

Στατιστικός Έλεγχος Υποθέσεων

Ένας νέος τύπος τσιγάρων βρίσκεται στο στάδιο ποιοτικού ελέγχου. Αν το τμήμα ποιοτικού ελέγχου της καπνοβιομηχανίας παραγωγής, ενδιαφέρεται να γνωρίζει τη μέση ποσότητα νικοτίνης που περιέχεται στα νέου τύπου τσιγάρα, μπορεί να υπολογίσει ένα *διάστημα εμπιστοσύνης* και να πάρει έτσι μια εκτίμηση για την άγνωστη μέση ποσότητα νικοτίνης. Στην περίπτωση όμως, που ενδιαφέρεται να γνωρίζει μόνο αν στα νέου τύπου τσιγάρα η μέση ποσότητα νικοτίνης δεν υπερβαίνει ένα μέγιστο επιτρεπτό όριο, τότε πρέπει να κάνει κατάλληλο **στατιστικό έλεγχο υποθέσεων** ώστε να μπορεί να αποφασίσει μεταξύ των υποθέσεων:

- Η μέση ποσότητα νικοτίνης δεν υπερβαίνει το μέγιστο επιτρεπτό όριο.
- Η μέση ποσότητα νικοτίνης υπερβαίνει το μέγιστο επιτρεπτό όριο.

Ο στατιστικός έλεγχος υποθέσεων (*hypothesis testing*) είναι μια συμπερασματική διαδικασία/μέθοδος που προσφέρει η *Στατιστική Συμπερασματολογία* και βρίσκει εφαρμογή σε στοχαστικά προβλήματα απόφασης μεταξύ δύο εναλλακτικών υποθέσεων. Η μία υπόθεση έχει επικρατήσει να συμβολίζεται με H_0 και ονομάζεται *μηδενική υπόθεση (null hypothesis)*, και η άλλη με H_1 και ονομάζεται *εναλλακτική υπόθεση (alternative hypothesis)*.

Αναγκαία προϋπόθεση για τη σωστή εφαρμογή των *στατιστικών ελέγχων* και κυρίως για τη σωστή ερμηνεία των αποτελεσμάτων τους, είναι η κατανόηση της λογικής και του νοήματός τους. Στη συνέχεια, αυτό θα προσπαθήσουμε. Να αναδείξουμε τη λογική, το νόημα και τα όρια εφαρμογής τους.

1. Βασικές Έννοιες

Η γενική ιδέα της διαδικασίας *στατιστικού ελέγχου υποθέσεων* είναι η εξής: θέτουμε ως *μηδενική υπόθεση* (H_0) αυτή για την οποία αμφιβάλλουμε, αυτή που αμφισβητείται, και εξετάζουμε αν ένα τυχαίο δείγμα που παίρνουμε από τον πληθυσμό συνηγορεί-δίνει αποδείξεις υπέρ της απόρριψής της, έναντι της *εναλλακτικής* (H_1).

Δηλαδή, η H_0 , απορρίπτεται ή δεν απορρίπτεται με βάση το **τι παρατηρείται** στο τυχαίο δείγμα που πήραμε από τον πληθυσμό. Πιο συγκεκριμένα, υποθέτοντας ότι η H_0 είναι αληθής, αν αυτό που παρατηρείται στο δείγμα είναι ακραίο, δηλαδή, αν έχει πολύ μικρή πιθανότητα να συμβεί, τότε απορρίπτουμε την H_0 . Σε αντίθετη περίπτωση, δηλαδή, αν αυτό που παρατηρείται στο δείγμα δεν είναι ακραίο-σπάνιο (όταν είναι αληθής η H_0) τότε το δείγμα που πήραμε δε μας δίνει αρκετές ενδείξεις για την απόρριψη της H_0 και «αποτυγχάνουμε να την απορρίψουμε».

Βέβαια, με αυτή τη στρατηγική παίρνουμε «ρίσκο», γιατί και τα ... ακραία, έστω και με πολύ μικρή πιθανότητα, μπορεί να συμβούν.

Πιο συγκεκριμένα, με την υπόθεση ότι η H_0 είναι αληθής, αν κρίνουμε ότι αυτό που παρατηρείται στο τυχαίο δείγμα είναι ακραίο και την απορρίψουμε, τότε ακριβώς ένα από τα παρακάτω μπορεί να συνέβη:

- (α) είτε η H_0 πράγματι δεν είναι αληθής, όποτε αποφασίσαμε σωστά,
- (β) είτε η H_0 είναι αληθής και το ακραίο οφείλεται στην τύχη, δηλαδή, συνέβη κάτι σπάνιο (εμφανίστηκε ένα δείγμα που σπάνια εμφανίζεται). Στην περίπτωση αυτή,

απορρίψαμε λανθασμένα την H_0 . Αυτό το σφάλμα ονομάζεται **σφάλμα τύπου I (type I error)**. Εφόσον, υπό την H_0 , το ακραίο υπάρχει πιθανότητα, έστω πολύ μικρή π.χ. 0.0001, να συμβεί, τότε, απορρίπτουμε λανθασμένα την H_0 με πιθανότητα 0.0001.

Ανάλογα, είναι δυνατόν, λανθασμένα να μην απορρίψουμε την H_0 . Δηλαδή, να αποτύχουμε να απορρίψουμε την H_0 , ενώ είναι αληθής η H_1 . Αυτό το σφάλμα ονομάζεται **σφάλμα τύπου II (type II error)**. Το «ρίσκο», επομένως, είναι διπλό, με πιθανότητα,

- **λανθασμένης απόρριψης της H_0 ,**
 $P(\text{σφάλμα τύπου I}) = P(\text{απόρριψη της } H_0 \mid \text{αληθής η } H_0)$ και
- **λανθασμένης μη απόρριψης της H_0 ,**
 $P(\text{σφάλμα τύπου II}) = P(\text{μη απόρριψη της } H_0 \mid \text{αληθής η } H_1).$

Είναι φανερό, ότι για να προχωρήσουμε πρέπει να αποσαφηνιστεί, α) τι εννοούμε επακριβώς όταν λέμε «αυτό που παρατηρείται στο δείγμα»; Πώς εκφράζεται; Μπορεί να μετρηθεί-ποσοτικοποιηθεί; β) Πώς κρίνουμε ότι «αυτό που παρατηρείται στο δείγμα» είναι ή όχι «ακραίο»; Δηλαδή, με ποιον σαφή κανόνα θεωρείται το παρατηρούμενο στο δείγμα «ακραίο»; Επίσης, πρέπει να απαντήσουμε στα εύλογα ερωτήματα: Πώς υπολογίζονται οι πιθανότητες *σφάλματος τύπου I* και *σφάλματος τύπου II*; Μπορούν να ελαχιστοποιηθούν; Σχετίζονται με κάποιο τρόπο; Μπορούμε να τις θέσουμε υπό τον έλεγχό μας;

Για να απαντήσουμε στα ερωτήματα αυτά, ας χρησιμοποιήσουμε ένα συγκεκριμένο παράδειγμα. Θα μας βοηθήσει στην κατανόηση.

Παράδειγμα: Το όριο αντοχής ενός τύπου καλωδίων είναι τυχαία μεταβλητή X , με μέση τιμή $\mu = 1500 \text{ Kgr}$ και τυπική απόκλιση $\sigma = 175 \text{ Kgr}$. Το εργοστάσιο που κατασκευάζει αυτόν τον τύπο καλωδίων ισχυρίζεται ότι βελτίωσε τα υλικά που χρησιμοποιεί και πλέον το όριο αντοχής των καλωδίων έχει αυξηθεί.

Για να ελεγχθεί ο ισχυρισμός του εργοστασίου, ως μηδενική υπόθεση θέτουμε την $H_0 : \mu = 1500 \text{ Kgr}$, δηλαδή, αυτήν η οποία αμφισβητείται από τον ισχυρισμό που ελέγχουμε.

Γενικά, η H_0 δηλώνει ότι στον πληθυσμό η κατάσταση παραμένει αμετάβλητη, δεν υπάρχει αλλαγή/διαφορά ή αλλιώς, ότι η ανεξάρτητη μεταβλητή δεν έχει επίδραση στην εξαρτημένη μεταβλητή για τον πληθυσμό¹ (στο παράδειγμά μας, ότι η βελτίωση των υλικών δεν έχει επίδραση στο όριο αντοχής των καλωδίων).

Ένας δεύτερος κανόνας για τον καθορισμό της H_0 , που έχει επίσης καθιερωθεί στη διεθνή επιστημονική πρακτική, είναι ο εξής: Ως μηδενική υπόθεση θέτουμε την υπόθεση της οποίας η λανθασμένη απόρριψη εγκυμονεί τους περισσότερους κινδύνους. Δηλαδή, αυτή που απαιτεί μεγαλύτερη προστασία από σφάλμα τύπου I.

¹ Για αυτό έχει επικρατήσει να λέγεται μηδενική υπόθεση (υποθέτουμε μηδενική αλλαγή/διαφορά στην τιμή της ελεγχόμενης παραμέτρου).

Ως εναλλακτική θέτουμε την $H_1 : \mu > 1500 \text{ Kgr}$, δηλαδή, η H_1 δηλώνει ότι η βελτίωση των υλικών επηρεάζει, και ειδικότερα αυξάνει, το όριο αντοχής των καλωδίων.

Γενικά, η H_1 δηλώνει ότι στον πληθυσμό υπάρχει αλλαγή/διαφορά ή αλλιώς, ότι η ανεξάρτητη μεταβλητή έχει επίδραση στην εξαρτημένη μεταβλητή για τον πληθυσμό.

Ο έλεγχος που μόλις διατυπώσαμε, είναι ένας **μονόπλευρος** και ειδικότερα **δεξιόπλευρος** έλεγχος. Γενικότερα, οι έλεγχοι $H_0 : \mu \leq \mu_0$, $H_0 : \mu \geq \mu_0$ ονομάζονται $H_1 : \mu > \mu_0$, $H_1 : \mu < \mu_0$

μονόπλευροι (one-tailed) έλεγχοι (**δεξιόπλευρος** και **αριστερόπλευρος** αντίστοιχα)

και ο έλεγχος, $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$, ονομάζεται **αμφίπλευρος (two-tailed)**. Σημειώνουμε,

επίσης, ότι τα δύο σύνολα τιμών της παραμέτρου που ελέγχουμε (στο παράδειγμά μας, της μ) που ορίζουν οι δύο υποθέσεις, πρέπει προφανώς να είναι ξένα μεταξύ τους (ή το ένα άρνηση του άλλου). Τέλος, υπογραμμίζουμε ότι και **οι δύο υποθέσεις αναφέρονται στον πληθυσμό** γι' αυτό δηλώνονται με όρους παραμέτρων του πληθυσμού.

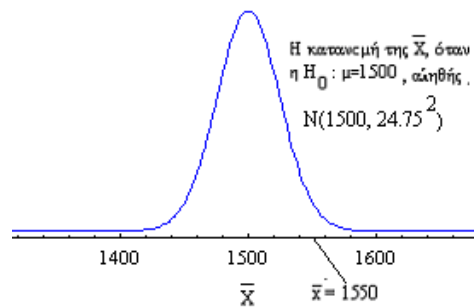
Όπως ήδη έχουμε αναφέρει και στην *Εισαγωγή*, στη στατιστική προσέγγιση προβλημάτων ελέγχεται η συμφωνία θεωρίας και εμπειρίας. Έτσι, στο παράδειγμά μας, αφού διατυπώσαμε την υπόθεση ότι η άγνωστη μέση τιμή του πληθυσμού των ορίων αντοχής των καλωδίων μετά τη βελτίωση των υλικών είναι 1500 Kgr ($H_0 : \mu = 1500 \text{ Kgr}$), παίρνουμε ένα τυχαίο δείγμα καλωδίων από το σύνολο της παραγωγής του εργοστασίου και μετράμε το όριο αντοχής κάθε καλωδίου του δείγματος. Για τις ανάγκες του παραδείγματος, έστω ότι ένα τυχαίο δείγμα X_1, X_2, \dots, X_n μεγέθους $n=50$, μας έδωσε τις μετρήσεις x_1, x_2, \dots, x_{50} με $\bar{x} = 1550 \text{ Kgr}$.

Η «εμπειρία», δηλαδή, αυτό που παρατηρείται στο δείγμα, συμφωνεί άραγε με την υπόθεση $H_0 : \mu = 1500 \text{ Kgr}$, δηλαδή, με ό,τι αυτή συνεπάγεται για το δείγμα (με βάση τη θεωρία πιθανοτήτων) ή μήπως δίνει αποδείξεις εναντίον της H_0 και υπέρ της H_1 .

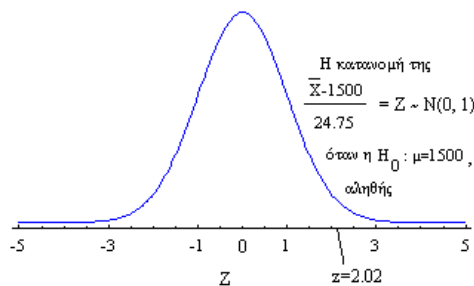
Για να απαντήσουμε σε αυτό το ερώτημα, πρέπει, πρώτα απ' όλα, να κατασκευάσουμε/επιλέξουμε μια κατάλληλη *στατιστική συνάρτηση (statistic)* $T = T(X_1, X_2, \dots, X_n)$, δηλαδή, μια συνάρτηση του δείγματος-δειγματοσυνάρτηση, ώστε να ποσοτικοποιήσουμε «αυτό που παρατηρείται στο δείγμα» και η οποία, υπό την H_0 , δηλαδή όταν ισχύει η H_0 , να ακολουθεί γνωστή κατανομή (χωρίς άγνωστες παραμέτρους) ώστε να μπορούμε να υπολογίσουμε τις απαιτούμενες για τον έλεγχο πιθανότητες.

Στο παράδειγμά μας, που αφορά στον έλεγχο της μέσης τιμής, μ , του πληθυσμού, είναι λογικό να επιλέξουμε ως *στατιστική συνάρτηση T*, τη *δειγματική μέση τιμή* $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ της οποίας η κατανομή, υπό την $H_0 : \mu = 1500 \text{ Kgr}$, είναι

γνωστή, αφού το μέγεθος του δείγματος που πήραμε είναι αρκετά μεγάλο και επομένως από το Κ.Ο.Θ. έχουμε² $\bar{X} \sim N(1500, \frac{175^2}{50})$ ή $\bar{X} \sim N(1500, 24.75^2)$.



Εναλλακτικά, ως στατιστική συνάρτηση T , μπορούμε να επιλέξουμε την $Z = \frac{\bar{X} - 1500}{175/\sqrt{50}} = \frac{(\bar{X} - 1500) \cdot \sqrt{50}}{175} \sim N(0, 1)$.



Έτσι, «αυτό που παρατηρείται στο δείγμα», στο παράδειγμά μας εκφράζεται από τη στατιστική συνάρτηση \bar{X} με τιμή, στο συγκεκριμένο δείγμα που πήραμε, $\bar{x} = 1550 \text{ Kgr}$ ή, ισοδύναμα, από την $Z = \frac{(\bar{X} - 1500) \cdot \sqrt{50}}{175}$, με τιμή $z = \frac{(1550 - 1500) \cdot \sqrt{50}}{175} = 2.02$.

Ας δούμε τώρα πώς – με ποιον κανόνα – ορίζουμε το «ακραίο».

1^{ος} τρόπος: Επιλέγουμε-(προ)καθορίζουμε το ανεκτό μέγεθος σφάλματος τύπου I

Αν η $H_0 : \mu = 1500 \text{ Kgr}$ είναι αληθής, είναι λογικό να αναμένουμε ότι η τιμή της στατιστικής συνάρτησης \bar{X} στο δείγμα που πήραμε (δηλαδή, η μέση τιμή του δείγματος) θα είναι κοντά στην τιμή 1500 Kgr . Αντιθέτως, αν η $H_0 : \mu = 1500 \text{ Kgr}$ δεν είναι αληθής, αναμένουμε η μέση τιμή του δείγματος να είναι μακριά (προς την κατεύθυνση της H_1 , δηλαδή δεξιότερα) του 1500 . Ένας λογικός, επομένως, έλεγχος είναι ο εξής: ορίζουμε μια τιμή c με βάση την οποία θα κρίνεται αν η δειγματική μέση τιμή βρίσκεται μακριά από την $\mu = 1500 \text{ Kgr}$, δηλαδή θα θεωρείται ακραία. Έτσι, αν στο παράδειγμά μας επιλέξουμε $c = 1530 \text{ Kgr}$ τότε επειδή $\bar{x} = 1550 > 1530$, αυτό που παρατηρείται στο δείγμα κρίνεται ακραίο και η H_0 απορρίπτεται. Το κριτήριο αυτό είναι φυσικά λογικό, όμως, πόσο λογική-εύλογη είναι η αυθαίρετη τιμή $c = 1530 \text{ Kgr}$ με την οποία οριοθετήσαμε τις ακραίες από της μη ακραίες δειγματικές μέσες τιμές. Αν, για παράδειγμα, επιλέξουμε $c = 1570 \text{ Kgr}$, τότε $\bar{x} = 1550 < 1570$ δηλαδή τώρα το

² Κατά προσέγγιση

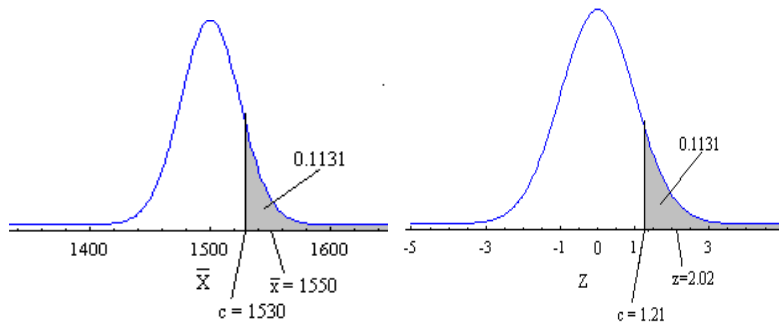
παρατηρούμενο στο δείγμα δεν κρίνεται ακραίο και το δείγμα δεν υποστηρίζει απόρριψη της H_0 . Τίθεται, επομένως, το ερώτημα: πώς επιλέγουμε την τιμή της σταθεράς c ;

Πριν απαντήσουμε σε αυτό το εύλογο ερώτημα, ας υπολογίσουμε την πιθανότητα να κάνουμε σφάλμα τύπου I στην περίπτωση που επιλέξουμε $c = 1530 \text{ Kgr}$ και αντίστοιχα στην περίπτωση που επιλέξουμε $c = 1570 \text{ Kgr}$.

Για $c = 1530 \text{ Kgr}$, έχουμε:

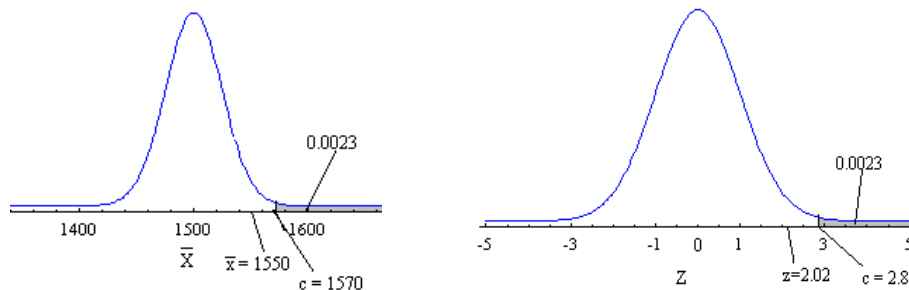
$$P(\text{σφάλμα τύπου } I) = P(\text{απόρριψη της } H_0 \mid \text{αληθής η } H_0) = P(\bar{X} \geq 1530 \mid \mu = 1500) =$$

$$= P\left(\frac{\bar{X} - 1500}{175/\sqrt{50}} \geq \frac{1530 - 1500}{175/\sqrt{50}}\right) = P(Z \geq 1.21) = 1 - \Phi(1.21) = 0.1131$$



Ομοίως, για $c = 1570 \text{ Kgr}$, έχουμε:

$$P(\bar{X} \geq 1570 \mid \mu = 1500) = P\left(\frac{\bar{X} - 1500}{175/\sqrt{50}} \geq \frac{1570 - 1500}{175/\sqrt{50}}\right) = P(Z \geq 2.83) = 1 - \Phi(2.83) = 0.0023$$



Και για οποιοδήποτε c , έχουμε:

$$P(\bar{X} \geq c \mid \mu = 1500) = P\left(\frac{\bar{X} - 1500}{175/\sqrt{50}} \geq \frac{c - 1500}{175/\sqrt{50}}\right) = P(Z \geq \frac{(c - 1500)\sqrt{50}}{175}) = 1 - \Phi\left(\frac{(c - 1500)\sqrt{50}}{175}\right)$$

Από τα παραπάνω, είναι φανερό ότι η τιμή της σταθεράς c επηρεάζει (ακριβέστερα, καθορίζει) την πιθανότητα σφάλματος τύπου I που κάνουμε. Έτσι, με κριτήριο τον έλεγχο του μεγέθους του σφάλματος τύπου I (θυμηθείτε και πώς ορίζουμε την H_0), μπορούμε να επιλέξουμε την τιμή της c ως εξής:

Ορίζουμε ένα **μέγιστο ανεκτό μέγεθος σφάλματος τύπου I** και με βάση αυτό υπολογίζουμε την τιμή της c . Με αυτό τον τρόπο, καθορίζουμε έναν απολύτως σαφή κανόνα για να κρίνουμε αν αυτό που παρατηρείται στο δείγμα, δηλαδή η τιμή της στατιστικής συνάρτησης T (στο παράδειγμά μας, της \bar{X} ή της $Z = \frac{(\bar{X} - 1500) \cdot \sqrt{50}}{175}$), είναι

«ακραία» ή όχι, και πλέον, **αποφασίζουμε για την απόρριψη ή τη μη απόρριψη της H_0 , με κριτήριο ένα προκαθορισμένο μέγεθος σφάλματος τύπου I.**

Το ανεκτό επίπεδο σφάλματος τύπου I που προκαθορίζουμε, συμβολίζεται με α και ονομάζεται **επίπεδο σημαντικότητας (level of significance)** του ελέγχου (γιατί από αυτό προκύπτει η τιμή της c που ορίζει αν αυτό που παρατηρείται στο δείγμα είναι σημαντικό-σημαντική απόδειξη για να υποστηρίξει την απόρριψη της H_0). Συνήθως το επίπεδο σημαντικότητας, α , ορίζεται ίσο με 0.01 ή 0.05.

Ας ολοκληρώσουμε τον έλεγχο, στο παράδειγμά μας, θέτοντας **επίπεδο σημαντικότητας $\alpha = 0.05$** . Πρέπει να επιλέξουμε τιμή c τέτοια ώστε:

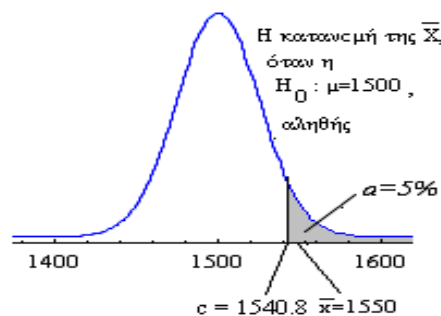
$$P(\bar{X} \geq c \mid \mu = 1500) \leq 0.05 \Leftrightarrow P\left(\frac{\bar{X} - 1500}{175/\sqrt{50}} \geq \frac{c - 1500}{175/\sqrt{50}}\right) \leq 0.05$$

$$\Leftrightarrow P\left(Z \geq \frac{(c - 1500)\sqrt{50}}{175}\right) \leq 0.05 \Leftrightarrow 1 - \Phi\left(\frac{(c - 1500)\sqrt{50}}{175}\right) \leq 0.05$$

$$\Leftrightarrow \Phi\left(\frac{(c - 1500)\sqrt{50}}{175}\right) \geq 0.95 \Leftrightarrow \frac{(c - 1500)\sqrt{50}}{175} \geq z_{0.05} = 1.645$$

$$\Leftrightarrow c \geq 1500 + 1.645 \frac{175}{\sqrt{50}} = 1540.8$$

Έτσι, επιλέγοντας $c = 1540.8$ έχουμε $\bar{x} = 1550 > 1540.8$ και επομένως απορρίπτουμε την H_0 με πιθανότητα εσφαλμένης απόρριψης το πολύ 5%.

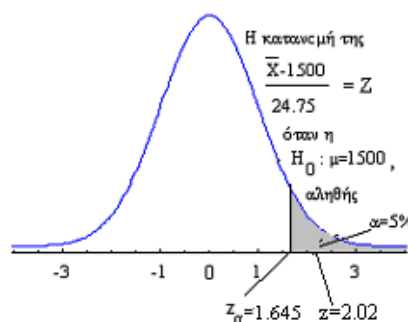


Ισοδύναμα, αν ως **στατιστική συνάρτηση** επιλέξουμε την $Z = \frac{(\bar{X} - 1500) \cdot \sqrt{50}}{175}$, έχουμε:

$P(Z \geq c) \leq 0.05 \Rightarrow c = z_{0.05} = 1.645$, δηλαδή, ως τιμή της c επιλέγουμε το $\alpha = 0.05$ άνω ποσοστιαίο σημείο της τυποποιημένης κανονικής κατανομής, $z_{0.05}$, και επειδή η

τιμή της **στατιστικής συνάρτησης** στο δείγμα, $z = \frac{(1550 - 1500) \cdot \sqrt{50}}{175} = 2.02$, είναι

μεγαλύτερη από την $c = z_{0.05} = 1.645$, δηλαδή $z = 2.02 > z_{0.05} = 1.645$, απορρίπτουμε την H_0 με πιθανότητα εσφαλμένης απόρριψης το πολύ 5%.



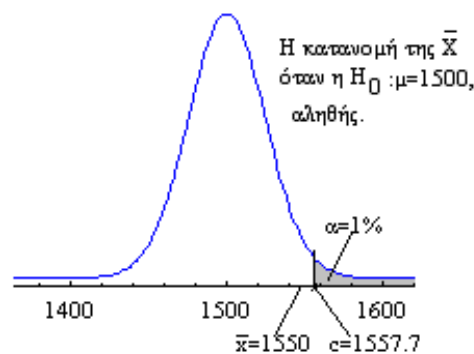
Αν η φύση του προβλήματος που εξετάζουμε επιβάλλει μεγαλύτερη «προστασία» από σφάλμα τύπου I, δηλαδή από εσφαλμένη απόρριψη της H_0 , τότε πρέπει να είμαστε πιο «συντηρητικοί» στην απόρριψη της H_0 και αυτό το επιτυγχάνουμε καθορίζοντας μικρότερο ανεκτό επίπεδο σφάλματος τύπου I, δηλαδή, μικρότερο επίπεδο σημαντικότητας. Έτσι, στο παράδειγμά μας, αν επιβάλλεται πιο αυστηρός έλεγχος του ισχυρισμού του εργοστασίου, κάνουμε τον έλεγχο σε μικρότερο επίπεδο σημαντικότητας, δηλαδή, κάνουμε τον έλεγχο με μικρότερη ανοχή σε εσφαλμένη απόρριψη της H_0 , π.χ. με $\alpha = 0.01$. Στην περίπτωση αυτή έχουμε:

$$P(\bar{X} \geq c \mid \mu = 1500) \leq 0.01 \Leftrightarrow P\left(\frac{\bar{X} - 1500}{175/\sqrt{50}} \geq \frac{c - 1500}{175/\sqrt{50}}\right) \leq 0.01$$

$$\Leftrightarrow P\left(Z \geq \frac{(c - 1500)\sqrt{50}}{175}\right) \leq 0.01 \Leftrightarrow 1 - \Phi\left(\frac{(c - 1500)\sqrt{50}}{175}\right) \leq 0.01$$

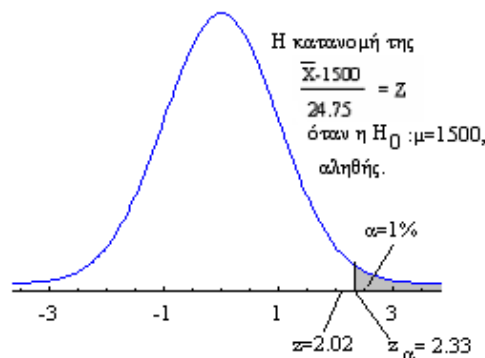
$$\Leftrightarrow \Phi\left(\frac{(c - 1500)\sqrt{50}}{175}\right) \geq 0.99 \Leftrightarrow \frac{(c - 1500)\sqrt{50}}{175} \geq z_{0.01} = 2.33 \Leftrightarrow c \geq 1557.7$$

Έτσι, για $\alpha = 0.01$ είναι $c = 1557.7$ και επειδή η $\bar{x} = 1550$ δεν είναι μεγαλύτερη από αυτή, σε επίπεδο σημαντικότητας $\alpha = 0.01$ δεν απορρίπτουμε την H_0 .



Ισοδύναμα, αν ως *στατιστική συνάρτηση* επιλέξουμε την $Z = \frac{(\bar{X} - 1500) \cdot \sqrt{50}}{175}$, έχουμε:

$P(Z \geq c) \leq 0.01 \Rightarrow c = z_{0.01} = 2.33$ και επειδή η τιμή της *στατιστικής συνάρτησης* στο δείγμα, $z = 2.02$, δεν είναι μεγαλύτερη από την $c = z_{0.01} = 2.33$, σε επίπεδο σημαντικότητας $\alpha = 0.01$ δεν απορρίπτουμε την H_0 .



Η σταθερά c ονομάζεται **κρίσιμη τιμή ή όριο απόρριψης (critical value, rejection limit)** γιατί με βάση αυτή κρίνεται αν μια τιμή της *στατιστικής συνάρτησης*, T , είναι ακραία ή όχι. Ανάλογα, η *στατιστική συνάρτηση* T , ονομάζεται **στατιστική**

συνάρτηση ελέγχου (test statistic) και οι τιμές της για τις οποίες απορρίπτεται η H_0 ορίζουν την **κρίσιμη περιοχή ή περιοχή απόρριψης (critical region, rejection region)**. Όταν απορρίπτεται η H_0 , το δείγμα χαρακτηρίζεται **στατιστικά σημαντικό (statistically significant)** και έχει την έννοια ότι διαφέρει σημαντικά από αυτό που αναμενόταν από την H_0 .

Στο παράδειγμά μας, σε επίπεδο σημαντικότητας $\alpha = 0.05$, η κρίσιμη τιμή είναι $c = 1540.8$ ή, ισοδύναμα, $c = z_{0.05} = 1.645$. Η κρίσιμη περιοχή ή περιοχή απόρριψης

είναι $C = \{\bar{x} : \bar{x} \geq 1500 + z_{0.05} \frac{175}{\sqrt{50}} = 1540.8\} = [1540.8, +\infty)$ ή, ισοδύναμα,

$C = \{z : z = \frac{(\bar{x} - 1500) \cdot \sqrt{50}}{175} \geq z_{0.05} = 1.645\} = [1.645, +\infty)$ και τα ευρήματα στο δείγμα ($\bar{x} = 1550 \text{ kgr}$ ή, ισοδύναμα, $z = 2.02$), σε επίπεδο σημαντικότητας $\alpha = 0.05$, είναι στατιστικά σημαντικά.

Σχόλιο. Επισημαίνουμε ότι, θέτοντας μικρότερο επίπεδο σημαντικότητας, απαιτούμε πιο «σημαντικές αποδείξεις» για την απόρριψη της H_0 και τον χαρακτηρισμό των ευρημάτων μας στο δείγμα ως στατιστικά σημαντικών. Έτσι, μπορεί, σε κάποιο επίπεδο σημαντικότητας α , π.χ. $\alpha = 0.05$, να απορρίπτουμε την H_0 και σε κάποιο μικρότερο, π.χ. $\alpha = 0.01$, να μην την απορρίπτουμε γιατί απαιτούμε σημαντικότερες αποδείξεις. Όσο πιο μικρό είναι το επίπεδο σημαντικότητας στο οποίο μπορούμε να απορρίψουμε την H_0 , τόσο πιο σημαντική είναι η τιμή της στατιστικής συνάρτησης ελέγχου που παρατηρείται στο δείγμα, με την έννοια ότι δίνει πιο ισχυρές αποδείξεις εναντίον της H_0 . Άρα, όσο **πιο μικρό** είναι το επίπεδο σημαντικότητας στο οποίο μπορούμε να απορρίψουμε την H_0 , τόσο **πιο σημαντικό**, στατιστικά, είναι το αποτέλεσμα του ελέγχου. Τέλος, είναι προφανές, ότι αν η H_0 απορρίπτεται σε κάποιο επίπεδο σημαντικότητας α , τότε επίσης απορρίπτεται σε οποιοδήποτε μεγαλύτερο, ενώ αν δεν απορρίπτεται σε κάποιο επίπεδο σημαντικότητας α , τότε επίσης δεν απορρίπτεται σε οποιοδήποτε μικρότερο.

Σημείωση: Ας δούμε, και με μια άλλη διατύπωση, τι σημαίνει «κάνω σφάλμα τύπου I». Έστω ότι κάνω τον έλεγχο σε επίπεδο σημαντικότητας α και ότι η μηδενική υπόθεση είναι αληθής. Τότε, από όλα τα δείγματα μεγέθους n που μπορώ να πάρω από τον πληθυσμό, ποσοστό το πολύ α από αυτά θα δώσουν τιμή της στατιστικής συνάρτησης ελέγχου που απορρίπτει τη μηδενική υπόθεση (εν προκειμένω εσφαλμένα).

Συνοψίζοντας, ο έλεγχος του παραδείγματός μας, με τη διαδικασία που περιγράψαμε, έγινε σε έξι βήματα:

1^ο Βήμα

Ορίσαμε τις δύο υποθέσεις (σύμφωνα με όσα αναφέρθηκαν):

$$H_0 : \mu = 1500 \text{ Kgr}, H_1 : \mu > 1500 \text{ Kgr}$$

2^ο Βήμα

Ορίσαμε το επίπεδο σημαντικότητας α του ελέγχου: $\alpha = 0.05$

3^ο Βήμα

Ορίσαμε τη στατιστική συνάρτηση ελέγχου: την \bar{X} ή, ισοδύναμα,

$$\text{την } Z = \frac{(\bar{X} - 1500) \cdot \sqrt{50}}{175}.$$

4^ο Βήμα

Επιλέξαμε από τον πληθυσμό ένα τυχαίο δείγμα και υπολογίσαμε την τιμή της στατιστικής συνάρτησης ελέγχου: $\bar{x} = 1550 \text{ Kgr}$ ή, ισοδύναμα, $z = 2.02$ (μέγεθος δείγματος, $n = 50$).

5^ο Βήμα

Ορίσαμε την κρίσιμη περιοχή (περιοχή απόρριψης) του ελέγχου:

$$C = \{\bar{x} : \bar{x} \geq 1500 + \frac{175}{\sqrt{50}} \cdot z_{0.05}\} = \{\bar{x} : \bar{x} \geq 1500 + \frac{175}{\sqrt{50}} \cdot 1.645\} = [1540.8, +\infty),$$

ή, ισοδύναμα,

$$C = \{z : z = \frac{\bar{x} - 1500}{175/\sqrt{50}} \geq z_{0.05} = 1.645\} = [1.645, +\infty).$$

6^ο Βήμα

Εξετάσαμε αν η τιμή της στατιστικής συνάρτησης ελέγχου βρίσκεται ή όχι στην κρίσιμη περιοχή (περιοχή απόρριψης) του ελέγχου και αποφασίσαμε με πιθανότητα σφάλματος τύπου I, $\alpha = 0.05$, για την απόρριψη ή όχι της μηδενικής υπόθεσης: επειδή $\bar{x} = 1550 \in [1540.8, +\infty)$ ή, ισοδύναμα, επειδή $z = 2.02 \in [1.645, +\infty)$, σε επίπεδο σημαντικότητας $\alpha = 0.05$, απορρίψαμε την $H_0 : \mu = 1500 \text{ Kgr}$.

Προσοχή: Στη διατύπωση του αποτελέσματος πρέπει οπωσδήποτε να αναφέρεται το επίπεδο σημαντικότητας του ελέγχου γιατί με βάση αυτό κρίνεται αν αυτό που παρατηρείται στο δείγμα είναι στατιστικά σημαντικό ή όχι και κατά συνέπεια αν η μηδενική υπόθεση απορρίπτεται ή δεν απορρίπτεται. Στη διατύπωση του αποτελέσματος θα αναφερθούμε και στη συνέχεια.

Σημείωση: Διευκρινίζουμε ότι όταν λέμε «περιοχή απόρριψης», πάντοτε εννοούμε «περιοχή απόρριψης της H_0 ».

Όπως, ήδη, έχουμε αναφέρει, με αυτόν τον τρόπο που εργασθήκαμε, πετύχαμε να θέσουμε υπό τον έλεγχό μας το σφάλμα τύπου I, δηλαδή, να αποφασίσουμε με γνωστή-προκαθορισμένη πιθανότητα εσφαλμένης απόρριψης της H_0 . Ένας παρεμφερής τρόπος χειρισμού του σφάλματος τύπου I είναι ο ακόλουθος.

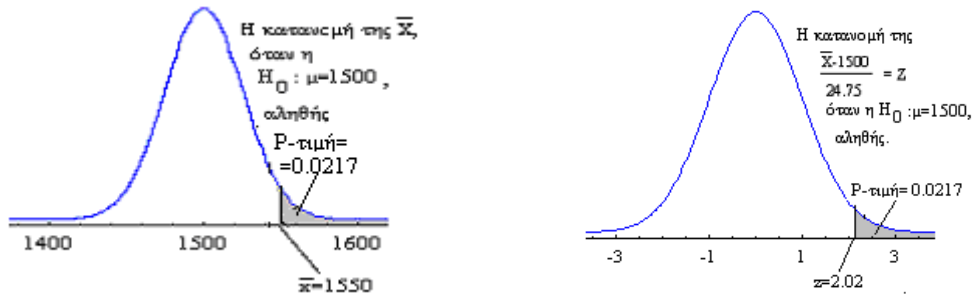
2^{ος} τρόπος: Υπολογίζουμε την P-Τιμή (P-Value) του δείγματος

Με δεδομένο ότι η $H_0 : \mu = 1500$ είναι αληθής, υπολογίζουμε την πιθανότητα να εμφανισθεί η τιμή $\bar{x} = 1550 \text{ Kgr}$ που εμφανίστηκε στο δείγμα ή κάποια μεγαλύτερη της (δηλαδή, κάποια τιμή προς την κατεύθυνση της H_1). Ζητάμε την πιθανότητα $P(\bar{X} \geq 1550 / H_0)$ ή $P(\bar{X} \geq 1550 / \mu = 1500)$ και επειδή γνωρίζουμε την κατανομή της \bar{X} έχουμε,

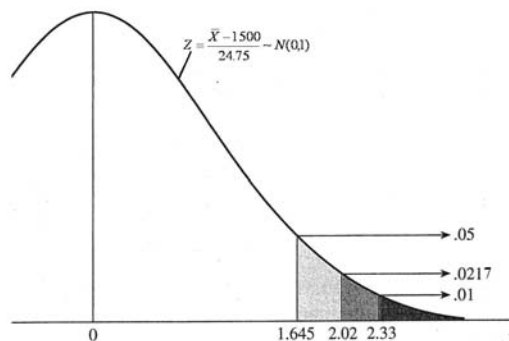
$$P(\bar{X} \geq 1550 | \mu = 1500) = P\left(\frac{\bar{X} - 1500}{175/\sqrt{50}} \geq \frac{1550 - 1500}{175/\sqrt{50}}\right) = P(Z \geq 2.02) = 1 - \Phi(2.02) = 0.0217$$

Αυτή η πιθανότητα ονομάζεται **P-Τιμή (P-Value)** του δείγματος και είναι η πιθανότητα να εμφανισθεί η τιμή της στατιστικής συνάρτησης ελέγχου που εμφανίστηκε (στο παράδειγμά μας, $\bar{x} = 1550 \text{ Kgr}$ ή $z = 2.02$) ή κάποια πιο μακριά (πιο ακραία), προς την κατεύθυνση της H_1 , **δεδομένου ότι η H_0 είναι αληθής**. Έτσι,

υπολογίζοντας την *P*-τιμή του δείγματος, γνωρίζουμε πόσο πιθανή ήταν η εμφάνιση του δείγματος που πήραμε με την υπόθεση ότι η H_0 είναι αληθής. Επομένως, **όσο πιο μικρή είναι η *P*-Τιμή** τόσο ισχυρότερες ενδείξεις **εναντίον** της H_0 προκύπτουν από το συγκεκριμένο τυχαίο δείγμα ή αλλιώς **τόσο πιο σημαντική είναι η τιμή της στατιστικής συνάρτησης ελέγχου** που δίνει το δείγμα. Στο παράδειγμά μας, υπολογίσαμε ότι η *P*-Τιμή του δείγματος που πήραμε, είναι ίση με 0.0217 ή 2.17%.



Επομένως, αν κάνουμε τον έλεγχο σε *επίπεδο σημαντικότητας* $\alpha = 0.01 = 1\%$, δηλαδή, αν θέλουμε πιο «σημαντικές αποδείξεις» εναντίον της H_0 από αυτές που παρατηρούνται στο δείγμα, τότε δεν την απορρίπτουμε, ενώ αν κάνουμε τον έλεγχο σε *επίπεδο σημαντικότητας* $\alpha = 0.05 = 5\%$ τότε την απορρίπτουμε (γιατί στην περίπτωση αυτή, απαιτούμε λιγότερο «σημαντικές αποδείξεις» εναντίον της H_0)³. Στο επόμενο σχήμα έχουμε μεγεθύνει τη δεξιά ουρά της κατανομής της Z και φαίνονται ευκρινώς οι περιοχές που αντιστοιχούν στο $\alpha = 0.05$, στην P -τιμή = 0.0217 και στο $\alpha = 0.01$



Έτσι, υπολογίζοντας την *P*-Τιμή, μπορούμε άμεσα να την συγκρίνουμε με όποιο α και αν επιλέξουμε και να αποφασίσουμε για την απόρριψη ή όχι της H_0 . Ειδικότερα,

- αν $P\text{-Τιμή} \leq \alpha$, τότε, σε επίπεδο σημαντικότητας α , η H_0 **απορρίπτεται**.
- αν $P\text{-Τιμή} > \alpha$, τότε, σε επίπεδο σημαντικότητας α , η H_0 **δεν απορρίπτεται**.

Συνοψίζοντας, από τα παραπάνω, είναι προφανές, ότι

1. Η *P*-τιμή μπορεί να ορισθεί και ως εξής: ***P*-Τιμή είναι η ελάχιστη τιμή του επιπέδου σημαντικότητας για την οποία απορρίπτεται η H_0 .**
2. Η *P*-τιμή είναι ένα μέτρο το οποίο εκφράζει πόσο ισχυρές είναι οι αποδείξεις που προκύπτουν από το δείγμα, εναντίον της H_0 .

Σημείωση: Στη βιβλιογραφία, για την *P*-Τιμή, χρησιμοποιείται και ο όρος, παρατηρούμενο επίπεδο σημαντικότητας (*observed significance level*). Τον αναφέρουμε, όμως, δεν τον συνιστούμε.

³ Θυμηθείτε ότι μικρότερο α σημαίνει ότι απαιτούνται πιο σημαντικές αποδείξεις εναντίον της H_0 .

Σε αυτή την ενότητα, προσπαθήσαμε, με ένα παράδειγμα, να περιγράψουμε, να εφαρμόσουμε και κυρίως να αναδείξουμε το νόημα και τη λογική της γενικής διαδικασίας *στατιστικού ελέγχου υποθέσεων*. Βέβαια, στο παράδειγμα που χρησιμοποιήσαμε, ο έλεγχος είναι ένας *μονόπλευρος, δεξιόπλευρος έλεγχος για τη μέση τιμή, μ* , ενός πληθυσμού του οποίου γνωρίζουμε τη *διασπορά, σ^2* , και το τυχαίο δείγμα που χρησιμοποιήσαμε είναι αρκετά μεγάλο ώστε η προσέγγιση που παίρνουμε από το Κ.Ο.Θ. για την κατανομή της *στατιστικής συνάρτησης ελέγχου* να είναι ικανοποιητική. Δηλαδή, είναι μια ειδική-συγκεκριμένη περίπτωση ελέγχου για τη *μέση τιμή* ενός πληθυσμού. Όμως, η μέθοδος που αναλύσαμε είναι γενική. Δεν αλλάζει αν, αντί *μονόπλευρος*, ο έλεγχος είναι *αμφίπλευρος* ή αντί στη *μέση τιμή, μ* , αφορά στη *διασπορά, σ^2* , ενός πληθυσμού, ή αν το δείγμα είναι αρκετά μεγάλο ή όχι, ή αντί στη *μέση τιμή* ενός πληθυσμού αφορά στη διαφορά $\mu_1 - \mu_2$ των μέσων τιμών μ_1, μ_2 δύο πληθυσμών, κ.ο.κ. Οι διάφορες περιπτώσεις *στατιστικών ελέγχων* διαφοροποιούνται, ή στην επιλογή *στατιστικής συνάρτησης ελέγχου* ή/και στη μορφή της *περιοχής απόρριψης* ($[c, +\infty)$ ή $(-\infty, c]$ ή $(-\infty, c_1] \cup [c_2, +\infty)$, αντίστοιχα).

Στην επόμενη ενότητα, δίνουμε τη *στατιστική συνάρτηση ελέγχου* και την *περιοχή απόρριψης* για διάφορες περιπτώσεις που μπορεί να εμφανισθούν στον *στατιστικό έλεγχο* της μέσης τιμής, μ , ενός πληθυσμού.

2. Στατιστικοί έλεγχοι υποθέσεων για τη μέση τιμή, μ , ενός πληθυσμού

Θα αναφερθούμε στον έλεγχο της υπόθεσης, $H_0 : \mu = \mu_0$, δηλαδή, της υπόθεσης ότι η άγνωστη μέση τιμή, μ , ενός πληθυσμού έχει τιμή μ_0 .

Ειδικότερα, θα δώσουμε τη *στατιστική συνάρτηση ελέγχου* στις ακόλουθες περιπτώσεις **(α)**, **(β)** και **(γ)** όπου ο πληθυσμός του οποίου ελέγχουμε τη μέση τιμή ακολουθεί αντίστοιχα,

- κανονική κατανομή με γνωστή διασπορά
- κανονική κατανομή με άγνωστη διασπορά
- οποιαδήποτε κατανομή με άγνωστη διασπορά και το μέγεθος δείγματος είναι μεγάλο.

(α) Ο πληθυσμός ακολουθεί κανονική κατανομή με γνωστή διασπορά.

Έστω τυχαίο δείγμα X_1, X_2, \dots, X_n από ένα πληθυσμό που ακολουθεί κανονική κατανομή με γνωστή διασπορά σ^2 και μέση τιμή $\mu = \mu_0$ (ελεγχόμενη), δηλαδή, $X_i \sim N(\mu_0, \sigma^2)$, $i = 1, 2, \dots, n$. Από τη θεωρία πιθανοτήτων γνωρίζουμε ότι στην περίπτωση αυτή, η κατανομή της τυχαίας μεταβλητής $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$, ανεξαρτήτως του μεγέθους του δείγματος, είναι κανονική κατανομή με $\bar{X} \sim N(\mu_0, \frac{\sigma^2}{n})$ και επομένως $\frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} = Z \sim N(0, 1)$.

Επειδή η *διασπορά, σ^2* , του πληθυσμού είναι γνωστή, στην $Z = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}$ δεν υπάρχουν άγνωστοι παράμετροι και επομένως η τιμή της, z , μπορεί να υπολογισθεί από το δείγμα. Έτσι, εργαζόμενοι όπως στο παράδειγμά μας, αν \bar{x} η τιμή της \bar{X} στο δείγμα, έχουμε:

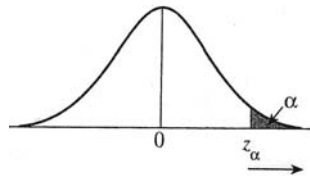
Σε επίπεδο σημαντικότητας α , απορρίπτουμε την $H_0 : \mu = \mu_0$,

- έναντι της $H_1 : \mu > \mu_0$, όταν,

$$\bar{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha,$$

ή, ισοδύναμα, όταν,

$$z = \frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} \geq z_\alpha$$

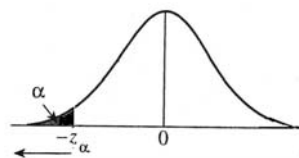


- έναντι της $H_1 : \mu < \mu_0$, όταν,

$$\bar{x} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_\alpha,$$

ή, ισοδύναμα, όταν,

$$z = \frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} \leq -z_\alpha$$

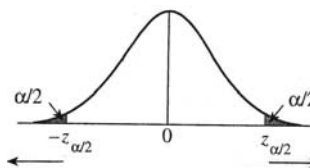


- έναντι της $H_1 : \mu \neq \mu_0$, όταν,

$$\bar{x} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \quad \text{ή} \quad \bar{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$$

ή, ισοδύναμα, όταν,

$$|z| = \frac{|\bar{x} - \mu_0| \sqrt{n}}{\sigma} \geq z_{\alpha/2}$$



Σημείωση: Στην περίπτωση που η διασπορά, σ^2 , του πληθυσμού είναι γνωστή και το μέγεθος του δείγματος, n , είναι μεγάλο (θεωρητικά $n \rightarrow +\infty$, στην πράξη $n \geq 30$), οι παραπάνω περιοχές απόρριψης ισχύουν για οποιονδήποτε πληθυσμό, όχι κατ' ανάγκη κανονικό (βλ. Κ.Ο.Θ.). Όμως, στην περίπτωση αυτή, οι αντίστοιχοι έλεγχοι είναι κατά προσέγγιση επιπέδου σημαντικότητας α , γιατί η κατανομή της στατιστικής συνάρτησης

ελέγχου \bar{X} ή, ισοδύναμα, της $Z = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}$ **δεν είναι**, στην περίπτωση αυτή,

κανονική αλλά προσεγγίζεται από την κανονική. Φυσικά, όσο μεγαλύτερο είναι το δείγμα, τόσο καλύτερη είναι η προσέγγιση.

Η υπόθεση που κάναμε ότι η διασπορά, σ^2 , του πληθυσμού είναι γνωστή, δεν είναι μια ρεαλιστική υπόθεση. Στην πράξη, η διασπορά του πληθυσμού συνήθως είναι άγνωστη. Οι δύο περιπτώσεις που ακολουθούν, αναφέρονται στο πώς εργαζόμαστε όταν η διασπορά του πληθυσμού είναι άγνωστη.

(β) Ο πληθυσμός ακολουθεί κανονική κατανομή με άγνωστη διασπορά.

Έστω τυχαίο δείγμα X_1, X_2, \dots, X_n από ένα πληθυσμό που ακολουθεί κανονική κατανομή με άγνωστη διασπορά σ^2 και μέση τιμή $\mu = \mu_0$ (ελεγχόμενη), δηλαδή, $X_i \sim N(\mu_0, \sigma^2)$, $i = 1, 2, \dots, n$.

Επειδή η διασπορά του πληθυσμού, σ^2 , είναι άγνωστη, δε μπορούμε ως στατιστική συνάρτηση ελέγχου να χρησιμοποιήσουμε την $Z = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}$ γιατί δε μπορούμε να υπολογίσουμε την τιμή της, z , από το δείγμα. Γι' αυτό, εκτιμάμε την άγνωστη διασπορά, σ^2 , από την (αμερόληπτη) δειγματική διασπορά $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ και ως στατιστική συνάρτηση ελέγχου, χρησιμοποιούμε την $T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}$, η οποία, είναι γνωστό ότι όταν $X_i \sim N(\mu_0, \sigma^2)$, $i = 1, 2, \dots, n$, και ανεξαρτήτως του μεγέθους του δείγματος, ακολουθεί την κατανομή t_{n-1} (την κατανομή Student με $n-1$ βαθμούς ελευθερίας). Δηλαδή,

$$T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S} \sim t_{n-1}.$$

Είναι επομένως λογικό, στην περίπτωση που εξετάζουμε, οι περιοχές απόρριψης να ορίζονται με βάση το άνω α ή το άνω $\alpha/2$ ποσοστιαίο σημείο της κατανομής t_{n-1} ($t_{n-1;\alpha}$ και $t_{n-1;\alpha/2}$, αντίστοιχα). Έτσι, έχουμε:

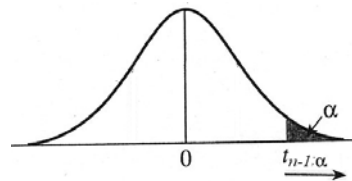
Σε επίπεδο σημαντικότητας α , απορρίπτουμε την $H_0 : \mu = \mu_0$,

- έναντι της $H_1 : \mu > \mu_0$, όταν,

$$\bar{x} \geq \mu_0 + \frac{S}{\sqrt{n}} t_{n-1;\alpha}$$

ή, ισοδύναμα, όταν,

$$t = \frac{(\bar{x} - \mu_0)\sqrt{n}}{S} \geq t_{n-1;\alpha}$$

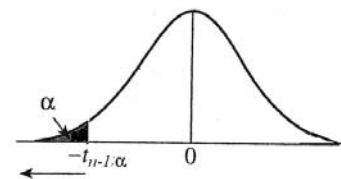


- έναντι της $H_1 : \mu < \mu_0$, όταν,

$$\bar{x} \leq \mu_0 - \frac{S}{\sqrt{n}} t_{n-1;\alpha}$$

ή, ισοδύναμα, όταν,

$$t = \frac{(\bar{x} - \mu_0)\sqrt{n}}{S} \leq -t_{n-1;\alpha}$$

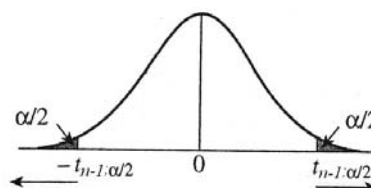


- έναντι της $H_1 : \mu \neq \mu_0$, όταν,

$$\bar{x} \leq \mu_0 - \frac{S}{\sqrt{n}} t_{n-1;\alpha/2} \quad \text{ή} \quad \bar{x} \geq \mu_0 + \frac{S}{\sqrt{n}} t_{n-1;\alpha/2}$$

ή, ισοδύναμα, όταν,

$$|t| = \frac{|(\bar{x} - \mu_0)|\sqrt{n}}{S} \geq t_{n-1;\alpha/2}$$



Στο σημείο αυτό, αξίζει να αναφέρουμε λίγα στοιχεία για την προέλευση της κατανομής *Student*.

Η κατανομή t ή κατανομή *Student*, προτάθηκε το 1908 από τον *William Gosset*. Στη σχετική εργασία που δημοσίευσε ("The probable error of a mean", *Biometrika*, 6, 1-25, 1908), για λόγους εμπορικού απορρήτου, χρησιμοποίησε το ψευδώνυμο **Student**. Ο *Gosset* είχε σπουδάσει Χημεία και Μαθηματικά και εργαζόταν στο Δουβλίνο για τη γνωστή ζυθοποιία *Guinness* σε θέματα στατιστικής συμπερασματολογίας. Τα προβλήματα στατιστικής συμπερασματολογίας που μελετούσε, αφορούσαν πειράματα στο κτήμα της ζυθοποιίας για τη **βελτίωση ποικιλιών κριθαριού!!!** Αφειρητία της ιδέας του *Gosset* ήταν μια απλή παρατήρηση. Επειδή ήταν αναγκασμένος να εργάζεται με

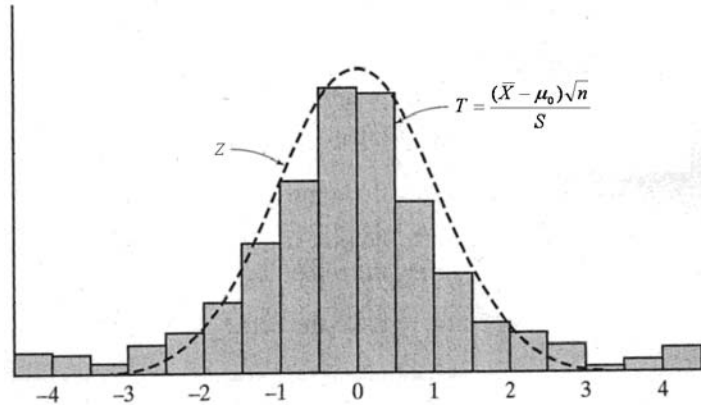
μικρού μεγέθους δείγματα, παρατήρησε ότι η κατανομή της $T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}$ για μικρά

δείγματα, έχει μεν κωδωνοειδή μορφή και είναι συμμετρική ως προς τον κατακόρυφο άξονα στο 0 όπως η $Z \sim N(0, 1)$, όμως έχει πιο «παχιές» ουρές. Παρατήρησε, δηλαδή, ότι για μικρά δείγματα, η $Z \sim N(0, 1)$, για πολύ μικρές και πολύ μεγάλες τιμές της

$T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}$ προβλέπει μικρότερες πιθανότητες από αυτές με τις οποίες συμβαίνουν. Δείτε το σχήμα που ακολουθεί, όπου φαίνεται η $Z \sim N(0, 1)$ και το

ιστόγραμμα 500 τιμών της $T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}$ που υπολογίσθηκαν από 500 δείγματα

μεγέθους $n = 4$ το καθένα από μια κανονική κατανομή. Η παρατήρηση του *Gosset* είναι προφανής.

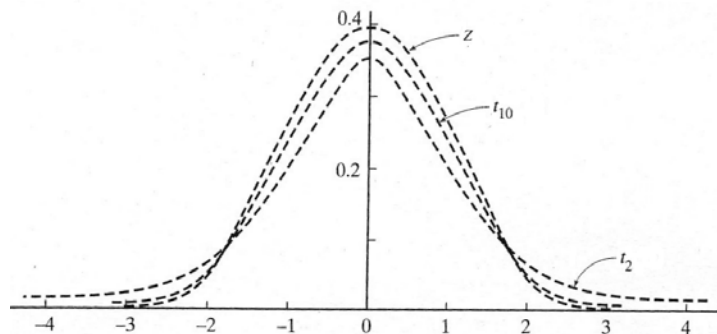


Παρατηρείστε, επίσης, στο παρακάτω σχήμα, την $Z \sim N(0, 1)$ και την t_n για $n = 2$ και $n = 10$. Η t_n είναι πιο πεπλατυσμένη (πιο χαμηλή κορυφή και πλησιάζει τον οριζόντιο άξονα πιο αργά). Επίσης, παρατηρείστε ότι η μορφή της επηρεάζεται από το μέγεθος του δείγματος n . Η t_n παρουσιάζει αυτή τη μεταβλητότητα ως προς το n , γιατί

η διασπορά της, $V(t_n) = \frac{n}{n-2}$, $n > 2$, εξαρτάται από το n . Γι'αυτό, όσο το n

αυξάνεται, η μεταβλητότητά της ελαττώνεται και πλησιάζει την $Z \sim N(0, n/(n-2))$ ⁴.

⁴ Σκεφθείτε επίσης ότι η t_n επηρεάζεται από δύο τυχαίες ποσότητες (\bar{X} και S) ενώ η $Z = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}$ μόνο από την \bar{X} . Όσο το n αυξάνεται η εκτίμηση της σ από την S γίνεται καλύτερη γιατί χρησιμοποιείται περισσότερη πληροφορία!



Η αξία και η χρησιμότητα της κατανομής t στον έλεγχο στατιστικών υποθέσεων αναγνωρίστηκε και αναδείχθηκε, πολλά χρόνια αργότερα, από τον διακεκριμένο στατιστικό και γενετιστή R. A. Fisher ("Applications of "Student's" distribution", *Metron* **5**: 90–104, 1925). Έκτοτε, η κατανομή t είναι γνωστή και ως κατανομή Student (Student distribution) και οι σχετικοί έλεγχοι στατιστικών υποθέσεων ως t -tests.

Σημείωση: Σημειώνουμε, τέλος, ότι παρότι το t -test προϋποθέτει να είναι κανονικός ο πληθυσμός του οποίου ελέγχουμε τη μέση τιμή και από τον οποίο παίρνουμε το δείγμα, εντούτοις, στην πράξη αποδεικνύεται «ανθεκτικό» σε αυτή την υπόθεση. Δηλαδή, το επίπεδο σημαντικότητας του ελέγχου είναι κοντά στο α ακόμη και αν η υπόθεση της κανονικότητας του πληθυσμού δεν ικανοποιείται. Φυσικά, αυτό δε συμβαίνει αν η κατανομή του πληθυσμού απέχει δραματικά από την κανονική κατανομή (σοβαρή ασυμμετρία, πολυκόρυφη κ.τ.λ.) και το μέγεθος του δείγματος είναι πολύ μικρό.

(γ) Ο πληθυσμός ακολουθεί οποιαδήποτε κατανομή με άγνωστη διασπορά και το μέγεθος του δείγματος είναι μεγάλο.

Έστω τυχαίο δείγμα X_1, X_2, \dots, X_n από ένα πληθυσμό με κατανομή F (οποιαδήποτε), με άγνωστη διασπορά σ^2 και μέση τιμή $\mu = \mu_0$ (ελεγχόμενη).

Με την υπόθεση ότι το μέγεθος του δείγματος είναι μεγάλο (θεωρητικά $n \rightarrow +\infty$, στην πράξη $n \geq 30$) μπορεί να αποδειχθεί⁵ ότι η στατιστική συνάρτηση

$T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}$ προσεγγίζεται ικανοποιητικά από την $Z \sim N(0, 1)$. Δηλαδή,

$$T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S} \rightarrow Z \sim N(0,1).$$

Επομένως, σε επίπεδο σημαντικότητας α , απορρίπτουμε την $H_0 : \mu = \mu_0$,

- έναντι της $H_1 : \mu > \mu_0$, όταν,

$$\bar{x} \geq \mu_0 + \frac{s}{\sqrt{n}} z_\alpha,$$

ή, ισοδύναμα, όταν,

$$z = \frac{(\bar{x} - \mu_0)\sqrt{n}}{s} \geq z_\alpha.$$

- έναντι της $H_1 : \mu < \mu_0$, όταν,

$$\bar{x} \leq \mu_0 - \frac{s}{\sqrt{n}} z_\alpha,$$

⁵ Η απόδειξη παραλείπεται. Σημειώνουμε μόνο ότι για την απόδειξη χρησιμοποιείται το Κ.Ο.Θ. και όχι μόνο!

ή, ισοδύναμα, όταν,

$$z = \frac{(\bar{x} - \mu_0)\sqrt{n}}{s} \leq -z_\alpha .$$

- έναντι της $H_1 : \mu \neq \mu_0$, όταν,

$$\bar{x} \leq \mu_0 - \frac{s}{\sqrt{n}} z_{\alpha/2} \quad \text{ή} \quad \bar{x} \geq \mu_0 + \frac{s}{\sqrt{n}} z_{\alpha/2}$$

ή, ισοδύναμα, όταν,

$$|z| = \frac{|(\bar{x} - \mu_0)|\sqrt{n}}{s} \geq z_{\alpha/2} .$$

Σχόλιο: Επειδή, στην περίπτωση αυτή, η κατανομή της $T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}$ δεν είναι κανονική $N(0, 1)$, αλλά **προσεγγίζεται** από την $N(0, 1)$, οι έλεγχοι είναι επιπέδου σημαντικότητας α , κατά προσέγγιση. Φυσικά, όσο μεγαλύτερο είναι το δείγμα, τόσο καλύτερη είναι η προσέγγιση.

Ερώτηση: Αν ο πληθυσμός είναι κανονικός, με άγνωστη διασπορά και το μέγεθος του δείγματος είναι μεγάλο, τότε προφανώς εφαρμόζεται ο έλεγχος της περίπτωσης (β) αλλά και της περίπτωσης (γ). Τι λέτε, τίθεται ουσιαστικό δίλημμα επιλογής ελέγχου⁶;

Ας δούμε τώρα μερικές ασκήσεις και εφαρμογές. Θα μας βοηθήσουν να εξοικειωθούμε στη διάκριση των παραπάνω περιπτώσεων, που ίσως φαντάζουν λαβύρινθος. Όμως, δεν είναι!

Εφαρμογή-1: Στη βιβλιογραφία αναφέρεται ότι η μέση ετήσια παραγωγή γάλακτος μιας συγκεκριμένης φυλής αγελάδων είναι 4000Kgr (ανά αγελάδα). Ένας ερευνητής θέλει να ελέγξει αν στις κτηνοτροφικές μονάδες της Μακεδονίας και της Θράκης οι αγελάδες της συγκεκριμένης φυλής έχουν τη μέση ετήσια απόδοση που αναφέρεται στη βιβλιογραφία. Για το σκοπό αυτό και με βάση ένα σχέδιο τυχαίας δειγματοληψίας, επέλεξε ένα δείγμα 40 αγελάδων της συγκεκριμένης φυλής από μονάδες της Μακεδονίας και της Θράκης και κατέγραφε κάθε μέρα, επί ένα έτος, την παραγωγή γάλακτος κάθε μιας αγελάδας του δείγματος. Η μέση ετήσια παραγωγή των 40 αγελάδων, βρέθηκε 3910Kgr με τυπική απόκλιση 250Kgr.

Θα κάνουμε κατάλληλο στατιστικό έλεγχο για να ελέγξουμε, σε επίπεδο σημαντικότητας $\alpha = 0.05$, αν αυτό που παρατηρήθηκε στο δείγμα υποστηρίζει ότι η μέση ετήσια απόδοση των αγελάδων της συγκεκριμένης φυλής στη Μακεδονία και τη Θράκη διαφέρει από τη μέση ετήσια απόδοση που αναφέρεται στη βιβλιογραφία.

Ο πληθυσμός του οποίου θα ελέγξουμε τη μέση τιμή είναι οι ετήσιες αποδόσεις γάλακτος όλων των αγελάδων της συγκεκριμένης φυλής που εκτρέφονται στη Μακεδονία και τη Θράκη. Ας συμβολίσουμε με X την ετήσια παραγωγή γάλακτος σε Kgr, μιας οποιασδήποτε αγελάδας της συγκεκριμένης φυλής στη Μακεδονία και τη Θράκη και με X_1, X_2, \dots, X_{40} τις ετήσιες αποδόσεις τυχαίου δείγματος 40 αγελάδων από τον πληθυσμό. Στο συγκεκριμένο δείγμα που πήρε ο ερευνητής, οι τιμές του δείγματος, x_1, x_2, \dots, x_{40} , έδωσαν $\bar{x} = 3910kgr$ με $s = 250kgr$.

⁶ Θυμηθείτε ότι για μεγάλα n ισχύει: $t_{n,\alpha} \approx z_\alpha$

Ως μηδενική υπόθεση θέτουμε αυτή που αμφισβητείται από τον ερευνητή (γι' αυτό την ελέγχει) δηλαδή την: $H_0 : \mu = 4000 \text{ Kgr}$. Ως εναλλακτική θέτουμε την $H_1 : \mu \neq 4000 \text{ Kgr}$, γιατί ο ερευνητής θέλει να ελέγξει πιθανή διαφοροποίηση της μέσης απόδοσης και όχι διαφοροποίησή της προς κάποια κατεύθυνση (αύξηση ή μείωση).

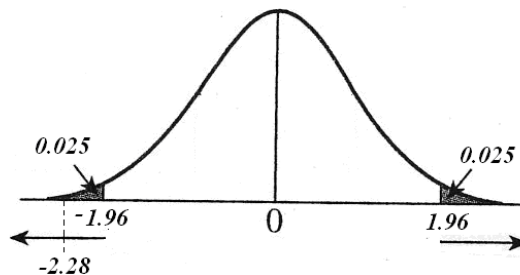
Ως στατιστική συνάρτηση ελέγχου θα χρησιμοποιήσουμε την $\frac{(\bar{X} - \mu_0)\sqrt{n}}{S} \rightarrow Z \sim N(0,1)$ γιατί η διασπορά του πληθυσμού είναι άγνωστη και το μέγεθος του δείγματος είναι $n = 40 > 30$ (περίπτωση (γ)).

Επειδή ο έλεγχος είναι αμφίπλευρος, σε επίπεδο σημαντικότητας $\alpha = 0.05$, η περιοχή απόρριψης είναι: $|z| \geq z_{0.05/2}$ ή $|z| \geq z_{0.025}$ ή $|z| \geq 1.96 \Leftrightarrow z \geq 1.96$ ή $z \leq -1.96$.

Υπολογίζουμε την τιμή της στατιστικής συνάρτησης ελέγχου στο δείγμα. Έχουμε:

$$z = \frac{(\bar{x} - \mu_0)\sqrt{n}}{s} = \frac{(3910 - 4000)\sqrt{40}}{250} \approx -2.28.$$

Ελέγχουμε αν η τιμή της στατιστικής συνάρτησης ελέγχου που βρήκαμε, βρίσκεται στην περιοχή απόρριψης. Πράγματι, επειδή $z = -2.28 \leq -1.96$, η τιμή της στατιστικής συνάρτησης ελέγχου βρίσκεται στην περιοχή απόρριψης και επομένως σε επίπεδο σημαντικότητας $\alpha = 0.05$ απορρίπτουμε τη μηδενική υπόθεση.



Συμπέρασμα: Σε επίπεδο σημαντικότητας $\alpha = 0.05$, το δείγμα δίνει στατιστικά σημαντικές αποδείξεις ότι η μέση ετήσια απόδοση των αγελάδων της συγκεκριμένης φυλής στη Μακεδονία και τη Θράκη διαφέρει από τη μέση ετήσια απόδοση που αναφέρεται στη βιβλιογραφία.

Αλλιώς: Το δείγμα δίνει στατιστικά σημαντικές αποδείξεις ότι η μέση ετήσια απόδοση των αγελάδων της συγκεκριμένης φυλής στη Μακεδονία και τη Θράκη διαφέρει από τη μέση ετήσια απόδοση που αναφέρεται στη βιβλιογραφία. Η πιθανότητα το συμπέρασμα αυτό να είναι λάθος είναι το πολύ 5%.

Παρατήρηση-1: Αν ο ερευνητής έχει υπόνοιες ότι η μέση ετήσια παραγωγή γάλακτος των αγελάδων της συγκεκριμένης φυλής στη Μακεδονία και τη Θράκη, είναι μικρότερη από την αναφερόμενη στη βιβλιογραφία, τότε πρόκειται για άλλο πρόβλημα, για άλλο ερευνητικό ερώτημα. Στην περίπτωση αυτή πρέπει να γίνει ο έλεγχος της $H_0 : \mu = 4000 \text{ Kgr}$ έναντι της $H_1 : \mu < 4000 \text{ Kgr}$. **Ερώτηση:** Τι λέτε, είναι απαραίτητο να κάνουμε τον έλεγχο ή μήπως μπορούμε να συμπεράνουμε το αποτέλεσμα του από το αποτέλεσμα του αμφίπλευρου ελέγχου που ήδη κάναμε;

Παρατήρηση-2: Αν κάνουμε τον έλεγχο σε επίπεδο σημαντικότητας $\alpha = 0.01$, η περιοχή απόρριψης είναι: $|z| \geq z_{0.01/2}$ ή $|z| \geq z_{0.005}$ ή $|z| \geq 2.58 \Leftrightarrow z \geq 2.58$ ή $z \leq -2.58$.

Η τιμή, $z = -2.28$, της στατιστικής συνάρτησης ελέγχου, φυσικά, δεν αλλάζει και επειδή τώρα δε βρίσκεται στην περιοχή απόρριψης, η μηδενική υπόθεση σε επίπεδο σημαντικότητας $\alpha = 0.01$ δεν απορρίπτεται. Δηλαδή, η διαφορά των 90Kgr (μεταξύ δειγματικού μέσου, $\bar{x} = 3910\text{kgr}$, και μηδενικής υπόθεσης, $\mu = 4000\text{ Kgr}$) τώρα δεν κρίνεται ως στατιστικά σημαντική. Αυτό, φυσικά, δεν είναι παράδοξο αφού θέτοντας $\alpha = 0.01$ απαιτούμε πλέον πιο ισχυρές αποδείξεις εναντίον της μηδενικής υπόθεσης.

Ερώτηση: Σε επίπεδο σημαντικότητας $\alpha = 0.02$ ή $\alpha = 0.03$ είναι άραγε στατιστικά σημαντική αυτή η παρατηρούμενη διαφορά; Για να απαντήσουμε, μπορούμε φυσικά να συγκρίνουμε την τιμή της στατιστικής συνάρτησης ελέγχου με την αντίστοιχη, για κάθε περίπτωση, κρίσιμη τιμή. Μπορούμε όμως να κάνουμε κάτι καλύτερο και να δώσουμε μια πληρέστερη απάντηση. Να υπολογίσουμε την P-τιμή του δείγματος, δηλαδή, το ελάχιστο επίπεδο σημαντικότητας για το οποίο απορρίπτεται η μηδενική υπόθεση ή αλλιώς, να υπολογίσουμε πόσο σημαντική (... επιτέλους) είναι αυτή η τιμή που εμφανίστηκε στο συγκεκριμένο τυχαίο δείγμα. Έχουμε:

$$P\text{-τιμή} = P(|Z| \geq 2.28) = P(Z \geq 2.28) + P(Z \leq -2.28) = [1 - \Phi(2.28)] + [1 - \Phi(2.28)] = 0.0226.$$

Έτσι, σε επίπεδο σημαντικότητας $\alpha = 0.01$ και $\alpha = 0.02$ δεν απορρίπτουμε τη μηδενική υπόθεση ενώ σε επίπεδο σημαντικότητας $\alpha = 0.03$ την απορρίπτουμε.

Άσκηση-1: Από έναν πληθυσμό που ακολουθεί κανονική κατανομή, πήραμε ένα δείγμα μεγέθους $n = 9$, με $\bar{x} = 60$ μονάδες και $s = 12$ μονάδες. Ας κάνουμε, σε επίπεδο σημαντικότητας $\alpha = 0.05$, τον έλεγχο της μηδενικής υπόθεσης $H_0 : \mu = 65$ έναντι της εναλλακτικής $H_1 : \mu \neq 65$.

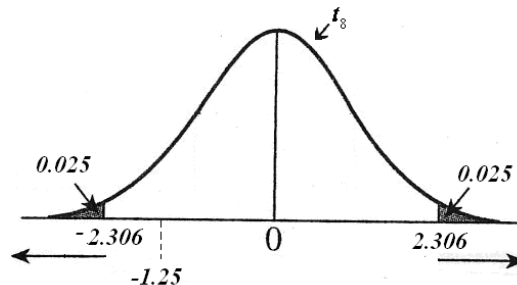
Προφανώς, κατάλληλο είναι το *t-test* (κανονικός πληθυσμός με άγνωστη διασπορά, περίπτωση (β)).

Ο έλεγχος είναι αμφίπλευρος και επομένως, σε επίπεδο σημαντικότητας $\alpha = 0.05$, η περιοχή απόρριψης είναι:

$$|t| \geq t_{8,0.025} \Leftrightarrow |t| \geq 2.306 \Leftrightarrow t \geq 2.306 \text{ ή } t \leq -2.306.$$

Επειδή $t = \frac{(60 - 65)\sqrt{9}}{12} = -1.25$, η τιμή της στατιστικής συνάρτησης ελέγχου δεν

βρίσκεται στην περιοχή απόρριψης και επομένως, σε επίπεδο σημαντικότητας $\alpha = 0.05$, δεν απορρίπτουμε τη μηδενική υπόθεση.

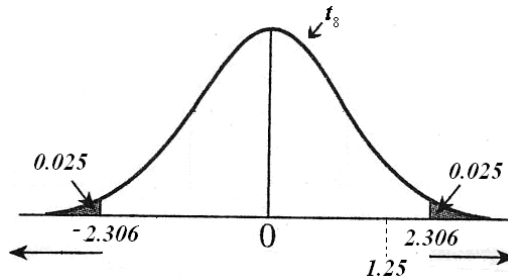


Ερώτηση: Μη απορρίπτοντας την $H_0 : \mu = 65$, αποδείξαμε άραγε ότι είναι αληθής; Δηλαδή, αποδεχόμαστε ότι η μέση τιμή, μ , του πληθυσμού είναι ίση με 65 και είμαστε βέβαιοι γι' αυτό;

Απάντηση: Όχι! Δεν αποδείξαμε ότι $\mu = 65$. Απλώς αποτύχαμε να απορρίψουμε την $H_0 : \mu = 65$. Γι' αυτό, στο συμπέρασμα δε γράψαμε ότι αποδεχόμαστε τη μηδενική

υπόθεση αλλά ότι δεν την απορρίπτουμε. Για να γίνει αυτό κατανοητό, ας κάνουμε, σε επίπεδο σημαντικότητας $\alpha = 0.05$, τον έλεγχο της $H_0 : \mu = 55$ έναντι της $H_1 : \mu \neq 55$.

Η τιμή της στατιστικής συνάρτησης ελέγχου είναι $t = \frac{(60 - 55)\sqrt{9}}{12} = 1.25$. Η περιοχή απόρριψης είναι όπως και προηγουμένως, $t \geq 2.306$ ή $t \leq -2.306$, και επομένως η μηδενική υπόθεση $H_0 : \mu = 55$, σε επίπεδο σημαντικότητας $\alpha = 0.05$, επίσης δεν απορρίπτεται.



Δηλαδή, σε επίπεδο σημαντικότητας $\alpha = 0.05$, τόσο η $H_0 : \mu = 65$ όσο και η $H_0 : \mu = 55$ δεν απορρίπτονται. Επομένως, αν γράψουμε ότι αποδεχόμαστε τη μηδενική, τι αποδεχόμαστε; Ότι η μέση τιμή είναι 65 ή ότι είναι 55; Η απάντηση είναι η εξής: όπως έχουμε αναφέρει, όταν σε ένα στατιστικό έλεγχο απορρίπτουμε τη μηδενική υπόθεση όπως και όταν δεν την απορρίπτουμε, δεν είμαστε βέβαιοι για το συμπέρασμά μας. Είναι πιθανόν να κάνουμε σφάλμα τύπου I ή σφάλμα τύπου II, αντίστοιχα. Την πιθανότητα σφάλματος τύπου I, δηλαδή, την πιθανότητα να κάνουμε σφάλμα όταν απορρίπτουμε τη μηδενική τη γνωρίζουμε. Είναι το πολύ α και τη δηλώνουμε. Όταν δεν απορρίπτουμε τη μηδενική δεν είναι σωστό στο συμπέρασμά μας να γράψουμε ότι «αποδεχόμαστε τη μηδενική υπόθεση» χωρίς να έχουμε υπολογίσει και να δηλώνουμε την πιθανότητα αυτό το συμπέρασμα να είναι λάθος, δηλαδή, χωρίς να έχουμε υπολογίσει την πιθανότητα σφάλματος τύπου II. Και αυτό γιατί αποδοχή σημαίνει απόδειξη-βεβαιότητα κάτι το οποίο δε συμβαίνει αφού υπάρχει πιθανότητα το συμπέρασμά μας αυτό να είναι λάθος. Επειδή, όπως θα δούμε στην επόμενη ενότητα, ο υπολογισμός της πιθανότητας σφάλματος τύπου II, συνήθως, δεν είναι εφικτός (γιατί είναι συνάρτηση της πραγματικής τιμής της παραμέτρου που ελέγχουμε), όταν η μηδενική υπόθεση, σε επίπεδο σημαντικότητας α , δεν απορρίπτεται, στο συμπέρασμα πρέπει να γράφουμε «η μηδενική υπόθεση, σε επίπεδο σημαντικότητας α , δεν απορρίπτεται» ή ακριβέστερα, «σε επίπεδο σημαντικότητας α , αποτύχαμε να απορρίψουμε τη μηδενική υπόθεση» και να αποφεύγουμε να γράφουμε «σε επίπεδο σημαντικότητας α αποδεχόμαστε τη μηδενική υπόθεση». Συμπληρωματικά με το αποτέλεσμα του ελέγχου, και προκειμένου να έχουμε μια εκτίμηση της άγνωστης μέσης τιμής που ελέγχουμε, μπορούμε να υπολογίσουμε ένα $(1 - \alpha) \cdot 100\%$ διάστημα εμπιστοσύνης. Στην περίπτωση που εξετάζουμε, ένα 95% διάστημα εμπιστοσύνης για την άγνωστη μέση τιμή, μ , του πληθυσμού είναι:

$$\bar{x} \pm \frac{s}{\sqrt{n}} \cdot t_{n-1; \alpha/2} \quad \text{ή} \quad 60 \pm \frac{12}{\sqrt{9}} \cdot t_{8; 0.025} \quad \text{ή} \quad 60 \pm 9.224 \quad \text{ή} \quad [50.776, 69.224].$$

Δηλαδή, με βάση αυτό που παρατηρείται στο δείγμα, η μηδενική υπόθεση $H_0 : \mu = 65$ (ή η μηδενική υπόθεση $H_0 : \mu = 55$) σε επίπεδο σημαντικότητας 5% δεν απορρίπτεται και το διάστημα $[50.776, 69.224]$, με πιθανότητα 95%, περιέχει η άγνωστη μέση τιμή, μ , του πληθυσμού. Παρατηρείστε ότι τόσο η τιμή 55 όσο και η τιμή 65 βρίσκονται εντός του 95% διαστήματος εμπιστοσύνης.

Ερώτηση: Τι λέτε, σχετίζεται το διάστημα εμπιστοσύνης με την περιοχή μη απόρριψης της μηδενικής υπόθεσης;

Σχόλιο για το νόημα της μη απόρριψης της μηδενικής υπόθεσης: Κάτι ανάλογο με τη διαδικασία ελέγχου στατιστικών υποθέσεων που περιγράψαμε, συμβαίνει και στη διαδικασία λήψης δικαστικών αποφάσεων. Όταν ένας πολίτης οδηγείται σε δίκη, αυτό συμβαίνει γιατί αμφισβητείται η αθωότητά του. Οι δικαστές θέτουν ως μηδενική υπόθεση ότι ο κατηγορούμενος πολίτης είναι αθώος⁷ (δηλαδή, αυτή που αμφισβητείται) και ως εναλλακτική ότι είναι ένοχος. Η δικαστική διαδικασία σκοπό έχει να διαπιστώσει αν υπάρχουν σημαντικά αποδεικτικά στοιχεία εναντίον της αθωότητας του κατηγορουμένου, δηλαδή, εναντίον της μηδενικής υπόθεσης. Αν δεν προκύψουν τέτοια στοιχεία η μηδενική υπόθεση δεν απορρίπτεται και ο κατηγορούμενος απαλλάσσεται των κατηγοριών. Αυτό δε σημαίνει ότι, κατ' ανάγκη, αποδείχθηκε η αθωότητά του. Σημαίνει ότι δεν βρέθηκαν σημαντικά στοιχεία εναντίον της αθωότητάς του.

Άσκηση-2: Από έναν πληθυσμό με άγνωστη κατανομή και άγνωστη διασπορά, πήραμε ένα τυχαίο δείγμα μεγέθους $n = 36$. Από παλαιότερες έρευνες είναι γνωστό ότι η μέση τιμή του πληθυσμού είναι $\mu = 83$, όμως υπάρχουν υπόνοιες ότι έχει αλλάξει. Το δείγμα που πήραμε έδωσε, $\bar{x} = 86.2$ και $s = 10$. **α)** Να γίνει, σε επίπεδο σημαντικότητας $\alpha = 0.05$, κατάλληλος στατιστικός έλεγχος για τη μέση τιμή του πληθυσμού. **β)** Αν αλλαγή της μέσης τιμής σημαίνει μόνο αύξηση, αλλάζει κάτι στον έλεγχο που πρέπει να κάνουμε; Στο συμπέρασμα;

α) Με βάση όσα έχουμε αναφέρει για τον καθορισμό των δύο υποθέσεων, πρέπει να κάνουμε τον έλεγχο της $H_0 : \mu = 83$ έναντι της $H_1 : \mu \neq 83$. Παρότι δε γνωρίζουμε αν είναι κανονική η κατανομή του πληθυσμού ούτε και τη διασπορά του, επειδή το μέγεθος του δείγματος είναι μεγάλο, σε επίπεδο σημαντικότητας $\alpha = 0.05$, η περιοχή

απόρριψης του ελέγχου είναι $|z| = \frac{|\bar{x} - \mu_0| \sqrt{n}}{s} \geq z_{0.025}$ (περίπτωση (γ)), δηλαδή,
 $z \geq z_{0.025}$ ή $z \leq -z_{0.025} \Leftrightarrow z \geq 1.96$ ή $z \leq -1.96$.

Η τιμή της στατιστικής συνάρτησης ελέγχου είναι $z = \frac{(86.2 - 83)\sqrt{36}}{10} = 1.92$ και

επειδή, προφανώς, δεν ανήκει στην περιοχή απόρριψης, σε επίπεδο σημαντικότητας $\alpha = 0.05$, η μηδενική υπόθεση δεν απορρίπτεται. Δηλαδή, αυτό που παρατηρείται στο δείγμα, σε επίπεδο σημαντικότητας $\alpha = 0.05$, δε δίνει στατιστικά σημαντικές αποδείξεις ότι έχει αλλάξει η μέση τιμή.

β) Είναι προφανές, ότι στην περίπτωση αυτή, σε επίπεδο σημαντικότητας $\alpha = 0.05$, πρέπει να κάνουμε τον έλεγχο της ίδιας μηδενικής υπόθεσης $H_0 : \mu = 83$, έναντι όμως της εναλλακτικής $H_1 : \mu > 83$. Επειδή τώρα πρόκειται για μονόπλευρο-δεξιόπλευρο έλεγχο, η περιοχή απόρριψης είναι, $z \geq z_{0.05}$ ή $z \geq 1.645$ και επειδή για την τιμή της στατιστικής συνάρτησης ελέγχου έχουμε $z = 1.92 \geq 1.645$, η μηδενική υπόθεση, σε επίπεδο σημαντικότητας $\alpha = 0.05$, απορρίπτεται. Δηλαδή, αυτό που παρατηρείται στο δείγμα, σε επίπεδο σημαντικότητας $\alpha = 0.05$, δίνει στατιστικά σημαντικές αποδείξεις ότι η μέση τιμή έχει αυξηθεί!

⁷ Έτσι προβλέπεται από το δικαϊκό μας σύστημα (... ακόμη.): «ο κατηγορούμενος είναι αθώος μέχρι αποδείξεως του εναντίου». Ας ελπίσουμε ότι δε θα επιστρέψουμε σε μεθόδους ιεράς εξέτασης όπου ο κατηγορούμενος έπρεπε να αποδείξει την αθωότητά του...

Ερώτηση: Εξηγείστε, με βάση τη λογική της διαδικασίας ελέγχου, γιατί τα αποτελέσματα των δύο ελέγχων που κάναμε στα (α) και (β) δεν είναι αντιφατικά⁸.

Εφαρμογή-2: Τα βιομηχανικά απόβλητα που ρίχνονται στα ποτάμια, απορροφούν το διαλυμένο στο νερό οξυγόνο, με συνέπεια αυτό να μειώνεται και όταν η μέση τιμή του δεν υπερβαίνει τα 5ppm, να δημιουργείται σοβαρό πρόβλημα επιβίωσης των υδρόβιων οργανισμών. Το πρόβλημα αυτό είχε διαπιστωθεί, πριν από αρκετά χρόνια, και στον ποταμό Καλαμά. Για την αντιμετώπισή του εφαρμόστηκε ειδικό πρόγραμμα αποκατάστασης και προστασίας του ποταμού. Ένας φοιτητής, στο πλαίσιο της πτυχιακής του εργασίας που είχε σκοπό να διερευνήσει αν απέδωσαν τα μέτρα προστασίας, έπρεπε, μεταξύ άλλων δεικτών, να μελετήσει την ποσότητα διαλυμένου οξυγόνου στα νερά του ποταμού. Για το σκοπό αυτό, πήρε, με βάση κατάλληλο σχέδιο τυχαίας δειγματοληψίας, μετρήσεις από 10 σημεία της κοίτης του ποταμού. Οι 10 μετρήσεις έδωσαν τις εξής τιμές διαλυμένου οξυγόνου(σε ppm): 5, 5.1, 5.2, 5.1, 4.9, 5.3, 5, 5.2, 5.1, 5.1. Με βάση αυτές τις τιμές, μπορεί ο φοιτητής να συμπεράνει ότι στον ποταμό Καλαμά η μέση ποσότητα διαλυμένου οξυγόνου είναι πλέον μεγαλύτερη από 5ppm;

Ο φοιτητής μελετάει την ποσότητα, X , διαλυμένου οξυγόνου στα νερά του ποταμού Καλαμά με βάση ένα τυχαίο δείγμα X_1, X_2, \dots, X_{10} , 10 μετρήσεων.

Αν μ είναι η άγνωστη μέση τιμή της τυχαίας μεταβλητής X , πρέπει να κάνει κατάλληλο στατιστικό έλεγχο για να ελέγξει αν οι 10 τιμές x_1, x_2, \dots, x_{10} που έδωσε το συγκεκριμένο δείγμα που πήρε, υποστηρίζουν την απόρριψη της μηδενικής υπόθεσης $H_0 : \mu = 5ppm$ ή, πιο σωστά, της $H_0 : \mu \leq 5ppm$, υπέρ της εναλλακτικής $H_1 : \mu > 5ppm$ ⁹.

Η κατανομή του πληθυσμού¹⁰ δεν είναι γνωστή. Επίσης, η διασπορά του δεν είναι γνωστή και το μέγεθος του δείγματος είναι μικρό ($n = 10 < 30$). Αυτή η περίπτωση δεν εντάσσεται στις περιπτώσεις (α), (β) ή (γ) που μελετήσαμε προηγουμένως. Αν το μέγεθος του δείγματος ήταν μεγάλο, ως περιοχή απόρριψης θα μπορούσαμε να πάρουμε την αντίστοιχη, για τον έλεγχο που κάνουμε, της περίπτωσης (γ). Όμως δεν είναι. Επίσης, αν γνωρίζαμε ότι η κατανομή του πληθυσμού είναι κανονική, θα εφαρμόζαμε το t -test (περίπτωση (β)). Τι κάνουμε επομένως;

Με βάση όσα μέχρι τώρα γνωρίζουμε, ένα δρόμο έχουμε.¹¹ Να ανατρέξουμε στη βιβλιογραφία και να αναζητήσουμε, από ανάλογες έρευνες, πληροφορίες για την κατανομή της ποσότητας διαλυμένου οξυγόνου στα νερά ποταμών με συνθήκες ανάλογες του Καλαμά. Τέτοιες έρευνες πράγματι βρέθηκαν και από αυτές προκύπτει ότι η κατανομή διαλυμένου οξυγόνου προσομοιάζει με την κανονική και σε κάθε

⁸ Σκεφθείτε ότι, παρότι τόσο ο αμφίπλευρος όσο και ο δεξιόπλευρος έλεγχος έγιναν στο ίδιο επίπεδο σημαντικότητας, εντούτοις στον δεξιόπλευρο είμαστε πιο ανεκτικοί σε σφάλμα λανθασμένης απόρριψης της μηδενικής.

⁹ Σημειώστε ότι η περιοχή απόρριψης της μηδενικής δεν αλλάζει αν αντί της $H_0 : \mu = 5ppm$ θεωρήσουμε την $H_0 : \mu \leq 5ppm$.

¹⁰ Ο πληθυσμός του οποίου ελέγχουμε τη μέση τιμή είναι οι τιμές διαλυμένου οξυγόνου στην κοίτη του ποταμού.

¹¹ Η Στατιστική προσφέρει και άλλο δρόμο. Με κατάλληλους στατιστικούς ελέγχους αλλά και με κατάλληλες γραφικές μεθόδους και εργαλεία μπορούμε να ελέγξουμε αν το δείγμα μας προέρχεται από κανονικό πληθυσμό και αν αυτό δε συμβαίνει μπορούμε να εφαρμόσουμε μη παραμετρικούς ελέγχους. Σε όλα αυτά θα αναφερθούμε σε άλλη ενότητα.

περίπτωση δεν παρουσιάζει σοβαρές ασυμμετρίες. Με βάση αυτή την πληροφορία και δεδομένου ότι το μέγεθος του δείγματος δεν είναι πολύ μικρό, μπορούμε να εφαρμόσουμε το t -test (περίπτωση (β)) αφού όπως έχουμε αναφέρει η εμπειρία έχει δείξει ότι αυτό είναι «ανθεκτικό» στην υπόθεση της κανονικότητας του πληθυσμού.

Η συνέχεια είναι πλέον γνωστή. Ορίζουμε το επίπεδο σημαντικότητας του ελέγχου, έστω $\alpha = 0.05$, και υπολογίζουμε την τιμή της στατιστικής συνάρτησης ελέγχου

$$T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S}$$
 αφού προηγουμένως υπολογίσουμε την τιμή \bar{x} της \bar{X} και την τιμή s της S στο δείγμα που πήραμε.

Έτσι έχουμε, $\bar{x} = 5.1 ppm$ και $s = 0.115 ppm$ και επομένως, $t = \frac{(5.1 - 5)\sqrt{10}}{0.115} = 2.75$.

Ελέγχουμε αν η τιμή της στατιστικής συνάρτησης ελέγχου που βρήκαμε βρίσκεται στην περιοχή απόρριψης. Ο έλεγχος είναι δεξιόπλευρος και επομένως, σε επίπεδο σημαντικότητας $\alpha = 0.05$, η περιοχή απόρριψης είναι: $t \geq t_{9,0.05}$ ή $t \geq 1.833$ και επειδή $t = 2.75 \geq 1.833$, η τιμή της στατιστικής συνάρτησης ελέγχου βρίσκεται στην περιοχή απόρριψης και επομένως σε επίπεδο σημαντικότητας $\alpha = 0.05$ απορρίπτουμε τη μηδενική υπόθεση.

Συμπέρασμα: Σε επίπεδο σημαντικότητας $\alpha = 0.05$, το δείγμα δίνει στατιστικά σημαντικές αποδείξεις ότι η μέση ποσότητα διαλυμένου οξυγόνου στον ποταμό Καλαμά είναι πλέον μεγαλύτερη από $5 ppm$ ¹².

Ερώτηση: Ο φοιτητής με τον έλεγχο που έκανε, απέρριψε τη μηδενική υπόθεση ότι η μέση ποσότητα διαλυμένου οξυγόνου είναι μικρότερη ή το πολύ ίση με $5 ppm$ υπέρ της εναλλακτικής ότι είναι μεγαλύτερη από $5 ppm$ με πιθανότητα αυτό το συμπέρασμα να είναι λάθος το πολύ 5%. Όμως, για τον επιβλέποντα καθηγητή, αυτό το συμπέρασμα δεν είναι αρκετό και του ζήτησε μια εκτίμηση της μέσης ποσότητας διαλυμένου οξυγόνου. Τι πρέπει να κάνει ο φοιτητής;

Άσκηση-3 (κατανόησης): Ένας ερευνητής έκανε ένα μονόπλευρο στατιστικό έλεγχο, σε επίπεδο σημαντικότητας $\alpha = 0.01$, και το αποτέλεσμα του ελέγχου ήταν ότι η μηδενική υπόθεση απορρίπτεται. Ένας άλλος ερευνητής, χρησιμοποιώντας το ίδιο δείγμα έκανε έναν αμφίπλευρο έλεγχο σε επίπεδο σημαντικότητας $\alpha = 0.05$, και το αποτέλεσμα του ελέγχου ήταν ότι η μηδενική υπόθεση δεν απορρίπτεται. Είναι δυνατόν να είναι σωστά και τα δύο αποτελέσματα;

Απάντηση: Όχι. Σκεφθείτε γιατί.

Ερώτηση: Αν, σε επίπεδο σημαντικότητας α , η μηδενική υπόθεση $H_0 : \mu = \mu_0$ απορρίπτεται υπέρ της εναλλακτικής $H_1 : \mu > \mu_0$ τότε, στο ίδιο επίπεδο σημαντικότητας, πρέπει απαραίτητως να απορρίπτεται και υπέρ της $H_1 : \mu \neq \mu_0$;

Απάντηση: Όχι! Σκεφθείτε γιατί.

¹² Το δείγμα που χρησιμοποιήσαμε είναι δυστυχώς... υποθετικό.

3. Η πιθανότητα σφάλματος τύπου II και η ισχύς ενός στατιστικού ελέγχου

Στη διαδικασία στατιστικού ελέγχου υποθέσεων που περιγράψαμε στα προηγούμενα, δεν αναφερθήκαμε καθόλου στο τι συμβαίνει με την πιθανότητα *λανθασμένης μη απόρριψης* της H_0 , δηλαδή, στην πιθανότητα σφάλματος τύπου II. Η πιθανότητα αυτή συμβολίζεται με β . Έτσι, ενώ αν απορρίψουμε την H_0 , γνωρίζουμε με ποια πιθανότητα αυτή η απόφασή μας μπορεί να είναι λάθος (είναι το πολύ α), αντίθετα, αν δεν απορρίψουμε την H_0 , με όσα μέχρι τώρα αναφέραμε, δεν γνωρίζουμε με ποια πιθανότητα αυτή η απόφασή μας μπορεί να είναι λάθος, αφού δεν υπολογίσαμε την πιθανότητα σφάλματος τύπου II,

$$\beta = P(\text{σφάλμα τύπου II}) = P(\text{μη απόρριψη της } H_0 \mid \text{αληθής η } H_1).$$

Φροντίσαμε, δηλαδή, για την «προστασία» από σφάλμα τύπου I και δεν ασχοληθήκαμε με το σφάλμα τύπου II, δηλαδή, με το σφάλμα που κάνουμε όταν, ενώ είναι αληθής η H_1 , αποτυγχάνουμε να απορρίψουμε την H_0 . Κατά συνέπεια, δε γνωρίζουμε και την πιθανότητα,

$$1 - \beta = P(\text{απόρριψη της } H_0 \mid \text{αληθής η } H_1),$$

δηλαδή, την ικανότητα του ελέγχου να «διακρίνει-αναγνωρίζει» υπαρκτές σημαντικές διαφορές του δείγματος από την H_0 και έτσι να μην αποτυγχάνει να την απορρίψει. Η πιθανότητα, $1 - \beta$, ονομάζεται *ισχύς (power)* του ελέγχου ή, ακριβέστερα (θα δούμε στη συνέχεια γιατί), *συνάρτηση ισχύος (power function)* του ελέγχου. Μεγαλύτερη *ισχύς* σημαίνει μεγαλύτερη πιθανότητα να μην αποτύχουμε να απορρίψουμε την H_0 όταν είναι αληθής η H_1 (και επομένως πιο καλός έλεγχος).

Σε αυτή την ενότητα θα δούμε πώς μπορούμε να υπολογίσουμε την πιθανότητα *λανθασμένης μη απόρριψης* της H_0 , β , και, κατά συνέπεια, την *ισχύ*, $1 - \beta$, του ελέγχου.

Στο παράδειγμά μας με τη μέση αντοχή των καλωδίων, στον έλεγχο της $H_0 : \mu = 1500 \text{ Kgr}$, έναντι της $H_1 : \mu > 1500 \text{ Kgr}$, σε επίπεδο σημαντικότητας $\alpha = 0.01$, δεν απορρίψαμε την H_0 . Ας υπολογίσουμε την πιθανότητα, η απόφασή μας αυτή να είναι λανθασμένη, δηλαδή, την πιθανότητα σφάλματος τύπου II, β .

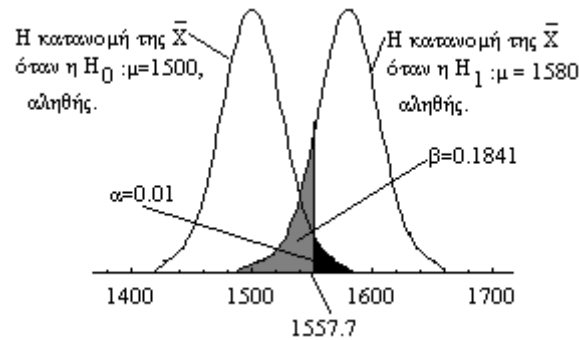
Αν μεταξύ των δύο υποθέσεων, αληθής είναι η $H_1 : \mu > 1500 \text{ Kgr}$, δηλαδή, αν η πραγματική-αληθής μέση τιμή, μ , της αντοχής των καλωδίων μετά τη βελτίωση των υλικών είναι ένας αριθμός μ_1 μεγαλύτερος των 1500 Kgr , τότε ζητάμε την πιθανότητα β , να μην απορρίψουμε την $H_0 : \mu = 1500 \text{ Kgr}$ (ενώ θα έπρεπε, αφού αληθής είναι η H_1). Έχουμε:

$$\begin{aligned} \beta &= P(\text{σφάλμα τύπου II}) = P(\text{μη απόρριψη της } H_0 \mid \text{αληθής η } H_1 : \mu = \mu_1 > 1500) = \\ &= P(\bar{X} < 1557.7 \mid \mu = \mu_1) = P\left(\frac{\bar{X} - \mu_1}{175/\sqrt{50}} < \frac{1557.7 - \mu_1}{175/\sqrt{50}}\right) = P\left(Z < \frac{1557.7 - \mu_1}{24.75}\right) = \Phi\left(\frac{1557.7 - \mu_1}{24.75}\right). \end{aligned}$$

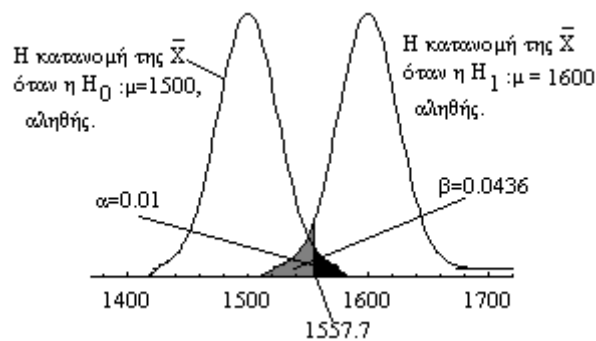
$$\text{Δηλαδή, } \beta = \Phi\left(\frac{1557.7 - \mu_1}{24.75}\right), \quad \mu_1 > 1500.$$

Παρατηρούμε, ότι η πιθανότητα σφάλματος τύπου II, β , εξαρτάται από την πραγματική τιμή μ_1 της άγνωστης παραμέτρου μ . Έτσι, κάνοντας τον έλεγχο σε επίπεδο σημαντικότητας $\alpha = 0.01$, αν η πραγματική τιμή είναι $\mu_1 = 1580$, η πιθανότητα σφάλματος τύπου II είναι:

$$\beta = \Phi\left(\frac{1557.7 - 1580}{24.75}\right) = \Phi(-0.90) = 1 - \Phi(0.90) = 0.1841,$$



ενώ αν $\mu_1 = 1600$ είναι, $\beta = \Phi\left(\frac{1557.7 - 1600}{24.75}\right) = \Phi(-1.71) = 1 - \Phi(1.71) = 0.0436$.



Δηλαδή, στο παράδειγμά μας, όσο πιο μακριά από την $H_0 : \mu = 1500$ Kgr (προς μεγαλύτερες τιμές), βρίσκεται η πραγματική τιμή της άγνωστης παραμέτρου μ , τόσο η πιθανότητα σφάλματος τύπου II ελαττώνεται¹³.

Αντίστοιχα, η ισχύς του ελέγχου,

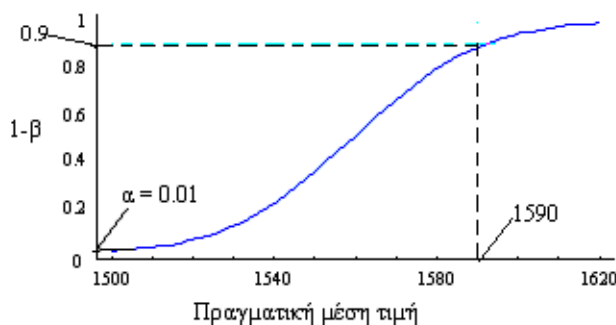
$$1 - \beta = 1 - \Phi\left(\frac{1557.7 - \mu_1}{24.75}\right), \mu_1 > 1500,$$

εξαρτάται και αυτή από την πραγματική τιμή της άγνωστης παραμέτρου, μ , και μάλιστα, στο παράδειγμά μας, είναι μια αύξουσα συνάρτηση γιατί η τιμή της αυξάνεται όταν η πραγματική τιμή της άγνωστης παραμέτρου μ , αυξάνεται. Έτσι, όσο πιο μακριά από την $H_0 : \mu = 1500$ Kgr (προς μεγαλύτερες τιμές), βρίσκεται η πραγματική τιμή της άγνωστης παραμέτρου μ , τόσο αυξάνεται η ικανότητα του ελέγχου να «αναγνωρίζει» σημαντικές διαφορές του δείγματος από την H_0 και να μην αποτυγχάνει να την απορρίψει σωστά¹⁴.

Η γραφική παράσταση της συνάρτησης ισχύος ονομάζεται *καμπύλη ισχύος (power curve)* του ελέγχου. Στο σχήμα που ακολουθεί φαίνεται η *καμπύλη ισχύος* του ελέγχου του παραδείγματός μας για $\alpha = 0.01$.

¹³ Είναι λογικό;

¹⁴ Είναι λογικό;



Από την καμπύλη ισχύος φαίνεται ότι, όσο αυξάνεται η πραγματική τιμή της μ , η ισχύς του ελέγχου τείνει προς το 1. Επίσης, όταν η πραγματική τιμή της παραμέτρου μ , τείνει προς την τιμή 1500, η ισχύς του ελέγχου μειώνεται και τείνει προς το $0.01 = \alpha$!!

Σχόλιο για τη χρησιμότητα της καμπύλης ισχύος: Με χρήση κατάλληλου λογισμικού, είναι πολύ εύκολο να πάρουμε την καμπύλη ισχύος ενός στατιστικού ελέγχου. Έτσι, έχουμε στη διάθεσή μας μια γραφική αναπαράσταση της «αποδοτικότητας» του ελέγχου, δηλαδή, της ικανότητάς του να απορρίπτει σωστά τη μηδενική υπόθεση. Αν, για παράδειγμα, το εργοστάσιο ισχυρισθεί ότι η μέση αντοχή των καλωδίων με τα νέα υλικά αυξήθηκε και μάλιστα τώρα πλέον είναι ίση με 1590Kgr, και πράγματι είναι έτσι, τότε, από την καμπύλη ισχύος του ελέγχου και χωρίς άλλους υπολογισμούς εύκολα διαπιστώνουμε ότι η πιθανότητα να διακρίνει σωστά ο έλεγχος τις δύο υποθέσεις και να απορριφθεί σωστά η $H_0 : \mu = 1500 \text{ Kgr}$ υπέρ της $H_1 : \mu = \mu_1 = 1590 \text{ Kgr}$ είναι περίπου 90%. Επίσης, από την καμπύλη ισχύος, μπορούμε να δούμε πόσο γρήγορα αυξάνει η ισχύς του ελέγχου και να συγκρίνουμε γραφικά την ισχύ του με την ισχύ κάποιου άλλου ελέγχου για κάθε τιμή της παραμέτρου που ορίζει η H_1 . Αναφέρουμε, τέλος, χωρίς απόδειξη, ότι ο έλεγχος που εφαρμόσαμε στο παράδειγμά μας, είναι ο πλέον ισχυρός από οποιονδήποτε άλλο, δηλαδή, οδηγεί στη μικρότερη δυνατή πιθανότητα σφάλματος τύπου II για κάθε τιμή της παραμέτρου που ορίζει η H_1 .

Από τον ορισμό της πιθανότητας σφάλματος τύπου II, β , είναι προφανές ότι αυτή επηρεάζεται από το επίπεδο σημαντικότητας, α , του ελέγχου (αφού η κρίσιμη τιμή του ελέγχου προκύπτει από το α). Όμως, πώς επηρεάζεται;

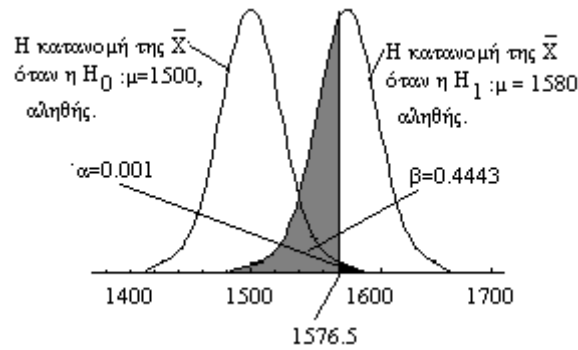
Ας κάνουμε τον έλεγχο του παραδείγματός μας σε μικρότερο επίπεδο σημαντικότητας, $\alpha = 0.001$. Στην περίπτωση αυτή, φυσικά, η $H_0 : \mu = 1500 \text{ Kgr}$ δεν απορρίπτεται¹⁵.

Η κρίσιμη τιμή τώρα είναι $c = 1500 + z_{0.001} \frac{175}{\sqrt{50}} = 1576.5$ ή, ισοδύναμα,

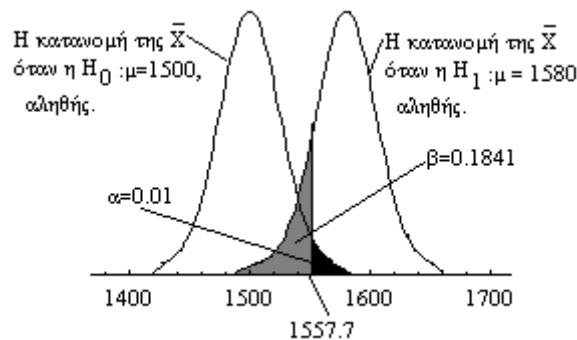
$c = z_{0.001} = 3.09$ και επομένως, $\beta = \Phi\left(\frac{1576.5 - \mu_1}{24.75}\right)$, $\mu_1 > 1500$. Έτσι, αν για

παράδειγμα, $\mu_1 = 1580$, τότε $\beta = \Phi\left(\frac{1576.5 - 1580}{24.75}\right) = \Phi(-0.14) = 0.4443$.

¹⁵ Αφού, όπως είδαμε, δεν απορρίπτεται σε $\alpha = 0.01$.



Συγκρίνοντας αυτή την τιμή του β , με την αντίστοιχη που υπολογίσαμε όταν κάναμε τον έλεγχο σε επίπεδο σημαντικότητας $\alpha = 0.01$,



παρατηρούμε ότι, ελαττώνοντας την πιθανότητα α , δηλαδή, κάνοντας πιο αυστηρό τον κανόνα απόρριψης της μηδενικής υπόθεσης, η πιθανότητα σφάλματος τύπου II, αυξάνεται, δηλαδή, αυξάνεται η πιθανότητα να αποτύχουμε να την απορρίψουμε ενώ έπρεπε.

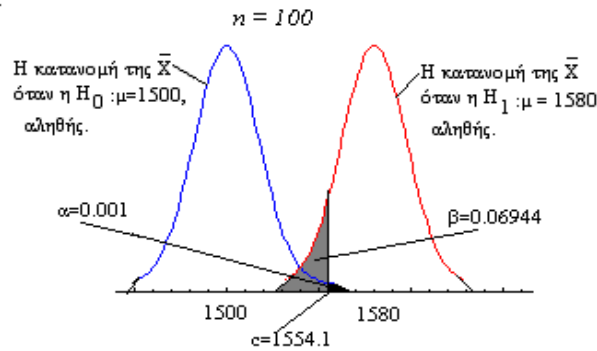
Από την τελευταία παρατήρηση, είναι φανερό ότι δε μπορούμε να θέσουμε υπό τον έλεγχό μας συγχρόνως και την πιθανότητα σφάλματος τύπου I και την πιθανότητα σφάλματος τύπου II, γι'αυτό και στη διαδικασία ελέγχου που εφαρμόσαμε αποφασίσαμε να θέσουμε υπό τον έλεγχό μας τη μία από τις δύο πιθανότητες σφάλματος (την πιθανότητα σφάλματος τύπου I). Τι κάνουμε όμως με την πιθανότητα β ; Την αφήνουμε «ανεξέλεγκτη»; Δηλαδή, πώς μπορούμε, για δεδομένη πιθανότητα σφάλματος τύπου I, α , να μειώσουμε την πιθανότητα σφάλματος τύπου II, β ; Στο πλαίσιο του μαθήματος, δε μπορούμε να απαντήσουμε με πληρότητα σε αυτό το ερώτημα. Ας δούμε, στο παράδειγμά μας, πώς μπορούμε για δεδομένο α , να ελαττώσουμε το β και κατά συνέπεια να αυξήσουμε την ισχύ, $1 - \beta$, του ελέγχου.

Ένας τρόπος, πρακτικά εφικτός¹⁶, είναι η αύξηση του μεγέθους του δείγματος. Με αυτόν τον τρόπο, επιτυγχάνουμε να μειώσουμε τη διασπορά $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$ της στατιστικής συνάρτησης ελέγχου \bar{X} . Αυτό, θεωρητικά, επιτυγχάνεται και με μείωση της διασποράς, σ^2 , της X , διατηρώντας το μέγεθος του δείγματος, n , σταθερό. Όμως, είναι φανερό, ότι αυτός ο τρόπος έχει μικρή πρακτική αξία.

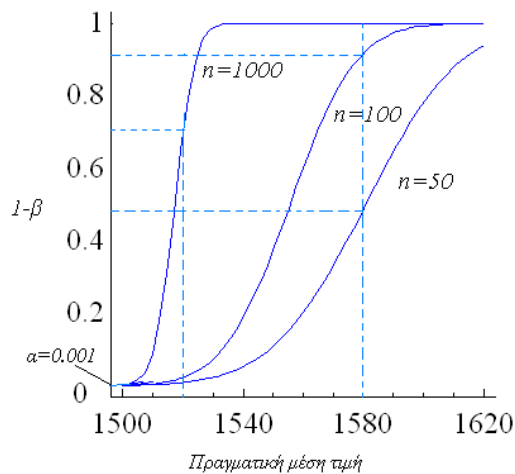
Ας υποθέσουμε, στο παράδειγμά μας, ότι παίρνουμε δείγμα μεγέθους $n = 100$ με $\bar{x} = 1550$ Kgr. Σε επίπεδο σημαντικότητας $\alpha = 0.001$, η κρίσιμη τιμή είναι

¹⁶ Όμως όχι πάντα.

$c = 1500 + z_{0.001} \frac{175}{\sqrt{100}} = 1554.1$ και επειδή η τιμή, $\bar{x} = 1550$, της \bar{X} δεν είναι μεγαλύτερη της κρίσιμης τιμής, η $H_0 : \mu = 1500 \text{ Kgr}$ δεν απορρίπτεται σε επίπεδο σημαντικότητας $\alpha = 0.001$. Η πιθανότητα σφάλματος τύπου II, β , είναι $\beta = \Phi\left(\frac{1554.1 - \mu_1}{17.5}\right)$, $\mu_1 > 1500$ και η τιμή της, για $\mu_1 = 1580$, είναι: $\beta = \Phi\left(\frac{1554.1 - 1580}{17.5}\right) = \Phi(-1.48) = 0.06944$.



Στον αντίστοιχο έλεγχο με δείγμα μεγέθους $n = 50$, και για $\mu_1 = 1580$, βρήκαμε μέγεθος σφάλματος τύπου II, $\beta = 0.4443$. Αυξάνοντας, επομένως, το μέγεθος του δείγματος, παρατηρούμε ότι η πιθανότητα σφάλματος τύπου II, β , μειώνεται. Αντίστοιχα, η ισχύς του ελέγχου αυξάνεται. Στο σχήμα που ακολουθεί φαίνεται η καμπύλη ισχύος του ελέγχου για $n = 50$, $n = 100$ και $n = 1000$.



Από τις καμπύλες ισχύος του ελέγχου, είναι φανερό, ότι για οποιοδήποτε $\mu_1 > 1500$, η ισχύς του ελέγχου αυξάνεται όταν το μέγεθος του δείγματος αυξάνεται. Παρατηρείστε ότι για πραγματική μέση τιμή $\mu_1 = 1580 \text{ Kgr}$ η ισχύς του ελέγχου για $n = 50$ είναι 55.57%, για $n = 100$ αυξάνεται σε 93% και για $n = 1000$ σε 100%. Αυτό φαίνεται να είναι ένα καλό χαρακτηριστικό της διαδικασίας ελέγχου που εφαρμόσαμε αφού, ορίζοντας το επιθυμητό α και αυξάνοντας το μέγεθος του δείγματος αυξάνουμε την πιθανότητα να απορριφθεί σωστά μια μη αληθής μηδενική υπόθεση. Όμως, δυστυχώς, υπάρχει και η «σκοτεινή πλευρά». Σκεφθείτε ότι με αυτό τον τρόπο μπορεί να απορριφθεί με βεβαιότητα (με πιθανότητα 100%) οποιαδήποτε μη αληθής μηδενική υπόθεση, όσο κοντά και αν είναι στην πραγματική τιμή μ_1 . Και αυτό είναι «κακό» γιατί μπορεί να λειτουργήσει παραπλανητικά στην ερμηνεία του αποτελέσματος, αφού ακόμη και μια πολύ μικρή διαφορά της πραγματικής τιμής από

την τιμή που ορίζει η μηδενική υπόθεση, με επιλογή κατάλληλου μεγέθους δείγματος, μπορεί να αναδειχθεί ως στατιστικά σημαντική ενώ πρακτικά η διαφορά αυτή μπορεί να μην έχει κανένα σημαντικό αντίκρουσμα. Δηλαδή, κάτι πρακτικά ασήμαντο να αναδειχθεί σημαντικό στατιστικά! Δείτε, για παράδειγμα, στο σχήμα, την ισχύ του ελέγχου για πραγματική τιμή $\mu_1 = 1520 \text{ Kgr}$, δηλαδή, για πραγματική τιμή πιο κοντά στην $H_0 : \mu = 1500 \text{ Kgr}$. Για μέγεθος δείγματος 50 η ισχύς του ελέγχου είναι πολύ κοντά στο μηδέν. Για μέγεθος δείγματος 100, είναι κοντά στο μηδέν επίσης. Όμως, για μέγεθος δείγματος 1000 αυξάνει στο 90% περίπου και για ακόμη μεγαλύτερο μέγεθος δείγματος θα φθάσει το 100%. Ανάλογα, μπορούμε να επιτύχουμε απόρριψη της $H_0 : \mu = 1500 \text{ Kgr}$ όσο κοντά και αν αυτή βρίσκεται στην πραγματική τιμή (π.χ. αν $\mu_1 = 1500.1 \text{ Kgr}$). Απαιτείται επομένως προσοχή στην ερμηνεία στατιστικά σημαντικών ευρημάτων ιδιαίτερα όταν χρησιμοποιούνται πολύ μεγάλα δείγματα. Στατιστική συμπερασματολογία δε σημαίνει απλή εφαρμογή αλγοριθμικών διαδικασιών. Για τη σωστή αξιοποίηση των μεθόδων που προσφέρει η στατιστική συμπερασματολογία, απαιτείται κατανόηση του νοήματός τους, της λογικής τους και κυρίως των ορίων τους.

Στην ενότητα αυτή, ολοκληρώσαμε την παρουσίαση των βασικών εννοιών των στατιστικών ελέγχων υποθέσεων, χρησιμοποιώντας, ως παράδειγμα, τον έλεγχο της μέσης τιμής, μ , ενός πληθυσμού. Στην ενότητα που ακολουθεί, παρουσιάζουμε τον έλεγχο της μέσης τιμής, p , της κατανομής *Bernoulli* (γνωστή και ως διωνυμικό ποσοστό).

4. Στατιστικός έλεγχος υποθέσεων για τη μέση τιμή, p , της κατανομής *Bernoulli*

Αν X_1, X_2, \dots, X_n ένα τυχαίο δείγμα από έναν πληθυσμό που ακολουθεί την κατανομή *Bernoulli* με παράμετρο p , ο έλεγχος της μηδενικής υπόθεσης, $H_0 : p = p_0$ είναι μια ειδική περίπτωση ελέγχου της μέσης τιμής, μ , ενός πληθυσμού αφού, όπως γνωρίζουμε, αν μια τυχαία μεταβλητή ακολουθεί την κατανομή *Bernoulli* με παράμετρο p , η μέση τιμή της είναι ίση με p , δηλαδή, $\mu = p$.

Ας θυμηθούμε το πρόβλημα που χρησιμοποιήσαμε, ως εισαγωγικό παράδειγμα, όταν μιλήσαμε για τη διωνυμική κατανομή.

«Ο γεωπόνος ενός φυτώριου ισχυρίζεται ότι ποσοστό 90% ενός είδους φυτών που παράγει δίνει περισσότερους από 5 καρπούς/φυτό. Ένας αγρότης που είχε προμηθευθεί από το συγκεκριμένο φυτώριο μεγάλο αριθμό φυτών αυτού του είδους, θέλησε να ελέγξει τον ισχυρισμό του γεωπόνου. Για το σκοπό αυτό, κατά τη συγκομιδή, επέλεξε τυχαία 52 φυτά και μέτρησε τους καρπούς κάθε φυτού. Από τα 52 φυτά που εξέτασε, περισσότερους από 5 καρπούς είχαν μόνο τα 38, γεγονός που δημιούργησε αμφιβολίες στον αγρότη για τον ισχυρισμό του γεωπόνου. Είναι, άραγε, δικαιολογημένες οι αμφιβολίες του αγρότη;»

Η απάντηση που είχαμε δώσει στο πρόβλημα ήταν,

«Με την υπόθεση ότι αληθεύει ο ισχυρισμός του γεωπόνου, η πιθανότητα να συμβεί αυτό που συνέβη στον έλεγχο που έκανε ο αγρότης είναι πολύ μικρή (σχεδόν μηδενική). Άρα οι αμφιβολίες του αγρότη έχουν βάση.»

Είναι φανερό, ότι οι αμφιβολίες του αγρότη θέτουν ένα πρόβλημα στατιστικού ελέγχου υποθέσεων για το πραγματικό ποσοστό p των φυτών (επί του συνόλου των

φυτών) που παράγουν περισσότερους από 5 καρπούς και έτσι, μπορούμε πλέον να δώσουμε ένα ασφαλέστερο συμπέρασμα.

Αυτό που αμφισβητείται είναι ο ισχυρισμός του γεωπόνου ότι το ποσοστό των φυτών που παράγουν περισσότερους από 5 καρπούς είναι 0.9. Επομένως, πρέπει να γίνει ο έλεγχος της μηδενικής υπόθεσης $H_0 : p = 0.9$ έναντι της εναλλακτικής $H_1 : p < 0.9$.

Ο έλεγχος αυτός, είναι λογικό, να γίνει με βάση το ποσοστό των «επιτυχιών» $\frac{x_1+x_2+\dots+x_{52}}{52} = \frac{38}{52} = 0.73$ στο τυχαίο δείγμα, X_1, X_2, \dots, X_{52} , των $n = 52$ δοκιμών¹⁷. Επομένως, για να προχωρήσουμε, πρέπει να γνωρίζουμε την κατανομή του δειγματικού ποσοστού $\frac{X_1+X_2+\dots+X_{52}}{52}$ (στην απάντηση που δώσαμε όταν μιλήσαμε για τη διωνυμική κατανομή, είχαμε υπολογίσει την πιθανότητα ο αριθμός των «επιτυχιών», $X_1 + X_2 + \dots + X_{52}$, στις 52 δοκιμές, να είναι 38 και επομένως έπρεπε να γνωρίζουμε την κατανομή της $X_1 + X_2 + \dots + X_{52}$ που εξ' ορισμού είναι η διωνυμική, $B(52, 0.9)$).

Γενικά, για τον έλεγχο της μηδενικής υπόθεσης $H_0 : p = p_0$, πρέπει να γνωρίζουμε την κατανομή του δειγματικού ποσοστού (δειγματικής μέσης τιμής) $\frac{X_1+X_2+\dots+X_n}{n}$.

Υποθέτοντας ότι η μηδενική υπόθεση, $H_0 : p = p_0$, είναι αληθής, δηλαδή, ότι το τυχαίο δείγμα, X_1, X_2, \dots, X_n , προέρχεται από έναν πληθυσμό που ακολουθεί την κατανομή *Bernoulli* με μέση τιμή $\mu = p_0$ και διασπορά $\sigma^2 = p_0(1 - p_0)$, η τυχαία μεταβλητή $\frac{X_1+X_2+\dots+X_n}{n}$, δηλαδή, η δειγματική μέση τιμή \bar{X} , για μεγάλα n , προσεγγίζεται ικανοποιητικά (βλ. *Κ.Ο.Θ.*) από την κανονική κατανομή με μέση τιμή p_0 και διασπορά $\frac{p_0(1-p_0)}{n}$. Κατά προσέγγιση, δηλαδή, έχουμε,

$$\bar{X} = \frac{X_1+X_2+\dots+X_n}{n} \sim N\left(p_0, \frac{p_0(1-p_0)}{n}\right).$$

Επομένως, για τον έλεγχο της μηδενικής υπόθεσης $H_0 : p = p_0$ μπορούμε, ως στατιστική συνάρτηση ελέγχου, να χρησιμοποιήσουμε την:

$$\frac{\bar{X} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \rightarrow Z \sim N(0, 1).$$

Σημειώνουμε ότι η τιμή, $\frac{x_1+x_2+\dots+x_n}{n}$, του δειγματικού ποσοστού $\frac{X_1+X_2+\dots+X_n}{n}$, δηλαδή της δειγματικής μέσης τιμής \bar{X} , στο συγκεκριμένο τυχαίο

¹⁷ Θυμηθείτε ότι η τυχαία μεταβλητή X_i είναι δίτιμη και εκφράζει τον αριθμό των επιτυχιών στη δοκιμή i , επομένως παίρνει τιμή ή 0 (καμία επιτυχία) ή 1 (μία επιτυχία). Έτσι, το άθροισμα $X_1 + X_2 + \dots + X_n$ εκφράζει τον αριθμό των επιτυχιών σε n δοκιμές.

δείγμα που παίρνουμε, έχει επικρατήσει να συμβολίζεται με \hat{p} (αντί \bar{x}) και αντίστοιχα, η τιμή της στατιστικής συνάρτησης ελέγχου με $z = \frac{(\hat{p} - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$.

Έτσι, σε επίπεδο σημαντικότητας α , για μεγάλο μέγεθος δείγματος n (θεωρητικά $n \rightarrow +\infty$, στην πράξη $n \cdot p_0 > 5$ και $n \cdot (1-p_0) > 5$ ή $n \cdot p_0 \cdot (1-p_0) \geq 10$)¹⁸, απορρίπτουμε την $H_0 : p = p_0$,

- έναντι της $H_1 : p > p_0$, όταν,

$$z = \frac{(\hat{p} - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} \geq z_\alpha$$

- έναντι της $H_1 : p < p_0$, όταν,

$$z = \frac{(\hat{p} - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} \leq -z_\alpha$$

- έναντι της $H_1 : p \neq p_0$, όταν,

$$|z| = \frac{|(\hat{p} - p_0)\sqrt{n}|}{\sqrt{p_0(1-p_0)}} \geq z_{\alpha/2} \Leftrightarrow \frac{(\hat{p} - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} \geq z_{\alpha/2} \quad \text{ή} \quad \frac{(\hat{p} - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} \leq -z_{\alpha/2}$$

Ας ολοκληρώσουμε, στο παράδειγμά μας, τον έλεγχο της $H_0 : p = 0.9$ έναντι της εναλλακτικής $H_1 : p < 0.9$ σε επίπεδο σημαντικότητας $\alpha = 0.05$.

Το δειγματικό ποσοστό είναι $\hat{p} = \frac{38}{52} = 0.73$ και επομένως η τιμή της στατιστικής

συνάρτησης ελέγχου στο δείγμα είναι $z = \frac{(\hat{p} - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} = \frac{(0.73 - 0.9)\sqrt{52}}{\sqrt{0.9 \cdot (1-0.9)}} = 4.09$. Η

περιοχή απόρριψης του ελέγχου είναι $z \geq z_{0.05}$ ή $z \geq 1.645$ και επειδή, $z = 4.09 \geq 1.645$, σε επίπεδο σημαντικότητας $\alpha = 0.05$ απορρίπτουμε τη μηδενική υπόθεση. Δηλαδή, ο αγρότης, με πιθανότητα να κάνει λάθος το πολύ 5%, αμφιβάλει δικαιολογημένα.

Σημείωση: Αν το μέγεθος του δείγματος δεν είναι μεγάλο, η κανονική προσέγγιση που χρησιμοποιήσαμε δε μπορεί να εφαρμοσθεί. Στην περίπτωση αυτή, για να ορισθεί η περιοχή απόρριψης μπορεί να χρησιμοποιηθεί η ακριβής κατανομή του αθροίσματος $X_1 + X_2 + \dots + X_n \sim B(n, p_0)$ (δες και την Άσκηση-1 όταν μιλήσαμε για τη διωνυμική κατανομή). Όμως, στην περίπτωση αυτή, με τη διαδικασία ελέγχου που περιγράψαμε, δε μπορούμε πάντοτε να πετύχουμε ένα προκαθορισμένο επίπεδο σημαντικότητας!!! Αυτό το πρόβλημα αντιμετωπίζεται, όμως, δε θα επεκταθούμε περισσότερο. Σημειώνουμε, τέλος, ότι σε προβλήματα ελέγχου ποσοστού, στην πράξη συνήθως δεν αντιμετωπίζουμε πρόβλημα μικρών δειγμάτων.

Στην ενότητα που ακολουθεί, παρουσιάζουμε τον έλεγχο της διασποράς, σ^2 , ενός πληθυσμού που ακολουθεί κανονική κατανομή. Ολοκληρώνουμε έτσι, την παρουσίαση των ελέγχων που αναφέρονται σε έναν πληθυσμό.

¹⁸ Θυμηθείτε ότι στην περίπτωση αυτή, το πόσο μεγάλο πρέπει να είναι το n , επηρεάζεται από την τιμή του p .

5. Στατιστικός έλεγχος υποθέσεων για τη διασπορά, σ^2 , ενός πληθυσμού που ακολουθεί κανονική κατανομή

Όπως ήδη έχουμε διαπιστώσει, για τον έλεγχο της μέσης τιμής ενός πληθυσμού, όταν δε γνωρίζουμε τη διασπορά του (που είναι και το σύνηθες) είναι αναγκαία μια εκτίμησή της. Επίσης, στους ελέγχους που αφορούν στη σύγκριση των μέσων τιμών δύο πληθυσμών, όπως θα διαπιστώσουμε στη συνέχεια, κάνουμε υποθέσεις για τις διασπορές τους. Ο έλεγχος, όμως, και η εκτίμηση της διασποράς ενός πληθυσμού δεν προκύπτει μόνο ως ένα δευτερογενές ερώτημα. Σε πολλά προβλήματα (όπως, στον έλεγχο της ακρίβειας οργάνων μέτρησης, στον έλεγχο προδιαγραφών και γενικότερα στον έλεγχο ποιότητας), μας ενδιαφέρει ο ίδιος ο έλεγχος της διασποράς (καθαυτός). Ας δούμε ένα παράδειγμα.

Μια αυτόματη μηχανή συσκευάζει καλαμπόκι σε τσουβάλια των 25Kgr. Η μηχανή έχει ρυθμισθεί έτσι, ώστε τα βάρη του καλαμποκιού στα τσουβάλια να έχουν τυπική απόκλιση 1.5Kgr. Επίσης, έχει παρατηρηθεί, ότι τα βάρη αυτά ακολουθούν κανονική κατανομή. Ο υπεύθυνος παραγωγής υποψιάζεται ότι η μηχανή έχει απορυθμισθεί και θέλει να ελέγξει αν η τυπική απόκλιση των βαρών είναι πράγματι 1.5Kgr. Για το σκοπό αυτό, από την παραγωγή μιας ημέρας, επέλεξε τυχαία 30 τσουβάλια, κατέγραψε τα βάρη τους και υπολόγισε το μέσο βάρος τους, $\bar{x} = 24.8Kgr$, και την τυπική τους απόκλιση, $s = 1.6Kgr$. Σε επίπεδο σημαντικότητας 5%, υποστηρίζουν αυτά τα δεδομένα τις υποψίες του υπεύθυνου παραγωγής;

Είναι φανερό ότι πρέπει να κάνουμε τον έλεγχο της μηδενικής υπόθεσης $H_0 : \sigma^2 = 1.5^2$, έναντι της εναλλακτικής υπόθεσης $H_0 : \sigma^2 \neq 1.5^2$.

Ας δούμε, γενικότερα, με ποια στατιστική συνάρτηση ελέγχου μπορούμε να κάνουμε τον έλεγχο της μηδενικής υπόθεσης $H_0 : \sigma^2 = \sigma_0^2$.

Όταν μιλήσαμε για τις δειγματοσυναρτήσεις, είδαμε ότι αν X_1, X_2, \dots, X_n ένα τυχαίο δείγμα από έναν πληθυσμό που ακολουθεί κανονική κατανομή με διασπορά σ^2 , τότε για οποιοδήποτε μέγεθος δείγματος n , η τυχαία μεταβλητή

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

ακολουθεί την κατανομή χ_{n-1}^2 , δηλαδή, μια κατανομή *χι τετράγωνο (chi-square)* με $n-1$ βαθμούς ελευθερίας. Υπενθυμίζουμε ότι με S^2 συμβολίζουμε την (αμερόληπτη)

$$\text{δειγματική διασπορά } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Με την υπόθεση ότι η μηδενική υπόθεση $H_0 : \sigma^2 = \sigma_0^2$ είναι αληθής, η τιμή της

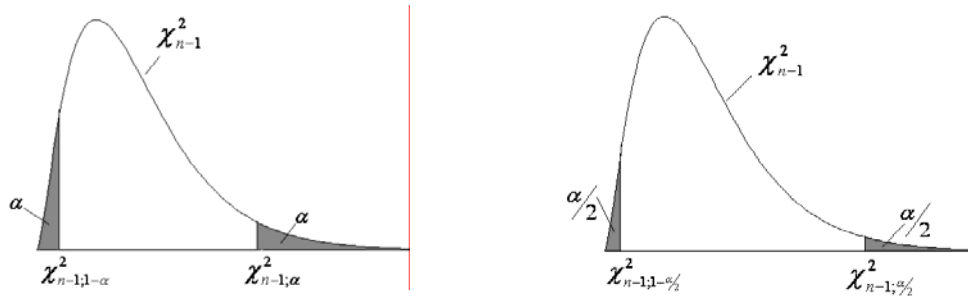
$$\text{δειγματοσυναρτησης } \chi^2 = \frac{(n-1)S^2}{\sigma^2} \text{ μπορεί να υπολογισθεί από το δείγμα (αφού } \sigma^2$$

γνωστή, ίση με σ_0^2) και επειδή η κατανομή της είναι γνωστή, είναι λογικό, ως

$$\text{στατιστική συνάρτηση ελέγχου να χρησιμοποιήσουμε την } \chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi_{n-1}^2.$$

Είναι επίσης λογικό, ως κρίσιμα σημεία, για τον ορισμό των περιοχών απόρριψης, σε επίπεδο σημαντικότητας α , να προκύψουν τα σημεία $\chi_{n-1, \alpha}^2$, $\chi_{n-1, 1-\alpha}^2$, $\chi_{n-1, \frac{\alpha}{2}}^2$ και $\chi_{n-1, 1-\frac{\alpha}{2}}^2$ της κατανομής χ_{n-1}^2 . Όπως φαίνεται στα σχήματα που ακολουθούν, δεξιά

των σημείων $\chi^2_{n-1;\alpha}$ και $\chi^2_{n-1;\alpha/2}$ ορίζονται περιοχές εμβαδού, αντίστοιχα, α και $\alpha/2$. Ομοίως, δεξιά των σημείων $\chi^2_{n-1;1-\alpha}$ και $\chi^2_{n-1;1-\alpha/2}$ ορίζονται περιοχές εμβαδού, αντίστοιχα, $1-\alpha$ και $1-\alpha/2$ και επομένως αριστερά, περιοχές εμβαδού α και $\alpha/2$.



Έτσι, σε επίπεδο σημαντικότητας α , απορρίπτουμε τη μηδενική υπόθεση $H_0 : \sigma^2 = \sigma_0^2$,

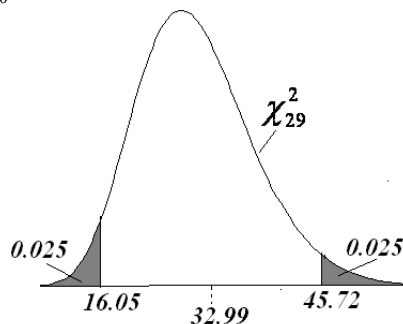
- έναντι της $H_1 : \sigma^2 > \sigma_0^2$, όταν,
$$\frac{(n-1)s^2}{\sigma_0^2} \geq \chi^2_{n-1;\alpha}$$
- έναντι της $H_1 : \sigma^2 < \sigma_0^2$, όταν,
$$\frac{(n-1)s^2}{\sigma_0^2} \leq \chi^2_{n-1;1-\alpha}$$
- έναντι της $H_1 : \sigma^2 \neq \sigma_0^2$, όταν,
$$\frac{(n-1)s^2}{\sigma_0^2} \geq \chi^2_{n-1;\alpha/2} \quad \text{ή} \quad \frac{(n-1)s^2}{\sigma_0^2} \leq \chi^2_{n-1;1-\alpha/2}$$

Ας ολοκληρώσουμε τον έλεγχο του παραδείγματός μας.

Σε επίπεδο σημαντικότητας $\alpha = 0.05$, η περιοχή απόρριψης του ελέγχου είναι

$$\frac{(n-1)s^2}{\sigma_0^2} \geq \chi^2_{29;0.025} \quad \text{ή} \quad \frac{(n-1)s^2}{\sigma_0^2} \leq \chi^2_{29;0.975} \quad \Leftrightarrow$$

$$\frac{(n-1)s^2}{\sigma_0^2} \geq 45.72 \quad \text{ή} \quad \frac{(n-1)s^2}{\sigma_0^2} \leq 16.047.$$



Επειδή η τιμή της στατιστικής συνάρτησης ελέγχου στο δείγμα, $\frac{29 \cdot 1.6^2}{1.5^2} = 32.99$, δε

βρίσκεται στην περιοχή απόρριψης, η μηδενική υπόθεση, σε επίπεδο σημαντικότητας $\alpha = 0.05$, δεν απορρίπτεται, δηλαδή, τα δεδομένα του δείγματος, σε επίπεδο σημαντικότητας $\alpha = 0.05$, δε δίνουν επαρκείς αποδείξεις ότι η μηχανή έχει απορρυθμισθεί.

Η μέθοδος στατιστικού ελέγχου υποθέσεων που περιγράψαμε, εφαρμόζεται και για τη σύγκριση παραμέτρων δύο πληθυσμών. Οι δύο επόμενες ενότητες αναφέρονται σε αυτές τις περιπτώσεις ελέγχων.

6. Στατιστικός έλεγχος υποθέσεων για τη διαφορά των μέσων τιμών δύο πληθυσμών

Συχνά, σε έρευνες και μελέτες, παρουσιάζεται ανάγκη να αποφασίσουμε μεταξύ δύο υποθέσεων για τη διαφορά, $\mu_1 - \mu_2$, της μέσης τιμής μ_2 ενός πληθυσμού από τη μέση τιμή μ_1 ενός άλλου πληθυσμού. Για παράδειγμα, σε μια μελέτη για την περιεκτικότητα των ζαχαρότευτλων σε σάκχαρα, πρέπει να αποφασίσουμε αν η μέση περιεκτικότητα μιας ποικιλίας ζαχαρότευτλων σε σάκχαρα, μ_1 , είναι ίση ή όχι με τη μέση περιεκτικότητα μιας άλλης ποικιλίας ζαχαρότευτλων, μ_2 , δηλαδή αν $\mu_1 - \mu_2 = 0$ ή αν $\mu_1 - \mu_2 \neq 0$. Επίσης, σε μια πειραματική έρευνα για την ποσότητα γάλακτος που παράγει μια συγκεκριμένη φυλή αγελάδων, πρέπει να ελέγξουμε αν με μια συγκεκριμένη βελτίωση του διαιτολογίου η μέση ημερήσια παραγωγή ανά αγελάδα αυξάνεται περισσότερο από 5Kgr. Δηλαδή, αν μ_1 είναι η μέση ημερήσια παραγωγή όταν οι αγελάδες εκτρέφονται με βελτιωμένο διαιτολόγιο και μ_2 η μέση ημερήσια παραγωγή όταν οι αγελάδες δεν εκτρέφονται με βελτιωμένο διαιτολόγιο, πρέπει να ελέγξουμε αν $\mu_1 - \mu_2 \leq 5$ ή αν $\mu_1 - \mu_2 > 5$. Σε μια άλλη έρευνα, ενδιαφερόμαστε να συγκρίνουμε το μέσο επίπεδο ενός αιματολογικού δείκτη στους ενήλικες άνδρες που πάσχουν από κάποια συγκεκριμένη ασθένεια, με το μέσο επίπεδο του αιματολογικού δείκτη στους ενήλικες άνδρες που δεν πάσχουν από τη συγκεκριμένη ασθένεια.

Είναι προφανές, ότι τέτοιου είδους προβλήματα είναι προβλήματα στατιστικού ελέγχου υποθέσεων. Μπορούμε όμως, και σε αυτές τις περιπτώσεις, να εφαρμόσουμε τη μέθοδο στατιστικού ελέγχου υποθέσεων που περιγράψαμε στα προηγούμενα; Η απάντηση είναι ναι. Η μέθοδος στατιστικού ελέγχου υποθέσεων που εφαρμόζουμε και σε αυτές τις περιπτώσεις ελέγχων, δε διαφέρει, ως προς τη λογική της, από αυτήν που εφαρμόσαμε για τον έλεγχο μιας παραμέτρου ενός μόνο πληθυσμού. Με μια, όμως, προφανή διαφορά. Η απόφασή μας για το αν θα απορρίπτουμε ή δε θα απορρίπτουμε τη μηδενική υπόθεση, δε θα βασίζεται πλέον σε ένα αλλά σε δύο τυχαία δείγματα (ένα από κάθε πληθυσμό) και επομένως, όπως είδαμε όταν μιλήσαμε για διαστήματα εμπιστοσύνης που αφορούν παραμέτρους δύο πληθυσμών, πρέπει να διακρίνουμε την περίπτωση που τα δύο δείγματα είναι ανεξάρτητα μεταξύ τους από την περίπτωση που τα δύο δείγματα δεν είναι ανεξάρτητα.

6α. Ανεξάρτητα δείγματα (Independent samples)

Ας δούμε ένα παράδειγμα.

Παράδειγμα-1: Η εταιρεία ύδρευσης μιας πόλης, ενδιαφέρεται να γνωρίζει τη συγκέντρωση μολύβδου στο πόσιμο νερό που φθάνει, μέσω του δικτύου της, στους καταναλωτές και για το σκοπό αυτό κάνει τακτικούς δειγματοληπτικούς ελέγχους. Επειδή το νερό που φθάνει στις διάφορες περιοχές (πολεοδομικούς τομείς) της πόλης (κέντρο, ανατολικά προάστια, νότια προάστια, κ.τ.λ.) δεν προέρχεται από τους ίδιους ταμιευτήρες όπως επίσης και το δίκτυο ύδρευσης δε βρίσκεται στην ίδια κατάσταση σε όλους τους τομείς της πόλης (ως προς την παλαιότητα, τη συντήρηση, τα υλικά κατασκευής κ.τ.λ.), οι δειγματοληπτικοί έλεγχοι γίνονται ανά πολεοδομικό τομέα. Σε

πρόσφατο δειγματοληπτικό έλεγχο, ένα δείγμα 100 παροχών από το κέντρο της πόλης έδωσε μέση συγκέντρωση μολύβδου 36ppm με τυπική απόκλιση 6ppm και ένα δείγμα 90 παροχών από τα ανατολικά προάστια έδωσε μέση συγκέντρωση 34.1ppm με τυπική απόκλιση 5.9ppm. Αραγε, αυτή η διαφορά, των 1.9ppm, μεταξύ των δύο δειγμάτων, είναι στατιστικά σημαντική; Δηλαδή, προέκυψε γιατί η μέση συγκέντρωση μολύβδου στο κέντρο της πόλης, πράγματι διαφέρει από τη μέση συγκέντρωση στα ανατολικά προάστια, ή προέκυψε τυχαία (οφείλεται στην τύχη);

Είναι προφανές ότι πρόκειται για ένα πρόβλημα σύγκρισης των μέσων τιμών δύο πληθυσμών με βάση δύο δείγματα, ένα από κάθε πληθυσμό. Ο ένας πληθυσμός, ας τον συμβολίσουμε με A , είναι οι τιμές της συγκέντρωσης μολύβδου στο πόσιμο νερό που φθάνει σε όλες τις παροχές του κέντρου της πόλης και ο άλλος, έστω B , οι τιμές της συγκέντρωσης μολύβδου στο πόσιμο νερό που φθάνει σε όλες τις παροχές των ανατολικών προαστίων της πόλης. Ας συμβολίσουμε με μ_A την άγνωστη μέση τιμή του πληθυσμού A και με μ_B , την επίσης άγνωστη, μέση τιμή του πληθυσμού B .

Για να απαντήσουμε στο πρόβλημα που τίθεται, πρέπει να κάνουμε τον έλεγχο της μηδενικής υπόθεσης $H_0 : \mu_A - \mu_B = 0$, έναντι της εναλλακτικής $H_1 : \mu_A - \mu_B \neq 0$. Ο έλεγχος, φυσικά, θα γίνει με βάση αυτό που παρατηρείται στα δύο δείγματα, στο δείγμα X_1, X_2, \dots, X_{100} από τον πληθυσμό A και στο δείγμα Y_1, Y_2, \dots, Y_{90} από τον πληθυσμό B .

Ως στατιστική συνάρτηση ελέγχου, είναι λογικό, να χρησιμοποιήσουμε την τυχαία μεταβλητή $\bar{X} - \bar{Y}$, δηλαδή, την αντίστοιχη διαφορά των δειγματικών μέσων \bar{X} και \bar{Y} .

Όταν μιλήσαμε για τις δειγματοσυναρτήσεις, είδαμε ότι αν X_1, X_2, \dots, X_{n_1} και Y_1, Y_2, \dots, Y_{n_2} , δύο τυχαία δείγματα που έχουν ληφθεί ανεξάρτητα το ένα από το άλλο από δύο πληθυσμούς A και B αντίστοιχα, τότε γνωρίζουμε ή μπορούμε ικανοποιητικά να προσεγγίσουμε την κατανομή της διαφοράς, $\bar{X} - \bar{Y}$, των αντίστοιχων δειγματικών μέσων, στις ακόλουθες περιπτώσεις:

- όταν τα δείγματα προέρχονται από κανονικούς πληθυσμούς με γνωστές διασπορές
- όταν τα δείγματα προέρχονται από κανονικούς πληθυσμούς με άγνωστες αλλά ίσες διασπορές
- όταν τα δείγματα είναι μεγάλα και οι διασπορές των πληθυσμών είναι γνωστές (και οι πληθυσμοί όχι κατ' ανάγκη κανονικοί)
- όταν τα δείγματα είναι μεγάλα και οι διασπορές των πληθυσμών είναι άγνωστες (και οι πληθυσμοί όχι κατ' ανάγκη κανονικοί).

Πριν ολοκληρώσουμε τον έλεγχο του παραδείγματός μας, θα δώσουμε, για κάθε μια από τις παραπάνω περιπτώσεις, την περιοχή απόρριψης, για τον έλεγχο της μηδενικής υπόθεσης,

$$H_0 : \mu_1 - \mu_2 = \delta$$

όπου, δ , η συγκεκριμένη τιμή της διαφοράς των μέσων τιμών των δύο πληθυσμών που ελέγχουμε (στο παράδειγμά μας, είναι $\delta = 0$).

(α) Τα δείγματα προέρχονται από κανονικούς πληθυσμούς με γνωστές διασπορές

Έστω δύο ανεξάρτητα τυχαία δείγματα, X_1, X_2, \dots, X_{n_1} και Y_1, Y_2, \dots, Y_{n_2} από δύο κανονικούς πληθυσμούς με γνωστές διασπορές, δηλαδή, $X_1, X_2, \dots, X_{n_1} \sim N(\mu_1, \sigma_1^2)$ και $Y_1, Y_2, \dots, Y_{n_2} \sim N(\mu_2, \sigma_2^2)$ με τις διασπορές σ_1^2 και σ_2^2 γνωστές. Στην περίπτωση αυτή είναι γνωστό ότι η κατανομή της τυχαίας μεταβλητής $\bar{X} - \bar{Y}$ είναι κανονική με $\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$ και επομένως, $\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = Z \sim N(0,1)$.

Επειδή, οι διασπορές σ_1^2 και σ_2^2 είναι γνωστές, με την υπόθεση ότι η μηδενική υπόθεση είναι αληθής, δηλαδή, με την υπόθεση ότι η διαφορά των μέσων τιμών είναι ίση με δ , στην $Z = \frac{\bar{X} - \bar{Y} - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ δεν υπάρχουν άγνωστοι παράμετροι και επομένως

μπορεί να χρησιμοποιηθεί ως στατιστική συνάρτηση ελέγχου αφού η τιμή της, z , μπορεί να υπολογισθεί από τα δείγματα και η κατανομή της είναι γνωστή. Έτσι, για τον έλεγχο της μηδενικής υπόθεσης $H_0 : \mu_1 - \mu_2 = \delta$, έχουμε:

Σε επίπεδο σημαντικότητας α , απορρίπτουμε την $H_0 : \mu_1 - \mu_2 = \delta$,

- έναντι της $H_0 : \mu_1 - \mu_2 > \delta$, όταν,

$$z = \frac{\bar{x} - \bar{y} - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \geq z_\alpha$$

- έναντι της $H_0 : \mu_1 - \mu_2 < \delta$, όταν,

$$z = \frac{\bar{x} - \bar{y} - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq -z_\alpha$$

- έναντι της $H_0 : \mu_1 - \mu_2 \neq \delta$, όταν,

$$|z| = \frac{|\bar{x} - \bar{y} - \delta|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \geq z_{\alpha/2}$$

Σημείωση: Στην περίπτωση που οι διασπορές των πληθυσμών, σ_1^2 και σ_2^2 , είναι γνωστές και τα μεγέθη των δειγμάτων, n_1 και n_2 , είναι μεγάλα (θεωρητικά $n_1 \rightarrow +\infty$ και $n_2 \rightarrow +\infty$, στην πράξη $n_1 \geq 30$ και $n_2 \geq 30$), οι παραπάνω περιοχές απόρριψης ισχύουν για οτιδήποτε πληθυσμούς, όχι κατ' ανάγκη κανονικούς. Όμως, στην περίπτωση αυτή, οι αντίστοιχοι έλεγχοι είναι επιπέδου σημαντικότητας α κατά προσέγγιση. Φυσικά, όσο μεγαλύτερα είναι τα n_1 και n_2 , τόσο καλύτερη είναι η προσέγγιση.

(β) Τα δείγματα προέρχονται από κανονικούς πληθυσμούς με άγνωστες αλλά ίσες διασπορές

Έστω δύο ανεξάρτητα τυχαία δείγματα, X_1, X_2, \dots, X_{n_1} και Y_1, Y_2, \dots, Y_{n_2} από δύο κανονικούς πληθυσμούς με άγνωστες αλλά ίσες διασπορές, δηλαδή, $X_1, X_2, \dots, X_{n_1} \sim N(\mu_1, \sigma^2)$ και $Y_1, Y_2, \dots, Y_{n_2} \sim N(\mu_2, \sigma^2)$. Στην περίπτωση αυτή (όπως είδαμε όταν μιλήσαμε για τις δειγματοσυναρτήσεις), η τυχαία μεταβλητή, $T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, για οτιδήποτε n_1 και n_2 (μικρά ή μεγάλα), ακολουθεί την

κατανομή $t_{n_1+n_2-2}$ (κατανομή Student με $n_1 + n_2 - 2$ βαθμούς ελευθερίας), δηλαδή,

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

όπου, $S^2 = \frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{n_1 + n_2 - 2}$ μια εκτιμήτρια, με βάση και τα δύο δείγματα, της κοινής διασποράς σ^2 των δύο πληθυσμών.

Με την υπόθεση ότι η μηδενική υπόθεση είναι αληθής, δηλαδή, με την υπόθεση ότι η διαφορά των μέσων τιμών είναι ίση με δ , στην $T = \frac{\bar{X} - \bar{Y} - \delta}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ δεν υπάρχουν

άγνωστοι παράμετροι και επομένως μπορεί να χρησιμοποιηθεί ως στατιστική συνάρτηση ελέγχου αφού η τιμή της, t , μπορεί να υπολογισθεί από τα δείγματα και η κατανομή της είναι γνωστή. Έτσι, για τον έλεγχο της μηδενικής υπόθεσης $H_0 : \mu_1 - \mu_2 = \delta$, έχουμε:

Σε επίπεδο σημαντικότητας α , απορρίπτουμε την $H_0 : \mu_1 - \mu_2 = \delta$,

- έναντι της $H_0 : \mu_1 - \mu_2 > \delta$, όταν,

$$t = \frac{\bar{x} - \bar{y} - \delta}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \geq t_{n_1+n_2-2; \alpha}$$

- έναντι της $H_0 : \mu_1 - \mu_2 < \delta$, όταν,

$$t = \frac{\bar{x} - \bar{y} - \delta}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq -t_{n_1+n_2-2; \alpha}$$

- έναντι της $H_0 : \mu_1 - \mu_2 \neq \delta$, όταν,

$$|t| = \frac{|\bar{x} - \bar{y} - \delta|}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \geq t_{n_1+n_2-2; \frac{\alpha}{2}}$$

Σημείωση: Για τις δύο προϋποθέσεις εφαρμογής του ελέγχου αυτής της περίπτωσης (κανονικοί πληθυσμοί και ίσες διασπορές) επισημαίνουμε τα εξής: α) Ακόμη και αν οι πληθυσμοί δεν είναι κανονικοί (και εφόσον, φυσικά, δεν απέχουν δραματικά από την κανονική κατανομή), ο έλεγχος δίνει καλά αποτελέσματα (δες και σχετικό σχόλιο στην

περίπτωση του *t*-test για ένα πληθυσμό). β) Αν οι δύο πληθυσμοί δεν έχουν ίσες διασπορές, ο παραπάνω έλεγχος δε μπορεί να εφαρμοσθεί. Για το σχετικό πρόβλημα (γνωστό ως πρόβλημα των Behrens-Fisher), μόνο προσεγγιστικές λύσεις έχουν προταθεί και παραμένει ανοιχτό.

(γ) Τα δείγματα είναι μεγάλα και προέρχονται από πληθυσμούς όχι κατ' ανάγκη κανονικούς

Έστω δύο ανεξάρτητα τυχαία δείγματα, X_1, X_2, \dots, X_{n_1} και Y_1, Y_2, \dots, Y_{n_2} από δύο πληθυσμούς με κατανομές, αντίστοιχα, F_1 και F_2 (οποιοσδήποτε). Υποθέτουμε ότι τα *μεγέθη των δειγμάτων*, n_1 και n_2 , είναι μεγάλα (θεωρητικά $n_1 \rightarrow +\infty$ και $n_2 \rightarrow +\infty$, στην πράξη $n_1 \geq 30$ και $n_2 \geq 30$).

Αν οι διασπορές σ_1^2 και σ_2^2 των δύο πληθυσμών είναι γνωστές, όπως ήδη αναφέραμε, εφαρμόζεται (κατά προσέγγιση) ο έλεγχος της περίπτωσης **(α)**.

Αν οι διασπορές σ_1^2 και σ_2^2 των δύο πληθυσμών είναι άγνωστες, ως *στατιστική συνάρτηση ελέγχου* χρησιμοποιούμε την τυχαία μεταβλητή $\frac{\bar{X} - \bar{Y} - \delta}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ (όπου S_1^2 και

S_2^2 οι δειγματικές διασπορές-αμερόληπτες εκτιμήτριες των σ_1^2 και σ_2^2) η οποία ακολουθεί, κατά προσέγγιση, την *τυποποιημένη κανονική κατανομή*, δηλαδή,

$$\frac{\bar{X} - \bar{Y} - \delta}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \rightarrow Z \sim N(0,1)$$

Έτσι, για τον έλεγχο της μηδενικής υπόθεσης $H_0 : \mu_1 - \mu_2 = \delta$, έχουμε:

Σε επίπεδο σημαντικότητας α , απορρίπτουμε την $H_0 : \mu_1 - \mu_2 = \delta$,

- έναντι της $H_0 : \mu_1 - \mu_2 > \delta$, όταν,

$$z = \frac{\bar{x} - \bar{y} - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \geq z_\alpha$$

- έναντι της $H_0 : \mu_1 - \mu_2 < \delta$, όταν,

$$z = \frac{\bar{x} - \bar{y} - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \leq -z_\alpha$$

- έναντι της $H_0 : \mu_1 - \mu_2 \neq \delta$, όταν,

$$|z| = \frac{|\bar{x} - \bar{y} - \delta|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \geq z_{\alpha/2}$$

Σημείωση: Επειδή στην περίπτωση αυτή, η στατιστική συνάρτηση ελέγχου,

$\frac{\bar{X} - \bar{Y} - \delta}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \rightarrow Z \sim N(0,1)$, δεν είναι κανονική αλλά προσεγγίζεται ικανοποιητικά από

την κανονική, οι αντίστοιχοι έλεγχοι είναι επιπέδου σημαντικότητας α κατά προσέγγιση. Φυσικά, όσο μεγαλύτερα είναι τα n_1 και n_2 , τόσο καλύτερη είναι η προσέγγιση.

Ας ολοκληρώσουμε, τον έλεγχο του παραδείγματός μας, σε επίπεδο σημαντικότητας $\alpha = 5\%$.

Από τη διατύπωση του προβλήματος, είναι προφανές ότι τα δύο δείγματα είναι μεταξύ τους ανεξάρτητα. Επίσης, οι κατανομές και οι διασπορές των πληθυσμών από τους οποίους προέρχονται τα δείγματα είναι άγνωστες.

Επειδή τα μεγέθη των δειγμάτων είναι μεγάλα ($n_A = 100 \geq 30$ και $n_B = 90 \geq 30$), πρόκειται για έλεγχο της περίπτωσης (γ) και επομένως, σε επίπεδο σημαντικότητας $\alpha = 5\%$, η περιοχή απόρριψης της $H_0 : \mu_A - \mu_B = 0$, έναντι της $H_1 : \mu_A - \mu_B \neq 0$, είναι:

$$|z| = \frac{|\bar{x} - \bar{y} - \delta|}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} \geq z_{0.025} \Leftrightarrow \frac{\bar{x} - \bar{y} - \delta}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} \geq z_{0.025} \quad \text{ή} \quad \frac{\bar{x} - \bar{y} - \delta}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} \leq -z_{0.025}$$

όπου, $\delta = 0$, $\bar{x} = 36 \text{ ppm}$, $\bar{y} = 34.1 \text{ ppm}$, $s_A = 6 \text{ ppm}$, $s_B = 5.9 \text{ ppm}$, $n_A = 100$ και $n_B = 90$.

Επειδή, $z = \frac{\bar{x} - \bar{y} - \delta}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} = \frac{36 - 34.1 - 0}{\sqrt{\frac{6^2}{100} + \frac{5.9^2}{90}}} \approx 2.2 \geq z_{0.025} = 1.96$, η τιμή της στατιστικής

συνάρτησης ελέγχου βρίσκεται στην περιοχή απόρριψης και επομένως, σε επίπεδο σημαντικότητας $\alpha = 5\%$, απορρίπτουμε τη μηδενική υπόθεση. Δηλαδή, τα ευρήματα στα δύο δείγματα δίνουν στατιστικά σημαντικές αποδείξεις ότι η μέση συγκέντρωση μολύβδου στο πόσιμο νερό διαφέρει στις δύο περιοχές της πόλης. Η πιθανότητα αυτό το συμπέρασμα να είναι λάθος, είναι το πολύ 5%.

Παρατήρηση-1: Αν η εταιρεία ύδρευσης θέλει να ελέγξει αν η μέση συγκέντρωση μολύβδου στο πόσιμο νερό που φθάνει στο κέντρο της πόλης (πληθυσμός A) είναι μεγαλύτερη από τη μέση συγκέντρωση μολύβδου στο πόσιμο νερό που φθάνει στα ανατολικά προάστια, πρέπει να κάνει τον έλεγχο της ίδιας μηδενικής υπόθεσης $H_0 : \mu_A - \mu_B = 0$, έναντι όμως της εναλλακτικής $H_1 : \mu_A - \mu_B > 0$. Φυσικά, και στην περίπτωση αυτή, η μηδενική υπόθεση, σε επίπεδο σημαντικότητας 5%, απορρίπτεται (αφού στον αμφίπλευρο έλεγχο απορρίπτεται). Δηλαδή, τα ευρήματα στα δύο δείγματα δίνουν στατιστικά σημαντικές αποδείξεις ότι η μέση συγκέντρωση μολύβδου στο πόσιμο νερό που φθάνει στο κέντρο, είναι μεγαλύτερη από τη μέση συγκέντρωση μολύβδου στο πόσιμο νερό που φθάνει στα ανατολικά προάστια. Η πιθανότητα αυτό το συμπέρασμα να είναι λάθος, είναι το πολύ 5%.

Παρατήρηση-2: Η εταιρεία ύδρευσης θέλει να ελέγξει αν η διαφορά των μέσων συγκεντρώσεων μολύβδου μεταξύ των δύο περιοχών ξεπερνά τα 2 ppm. Στην περίπτωση αυτή πρέπει να κάνει τον έλεγχο της μηδενικής υπόθεσης $H_0 : \mu_A - \mu_B = 2$ έναντι της εναλλακτικής $H_1 : \mu_A - \mu_B > 2$. Σε επίπεδο σημαντικότητας $\alpha = 5\%$, η περιοχή απόρριψης της μηδενικής υπόθεσης είναι,

$$z = \frac{\bar{x} - \bar{y} - \delta}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} \geq z_{0.05} = 1.645$$

και επειδή η τιμή $z = \frac{36 - 34.1 - 2}{\sqrt{\frac{6^2}{100} + \frac{5.9^2}{90}}} = -0.1157$ της στατιστικής συνάρτησης ελέγχου,

δε βρίσκεται στην περιοχή απόρριψης, η μηδενική υπόθεση, σε επίπεδο σημαντικότητας $\alpha = 5\%$, δεν απορρίπτεται. Δηλαδή, σε επίπεδο σημαντικότητας 5% , τα ευρήματα στα δύο δείγματα δε δίνουν στατιστικά σημαντικές αποδείξεις ότι η διαφορά των μέσων συγκεντρώσεων μολύβδου μεταξύ των δύο περιοχών ξεπερνά τα $2ppm$.

Ας δούμε δύο ακόμη παραδείγματα.

Παράδειγμα-2: Οι αγρότες σε μια αγροτική περιοχή καλλιεργούν παραδοσιακά την ποικιλία A ενός φυτού. Την τελευταία χρονιά, τα κτήματα 10 αγροτών από αυτή την περιοχή (που επελέγησαν σύμφωνα με ένα σχέδιο τυχαίας δειγματοληψίας) είχαν μέση στρεμματική απόδοση $8.5Kgr/στρέμμα$ με τυπική απόκλιση $1.2Kgr/στρέμμα$. Σε μια γειτονική αγροτική περιοχή οι αγρότες καλλιεργούν μια άλλη ποικιλία του φυτού, έστω B . Η μέση στρεμματική απόδοση στα κτήματα 15 αγροτών αυτής της περιοχής (που επελέγησαν επίσης τυχαία) ήταν την τελευταία χρονιά $11Kgr/στρέμμα$ με τυπική απόκλιση $1.1Kgr/στρέμμα$. Άραγε, η διαφορά που παρατηρείται μεταξύ των δύο δειγμάτων είναι στατιστικά σημαντική ή μήπως οφείλεται στην τύχη.

Ας συμβολίσουμε με X την απόδοση (ανά στρέμμα) της ποικιλίας A και με Y την απόδοση (ανά στρέμμα) της ποικιλίας B . Αν μ_A η άγνωστη μέση τιμή της X και μ_B η επίσης άγνωστη μέση τιμή της Y , είναι λογικό, να κάνουμε τον έλεγχο της μηδενικής υπόθεσης $H_0 : \mu_A - \mu_B = 0$, έναντι της εναλλακτικής $H_1 : \mu_A - \mu_B < 0$. Δηλαδή, να ελέγξουμε αν αυτό που παρατηρείται στα δύο δείγματα υποστηρίζει απόρριψη της μηδενικής υπόθεσης ότι οι μέσες αποδόσεις είναι ίσες έναντι της εναλλακτικής ότι η μέση απόδοση της ποικιλίας A είναι μικρότερη από τη μέση απόδοση της ποικιλίας B . Φυσικά, όπως έχουμε εξηγήσει, αυτό πρέπει να γίνει σε κάποιο επίπεδο σημαντικότητας, έστω $\alpha = 5\%$ (αφού δεν είναι δυνατόν να είμαστε βέβαιοι για την απάντησή μας).

Από τη διατύπωση του προβλήματος, είναι προφανές ότι τα δύο δείγματα X_1, X_2, \dots, X_{10} και Y_1, Y_2, \dots, Y_{15} έχουν ληφθεί το ένα ανεξάρτητα από το άλλο και επομένως θα εξετάσουμε σε ποια από τις τρεις περιπτώσεις ελέγχων για ανεξάρτητα δείγματα που παρουσιάσαμε προηγουμένως μπορεί να ενταχθεί ο έλεγχος του προβλήματός μας. Επειδή τα δείγματα είναι μικρά και για τους πληθυσμούς από τους οποίους προέρχονται δεν γνωρίζουμε αν είναι κανονικοί (ούτε κάτι για τις διασπορές τους), θα χρησιμοποιήσουμε την περιοχή απόρριψης της περίπτωσης (β) κάνοντας όμως δύο παραδοχές: τα δείγματα προέρχονται από κανονικούς πληθυσμούς με ίσες διασπορές¹⁹.

¹⁹ Δείτε τη σχετική σημείωση στην περίπτωση (β) στα προηγούμενα. Δείτε επίσης όσα αναφέραμε για ανάλογο πρόβλημα που αντιμετωπίσαμε στην Εφαρμογή-2 (βιομηχανικά απόβλητα στον Καλαμά) στη 2^η Ενότητα.

Έτσι, σε επίπεδο σημαντικότητας $\alpha = 5\%$, η περιοχή απόρριψης της $H_0 : \mu_A - \mu_B = 0$, έναντι της $H_1 : \mu_A - \mu_B < 0$ είναι:

$$t = \frac{\bar{x} - \bar{y} - \delta}{s \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \leq -t_{n_A+n_B-2;0.05}$$

όπου, $\delta = 0$, $\bar{x} = 8.5 \text{Kgr} / \text{στρ.}$, $\bar{y} = 11 \text{Kgr} / \text{στρ.}$, $s^2 = \frac{(n_A - 1) \cdot s_A^2 + (n_B - 1) \cdot s_B^2}{n_A + n_B - 2}$,

$n_A = 10$, $n_B = 15$, $s_A = 1.2 \text{Kgr} / \text{στρ.}$ και $s_B = 1.1 \text{Kgr} / \text{στρ.}$.

Δηλαδή,

$$t = \frac{\bar{x} - \bar{y} - \delta}{s \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \leq -t_{23;0.05} \quad \text{ή} \quad t = \frac{\bar{x} - \bar{y} - \delta}{s \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \leq -1.714$$

Υπολογίζουμε την τιμή της εκτιμήτριας της κοινής διασποράς,

$$s^2 = \frac{(n_A - 1) \cdot s_A^2 + (n_B - 1) \cdot s_B^2}{n_A + n_B - 2} = \frac{9 \cdot 1.2^2 + 14 \cdot 1.1^2}{10 + 15 - 2} = 1.3 \Rightarrow s = 1.14 \quad \text{και} \quad \text{επομένως} \quad \eta$$

τιμή της στατιστικής συνάρτησης ελέγχου είναι,

$$t = \frac{\bar{x} - \bar{y} - \delta}{s \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} = \frac{8.5 - 11 - 0}{1.14 \cdot \sqrt{\frac{1}{10} + \frac{1}{15}}} = \frac{-2.5}{0.47} = -5.37.$$

Επειδή, $t = -5.37 \leq -1.714$, η τιμή της στατιστικής συνάρτησης ελέγχου βρίσκεται στην περιοχή απόρριψης, και επομένως σε επίπεδο σημαντικότητας $\alpha = 5\%$ απορρίπτουμε τη μηδενική υπόθεση. Δηλαδή, τα ευρήματα στα δύο δείγματα δίνουν στατιστικά σημαντικές αποδείξεις ότι η μέση στρεμματική απόδοση της ποικιλίας A είναι μικρότερη από τη μέση στρεμματική απόδοση της ποικιλίας B. Η πιθανότητα αυτό το συμπέρασμα να είναι λάθος, είναι το πολύ 5%.

Ερώτηση-1: Μήπως τα ευρήματα στα δύο δείγματα είναι ακόμη πιο σημαντικά; Πόσο σημαντικά είναι; Σκεφθείτε τι πρέπει να κάνετε για να απαντήσετε σε αυτό το ερώτημα.

Ερώτηση-2: Μόλις διαπιστώσαμε ότι τα ευρήματα στα δύο δείγματα δίνουν στατιστικά σημαντικές αποδείξεις ότι η μέση στρεμματική απόδοση της ποικιλίας A είναι μικρότερη από τη μέση στρεμματική απόδοση της ποικιλίας B. Αν μάλιστα απαντήσουμε και στην προηγούμενη Ερώτηση-1, θα διαπιστώσουμε ότι είναι πολύ σημαντικές αυτές οι αποδείξεις (P -Τιμή < 0.0001). Βέβαια, μπορεί η διαφορά στις μέσες αποδόσεις να είναι (με βάση τα ευρήματα στα δύο δείγματα) στατιστικά σημαντική όμως για τους αγρότες που καλλιεργούν την ποικιλία A, αυτή η διαφορά είναι σημαντική μόνο αν είναι μεγαλύτερη από 2Kgr/στρέμμα γιατί κρίνουν ότι μόνο τότε αξίζει να αλλάξουν τις συνθήκες τους... Τι έλεγχο πρέπει να κάνουμε για να δώσουμε μια απάντηση στους αγρότες;

Ερώτηση-3: Αν κάποιος αγρότης ισχυρισθεί ότι η απόδοση της ποικιλίας B είναι μεγαλύτερη γιατί στη γειτονική περιοχή που καλλιεργείται, οι καλλιεργητικές συνθήκες είναι πιο ευνοϊκές, τι μπορούμε να απαντήσουμε;

Απάντηση: Δίνεται στο τέλος της ενότητας.

Παράδειγμα-3: Προκειμένου ένας φοιτητής του Τμήματος Επιστήμης και Τεχνολογίας Τροφίμων του Γεωπονικού Πανεπιστημίου Αθηνών να συγκρίνει, στο πλαίσιο της πτυχιακής του εργασίας, την αποτελεσματικότητα δύο διαιτολογίων στη μείωση του βάρους, εργάστηκε ως εξής. Επέλεξε τυχαία 10 άτομα (άνδρες και γυναίκες) τα οποία δέχθηκαν να συμμετέχουν στην έρευνα και στα 5 από αυτά τα οποία επίσης επέλεξε τυχαία (από τα 10) έδωσε να ακολουθήσουν το ένα διαιτολόγιο, έστω *A*, ενώ στα υπόλοιπα 5 έδωσε το άλλο διαιτολόγιο, έστω *B*. Στον πίνακα που ακολουθεί φαίνεται, για καθένα από τα δέκα άτομα που επελέγησαν, ο κωδικός, το φύλο, η ηλικία, το ύψος, το βάρος και ο δείκτης μάζας σώματος (BMI)²⁰.

Κωδικός Ατόμου	Φύλο	Ηλικία (έτη)	Ύψος (m)	Βάρος (Kgr)	BMI
1	A	65	1.77	76.14	24.30
2	Γ	41	1.65	61.58	22.62
3	A	23	1.83	97.04	28.98
4	Γ	63	1.62	77.26	29.44
5	Γ	59	1.59	71.66	28.35
6	A	22	1.89	94.43	26.44
7	Γ	19	1.56	66.44	27.30
8	Γ	38	1.68	63.45	22.48
9	A	62	1.74	79.13	26.14
10	Γ	23	1.62	72.78	27.73

Τρεις μήνες μετά την εφαρμογή των διαιτολογίων, η μέση μείωση του βάρους για το πρώτο δείγμα (διαιτολόγιο *A*) ήταν 3.1Kgr με τυπική απόκλιση 1.3Kgr και για το δεύτερο δείγμα (διαιτολόγιο *B*), η μέση μείωση ήταν 1.7Kgr με τυπική απόκλιση 1.2Kgr. Τα ευρήματα στα δύο δείγματα, μαρτυρούν άραγε ότι το διαιτολόγιο *A* είναι πιο αποτελεσματικό από το *B*; ($\alpha = 5\%$).

Ας συμβολίσουμε με *X* τη μείωση του βάρους όταν εφαρμόζεται το διαιτολόγιο *A* και με *Y* τη μείωση του βάρους όταν εφαρμόζεται το διαιτολόγιο *B*. Έστω μ_A η άγνωστη μέση τιμή της *X* και μ_B η επίσης άγνωστη μέση τιμή της *Y*. Σύμφωνα με το ερώτημα που τίθεται, πρέπει να κάνουμε τον έλεγχο της μηδενικής υπόθεσης $H_0 : \mu_A - \mu_B = 0$, έναντι της εναλλακτικής $H_1 : \mu_A - \mu_B > 0$.

Από τη διατύπωση του προβλήματος είναι προφανές ότι τα δύο δείγματα έχουν ληφθεί το ένα ανεξάρτητα από το άλλο και επομένως μπορούμε να εφαρμόσουμε το *t-test* για δύο ανεξάρτητα δείγματα. Κάνοντας τις παραδοχές που κάναμε και στο παράδειγμα-2, μπορούμε να χρησιμοποιήσουμε την περιοχή απόρριψης της περίπτωσης (β)²¹:

²⁰ Ο Δείκτης Μάζας Σώματος (Body Mass Index) ορίζεται ως εξής: $BMI = \frac{\text{βάρος } Kgr}{(\text{ύψος})^2 m^2}$. Αν για ένα

άτομο η τιμή του BMI είναι κάτω από 18.5 τότε βρίσκεται στην ομάδα των ατόμων με **κάτω του φυσιολογικού βάρους**, αν είναι μεταξύ 18.5 και 25 βρίσκεται στην ομάδα ατόμων με **φυσιολογικό βάρος**, αν είναι μεταξύ 25 και 30 βρίσκεται στην ομάδα των **υπέρβαρων** ατόμων, αν είναι μεταξύ 30 και 40 βρίσκεται στην ομάδα των **παχύσαρκων** ατόμων και αν είναι πάνω από 40 βρίσκεται στην ομάδα των **σοβαρά παχύσαρκων**.

²¹ Υπενθυμίζουμε και πάλι ότι η Στατιστική προσφέρει μεθόδους και εργαλεία για τον έλεγχο αυτών των παραδοχών καθώς και εναλλακτικούς ελέγχους στην περίπτωση που αυτές οι παραδοχές δεν ευσταθούν. Όμως αυτά θα τα δούμε στα επόμενα, σε άλλες ενότητες.

$$t = \frac{\bar{x} - \bar{y} - \delta}{s \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \geq t_{n_1+n_2-2; \alpha}$$

όπου, $\delta = 0$, $\bar{x} = 3.1Kgr$, $\bar{y} = 1.7Kgr$, $s^2 = \frac{(n_A - 1) \cdot s_A^2 + (n_B - 1) \cdot s_B^2}{n_A + n_B - 2}$, $n_A = 5$, $n_B = 5$, $s_A = 1.3Kgr$, $s_B = 1.2Kgr$ και $\alpha = 5\%$.

Δηλαδή, η περιοχή απόρριψης είναι,

$$t = \frac{\bar{x} - \bar{y}}{s \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \geq t_{8; 0.05} \quad \text{ή} \quad t = \frac{\bar{x} - \bar{y}}{s \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \geq 1.860.$$

Υπολογίζουμε την τιμή της εκτιμήτριας της κοινής διασποράς,

$$s^2 = \frac{(n_A - 1) \cdot s_A^2 + (n_B - 1) \cdot s_B^2}{n_A + n_B - 2} = \frac{4 \cdot 1.3^2 + 4 \cdot 1.2^2}{5 + 5 - 2} = 1.57 \Rightarrow s = 1.25 \quad \text{και επομένως η}$$

τιμή της στατιστικής συνάρτησης ελέγχου είναι,

$$t = \frac{\bar{x} - \bar{y}}{s \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} = \frac{3.1 - 1.7}{1.25 \cdot \sqrt{\frac{1}{5} + \frac{1}{5}}} = \frac{1.4}{0.79} = 1.77.$$

Επειδή, η τιμή της στατιστικής συνάρτησης ελέγχου, $t = 1.77$, δε βρίσκεται στην περιοχή απόρριψης, σε επίπεδο σημαντικότητας $\alpha = 5\%$ δεν απορρίπτουμε τη μηδενική υπόθεση ότι τα δύο διαιτολόγια έχουν την ίδια αποτελεσματικότητα.

Παρατήρηση: Ο επιβλέπων καθηγητής, όταν ο φοιτητής του παρουσίασε το παραπάνω συμπέρασμα, ζήτησε να δει για κάθε δείγμα τα πρωτογενή δεδομένα, δηλαδή, τη μείωση του βάρους καθενός από τα 5 άτομα κάθε δείγματος. Ζήτησε επίσης να δει το φύλο, την ηλικία, το ύψος, το αρχικό βάρος και το δείκτη μάζας σώματος κάθε ατόμου. Όλα τα στοιχεία που ζήτησε ο καθηγητής φαίνονται στον πίνακα που ακολουθεί.

Κωδικός Ατόμου	Φύλο	Ηλικία (έτη)	Ύψος (m)	Βάρος (Kgr)	BMI	Μείωση Βάρους (Kgr)
Διαιτολόγιο Α						
2	Γ	41	1.65	61.58	22.62	1.46
4	Γ	63	1.62	77.26	29.44	3.92
7	Γ	19	1.56	66.44	27.30	4.74
9	Α	62	1.74	79.13	26.14	2.28
3	Α	23	1.83	97.04	28.98	3.10
Διαιτολόγιο Β						
5	Γ	59	1.59	71.66	28.35	3.22
6	Α	22	1.89	94.43	26.44	1.70
1	Α	65	1.77	76.14	24.30	0.18
10	Γ	23	1.62	72.78	27.73	2.46
8	Γ	38	1.68	63.45	22.48	0.94

Αφού ο καθηγητής μελέτησε τα στοιχεία του πίνακα, είπε στον φοιτητή ότι δεν είναι ικανοποιημένος και πρέπει να ξανασκεφθεί το πρόβλημα.

Ας δούμε τι προβλήματα είχε ο καθηγητής και τι μπορεί να κάνει ο φοιτητής.

Ο καθηγητής παρατήρησε ότι η διαφορά $\bar{x} - \bar{y}$ είναι εμφανώς διαφορετική από το μηδέν και μάλιστα, η μέση μείωση του βάρους στο δείγμα που δόθηκε το διαιτολόγιο

A ($\bar{x} = 3.1\text{kg}$) είναι σχεδόν διπλάσια από τη μέση μείωση του βάρους στο δείγμα που δόθηκε το διαιτολόγιο B ($\bar{y} = 1.7\text{kg}$). Εντούτοις, με το t -test για δύο ανεξάρτητα δείγματα που εφάρμοσε ο φοιτητής, η διαφορά αυτή σε επίπεδο σημαντικότητας $\alpha = 5\%$ δεν κρίθηκε στατιστικά σημαντική. Βέβαια, παρατηρώντας και τις τυπικές αποκλίσεις στα δύο δείγματα (1.3 και 1.2Kg αντίστοιχα) ο καθηγητής κατάλαβε τι έχει συμβεί.

Ας δούμε πιο προσεκτικά τον τύπο που δίνει την τιμή t στον έλεγχο για δύο ανεξάρτητα δείγματα που χρησιμοποίησε ο φοιτητής,

$$t = \frac{\bar{x} - \bar{y}}{s \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$$

Αν οι τυπικές αποκλίσεις s_A και s_B είναι μεγάλες, θα είναι, επίσης, μεγάλη και η τιμή s^2 της εκτιμήτριας της κοινής διασποράς, δηλαδή, το s που βρίσκεται στον παρανομαστή του τύπου είναι επίσης αρκετά μεγάλο. Όμως, αν το s είναι αρκετά μεγάλο, η τιμή t της στατιστικής συνάρτησης ελέγχου μπορεί να γίνει αρκετά μικρή ακόμη και αν η διαφορά $\bar{x} - \bar{y}$ είναι εμφανώς διαφορετική από το μηδέν (ιδιαίτερα αν και τα μεγέθη των δειγμάτων είναι μικρά). Δηλαδή, μια μεγάλη μεταβλητότητα εντός των δειγμάτων είναι ικανή να «υπονομεύσει/κρύψει» τη μεταβλητότητα μεταξύ των δειγμάτων. Κατά συνέπεια, είναι δυνατόν η μηδενική υπόθεση να μην απορριφθεί (ενώ θα έπρεπε) επειδή υπάρχει μεγάλη μεταβλητότητα εντός των δειγμάτων. Ακριβώς αυτό συνέβη και στο συγκεκριμένο πρόβλημα της πτυχιακής εργασίας του φοιτητή. Τα δείγματα (και τα δύο) έχουν μεγάλη μεταβλητότητα (ο συντελεστής μεταβλητότητας²² στα δύο δείγματα είναι, αντίστοιχα, 41.9% και 70.6%). Έτσι, η μεγάλη τιμή του s , σε συνδυασμό με τα μικρά μεγέθη των δύο δειγμάτων, «υπονόμισαν» τη διαφορά μεταξύ των δύο δειγμάτων και μάλιστα αρκετά ώστε να μην απορριφθεί η μηδενική υπόθεση!

Σημείωση: Για το πώς αξιοποιείται η σχέση της μεταβλητότητας μεταξύ δύο (ή περισσότερων) ανεξάρτητων δειγμάτων με τη μεταβλητότητα εντός των δειγμάτων θα αναφερθούμε και πάλι στα επόμενα όταν θα μιλήσουμε για την Ανάλυση Διασποράς.

Ο καθηγητής ζήτησε να δει τα αναλυτικά στοιχεία που φαίνονται στον πίνακα, για να ελέγξει αν η μεγάλη μεταβλητότητα που παρατηρήθηκε εντός των δειγμάτων μπορεί να ερμηνευθεί/αιτιολογηθεί με όρους της επιστημονικής περιοχής που εντάσσεται το συγκεκριμένο πρόβλημα. Διαπίστωσε ότι πράγματι ερμηνεύεται. Η μείωση του βάρους τους τρεις πρώτους μήνες, οποιαδήποτε δίαιτα και αν εφαρμοσθεί, επηρεάζεται και από κάποια χαρακτηριστικά όπως ο ΒΜΙ, η ηλικία και το φύλο. Αν παρατηρήσουμε τα στοιχεία του πίνακα, εύκολα θα διαπιστώσουμε ότι υπάρχουν άτομα με τα χαρακτηριστικά αυτά πολύ διαφορετικά. Αυτό συμβαίνει και στα δύο δείγματα (λογικό, αφού τα δείγματα είναι τυχαία). Για παράδειγμα, στο δείγμα που ακολούθησε το διαιτολόγιο A , το άτομο με κωδικό 2 (η πειραματική μονάδα 2) είναι γυναίκα, μέσης ηλικίας με φυσιολογικό βάρος ενώ το άτομο με κωδικό 9 είναι άνδρας, ηλικιωμένος και υπέρβαρος. Με οποιαδήποτε δίαιτα, η μείωση του βάρους μετά τρεις μήνες στα δύο αυτά άτομα θα διαφέρει αρκετά. Επίσης στο δείγμα που ακολούθησε το διαιτολόγιο B , το άτομο με κωδικό 10 είναι γυναίκα, νέα και υπέρβαρη ενώ το

²² $CV = \frac{s}{\bar{x}} \cdot 100\%$

άτομο με κωδικό 1 είναι *άνδρας, ηλικιωμένος με φυσιολογικό βάρος*. Είναι επομένως λογικές και αναμενόμενες οι μεγάλες μεταβλητότητες εντός των δύο δειγμάτων.

Αφού πλέον κατανοήσαμε τι συνέβη, πρέπει να δούμε πώς μπορούμε να αντιμετωπίσουμε το πρόβλημα. Δηλαδή, την «υπονόμηση» του *t-test* με δύο ανεξάρτητα δείγματα από τη μεγάλη μεταβλητότητα εντός των δύο δειγμάτων.

1^η επιλογή: Μπορούμε να επαναλάβουμε το πείραμα με δύο ανεξάρτητα και πάλι δείγματα, αλλά μεγαλύτερου μεγέθους. Για να κατανοήσετε γιατί είναι λογική αυτή η επιλογή, δείτε πάλι τον τύπο της *στατιστικής συνάρτησης ελέγχου* και θυμηθείτε όσα είχαμε αναφέρει όταν μιλήσαμε για την *ισχύ* ενός ελέγχου, δηλαδή, την ικανότητά του να απορρίπτει τη *μηδενική υπόθεση* όταν είναι αληθής η *εναλλακτική*. Σκεφθείτε επίσης όσα είχαμε αναφέρει στην εισαγωγή για τις πηγές των *δειγματοληπτικών σφαλμάτων*.

Για να δούμε ένα αριθμητικό παράδειγμα, και επειδή δεν έχουμε μετρήσεις από μεγαλύτερα δείγματα, ας θεωρήσουμε ότι τα ευρήματα $\bar{x} = 3.1\text{kg}$, $\bar{y} = 1.7\text{kg}$, $s_A = 1.3\text{kg}$ και $s_B = 1.2\text{kg}$ δεν προέρχονται από δύο ανεξάρτητα δείγματα μεγέθους 5 αλλά από δύο ανεξάρτητα δείγματα μεγέθους 10. Στην περίπτωση αυτή, η *περιοχή απόρριψης* είναι,

$$t = \frac{\bar{x} - \bar{y}}{s \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \geq t_{18;0.05} \quad \text{ή} \quad t = \frac{\bar{x} - \bar{y}}{s \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \geq 1.734$$

και η τιμή της *στατιστικής συνάρτησης ελέγχου* είναι,

$$t = \frac{\bar{x} - \bar{y}}{s \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} = \frac{3.1 - 1.7}{1.25 \cdot \sqrt{\frac{1}{10} + \frac{1}{10}}} = \frac{1.4}{0.79} = 2.5.$$

Προφανώς, η τιμή της *στατιστικής συνάρτησης ελέγχου*, $t = 2.5$, βρίσκεται στην *περιοχή απόρριψης* και επομένως, τώρα, σε *επίπεδο σημαντικότητας* $\alpha = 5\%$, απορρίπτουμε τη *μηδενική υπόθεση* ότι τα δύο διαιτολόγια έχουν την ίδια αποτελεσματικότητα. Δηλαδή, αυξάνοντας τα μεγέθη των δειγμάτων αυξήσαμε την *ισχύ* του ελέγχου. Όμως, ας έχουμε πάντοτε υπόψη μας και όσα αναφέραμε για τη «σκοτεινή πλευρά» των μεγάλων δειγμάτων όταν μιλήσαμε για την *ισχύ* των ελέγχων.

2^η επιλογή: Μπορούμε να σχεδιάσουμε το πείραμα διαφορετικά! Να κάνουμε τη σύγκριση των δύο διαιτολογίων όχι με *ανεξάρτητα δείγματα* αλλά με *δείγματα από ζευγαρωτές παρατηρήσεις (εξαρτημένα δείγματα)*. Ας δούμε αυτή την επιλογή.

6β. Εξαρτημένα δείγματα ή δείγματα από ζευγαρωτές παρατηρήσεις (Dependent samples, paired samples, matched samples)

Όπως ήδη εξηγήσαμε, η μείωση του βάρους (υπό οποιαδήποτε διαίτα) επηρεάζεται και από παράγοντες όπως ο BMI, η ηλικία και το φύλο. Έτσι, η μείωση του βάρους π.χ. του ατόμου με κωδικό 7 που βρέθηκε ίση με 4.74kg , δεν είναι αποτέλεσμα μόνο της επίδρασης του διαιτολογίου *A* που ακολούθησε αυτό το άτομο αλλά και της επίδρασης των συγκεκριμένων χαρακτηριστικών του (*γυναίκα, νέα, υπέρβαρη*). Δηλαδή, στην τιμή 4.74kg είναι ενσωματωμένη, εκτός από την επίδραση του διαιτολογίου *A* (που μας είναι άγνωστη και τη διερευνούμε), και η επίδραση των συγκεκριμένων χαρακτηριστικών του ατόμου (που επίσης είναι άγνωστη). Επίσης, η μείωση του βάρους του ατόμου με κωδικό 2 είναι ίση με 1.46kg , αρκετά μικρότερη

από τη μείωση του βάρους του ατόμου 7 παρότι και τα δύο άτομα ακολούθησαν την ίδια δίαιτα. Αυτό συνέβη γιατί τα συγκεκριμένα χαρακτηριστικά στο άτομο με κωδικό 2 (γυναίκα, μέσης ηλικίας, με φυσιολογικό βάρος) διαφέρουν αρκετά από αυτά στο άτομο με κωδικό 7. Αντίστοιχες διαφοροποιήσεις στη μείωση του βάρους παρατηρούνται και στα άτομα με διαφορετικά τα συγκεκριμένα χαρακτηριστικά, που ακολούθησαν τη δίαιτα Β.

Στα ανεξάρτητα τυχαία δείγματα (επιλογή-1) η κεντρική ιδέα είναι ότι ως προς τους «εξωγενείς» παράγοντες που επηρεάζουν το αποτέλεσμα (δηλαδή, εκτός αυτού του οποίου μελετάμε την επίδραση) αναμένεται να επιτευχθεί και στα δύο δείγματα μια μέση αντιπροσωπευτικότητα, ακριβώς επειδή αυτά είναι τυχαία (στο παράδειγμά μας, ως προς τον BMI, την ηλικία και το φύλο). Φυσικά, όσο μεγαλύτερα είναι τα τυχαία δείγματα και όσο πιο ομοιογενής (ως προς αυτούς τους παράγοντες) είναι ο πληθυσμός από τον οποίο προέρχονται, τόσο καλύτερα επιτυγχάνεται αυτή η αντιπροσωπευτικότητα. Βέβαια, στις μετρήσεις που παίρνουμε (και στα δύο δείγματα) η άγνωστη επίδραση των «εξωγενών» παραγόντων παραμένει ενσωματωμένη. Εύλογα, επομένως, γεννάται το ερώτημα: μήπως μπορούμε να «προστατευθούμε» από αυτή την ανεπιθύμητη επίδραση ώστε η σύγκριση να αφορά μόνο την επίδραση του παράγοντα που μελετάμε; Η απάντηση είναι ναι και πολύ απλή.

Φυσικά, δε μπορούμε να «εξαφανίσουμε» τη διαφορετικότητα και την ποικιλομορφία ως προς τους εξωγενείς παράγοντες. Όμως μπορούμε, τις δύο διαφορετικές αγωγές/θεραπείες/μεταχειρίσεις/επεμβάσεις (*treatments*) των οποίων την αποτελεσματικότητα μελετάμε (στο παράδειγμά μας τα διαιτολόγια Α και Β) να τις εφαρμόσουμε υπό τις ίδιες συνθήκες δηλαδή σε συνθήκες που αυτοί οι παράγοντες είναι ίδιοι ή σχεδόν ίδιοι. Αν για παράδειγμα, πρόκειται για άτομα, να εφαρμόσουμε και τις δύο αγωγές στο ίδιο άτομο σε διαφορετικές χρονικές περιόδους ή από μία σε καθένα από δύο άτομα με ίδιους (ή σχεδόν ίδιους) τους υπόλοιπους παράγοντες που επιδρούν στο αποτέλεσμα. Με αυτό τον τρόπο δημιουργούμε n ζεύγη παρατηρήσεων (x_i, y_i) , $i = 1, 2, \dots, n$, όπου x_i είναι η μέτρηση που παίρνουμε με εφαρμογή της μιας αγωγής στο ένα άτομο του ζεύγους και y_i η μέτρηση που παίρνουμε με εφαρμογή της άλλης αγωγής στο άλλο άτομο του ίδιου ζεύγους (που μπορεί να είναι και το ίδιο κάποια άλλη χρονική στιγμή ή περίοδο).

Έτσι, στο παράδειγμά μας, ο φοιτητής θα μπορούσε, να εργασθεί ως εξής: να δημιουργήσει ζεύγη ατόμων με ίδια (ή σχεδόν ίδια) τα χαρακτηριστικά (τους παράγοντες) που επηρεάζουν τη μείωση του βάρους (πέραν του διαιτολογίου), για παράδειγμα, ένα ζεύγος αποτελούμενο από δύο γυναίκες, μέσης ηλικίας με φυσιολογικό βάρος, ένα άλλο ζεύγος αποτελούμενο από δύο άνδρες, νέους και υπέρβαρους, ένα τρίτο από δύο γυναίκες, ηλικιωμένες και υπέρβαρες, κ.τλ. και στο ένα άτομο κάθε ζεύγους να δώσει το διαιτολόγιο Α ενώ στο δεύτερο άτομο κάθε ζεύγους να δώσει το διαιτολόγιο Β. Εναλλακτικά, θα μπορούσε να επιλέξει ένα τυχαίο δείγμα ατόμων και σε κάθε άτομο να δώσει για τρεις μήνες το διαιτολόγιο Α και για ένα επόμενο τρίμηνο να δώσει σε καθένα επίσης άτομο το διαιτολόγιο Β.

Με αυτό τον τρόπο, εντός κάθε ζεύγους, τα αποτελέσματα των δύο διαφορετικών επεμβάσεων (διαιτολογίων στο παράδειγμά μας) μετρώνται σε ίδιες συνθήκες. Είναι προφανές, ότι εργαζόμενοι έτσι, επιτυγχάνουμε το εξής: για κάθε ζεύγος παρατηρήσεων (x_i, y_i) η διαφορά $x_i - y_i$ εκφράζει μόνο (ακριβέστερα, κυρίως) την

επίδραση της επέμβασης (του διαιτολογίου) αφού τα άλλα χαρακτηριστικά που επηρεάζουν τη μείωση του βάρους είναι και στα δύο άτομα του ζεύγους ίδια (ή σχεδόν ίδια) και επομένως η άγνωστη επίδρασή τους, που φυσικά είναι ενσωματωμένη στα x_i και y_i , όσο μεγάλη ή μικρή και αν είναι, θα είναι ίδια και στο x_i και στο y_i και με την αφαίρεση θα απαλείφεται²³. Έτσι, η σύγκριση των δύο επεμβάσεων γίνεται πλέον εντός κάθε ζεύγους σε ίδιες συνθήκες, ενώ επιτρέπεται και η ύπαρξη ποικίλων συνθηκών αφού τα ζεύγη μεταξύ τους είναι ανόμοια.

Ας δούμε τώρα πώς θα κάνουμε τον απαιτούμενο *στατιστικό έλεγχο*. Έστω X η μέτρηση που παίρνουμε υπό την επέμβαση A και Y η μέτρηση που παίρνουμε υπό την επέμβαση B (στο παράδειγμά μας, η μείωση του βάρους υπό το διαιτολόγιο A και B αντίστοιχα). Παίρνοντας τα δύο δείγματα X_1, X_2, \dots, X_n και Y_1, Y_2, \dots, Y_n με τον τρόπο που περιγράψαμε (τα οποία προφανώς πρέπει να είναι του ίδιου μεγέθους $n = n_A = n_B$) μπορούμε να τα θεωρήσουμε ως ζεύγη $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ τα οποία είναι ανεξάρτητα το ένα από το άλλο ενώ τα X_i και Y_i εντός του ίδιου ζεύγους δεν είναι ανεξάρτητα (δε μπορούν να θεωρηθούν ανεξάρτητα). Για κάθε ζεύγος (X_i, Y_i) , σχηματίζουμε τη διαφορά $D_i = X_i - Y_i$ και πλέον μπορούμε να εργασθούμε με ένα δείγμα, αυτό των διαφορών D_1, D_2, \dots, D_n , το οποίο θεωρούμε ότι προέρχεται από ένα θεωρητικό πληθυσμό (τον πληθυσμό των διαφορών) με *μέση τιμή* $\mu_D = \mu_A - \mu_B$ (γιατί;) όπου μ_A η *μέση τιμή* της X και μ_B η *μέση τιμή* της Y . Έτσι, ο (γνωστός) έλεγχος της *μηδενικής υπόθεσης*,

$$H_0 : \mu_D = 0$$

είναι ισοδύναμος με τον έλεγχο της *μηδενικής υπόθεσης* $H_0 : \mu_A - \mu_B = 0$ ή γενικότερα, ο έλεγχος της *μηδενικής υπόθεσης*,

$$H_0 : \mu_D = \delta$$

είναι ισοδύναμος με τον έλεγχο της *μηδενικής υπόθεσης* $H_0 : \mu_A - \mu_B = \delta$.

Αν η *διασπορά* του πληθυσμού των διαφορών, σ_D^2 , είναι άγνωστη (που είναι και το σύνηθες) ως *στατιστική συνάρτηση ελέγχου* μπορούμε (κατά τα γνωστά) να χρησιμοποιήσουμε την

$$T = \frac{(\bar{D} - \delta) \cdot \sqrt{n}}{S_D}$$

όπου, \bar{D} και S_D^2 η *μέση τιμή* και η *διασπορά* του δείγματος των διαφορών αντίστοιχα.

Αν επομένως το δείγμα των διαφορών προέρχεται από κανονικό πληθυσμό, η *μηδενική υπόθεση* $H_0 : \mu_D = \delta$, απορρίπτεται σε *επίπεδο σημαντικότητας* α ,

- έναντι της $H_1 : \mu_D > \delta$, όταν, $t = \frac{(\bar{d} - \delta) \cdot \sqrt{n}}{s_d} \geq t_{n-1; \alpha}$
- έναντι της $H_1 : \mu_D < \delta$, όταν, $t = \frac{(\bar{d} - \delta) \cdot \sqrt{n}}{s_d} \leq -t_{n-1; \alpha}$

²³ Αυτό που μόλις περιγράψαμε θα το δούμε και διατυπωμένο ως *στοχαστικό μοντέλο* όταν θα μιλήσουμε στα επόμενα για την *Ανάλυση Διασποράς*.

- έναντι της $H_1 : \mu_D \neq \delta$, όταν, $t = \frac{|\bar{d} - \delta| \cdot \sqrt{n}}{s_d} \geq t_{n-1; \frac{\alpha}{2}}$

Με \bar{d} και s_D^2 συμβολίζουμε, αντίστοιχα, την τιμή της τυχαίας μεταβλητής \bar{D} και την τιμή της τυχαίας μεταβλητής S_D^2 στο συγκεκριμένο δείγμα διαφορών που πήραμε.

Σημείωση: Σημειώνουμε ότι ισχύουν και εδώ όλες οι παρατηρήσεις που αναφέραμε όταν μιλήσαμε για τον έλεγχο της μέσης τιμής ενός πληθυσμού.

Ο έλεγχος που μόλις περιγράψαμε, στη βιβλιογραφία είναι γνωστός ως, **έλεγχος t κατά ζεύγη (paired t test)**.

Ας δούμε δύο παραδείγματα (στο πρόβλημα του φοιτητή θα επανέλθουμε αργότερα).

Παράδειγμα-4: Ένας ερευνητής θέλει να συγκρίνει τις αποδόσεις (ανά στρέμμα) δύο ποικιλιών σταριού στον κάμπο της Θεσσαλίας. Για το σκοπό αυτό, σχεδίασε ένα πείραμα ως εξής: Επέλεξε 10 αγρούς σε δέκα διαφορετικές τοποθεσίες του Θεσσαλικού κάμπου και κάθε αγρό τον χώρισε σε δύο αγροτεμάχια ίδιου σχήματος και ίδιου εμβαδού. Στο ένα αγροτεμάχιο κάθε αγρού καλλιέργησε στάρι της μιας ποικιλίας, έστω A , και στο άλλο αγροτεμάχιο καλλιέργησε στάρι της άλλης ποικιλίας, έστω B . Σε ποιο από τα δύο αγροτεμάχια κάθε αγρού καλλιέργησε την ποικιλία A και σε ποιο την ποικιλία B το αποφάσισε με τυχαίο τρόπο. Επίσης φρόντισε στα δύο αγροτεμάχια κάθε αγρού να υπάρχουν ίδιες καλλιεργητικές συνθήκες και ίδιες συνθήκες συγκομιδής (γονιμότητα εδάφους, υγρασία, προσανατολισμός, χρόνος σποράς, καλλιεργητική μέθοδος, λίπανση, ημέρα θερισμού, κ.τλ.). Στον πίνακα που ακολουθεί φαίνεται η απόδοση των δύο ποικιλιών σε καθέναν από τους 10 πειραματικούς αγρούς.

Αγρός (i)	1	2	3	4	5	6	7	8	9	10
Απόδοση ποικιλίας A (Kgr/στρ.) (x_i)	500	650	490	570	555	545	535	635	625	540
Απόδοση ποικιλίας B (Kgr/στρ.) (y_i)	455	620	455	610	505	495	515	600	600	510

Τα ευρήματα στα δύο δείγματα μαρτυρούν άραγε ότι οι μέσες αποδόσεις των δύο ποικιλιών στον κάμπο της Θεσσαλίας διαφέρουν; ($\alpha = 5\%$)

Ο ερευνητής γνωρίζει ότι η απόδοση της καλλιέργειας επηρεάζεται, εκτός από τον παράγοντα ποικιλία, την επίδραση του οποίου θέλει να μελετήσει, και από άλλους παράγοντες όπως η γονιμότητα του εδάφους, η υγρασία, η λίπανση, κ.τλ. Έτσι, για να αναδειχθεί μια πιθανή διαφορά στην απόδοση που οφείλεται μόνο (ή κυρίως) στη διαφορετική ποικιλία, σχεδίασε το πείραμα έτσι ώστε η καλλιέργεια των δύο ποικιλιών να γίνεται κάθε φορά υπό τις ίδιες συνθήκες. Φρόντισε δηλαδή, σε καθένα από τα δέκα διαφορετικά ζεύγη αγροτεμαχίων να υπάρχουν ίδιες καλλιεργητικές συνθήκες και να δημιουργήσει έτσι δέκα ζεύγη παρατηρήσεων. Πρόκειται επομένως για σύγκριση κατά ζεύγη.

Έστω X η απόδοση (σε Kgr/στρ.) της ποικιλίας A και Y η απόδοση (σε Kgr/στρ.) της ποικιλίας B . Για κάθε ζεύγος παρατηρήσεων (x_i, y_i) , $i = 1, 2, \dots, 10$, παίρνουμε τη διαφορά $d_i = x_i - y_i$.

Αγρός (i)	1	2	3	4	5	6	7	8	9	10
Απόδοση ποικιλίας A (Kgr/στρ.) (x_i)	500	650	490	570	555	545	535	635	625	540
Απόδοση ποικιλίας B (Kgr/στρ.) (y_i)	455	620	455	610	505	495	515	600	600	510
Διαφορά (Kgr/στρ.) ($d_i = x_i - y_i$)	45	30	35	-40	50	50	20	35	25	30

Αν μ_A η μέση τιμή της X και μ_B η μέση τιμή της Y , θα κάνουμε τον έλεγχο της μηδενικής υπόθεσης

$$H_0 : \mu_A - \mu_B = 0$$

έναντι της εναλλακτικής,

$$H_1 : \mu_A - \mu_B \neq 0$$

ή, όπως εξηγήσαμε προηγουμένως, τον έλεγχο της μηδενικής υπόθεσης,

$$H_0 : \mu_D = 0$$

έναντι της εναλλακτικής,

$$H_1 : \mu_D \neq 0$$

όπου, $\mu_D = \mu_A - \mu_B$ η μέση τιμή της $D = X - Y$.

Ο έλεγχος θα γίνει με βάση το δείγμα των διαφορών $d_i = x_i - y_i$, $i = 1, 2, \dots, 10$ που δημιουργήσαμε. Έτσι, με την υπόθεση ότι αυτό προέρχεται από κανονικό πληθυσμό (δηλαδή ότι $D = X - Y \sim N(\mu_D, \sigma_D^2)$) η περιοχή απόρριψης του έλεγχου σε επίπεδο σημαντικότητας $\alpha = 5\%$ είναι,

$$t = \frac{|\bar{d} - 0| \cdot \sqrt{10}}{s_d} \geq t_{9,0.025} \quad \text{ή} \quad t = \frac{|\bar{d}| \cdot \sqrt{10}}{s_d} \geq 2.262.$$

Επειδή για το συγκεκριμένο δείγμα των διαφορών που πήραμε έχουμε,

$$\bar{d} = \frac{\sum_{i=1}^{10} d_i}{10} = \frac{45 + 30 + 35 - 40 + 50 + \dots + 30}{10} = \frac{280}{10} = 28$$

και

$$s_D^2 = \frac{\sum_{i=1}^{10} (d_i - 28)^2}{9} = \frac{(45 - 28)^2 + (30 - 28)^2 + \dots + (30 - 28)^2}{9} = 673.33,$$

η τιμή της στατιστικής συνάρτησης ελέγχου στο συγκεκριμένο δείγμα είναι

$$t = \frac{|28| \cdot \sqrt{10}}{\sqrt{673.33}} = 3.41$$

και επειδή αυτή βρίσκεται στην περιοχή απόρριψης, η μηδενική υπόθεση, σε επίπεδο σημαντικότητας $\alpha = 5\%$, απορρίπτεται υπέρ της εναλλακτικής. Δηλαδή, τα πειραματικά δεδομένα υποστηρίζουν ότι στον κάμπο της Θεσσαλίας, η μέση απόδοση της ποικιλίας A διαφέρει από τη μέση απόδοση της ποικιλίας B. Η πιθανότητα αυτό το συμπέρασμα να είναι λάθος είναι το πολύ 5%.

Παράδειγμα-5: Ένας ερευνητής θέλει να ελέγξει αν η ασπιρίνη επηρεάζει-μεταβάλλει την τιμή ενός αιματολογικού δείκτη (prothrombin time) ο οποίος σχετίζεται με την πηκτικότητα του αίματος και τη δημιουργία θρόμβων (η τιμή του δείκτη μετράται σε

δευτερόλεπτα). Για το σκοπό αυτό, επέλεξε ένα τυχαίο δείγμα 12 ατόμων και μέτρησε, για κάθε άτομο, την τιμή του δείκτη πριν και τρεις ώρες μετά τη λήψη δύο δισκίων ασπιρίνης (650mg). Οι σχετικές μετρήσεις φαίνονται στον πίνακα που ακολουθεί.

Άτομο (i)	1	2	3	4	5	6	7	8	9	10	11	12
Prothrombin time (sec) Πριν (x_i)	12.3	12	12	13	13	12.5	11.3	11.8	11.5	11	11	11.3
Prothrombin time (sec) Μετά (y_i)	12	12.3	12.5	12	13	12.5	10.3	11.3	11.5	11.5	11	11.5

Σε επίπεδο σημαντικότητας $\alpha = 5\%$, υποστηρίζουν τα πειραματικά δεδομένα ότι η μέση τιμή του δείκτη πριν και μετά τη λήψη των δισκίων ασπιρίνης διαφέρουν;

Έστω X και Y η τιμή του δείκτη πριν και τρεις ώρες μετά τη λήψη των δισκίων ασπιρίνης, αντίστοιχα. Ο ερευνητής γνωρίζει ότι η πήξη του αίματος είναι αποτέλεσμα μιας πολύπλοκης διαδικασίας, γι' αυτό σχεδίασε το πείραμα έτσι ώστε η σύγκριση των τιμών του δείκτη να γίνει σε ίδιες πειραματικές συνθήκες. Έτσι, δημιούργησε 12 ζεύγη παρατηρήσεων (x_i, y_i) , $i = 1, 2, \dots, 12$, όπου σε κάθε ζεύγος (x_i, y_i) η παρατήρηση x_i είναι η τιμή του δείκτη στο άτομο i πριν τη λήψη των δισκίων ασπιρίνης και y_i είναι η τιμή του δείκτη στο ίδιο άτομο (i) τρεις ώρες μετά τη λήψη των δισκίων ασπιρίνης. Πρόκειται, επομένως, για σύγκριση δύο μέσων με ζευγαρωτές παρατηρήσεις. Έτσι, υπολογίζουμε τις δειγματικές διαφορές $d_i = x_i - y_i$, $i = 1, 2, \dots, 12$ και εργαζόμαστε όπως στο προηγούμενο παράδειγμα.

Άτομο (i)	1	2	3	4	5	6	7	8	9	10	11	12
Prothrombin time (sec) Πριν (x_i)	12.3	12	12	13	13	12.5	11.3	11.8	11.5	11	11	11.3
Prothrombin time (sec) Μετά (y_i)	12	12.3	12.5	12	13	12.5	10.3	11.3	11.5	11.5	11	11.5
Διαφορά (sec.) ($d_i = x_i - y_i$)	0.3	-0.3	-0.5	1	0	0	1	0.5	0	-0.5	0	-0.2

Αν μ_1 είναι η μέση τιμή της X , μ_2 η μέση τιμή της Y και $\mu_D = \mu_1 - \mu_2$ η μέση τιμή της $D = X - Y$, θα κάνουμε τον έλεγχο της μηδενικής υπόθεσης,

$$H_0 : \mu_D = 0$$

έναντι της εναλλακτικής,

$$H_1 : \mu_D \neq 0.$$

Ο έλεγχος θα γίνει με βάση το δείγμα των διαφορών $d_i = x_i - y_i$, $i = 1, 2, \dots, 12$. Έτσι, με την υπόθεση ότι αυτό προέρχεται από κανονικό πληθυσμό, η περιοχή απόρριψης του έλεγχου σε επίπεδο σημαντικότητας $\alpha = 5\%$ είναι,

$$t = \frac{(\bar{d} - 0) \cdot \sqrt{12}}{s_d} \geq t_{11;0.025} \quad \text{ή} \quad t = \frac{|\bar{d}| \cdot \sqrt{12}}{s_d} \geq 2.201.$$

Επειδή για το συγκεκριμένο δείγμα των διαφορών έχουμε,

$$\bar{d} = \frac{\sum_{i=1}^{12} d_i}{12} = \frac{0.3 - 0.3 - 0.5 + 1 + \dots + 0 - 0.2}{12} = \frac{1.3}{12} = 0.108$$

και

$$s_D^2 = \frac{\sum_{i=1}^{12} (d_i - 28)^2}{11} = \frac{(0.3 - 0.108)^2 + (-0.3 - 0.108)^2 + \dots + (-0.2 - 0.108)^2}{9} = 0.257$$

η τιμή της στατιστικής συνάρτησης ελέγχου στο συγκεκριμένο δείγμα είναι

$$t = \frac{|0.108| \cdot \sqrt{12}}{\sqrt{0.257}} = 0.74$$

και επειδή αυτή δε βρίσκεται στην περιοχή απόρριψης, η μηδενική υπόθεση, σε επίπεδο σημαντικότητας $\alpha = 5\%$, δεν απορρίπτεται υπέρ της εναλλακτικής. Δηλαδή, τα πειραματικά δεδομένα, σε επίπεδο σημαντικότητας $\alpha = 5\%$, δεν υποστηρίζουν ότι η μέση τιμή του δείκτη πριν τη λήψη των δισκίων ασπιρίνης διαφέρει από τη μέση τιμή του δείκτη τρεις ώρες μετά τη λήψη των δισκίων ασπιρίνης.

Στα προηγούμενα παρουσιάσαμε πώς γίνεται η σύγκριση των μέσων τιμών δύο πληθυσμών με ανεξάρτητα δείγματα και πώς γίνεται με ζευγαρωτές παρατηρήσεις. Θα ολοκληρώσουμε αυτή την ενότητα με δύο εύλογα ερωτήματα:

- αν ένας ερευνητής σχεδιάζει να συγκρίνει τις μέσες τιμές δύο πληθυσμών, τι είναι καλύτερο να χρησιμοποιήσει, δύο ανεξάρτητα δείγματα ή ζευγαρωτές παρατηρήσεις;
- αν ο ερευνητής αποφασίσει να συγκρίνει τους μέσους των πληθυσμών με ζευγαρωτές παρατηρήσεις, με ποιο κριτήριο πρέπει να δημιουργηθούν τα ζεύγη.

Στη συνέχεια θα συζητήσουμε το δεύτερο ερώτημα και στο τέλος της ενότητας το πρώτο.

Κριτήρια για τη δημιουργία ζευγών παρατηρήσεων

Με βάση όσα ήδη έχουμε αναφέρει, η απάντηση είναι προφανής. Το ζευγάρι (συνταίριασμα, αντιστοίχιση) των υποκειμένων των οποίων ο ερευνητής ενδιαφέρεται να μετρήσει κάποιο χαρακτηριστικό τους (ή των πειραματικών μονάδων επί των οποίων ενδιαφέρεται να κάνει κάποια επέμβαση για να μετρήσει επίσης κάποιο χαρακτηριστικό τους), πρέπει να γίνει με βάση χαρακτηριστικά ή πειραματικές συνθήκες που επηρεάζουν την τιμή του χαρακτηριστικού που ενδιαφέρεται να μελετήσει. Επιπλέον, αυτά τα χαρακτηριστικά ή οι πειραματικές συνθήκες πρέπει να είναι (κατά το δυνατόν) όμοια/όμοιες ώστε η επίδρασή τους εντός του ίδου ζεύγους να είναι ίδια. Αν αυτό δε συμβαίνει, δηλαδή αν τα ζεύγη δημιουργηθούν με βάση χαρακτηριστικά (ή συνθήκες) τα οποία εντός του ίδου ζεύγους δεν έχουν την ίδια επίδραση στην τιμή του χαρακτηριστικού που μελετάμε, τότε, όπως θα εξηγήσουμε στη συνέχεια, σε σχέση με τη σύγκριση με ανεξάρτητα δείγματα, σπαταλάμε βαθμούς ελευθερίας χωρίς αντίστοιχο όφελος!

Ας δούμε κάποιες χαρακτηριστικές περιπτώσεις δημιουργίας ζευγαρωτών παρατηρήσεων:

1. Τα ζεύγη δημιουργούνται από δύο παρατηρήσεις που γίνονται στην ίδια πειραματική μονάδα πριν και μετά από κάποια επέμβαση. Είναι ίσως ο πιο «δημοφιλής» τρόπος καθορισμού ζευγαρωτών παρατηρήσεων γνωστός ως «**πριν και μετά**» («**before and after**»). Χρησιμοποιείται για τη σύγκριση δύο θεραπευτικών αγωγών ή δύο διαιτολογιών και σε άλλα ανάλογα προβλήματα. Για παράδειγμα, συγκρίνουμε τα επίπεδα αιμοσφαιρίνης πριν και μετά από ένα πρόγραμμα σωματικών ασκήσεων. Στο παράδειγμα-5 προηγουμένως, συγκρίναμε «πριν και μετά» ζευγαρωτές παρατηρήσεις.

2. Τα ζεύγη δημιουργούνται από δύο παρατηρήσεις που γίνονται στο ίδιο υποκείμενο όχι όμως κατ' ανάγκη σε χρονική απόσταση (πριν και μετά). Για παράδειγμα, συγκρίνουμε την πίεση του αίματος στον αριστερό και το δεξιό βραχίονα, το μήκος του εμπρός αριστερού και του πίσω αριστερού ποδιού των ελαφιών, το ύψος των δένδρων μιας δασικής έκτασης όταν αυτό μετράται με τη μέθοδο *A* και όταν μετράται με τη μέθοδο *B*, τη συγκέντρωση οξειδίου του αζώτου και τη συγκέντρωση υδρογονανθράκων στον ίδιο σταθμό, την υγρασία στο πάνω μέρος και στο μέσο του ίδιου σταχυού, κ.ά.
3. Για τη σύγκριση δύο μεθόδων ανάλυσης ή γενικότερα δύο διαφορετικών επεμβάσεων/μεταχειρίσεων, ένα υλικό χωρίζεται σε δύο ίσα μέρη και στο ένα εφαρμόζεται η μια μέθοδος και στο άλλο η άλλη. Για παράδειγμα, κομμάτι συνθετικού νήματος κόβεται στη μέση και στο ένα κομμάτι γίνεται ειδική χημική επεξεργασία για την αύξηση της αντοχής του και στο άλλο όχι.
4. Όταν η φύση του προβλήματος δεν επιτρέπει να πάρουμε τις δύο παρατηρήσεις από το ίδιο υποκείμενο, τα ζεύγη δημιουργούνται από δύο παρατηρήσεις που γίνονται σε υποκείμενα ή πειραματικές συνθήκες που επιδιώκεται να είναι όσο το δυνατόν όμοια/όμοιες. Για παράδειγμα, οι παρατηρήσεις που πήραμε από τα αγροτεμάχια στο παράδειγμα-4. Επίσης, αν θέλουμε να συγκρίνουμε δύο αντιδιαβρωτικά επιστρώματα σωλήνων, δημιουργούμε ζεύγη σωλήνων όπου ο ένας σωλήνας κάθε ζεύγους επιστρώνεται με το ένα επίστρωμα και ο άλλος με το άλλο. Οι σωλήνες τοποθετούνται στο έδαφος ανά ζεύγη, στο ίδιο βάθος και για ίδιο χρονικό διάστημα ώστε ανά ζεύγος να έχουμε όμοιες κατά το δυνατό συνθήκες (που επηρεάζουν τη διάβρωση). Μπορούμε επίσης, να χρησιμοποιήσουμε ζεύγη διδύμων εφόσον φυσικά είναι εφικτό, ή ζεύγη συζύγων (όχι γενικός αλλά σε κατάλληλα προβλήματα).

Πριν απαντήσουμε και στο δεύτερο ερώτημα ας ξαναδούμε το πρόβλημα του φοιτητή.

Συνέχεια του Παραδείγματος-3: Όπως εξηγήσαμε στα προηγούμενα, ο φοιτητής έπρεπε να έχει κάνει τη σύγκριση της μέσης μείωσης του βάρους, που επιτυγχάνεται με τα διαιτολόγια *A* και *B*, με ζευγαρωτές παρατηρήσεις. Αυτό σημαίνει ότι πρέπει να σχεδιάσει και να εκτελέσει το πείραμα από την αρχή. Δηλαδή, να δημιουργήσει ζεύγη όπου, σε κάθε ζεύγος, τα δύο άτομα να έχουν τα χαρακτηριστικά *BMI*, ηλικία και φύλο που επηρεάζουν τη μείωση του βάρους όσο το δυνατόν όμοια. Στη συνέχεια, και αφού έχει καθορίσει τα ζεύγη, στο ένα άτομο κάθε ζεύγους να δώσει το διαιτολόγιο *A* και στο άλλο άτομο του ίδιου ζεύγους το διαιτολόγιο *B*. Όμως ο φοιτητής στάθηκε τυχερός. Δε θα χρειασθεί να ξανακάνει το πείραμα. Τα πέντε άτομα στα οποία έδωσε το διαιτολόγιο *A* και τα πέντε άτομα στα οποία έδωσε το διαιτολόγιο *B*, από ευτυχή συγκυρία, δημιουργούν με βάση το κριτήριο που αναφέραμε προηγουμένως (όμοια ή σχεδόν όμοια τα χαρακτηριστικά *BMI*, ηλικία και φύλο που επιδρούν στη μείωση του βάρους) τα παρακάτω πέντε ζεύγη.

Ζεύγος	Χαρακτηριστικά
(2, 8)	Γυναίκες, μέσης ηλικίας, φυσιολογικό βάρος
(3, 6)	Άνδρες, νέοι, υπέρβαροι
(7, 10)	Γυναίκες, νέες, υπέρβαρες
(9, 1)	Άνδρες, ηλικιωμένοι, υπέρβαροι
(4, 5)	Γυναίκες, ηλικιωμένες, υπέρβαρες

Έτσι, έχει στη διάθεσή του τα παρακάτω ζεύγη παρατηρήσεων:

Ζεύγος (i)	1	2	3	4	5
	(2, 8)	(3, 6)	(7, 10)	(9, 1)	(4, 5)
Μείωση του βάρους με το διαιτολόγιο A (Kgr) (x_i)	1.46	3.10	4.74	2.28	3.92
Μείωση του βάρους με το διαιτολόγιο B (Kgr) (y_i)	0.94	1.70	2.46	0.18	3.22

Προσοχή: Πριν συνεχίσουμε, πρέπει να τονίσουμε ότι, με ποιο σχέδιο θα κάνουμε τη σύγκριση (με ζευγαρωτές παρατηρήσεις ή με ανεξάρτητα δείγματα), το αποφασίζουμε **πριν** πάρουμε τις παρατηρήσεις και όχι αφού τις πάρουμε. Δηλαδή, αν πάρουμε ζευγαρωτές παρατηρήσεις δε μπορούμε εκ των υστέρων να κάνουμε τον έλεγχο για ανεξάρτητα δείγματα (γιατί προφανώς τα δείγματα που πήραμε δεν είναι ανεξάρτητα). Επίσης, αν πάρουμε παρατηρήσεις από δύο ανεξάρτητα δείγματα (όπως ο φοιτητής) μάλλον δε θα μπορούμε εκ των υστέρων να τις επεξεργαστούμε ως ζευγαρωτές γιατί είναι πολύ πιθανό ότι δε θα μπορέσουμε να δημιουργήσουμε ζεύγη από τα δύο τυχαία δείγματα (ο φοιτητής, είπαμε, αποτελεί εξαίρεση ... στάθηκε τυχερός).

Ας κάνουμε τώρα τον έλεγχο t-κατά ζεύγη. Ο έλεγχος, κατά τα γνωστά, θα γίνει με βάση το δείγμα των διαφορών $d_i = x_i - y_i$, $i = 1, 2, \dots, 5$.

Ζεύγος (i)	1	2	3	4	5
	(2, 8)	(3, 6)	(7, 10)	(9, 1)	(4, 5)
Μείωση του βάρους με το διαιτολόγιο A (Kgr) (x_i)	1.46	3.10	4.74	2.28	3.92
Μείωση του βάρους με το διαιτολόγιο B (Kgr) (y_i)	0.94	1.70	2.46	0.18	3.22
Διαφορά (Kgr) ($d_i = x_i - y_i$)	0.52	1.40	2.28	2.10	0.70

Αν μ_A η μέση τιμή της X και μ_B η μέση τιμή της Y , θα κάνουμε τον έλεγχο της μηδενικής υπόθεσης, $H_0 : \mu_D = 0$ έναντι της εναλλακτικής, $H_1 : \mu_D \neq 0$, όπου, $\mu_D = \mu_A - \mu_B$ η μέση τιμή της $D = X - Y$. Με την υπόθεση ότι το δείγμα των διαφορών προέρχεται από κανονικό πληθυσμό, η περιοχή απόρριψης του ελέγχου σε επίπεδο σημαντικότητας $\alpha = 5\%$ είναι,

$$t = \frac{|\bar{d} - 0| \cdot \sqrt{5}}{s_d} \geq t_{4;0.025} \quad \text{ή} \quad t = \frac{|\bar{d}| \cdot \sqrt{5}}{s_d} \geq 2.776.$$

Επειδή για το συγκεκριμένο δείγμα των διαφορών που πήραμε έχουμε,

$$\bar{d} = \frac{\sum_{i=1}^5 d_i}{5} = \frac{0.52 + 1.4 + 2.28 + 2.1 + 0.7}{5} = 1.4$$

και

$$s_D^2 = \frac{\sum_{i=1}^5 (d_i - 1.4)^2}{4} = \frac{(0.52 - 1.4)^2 + \dots + (0.7 - 1.4)^2}{4} = 0.632,$$

η τιμή της στατιστικής συνάρτησης ελέγχου στο συγκεκριμένο δείγμα είναι

$$t = \frac{|1.4| \cdot \sqrt{5}}{\sqrt{0.632}} = 3.937$$

και επειδή αυτή βρίσκεται στην περιοχή απόρριψης, η μηδενική υπόθεση, σε επίπεδο σημαντικότητας $\alpha = 5\%$, απορρίπτεται υπέρ της εναλλακτικής. Δηλαδή, τα πειραματικά δεδομένα υποστηρίζουν ότι τα δύο διαιτολόγια δεν έχουν την ίδια αποτελεσματικότητα. Η πιθανότητα αυτό το συμπέρασμα να είναι λάθος είναι το πολύ 5%.

Αν κάνουμε τον έλεγχο έναντι της εναλλακτικής υπόθεσης, $H_1 : \mu_D > 0$, προφανώς (γιατί;) η μηδενική υπόθεση, σε επίπεδο σημαντικότητας $\alpha = 5\%$, και πάλι θα απορριφθεί. Δηλαδή, τα πειραματικά δεδομένα υποστηρίζουν ότι το διαιτολόγιο Α είναι πιο αποτελεσματικό από το διαιτολόγιο Β. Η πιθανότητα αυτό το συμπέρασμα να είναι λάθος είναι το πολύ 5%.

Παρατηρείστε ότι αν και οι βαθμοί ελευθερίας στον έλεγχο t κατά ζεύγη είναι 4, δηλαδή, οι μισοί από τους βαθμούς ελευθερίας που έχουμε στον έλεγχο t για δύο ανεξάρτητα δείγματα ($n_A + n_B - 2 = 5 + 5 - 2 = 8$), εντούτοις η μηδενική υπόθεση απορρίφθηκε. Αυτό συνέβη γιατί η διασπορά $s_d^2 = 0.632$ του δείγματος των διαφορών είναι πολύ μικρότερη από την κοινή διασπορά, $s^2 = 1.57$, των δύο ανεξάρτητων δειγμάτων και έτσι, παρότι οι βαθμοί ελευθερίας στον έλεγχο t κατά ζεύγη είναι λιγότεροι, αυτό δεν εμπόδισε να αναδειχθεί στατιστικά σημαντική η διαφορά που παρατηρήθηκε στους δειγματικούς μέσους. Θυμίζουμε τον τύπο που δίνει την τιμή t στον έλεγχο t κατά ζεύγη,

$$t = \frac{\bar{d} \cdot \sqrt{n}}{s_d} = \frac{\bar{x} - \bar{y}}{s_d / \sqrt{n}}$$

Όσο μικρότερος είναι ο παρανομαστής s_d / \sqrt{n} τόσο αυξάνει η τιμή t και μικραίνει επομένως η πιθανότητα να «κρύψει» ο παρανομαστής την παρατηρούμενη διαφορά $\bar{d} = \bar{x} - \bar{y}$ (που βρίσκεται στον αριθμητή). Θυμίζουμε επίσης ότι, για το ίδιο επίπεδο σημαντικότητας α , όσο λιγότεροι είναι οι βαθμοί ελευθερίας σε ένα t -test, τόσο απομακρύνεται η κριτική τιμή από το μηδέν και επομένως τόσο αυξάνει η πιθανότητα σφάλματος τύπου II, δηλαδή να μην απορριφθεί η μηδενική υπόθεση ενώ είναι αληθής η εναλλακτική. Εντούτοις, στο παράδειγμά μας, ήταν τόσο μεγάλη η μείωση της μεταβλητότητας που η μείωση των βαθμών ελευθερίας «αντισταθμίστηκε».

Ας δούμε τώρα και το δεύτερο ερώτημα.

Ανεξάρτητα δείγματα ή ζευγαρωτές παρατηρήσεις:

Παρότι η απάντηση στο ερώτημα αυτό προκύπτει από όσα ήδη έχουμε αναφέρει στα προηγούμενα, αξίζει να το ξανασυζητήσουμε.

Σε προβλήματα όπως αυτά των παραδειγμάτων 3, 4 και 5, η σύγκριση κατά ζεύγη, παρότι γίνεται με τους μισούς βαθμούς ελευθερίας ($n-1$) σε σχέση με τη σύγκριση με ανεξάρτητα δείγματα ($n_A + n_B - 2 = n + n - 2 = 2 \cdot (n - 1)$) εντούτοις μπορεί να μειώσει την πιθανότητα σφάλματος τύπου II, ιδιαίτερα όταν η μεταβλητότητα στο δείγμα των διαφορών είναι αρκετά μικρότερη από τη μεταβλητότητα των δύο ανεξάρτητων δειγμάτων. Βέβαια, για να συμβεί αυτό, τα ζεύγη πρέπει να δημιουργούνται με βάση κάποια όμοια χαρακτηριστικά/συνθήκες που έχουν επίδραση στο χαρακτηριστικό που μελετάμε. Διαφορετικά, σπαταλάμε βαθμούς ελευθερίας χωρίς «αντισταθμιστικό» όφελος. Ας δούμε το εξής παράδειγμα: Ένας ερευνητής θέλει να συγκρίνει το μέσο βαθμό ικανοποίησης (σε μια κλίμακα από 0 έως 100) από μια επέμβαση στα μάτια

όταν αυτή γίνει με τη μέθοδο A και όταν γίνει με τη μέθοδο B. Ας υποθέσουμε ότι ο ερευνητής δημιουργεί ζευγαρωτές παρατηρήσεις χρησιμοποιώντας τα ζεύγη του παραδείγματος-3, δηλαδή τα ζεύγη,

Ζεύγος	Χαρακτηριστικά
(2, 8)	Γυναίκες, μέσης ηλικίας, φυσιολογικό βάρος
(3, 6)	Άνδρες, νέοι, υπέρβαροι
(7, 10)	Γυναίκες, νέες, υπέρβαρες
(9, 1)	Άνδρες, ηλικιωμένοι, υπέρβαροι
(4, 5)	Γυναίκες, ηλικιωμένες, υπέρβαρες

Στο πρώτο άτομο κάθε ζεύγους η επέμβαση γίνεται με τη μέθοδο A και στο δεύτερο άτομο κάθε ζεύγους με τη μέθοδο B. Σε κάθε ζεύγος, τα χαρακτηριστικά, φύλο, ηλικία και BMI είναι όμοια ή σχεδόν όμοια και με βάση αυτά δημιουργήθηκαν τα ζεύγη. Όμως, αυτά τα χαρακτηριστικά δε μας λένε κάτι για την επίδραση στον παράγοντα που μελετάμε (βαθμός ικανοποίησης από την επέμβαση) και επομένως η επίδραση της «ατομικότητας» στο αποτέλεσμα, όποια και αν είναι, δε μπορεί να θεωρηθεί ίδια στα δύο άτομα κάθε ζεύγους. Αυτό έχει την εξής σοβαρή συνέπεια: δημιουργώντας τις διαφορές, η επίδραση της «ατομικότητας», δεν απαλείφεται και επομένως η μεταβλητότητα δεν ελαττώνεται, δηλαδή, δεν έχουμε το πλεονέκτημα της σύγκρισης κατά ζεύγη που είναι ο περιορισμός της μεταβλητότητας, ενώ έχουμε τις αρνητικές συνέπειες στην ισχύ του ελέγχου από τη μείωση των βαθμών ελευθερίας κατά το ήμισυ.

Τέλος, σχετικά με την ερώτηση-3 που θέσαμε στο παράδειγμα-2, είναι πλέον φανερό ότι αν η σύγκριση της μέσης απόδοσης των δύο ποικιλιών είχε γίνει με ένα πειραματικό σχέδιο όπως αυτό του παραδείγματος-4, δηλαδή κατά ζεύγη, όπου η σύγκριση θα γινόταν σε ίδιες καλλιεργητικές συνθήκες, θα είχαμε περιορίσει πολύ τον κίνδυνο να γίνει τέτοιο λάθος. Γι' αυτό, σε προβλήματα όπως αυτά των παραδειγμάτων 3 και 4, η σύγκριση συνήθως γίνεται με τυχαιοποιημένα μπλοκ (randomized blocks)-ένα πολύ γνωστό πειραματικό σχέδιο-μερική περίπτωση του οποίου είναι η σύγκριση κατά ζεύγη.