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 γνωρίζουμε αν κατά μέσο όρο ο χρόνος ζωής αντλίας θερμότητας κάποιου τύπου είναι μεγαλύτερος από μια συγκεκριμένη τιμή. Βέβαια μπορούμε να απαντήσουμε σ' ένα τέτοιο ερώτημα εκτιμώντας το κατάλληλο 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 $\mathrm{H}_0 : \vartheta = \vartheta_0$. Η **εναλλακτική υπόθεση** (alternative hypothesis) H_1 την οποία δεχόμαστε αν απορρίψουμε τη H_0 μπορεί να είναι:

1. $\mathrm{H}_1 : \vartheta \neq \vartheta_0$, 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. $\mathrm{H}_1 : \vartheta < \vartheta_0$, 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 ($\mathrm{H}_0 : \vartheta = \vartheta_0$) και στις μικρότερες από τη ϑ_0 ($\mathrm{H}_1 : \vartheta < \vartheta_0$). Γι αυτό και είναι σωστότερο σ' αυτήν την

περίπτωση να θεωρούμε πως $\mathrm{H}_0 : \vartheta \geq \vartheta_0$, έστω κι αν πρακτικά η ϑ δε μπορεί να πάρει τιμές μεγαλύτερες της ϑ_0.

3. $\mathrm{H}_1 : \vartheta > \vartheta_0$, 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 ($\mathrm{H}_0 : \mu_1 \geq \mu_2$ και $\mathrm{H}_1 : \mu_1 < \mu_2$) 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 αλλού μηχανήματος της αγοράς κι *ελπίζουμε* το μηχάνημα μας να αποδίδει καλύτερα, ο έλεγχος δε θα πρέπει να είναι μονόπλευρος ($\mathrm{H}_0 : \mu_1 \leq \mu_3$ και $\mathrm{H}_1 : \mu_1 > \mu_3$) αλλά δίπλευρος ($\mathrm{H}_0 : \mu_1 = \mu_3$ και $\mathrm{H}_1 : \mu_1 \neq \mu_3$) 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

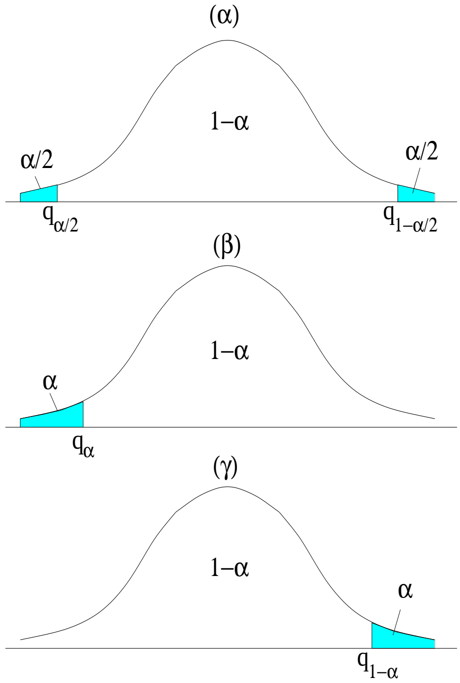
Η μηδενική υπόθεση $\mathrm{H}_0 : \vartheta = \vartheta_0$ υποδηλώνει ότι η εκτίμηση ϑ_0 της ϑ είναι σωστή. Επίσης, τιμές της ϑ 'κοντά' στη ϑ_0 υποστηρίζουν την ορθότητα της H_0 ενώ τιμές της ϑ 'μακριά' από τη ϑ_0 δεν την υποστηρίζουν. Έτσι, χωρίζουμε τις δυνατές τιμές της παραμέτρου ϑ σε αυτές για 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 *μή-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 $R = \{q| \; q < q_{\alpha/2} \;\vee\; q > q_{1-\alpha/2}\}$, 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, $R = \{q| \; q < q_\alpha\}$ για την αριστερή πλευρά και $R = \{q| \; q > q_{1-\alpha}\}$ 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 ελέγχου γίνεται με βάση το δείγμα που δίνει και την εκτίμηση $\hat{\vartheta}$ για την παράμετρο ϑ. Βέβαια αφού έχουμε σχηματίσει την R ως προς την q, χρησιμοποιούμε αντί της τιμής $\hat{\vartheta}$ 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 $\mathrm{H}_0 : \vartheta = \vartheta_0$ είναι

Επιλογή του επιπέδου σημαντικότητας α
Καθορισμός της κατάλληλης στατιστικής ελέγχου q
$\Downarrow$
Υπολογισμός κρίσιμης(-ων) τιμής(-ών) της q από τον αντίστοιχο στατιστικό πίνακα και καθορισμός της περιοχής απόρριψης R
$\Downarrow$
Υπολογισμός της στατιστικής ελέγχου $\tilde{q}$ από το δείγμα
$\Downarrow$
Απόρριψη της H_0 αν $\tilde{q}$ ανήκει στην R.
Μή απόρριψη της H_0 αν $\tilde{q}$ δεν ανήκει στην R.

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 $\mathrm{H}_0 : \mu = \mu_0$ η στατιστική ελέγχου είναι

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

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_a είναι τέτοια ώστε $\mathrm{P}(z < z_a) = \Phi(z_a) = a$. Λόγω συμμετρικότητας της τυπικής κανονικής κατανομής γύρω από το μηδέν, είναι $z_a = -z_{1-a}$ (δες Παράγραφο 2.2.1).] Η περιοχή απόρριψης για επίπεδο σημαντικότητας a είναι ανάλογα με την εναλλακτική υπόθεση

1. $\mathrm{H}_1 : \mu \neq \mu_0$, δίπλευρος έλεγχος και $R = \{z|\ z < z_{a/2}\ \vee\ z > z_{1-a/2}\}$ ή $R = \{z|\ |z| > z_{1-a/2}\}$.
2. $\mathrm{H}_1 : \mu < \mu_0$, μονόπλευρος έλεγχος από αριστερά και $R = \{z|\ z < z_a\} = \{z|\ z < -z_{1-a}\}$.
3. $\mathrm{H}_1 : \mu > \mu_0$, μονόπλευρος έλεγχος από δεξιά και $R = \{z|\ z > z_{1-a}\}$.

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 a, αλλιώς λέμε ότι δεν έχουμε αρκετές ενδείξεις για να την απορρίψουμε.

Η p-τιμή για τα τρία είδη ελέγχου είναι

1. $\mathrm{H}_1 : \mu \neq \mu_0, \quad p = 2\,\mathrm{P}(z > |\tilde{z}|) = 2\,(1 - \mathrm{P}(z < |\tilde{z}|))$.
2. $\mathrm{H}_1 : \mu < \mu_0, \quad p = \mathrm{P}(z < \tilde{z}) = 1 - \mathrm{P}(z < -\tilde{z})$.
3. $\mathrm{H}_1 : \mu > \mu_0, \quad p = \mathrm{P}(z > \tilde{z}) = 1 - \mathrm{P}(z < \tilde{z})$.

Παράδειγμα 3.1. *Χρησιμοποιούμε τα δεδομένα από το Παράδειγμα 1.2 και θέλουμε να ελέγξουμε αν το μέσο ποσοστό του πορώδους ηλίου μ στο γαιάνθρακα από το κοίτασμα A είναι 6. Υποθέτουμε ότι γνωρίζουμε πως η διασπορά είναι γνωστή και είναι $\sigma^2 = 0.4$. Τα δεδομένα του πορώδες ηλίου για το κοίτασμα γαιάνθρακα A δίνονται στον Πίνακα 1.3. Οι υποθέσεις είναι $H_0 : \mu = 6$ και $H_1 : \mu \neq 6$. Για το επίπεδο σημαντικότητας $a = 0.05$ η κρίσιμη τιμή είναι $z_{0.975} = 1.96$ και η περιοχή απόρριψης της H_0 είναι $R = \{z|\ |z| > 1.96\}$. 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 το μέσο ποσοστό πορώδους ηλίου στο γαιάνθρακα από το κοίτασμα A δε μπορεί να είναι 6. 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 σε επίπεδο 99.1% να πούμε πως το μέσο ποσοστό του πορώδους ηλίου δε μπορεί να είναι 6.

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

$$q = t \equiv \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim \mathrm{t}_{n-1}. \tag{3.2}$$

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. $\mathrm{H}_1 : \mu \neq \mu_0, \quad R = \{t \mid |t| > t_{n-1,1-\alpha/2}\}$.

2. $\mathrm{H}_1 : \mu < \mu_0, \quad R = \{t \mid t < -t_{n-1,1-\alpha}\}$.

3. $\mathrm{H}_1 : \mu > \mu_0, \quad R = \{t \mid t > t_{n-1,1-\alpha}\}$.

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 το πορώδες ηλίου του γαιάνθρακα από το κοίτασμα 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 $\mathrm{H}_0 : \mu = \mu_0$ 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 $\mathrm{H}_0 : \sigma^2 = \sigma_0^2$ 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)

$$q = \chi^2 \equiv \frac{(n-1)s^2}{\sigma_0^2} \sim \mathcal{X}_{n-1}^2, \tag{3.3}$$

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

Πίνακας 3.1: Στατιστική ελέγχου υπόθεσης $\mathrm{H}_0 : \mu = \mu_0$ 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 $\mathcal{X}^2$ 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. $\mathrm{H}_1 : \sigma^2 \neq \sigma_0^2, \quad R = \{\chi^2 | \ \chi^2 < \chi^2_{n-1,\alpha/2} \ \vee \ \chi^2 > \chi^2_{n-1,1-\alpha/2}\}$.
2. $\mathrm{H}_1 : \sigma^2 < \sigma_0^2, \quad R = \{\chi^2 | \ \chi^2 < \chi^2_{n-1,\alpha}\}$.
3. $\mathrm{H}_1 : \sigma^2 > \sigma_0^2, \quad R = \{\chi^2 | \ \chi^2 > \chi^2_{n-1,1-\alpha}\}$.

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. $\mathrm{H}_1 : \sigma^2 \neq \sigma_0^2, \quad p = \mathrm{P}(\chi^2 < \tilde{\chi}^2 \ \vee \ \chi^2 > \tilde{\chi}^2\}$.
2. $\mathrm{H}_1 : \sigma^2 < \sigma_0^2, \quad p = \mathrm{P}(\chi^2 < \tilde{\chi}^2\}$.
3. $\mathrm{H}_1 : \sigma^2 > \sigma_0^2, \quad p = \mathrm{P}(\chi^2 > \tilde{\chi}^2\}$.

Παράδειγμα 3.3. *Θέλουμε να ελέγξουμε αν θα μπορούσαμε να πούμε με μεγάλη σιγουριά (επίπεδο εμπιστοσύνης 99%) πως η τυπική απόκλιση σ του πορώδους ηλίου γαιάνθρακα μπορεί να ξεπεράσει το επίπεδο της μιας ποσοστιαίας μονάδας. 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 του πορώδους ηλίου του γαιάνθρακα από το κοίτασμα A μεγέθους $n = 25$ (Πίνακας 1.3) και η δειγματική διασπορά είναι $s^2 = 0.375$. Η κρίσιμη τιμή της στατιστικής χ^2 για $\alpha = 0.01$ είναι $\chi^2_{24,0.99} = 42.98$ 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 του ποσοστού του πορώδους ηλίου σε γαιάνθρακα να ξεπεράσει το επίπεδο της μονάδας 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 $\mathrm{H}_0 : p = p_0$ και η στατιστική ελέγχου είναι (δες επίσης Παράγραφο 2.2.3 για το διάστημα εμπιστοσύνης της p)

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

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. $\mathrm{H}_1 : p \neq p_0$, $R = \{z | \, |z| > z_{1-\alpha/2}\}$.
2. $\mathrm{H}_1 : p < p_0$, $R = \{z | \, z < -z_{1-\alpha}\}$.
3. $\mathrm{H}_1 : p > p_0$, $R = \{z | \, z > z_{1-\alpha}\}$.

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 ράβδοι σε δείγμα 100 ραβδών). Η κρίσιμη τιμή για τη στατιστική ελέγχου σε επίπεδο σημαντικότητας $\alpha = 0.05$ είναι $z_{0.975} = 1.96$ κι η περιοχή απόρριψης είναι $R = \{z | \, |z| > 1.96\}$. Η εκτίμηση της αναλογίας από το δείγμα των 100 ραβδών είναι $\hat{p} = 0.12$ κι από τη σχέση (3.4) βρίσκουμε τη στατιστική ελέγχου 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* $a = 0.05$ *είναι* $z_{0.95} = 1.65$ *κι η περιοχή απόρριψης είναι* $R = \{z | \; z < -1.65\}$. *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 η περιεκτικότητα ραδιενέργειας σε χάλυβα από δύο εργοστάσια. 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 $\mathrm{H}_0 : \mu_1 = \mu_2$ ή ισοδύναμα $\mathrm{H}_0 : \mu_1 - \mu_2 = 0$. 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 $\mathrm{H}_0 : \mu_1 - \mu_2 = 0$, η στατιστική του παραμετρικού ελέγχου είναι

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

Οι κρίσιμες τιμές κι οι περιοχές απόρριψης για δίπλευρο και μονόπλευρο έλεγχο είναι όπως και για τον ελέγχο 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

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \mathrm{t}_{n_1+n_2-2} \tag{3.6}$$

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 ελέγχο 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 οι γαιάνθρακες από δύο κοίτασματα A και B έχουν το ίδιο πορώδες ηλίου. Γι αυτό κάνουμε δίπλευρο έλεγχο για τα μέσα ποσοστά πορώδους ηλίου μ_1 και μ_2 για τους γαιάνθρακες από τα κοιτάσματα 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 του πορώδες ηλίου γαιάνθρακα από τα κοιτάσματα A και B του Πίνακα 1.3. Υποθέτουμε πως το ποσοστό πορώδες ηλίου του γαιάνθρακα και από τα δύο κοιτάσματα ακολουθεί κανονική κατανομή, το οποίο φαίνεται να ισχύει κι από τα ιστογράμματα και τα θηκογράμματα για το πορώδες ηλίου των δύο δειγμάτων (δες Σχήματα 1.2 και 2.7). Οι δειγματικές μέσες τιμές είναι $\bar{x}_1 = 5.67$ και $\bar{x}_2 = 5.38$ για το κοίτασμα A και B αντίστοιχα.*

Πρώτα θεωρούμε πως από παλιότερες μετρήσεις γνωρίζουμε πως η διασπορά του ποσοστού πορώδους ηλίου είναι ίδια και για τα δύο κοιτάσματα γαιάνθρακα και είναι $\sigma^2 = 0.4$. 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 το μέσο ποσοστό πορώδους ηλίου στο γαιάνθρακα του κοιτάσματος A είναι διαφορετικό από αυτό στο κοίτασμα B. 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 του ποσοστού πορώδους ηλίου για τα δύο κοιτάσματα γαιάνθρακα είναι κοινή αλλά δεν εμπιστευόμαστε την εμπειρική εκτίμηση $\sigma^2 = 0.4$. Άρα θα πρέπει *να χρησιμοποιήσουμε στον έλεγχο μας τις εκτιμήσεις των διασπορών για τους δύο τύπους. Στο Παράδειγμα 2.15 βρήκαμε* $s_1^2 = 0.375$ *για τον τύπο A,* $s_2^2 = 0.326$ *για τον τύπο B και* $s^2 = 0.353$ *για την κοινή διασπορά* σ^2 *(*$s = 0.594$*).*

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* $R = \{t| \, |t| > 2.02\}$*. 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 $\mathrm{H}_0 : \mu_1 - \mu_2 = 0$ 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		$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathrm{N}(0,1)$
γνωστές	μη κανονική	μεγάλα	$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathrm{N}(0,1)$
γνωστές	μη κανονική	μικρά	—
άγνωστες ίσες/άνισες		μεγάλα	$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim \mathrm{N}(0,1)$
άγνωστες άνισες	κανονική	μικρά	$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \mathrm{t}_{n_1+n_2-2}$
άγνωστες άνισες	μη κανονική	μικρά	—
άγνωστες άνισες		μικρά	—

Πίνακας 3.2: Έλεγχος της $\mathrm{H}_0 : \mu_1 - \mu_2 = 0$ 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 α-νεξάρτητοι 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 περίπτωση διάβρωσης από νερό πλημμύρας, μας ενδιαφέρει να συγκρίνουμε την ανθεκτικότητα τους πριν και μετά κάποια πλημμύρα. Βέβαια αυτό είναι δυνατόν αν πρώτα μετρήσαμε την ανθεκτικότητα των δοκών σ' ένα δείγμα κτιρίων πριν τη πλημμύρα και επαναλάβαμε τις ίδιες μετρήσεις σε κάποιο χρονικό διάστημα μετά την πλημμύρα. Είναι φανερό ότι οι τ.μ. X_1 και X_2 είναι εξαρτημένες (η ανθεκτικότητα των δοκών ενός κτιρίου μετά τη πλημμύρα καθορίζεται ως ένα βαθμό από την ανθεκτικότητα τους πριν τη πλημμύρα).

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 τη πλημμύρα) 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 $\mathrm{H}_0 : \mu_D = 0$ 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 $\mathrm{H}_0 : p_1 - p_2 = 0$, 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

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

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

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

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

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

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* $R = \{z | \, |z| > 1.96\}$. *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 απόθηκη 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* $R = \{z | \, z < -1.65\}$ *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 αλλά 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

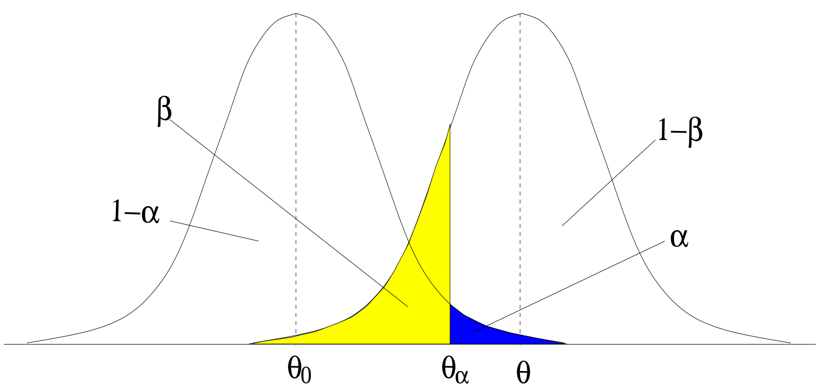
Ο έλεγχος υπόθεσης καταλήγει στην απόφαση επιλογής της μίας από τις δύο υποθέσεις, της μηδενικής H_0 ή της εναλλακτικής H_1. Κάθε μία από τις δύο υποθέσεις μπορεί να είναι σωστή ή λανθασμένη. Συνδυάζοντας τις δύο δυνατές αποφάσεις του ελέγχου με τις δύο δυνατές πραγματικές καταστάσεις για την στατιστική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

1. *Σωστή απόφαση*: Αποδεχόμαστε την H_0 όταν η H_0 είναι σωστή. Η πιθανότητα αυτής της απόφασης είναι
$$\text{P(αποδοχή της } H_0 \mid H_0 \text{ σωστή)} = 1 - \alpha.$$

2. **Σφάλμα τύπου ΙΙ** (type II error): Αποδεχόμαστε την H_0 όταν η H_0 είναι λανθασμένη. Η πιθανότητα αυτού του σφάλματος είναι
$$\text{P(αποδοχή της } H_0 \mid H_0 \text{ λανθασμένη)} = \beta.$$

3. **Σφάλμα τύπου Ι** (type I error): Απορρίπτουμε την H_0 όταν η H_0 είναι σωστή. Η πιθανότητα αυτού του σφάλματος είναι το επίπεδο σημαντικότητας
$$\text{P(απόρριψη της } H_0 \mid H_0 \text{ σωστή)} = \alpha.$$

4. *Σωστή απόφαση*: Απορρίπτουμε την H_0 και η H_0 είναι λανθασμένη. Η πιθανότητα αυτής της απόφασης είναι
$$\text{P(απόρριψη της } H_0 \mid H_0 \text{ λανθασμένη)} = 1 - \beta$$
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 $\hat{\vartheta}$ της ϑ, αριστερά στο Σχήμα 3.2, είναι σύμφωνη με τη μηδενική υπόθεση $\mathrm{H}_0 : \vartheta = \vartheta_0$ 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 $\hat{\vartheta}$ (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 βαθειά σκούρη περιοχή) είναι η περιοχή λανθασμένης απόρριψης της H_0 με πιθανότητα α (σφάλμα τύπου I). Θεωρώντας τη δεύτερη κατανομή της $\hat{\vartheta}$ (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 αντισταθμίσουμε 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% ($\mathrm{H}_0 : p \geq 0.19$). 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 $\mathrm{H}_0 : p \geq 0.19$ 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 ράβδους 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 γι 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. Τρείς 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 $\hat{\vartheta}$*: 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 $\vartheta-\vartheta_0$*: 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 το μέσο ποσοστό πορώδους ηλίου του γαιάνθρακα μ μπορεί να είναι 6 ($H_0: \mu=\mu_0=6$). Υποθέσαμε γνωστή διασπορά $\sigma^2=0.4$ 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

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

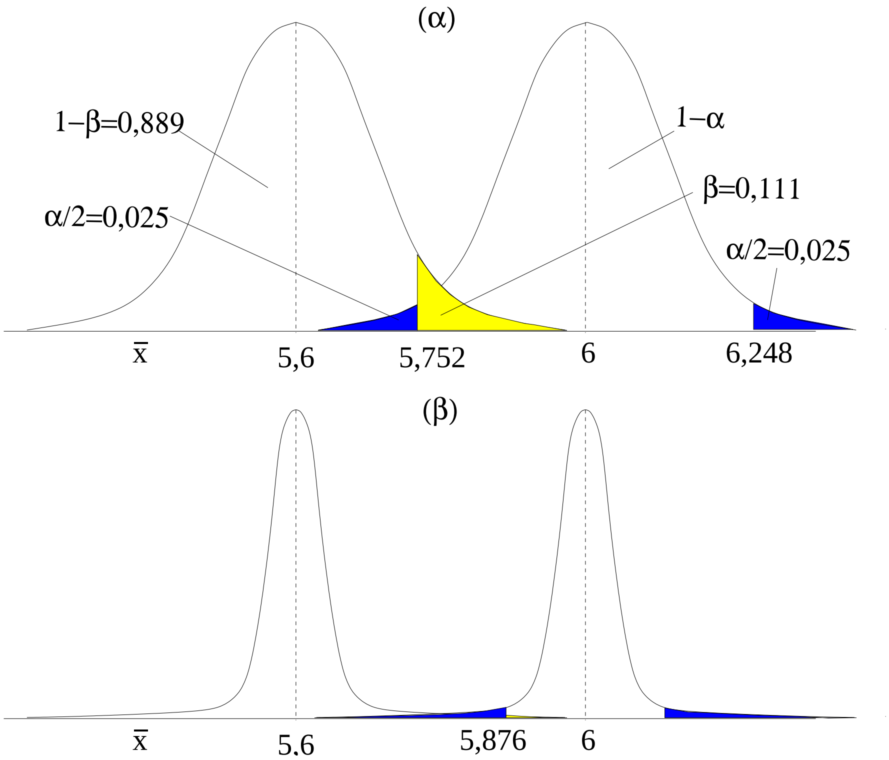
και για τη δεξιά κρίσιμη τιμή

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

Άρα η πιθανότητα του σφάλματος τύπου I ορίζεται ως

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

όπως φαίνεται στο Σχήμα 3.3α από τις βαθειά 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. *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 σκούρη 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 το τυπικό ποσοστό πορώδους ηλίου που θα πρέπει να έχει ο γαιάνθρακας για να θεωρείται αποδεκτός).*

- *Υποθέτουμε ότι η διασπορά του ποσοστού πορώδους ηλίου στο γαιάνθρακα είναι μικρότερη ($\sigma^2 < 0.4$). Η μεταβλητότητα του ποσοστού πορώδους ηλίου όμως είναι ένα φυσικό φαινόμενο και δε μπορούμε να την ορίσουμε όπως θέλουμε (μπορούμε βέβαια να την εκτιμήσουμε λάθος!).*

- *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*

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

  Όπως φαίνεται κι από το Σχήμα 3.3β αυτό μειώνει σημαντικά την πιθανότητα β σφάλματος τύπου ΙΙ 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 $\mathrm{H}_0 : \theta = \theta_0$ 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 του πορώδους ηλίου σε γαιάνθρακα με δειγματική μέση τιμή $\bar{x} = 5.67$, βρήκαμε το 95% διάστημα εμπιστοσύνης*

[5.42, 5.92] *για το μέσο ποσοστό πορώδους ηλίου 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 μικρά ποσοστά πορώδους ηλίου).*

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 κοιτάσματα γαιάνθρακα A και B, μας ενδιαφέρει να δούμε αν έχουν το ίδιο μέσο ποσοστό πορώδες ηλίου κι αν όχι να εκτιμήσουμε πόση είναι η διαφορά τους. Ο έλεγχος της $\mathrm{H}_0 : \mu_1 - \mu_2 = 0$ 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 το μέσο ποσοστό πορώδους ηλίου μ_1 του γαιάνθρακα από το κοίτασμα A είναι μεγαλύτερο από το αντίστοιχο μ_2 για το κοίτασμα B, 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 εκτιμούμενη τιμή της αναλογίας από το δείγμα), ενώ για το δίπλευρο έλεγχο θεωρείται γνωστή κάτω από τη μηδενική υπόθεση $\mathrm{H}_0 : p = p_0$ 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.