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.

12.1 Βασικές έννοιες

Η γενική ιδέα της διαδικασίας *στατιστικού ελέγχου υποθέσεων* είναι η εξής: θέτουμε ως *μηδενική υπόθεση* (H_0) αυτή για την οποία αμφιβάλουμε, 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 (εμφανίσθηκε ένα δείγμα που σπάνια εμφανίζεται). Στην περίπτωση αυτή απορρίψαμε λανθασμένα την H_0. Αυτό το σφάλμα ονομάζεται ***σφάλμα τύπου Ι (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 ΙΙ (type II error)***. Το «ρίσκο» επομένως είναι διπλό, με πιθανότητα

- **λανθασμένης απόρριψης** της H_0,
  P(σφάλμα τύπου Ι) = P(απόρριψη της H_0 *| αληθής η* H_0*)* και
- **λανθασμένης μη απόρριψης** της H_0,
  P(σφάλμα τύπου ΙΙ) = P(μη απόρριψη της H_0 *| 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 Ι* και *σφάλματος τύπου ΙΙ*; 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 12.1.1: *Το όριο αντοχής ενός τύπου καλωδίων είναι τυχαία μεταβλητή Χ με μέση τιμή* $\mu = 1500\,Kg$ *και τυπική απόκλιση* $\sigma = 175\,Kg$*. 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

$$H_0 : \mu = 1500\,Kg$$

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*[1] *(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 Ι.*

Ως εναλλακτική θέτουμε την

$$H_1 : \mu > 1500\,Kg$$

δηλαδή, η H_1 δηλώνει ότι η βελτίωση των υλικών επηρεάζει και ειδικότερα αυξάνει το όριο αντοχής των καλωδίων.

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

Γενικά, η H_1 δηλώνει ότι στον πληθυσμό υπάρχει αλλαγή/διαφορά.

Ο έλεγχος που μόλις διατυπώσαμε, είναι ένας ***μονόπλευρος*** και ειδικότερα ***δεξιόπλευρος*** έλεγχος. Γενικότερα, οι έλεγχοι

$$\begin{array}{lcl} H_0 : \mu \le \mu_0 & & H_0 : \mu \ge \mu_0 \\ & \text{και} & \\ H_1 : \mu > \mu_0 & & H_1 : \mu < \mu_0 \end{array}$$

ονομάζονται ***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, 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 $1500Kg$ ($H_0 : \mu = 1500\,Kg$), 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 $X_1, X_2, ... X_\nu$ μεγέθους $\nu = 50$ μας έδωσε τις μετρήσεις $x_1, x_2, ..., x_{50}$ με $\bar{x} = 1550\,Kg$.

Η «εμπειρία», δηλαδή αυτό που παρατηρείται στο δείγμα, συμφωνεί άραγε με την υπόθεση

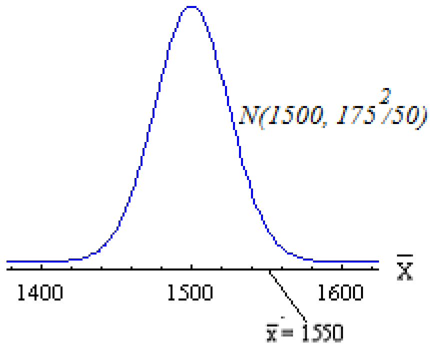
$$H_0 : \mu = 1500\,Kg$$

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* $T = T(X_1, X_2, \ldots, X_\nu)$, 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* τον *δειγματικό μέσο* $\bar{X}$ του οποίου η κατανομή, υπό την $H_0 : \mu = 1500\,Kg$, είναι γνωστή αφού το μέγεθος του δείγματος που πήραμε είναι αρκετά μεγάλο και επομένως από το *Κ.Ο.Θ.*, κατά προσέγγιση, έχουμε

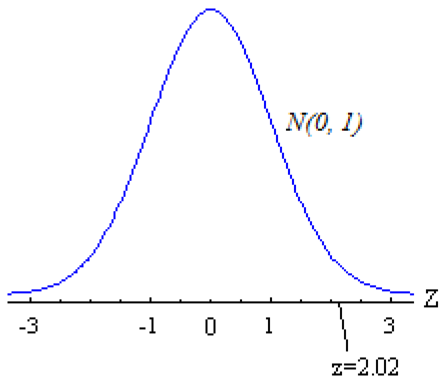
$$\bar{X} \sim N(1500, \frac{175^2}{50}) \text{ ή } \bar{X} \sim N(1500,\ 24.75^2).$$



Σχήμα 12.1.1
Η κατανομή του δειγματικού μέσου $\overline{X}$
όταν η $H_0 : \mu = 1500$ *είναι αληθής*

Εναλλακτικά, ως *στατιστική συνάρτηση* μπορούμε να επιλέξουμε την

$$Z = \frac{\overline{X} - 1500}{175/\sqrt{50}} = \frac{(\overline{X} - 1500)\sqrt{50}}{175} \sim N(0,1).$$



Σχήμα 12.1.2
Η κατανομή της στατιστικής συνάρτησης $(\overline{X} - 1500)\sqrt{50}/175$
όταν η $H_0 : \mu = 1500$ *είναι αληθής*

Έτσι, *«αυτό που παρατηρείται στο δείγμα»,* στο παράδειγμά μας εκφράζεται από τη *στατιστική συνάρτηση* $\overline{X}$ με τιμή, στο συγκεκριμένο δείγμα που πήραμε, $\bar{x} = 1550\,Kg$ *(Σχήμα 12.1.1)* ή ισοδύναμα, από την

$$Z = \frac{(\overline{X} - 1500)\sqrt{50}}{175}$$

με τιμή

$$z = \frac{(1550 - 1500)\sqrt{50}}{175} = 2.02 \quad \textit{(Σχήμα 12.1.2).}$$

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

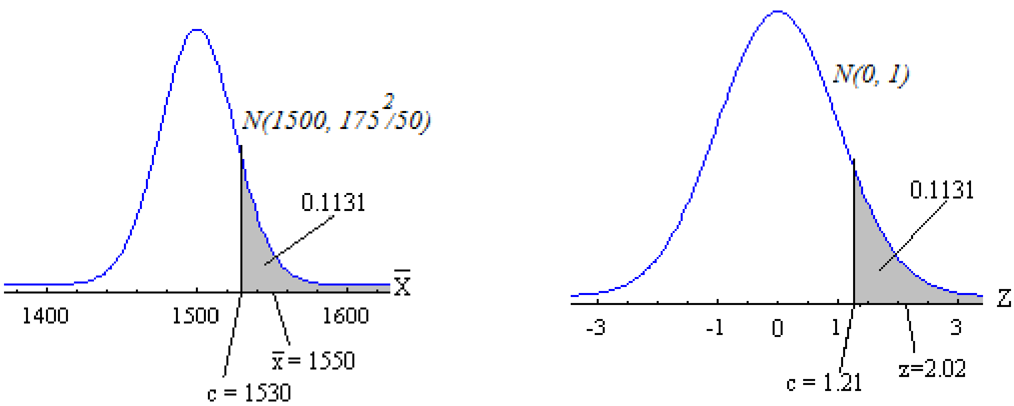
1ος τρόπος: Επιλέγουμε-(προ)καθορίζουμε το ανεκτό μέγεθος σφάλματος τύπου Ι

Αν η $H_0 : \mu = 1500\,Kg$ είναι αληθής, είναι λογικό να αναμένουμε ότι η τιμή της *στατιστικής συνάρτησης* $\overline{X}$ στο δείγμα που πήραμε (δηλαδή, ο *δειγματικός μέσος*) θα είναι κοντά στην τιμή 1500*Kg*. Αντιθέτως, αν η $H_0 : \mu = 1500\,Kg$ δεν είναι αληθής αναμένουμε ο *δειγματικός μέσος* 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 ο *δειγματικός μέσος* βρίσκεται μακριά από την $\mu = 1500\,Kg$, δηλαδή θα θεωρείται ακραίος. Έτσι, αν στο παράδειγμά μας επιλέξουμε $c = 1530\,Kg$ 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 $c = 1530\,Kg$ με την οποία οριοθετήσαμε τις ακραίες από της μη ακραίες τιμές του *δειγματικού μέσου*. Αν για παράδειγμα, επιλέξουμε $c = 1570\,Kg$ τότε $\bar{x} = 1550 < 1570$ δηλαδή τώρα το 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 $c = 1530\,Kg$ και αντίστοιχα στην περίπτωση που επιλέξουμε $c = 1570\,Kg$.

Για $c = 1530\,Kg$ έχουμε *(Σχήμα 12.1.3)*

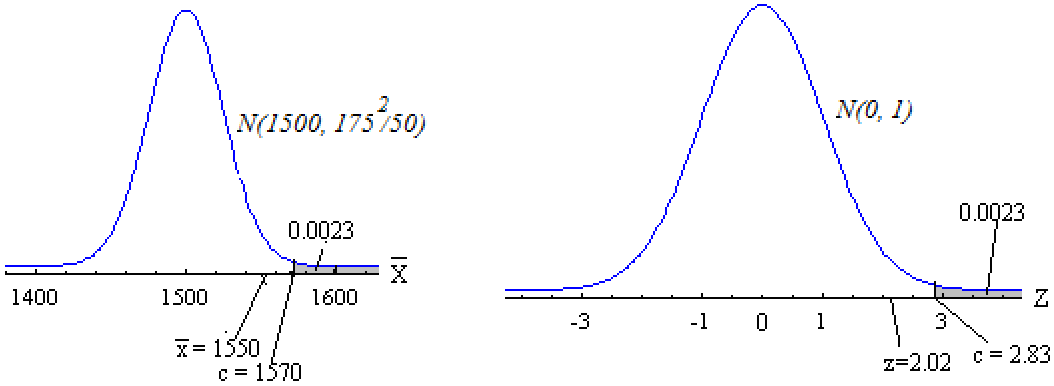
$$P(\text{σφάλμα τύπου I}) = P(\text{απόρριψη της } H_0 \mid \text{αληθής η } H_0) = P(\bar{X} \geq 1530 \mid \mu = 1500) =$$
$$= P\left(\frac{\bar{X} - 1500}{175/\sqrt{50}} \geq \frac{1530 - 1500}{175/\sqrt{50}}\right) = P(Z \geq 1.21) = 1 - \Phi(1.21) = 0.1131.$$



Σχήμα 12.1.3
Η πιθανότητα να κάνουμε σφάλμα τύπου I
αν ορίσουμε $c = 1530\,Kg$

Ομοίως, για $c = 1570\,Kg$ έχουμε *(Σχήμα 12.1.4)*

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



Σχήμα 12.1.4
Η πιθανότητα να κάνουμε σφάλμα τύπου Ι
αν ορίσουμε $c = 1570\,Kg$

Και για οποιοδήποτε *c* έχουμε

$$P(\overline{X} \geq c \mid \mu = 1500) = P(\frac{\overline{X} - 1500}{175/\sqrt{50}} \geq \frac{c-1500}{175/\sqrt{50}}) = P(Z \geq \frac{(c-1500)\sqrt{50}}{175}) =$$

$$= 1 - \Phi\left(\frac{(c-1500)\sqrt{50}}{175}\right).$$

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

Ορίζουμε ένα **μέγιστο αποδεκτό επίπεδο για την πιθανότητα να συμβεί *σφάλμα τύπου Ι*** 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* που επιλέξαμε (στο παράδειγμά μας της $\overline{X}$ ή της $Z = (\overline{X} - 1500)\sqrt{50}/175$), είναι «ακραία» ή όχι, και πλέον, **αποφασίζουμε για την απόρριψη ή τη μη απόρριψη της H_0, με κριτήριο ένα προκαθορισμένο επίπεδο της πιθανότητας να συμβεί *σφάλμα τύπου Ι.***

Το αποδεκτό/ανεκτό επίπεδο πιθανότητας σφάλματος τύπου Ι 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

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

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

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

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

Έτσι, επιλέγοντας $c = 1540.8$ *(Σχήμα 12.1.5α)* έχουμε $\bar{x} = 1550 > 1540.8$ και επομένως απορρίπτουμε την H_0 με πιθανότητα εσφαλμένης απόρριψης το πολύ 0.05.

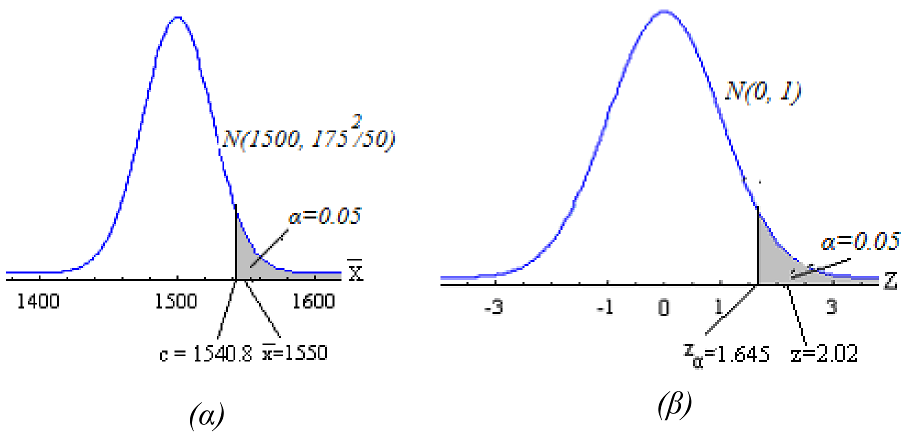
Ισοδύναμα, αν ως *στατιστική συνάρτηση* επιλέξουμε την

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

έχουμε

$$P(Z \geq c) \leq 0.05 \Rightarrow c = z_{0.05} = 1.645$$

δηλαδή, ως τιμή της c επιλέγουμε το 0.05-*άνω ποσοστιαίο σημείο* της τυποποιημένης κανονικής κατανομής $z_{0.05}$, και επειδή $z = 2.02 > z_{0.05} = 1.645$, απορρίπτουμε την H_0 με πιθανότητα εσφαλμένης απόρριψης το πολύ 0.05 *(Σχήμα 12.1.5β)*.



Σχήμα 12.1.5
Έλεγχος της $H_0: \mu = 1500\,Kg$ έναντι της $H_1: \mu > 1500\,Kg$
σε επίπεδο σημαντικότητας $\alpha = 0.05$

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 ο δειγματικός μέσος $\bar{x} = 1550$ δεν είναι μεγαλύτερος από αυτή την τιμή, σε επίπεδο σημαντικότητας $\alpha = 0.01$ δεν απορρίπτουμε την H_0 *(Σχήμα 12.1.6α)*.

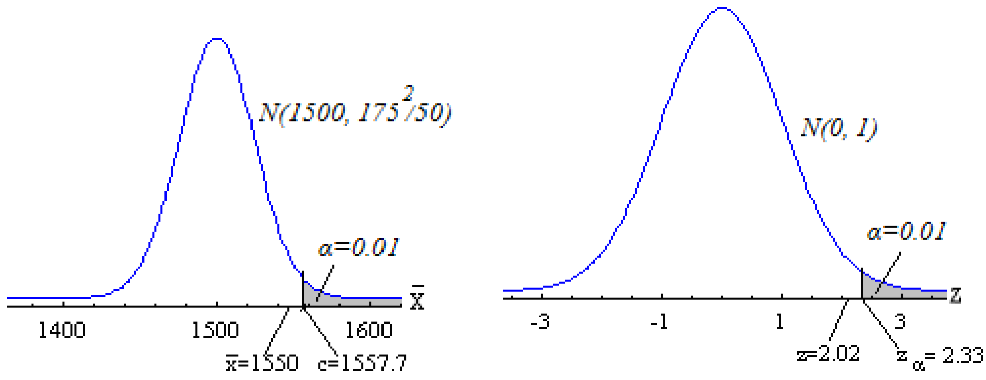
Ισοδύναμα, αν ως *στατιστική συνάρτηση* επιλέξουμε την

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

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 *(Σχήμα 12.1.6β)*.



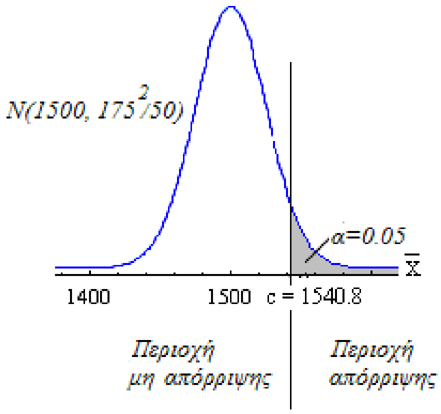
Σχήμα 12.1.6

Έλεγχος της $H_0 : \mu = 1500\, Kg$ *έναντι της* $H_1 : \mu > 1500\, Kg$

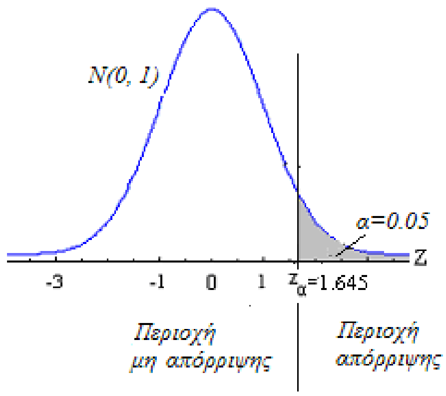
σε επίπεδο σημαντικότητας $\alpha = 0.01$

Η σταθερά c ονομάζεται ***κρίσιμη τιμή*** ή ***όριο απόρριψης (critical value, rejection limit)*** γιατί με βάση αυτή κρίνεται αν *«αυτό που παρατηρείται στο δείγμα»* είναι ακραίο ή όχι. Ανάλογα, η *στατιστική συνάρτηση* την οποία επιλέγουμε για να εκφράσει *«αυτό που παρατηρείται στο δείγμα»,* ονομάζεται ***στατιστική συνάρτηση ελέγχου (test statistic)*** και οι τιμές της για τις οποίες, σε προκαθορισμένο *επίπεδο σημαντικότητας* α, απορρίπτεται η H_0 ορίζουν την ***κρίσιμη περιοχή*** ή ***περιοχή απόρριψης (critical region, rejection region)***.

Δηλαδή, σε *επίπεδο σημαντικότητας* α, η *περιοχή απόρριψης* ορίζεται από τις τιμές της *στατιστικής συνάρτησης ελέγχου* οι οποίες, όταν είναι αληθής η H_0, *εμφανίζονται με πιθανότητα* α. Δείτε στο *Σχήμα 12.1.7* την *περιοχή απόρριψης* (και την *περιοχή μη απόρριψης*) του δεξιόπλευρου ελέγχου της $H_0 : \mu = 1500$ σε *επίπεδο σημαντικότητας* $\alpha = 0.05$.



(α)
Στατιστική συνάρτηση ελέγχου
ο δειγματικός μέσος
$\bar{X} \sim N(1500,\ 175^2/50)$



(β)
Στατιστική συνάρτηση ελέγχου η
$Z = (\bar{X} - 1500)\sqrt{50}/175 \sim N(0,1)$

Σχήμα 12.1.7
Περιοχή απόρριψης και περιοχή μη απόρριψης
δεξιόπλευρου ελέγχου σε $\alpha = 0.05$

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

Όταν η H_0 απορρίπτεται, το δείγμα χαρακτηρίζεται ***στατιστικά σημαντικό (statistically significant)*** και έχει την έννοια ότι δίνει σημαντικές αποδείξεις εναντίον της H_0 (αυτό που παρατηρούμε στο δείγμα διαφέρει σημαντικά από αυτό που αναμένουμε να εμφανισθεί αν θεωρήσουμε την H_0 αληθή).

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 το διάστημα

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

ή ισοδύναμα, το διάστημα

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

και τα ευρήματα στο δείγμα ($\bar{x} = 1550kg$ ή ισοδύναμα, $z = 2.02$), σε *επίπεδο σημαντικότητας* $\alpha = 0.05$, είναι *στατιστικά σημαντικά (Σχήμα 12.1.5).*

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

1

1^o Βήμα: Ορίσαμε τις δύο υποθέσεις:

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

2^o Βήμα: Ορίσαμε το *επίπεδο σημαντικότητας* α του ελέγχου:

$$\alpha = 0.05$$

3^o Βήμα: Επιλέξαμε τη *στατιστική συνάρτηση ελέγχου:*

$$\text{την } \bar{X}, \text{ ή ισοδύναμα, την } Z = \frac{(\bar{X} - 1500)\sqrt{50}}{175}$$

4^o Βήμα: Υπολογίσαμε την τιμή της *στατιστικής συνάρτησης ελέγχου:*

$$\bar{x} = 1550\,Kg, \text{ ή ισοδύναμα, } z = 2.02$$

5^o Βήμα: Για το συγκεκριμένο *επίπεδο σημαντικότητας* α που ορίσαμε, προσδιορίσαμε την *κρίσιμη περιοχή του ελέγχου (την περιοχή απόρριψης της* H_0*)*:

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

ή ισοδύναμα

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

6^o Βήμα: Εξετάσαμε αν η τιμή της *στατιστικής συνάρτησης ελέγχου* βρίσκεται ή όχι στην *κρίσιμη περιοχή του ελέγχου (την περιοχή απόρριψης της* H_0*)* και αποφασίσαμε (με πιθανότητα *σφάλματος τύπου Ι* το πολύ α) για την απόρριψη ή όχι της μηδενικής υπόθεσης:

Επειδή

$$\bar{x} = 1550 \in [1540.8, +\infty)$$

ή ισοδύναμα, επειδή

$$z = 2.02 \in [1.645, +\infty)$$

απορρίψαμε την $H_0 : \mu = 1500\,Kg,$ σε *επίπεδο σημαντικότητας* $\alpha = 0.05$.

Σχόλιο 12.1.1 (για το νόημα του επιπέδου σημαντικότητας): *α) Στη διατύπωση του συμπεράσματος του ελέγχου πρέπει προφανώς να αναφέρεται οπωσδήποτε το επίπεδο σημαντικότητας στο οποίο έγινε ο έλεγχος 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. β) 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 απορρίπτεται η* H_0 *και σε κάποιο μικρότερο, π.χ.* $\alpha = 0.01$, *να μην απορρίπτεται γιατί απαιτούνται σημαντικότερες αποδείξεις. Όσο πιο μικρό είναι το επίπεδο σημαντικότητας στο οποίο απορρίπτεται η* 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 μπορεί να απορριφθεί η* 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 12.1.1 (τι σημαίνει κάνω σφάλμα τύπου Ι): *Ας δούμε τι σημαίνει «κάνω σφάλμα τύπου Ι» και με μια άλλη διατύπωση. Έστω ότι κάνω τον έλεγχο σε επίπεδο σημαντικότητας* α *και ότι η μηδενική υπόθεση είναι αληθής. Τότε, από όλα τα δείγματα μεγέθους* ν *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).*

■

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. Δίνει (ευθεία) απάντηση στο ερώτημα ***«πόσο στατιστικά σημαντικό είναι το δείγμα;»***

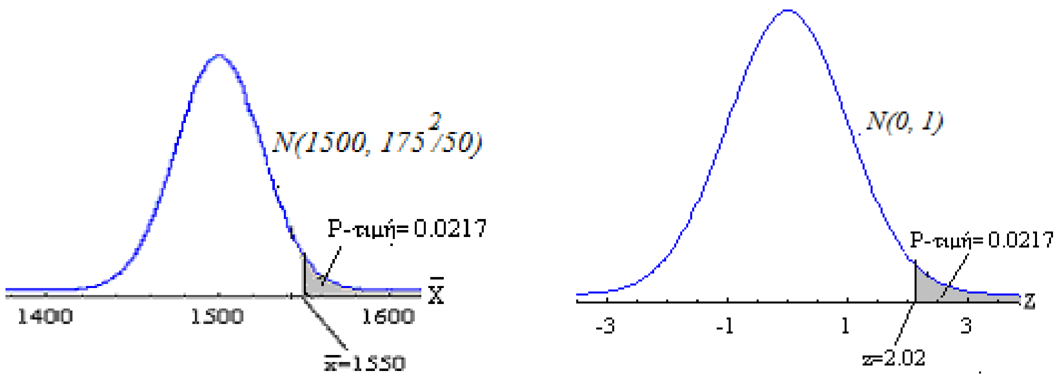
2^{ος} τρόπος: Υπολογίζουμε την P-Τιμή του ελέγχου

Με δεδομένο ότι η $H_0 : \mu = 1500\, Kg$ είναι αληθής, υπολογίζουμε την πιθανότητα να εμφανισθεί η τιμή $\bar{x} = 1550\, Kg$ που εμφανίσθηκε στο δείγμα ή κάποια μεγαλύτερή της (δηλαδή, προς την κατεύθυνση της H_1).

Ζητάμε την πιθανότητα $P(\bar{X} \geq 1550 \mid H_0)$ ή $P(\bar{X} \geq 1550 \mid \mu = 1500)$ και επειδή γνωρίζουμε την κατανομή της $\bar{X}$ έχουμε

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

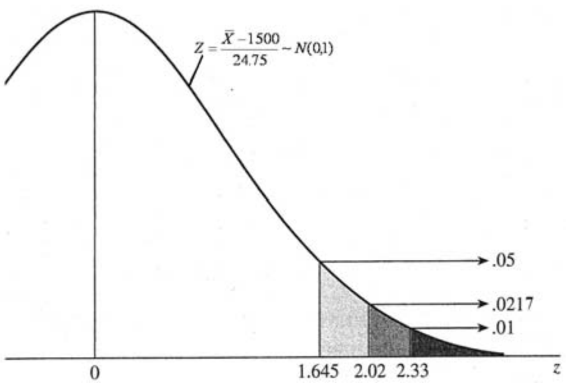
Αυτή η πιθανότητα ονομάζεται ***P-Τιμή (P-Value)*** ή ***κρίσιμο επίπεδο (critical level)*** του ελέγχου και είναι η πιθανότητα να εμφανισθεί η τιμή της *στατιστικής συνάρτησης ελέγχου* που εμφανίσθηκε (στο παράδειγμά μας $\bar{x} = 1550\, Kg$ *ή* $z = 2.02$) ή κάποια πιο μακριά (πιο ακραία) προς την κατεύθυνση της H_1, **δεδομένου ότι η *H$_o$* είναι αληθής.** Έτσι, υπολογίζοντας την *P-τιμή* του ελέγχου*,* 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 ελέγχου είναι ίση με 0.0217 ή 2.17% *(Σχήμα 12.1.8).*



Σχήμα 12.1.8

Η P-τιμή του ελέγχου της $H_0 : \mu = 1500\, Kg$ *έναντι της* $H_1 : \mu > 1500\, Kg$

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)[2]. Στο *Σχήμα 12.1.9* έχουμε μεγεθύνει τη δεξιά ουρά της κατανομής της *Z* και φαίνονται ευκρινώς οι περιοχές που αντιστοιχούν στο $\alpha = 0.05$, στην $P-\text{τιμή} = 0.0217$ και στο $\alpha = 0.01$.



Σχήμα 12.1.9
$\alpha = 0.05 > 0.0217$ *ενώ* $\alpha = 0.01 < 0.0217$

Έτσι, υπολογίζοντας την *P-Τιμή* του ελέγχου, μπορούμε άμεσα να την συγκρίνουμε με οποιοδήποτε α και αν επιλέξουμε και να αποφασίσουμε για την απόρριψη ή όχι της H_0. Βέβαια, ο κανόνας απόφασης διαμορφώνεται πλέον ως εξής:

- αν $\alpha \geq$ *P-Τιμή*, τότε σε επίπεδο σημαντικότητας α, η H_0 **απορρίπτεται**.
- αν $\alpha <$ *P-Τιμή*, τότε σε επίπεδο σημαντικότητας α, η H_0 **δεν απορρίπτεται**.

Συνοψίζοντας, από τα παραπάνω είναι προφανές ότι η *P-τιμή*

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

Σημείωση 12.1.2: *Στη βιβλιογραφία για την P-Τιμή χρησιμοποιείται και ο όρος* ***παρατηρούμενο επίπεδο σημαντικότητας (observed significance level)***. *Τον αναφέρουμε, όμως δεν τον συνιστούμε.*

■

Στα προηγούμενα προσπαθήσαμε 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 *διακύμανση* 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 *διακύμανση* σ^2 ενός πληθυσμού, ή αν το δείγμα είναι αρκετά μεγάλο ή όχι, ή αντί στη *μέση τιμή* ενός πληθυσμού αφορά

[2] *Θυμηθείτε ότι μικρότερο α σημαίνει ότι απαιτούνται πιο σημαντικές αποδείξεις εναντίον της* H_0.

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 $(-\infty, c_1] \cup [c_2, +\infty)$).

Στη συνέχεια 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.

12.2 Στατιστικός έλεγχος υποθέσεων για τη μέση τιμή ενός πληθυσμού

Θα αναφερθούμε στον έλεγχο της υπόθεσης

$$H_0 : \mu = \mu_0$$

δηλαδή, στον έλεγχο της υπόθεσης ότι η άγνωστη *μέση τιμή* μ ενός πληθυσμού έχει τιμή μ_0. Ειδικότερα, θα δώσουμε τη *στατιστική συνάρτηση ελέγχου* στις ακόλουθες περιπτώσεις που όπως είδαμε στα προηγούμενα *(στο Α΄ Μέρος και στο 10^o Κεφάλαιο),* γνωρίζουμε επακριβώς ή μπορούμε να προσεγγίσουμε την κατανομή του δειγματικού μέσου $\overline{X}$.

α) Όταν ο πληθυσμός είναι κανονικός.

β) Όταν το μέγεθος του δείγματος είναι μεγάλο (και ο πληθυσμός όχι κατ' ανάγκη κανονικός).

12.2.1 Ο πληθυσμός είναι κανονικός

Θα διακρίνουμε την περίπτωση που η *διακύμανση* του πληθυσμού είναι γνωστή από την περίπτωση που δεν είναι γνωστή.

(α) Ο πληθυσμός είναι κανονικός με γνωστή διακύμανση

Έστω *τυχαίο δείγμα* $X_1, X_2, \ldots, X_\nu$ από έναν κανονικό *πληθυσμό* με γνωστή *διακύμανση* σ^2 και *μέση τιμή* $\mu = \mu_0$ (ελεγχόμενη).

Επειδή, $X_i \sim N(\mu_0, \sigma^2)$, $i = 1,2,\ldots,\nu$, η κατανομή του δειγματικού μέσου

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_\nu}{\nu}$$

είναι, όπως είδαμε στο Α΄ Μέρος, *κανονική* (ανεξαρτήτως του μεγέθους του δείγματος) με

$$\overline{X} \sim N(\mu_0, \frac{\sigma^2}{\nu})$$

και επομένως

$$\frac{(\overline{X} - \mu_0)\sqrt{\nu}}{\sigma} = Z \sim N(0, 1).$$

Επειδή η *διακύμανση* σ^2 του πληθυσμού είναι γνωστή, στη συνάρτηση

$$Z = \frac{(\overline{X} - \mu_0)\sqrt{\nu}}{\sigma}$$

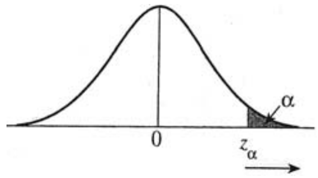
δεν υπάρχουν άγνωστοι παράμετροι και επομένως η τιμή της μπορεί να υπολογισθεί από το δείγμα. Έτσι, εργαζόμενοι όπως στο παράδειγμά μας, αν $\bar{x}$ η τιμή της $\overline{X}$ για συγκεκριμένη πραγματοποίηση του δείγματος, σε *επίπεδο σημαντικότητας* α, απορρίπτουμε την $H_0 : \mu = \mu_0$

- έναντι της $H_1: \mu > \mu_0$, όταν

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

  ή ισοδύναμα, όταν

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

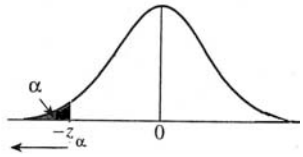


- έναντι της $H_1: \mu < \mu_0$, όταν

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

  ή ισοδύναμα, όταν

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

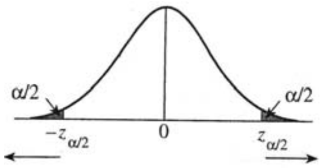


- έναντι της $H_1: \mu \neq \mu_0$, όταν

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

  ή ισοδύναμα, όταν

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



Η υπόθεση που κάναμε ότι η *διακύμανση* σ^2 του πληθυσμού είναι γνωστή, δεν είναι μια ιδιαίτερα ρεαλιστική υπόθεση. Στην πράξη, η *διακύμανση* του πληθυσμού συνήθως είναι άγνωστη. Ας δούμε πώς εργαζόμαστε στην περίπτωση αυτή.

(β) Ο πληθυσμός είναι κανονικός με άγνωστη διακύμανση

Έστω *τυχαίο δείγμα* $X_1, X_2, \ldots, X_\nu$ από έναν *πληθυσμό* που ακολουθεί *κανονική κατανομή* με άγνωστη *διακύμανση* σ^2 και *μέση τιμή* $\mu = \mu_0$ (ελεγχόμενη), δηλαδή,

$$X_i \sim N(\mu_0, \sigma^2), \;\; i = 1,2,\ldots,\nu\,.$$

Επειδή η *διακύμανση* του πληθυσμού είναι άγνωστη, δε μπορούμε ως *στατιστική συνάρτηση ελέγχου* να χρησιμοποιήσουμε την

$$Z = \frac{(\bar{X} - \mu_0)\sqrt{\nu}}{\sigma}$$

γιατί δε μπορούμε να υπολογίσουμε την τιμή της. Γι' αυτό, εκτιμάμε την άγνωστη *διακύμανση* σ^2 από την *(αμερόληπτη) δειγματική διακύμανση*

$$S^2 = \frac{1}{\nu - 1}\sum_{i=1}^{\nu}(X_i - \bar{X})^2$$

και ως *στατιστική συνάρτηση ελέγχου* χρησιμοποιούμε την

$$T = \frac{(\overline{X} - \mu_0)\sqrt{\nu}}{S}$$

η οποία είναι γνωστό *(10ο Κεφάλαιο)* ότι όταν $X_i \sim N(\mu_0, \sigma^2)$, $i = 1,2,\ldots,\nu$, και ανεξαρτήτως του μεγέθους του δείγματος, ακολουθεί *t-κατανομή* με $\nu - 1$ *βαθμούς ελευθερίας*. Δηλαδή

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

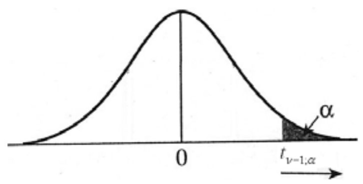
Είναι επομένως λογικό, οι *περιοχές απόρριψης* τώρα να ορίζονται με βάση το *άνω* α ή *το άνω* $\alpha/2$ *ποσοστιαίο σημείο* της κατανομής $t_{\nu-1}$ ($t_{\nu-1;\alpha}$ και $t_{\nu-1;\alpha/2}$ αντίστοιχα). Έτσι, σε *επίπεδο σημαντικότητας* α, απορρίπτουμε την $H_0 : \mu = \mu_0$

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

  $$\overline{x} \geq \mu_0 + \frac{s}{\sqrt{\nu}} t_{\nu-1;\alpha}$$

  ή ισοδύναμα, όταν

  $$t = \frac{(\overline{x} - \mu_0)\sqrt{\nu}}{s} \geq t_{\nu-1;\alpha}$$

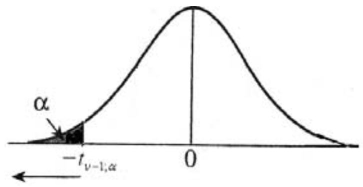


- έναντι της $H_1 : \mu < \mu_0$, όταν

  $$\overline{x} \leq \mu_0 - \frac{s}{\sqrt{\nu}} t_{\nu-1;\alpha}$$

  ή ισοδύναμα, όταν

  $$t = \frac{(\overline{x} - \mu_0)\sqrt{\nu}}{s} \leq -t_{\nu-1;\alpha}$$

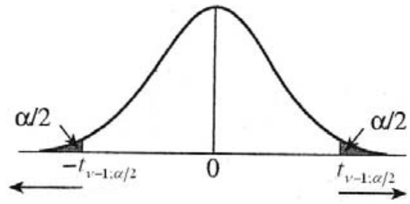


- έναντι της $H_1 : \mu \neq \mu_0$, όταν

  $$\overline{x} \leq \mu_0 - \frac{s}{\sqrt{\nu}} t_{\nu-1;\alpha/2} \quad \text{ή} \quad \overline{x} \geq \mu_0 + \frac{s}{\sqrt{\nu}} t_{\nu-1;\alpha/2}$$

  ή ισοδύναμα, όταν

  $$|t| = \frac{|\,\overline{x} - \mu_0\,|\sqrt{\nu}}{s} \geq t_{\nu-1;\alpha/2}$$



Σημείωση 12.2.1: *Όπως σημειώσαμε και στα προηγούμενα (10ο Κεφάλαιο) η κατανομή t είναι γνωστή και ως κατανομή Student (Student distribution). Επίσης, οι σχετικοί έλεγχοι στατιστικών υποθέσεων ονομάζονται t-tests. Σημειώνουμε επίσης, 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 και εφόσον βέβαια, η κατανομή του πληθυσμού δεν απέχει δραματικά*

από την κανονική (σοβαρή ασυμμετρία, πολυκόρυφη κτλ.) και το μέγεθος του δείγματος δεν είναι πολύ μικρό.

■

12.2.2 Το μέγεθος του δείγματος είναι μεγάλο

Θα διακρίνουμε και πάλι την περίπτωση που η *διακύμανση* του πληθυσμού είναι γνωστή από την περίπτωση που δεν είναι γνωστή.

(α) Το μέγεθος του δείγματος είναι μεγάλο και η διακύμανση του πληθυσμού είναι γνωστή

Έστω *τυχαίο δείγμα* $X_1, X_2, \ldots, X_\nu$ από οποιαδήποτε κατανομή (όχι κατ' ανάγκη κανονική), με γνωστή *διακύμανση* σ^2 και *μέση τιμή* $\mu = \mu_0$ (ελεγχόμενη). Από τη θεωρία πιθανοτήτων (*Κ.Ο.Θ.*) γνωρίζουμε ότι για μεγάλο μέγεθος δείγματος ν (εν γένει, $\nu \geq 30$), κατά προσέγγιση έχουμε

$$\overline{X} \sim N(\mu_0, \frac{\sigma^2}{\nu})$$

και επομένως

$$Z = \frac{(\overline{X} - \mu_0)\sqrt{\nu}}{\sigma} \sim N(0,1)\,.$$

Έτσι, στην περίπτωση αυτή, ισχύει ό,τι έχουμε αναφέρει στην *Παράγραφο 12.2.1.* Βέβαια, οι αντίστοιχοι έλεγχοι πλέον είναι **κατά προσέγγιση** επιπέδου σημαντικότητας α, αφού η κατανομή της συνάρτησης ελέγχου $\overline{X}$ ή $Z = (\overline{X} - \mu_0)\sqrt{\nu}/\sigma$ **δεν είναι** στην περίπτωση αυτή κανονική αλλά **προσεγγίζεται** από την κανονική. Ασφαλώς, όσο μεγαλύτερο είναι το μέγεθος του δείγματος, τόσο καλύτερη είναι η προσέγγιση.

(β) Το μέγεθος του δείγματος είναι μεγάλο και η διακύμανση του πληθυσμού είναι άγνωστη

Έστω *τυχαίο δείγμα* $X_1, X_2, \ldots, X_\nu$ από οποιαδήποτε κατανομή (όχι κατ' ανάγκη κανονική), με άγνωστη *διακύμανση* σ^2 και *μέση τιμή* $\mu = \mu_0$ (ελεγχόμενη).

Αν το μέγεθος του δείγματος ν είναι μεγάλο (εν γένει, $\nu \geq 30$), όπως είδαμε στο *10^ο^ Κεφάλαιο*, η *στατιστική συνάρτηση*

$$T = \frac{(\overline{X} - \mu_0)\sqrt{\nu}}{S}$$

προσεγγίζεται ικανοποιητικά από την $Z \sim N(0, 1)$. Δηλαδή,

$$T = \frac{(\overline{X} - \mu_0)\sqrt{\nu}}{S} \to Z \sim N(0,1)\,.$$

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

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

$$\overline{x} \geq \mu_0 + \frac{s}{\sqrt{\nu}} z_\alpha\,,$$

  ή ισοδύναμα, όταν

$$z = \frac{(\overline{x} - \mu_0)\sqrt{\nu}}{s} \geq z_\alpha$$

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

$$\bar{x} \le \mu_0 - \frac{s}{\sqrt{\nu}} z_\alpha ,$$

ή ισοδύναμα, όταν

$$z = \frac{(\bar{x} - \mu_0)\sqrt{\nu}}{s} \le -z_\alpha$$

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

$$\bar{x} \le \mu_0 - \frac{s}{\sqrt{\nu}} z_{\alpha/2} \text{ ή } \bar{x} \ge \mu_0 + \frac{s}{\sqrt{\nu}} z_{\alpha/2}$$

ή ισοδύναμα, όταν

$$|z| = \frac{|\bar{x} - \mu_0|\sqrt{\nu}}{s} \ge z_{\alpha/2} .$$

Επειδή στην περίπτωση αυτή η κατανομή της $(\bar{X} - \mu_0)\sqrt{\nu}/S$ **δεν είναι** η $N(0,1)$ 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.

■

Για διευκόλυνσή μας, ας συνοψίσουμε τις προηγούμενες περιπτώσεις σε έναν πίνακα.

	$H_1 : \mu \ne \mu_0$	$H_1 : \mu > \mu_0$	$H_1 : \mu < \mu_0$	**Προϋποθέσεις**
Περιοχή απόρριψης της $H_0 : \mu = \mu_0$	$\lvert Z\rvert = \frac{\lvert\bar{X} - \mu_0\rvert}{\sigma/\sqrt{\nu}} \ge z_{\alpha/2}$	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{\nu}} \ge z_\alpha$	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{\nu}} \le -z_\alpha$	*Η διακύμανση σ^2 είναι γνωστή και ο πληθυσμός είναι κανονικός ή Η διακύμανση σ^2 είναι γνωστή και το ν μεγάλο*
	$\lvert Z\rvert = \frac{\lvert\bar{X} - \mu_0\rvert}{S/\sqrt{\nu}} \ge z_{\alpha/2}$	$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{\nu}} \ge z_\alpha$	$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{\nu}} \le -z_\alpha$	*Η διακύμανση σ^2 είναι άγνωστη και το ν μεγάλο (οτιδήποτε πληθυσμός)*
	$\lvert T\rvert = \frac{\lvert\bar{X} - \mu_0\rvert}{S/\sqrt{\nu}} \ge t_{\nu-1,\alpha/2}$	$T = \frac{\bar{X} - \mu_0}{S/\sqrt{\nu}} \ge t_{\nu-1,\alpha}$	$T = \frac{\bar{X} - \mu_0}{S/\sqrt{\nu}} \le -t_{\nu-1,\alpha}$	*Η διακύμανση, σ^2 άγνωστη και ο πληθυσμός είναι κανονικός (οτιδήποτε ν)*
	?	?	?	*Το ν είναι μικρό, ο πληθυσμός όχι κανονικός και η διακύμανση γνωστή ή άγνωστη*

Πίνακας 12.2.1
Στατιστικοί έλεγχοι υποθέσεων για τη μέση τιμή μ ενός πληθυσμού με ένα τυχαίο δείγμα μεγέθους ν και σε επίπεδο σημαντικότητας α

Ερώτηση: *Αν ο πληθυσμός είναι κανονικός με άγνωστη διασπορά και το μέγεθος του δείγματος είναι μεγάλο, τότε προφανώς εφαρμόζεται ο έλεγχος της Παραγράφου 12.2.1β αλλά και της Παραγράφου 12.2.2β. Τι λέτε, τίθεται δίλημμα επιλογής ελέγχου[3];*

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

---

[3] *Θυμηθείτε ότι για μεγάλα ν ισχύει:* $t_{\nu;\alpha} \cong z_\alpha$.

Παράδειγμα 12.2.1: *Στη βιβλιογραφία αναφέρεται ότι η μέση ετήσια παραγωγή γάλακτος μιας συγκεκριμένης φυλής αγελάδων είναι 4000Kg (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 40 αγελάδες της συγκεκριμένης φυλής από μονάδες της Μακεδονίας και της Θράκης και κατέγραφε κάθε μέρα, επί ένα έτος, την παραγωγή γάλακτος κάθε μιας αγελάδας. Η μέση ετήσια παραγωγή των 40 αγελάδων, βρέθηκε 3910Kg με τυπική απόκλιση 250Kg.*

Απάντηση: 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 *η κατανομή των ετήσιων αποδόσεων γάλακτος όλων των αγελάδων της συγκεκριμένης φυλής που εκτρέφονται στη Μακεδονία και τη Θράκη*. Ας συμβολίσουμε με X την ετήσια παραγωγή γάλακτος σε Kg μιας οποιασδήποτε αγελάδας της συγκεκριμένης φυλής στη Μακεδονία και τη Θράκη και με $X_1, X_2, \ldots, X_{40}$ τις ετήσιες αποδόσεις 40 αγελάδων τυχαία επιλεγμένων. Στο συγκεκριμένο δείγμα που πήρε ο ερευνητής, οι τιμές του δείγματος, $x_1, x_2, \ldots, x_{40}$, έδωσαν $\bar{x} = 3910kg$ με $s = 250kg$.

Ως *μηδενική υπόθεση* θέτουμε αυτή που αμφισβητείται από τον ερευνητή (γι' αυτό την ελέγχει) δηλαδή την

$$H_0 : \mu = 4000\,Kg.$$

Ως *εναλλακτική* θέτουμε την

$$H_1 : \mu \neq 4000\,Kg$$

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

$$Z = \frac{(\bar{X} - \mu_0)\sqrt{\nu}}{S}$$

γιατί η *διακύμανση* του πληθυσμού είναι άγνωστη και το μέγεθος του δείγματος είναι $\nu = 40 > 30$ (περίπτωση της *Παραγράφου 12.2.2β*).

Επειδή ο έλεγχος είναι *αμφίπλευρος*, σε *επίπεδο σημαντικότητας* $\alpha = 0.05$, η *περιοχή απόρριψης* είναι

$$|z| \geq z_{0.05/2} \;\; ή \;\; |z| \geq z_{0.025} \;\; ή \;\; |z| \geq 1.96 \Leftrightarrow z \geq 1.96 \;\; ή \;\; z \leq -1.96 .$$

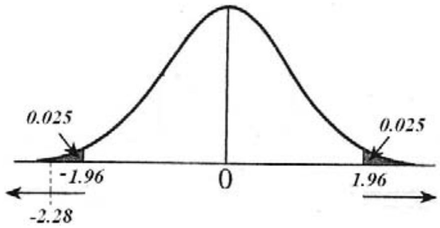
Υπολογίζουμε την τιμή της *στατιστικής συνάρτησης ελέγχου* στο δείγμα. Έχουμε

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

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 (Σχήμα 12.2.1)* και επομένως σε *επίπεδο σημαντικότητας* $\alpha = 0.05$ απορρίπτουμε τη *μηδενική υπόθεση*.



Σχήμα 12.2.1
Η τιμή $z = -2.28$ της συνάρτησης ελέγχου
βρίσκεται στην περιοχή απόρριψης της $H_0 : \mu = 4000\,Kg$

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, ή 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 0.05.*

Παρατήρηση 12.2.1: *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 $H_0 : \mu = 4000\,Kg$ έναντι της $H_1 : \mu < 4000\,Kg$. Τι λέτε, είναι απαραίτητο να κάνουμε αυτό τον έλεγχο ή μήπως μπορούμε να συμπεράνουμε το αποτέλεσμά του από το αποτέλεσμα του αμφίπλευρου ελέγχου που ήδη κάναμε;*

Παρατήρηση 12.2.2: *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 $|z| \geq 2.58$, δηλαδή, $z \geq 2.58$ ή $z \leq -2.58$. *Η τιμή $z = -2.28$ της 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 90Kg (μεταξύ δειγματικού μέσου $\bar{x} = 3910kg$ και μηδενικής υπόθεσης $\mu = 4000\,Kg$) 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. Άραγε, σε επίπεδο σημαντικότητας $\alpha = 0.02$ ή $\alpha = 0.03$ είναι 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 εμφανίσθηκε στο συγκεκριμένο τυχαίο δείγμα. Έχουμε*

$$P - \textit{τιμή} = P(|Z| \geq 2.28) = P(Z \geq 2.28) + P(Z \leq -2.28) = 0.0226.$$

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.

■

Παράδειγμα 12.2.2: *Από έναν κανονικό πληθυσμό πήραμε ένα τυχαίο δείγμα μεγέθους $\nu = 9$, με $\bar{x} = 60$ και $s = 12$. Ας κάνουμε, σε επίπεδο σημαντικότητας $\alpha = 0.05$, τον έλεγχο της μηδενικής υπόθεσης $H_0 : \mu = 65$ έναντι της εναλλακτικής $H_1 : \mu \neq 65$.*

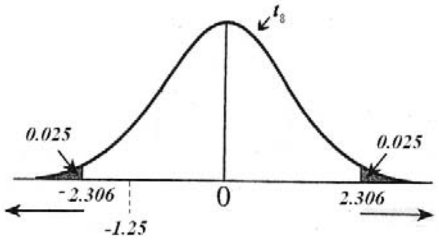
<u>Απάντηση:</u> Προφανώς, κατάλληλο είναι το *t-test* (περίπτωση της *Παραγράφου 12.2.1β*). Ο έλεγχος είναι αμφίπλευρος και επομένως, σε *επίπεδο σημαντικότητας* $\alpha = 0.05$, η *περιοχή απόρριψης* είναι

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

Επειδή

$$t = \frac{(\bar{x} - \mu_0)\sqrt{\nu}}{s} = \frac{(60-65)\sqrt{9}}{12} = -1.25$$

η τιμή της *στατιστικής συνάρτησης ελέγχου* δε βρίσκεται στην *περιοχή απόρριψης (Σχήμα 12.2.2)* και επομένως, σε *επίπεδο σημαντικότητας* $\alpha = 0.05$, δεν απορρίπτουμε τη *μηδενική υπόθεση*.



Σχήμα 12.2.2
Η τιμή $t = -1.25$ *της συνάρτησης ελέγχου*
δε βρίσκεται στην περιοχή απόρριψης της $H_0 : \mu = 65$

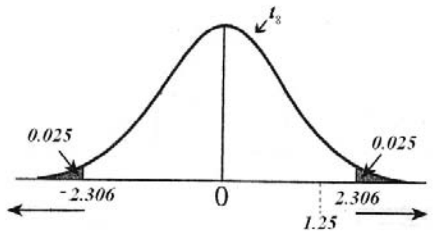
Παρατήρηση 12.2.3: *Μη απορρίπτοντας την* $H_0 : \mu = 65$, *αποδείξαμε άραγε ότι είναι αληθής; Δηλαδή, αποδεχόμαστε ότι η μέση τιμή* μ *του πληθυσμού είναι ίση με 65 και είμαστε βέβαιοι γι' αυτό; Η απάντηση είναι όχι! Δεν αποδείξαμε ότι* $\mu = 65$. *Απλώς αποτύχαμε να απορρίψουμε την* $H_0 : \mu = 65$. *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*

$$t = \frac{(\bar{x} - \mu_0)\sqrt{\nu}}{s} = \frac{(60-55)\sqrt{9}}{12} = 1.25 \,.$$

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 (Σχήμα 12.2.3).*



Σχήμα 12.2.3
Η τιμή $t = 1.25$ *της συνάρτησης ελέγχου*
δε βρίσκεται στην περιοχή απόρριψης της $H_0 : \mu = 55$

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 Ι ή σφάλμα τύπου ΙΙ, αντίστοιχα. Την πιθανότητα σφάλματος τύπου Ι, 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 ΙΙ. 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 στη συνέχεια, ο υπολογισμός επακριβώς της πιθανότητας σφάλματος τύπου ΙΙ, 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*

«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 $(1-\alpha)100\%$ *διάστημα εμπιστοσύνης. Στην περίπτωση που εξετάζουμε, ένα 95% διάστημα εμπιστοσύνης για την άγνωστη μέση τιμή* μ *του πληθυσμού είναι*

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

Έτσι, με βάση αυτό που παρατηρείται στο δείγμα, συμπεραίνουμε ότι η μηδενική υπόθεση $H_0 : \mu = 65$ *(όπως και η* $H_0 : \mu = 55$*) σε επίπεδο σημαντικότητας 5% δεν απορρίπτεται και το διάστημα* [50.776, 69.224], *με πιθανότητα 95%* περιέχει *την άγνωστη μέση τιμή* μ, *του πληθυσμού.*

Παρατηρείστε ότι τόσο η τιμή 55 όσο και η τιμή 65 βρίσκονται εντός του 95% διαστήματος εμπιστοσύνης. Τι λέτε, σχετίζεται το διάστημα εμπιστοσύνης με την περιοχή μη απόρριψης της μηδενικής υπόθεσης;

■

Σχόλιο 12.2.1 (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 αθώος[4] (δηλαδή, αυτή που αμφισβητείται) και ως εναλλακτική ότι είναι ένοχος. Η δικαστική διαδικασία σκοπό έχει να διαπιστώσει αν υπάρχουν σημαντικά αποδεικτικά στοιχεία εναντίον της*

[4] *Έτσι προβλέπεται από το δικαιακό 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...*

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.

■

Παράδειγμα 12.2.3: *Από έναν πληθυσμό με άγνωστη διακύμανση, πήραμε ένα τυχαίο δείγμα μεγέθους* $\nu = 36$. *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;*

<u>*Απάντηση:*</u> *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 ο πληθυσμός είναι κανονικός ούτε και τη *διακύμανση* του, επειδή το μέγεθος του δείγματος είναι μεγάλο, σε *επίπεδο σημαντικότητας* $\alpha = 0.05$, η *περιοχή απόρριψης* του ελέγχου είναι αυτή της *Παραγράφου 12.2.2β,*

$$|z| = \frac{|\bar{x} - \mu_0|\sqrt{\nu}}{s} \geq z_{0.025}$$

δηλαδή, $z \geq 1.96$ *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!

Ερώτηση: *Με βάση τη λογική της διαδικασίας ελέγχου, μπορείτε να εξηγήσετε γιατί τα αποτελέσματα των δύο ελέγχων που κάναμε στα (α) και (β) δεν είναι αντιφατικά*[5].

■

Παράδειγμα 12.2.4: *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*

---

[5] *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.*

των υδρόβιων οργανισμών. Το πρόβλημα αυτό είχε διαπιστωθεί, πριν από αρκετά χρόνια, και στον ποταμό Καλαμά. Για την αντιμετώπισή του εφαρμόσθηκε 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 αυτά τα δεδομένα, μπορεί ο φοιτητής να συμπεράνει ότι στον ποταμό Καλαμά η μέση ποσότητα διαλυμένου οξυγόνου είναι πλέον μεγαλύτερη από 5ppm;

Απάντηση: Ο φοιτητής μελετάει την ποσότητα *X* διαλυμένου οξυγόνου στα νερά του ποταμού Καλαμά με βάση ένα τυχαίο δείγμα $X_1, X_2, \ldots, X_{10}$, 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 $x_1, x_2, \ldots, x_{10}$ 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* $H_1 : \mu > 5ppm$ [6].

Η *κατανομή* του πληθυσμού[7] δε μας είναι γνωστή. Επίσης, η *διακύμανση* του δε μας είναι γνωστή και το μέγεθος του δείγματος είναι μικρό ($\nu = 10 < 30$). Αυτή η περίπτωση δεν εντάσσεται σε καμία από τις περιπτώσεις 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 *Παραγράφου 12.2.2β*. Όμως δεν είναι. Επίσης, αν γνωρίζαμε ότι ο πληθυσμός είναι κανονικός, θα εφαρμόζαμε το *t-test (Παράγραφος 12.2.1β)*. Τι κάνουμε επομένως;

Με βάση όσα μέχρι τώρα γνωρίζουμε, ένα δρόμο έχουμε.[8] 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 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 (Παράγραφος 12.2.1β)* 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*

---

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

[7] *Ο πληθυσμός του οποίου ελέγχουμε τη μέση τιμή είναι η κατανομή των τιμών διαλυμένου οξυγόνου στην κοίτη του ποταμού.*

[8] *Η Στατιστική προσφέρει και άλλες δυνατότητες. 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.*

$$T = \frac{(\overline{X} - \mu_0)\sqrt{\nu}}{S}$$

αφού προηγουμένως υπολογίσουμε την τιμή $\overline{x}$ της $\overline{X}$ και την τιμή s της S για τη συγκεκριμένη πραγματοποίηση του δείγματος.

Έτσι έχουμε, $\overline{x} = 5.1 ppm$ και $s = 0.115 ppm$ και επομένως

$$t = \frac{(\overline{x} - \mu_0)\sqrt{\nu}}{s} = \frac{(5.1 - 5)\sqrt{10}}{0.115} = 2.75 .$$

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

$$t \geq t_{9;0.05} \text{ ή } t \geq 1.833$$

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 5ppm*[9].

■

Παράδειγμα 12.2.5 (συνέχεια του Παραδείγματος 9.1.3): *Στον Πίνακα 12.2.2 φαίνεται για κάθε μια από 50 τυχαία επιλεγμένες γαλακτοπαραγωγές αγελάδες, ο χρόνος (σε μήνες) από την πρώτη εκδήλωση μιας συγκεκριμένης ασθένειας που προσβάλλει τις αγελάδες μέχρι την επανεμφάνισή της.*

2.1	1.7	0.8	0.8	4.1	8.7	1.4	2.9	1.9	2.7
4.4	2.2	5.5	7.0	1.8	0.2	1.0	0.9	4.0	0.7
2.0	6.5	0.7	4.3	0.2	5.6	2.4	1.4	1.3	1.2
0.5	3.9	7.4	3.3	8.8	0.3	2.0	5.7	0.8	2.6
9.9	1.6	2.8	1.0	0.6	1.3	0.8	5.9	0.9	0.4

Πίνακας 12.2.2
Οι χρόνοι επανεμφάνισης (σε μήνες) μιας ασθένειας
σε 50 γαλακτοπαραγωγές αγελάδες

Στο 9^o Κεφάλαιο είδαμε (Παραδείγματα 9.1.3&9.1.4 και Άσκηση 9.13) ότι ο μέσος και η τυπική απόκλιση του συγκεκριμένου δείγματος αντίστοιχα είναι $\overline{x} = 2.82$ *μήνες και* $s = 2.51$ *μήνες. Α πό τη σχετική βιβλιογραφία είναι γνωστό ότι ο μέσος χρόνος μέχρι την επανεμφάνιση της συγκεκριμένης ασθένειας είναι 3 μήνες. Μήπως τα συγκεκριμένα δεδομένα δίνουν σε επίπεδο σημαντικότητας 5%, στατιστικά σημαντικές αποδείξεις ότι ο μέσος χρόνος επανεμφάνισης της ασθένειας δεν είναι 3 μήνες αλλά λιγότερο;*

<u>*Απάντηση:*</u> Η κατανομή της τυχαίας μεταβλητής, έστω *X*, που εκφράζει το χρόνο επανεμφάνισης της ασθένειας, μας είναι άγνωστη. Δε γνωρίζουμε ούτε τη μορφή της ούτε κάποια παράμετρό της.

Ζητείται να κάνουμε σε επίπεδο σημαντικότητας 5%, κατάλληλο στατιστικό έλεγχο για να ελέγξουμε αν το συγκεκριμένο δείγμα δίνει στατιστικά σημαντικές αποδείξεις εναντίον της *μηδενικής υπόθεσης*

$$H_0 : \mu = 3 \text{ μήνες}$$

[9] *Το δείγμα που χρησιμοποιήσαμε είναι δυστυχώς... υποθετικό.*

και υπέρ της *εναλλακτικής*

$$H_1 : \mu < 3\,\textit{μήνες}.$$

Επειδή το δείγμα είναι μεγάλο *(*$\nu = 50 \geq 30$*)* και η διακύμανση σ^2 της X (του πληθυσμού) μας είναι άγνωστη, με βάση όσα προηγουμένως εξηγήσαμε *(περίπτωση 12.2.2β)*, η περιοχή απόρριψης του ελέγχου είναι

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{\nu}} \leq -z_{0.05} \text{ ή } z = \frac{\bar{x} - \mu_0}{s/\sqrt{\nu}} \leq -1.645\,.$$

Η τιμή της *στατιστικής συνάρτησης ελέγχου* είναι

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{\nu}} = \frac{2.82 - 3}{2.51/\sqrt{50}} = -0.51$$

και επειδή αυτή δε βρίσκεται στην *περιοχή απόρριψης* (δεν είναι μικρότερη της *κρίσιμης τιμής* -1.645), η μηδενική υπόθεση σε επίπεδο σημαντικότητας 5%, με βάση το συγκεκριμένο δείγμα, δεν απορρίπτεται.

Παρατηρείστε ότι η τιμή $\mu = 3$ περιέχεται στο *95% διάστημα εμπιστοσύνης* [2.12, 3.52] που κατασκευάσαμε στο *Παράδειγμα 11.2.2.*

■

12.3 Πιθανότητα σφάλματος τύπου ΙΙ και ισχύς 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 ΙΙ*. 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, δε γνωρίζουμε με ποια πιθανότητα αυτή η απόφασή μας μπορεί να είναι λάθος, αφού δεν υπολογίσαμε την πιθανότητα *σφάλματος τύπου ΙΙ*

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

Φροντίσαμε, δηλαδή, για την «προστασία» από *σφάλμα τύπου Ι* και δεν ασχοληθήκαμε με το *σφάλμα τύπου ΙΙ,* 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

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

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 εισαγωγικό παράδειγμά μας *(Παράδειγμα 12.1.1)*, στον έλεγχο της $H_0 : \mu = 1500\,Kg$ έναντι της $H_1 : \mu > 1500\,Kg$, 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 ΙΙ*, β.

Αν μεταξύ των δύο υποθέσεων αληθής είναι η $H_1 : \mu > 1500\,Kg$, 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*Kg*, τότε ζητάμε την πιθανότητα β, να μην απορρίψουμε την $H_0 : \mu = 1500\,Kg$ (ενώ θα έπρεπε, αφού αληθής είναι η H_1). Έχουμε

$$\beta = P(\textit{σφάλμα τύπου ΙΙ}) = P(\textit{μη απόρριψη της } H_0 \mid \textit{αληθής η } H_1 : \mu = \mu_1 > 1500) =$$

$$= P(\overline{X} < 1557.7 \mid \mu = \mu_1) = P(\frac{\overline{X} - \mu_1}{175/\sqrt{50}} < \frac{1557.7 - \mu_1}{175/\sqrt{50}}) = P(Z < \frac{1557.7 - \mu_1}{24.75}) =$$

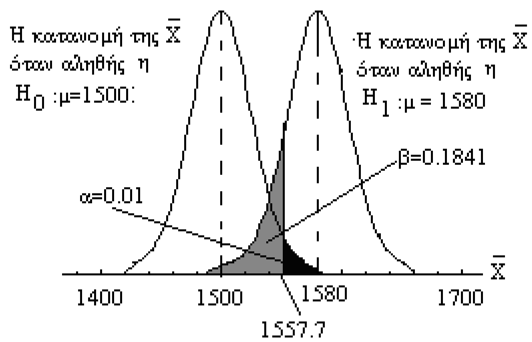
$$= \Phi(\frac{1557.7 - \mu_1}{24.75}).$$

Δηλαδή,

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

Παρατηρούμε ότι η πιθανότητα *σφάλματος τύπου ΙΙ*, 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 ΙΙ* είναι

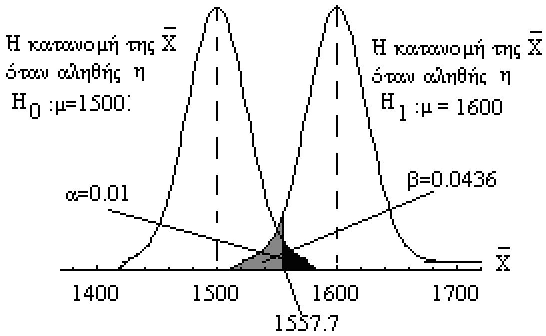
$$\beta = \Phi(\frac{1557.7 - 1580}{24.75}) = \Phi(-0.90) = 1 - \Phi(0.90) = 0.1841 \textit{ (Σχήμα 12.3.1)}$$



Σχήμα 12.3.1
Η πιθανότητα σφάλματος τύπου ΙΙ όταν $\alpha = 0.01$
και η αληθής τιμή είναι $\mu = 1580$

ενώ αν $\mu_1 = 1600$ η πιθανότητα *σφάλματος τύπου ΙΙ* είναι

$$\beta = \Phi(\frac{1557.7 - 1600}{24.75}) = \Phi(-1.71) = 1 - \Phi(1.71) = 0.0436 \textit{ (Σχήμα 12.3.2).}$$



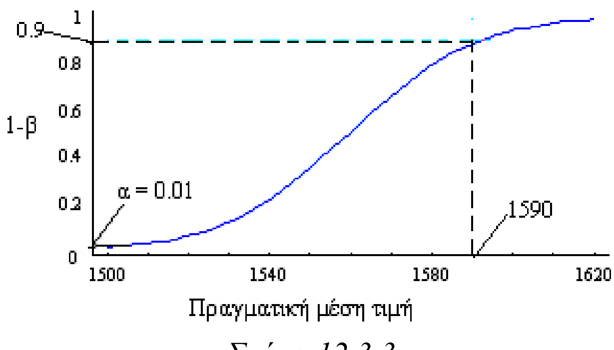
Σχήμα 12.3.2
Η πιθανότητα σφάλματος τύπου ΙΙ όταν $\alpha = 0.01$
και η αληθής τιμή είναι $\mu = 1600$

Δηλαδή, στο παράδειγμά μας, όσο πιο μακριά από την $H_0: \mu = 1500\,Kg$ (προς μεγαλύτερες τιμές), βρίσκεται η πραγματική τιμή της άγνωστης παραμέτρου μ, τόσο η πιθανότητα *σφάλματος τύπου ΙΙ* ελαττώνεται[10]. 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 $H_0: \mu = 1500\,Kg$ (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 σωστά.

Η γραφική παράσταση της *συνάρτησης ισχύος* ονομάζεται *καμπύλη ισχύος (power curve)* του ελέγχου. Στο *Σχήμα 12.3.3* φαίνεται η *καμπύλη ισχύος* του ελέγχου του παραδείγματός μας για $\alpha = 0.01$.



Σχήμα 12.3.3
Η καμπύλη ισχύος του ελέγχου της $H_0: \mu = 1500$
έναντι της $H_1: \mu > 1500$ *για* $\alpha = 0.01$

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

[10] *Είναι λογικό;*

παραμέτρου μ τείνει προς την τιμή 1500, η *ισχύς* του ελέγχου μειώνεται και τείνει προς το $0.01 = \alpha$.

Σχόλιο 12.3.1 (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 1590Kg, 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* $H_0 : \mu = 1500\,Kg$ *υπέρ της* $H_1 : \mu = \mu_1 = 1590\,Kg$ *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 ΙΙ για κάθε τιμή της παραμέτρου που ορίζει η* H_1.

■

Από τον ορισμό της πιθανότητας *σφάλματος τύπου ΙΙ,* 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 η $H_0 : \mu = 1500\,Kg$ ασφαλώς δεν απορρίπτεται[11]. Η *κρίσιμη τιμή* τώρα είναι

$$c = 1500 + z_{0.001}\frac{175}{\sqrt{50}} = 1576.5 \text{ ή ισοδύναμα, } c = z_{0.001} = 3.09$$

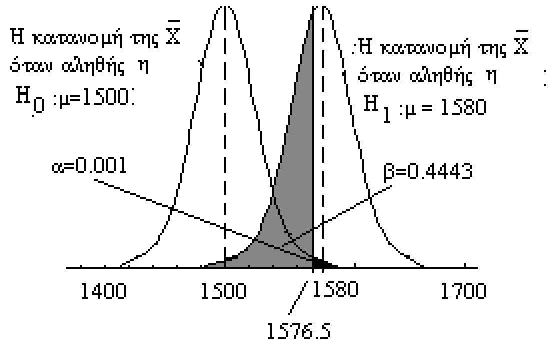
και επομένως

$$\beta = \Phi(\frac{1576.5 - \mu_1}{24.75}),\ \ \mu_1 > 1500.$$

Έτσι, αν για παράδειγμα, $\mu_1 = 1580$ τότε

$$\beta = \Phi(\frac{1576.5 - 1580}{24.75}) = \Phi(-0.14) = 0.4443.$$

[11] *Αφού, όπως είδαμε, δεν απορρίπτεται σε* $\alpha = 0.01$.



Σχήμα 12.3.4
Η πιθανότητα σφάλματος τύπου ΙΙ όταν $\alpha = 0.001$
και η αληθής τιμή είναι $\mu = 1580$

Συγκρίνοντας αυτή την τιμή, $\beta = 0.4443$ *(Σχήμα 12.3.4),* με την αντίστοιχη που υπολογίσαμε όταν κάναμε τον έλεγχο σε *επίπεδο σημαντικότητας* $\alpha = 0.01$ *(*$\beta = 0.0436$, *Σχήμα 12.3.2)* παρατηρούμε ότι ελαττώνοντας την πιθανότητα α, δηλαδή κάνοντας πιο αυστηρό τον κανόνα απόρριψης της *μηδενικής υπόθεσης*, η πιθανότητα *σφάλματος τύπου ΙΙ* 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 Ι* και την πιθανότητα *σφάλματος τύπου ΙΙ*, 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 Ι*). Τι κάνουμε όμως με την πιθανότητα β; Την αφήνουμε «ανεξέλεγκτη»; Δηλαδή, πώς μπορούμε για δεδομένη πιθανότητα *σφάλματος τύπου Ι*, να μειώσουμε την πιθανότητα *σφάλματος τύπου ΙΙ*; 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 εφικτός[12], είναι η αύξηση του μεγέθους του δείγματος. Με αυτόν τον τρόπο, επιτυγχάνουμε να μειώσουμε τη διακύμανση $\sigma^2_{\overline{X}} = \sigma^2/\nu$ της *στατιστικής συνάρτησης ελέγχου* $\overline{X}$. Αυτό, θεωρητικά επιτυγχάνεται και με μείωση της διακύμανσης, σ^2, της *X,* διατηρώντας το μέγεθος του δείγματος ν σταθερό. Όμως, είναι φανερό, ότι αυτός ο τρόπος έχει μικρή πρακτική αξία.

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

$$c = 1500 + z_{0.001}\frac{175}{\sqrt{100}} = 1554.1$$

και επειδή η τιμή $\bar{x} = 1550$ της $\overline{X}$ δεν είναι μεγαλύτερη της *κρίσιμης τιμής*, η $H_0 : \mu = 1500\,Kg$ δεν απορρίπτεται σε *επίπεδο σημαντικότητας* $\alpha = 0.001$.

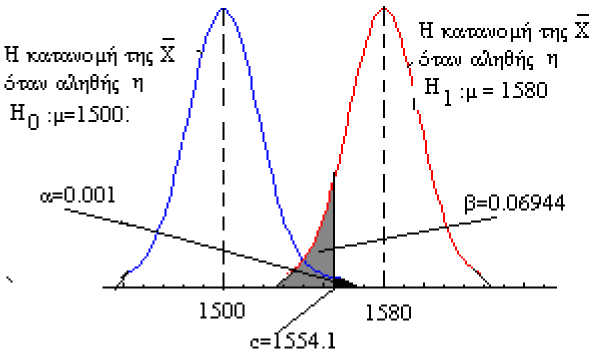
Η πιθανότητα *σφάλματος τύπου ΙΙ,* β, είναι

[12] *Όμως όχι πάντα.*

$$\beta = \Phi(\frac{1554.1-\mu_1}{17.5}), \quad \mu_1 > 1500$$

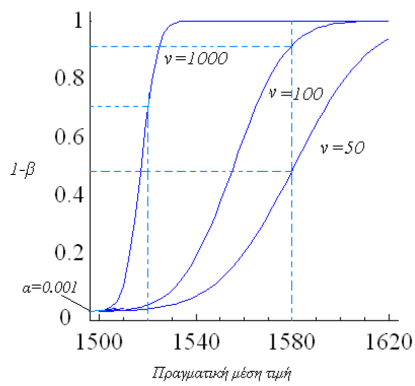
και η τιμή της για $\mu_1 = 1580$, είναι

$$\beta = \Phi(\frac{1554.1-1580}{17.5}) = \Phi(-1.48) = 0.06944 \text{ (Σχήμα 12.3.5).}$$



Σχήμα 12.3.5
Η πιθανότητα σφάλματος τύπου ΙΙ
όταν το μέγεθος του δείγματος αυξηθεί σε $\nu = 100$
(η αληθής τιμή είναι $\mu = 1580$ και $\alpha = 0.001$)

Στον αντίστοιχο έλεγχο με δείγμα μεγέθους $\nu = 50$, και για $\mu_1 = 1580$, βρήκαμε ότι πιθανότητα *σφάλματος τύπου ΙΙ* είναι $\beta = 0.4443$. Αυξάνοντας, επομένως, το μέγεθος του δείγματος, παρατηρούμε ότι η πιθανότητα *σφάλματος τύπου ΙΙ* μειώνεται. Αντίστοιχα, η *ισχύς* του ελέγχου αυξάνεται. Στο *Σχήμα 12.3.6* φαίνεται η *καμπύλη ισχύος* του ελέγχου για $\nu = 50$, $\nu = 100$ και $\nu = 1000$. 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* $\mu_1 = 1580\,Kg$ η *ισχύς* του ελέγχου για $\nu = 50$ είναι 55.57%, για $\nu = 100$ αυξάνεται σε 93% και για $\nu = 1000$ 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».



Σχήμα 12.3.6
Η καμπύλη ισχύος του ελέγχου της $H_0 : \mu = 1500$
έναντι της $H_1 : \mu > 1500$ για $\alpha = 0.001$
και $\nu = 50, 100, 1000$

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 αντίκρυσμα. Δηλαδή, κάτι πρακτικά ασήμαντο να αναδειχθεί *σημαντικό στατιστικά*! Δείτε, για παράδειγμα, στο σχήμα, την *ισχύ* του ελέγχου για πραγματική τιμή $\mu_1 = 1520\,Kg$, δηλαδή, για πραγματική τιμή πιο κοντά στην $H_0 : \mu = 1500\,Kg$. 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 70% περίπου και για ακόμη μεγαλύτερο μέγεθος δείγματος θα φθάσει το 100%. Ανάλογα, μπορούμε να επιτύχουμε απόρριψη της $H_0 : \mu = 1500\,Kg$ όσο κοντά και αν αυτή βρίσκεται στην πραγματική τιμή (π.χ. αν $\mu_1 = 1500.1Kg$). 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 νοήματος, της λογικής και κυρίως των ορίων τους.

■

Με την παρουσίαση της έννοιας της *ισχύος* ενός ελέγχου ολοκληρώσαμε την παρουσίαση των βασικών εννοιών των *στατιστικών ελέγχων υποθέσεων*. Στη συνέχεια, παρουσιάζουμε τον έλεγχο της *μέσης τιμής, p*, της κατανομής *Bernoulli,* που είναι γνωστός ως ***έλεγχος διωνυμικού ποσοστού***.

12.4 Στατιστικός έλεγχος υποθέσεων για το διωνυμικό ποσοστό

Αν $X_1, X_2, \ldots, X_\nu$ *τυχαίο δείγμα* από μια 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. Ο έλεγχος αυτός είναι επίσης γνωστός ως ***έλεγχος για τη διωμυμική παράμετρο p*** ή ***για το διωνυμικό ποσοστό p (Binomial parameter ή Binomial proportion).***

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

Παράδειγμα 12.4.1 (συνέχεια του Παραδείγματος 6.1.1): *Ο γεωπόνος ενός φυτώριου ισχυρίζεται ότι οι βολβοί τουλίπας που παράγονται στο φυτώριο βλαστάνουν σε ποσοστό 90%. Ένας αγρότης που είχε προμηθευτεί από το συγκεκριμένο φυτώριο μεγάλο αριθμό βολβών τουλίπας θέλησε να ελέγξει τον ισχυρισμό του γεωπόνου. Συγκεκριμένα, επέλεξε τυχαία 52 από τους βολβούς που φύτεψε και διαπίστωσε ότι βλάστησαν μόνο οι 38, γεγονός που του δημιούργησε αμφιβολίες για το αν πράγματι ευσταθεί ο ισχυρισμός του γεωπόνου. Είναι άραγε, δικαιολογημένες οι αμφιβολίες του αγρότη;*

Η απάντηση που δώσαμε στο *Παράδειγμα 6.1.1* δικαιώνει όπως είδαμε τις αμφιβολίες του αγρότη αφού όπως δείξαμε *«με την υπόθεση ότι αληθεύει ο ισχυρισμός του* *γεωπόνου, η πιθανότητα να βλαστήσουν μόνο 38 από τους 52 βολβούς είναι περίπου μηδενική».*

Υπέρ των αμφιβολιών του αγρότη συνηγορεί επίσης και το *95% διάστημα εμπιστοσύνης* (για το πραγματικό ποσοστό p των βολβών που βλαστάνουν) που κατασκευάσαμε στο *Παράδειγμα 11.3.2*.

Βέβαια, τώρα που πλέον γνωρίσαμε τους στατιστικούς ελέγχους, είναι φανερό ότι ουσιαστικά οι αμφιβολίες του αγρότη θέτουν ένα πρόβλημα *στατιστικού ελέγχου υποθέσεων* για το πραγματικό ποσοστό p των βολβών που βλαστάνουν.

Αυτό που αμφισβητείται είναι ο ισχυρισμός του γεωπόνου ότι το ποσοστό των βολβών που βλαστάνουν 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»

$$\frac{x_1+x_2+\ldots+x_{52}}{52} = \frac{38}{52} = 0.73$$

στο *τυχαίο δείγμα* $X_1, X_2, \ldots, X_{52}$ των $\nu = 52$ δοκιμών. Επομένως, για να προχωρήσουμε, πρέπει να γνωρίζουμε την κατανομή του *δειγματικού ποσοστού*

$$\frac{X_1+X_2+\ldots+X_{52}}{52} .$$

(Στην απάντηση που δώσαμε όταν μιλήσαμε για τη διωνυμική κατανομή, είχαμε υπολογίσει την πιθανότητα ο αριθμός των «επιτυχιών», $X_1,+X_2+\ldots+X_{52}$, στις 52 δοκιμές, να είναι 38 και επομένως έπρεπε να γνωρίζουμε την κατανομή της $X_1,+X_2+\ldots+X_{52}$ που εξ' ορισμού είναι η διωνυμική, $B(52, 0.9)$).

Γενικότερα, για τον έλεγχο της *μηδενικής υπόθεσης*

$$H_0 : p = p_0$$

πρέπει να γνωρίζουμε την κατανομή του *δειγματικού ποσοστού (δειγματικού μέσου)*

$$\hat{P} = \frac{X_1+X_2+\ldots+X_\nu}{\nu} .$$

Υποθέτοντας ότι η *μηδενική υπόθεση* $H_0 : p = p_0$, είναι αληθής, δηλαδή, ότι το τυχαίο δείγμα, $X_1, X_2, \ldots, X_\nu$ προέρχεται από έναν πληθυσμό που ακολουθεί την κατανομή *Bernoulli* με *μέση τιμή* $\mu = p_0$ και *διακύμανση* $\sigma^2 = p_0(1-p_0)$, το δειγματικό ποσοστό $\hat{P}$ για μεγάλα ν όπως γνωρίζουμε (*Κ.Ο.Θ.*) προσεγγίζεται ικανοποιητικά από την *κανονική* κατανομή με *μέση τιμή* p_0 και *διακύμανση*

$$\frac{\sigma^2}{\nu} = \frac{p_0(1-p_0)}{\nu} .$$

Κατά προσέγγιση, δηλαδή, έχουμε

$$\hat{P} = \frac{X_1+X_2+\ldots+X_\nu}{\nu} \sim N\left(p_0, \frac{p_0(1-p_0)}{\nu}\right).$$

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

$$\frac{\hat{P}-p_0}{\sqrt{\frac{p_0(1-p_0)}{\nu}}} \to Z \sim N(0,\,1)\,.$$

Υπενθυμίζουμε ότι η τιμή

$$\frac{x_1+x_2+\ldots+x_\nu}{\nu}$$

του *δειγματικού ποσοστού* $\hat{P}$ για συγκεκριμένη πραγματοποίηση $x_1, x_2, \ldots, x_\nu$ του τυχαίου δείγματος, έχει επικρατήσει να συμβολίζεται με $\hat{p}$ και αντίστοιχα η τιμή της *στατιστικής συνάρτησης ελέγχου* με

$$z = \frac{(\hat{p}-p_0)\sqrt{\nu}}{\sqrt{p_0(1-p_0)}}\,.$$

Έτσι, σε *επίπεδο σημαντικότητας* α, για μεγάλο μέγεθος δείγματος ν (στην πράξη, αν $\nu p_0 \geq 5$ και $\nu(1-p_0) \geq 5$ ή $\nu p_0(1-p_0) \geq 10$), απορρίπτουμε την $H_0 : p = p_0$

- έναντι της $H_1 : p > p_0$, όταν $z = \dfrac{(\hat{p}-p_0)\sqrt{\nu}}{\sqrt{p_0(1-p_0)}} \geq z_\alpha$
- έναντι της $H_1 : p < p_0$, όταν $z = \dfrac{(\hat{p}-p_0)\sqrt{\nu}}{\sqrt{p_0(1-p_0)}} \leq -z_\alpha$
- έναντι της $H_1 : p \neq p_0$, όταν

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

Ας ολοκληρώσουμε στο παράδειγμά μας τον έλεγχο της *μηδενικής υπόθεσης*

$$H_0 : p = 0.9$$

έναντι της *εναλλακτικής*

$$H_1 : p < 0.9$$

σε *επίπεδο σημαντικότητας* $\alpha = 0.05$.

Το *δειγματικό ποσοστό* είναι $\hat{p} = 38/52 = 0.73$ και επειδή $\nu\hat{p} = 52 \cdot 0.73 = 38 \geq 5$ και $\nu(1-\hat{p}) = 52 \cdot 0.27 = 14 \geq 5$ η *περιοχή απόρριψης* του ελέγχου είναι

$$z = \frac{(\hat{p}-p_0)\sqrt{\nu}}{\sqrt{p_0(1-p_0)}} \leq -z_{0.05} = -1.645\,.$$

Η τιμή της *στατιστικής συνάρτησης ελέγχου* είναι

$$z = \frac{(\hat{p}-p_0)\sqrt{\nu}}{\sqrt{p_0(1-p_0)}} = \frac{(0.73-0.9)\sqrt{52}}{\sqrt{0.9\cdot(1-0.9)}} = -4.09$$

και επομένως, σε *επίπεδο σημαντικότητας* $\alpha = 0.05$ απορρίπτουμε τη *μηδενική υπόθεση* αφού $z = -4.09 \leq -1.645$. Δηλαδή, ο αγρότης, με πιθανότητα να κάνει λάθος το πολύ 0.05, αμφιβάλει δικαιολογημένα.

■

Σημείωση 12.4.1: *Επισημαίνουμε, ότι οι παραπάνω έλεγχοι είναι επιπέδου σημαντικότητας* α ***κατά προσέγγιση****. 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* $X_1+X_2+\ldots+X_\nu \sim B(\nu,\, p_0)$. *Όμως, στην περίπτωση αυτή, με τη διαδικασία ελέγχου που περιγράψαμε δε μπορούμε πάντοτε να πετύχουμε ένα προκαθορισμένο επίπεδο σημαντικότητας! Το πρόβλημα αυτό αντιμετωπίζεται, όμως δε θα επεκταθούμε περισσότερο. Σημειώνουμε τέλος, ότι σε προβλήματα ελέγχου ποσοστού, στην πράξη συνήθως δεν αντιμετωπίζουμε πρόβλημα μικρών δειγμάτων.*

■

Στη συνέχεια, παρουσιάζουμε τον έλεγχο της *διακύμανσης* σ^2 ενός κανονικού πληθυσμού. Ολοκληρώνουμε έτσι, την παρουσίαση των ελέγχων που αναφέρονται σε έναν πληθυσμό.

12.5. Στατιστικός έλεγχος υποθέσεων για τη διακύμανση ενός κανονικού πληθυσμού

Σε πολλά προβλήματα (όπως, στον έλεγχο της ακρίβειας οργάνων μέτρησης, στον έλεγχο προδιαγραφών και γενικότερα στον έλεγχο ποιότητας), μας ενδιαφέρει ο έλεγχος της *διακύμανσης* ενός πληθυσμού. Ας δούμε ένα παράδειγμα.

Παράδειγμα 12.5.1: *Μια αυτόματη μηχανή συσκευάζει καλαμπόκι σε τσουβάλια των 25Kg. Η μηχανή έχει ρυθμισθεί έτσι, ώστε οι ποσότητες καλαμποκιού που συσκευάζονται ανά τσουβάλι να έχουν τυπική απόκλιση 1.5Kg. Επίσης, έχει παρατηρηθεί ότι οι ποσότητες αυτές ακολουθούν κανονική κατανομή. Ο υπεύθυνος παραγωγής υποψιάζεται ότι η μηχανή έχει απορυθμισθεί και θέλει να ελέγξει αν η τυπική απόκλιση των ποσοτήτων καλαμποκιού που συσκευάζονται ανά τσουβάλι είναι πράγματι 1.5Kg. Για το σκοπό αυτό, από την παραγωγή μιας ημέρας, επέλεξε τυχαία 30 τσουβάλια, κατέγραψε τα βάρη τους και υπολόγισε τον μέσο τους και την τυπική τους απόκλιση και αντίστοιχα βρήκε* $\bar{x} = 24.8Kg$ *και* $s = 1.6Kg$. *Σε επίπεδο σημαντικότητας 5%, υποστηρίζουν αυτά τα δεδομένα τις υποψίες του υπεύθυνου παραγωγής;*

<u>*Απάντηση:*</u> Είναι φανερό ότι πρέπει να κάνουμε τον έλεγχο της *μηδενικής υπόθεσης*

$$H_0 : \sigma^2 = 1.5^2$$

έναντι της *εναλλακτικής*

$$H_0 : \sigma^2 \neq 1.5^2 .$$

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

Στο *10^ο Κεφάλαιο* είδαμε ότι αν $X_1, X_2, \ldots, X_\nu$ ένα τυχαίο δείγμα από έναν κανονικό πληθυσμό με *διακύμανση* σ^2, τότε για οποιοδήποτε *μέγεθος δείγματος* ν, η τυχαία μεταβλητή

$$\frac{(\nu-1)S^2}{\sigma^2}$$

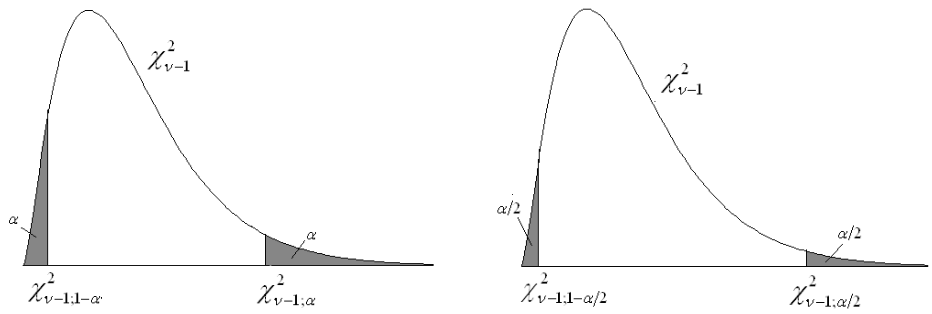
ακολουθεί την κατανομή $\chi^2_{\nu-1}$, δηλαδή, μια *κατανομή χι τετράγωνο* με $\nu-1$ *βαθμούς ελευθερίας*. Υπενθυμίζουμε ότι με S^2 συμβολίζουμε την *(αμερόληπτη) δειγματική διακύμανση*

$$S^2 = \frac{1}{\nu-1}\sum_{i=1}^{\nu}(X_i - \bar{X})^2 .$$

Με την υπόθεση ότι η *μηδενική υπόθεση* $H_0 : \sigma^2 = \sigma_0^2$ είναι αληθής, η τιμή της $(\nu - 1)S^2 / \sigma^2$ μπορεί να υπολογισθεί από το δείγμα (αφού σ^2 γνωστή, ίση με σ_0^2) και επειδή η κατανομή της είναι γνωστή, είναι λογικό, ως *στατιστική συνάρτηση ελέγχου* να χρησιμοποιήσουμε την

$$\frac{(\nu-1)S^2}{\sigma_0^2}.$$

Είναι επίσης λογικό, ως *κρίσιμα σημεία* για τον ορισμό των *περιοχών απόρριψης* σε *επίπεδο σημαντικότητας* α, να προκύψουν τα σημεία $\chi^2_{\nu-1;\alpha}$, $\chi^2_{\nu-1;1-\alpha}$, $\chi^2_{\nu-1;\alpha/2}$ και $\chi^2_{\nu-1;1-\alpha/2}$ της κατανομής $\chi^2_{\nu-1}$. Όπως φαίνεται στα *Σχήματα 12.5.1,* δεξιά των σημείων $\chi^2_{\nu-1;\alpha}$ και $\chi^2_{\nu-1;\alpha/2}$ ορίζονται περιοχές εμβαδού, αντίστοιχα, α και $\alpha/2$. Ομοίως, δεξιά των σημείων $\chi^2_{\nu-1;1-\alpha}$ και $\chi^2_{\nu-1;1-\alpha/2}$ ορίζονται περιοχές εμβαδού, αντίστοιχα, $1-\alpha$ και $1-\alpha/2$ και επομένως αριστερά, περιοχές εμβαδού α και $\alpha/2$.



Σχήμα 12.5.1
Άνω ποσοστιαία σημεία της κατανομής $\chi^2_{\nu-1}$

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 $\dfrac{(\nu-1)s^2}{\sigma_0^2} \geq \chi^2_{\nu-1;\alpha}$
- έναντι της $H_1 : \sigma^2 < \sigma_0^2$, όταν $\dfrac{(\nu-1)s^2}{\sigma_0^2} \leq \chi^2_{\nu-1;1-\alpha}$
- έναντι της $H_1 : \sigma^2 \neq \sigma_0^2$, όταν $\dfrac{(\nu-1)s^2}{\sigma_0^2} \geq \chi^2_{\nu-1;\alpha/2}$ ή $\dfrac{(\nu-1)s^2}{\sigma_0^2} \leq \chi^2_{\nu-1;1-\alpha/2}$.

Ας ολοκληρώσουμε τον έλεγχο του παραδείγματός μας.

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

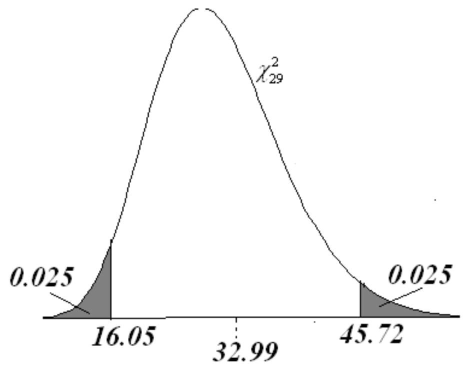
$$\frac{(\nu-1)s^2}{\sigma_0^2} \geq \chi^2_{29;0.025} \text{ ή } \frac{(\nu-1)s^2}{\sigma_0^2} \leq \chi^2_{29;0.975} \Leftrightarrow$$

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

Επειδή η τιμή της *στατιστικής συνάρτησης ελέγχου* στο δείγμα

$$\frac{(\nu-1)s^2}{\sigma_0^2}=\frac{29\cdot 1.6^2}{1.5^2}=32.99$$

δε βρίσκεται στην *περιοχή απόρριψης (Σχήμα 12.5.2)*, η *μηδενική υπόθεση*, σε *επίπεδο σημαντικότητας* $\alpha = 0.05$ δεν απορρίπτεται, δηλαδή, τα δεδομένα του δείγματος σε *επίπεδο σημαντικότητας* $\alpha = 0.05$ δε δίνουν στατιστικά σημαντικές αποδείξεις ότι η μηχανή έχει απορυθμισθεί.



Σχήμα 12.5.2
Η τιμή 32.99 της συνάρτησης ελέγχου
δε βρίσκεται στην περιοχή απόρριψης της $H_0 : \sigma^2 = 1.5^2$

■

Η μέθοδος *στατιστικού ελέγχου υποθέσεων* που περιγράψαμε, εφαρμόζεται και για τη σύγκριση παραμέτρων δύο πληθυσμών. Στη συνέχεια αναφερόμαστε σε αυτές τις περιπτώσεις ελέγχων.

12.6 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 5Kg. 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 τα *διαστήματα εμπιστοσύνης* που αφορούν παραμέτρους δύο πληθυσμών, πρέπει να διακρίνουμε την περίπτωση που *τα δύο δείγματα είναι ανεξάρτητα* από την περίπτωση που *τα δύο δείγματα δεν είναι ανεξάρτητα.*

12.6.1 Ανεξάρτητα δείγματα

Ας δούμε ένα παράδειγμα.

***Παράδειγμα 12.6.1**: 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, κτλ.) 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 κτλ.), 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. Άραγε, 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);*

Απάντηση: 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 *η κατανομή των τιμών της συγκέντρωσης μολύβδου στο πόσιμο νερό που φθάνει σε όλες τις παροχές του κέντρου της πόλης* και ο άλλος, έστω *B*, *η κατανομή των τιμών 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, ασφαλώς, θα γίνει με βάση αυτό που παρατηρείται στα δύο δείγματα, στο δείγμα $X_1, X_2, \ldots, X_{100}$ από τον πληθυσμό *A* και στο δείγμα $Y_1, Y_2, \ldots, Y_{90}$ από τον πληθυσμό *B*.

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

Με βάση όσα αναφέραμε στο *10^o Κεφάλαιο,* αν $X_1, X_2, \ldots, X_{\nu_1}$ και $Y_1, Y_2, \ldots, Y_{\nu_2}$ δύο *τυχαία δείγματα* που έχουν ληφθεί *ανεξάρτητα* το ένα από το άλλο από δύο

πληθυσμούς A και B αντίστοιχα, τότε γνωρίζουμε ή μπορούμε να προσεγγίσουμε ικανοποιητικά την κατανομή της διαφοράς, $\overline{X}-\overline{Y}$, των αντίστοιχων *δειγματικών μέσων*, στις ακόλουθες περιπτώσεις.
α) Όταν τα δείγματα προέρχονται από κανονικούς πληθυσμούς.
β) Όταν τα δείγματα είναι μεγάλα (και οι πληθυσμοί όχι κατ' ανάγκη κανονικοί).

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

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

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

Επίσης, όπως κάναμε και στους ελέγχους για τη *μέση τιμή* ενός πληθυσμού, θα διακρίνουμε τις περιπτώσεις που οι διακυμάνσεις των δύο πληθυσμών είναι γνωστές από τις περιπτώσεις που δεν είναι γνωστές.

(α) Οι πληθυσμοί είναι κανονικοί και οι διακυμάνσεις τους είναι γνωστές

Έστω δύο *ανεξάρτητα τυχαία δείγματα* $X_1, X_2, \ldots, X_{\nu_1}$ και $Y_1, Y_2, \ldots, Y_{\nu_2}$ από δύο *κανονικούς* πληθυσμούς με *διακυμάνσεις* σ_1^2 και σ_2^2 γνωστές. Επειδή

$$X_1, X_2, \ldots, X_{\nu_1} \sim N(\mu_1, \sigma_1^2) \text{ και } Y_1, Y_2, \ldots, Y_{\nu_2} \sim N(\mu_2, \sigma_2^2)$$

η κατανομή της τυχαίας μεταβλητής $\overline{X}-\overline{Y}$ όπως είδαμε στο *10^ο Κεφάλαιο* για οτιδήποτε ν_1 και ν_2 (μικρά ή μεγάλα) είναι *κανονική* με

$$\overline{X}-\overline{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{\nu_1} + \frac{\sigma_2^2}{\nu_2}\right)$$

και επομένως

$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{\nu_1}+\dfrac{\sigma_2^2}{\nu_2}}} = Z \sim N(0,1).$$

Επειδή, οι *διακυμάνσεις* 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, στη συνάρτηση

$$Z = \frac{\overline{X}-\overline{Y}-\delta}{\sqrt{\dfrac{\sigma_1^2}{\nu_1}+\dfrac{\sigma_2^2}{\nu_2}}}$$

δεν υπάρχουν άγνωστες παράμετροι και επομένως μπορεί να χρησιμοποιηθεί ως *στατιστική συνάρτηση ελέγχου* αφού η τιμή της μπορεί να υπολογισθεί από τα δείγματα και η κατανομή της είναι γνωστή. Έτσι, σε *επίπεδο σημαντικότητας* α, απορρίπτουμε την $H_0 : \mu_1 - \mu_2 = \delta$

- έναντι της $H_1 : \mu_1 - \mu_2 > \delta$, όταν $z = \dfrac{\overline{x}-\overline{y}-\delta}{\sqrt{\dfrac{\sigma_1^2}{\nu_1}+\dfrac{\sigma_2^2}{\nu_2}}} \geq z_\alpha$

• έναντι της $H_1 : \mu_1 - \mu_2 < \delta$, όταν $z = \dfrac{\bar{x} - \bar{y} - \delta}{\sqrt{\dfrac{\sigma_1^2}{\nu_1} + \dfrac{\sigma_2^2}{\nu_2}}} \le -z_\alpha$

• έναντι της $H_1 : \mu_1 - \mu_2 \ne \delta$, όταν $|z| = \dfrac{|\bar{x} - \bar{y} - \delta|}{\sqrt{\dfrac{\sigma_1^2}{\nu_1} + \dfrac{\sigma_2^2}{\nu_2}}} \ge z_{\alpha/2}$

(β) Οι πληθυσμοί είναι κανονικοί και οι διακυμάνσεις τους είναι άγνωστες και ίσες

Έστω δύο *ανεξάρτητα τυχαία δείγματα* $X_1, X_2, \ldots, X_{\nu_1}$ και $Y_1, Y_2, \ldots, Y_{\nu_2}$ από δύο *κανονικούς* πληθυσμούς με άγνωστες αλλά ίσες *διακυμάνσεις*, έστω δηλαδή

$$X_1, X_2, \ldots, X_{\nu_1} \sim N(\mu_1, \sigma^2) \text{ και } Y_1, Y_2, \ldots, Y_{\nu_2} \sim N(\mu_2, \sigma^2).$$

Στην περίπτωση αυτή, όπως είδαμε στο *10ο Κεφάλαιο,* η τυχαία μεταβλητή

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S\sqrt{\dfrac{1}{\nu_1} + \dfrac{1}{\nu_2}}},$$

για οτιδήποτε ν_1 και ν_2 (μικρά ή μεγάλα), ακολουθεί την κατανομή $t_{\nu_1+\nu_2-2}$ (την *t κατανομή* με $\nu_1 + \nu_2 - 2$ *βαθμούς ελευθερίας*), δηλαδή

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S\sqrt{\dfrac{1}{\nu_1} + \dfrac{1}{\nu_2}}} \sim t_{\nu_1+\nu_2-2}$$

όπου

$$S^2 = \frac{(\nu_1 - 1)S_1^2 + (\nu_2 - 1)S_2^2}{\nu_1 + \nu_2 - 2}$$

μια *εκτιμήτρια* της *κοινής διακύμανσης* σ^2 των δύο πληθυσμών με βάση και τα δύο δείγματα.

Με την υπόθεση ότι η *μηδενική υπόθεση* είναι αληθής, δηλαδή με την υπόθεση ότι η διαφορά των *μέσων τιμών* είναι ίση με δ , στη συνάρτηση

$$T = \frac{\bar{X} - \bar{Y} - \delta}{S\sqrt{\dfrac{1}{\nu_1} + \dfrac{1}{\nu_2}}}$$

δεν υπάρχουν άγνωστες παράμετροι και επομένως μπορεί να χρησιμοποιηθεί ως *στατιστική συνάρτηση ελέγχου* αφού η τιμή της μπορεί να υπολογισθεί από τα δείγματα και η κατανομή της είναι γνωστή. Έτσι, σε *επίπεδο σημαντικότητας* α*,* απορρίπτουμε την $H_0 : \mu_1 - \mu_2 = \delta$

• έναντι της $H_1 : \mu_1 - \mu_2 > \delta$, όταν $t = \dfrac{\bar{x} - \bar{y} - \delta}{s\sqrt{\dfrac{1}{\nu_1} + \dfrac{1}{\nu_2}}} \ge t_{\nu_1+\nu_2-2;\alpha}$

• έναντι της $H_1 : \mu_1 - \mu_2 < \delta$, όταν $t = \dfrac{\bar{x} - \bar{y} - \delta}{s\sqrt{\dfrac{1}{\nu_1} + \dfrac{1}{\nu_2}}} \le -t_{\nu_1+\nu_2-2;\alpha}$

- έναντι της $H_1 : \mu_1 - \mu_2 \neq \delta$, όταν $|t| = \dfrac{|\bar{x} - \bar{y} - \delta|}{s\sqrt{\dfrac{1}{\nu_1} + \dfrac{1}{\nu_2}}} \geq t_{\nu_1+\nu_2-2;\alpha/2}$

Σημείωση 12.6.1: *Για τις δύο προϋποθέσεις εφαρμογής του ελέγχου αυτής της περίπτωσης (κανονικοί πληθυσμοί και ίσες διακυμάνσεις) σημειώνουμε τα εξής. α) Ακόμη και αν οι πληθυσμοί δεν είναι κανονικοί (και εφόσον, ασφαλώς 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 διακυμάνσεις, 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 οι διακυμάνσεις των πληθυσμών είναι γνωστές

Έστω δύο *ανεξάρτητα τυχαία δείγματα* $X_1, X_2, \ldots, X_{\nu_1}$ και $Y_1, Y_2, \ldots, Y_{\nu_2}$ από δύο πληθυσμούς όχι κατ' ανάγκη κανονικούς (οποιουσδήποτε) και έστω επίσης ότι τα μεγέθη ν_1 και ν_2 των δειγμάτων είναι μεγάλα (στην πράξη, $\nu_1 \geq 30$ και $\nu_2 \geq 30$). Με βάση όσα αναφέραμε στο *10ο Κεφάλαιο,* επειδή τα μεγέθη των δειγμάτων είναι μεγάλα, **κατά προσέγγιση** έχουμε

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{\nu_1} + \dfrac{\sigma_2^2}{\nu_2}}} = Z \sim N(0,1)$$

επομένως, αν οι διακυμάνσεις των πληθυσμών σ_1^2 και σ_2^2 είναι γνωστές, ισχύουν και εφαρμόζονται όσα αναφέραμε στην περίπτωση (α). Βέβαια τώρα, οι αντίστοιχοι έλεγχοι είναι επιπέδου σημαντικότητας α **κατά προσέγγιση**. Ασφαλώς, όσο μεγαλύτερα είναι τα ν_1 και ν_2, τόσο καλύτερη είναι η προσέγγιση.

(δ) Τα δείγματα είναι μεγάλα και οι διακυμάνσεις των πληθυσμών είναι άγνωστες

Έστω δύο *ανεξάρτητα τυχαία δείγματα* $X_1, X_2, \ldots, X_{\nu_1}$ και $Y_1, Y_2, \ldots, Y_{\nu_2}$ από δύο πληθυσμούς όχι κατ' ανάγκη κανονικούς (οποιουσδήποτε) και έστω επίσης ότι τα μεγέθη ν_1 και ν_2 των δειγμάτων είναι μεγάλα (στην πράξη, $\nu_1 \geq 30$ και $\nu_2 \geq 30$). Με βάση όσα αναφέραμε στο *10ο Κεφάλαιο,* επειδή τα μεγέθη των δειγμάτων είναι μεγάλα, κατά προσέγγιση έχουμε

$$\frac{\bar{X} - \bar{Y} - \delta}{\sqrt{\dfrac{S_1^2}{\nu_1} + \dfrac{S_2^2}{\nu_2}}} \sim N(0,1)$$

(όπου S_1^2, S_2^2 οι *δειγματικές διακυμάνσεις*-αμερόληπτες εκτιμήτριες των σ_1^2, σ_2^2).

Επομένως, όταν οι *διακυμάνσεις* σ_1^2 και σ_2^2 των δύο πληθυσμών είναι άγνωστες, ως *στατιστική συνάρτηση ελέγχου* χρησιμοποιούμε την τυχαία μεταβλητή

$$\frac{\bar{X} - \bar{Y} - \delta}{\sqrt{\dfrac{S_1^2}{\nu_1} + \dfrac{S_2^2}{\nu_2}}}$$

και έτσι, σε *επίπεδο σημαντικότητας* α, απορρίπτουμε την $H_0 : \mu_1 - \mu_2 = \delta$

- έναντι της $H_1 : \mu_1 - \mu_2 > \delta$, όταν $z = \dfrac{\bar{x} - \bar{y} - \delta}{\sqrt{\dfrac{s_1^2}{\nu_1} + \dfrac{s_2^2}{\nu_2}}} \geq z_\alpha$
- έναντι της $H_1 : \mu_1 - \mu_2 < \delta$, όταν $z = \dfrac{\bar{x} - \bar{y} - \delta}{\sqrt{\dfrac{s_1^2}{\nu_1} + \dfrac{s_2^2}{\nu_2}}} \leq -z_\alpha$
- έναντι της $H_1 : \mu_1 - \mu_2 \neq \delta$, όταν $|z| = \dfrac{|\bar{x} - \bar{y} - \delta|}{\sqrt{\dfrac{s_1^2}{\nu_1} + \dfrac{s_2^2}{\nu_2}}} \geq z_{\alpha/2}$

Επειδή η στατιστική συνάρτηση ελέγχου δεν είναι κανονική αλλά προσεγγίζεται ικανοποιητικά από την κανονική, οι παραπάνω έλεγχοι είναι επιπέδου σημαντικότητας α **κατά προσέγγιση**. Ασφαλώς, όσο μεγαλύτερα είναι τα ν_1 και ν_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 μπορούν να θεωρηθούν *ανεξάρτητα* μεταξύ τους. Επίσης, οι κατανομές και οι *διακυμάνσεις* των πληθυσμών από τους οποίους προέρχονται τα δείγματα είναι άγνωστες.

Επειδή τα μεγέθη των δειγμάτων είναι μεγάλα ($\nu_A = 100 \geq 30$ και $\nu_B = 90 \geq 30$), πρόκειται για έλεγχο της περίπτωσης ***(δ)*** και επομένως, σε *επίπεδο σημαντικότητας* $\alpha = 5\%$, *η περιοχή απόρριψης* της *μηδενικής υπόθεσης*

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

έναντι της *εναλλακτικής*

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

είναι

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

όπου

$$\delta = 0,\ \bar{x} = 36,\ \bar{y} = 34.1,\ s_A = 6,\ s_B = 5.9,\ \nu_A = 100 \text{ και } \nu_B = 90.$$

Επειδή

$$z = \frac{\bar{x} - \bar{y} - \delta}{\sqrt{\dfrac{s_A^2}{\nu_A} + \dfrac{s_B^2}{\nu_B}}} = \frac{36 - 34.1 - 0}{\sqrt{\dfrac{6^2}{100} + \dfrac{5.9^2}{90}}} \cong 2.2 \geq z_{0.025} = 1.96$$

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 0.05.

Παρατήρηση 12.6.1: *Αν η εταιρεία ύδρευσης θέλει να ελέγξει αν η μέση συγκέντρωση μολύβδου στο πόσιμο νερό που φθάνει στο κέντρο της πόλης (πληθυσμός Α) είναι* *μεγαλύτερη από τη μέση συγκέντρωση μολύβδου στο πόσιμο νερό που φθάνει στα ανατολικά προάστια (πληθυσμός Β), πρέπει να κάνει τον έλεγχο της ίδιας μηδενικής υπόθεσης*

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

έναντι όμως της εναλλακτικής

$$H_1 : \mu_A - \mu_B > 0 .$$

Ασφαλώς 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 0.05.

Παρατήρηση 12.6.2: *Η εταιρεία ύδρευσης θέλει να ελέγξει αν η διαφορά των μέσων συγκεντρώσεων μολύβδου μεταξύ των δύο περιοχών ξεπερνά τα 2ppm. Στην περίπτωση αυτή πρέπει να κάνει τον έλεγχο της μηδενικής υπόθεσης*

$$H_0 : \mu_A - \mu_B = 2$$

έναντι της εναλλακτικής

$$H_0 : \mu_A - \mu_B > 2 .$$

Σε επίπεδο σημαντικότητας $\alpha = 5\%$, *η περιοχή απόρριψης της μηδενικής υπόθεσης είναι*

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

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.*

■

Στον *Πίνακα 12.6.1* που ακολουθεί συνοψίζουμε, για διευκόλυνσή μας, όλες τις προηγούμενες περιπτώσεις.

	$H_1 : \mu_1 - \mu_2 \neq \delta$	$H_1 : \mu_1 - \mu_2 > \delta$	$H_1 : \mu_1 - \mu_2 < \delta$	**Προϋποθέσεις**
Περιοχή απόρριψης της $H_0 : \mu_1 - \mu_2 = \delta$	$\lvert Z \rvert = \dfrac{\lvert \overline{X} - \overline{Y} - \delta \rvert}{\sqrt{\dfrac{\sigma_1^2}{\nu_1} + \dfrac{\sigma_2^2}{\nu_2}}} \geq z_{\alpha/2}$	$Z = \dfrac{\overline{X} - \overline{Y} - \delta}{\sqrt{\dfrac{\sigma_1^2}{\nu_1} + \dfrac{\sigma_2^2}{\nu_2}}} \geq z_{\alpha}$	$Z = \dfrac{\overline{X} - \overline{Y} - \delta}{\sqrt{\dfrac{\sigma_1^2}{\nu_1} + \dfrac{\sigma_2^2}{\nu_2}}} \leq -z_{\alpha}$	*Οι διακυμάνσεις,* σ_1^2, σ_2^2 *είναι γνωστές και οι πληθυσμοί είναι κανονικοί* *ή* *Οι διακυμάνσεις,* σ_1^2, σ_2^2 *είναι γνωστές και τα* ν_1, ν_2 *είναι μεγάλα*
	$\lvert Z \rvert = \dfrac{\lvert \overline{X} - \overline{Y} - \delta \rvert}{\sqrt{\dfrac{S_1^2}{\nu_1} + \dfrac{S_2^2}{\nu_2}}} \geq z_{\alpha/2}$	$Z = \dfrac{\overline{X} - \overline{Y} - \delta}{\sqrt{\dfrac{S_1^2}{\nu_1} + \dfrac{S_2^2}{\nu_2}}} \geq z_{\alpha}$	$Z = \dfrac{\overline{X} - \overline{Y} - \delta}{\sqrt{\dfrac{S_1^2}{\nu_1} + \dfrac{S_2^2}{\nu_2}}} \leq -z_{\alpha}$	*Οι διακυμάνσεις,* σ_1^2, σ_2^2 *είναι άγνωστες και τα* ν_1, ν_2 *είναι μεγάλα* *(οτιδήποτε πληθυσμοί)*
	$\lvert T \rvert = \dfrac{\lvert \overline{X} - \overline{Y} - \delta \rvert}{S\sqrt{\dfrac{1}{\nu_1} + \dfrac{1}{\nu_2}}} \geq t_{\nu,\alpha/2}$	$T = \dfrac{\overline{X} - \overline{Y} - \delta}{S\sqrt{\dfrac{1}{\nu_1} + \dfrac{1}{\nu_2}}} \geq t_{\nu,\alpha}$	$T = \dfrac{\overline{X} - \overline{Y} - \delta}{S\sqrt{\dfrac{1}{\nu_1} + \dfrac{1}{\nu_2}}} \leq -t_{\nu,\alpha}$	*Οι διακυμάνσεις,* σ_1^2, σ_2^2 *είναι άγνωστες και ίσες, οι πληθυσμοί είναι κανονικοί, και τα* ν_1, ν_2 *οτιδήποτε.* $\nu = \nu_1 + \nu_2 - 2$ $S^2 = \dfrac{(\nu_1 - 1)S_1^2 + (\nu_2 - 1)S_2^2}{\nu_1 + \nu_2 - 2}$
	?	?	?	*Τα* ν_1, ν_2 *μικρά, οι πληθυσμοί όχι κανονικοί και οι διακυμάνσεις γνωστές ή άγνωστες (ίσες ή όχι)*

Πίνακας 12.6.1
Στατιστικοί έλεγχοι υποθέσεων για τη διαφορά $\mu_1 - \mu_2$ *των μέσων τιμών δύο πληθυσμών με δύο ανεξάρτητα δείγματα μεγέθους* ν_1, ν_2 *αντίστοιχα και σε επίπεδο σημαντικότητας* α

Ας δούμε δύο ακόμη παραδείγματα.

Παράδειγμα 12.6.2: *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 απόδοση 8.5Kg/στρέμμα με τυπική απόκλιση 1.2Kg/στρέμμα. Σε μια γειτονική αγροτική περιοχή οι αγρότες καλλιεργούν μια άλλη ποικιλία του φυτού, έστω B. Η μέση απόδοση στα κτήματα 15 αγροτών αυτής της περιοχής (που επελέγησαν επίσης τυχαία) ήταν την τελευταία χρονιά 11Kg/στρέμμα με τυπική απόκλιση 1.1Kg/στρέμμα. Άραγε, η διαφορά που παρατηρείται μεταξύ των δύο δειγμάτων είναι στατιστικά σημαντική ή μήπως οφείλεται στην τύχη.*

Απάντηση: 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 $X_1, X_2, \ldots, X_{10}$ και $Y_1, Y_2, \ldots, Y_{15}$ έχουν ληφθεί το ένα *ανεξάρτητα* από το άλλο και επομένως θα εξετάσουμε σε ποια από τις 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 δε γνωρίζουμε αν είναι *κανονικοί* (ούτε κάτι για τις *διακυμάνσεις* τους), θα χρησιμοποιήσουμε την *περιοχή απόρριψης* της περίπτωσης ***(β)*** κάνοντας όμως δύο παραδοχές: ότι τα δείγματα προέρχονται από *κανονικούς* πληθυσμούς με ίσες *διακυμάνσεις*.

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

$$t = \frac{\bar{x} - \bar{y} - \delta}{s\sqrt{\dfrac{1}{\nu_A} + \dfrac{1}{\nu_B}}} \leq -t_{\nu_A + \nu_B - 2;0.05}$$

όπου

$$\delta = 0,\ \bar{x} = 8.5,\ \bar{y} = 11,\ \nu_A = 10,\ \nu_B = 15,\ \alpha = 0.05$$

και το s υπολογίζεται από τον τύπο

$$s^2 = \frac{(\nu_A - 1)s_A^2 + (\nu_B - 1)s_B^2}{\nu_A + \nu_B - 2}$$

όπου

$$s_A = 1.2 \text{ και } s_B = 1.1.$$

Άρα έχουμε

$$t = \frac{\bar{x} - \bar{y} - \delta}{s\sqrt{\frac{1}{\nu_A} + \frac{1}{\nu_B}}} \leq -t_{23;0.05} \text{ ή } t = \frac{\bar{x} - \bar{y} - \delta}{s\sqrt{\frac{1}{\nu_A} + \frac{1}{\nu_B}}} \leq -1.714 .$$

Υπολογίζουμε την τιμή της *εκτιμήτριας* της *κοινής διακύμανσης*,

$$s^2 = \frac{(\nu_A - 1)s_A^2 + (\nu_B - 1)s_B^2}{\nu_A + \nu_B - 2} = \frac{9 \cdot 1.2^2 + 14 \cdot 1.1^2}{10 + 15 - 2} = 1.3 \Rightarrow s = 1.14$$

και επομένως η τιμή της *στατιστικής συνάρτησης ελέγχου* είναι

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

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 Α* είναι μικρότερη από *τη μέση στρεμματική απόδοση της ποικιλίας Β*. Η πιθανότητα αυτό το συμπέρασμα να είναι λάθος, είναι το πολύ 0.05.

Ερώτηση: *Μόλις διαπιστώσαμε ότι τα ευρήματα στα δύο δείγματα δίνουν στατιστικά σημαντικές αποδείξεις ότι η μέση στρεμματική απόδοση της ποικιλίας Α είναι μικρότερη από τη μέση στρεμματική απόδοση της ποικιλίας Β. Μάλιστα, αν υπολογίσουμε την P-Τιμή του ελέγχου προκύπτει ότι οι αποδείξεις αυτές είναι πολύ σημαντικές (είναι πολύ μικρή, μικρότερη από 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 Α, αυτή η διαφορά είναι σημαντική μόνο αν είναι μεγαλύτερη από 2Kg/στρέμμα γιατί κρίνουν ότι μόνο τότε αξίζει να αλλάξουν τις συνήθειές τους. Τι έλεγχο πρέπει να κάνουμε για να δώσουμε μια απάντηση στους αγρότες;*

Σχόλιο 12.6.1: *Αν κάποιος αγρότης ισχυρισθεί ότι η απόδοση της ποικιλίας Β είναι μεγαλύτερη γιατί στη γειτονική περιοχή που καλλιεργείται, οι καλλιεργητικές συνθήκες είναι πιο ευνοϊκές, τι μπορούμε άραγε να απαντήσουμε; Απάντηση σε αυτό το εύλογο ερώτημα δίνουμε στα επόμενα όταν θα μιλήσουμε και για τη σύγκριση κατά ζεύγη.*

■

Παράδειγμα 12.6.3**:** *Προκειμένου ένας φοιτητής να συγκρίνει στο πλαίσιο της πτυχιακής του εργασίας την αποτελεσματικότητα δύο διαιτολογίων στη μείωση του βάρους, εργάσθηκε 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 Α, ενώ στα υπόλοιπα 5 έδωσε το άλλο διαιτολόγιο, έστω Β. Στον Πίνακα 12.6.2 φαίνεται για καθένα από τα δέκα άτομα που επελέγησαν, ο κωδικός, το φύλο, η ηλικία και ο δείκτης μάζας σώματος (BMI)*[13].

---

[13] *Ο Δείκτης Μάζας Σώματος (Body Mass Index) ορίζεται ως εξής:* $BMI = \beta\acute{\alpha}\rho o\varsigma / (\acute{\upsilon}\psi o\varsigma)^2$ *σε* Kg/m^2 *. Αν για ένα άτομο η τιμή του BMI είναι κάτω από 18.5 τότε βρίσκεται στην ομάδα των ατόμων με* ***βάρος κάτω του φυσιολογικού****, 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*

Κωδικός Ατόμου	**Φύλο**	**Ηλικία *(έτη)***	**BMI**
1	Α	63	24.30
2	Γ	40	22.62
3	Γ	22	28.98
4	Α	61	29.44
5	Α	58	28.35
6	Γ	23	26.44
7	Γ	19	27.30
8	Α	37	22.48
9	Α	60	26.14
10	Γ	22	27.73

Πίνακας 12.6.2
Ο κωδικός, το φύλο, η ηλικία και ο δείκτης μάζας σώματος καθενός από 10 τυχαία επιλεγμένα άτομα

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

Απάντηση: 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 έτσι, κάνοντας τις παραδοχές ότι προέρχονται από κανονικούς πληθυσμούς με ίσες διακυμάνσεις, μπορούμε να εφαρμόσουμε το *t-test για δύο ανεξάρτητα δείγματα* και να χρησιμοποιήσουμε ως *περιοχή απόρριψης* την[14]

$$t = \frac{\bar{x} - \bar{y} - \delta}{s\sqrt{\dfrac{1}{\nu_A} + \dfrac{1}{\nu_B}}} \geq t_{\nu_A + \nu_B - 2;\alpha}$$

όπου

$$\delta = 0 ,\ \bar{x} = 3.1 ,\ \bar{y} = 1.7 ,\ \nu_A = 5 ,\ \nu_B = 5 ,\ \alpha = 0.05$$

και το s υπολογίζεται από τον τύπο

$$s^2 = \frac{(\nu_A - 1)s_A^2 + (\nu_B - 1)s_B^2}{\nu_A + \nu_B - 2}$$

---

*μεταξύ 30 και 40 βρίσκεται στην ομάδα των **παχύσαρκων** ατόμων και αν είναι πάνω από 40 βρίσκεται στην ομάδα των **σοβαρά παχύσαρκων**.*

[14] *Υπενθυμίζουμε και πάλι ότι η στατιστική συμπερασματολογία προσφέρει μεθόδους και εργαλεία για τον έλεγχο αυτών των παραδοχών καθώς και εναλλακτικούς ελέγχους στην περίπτωση που αυτές οι παραδοχές δεν ευσταθούν.*

όπου

$$s_A = 1.3\,,\ s_B = 1.2\,.$$

Δηλαδή, η *περιοχή απόρριψης* είναι

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{\nu_A} + \frac{1}{\nu_B}}} \geq t_{8;0.05} \text{ ή } t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{\nu_A} + \frac{1}{\nu_B}}} \geq 1.860\,.$$

Υπολογίζουμε την τιμή της εκτιμήτριας της *κοινής διακύμανσης*,

$$s^2 = \frac{(\nu_A - 1)s_A^2 + (\nu_B - 1)s_B^2}{\nu_A + \nu_B - 2} = \frac{4 \cdot 1.3^2 + 4 \cdot 1.2^2}{5 + 5 - 2} = 1.57 \Rightarrow s = 1.25$$

και επομένως η τιμή της *στατιστικής συνάρτησης ελέγχου* είναι

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{\nu_A} + \frac{1}{\nu_B}}} = \frac{3.1 - 1.7}{1.25\sqrt{\frac{1}{5} + \frac{1}{5}}} = \frac{1.4}{0.79} = 1.77\,.$$

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 12.6.3: *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 και το δείκτη μάζας σώματος κάθε ατόμου. Όλα τα στοιχεία που ζήτησε ο καθηγητής φαίνονται στον Πίνακα 12.6.3 .*

Κωδικός Ατόμου	**Φύλο**	**Ηλικία *(έτη)***	**BMI**	**Μείωση Βάρους *(Kg)***
Διαιτολόγιο Α				
2	Γ	40	22.62	1.46
4	Α	61	29.44	3.92
7	Γ	19	27.30	4.74
9	Α	60	26.14	2.28
3	Γ	22	28.98	3.10
Διαιτολόγιο Β				
5	Α	58	28.35	3.22
6	Γ	23	26.44	1.70
1	Α	63	24.30	0.18
10	Γ	22	27.73	2.46
8	Α	37	22.48	0.94

Πίνακας 12.6.3
Τα στοιχεία του Πίνακα 12.6.2 ανά τύπο διαιτολογίου που εφαρμόσθηκε και συμπληρωμένα με τη μείωση του βάρους κάθε ατόμου

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

Ας δούμε τι προβλημάτισε τον καθηγητή 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

Α ($\bar{x} = 3.1kg$) είναι σχεδόν διπλάσια από τη μέση μείωση του βάρους στο δείγμα που δόθηκε το διαιτολόγιο *Β* ($\bar{y} = 1.7kg$). 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,

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\dfrac{1}{\nu_A} + \dfrac{1}{\nu_B}}}.$$

Αν οι *τυπικές αποκλίσεις* s_A και s_B είναι μεγάλες, θα είναι επίσης μεγάλη και η τιμή s^2 της εκτιμήτριας της *κοινής διακύμανσης,* 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 μεταβλητότητας CV* στα δύο δείγματα είναι 41.9% και 70.6% αντίστοιχα[15]). Έτσι, η μεγάλη τιμή του *s*, σε συνδυασμό με τα μικρά μεγέθη των δύο δειγμάτων, «υπονόμευσαν» τη διαφορά *μεταξύ* των δύο δειγμάτων και μάλιστα αρκετά ώστε να μην απορριφθεί η *μηδενική υπόθεση*!

Σημείωση 12.6.3: *Για το πώς αξιοποιείται η σχέση της μεταβλητότητας* ***μεταξύ*** *δύο (ή περισσοτέρων) ανεξάρτητων δειγμάτων με τη μεταβλητότητα* ***εντός*** *των δειγμάτων θα αναφερθούμε και πάλι στα επόμενα όταν θα μιλήσουμε για την Ανάλυση Διακύμανσης.*

Ο καθηγητής ζήτησε να δει τα αναλυτικά στοιχεία που φαίνονται στον *Πίνακα 12.6.3*, 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 BMI, 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 *Α*, το άτομο με κωδικό 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 *Β*, το άτομο με κωδικό 10 είναι *γυναίκα, νέα και υπέρβαρη* ενώ το άτομο

[15] *Υπενθυμίζουμε ότι* $CV = (s/|\bar{x}|)\cdot 100\%$

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 «υπονόμευση» του *t-test με δύο ανεξάρτητα δείγματα* από τη μεγάλη μεταβλητότητα εντός των δύο δειγμάτων.

1η 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 στο εισαγωγικό κεφάλαιο *(1ο Κεφάλαιο)* 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 $\bar{x} = 3.1kg$, $\bar{y} = 1.7kg$, $s_A = 1.3Kg$ και $s_B = 1.2Kg$ δεν προέρχονται από δύο *ανεξάρτητα* δείγματα μεγέθους 5 αλλά από δύο *ανεξάρτητα* δείγματα μεγέθους 10. Στην περίπτωση αυτή, η *περιοχή απόρριψης* είναι

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\dfrac{1}{\nu_A} + \dfrac{1}{\nu_B}}} \geq t_{18;0.05} \text{ ή } t = \frac{\bar{x} - \bar{y}}{s\sqrt{\dfrac{1}{\nu_A} + \dfrac{1}{\nu_B}}} \geq 1.734$$

και η τιμή της *στατιστικής συνάρτησης ελέγχου* είναι

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\dfrac{1}{\nu_A} + \dfrac{1}{\nu_B}}} = \frac{3.1 - 1.7}{1.25\sqrt{\dfrac{1}{10} + \dfrac{1}{10}}} = \frac{1.4}{0.56} = 2.5 .$$

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.

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

12.6.2 Εξαρτημένα δείγματα/Ζευγαρωτές παρατηρήσεις

Όπως ήδη εξηγήσαμε, η μείωση του βάρους (υπό οποιαδήποτε δίαιτα) 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 βάρος 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.

Ασφαλώς 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, μπορούμε 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 ν *ζεύγη* παρατηρήσεων (x_i, y_i), $i = 1,2,\ldots,\nu$, 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 ιδίου 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,* κτλ. 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 απαλείφεται. 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 $X_1, X_2, \ldots, X_\nu$ και $Y_1, Y_2, \ldots, Y_\nu$ με τον τρόπο που περιγράψαμε (τα οποία προφανώς πρέπει να είναι του ιδίου μεγέθους $\nu = \nu_A = \nu_B$) μπορούμε να τα θεωρήσουμε ως ζεύγη

$$(X_1, Y_1),\ (X_2, Y_2), \ldots, (X_\nu, Y_\nu)$$

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

$$D_1, D_2, \ldots, D_\nu$$

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 *διακύμανση* 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

$$T = \frac{(\overline{D} - \delta)\sqrt{\nu}}{S_D}$$

όπου, $\overline{D}$ και S_D^2 ο *μέσος* και η *διακύμανση* του δείγματος των <u>διαφορών</u> 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 $t = \dfrac{(\overline{d} - \delta)\sqrt{\nu}}{s_d} \geq t_{\nu-1;\alpha}$

- έναντι της $H_1 : \mu_D < \delta$, όταν $t = \dfrac{(\bar{d} - \delta)\sqrt{\nu}}{s_d} \leq -t_{\nu-1;\alpha}$
- έναντι της $H_1 : \mu_D \neq \delta$, όταν $| t |= \dfrac{|\bar{d} - \delta|\sqrt{\nu}}{s_d} \geq t_{\nu-1;\alpha/2}$.

Με $\bar{d}$ και s_d^2 συμβολίζουμε, αντίστοιχα, την τιμή της *τυχαίας μεταβλητής* $\overline{D}$ και την τιμή της *τυχαίας μεταβλητής* S_D^2 για συγκεκριμένη πραγματοποίηση του δείγματος των διαφορών.

Σημείωση 12.6.4: *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 12.6.4: *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 Α, και στο άλλο αγροτεμάχιο καλλιέργησε στάρι της άλλης ποικιλίας, έστω Β. Σε ποιο από τα δύο αγροτεμάχια κάθε αγρού καλλιέργησε την ποικιλία Α και σε ποιο την ποικιλία Β το αποφάσισε με τυχαίο τρόπο (π.χ. με βάση το αποτέλεσμα της ρίψης ενός νομίσματος). 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, κτλ.). Στον Πίνακα 12.6.4 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 Α (Kg/στρ.) (x_i)	500	650	490	570	555	545	535	635	625	540
Απόδοση ποικιλίας Β (Kg/στρ.) (y_i)	455	620	455	610	505	495	515	600	600	510

Πίνακας 12.6.4
Η απόδοση δύο ποικιλιών σταριού
σε καθέναν από 10 πειραματικούς αγρούς

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

<u>*Απάντηση:*</u> 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, κτλ. 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 *Kg/στρ.*) της ποικιλίας *A* και Y η απόδοση (σε *Kg/στρ.*) της ποικιλίας *B*. Για κάθε ζεύγος παρατηρήσεων (x_i, y_i), $i = 1,2,\ldots,10$, παίρνουμε τη διαφορά

$$d_i = x_i - y_i .$$

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

$$d_i = x_i - y_i , \; i = 1,2,\ldots,10$$

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 ελέγχου σε *επίπεδο σημαντικότητας* $\alpha = 5\%$ είναι

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

Επειδή για το συγκεκριμένο δείγμα των διαφορών που πήραμε έχουμε

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

και

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

η τιμή της *στατιστικής συνάρτησης ελέγχου* στο συγκεκριμένο δείγμα είναι

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

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 Α* διαφέρει από τη *μέση απόδοση της ποικιλίας Β*. Η πιθανότητα αυτό το συμπέρασμα να είναι λάθος είναι το πολύ 0.05.

■

Παράδειγμα 12.6.5: *Μια ερευνητική ομάδα σχεδίασε ένα πείραμα για 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)*[16]. *Για το σκοπό αυτό, επέλεξε ένα τυχαίο δείγμα 12 ατόμων και για κάθε άτομο μέτρησε την τιμή του δείκτη πριν και τρεις ώρες μετά τη λήψη δύο δισκίων ασπιρίνης (650mg). Οι σχετικές μετρήσεις φαίνονται στον Πίνακα 12.6.5.*

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

Πίνακας 12.6.5
Η τιμή του δείκτη prothrombin time σε 12 άτομα
πριν και (τρεις ώρες) μετά τη λήψη δύο δισκίων ασπιρίνης

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

Απάντηση: Έστω *X* και *Y* η τιμή του δείκτη πριν και τρεις ώρες μετά τη λήψη των δισκίων ασπιρίνης, αντίστοιχα. Οι ερευνητές γνωρίζουν ότι η πήξη του αίματος είναι αποτέλεσμα μιας πολύπλοκης διαδικασίας, γι' αυτό σχεδίασαν το πείραμα έτσι ώστε η σύγκριση των τιμών του δείκτη να γίνει σε ίδιες πειραματικές συνθήκες. Έτσι, δημιούργησαν 12 ζεύγη παρατηρήσεων (x_i, y_i), $i = 1,2,\ldots,12$, 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 τρεις ώρες μετά τη λήψη των δισκίων ασπιρίνης. Πρόκειται, επομένως, για σύγκριση δύο μέσων με *ζευγαρωτές παρατηρήσεις*. Έτσι, υπολογίζουμε τις δειγματικές διαφορές

$$d_i = x_i - y_i\,,\ i = 1,2,\ldots,12$$

και εργαζόμαστε όπως στο προηγούμενο παράδειγμα.

[16] *Yochem, Donald, and Roach, Darrell (1971). Aspirin: Effect on Thrombus Formulation Time and Prothrombin Time of Human Subjects. Angiology, 22, 70-76 (από το Larsen and Marx, 2006).*

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 $d_i = x_i - y_i$, $i = 1,2,\ldots,12$. Έτσι, με την υπόθεση ότι αυτό προέρχεται από *κανονικό* πληθυσμό, η *περιοχή απόρριψης* του έλεγχου σε *επίπεδο σημαντικότητας* $\alpha = 5\%$ είναι

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

Επειδή για το συγκεκριμένο δείγμα των διαφορών έχουμε

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

και

$$s_d^2 = \frac{\sum_{i=1}^{12}(d_i - 0.108)^2}{11} = \frac{(0.3-0.108)^2 + (-0.3-0.108)^2 + \ldots + (-0.2-0.108)^2}{11} = 0.257 ,$$

η τιμή της *στατιστικής συνάρτησης ελέγχου* στο συγκεκριμένο δείγμα είναι

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

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 θέτοντας 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 τις μέσες τιμές των δύο πληθυσμών με ζευγαρωτές παρατηρήσεις,* ***με ποιο κριτήριο πρέπει να δημιουργήσει τα ζεύγη****.*

Πρώτα θα συζητήσουμε το δεύτερο και στη συνέχεια το πρώτο.

Κριτήρια για τη δημιουργία ζευγών παρατηρήσεων

Με βάση όσα ήδη έχουμε αναφέρει, η απάντηση είναι προφανής. Το ζευγάρωμα (συνταίριασμα, αντιστοίχηση) των πειραματικών ή δειγματοληπτικών μονάδων 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 «αντισταθμιστικό» 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 διαιτολογίων και σε άλλα ανάλογα προβλήματα. Για παράδειγμα, συγκρίνουμε τα επίπεδα αιμοσφαιρίνης πριν και μετά από ένα πρόγραμμα σωματικών ασκήσεων. Στο *Παράδειγμα 12.6.5* προηγουμένως, εργασθήκαμε με «*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 *Α* και όταν μετράται με τη μέθοδο *Β*, 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 12.6.4*. 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 γενικώς, αλλά σε κατάλληλα προβλήματα).

■

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

Παράδειγμα 12.6.6 (συνέχεια του Παραδείγματος 12.6.3): *Όπως εξηγήσαμε στα προηγούμενα, ο φοιτητής έπρεπε να έχει κάνει τη σύγκριση της μέσης μείωσης του βάρους που επιτυγχάνεται με τα διαιτολόγια Α και Β, 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 Α και στο άλλο άτομο του ίδιου ζεύγους το διαιτολόγιο Β. Πράγματι αυτό έκανε. Δημιούργησε πέντε τέτοια ζεύγη και τα αποτελέσματα που πήρε φαίνονται στον Πίνακα 12.6.6.*

Ζεύγος (i)	*1*	*2*	*3*	*4*	*5*
Μείωση του βάρους με το διαιτολόγιο Α (Kg) (x_i)	1.48	3.12	4.70	2.30	3.94
Μείωση του βάρους με το διαιτολόγιο Β (Kg) (y_i)	0.96	1.72	2.42	0.20	3.24

Πίνακας 12.6.6
Η μείωση του βάρους 10 ατόμων που συγκροτούν 5 ζεύγη. Το ένα άτομο κάθε ζεύγους ακολούθησε το διαιτολόγιο Α και το άλλο το διαιτολόγιο Β.

Θα κάνουμε τον έλεγχο *t-κατά ζεύγη*. Ο έλεγχος, κατά τα γνωστά, θα γίνει με βάση το *δείγμα των διαφορών*

$$d_i = x_i - y_i,\ i = 1,2,\ldots,5.$$

Ζεύγος (i)	*1*	*2*	*3*	*4*	*5*
Μείωση του βάρους με το διαιτολόγιο Α (Kg) (x_i)	1.48	3.12	4.70	2.30	3.94
Μείωση του βάρους με το διαιτολόγιο Β (Kg) (y_i)	0.96	1.72	2.42	0.20	3.24
Διαφορά *(Kg)* ($d_i = x_i - y_i$)	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 έλεγχου σε *επίπεδο σημαντικότητας* $\alpha = 5\%$ είναι

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

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

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

η τιμή της *στατιστικής συνάρτησης ελέγχου* στο συγκεκριμένο δείγμα είναι,

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

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 0.05.

Αν κάνουμε τον έλεγχο έναντι της *εναλλακτικής* $H_1 : \mu_D > 0$, προφανώς (γιατί;) η *μηδενική υπόθεση*, σε *επίπεδο σημαντικότητας* $\alpha = 5\%$, και πάλι θα απορριφθεί. Δηλαδή, τα πειραματικά δεδομένα υποστηρίζουν ότι το διαιτολόγιο A είναι πιο αποτελεσματικό από το διαιτολόγιο B. Η πιθανότητα αυτό το συμπέρασμα να είναι λάθος είναι το πολύ 0.05.

■

Παρατήρηση 12.6.4: *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* $(\nu_A + \nu_B - 2 = 5 + 5 - 2 = 8)$, *εντούτοις η μηδενική υπόθεση απορρίφθηκε. Αυτό συνέβη γιατί η διακύμανση* $s_d^2 = 0.632$ *του δείγματος των διαφορών είναι πολύ μικρότερη από την κοινή διακύμανση,* $s^2 = 1.57$, *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,*

$$t = \frac{\bar{d}\sqrt{\nu}}{s_d} = \frac{\bar{x} - \bar{y}}{s_d / \sqrt{\nu}}\,.$$

Όσο μικρότερος είναι ο παρανομαστής $s_d / \sqrt{\nu}$ *τόσο αυξάνεται 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 ΙΙ, 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 «αντισταθμίσθηκε».*

■

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 12.6.4, 12.6.5* και *12.6.6* η *σύγκριση κατά ζεύγη*, παρότι γίνεται με τους μισούς *βαθμούς ελευθερίας* $(\nu - 1)$ σε σχέση με τη σύγκριση με *ανεξάρτητα δείγματα* $(\nu_A + \nu_B - 2 = \nu + \nu - 2 = 2(\nu - 1))$ εντούτοις, μπορεί να μειώσει την *πιθανότητα σφάλματος τύπου ΙΙ*, 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. Σκεφθείτε 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 Α και όταν γίνει με τη μέθοδο Β. Ας υποθέσουμε ότι ο ερευνητής δημιουργεί ζευγαρωτές παρατηρήσεις χρησιμοποιώντας ζεύγη όπως αυτά του Παραδείγματος 12.6.6.

Στο πρώτο άτομο κάθε ζεύγους η επέμβαση γίνεται με τη μέθοδο Α και στο δεύτερο άτομο κάθε ζεύγους με τη μέθοδο Β. Σε κάθε ζεύγος, τα χαρακτηριστικά, φύλο, ηλικία και ΒΜΙ είναι όμοια ή σχεδόν όμοια και με βάση αυτά δημιουργήθηκαν τα ζεύγη. Όμως, αυτά τα χαρακτηριστικά δε μας λένε κάτι για την επίδραση στο χαρακτηριστικό 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 το *Σχόλιο 12.6.1* που κάναμε για το συμπέρασμα που διατυπώσαμε στο *Παράδειγμα 12.6.2*, είναι πλέον φανερό ότι αν η σύγκριση της μέσης απόδοσης των δύο ποικιλιών είχε γίνει με ένα πειραματικό σχέδιο όπως αυτό του *Παραδείγματος 12.6.4*, 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 12.6.2* και *12.6.4*, η σύγκριση συνήθως γίνεται με βάση το *σχέδιο τυχαιοποιημένων πλήρων ομάδων (Randomized Complete Block Design)*-ένα πολύ γνωστό *πειραματικό σχέδιο*-ειδική περίπτωση του οποίου είναι η *σύγκριση κατά ζεύγη*.

■

12.7 Στατιστικός έλεγχος υποθέσεων για τη διαφορά δύο διωνυμικών ποσοστών με δύο ανεξάρτητα δείγματα

Έστω $X_1, X_2, \ldots, X_{\nu_1}$ και $Y_1, Y_2, \ldots, Y_{\nu_2}$ δύο *ανεξάρτητα τυχαία δείγματα* (μεγέθους ν_1 και ν_2 αντίστοιχα) που το καθένα προέρχεται από μια κατανομή *Bernoulli* με *παράμετρο (μέση τιμή)* p_1 και p_2 αντίστοιχα.

Έστω επίσης, $x_1, x_2, \ldots, x_{\nu_1}$ μια πραγματοποίηση του δείγματος $X_1, X_2, \ldots, X_{\nu_1}$ με *ποσοστό επιτυχιών* $\hat{p}_1$, $y_1, y_2, \ldots, y_{\nu_2}$ μια πραγματοποίηση του δείγματος $Y_1, Y_2, \ldots, Y_{\nu_2}$ με *ποσοστό επιτυχιών* $\hat{p}_2$ και $\hat{p}$ ο *σταθμισμένος μέσος* των $\hat{p}_1$, $\hat{p}_2$. Έστω δηλαδή,

$$\hat{p}_1 = \frac{\sum_{i=1}^{\nu_1} x_i}{\nu_1},\ \hat{p}_2 = \frac{\sum_{i=1}^{\nu_2} y_i}{\nu_2} \text{ και } \hat{p} = \frac{\nu_1 \hat{p}_1 + \nu_2 \hat{p}_2}{\nu_1 + \nu_2} = \frac{\sum_{i=1}^{\nu_1} x_i + \sum_{i=1}^{\nu_2} y_i}{\nu_1 + \nu_2}.$$

Για μεγάλα μεγέθη δειγμάτων, δηλαδή αν $\nu_i \hat{p}_i \geq 5$ και $\nu_i\left(1-\hat{p}_i\right) \geq 5$, $i=1,2$, εύκολα προκύπτει ότι σε *επίπεδο σημαντικότητας* α, η *μηδενική υπόθεση* $H_0 : p_1 = p_2$ απορρίπτεται

- έναντι της $H_1 : p_1 > p_2$, όταν $\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{\nu_1}+\dfrac{1}{\nu_2}\right)}} \geq z_\alpha$
- έναντι της $H_1 : p_1 < p_2$, όταν $\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{\nu_1}+\dfrac{1}{\nu_2}\right)}} \leq -z_\alpha$
- έναντι της $H_1 : p_1 \neq p_2$, όταν $\dfrac{\left|\hat{p}_1 - \hat{p}_2\right|}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{\nu_1}+\dfrac{1}{\nu_2}\right)}} \geq z_{\alpha/2}$.

Επισημαίνουμε ότι οι παραπάνω έλεγχοι είναι *επιπέδου σημαντικότητας* α **κατά προσέγγιση**.

■

Παράδειγμα 12.7.1: *Στο περιοδικό journal of Biology δημοσιεύθηκαν τα αποτελέσματα μιας έρευνας για το ποσοστό* p_1 *των ψαριών στη Μεσόγειο και το ποσοστό* p_2 *των ψαριών στον Ατλαντικό που έχουν προσβληθεί από παράσιτα. Στη Μεσόγειο, από 588 τυχαία επιλεγμένα ψάρια που εξετάσθηκαν βρέθηκαν μολυσμένα από παράσιτα τα 211 ενώ στον Ατλαντικό, από 123 τυχαία επιλεγμένα ψάρια που εξετάσθηκαν, βρέθηκαν μολυσμένα από παράσιτα τα 26.*

Υπενθυμίζουμε ότι αυτό το παράδειγμα, το χρησιμοποιήσαμε και στο *11^o Κεφάλαιο* για την *εκτίμηση* της *διαφοράς,* $p_1 - p_2$, των δύο *ποσοστών* με ένα *95% διάστημα εμπιστοσύνης (Παράδειγμα 11.6.1).*

Ας κάνουμε τώρα σε *επίπεδο σημαντικότητας* 5%, τον έλεγχο της *μηδενικής υπόθεσης*

$$H_0 : p_1 = p_2$$

έναντι της *εναλλακτικής*

$$H_1 : p_1 \neq p_2 .$$

Τα δύο *δειγματικά ποσοστά* είναι αντίστοιχα

$$\hat{p}_1 = \frac{211}{588} = 0.36 \text{ και } \hat{p}_2 = \frac{26}{123} = 0.21 .$$

Επίσης, ο σταθμισμένος μέσος των $\hat{p}_1$, $\hat{p}_2$ είναι

$$\hat{p} = \frac{211+26}{588+123} = \frac{237}{711} = 0.33$$

και επειδή

$\nu_1 \hat{p}_1 = 211 \geq 5$ **,** $\nu_1(1-\hat{p}_1) = 377 \geq 5$, $\nu_2 \hat{p}_2 = 26 \geq 5$ και $\nu_2(1-\hat{p}_2) = 97 \geq 5$,

η *απορριπτική περιοχή* του ελέγχου σε επίπεδο σημαντικότητας $\alpha = 0.05$, είναι

$$\frac{\left|\hat{p}_1 - \hat{p}_2\right|}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{\nu_1}+\dfrac{1}{\nu_2}\right)}} \geq z_{0.025} \text{ ή } \frac{\left|\hat{p}_1 - \hat{p}_2\right|}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{\nu_1}+\dfrac{1}{\nu_2}\right)}} \geq 1.96$$

και επειδή

$$\frac{|\hat{p}_1-\hat{p}_2|}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{\nu_1}+\frac{1}{\nu_2}\right)}}=\frac{0.15}{\sqrt{0.33\cdot 0.67\left(\frac{1}{588}+\frac{1}{123}\right)}}=3.22\geq 1.96$$

η *μηδενική υπόθεση* σε *επίπεδο σημαντικότητας 5%* απορρίπτεται. Δηλαδή, τα ευρήματα στα δύο δείγματα δίνουν στατιστικά σημαντικές αποδείξεις ότι το *ποσοστό των ψαριών στη Μεσόγειο που έχουν προσβληθεί από παράσιτα* δεν είναι ίδιο με το *ποσοστό των ψαριών στον Ατλαντικό που έχουν προσβληθεί από παράσιτα*. Η πιθανότητα το συμπέρασμα αυτό να είναι λάθος, είναι το πολύ 0.05.

Μια *εκτίμηση* του πόσο διαφέρουν τα δύο αυτά ποσοστά δώσαμε με το *95% διάστημα εμπιστοσύνης,* $[0.07, 0.23]$, που υπολογίσαμε στο *Παράδειγμα 11.6.1.*

Παρατηρείστε ότι το *95% διάστημα εμπιστοσύνης* $[0.07, 0.23]$, δεν περιέχει το μηδέν, κάτι το όποιο ασφαλώς συμφωνεί με το αποτέλεσμα του αμφίπλευρου ελέγχου για την ισότητα των ποσοστών (σε επίπεδο σημαντικότητας *5%*).

■

Ολοκληρώνουμε την ενότητα των *στατιστικών ελέγχων*, με τον έλεγχο των διακυμάνσεων δύο πληθυσμών.

12.8 Στατιστικός έλεγχος υποθέσεων για την ισότητα των διακυμάνσεων δύο κανονικών πληθυσμών

Όταν στα προηγούμενα μιλήσαμε για τον έλεγχο με δύο *ανεξάρτητα τυχαία δείγματα* των *μέσων τιμών* δύο *κανονικών* πληθυσμών με άγνωστες *διακυμάνσεις*, είδαμε ότι όταν τα δείγματα δεν είναι μεγάλα, ο προτεινόμενος *στατιστικός έλεγχος* μπορεί να εφαρμοσθεί μόνο αν οι άγνωστες *διακυμάνσεις* των πληθυσμών είναι ίσες. Την ίδια παραδοχή/υπόθεση, όπως είδαμε, απαιτείται να κάνουμε και για την κατασκευή *διαστήματος εμπιστοσύνης* για τη διαφορά των *μέσων τιμών* δύο πληθυσμών στην αντίστοιχη περίπτωση *(κανονικοί πληθυσμοί με άγνωστες διακυμάνσεις και τα δείγματα μικρά)*. Προκύπτει συνεπώς ανάγκη για τη σύγκριση των *διακυμάνσεων* δύο *κανονικών* πληθυσμών. Βέβαια, όπως εξηγήσαμε και όταν μιλήσαμε για τα *διαστήματα εμπιστοσύνης*, το πρόβλημα της σύγκρισης των *διακυμάνσεων* δύο πληθυσμών δεν προκύπτει μόνο ως ένα δευτερογενές πρόβλημα. Σε πολλές εφαρμογές, η σύγκριση των *διακυμάνσεων* δύο πληθυσμών μας ενδιαφέρει αυτή καθαυτή. Ας δούμε ένα τέτοιο παράδειγμα.

Παράδειγμα 12.8.1: *Δύο αυτόματες μηχανές συσκευασίας, έστω Α και Β, συσκευάζουν λίπασμα σε τσουβάλια των 50kg. Οι ποσότητες λιπάσματος που συσκευάζονται ανά τσουβάλι από κάθε μηχανή, είναι τυχαίες μεταβλητές, έστω Χ και Υ, αντίστοιχα. Μάλιστα, είναι γνωστό ότι οι Χ και Υ είναι κανονικές τυχαίες μεταβλητές. Ο υπεύθυνος παραγωγής, υποψιάζεται ότι η μεταβλητότητα της ποσότητας λιπάσματος που συσκευάζεται ανά τσουβάλι από τη μηχανή Α είναι μεγαλύτερη από τη μεταβλητότητα της ποσότητας λιπάσματος που συσκευάζεται ανά τσουβάλι από τη μηχανή Β και προκειμένου να ελέγξει αν πράγματι αυτό συμβαίνει, επέλεξε τυχαία 6 τσουβάλια από την παραγωγή κάθε μηχανής και τα ζύγισε. Τα δεδομένα που συγκέντρωσε φαίνονται στον Πίνακα 12.8.1. Σε επίπεδο σημαντικότητας 5%, υποστηρίζουν άραγε αυτά τα δεδομένα την υποψία του;*

Ποσότητα λιπάσματος (σε kg) που συσκευάσθηκε από τη μηχανή Α (x_i)	51.8	50.0	50.3	49.9	48.7	48.5
Ποσότητα λιπάσματος (σε kg) που συσκευάσθηκε από τη μηχανή Β (y_i)	49.3	49.2	50	48.8	49.2	49.6

Πίνακα 12.8.1
Ποσότητα λιπάσματος που συσκευάσθηκε σε καθένα από 12 τσουβάλια (6 από τη μηχανή Α και 6 από τη μηχανή Β)

Το γενικότερο πρόβλημα που τίθεται είναι το εξής: έστω δύο *ανεξάρτητα τυχαία δείγματα,* $X_1, X_2, \ldots, X_{\nu_1}$ και $Y_1, Y_2, \ldots, Y_{\nu_2}$ από δύο *κανονικούς* πληθυσμούς με άγνωστες *διακυμάνσεις,* σ_1^2 και σ_2^2, αντίστοιχα. Με ποια *στατιστική συνάρτηση* μπορούμε να κάνουμε τον έλεγχο της *μηδενικής υπόθεσης*

$$H_0 : \sigma_1^2 = \sigma_2^2\,;$$

Στο *10^ο^ Κεφάλαιο* είδαμε ότι αν

$$X_1, X_2, \ldots, X_{\nu_1} \sim N(\mu_1, \sigma_1^2) \text{ και } Y_1, Y_2, \ldots, Y_{\nu_2} \sim N(\mu_2, \sigma_2^2)$$

τότε η *τυχαία μεταβλητή*

$$\frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2}$$

ακολουθεί μια κατανομή *F* με $\nu_1 - 1$ και $\nu_2 - 1$ *βαθμούς ελευθερίας*, δηλαδή

$$\frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F_{\nu_1 - 1; \nu_2 - 1}.$$

Επομένως, με την υπόθεση ότι η *μηδενική υπόθεση* $H_0 : \sigma_1^2 = \sigma_2^2$ είναι αληθής, για τη *στατιστική συνάρτηση* S_1^2 / S_2^2, έχουμε

$$\frac{S_1^2}{S_2^2} \sim F_{\nu_1 - 1; \nu_2 - 1}$$

και επομένως μπορούμε να τη χρησιμοποιήσουμε ως *συνάρτηση ελέγχου.* Έτσι, μπορεί να αποδειχθεί ότι σε *επίπεδο σημαντικότητας* α, απορρίπτουμε τη *μηδενική υπόθεση* $H_0 : \sigma_1^2 = \sigma_2^2$

- έναντι της $H_1 : \sigma_1^2 > \sigma_2^2$, όταν $\dfrac{s_1^2}{s_2^2} \geq F_{\nu_1 - 1; \nu_2 - 1; \alpha}$
- έναντι της $H_1 : \sigma_1^2 < \sigma_2^2$, όταν $\dfrac{s_1^2}{s_2^2} \leq F_{\nu_1 - 1; \nu_2 - 1; 1-\alpha}$
- έναντι της $H_1 : \sigma_1^2 \neq \sigma_2^2$, όταν $\dfrac{s_1^2}{s_2^2} \geq F_{\nu_1 - 1; \nu_2 - 1; \alpha/2}$ ή $\dfrac{s_1^2}{s_2^2} \leq F_{\nu_1 - 1; \nu_2 - 1; 1-\alpha/2}$.

Υπενθυμίζουμε ότι με $F_{\nu_1 - 1; \nu_2 - 1; \alpha}$ και $F_{\nu_1 - 1; \nu_2 - 1; \alpha/2}$ συμβολίζουμε το *άνω* α και το *άνω* $\alpha/2$ *ποσοστιαίο σημείο* της κατανομής $F_{\nu_1 - 1; \nu_2 - 1}$ και με $F_{\nu_1 - 1; \nu_2 - 1; 1-\alpha/2}$ το *άνω* $1 - \alpha/2$ *ποσοστιαίο σημείο* της $F_{\nu_1 - 1; \nu_2 - 1}$. Υπενθυμίζουμε επίσης, ότι $F_{\nu_1 - 1; \nu_2 - 1; 1-\alpha/2} = 1 / F_{\nu_2 - 1; \nu_1 - 1; \alpha/2}$.

Μπορούμε πλέον να απαντήσουμε στο ερώτημα του *Παραδείγματος 12.8.1.*

Αν σ_1^2 *η διακύμανση των ποσοτήτων λιπάσματος που συσκευάζονται ανά τσουβάλι από τη μηχανή A* και σ_2^2 *η διακύμανση των ποσοτήτων λιπάσματος που συσκευάζονται ανά τσουβάλι από τη μηχανή B,* για να απαντήσουμε στο ερώτημα που τίθεται, πρέπει να κάνουμε σε *επίπεδο σημαντικότητας 5%,* τον έλεγχο της *μηδενικής υπόθεσης*

$$H_0 : \sigma_1^2 = \sigma_2^2$$

έναντι της *εναλλακτικής*

$$H_1 : \sigma_1^2 > \sigma_2^2 .$$

Επειδή τα δύο *τυχαία δείγματα* προέρχονται από *κανονικούς* πληθυσμούς και έχουν ληφθεί *ανεξάρτητα* το ένα από το άλλο, η *απορριπτική περιοχή* του ελέγχου, σε *επίπεδο σημαντικότητας* 5%, είναι

$$\frac{s_1^2}{s_2^2} \geq F_{5;5;0.05} \text{ ή } \frac{s_1^2}{s_2^2} \geq 5.05 .$$

Από τα δεδομένα εύκολα βρίσκουμε

$$\bar{x} = 49.87 ,\ \bar{y} = 49.35$$

και

$$s_1^2 = \frac{\sum_{i=1}^{6}(x_i - 49.87)^2}{5} = 1.43 ,\ s_2^2 = \frac{\sum_{i=1}^{6}(y_i - 49.35)^2}{5} = 0.17$$

και επειδή

$$\frac{s_1^2}{s_2^2} = \frac{1.43}{0.17} = 8.41 \geq 5.05$$

η *μηδενική υπόθεση*, σε *επίπεδο σημαντικότητας 5%,* απορρίπτεται. Δηλαδή, τα ευρήματα στα δύο δείγματα δίνουν στατιστικά σημαντικές αποδείξεις ότι η υποψία του υπεύθυνου παραγωγής είναι βάσιμη. Η πιθανότητα το συμπέρασμα αυτό να είναι λάθος, είναι το πολύ 0.05.

■

Ερωτήσεις κατανόησης

1. *Αν σε ένα στατιστικό έλεγχο υποθέσεων η μηδενική υπόθεση για επίπεδο σημαντικότητας 5% απορρίπτεται, τότε α) απορρίπτεται για οποιοδήποτε επίπεδο σημαντικότητας β) για οποιοδήποτε άλλο επίπεδο* σημαντικότητας δεν απορρίπτεται *γ) για επίπεδο σημαντικότητας 2% απορρίπτεται δ) για επίπεδο σημαντικότητας 2% δεν απορρίπτεται ε) για επίπεδο σημαντικότητας 2% μπορεί να απορρίπτεται ή να μην απορρίπτεται. Ποια (ή ποιες) από τις πέντε εκδοχές είναι σωστή (σωστές);*
2. *Η περιοχή μη απόρριψης της H_0 για επίπεδο σημαντικότητας α ορίζεται από τις τιμές της στατιστικής συνάρτησης ελέγχου οι οποίες, όταν είναι αληθής η H_0, έχουν πιθανότητα $1-\alpha$ να παρατηρηθούν/εμφανισθούν. Τι λέτε, αυτό είναι σωστό ή λάθος;*
3. *Σε ένα στατιστικό έλεγχο υποθέσεων, για να υπολογίσουμε την P-τιμή είναι αναγκαίο να έχουμε ορίσει το επίπεδο σημαντικότητας του ελέγχου. Αυτό είναι σωστό ή λάθος;*
4. *Σε ένα στατιστικό έλεγχο υποθέσεων για το ποσοστό (αναλογία) p σε έναν πληθυσμό, βρέθηκε $P - \tau\iota\mu\acute{\eta} = 0.038$. Αν για τον έλεγχο έχουμε επιλέξει επίπεδο*

σημαντικότητας 5%, η μηδενική υπόθεση απορρίπτεται ή όχι; Αν έχουμε επιλέξει 1% ;

5. *Από 60 φοιτητές ζητήθηκε να εκτελέσουν το πείραμα που περιγράφεται στο Παράδειγμα 12.6.5, δηλαδή από κάθε φοιτητή ζητήθηκε να επιλέξει ένα τυχαίο δείγμα 12 ατόμων, για κάθε άτομο να μετρήσει την τιμή του δείκτη prothrombin time πριν και (τρεις ώρες) μετά τη λήψη δύο δισκίων ασπιρίνης (650mg) και με βάση τα δεδομένα που θα συγκεντρώσει να ελέγξει σε επίπεδο σημαντικότητας 0.05 αν 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. Αν είναι γνωστό (δηλαδή αν έχει τεκμηριωθεί) ότι η λήψη δύο δισκίων ασπιρίνης δεν επηρεάζει την τιμή του δείκτη prothrombin time και αν το πείραμα και η απαιτούμενη στατιστική ανάλυση των πειραματικών δεδομένων έγιναν από όλους τους φοιτητές σωστά, μπορεί κάποιοι φοιτητές να κατέληξαν σε εσφαλμένο συμπέρασμα; Αν ναι, πόσοι (περίπου) θα ήταν αυτοί οι φοιτητές; Αν όχι, εξηγείστε γιατί όχι.*
6. *Μια γαλλική εισαγωγική εταιρεία νωπών λαχανικών, εισάγει στη Γαλλία νωπά λαχανικά από έναν ελληνικό αγροτικό συνεταιρισμό. Από κάθε παρτίδα νωπών λαχανικών που φθάνει στις εγκαταστάσεις της από το συνεταιρισμό, παίρνει ένα τυχαίο δείγμα και κάνει ποιοτικό έλεγχο. Η εταιρεία επιστρέφει την παρτίδα αν με βάση το δείγμα συμπεράνει ότι το ποσοστό των λαχανικών που είναι εκτός προδιαγραφών υπερβαίνει το 7%, αλλιώς αποδέχεται την παρτίδα. α) Ποια μηδενική υπόθεση και έναντι ποιας εναλλακτικής πρέπει να ελέγχει η εισαγωγική εταιρεία; β) Διατυπώστε με όρους του προβλήματος το σφάλμα τύπου Ι και το σφάλμα τύπου ΙΙ του ελέγχου που κάνει η εταιρεία. γ) Από τη σκοπιά της εταιρείας, ποιο σφάλμα είναι πιο σοβαρό; Αντίστοιχα, από τη σκοπιά του συνεταιρισμού;*
7. *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;*
8. *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 απαραιτήτως να απορρίπτεται και υπέρ της $H_1 : \mu \neq \mu_0$;*

■