοσύνης 90% και δικαιολογήστε την απάντησή σας.

10. Μια εταιρεία ταξί θέλει να κάνει μια μεγάλη αγορά ελαστικών για το στόλο της και πρέπει να διαλέξει μεταξύ δύο τύπων (δηλαδή μάρκων ελαστικών) Α και Β. Για να εκτιμήσει τη διαφορά στους δύο τύπους έγινε ένα πείραμα χρησιμοποιώντας 12 ελαστικά από κάθε τύπο. Μετρήθηκαν τα χιλιόμετρα που διένυσαν τα ελαστικά μέχρι να φθαρούν. Τα αποτελέσματα για τα δύο δείγματα είναι:

Τύπος Α : μέση τιμή $\bar{x}_1 = 36300\text{km}$, τυπική απόκλιση $s_1 = 5000\text{km}$

Τύπος Β : μέση τιμή $\bar{x}_2 = 40100\text{km}$, τυπική απόκλιση $s_2 = 4100\text{km}$

Θεωρούμε ότι η «ζωή» του ελαστικού και των δύο τύπων (Α και Β) ακολουθεί κανονική κατανομή αλλά με άγνωστη για μας διασπορά, που είναι κοινή και για τους δύο τύπους.

- (α) Υπολογίστε το 99% διάστημα εμπιστοσύνης για τη μέση «ζωή» ελαστικού του τύπου Α. Κάνετε το ίδιο για τον τύπο Β.
- (β) Αν θέλουμε να εκτιμήσουμε τη μέση «ζωή» ελαστικού του τύπου Α με μεγαλύτερη ακρίβεια και συγκεκριμένα με ακρίβεια $\pm 500\text{km}$, δηλαδή το 99% διάστημα εμπιστοσύνης της μέσης «ζωής» ελαστικού του τύπου Α να έχει εύρος 1000km , πόσες ακόμα μετρήσεις πρέπει να κάνουμε;
- (γ) Χρησιμοποιώντας έλεγχο υπόθεσης εξετάστε σε επίπεδο σημαντικότητας $\alpha = 0.05$ αν η μέση «ζωή» ελαστικού διαφέρει στους δύο τύπους.

11. Θέλουμε να μελετήσουμε τη διάρκεια ενός συγκεκριμένου υπεραστικού οδικού δρομολογίου. Έγιναν 11 μετρήσεις σε διαφορετικές μέρες και ώρες και βρέθηκαν τα εξής αποτελέσματα για το δείγμα :

μέση τιμή $\bar{x} = 125\text{min}$, τυπική απόκλιση $s = 15\text{min}$

Θεωρούμε ότι η διάρκεια του δρομολογίου ακολουθεί κανονική κατανομή αλλά με άγνωστη για μας διασπορά.

- (α) Υπολογίστε το 95% διάστημα εμπιστοσύνης για τη μέση διάρκεια του δρομολογίου.
- (β) Υπολογίστε το 95% διάστημα εμπιστοσύνης για την τυπική απόκλιση της διάρκειας του δρομολογίου.

- (γ) Από προηγούμενες μελέτες πιστεύεται πως η μέση διάρκεια αυτού του δρομολογίου είναι 120min και η τυπική απόκλιση 5min. Είναι αυτές οι τιμές αποδεκτές σε επίπεδο εμπιστοσύνης 95% με βάση αυτό το δείγμα ;
- (δ) Αν θα θέλαμε να εκτιμήσουμε τη μέση διάρκεια του δρομολογίου με μεγαλύτερη ακρίβεια και συγκεκριμένα με ακρίβεια $\pm 2\text{min}$, δηλαδή το 95% διάστημα εμπιστοσύνης της μέσης διάρκειας του δρομολογίου να έχει εύρος 4min, πόσες ακόμα μετρήσεις πρέπει να κάνουμε ;
12. Σε 144 δοκιμές σε κάποιο εργαστήριο, 48 κατέληξαν σε ανάφλεξη κάποιου συγκεκριμένου τύπου υποστρώματος από αναμμένο τσιγάρο. Έστω p η αναλογία όλων των δοκιμών που καταλήγουν σε ανάφλεξη του υποστρώματος.
- (α) Υπολογίστε το 99% διάστημα εμπιστοσύνης για την αναλογία p .
- (β) Αν θέλουμε το εύρος του 99% διαστήματος εμπιστοσύνης για την αναλογία να είναι 0.1 πόσες ακόμα δοκιμές πρέπει να κάνουμε ;
13. Σε μια μελέτη για ένα καινούριο σύστημα εκτόξευσης ρουκετών μικρού βεληνεκούς έγιναν δοκιμές με το παλιό και το καινούριο σύστημα. Σε δείγμα 60 πειραματικών εκτοξεύσεων με το παλιό σύστημα 44 ήταν πετυχημένες και σε 80 πειραματικές εκτοξεύσεις με το καινούριο σύστημα 72 ήταν πετυχημένες.
- (α) Υπολογίστε τα 95% διαστήματα εμπιστοσύνης για τις αναλογίες πετυχημένων εκτοξεύσεων με τα δύο συστήματα.
- (β) Σε επίπεδο εμπιστοσύνης 90% κάνετε στατιστικό έλεγχο για την υπόθεση ότι οι δύο αναλογίες δε διαφέρουν.
- (γ) Αν στον παραπάνω έλεγχο βρήκατε ότι υπάρχει διαφορά εκτιμείστε πόση είναι αυτή η διαφορά.

Συσχέτιση και Παλινδρόμηση

1. Στον παρακάτω πίνακα δίνεται για 10 σταθμούς ο αριθμός των ημερών σε ένα χρόνο που η θερμοκρασία έπεσε κάτω από 0°C και το υψόμετρο τους.

Υψόμετρο (μ)	1000	1050	1110	1220	1320	1380	1420	1560	1670	1950
Αριθμός ημερών	32	29	36	38	43	53	52	63	73	100

- (α) Σχηματίστε το κατάλληλο διάγραμμα διασποράς, υπολογίστε τον συντελεστή συσχέτισης και σχολιάστε αν φαίνεται ο αριθμός των ημερών να εξαρτάται γραμμικά από το υψόμετρο.
- (β) Υπολογίστε τις σημειακές εκτιμήσεις a και b των παραμέτρων α και β της ευθείας παλινδρόμησης (με τη μέθοδο των ελαχίστων τετραγώνων) και σχηματίστε στο παραπάνω διάγραμμα διασποράς την ευθεία ελαχίστων τετραγώνων.
- (γ) Υπολογίστε τη διασπορά των σφαλμάτων παλινδρόμησης και σχηματίστε στο διάγραμμα διασποράς ενδεικτικά όρια του τυπικού σφάλματος της εκτίμησης της ευθείας ελαχίστων τετραγώνων
- (δ) Με βάση το δείγμα, μπορείτε να εκτιμήσετε το μέσο αριθμό ημερών το χρόνο που η θερμοκρασία πέφτει κάτω από 0°C σε υψόμετρο 1500μ; Σε υψόμετρο 2200μ;
2. Σε περιοχές με υψηλά επίπεδα βροχοπτώσεων είναι σημαντικό να έχει το σκυροδέμα χαμηλό πορώδες, το οποίο μετριέται με το ποσοστό κενού στο υλικό. Ένας παράγοντας που επηρεάζει το πορώδες του σκυροδέματος είναι το βάρος μονάδας του σκυροδέματος που μετριέται σε pcf . Για κάθε ένα από 11 δοκίμια σκυροδέματος που συλλέχτηκαν τυχαία μετρήθηκαν το βάρος μονάδας και το πορώδες και οι παρατηρήσεις δίνονται στον παρακάτω πίνακα.

βάρος μονάδας	99.0	101.1	102.7	105.4	108.7	110.8	112.1	113.6	115.1	115.4	120.0
πορώδες	28.8	27.9	27.0	22.8	20.9	19.6	17.1	16.0	13.0	13.6	10.8

- (α) Υποθέτουμε ότι το πορώδες του σκυροδέματος Y εξαρτάται γραμμικά από το βάρος μονάδας του σκυροδέματος X ($E(Y|X = x) = \alpha + \beta x$). Σχηματίστε το κατάλληλο διάγραμμα διασποράς και σχολιάστε αν αυτή η υπόθεση φαίνεται σωστή με βάση το δείγμα των παρατηρήσεων του πίνακα.
- (β) Υπολογίστε τις σημειακές εκτιμήσεις a και b των παραμέτρων α και β της ευθείας παλινδρόμησης και σχηματίστε την ευθεία αυτή στο διάγραμμα διασποράς.
- (γ) Με βάση το δείγμα, εκτιμήστε (αν γίνεται) το μέσο πορώδες του σκυροδέματος που έχει βάρος μονάδας 110 και το ίδιο για βάρος μονάδας 130. Σχολιάστε με βάση τη διασπορά των σφαλμάτων παλινδρόμησης το τυπικό σφάλμα γύρω από όποια πρόβλεψη κάνετε.
3. Στα πλαίσια μιας έρευνας για τη μόλυνση από τους ρύπους των αυτοκινήτων διερευνήθηκε η σχέση μεταξύ της ροής κίνησης σε κάποιον κύριο αυτοκινητόδρομο (σε χιλιάδες αυτοκινήτων ανά 24 ώρες) και της περιεκτικότητας σε μόλυβδο στον φλοιό των

παρακείμενων δένδρων (σε $\mu g/g$) ξηρού βάρους, δηλαδή μg μολύβδου ανά g ξηρού φλοιού). Οι μετρήσεις της ροής κίνησης και της περιεκτικότητας μολύβδου που έγιναν σε 10 μέρες που επιλέχτηκαν τυχαία δίνονται στον παρακάτω πίνακα.

ροή κίνησης	8.3	8.5	12.0	12.1	17.0	17.8	23.0	24.0	24.3	33.6
περιεκτικότητα	227	312	362	521	640	728	945	738	759	1263

- (α) Σχηματίστε το κατάλληλο διάγραμμα διασποράς και υπολογίστε τον αντίστοιχο συντελεστή συσχέτισης. Σχολιάστε αν υπάρχει σχέση και τι είδους μεταξύ της κίνησης ροής και της περιεκτικότητας σε μόλυβδο καθώς κι αν φαίνεται η περιεκτικότητα σε μόλυβδο να εξαρτάται γραμμικά από την κίνηση ροής. Δικαιολογήστε την απάντησή σας με βάση τον συντελεστή συσχέτισης και το διάγραμμα διασποράς.
- (β) Υπολογίστε τις σημειακές εκτιμήσεις a και b των παραμέτρων α και β της ευθείας παλινδρόμησης (με τη μέθοδο των ελαχίστων τετραγώνων).
- (γ) Με βάση το δείγμα, εκτιμήστε τη μέση περιεκτικότητα σε μόλυβδο στο φλοιό των δένδρων όταν η ροή κίνησης στον αυτοκινητόδρομο σε ένα 24ωρο είναι 20 χιλιάδες. Εξηγήστε αν η εκτίμηση αυτή μπορεί να θεωρηθεί χρήσιμη.
4. Στα πλαίσια μιας μελέτης για την ποιότητα του νερού που απορρέει από εθνικούς δρόμους, έγινε μια έρευνα για τη σχέση της απορροής νερού από ένα δρόμο και της βροχόπτωσης. Μετρήθηκε η ποσότητα νερού από βροχόπτωση (σε m^3), έστω X , κι η ποσότητα απορροής (σε m^3) από το δρόμο, έστω Y , σε μια συγκεκριμένη περιοχή. Τα αποτελέσματα δίνονται στον παρακάτω πίνακα.

X	6	12	15	17	23	30	40	47	53	67	72	74	96	100	125
Y	6	10	20	15	30	30	26	46	38	40	48	53	70	85	92

- (α) Σχηματίστε το κατάλληλο διάγραμμα διασποράς και υπολογίστε τον συντελεστή αυτοσυσχέτισης. Σχολιάστε αν από το διάγραμμα διασποράς φαίνεται σωστή η υπόθεση πως η ποσότητα απορροής νερού εξαρτάται από την ποσότητα βροχόπτωσης.
- (β) Με βάση το δείγμα, υπολογίστε τη σημειακή εκτίμηση (αν γίνεται) της αληθινής μέσης ποσότητας απορροής νερού αυτού του δρόμου όταν η ποσότητα βροχόπτωσης είναι $80m^3$. Κάνετε το ίδιο για ποσότητα βροχόπτωσης $180m^3$.
5. Στα πλαίσια μελέτης για την ρύπανση του αέρα που έγιναν 10 διαφορετικές μετρήσεις της ταχύτητας του αέρα και του συντελεστή εξάτμισης σταγονιδίων καμένων καυσίμων σε μια προωθητική μηχανή. Οι μετρήσεις δίνονται στον παρακάτω πίνακα.

Ταχύτητα αέρα (cm/sec)	20	60	100	140	180	220	260	300	340	380
Εξάτμιση (mm^2/sec)	0.18	0.37	0.35	0.78	0.56	0.75	1.18	1.36	1.17	1.65

- (α) Σχηματίστε το κατάλληλο διάγραμμα διασποράς και σχολιάστε αν φαίνεται ο συντελεστής εξάτμισης σταγονιδίων καμένων καυσίμων να εξαρτάται γραμμικά από την ταχύτητα του αέρα.

- (β) Υπολογίστε τις σημειακές εκτιμήσεις a και b των παραμέτρων a και b της ευθείας παλινδρόμησης (με τη μέθοδο των ελαχίστων τετραγώνων) και σχηματίστε στο διάγραμμα διασποράς του ερωτήματος 5α' την ευθεία ελαχίστων τετραγώνων.
- (γ) Με βάση το δείγμα, εκτιμήστε (αν γίνεται) το μέσο συντελεστή εξάτμισης σταγονιδίων καμένων καυσίμων για ταχύτητα αέρα 190cm/sec και 500cm/sec.
6. Σε μια μελέτη για την επίδραση της βροχόπτωσης στον καθαρισμό του μολυσμένου αέρα, έγιναν 9 μετρήσεις καθημερινής βροχής (σε 0.01cm) και μολυσμένων σωματιδίων (σε mcg/cum) που απαλείφονται από την ατμόσφαιρα. Τα αποτελέσματα δίνονται στον παρακάτω πίνακα.

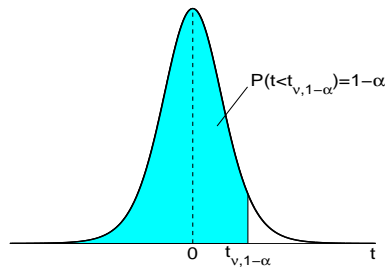
βροχή	2.1	3.8	4.3	4.5	5.2	5.6	5.9	6.1	7.5
μολυσμένα σωματίδια	141	132	126	121	118	118	116	114	108

- (α) Σχηματίστε το κατάλληλο διάγραμμα διασποράς και σχολιάστε αν φαίνεται να υπάρχει εξάρτηση της ποσότητας του μολυσμένου αέρα από το μέγεθος της βροχόπτωσης, και αν ναι, τι είδους εξάρτηση είναι.
- (β) Υπολογίστε τις σημειακές εκτιμήσεις a και b των παραμέτρων a και b της ευθείας παλινδρόμησης (με τη μέθοδο των ελαχίστων τετραγώνων) για το πρόβλημα της γραμμικής εξάρτησης της ποσότητας του μολυσμένου αέρα από το μέγεθος της βροχόπτωσης.
- (γ) Με βάση το δείγμα και το μοντέλο στο 6β', υπολογίστε τη σημειακή εκτίμηση (αν γίνεται) της αληθινής μέσης ποσότητας μολυσμένων σωματιδίων που απαλείφονται από τον αέρα όταν το μέγεθος της βροχόπτωσης είναι 5 (σε 0.01cm). Κάνετε το ίδιο για μέγεθος βροχόπτωσης 10 (σε 0.01cm).
7. Η χρήση αστικού λεωφορείου σε μια πόλη πιστεύεται ότι επηρεάζεται από το αντίτιμο του εισιτηρίου. Υπάρχουν στοιχεία για τη χρήση του αστικού λεωφορείου (σε ποσοστό μετακίνησης) μετά από αυξήσεις του εισιτηρίου. Στον παρακάτω πίνακα δίνονται 6 τιμές εισιτηρίων και το αντίστοιχο ποσοστό μετακίνησης με αστικό λεωφορείο.

Τιμή εισιτηρίου (σε €)	0.15	0.18	0.25	0.30	0.40	0.50
Ποσοστό κίνησης (σε %)	37	38	34	30	32	28

- (α) Σχηματίστε το κατάλληλο διάγραμμα διασποράς και σχολιάστε αν φαίνεται να υπάρχει εξάρτηση της χρήσης αστικού λεωφορείου από το αντίτιμο του εισιτηρίου, και αν ναι, τι είδους εξάρτηση είναι.
- (β) Υπολογίστε τις σημειακές εκτιμήσεις a και b των παραμέτρων a και b της ευθείας παλινδρόμησης (με τη μέθοδο των ελαχίστων τετραγώνων) για το πρόβλημα της γραμμικής εξάρτησης του ποσοστού χρήσης αστικού λεωφορείου από το αντίτιμο του εισιτηρίου. Σχηματίστε την ευθεία ελαχίστων τετραγώνων στο διάγραμμα διασποράς που σχηματίσατε στο ερώτημα 7α'.
- (γ) Με βάση το δείγμα και το μοντέλο στο ερώτημα 7β', υπολογίστε τη σημειακή εκτίμηση (αν γίνεται) της μέσου ποσοστού κίνησης με αστικό λεωφορείο αν το αντίτιμο είναι 0.35€. Κάνετε το ίδιο για τιμή εισιτηρίου 0.70€.

Στατιστικός Πίνακας Κατανομής Student



ν : βαθμοί ελευθερίας

$1 - \alpha$: τιμή της αθροιστικής συνάρτησης

Παράδειγμα:

$$\nu = 10, \alpha = 0.05 \implies t_{10,0.95} = 1.812$$

$$\nu = 10, \alpha = 0.001 \implies t_{10,0.999} = 4.144$$

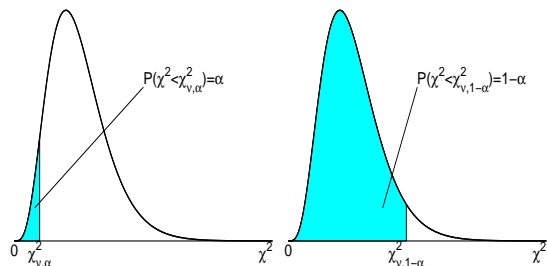
$$\nu = 20, \alpha = 0.025 \implies t_{20,0.975} = 2.086$$

ν	$1 - \alpha$						
	0.900	0.950	0.975	0.990	0.995	0.999	0.9995
1	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
32	1.309	1.694	2.037	2.449	2.738	3.365	3.622
34	1.307	1.691	2.032	2.441	2.728	3.348	3.601
36	1.306	1.688	2.028	2.434	2.719	3.333	3.582
38	1.304	1.686	2.024	2.429	2.712	3.319	3.566
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	1.299	1.676	2.009	2.403	2.678	3.261	3.496
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
100	1.290	1.660	1.984	2.364	2.626	3.174	3.390
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
150	1.287	1.655	1.976	2.351	2.609	3.145	3.357
200	1.286	1.653	1.972	2.345	2.601	3.131	3.340
300	1.284	1.650	1.968	2.339	2.592	3.118	3.323
400	1.284	1.649	1.966	2.336	2.588	3.111	3.315
∞	1.282	1.646	1.962	2.330	2.581	3.098	3.300

Στατιστικός Πίνακας Κατανομής χ^2

ν : βαθμοί ελευθερίας

α ή $1-\alpha$: τιμή αθροιστικής συνάρτησης



Παράδειγμα:

$$\nu = 10, \alpha = 0.05 \implies$$

$$\chi^2_{10,0.05} = 3.94$$

$$\chi^2_{10,0.95} = 18.31$$

$$\nu = 18, \alpha = 0.025 \implies$$

$$\chi^2_{18,0.025} = 8.23$$

$$\chi^2_{18,0.975} = 31.53$$

ν	α					$1 - \alpha$				
	0.001	0.005	0.010	0.025	0.050	0.950	0.975	0.990	0.995	0.999
1	0.00	0.00	0.00	0.00	0.00	3.84	5.02	6.63	7.88	10.83
2	0.00	0.01	0.02	0.05	0.10	5.99	7.38	9.21	10.60	13.82
3	0.02	0.07	0.11	0.22	0.35	7.81	9.35	11.34	12.84	16.27
4	0.09	0.21	0.30	0.48	0.71	9.49	11.14	13.28	14.86	18.47
5	0.21	0.41	0.55	0.83	1.15	11.07	12.83	15.09	16.75	20.52
6	0.38	0.68	0.87	1.24	1.64	12.59	14.45	16.81	18.55	22.46
7	0.60	0.99	1.24	1.69	2.17	14.07	16.01	18.48	20.28	24.32
8	0.86	1.34	1.65	2.18	2.73	15.51	17.53	20.09	21.95	26.12
9	1.15	1.73	2.09	2.70	3.33	16.92	19.02	21.67	23.59	27.88
10	1.48	2.16	2.56	3.25	3.94	18.31	20.48	23.21	25.19	29.59
11	1.83	2.60	3.05	3.82	4.57	19.68	21.92	24.72	26.76	31.26
12	2.21	3.07	3.57	4.40	5.23	21.03	23.34	26.22	28.30	32.91
13	2.62	3.57	4.11	5.01	5.89	22.36	24.74	27.69	29.82	34.53
14	3.04	4.07	4.66	5.63	6.57	23.68	26.12	29.14	31.32	36.12
15	3.48	4.60	5.23	6.26	7.26	25.00	27.49	30.58	32.80	37.70
16	3.94	5.14	5.81	6.91	7.96	26.30	28.85	32.00	34.27	39.25
17	4.42	5.70	6.41	7.56	8.67	27.59	30.19	33.41	35.72	40.79
18	4.90	6.26	7.01	8.23	9.39	28.87	31.53	34.81	37.16	42.31
19	5.41	6.84	7.63	8.91	10.12	30.14	32.85	36.19	38.58	43.82
20	5.92	7.43	8.26	9.59	10.85	31.41	34.17	37.57	40.00	45.31
21	6.45	8.03	8.90	10.28	11.59	32.67	35.48	38.93	41.40	46.80
22	6.98	8.64	9.54	10.98	12.34	33.92	36.78	40.29	42.80	48.27
23	7.53	9.26	10.20	11.69	13.09	35.17	38.08	41.64	44.18	49.73
24	8.08	9.89	10.86	12.40	13.85	36.42	39.36	42.98	45.56	51.18
25	8.65	10.52	11.52	13.12	14.61	37.65	40.65	44.31	46.93	52.62
26	9.22	11.16	12.20	13.84	15.38	38.89	41.92	45.64	48.29	54.05
27	9.80	11.81	12.88	14.57	16.15	40.11	43.19	46.96	49.64	55.48
28	10.39	12.46	13.56	15.31	16.93	41.34	44.46	48.28	50.99	56.89
29	10.99	13.12	14.26	16.05	17.71	42.56	45.72	49.59	52.34	58.30
30	11.59	13.79	14.95	16.79	18.49	43.77	46.98	50.89	53.67	59.70
32	12.81	15.13	16.36	18.29	20.07	46.19	49.48	53.49	56.33	62.49
34	14.06	16.50	17.79	19.81	21.66	48.60	51.97	56.06	58.96	65.25
36	15.32	17.89	19.23	21.34	23.27	51.00	54.44	58.62	61.58	67.99
38	16.61	19.29	20.69	22.88	24.88	53.38	56.90	61.16	64.18	70.70
40	17.92	20.71	22.16	24.43	26.51	55.76	59.34	63.69	66.77	73.40
50	24.67	27.99	29.71	32.36	34.76	67.50	71.42	76.15	79.49	86.66
60	31.74	35.53	37.48	40.48	43.19	79.08	83.30	88.38	91.95	99.61
100	61.92	67.33	70.06	74.22	77.93	124.34	129.56	135.81	140.17	149.45
120	77.76	83.85	86.92	91.57	95.70	146.57	152.21	158.95	163.65	173.62
150	102.11	109.14	112.67	117.98	122.69	179.58	185.80	193.21	198.36	209.26
200	143.84	152.24	156.43	162.73	168.28	233.99	241.06	249.45	255.26	267.54
300	229.96	240.66	245.97	253.91	260.88	341.40	349.87	359.91	366.84	381.43
400	318.26	330.90	337.16	346.48	354.64	447.63	457.31	468.72	476.61	493.13
∞	867.48	888.56	898.91	914.26	927.59	1074.68	1089.53	1106.97	1118.95	1143.92