

## ΚΕΦΑΛΑΙΟ 22

# ΓΕΝΙΚΗ ΘΕΩΡΗΣΗ ΤΩΝ ΕΛΕΓΧΩΝ ΥΠΟΘΕΣΕΩΝ

## ΓΕΝΙΚΕΥΣΕΙΣ ΚΑΙ ΣΥΓΚΡΙΣΕΙΣ

### Σύνοψη των Απαιτούμενων Ενεργειών στους Ελέγχους Υποθέσεων

Από όσα έχουμε δει μέχρι τώρα, οδηγούμαστε στο συμπέρασμα ότι τα βήματα που χρειάζεται να ακολουθήσει κανείς προκειμένου να προβεί σε έναν έλεγχο μιας στατιστικής υπόθεσης είναι τα εξής:

1. Καθορισμός της μηδενικής υπόθεσης  $H_0$ .
2. Καθορισμός της εναλλακτικής υπόθεσης  $H_1$ .
3. Επιλογή του επιπέδου σημαντικότητας  $\alpha$ .
4. Επιλογή του μεγέθους του δείγματος  $n$ .
5. Καθορισμός της κατάλληλης στατιστικής τεχνικής που θα χρησιμοποιηθεί και της αντίστοιχης στατιστικής συνάρτησης ελέγχου για το πρόβλημα.
6. Προσδιορισμός της κρίσιμης τιμής ή των κρίσιμων τιμών που θα διαχωρίσουν την περιοχή απόρριψης από την περιοχή αποδοχής (περιοχή μη απόρριψης).
7. Συγκέντρωση των δεδομένων και υπολογισμός της τιμής της στατιστικής συνάρτησης ελέγχου για τα συγκεκριμένα δεδομένα.
8. Καθορισμός του κατά πόσο η τιμή της στατιστικής συνάρτησης ελέγχου για το συγκεκριμένο δείγμα βρίσκεται στην περιοχή απόρριψης ή όχι.
9. Λήψη της κατάλληλης στατιστικής απόφασης.
10. Διατύπωση της στατιστικής αυτής απόφασης στην ορολογία του προβλήματος.

## Γενίκευση των Ελέγχων Υποθέσεων

Στους ελέγχους υποθέσεων που εξετάσαμε οι οποίοι αναφέρονταν σε σύγκριση πληθυσμών, χρησιμοποιήθηκε η μηδενική υπόθεση της ισότητας των παραμέτρων που μας ενδιέφεραν μεταξύ των δύο πληθυσμών (είτε οι παράμετροι αυτές ήταν μέσες τιμές, είτε αναλογίες, είτε διακυμάνσεις). Πολλοί συγγραφείς εξετάζουν την γενικότερη περίπτωση όπου η μηδενική υπόθεση αναφέρεται σε μία συγκεκριμένη τιμή της διαφοράς ή του λόγου των δύο παραμέτρων (π.χ.  $H_0: \mu_X - \mu_Y = \mu_D$ ,  $H_0: P_X - P_Y = P_D$ ,  $H_0: \sigma_X^2 / \sigma_Y^2 = \kappa$ ). Οι έλεγχοι αυτοί υποθέσεων γίνονται με τον ίδιο ακριβώς τρόπο όπως αυτόν που αναλύσαμε στις προηγούμενες ενότητες. Το μόνο σημείο στο οποίο διαφέρουν είναι η τιμή της στατιστικής συνάρτησης ελέγχου κάτω από την μηδενική υπόθεση.

Για παράδειγμα, στην περίπτωση της σύγκρισης των μέσων τιμών δύο κανονικών πληθυσμών με γνωστές διακυμάνσεις, η στατιστική συνάρτηση ελέγχου θα είναι η

$$Z_0 = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$$

**Παρατήρηση 1:** Τα κρίσιμα σημεία είναι τα ίδια με εκείνα που χρησιμοποιήσαμε στις προηγούμενες ενότητες. Ο λόγος που δεν παρουσιάσαμε την γενική αυτή περίπτωση είναι γιατί η γενική αυτή περίπτωση σπάνια συναντάται στην πράξη.

## Έλεγχοι Υποθέσεων και Ανάλυση Διακύμανσης

Τόσο στην κατασκευή διαστημάτων εμπιστοσύνης όσο και στους ελέγχους υποθέσεων για τις μέσες τιμές δύο πληθυσμών, είδαμε ότι, ανάλογα με την μορφή του προβλήματος, μπορούμε να επιλέξουμε είτε δείγματα με παρατηρήσεις κατά ζεύγη είτε ανεξάρτητα δείγματα. Όπως είδαμε, η χρησιμοποίηση των παρατηρήσεων κατά ζεύγη ελαχιστοποιεί την διακύμανση (και επομένως το λάθος) που οφείλεται σε άλλους παράγοντες και εξετάζει

κυρίως εκείνην που οφείλεται στο συγκεκριμένο χαρακτηριστικό των δύο πληθυσμών το οποίο μελετάμε.

Το ποια μέθοδος θα επιλεγεί εξαρτάται από τον ερευνητή και αποτελεί μέρος της στατιστικής διαδικασίας που ονομάζεται *σχεδιασμός του πειράματος (design of the experiment)*. Οι σχεδιασμοί πειραμάτων αποτελούν ιδιαίτερο πεδίο μελέτης στην Στατιστική.

Ο σχεδιασμός που αναφέρεται σε επιλογή δύο δειγμάτων με παρατηρήσεις κατά ζεύγη για τη σύγκριση δύο πληθυσμών είναι ειδική περίπτωση ενός γενικότερου σχεδιασμού που χρησιμοποιείται όταν χρειάζεται να συγκριθούν οι μέσες τιμές ενός αριθμού πληθυσμών με ελαχιστοποίηση της διακύμανσης μεταξύ των ζευγών που χρησιμοποιούνται για σύγκριση. Ο σχεδιασμός αυτός ονομάζεται *σχεδιασμός τυχαιοποιημένων πλαισίων (μπλοκ) (randomized block design)* και εξετάζεται αναλυτικότερα στο μέρος εκείνο της Στατιστικής που ασχολείται με την *ανάλυση διακύμανσης (analysis of variance)*.

### **Η Υπόθεση της Κανονικότητας στους Ελέγχους Υποθέσεων**

Για τους περισσότερους από τους ελέγχους υποθέσεων για παραμέτρους που εξετάσαμε - αλλά και για την κατασκευή διαστημάτων εμπιστοσύνης - είδαμε ότι είναι απαραίτητη η επίκληση της υπόθεσης ότι ο πληθυσμός (ή οι πληθυσμοί) που μελετάμε ακολουθούν την κανονική κατανομή. Αυτό συμβαίνει κυρίως όταν χρησιμοποιούμε την κατανομή  $t$  και τα μεγέθη των δειγμάτων είναι μικρά. Το ίδιο απαραίτητη είναι η υπόθεση της κανονικότητας για την εξαγωγή συμπερασμάτων που αναφέρονται στην διακύμανση (ή στις διακυμάνσεις) πληθυσμών. Δοθέντος ότι η Στατιστική είναι μια επιστήμη που στηρίζεται κυρίως σε πειραματικά δεδομένα, είναι φυσικό ότι η υπόθεση της κανονικότητας δεν είναι δυνατόν να ικανοποιείται απόλυτα. Το πρόβλημα του κατά ποσόν ισχύει η στατιστική συμπερασματολογία, όταν κάποιες υποθέσεις (που αναφέρονται κυρίως στις κατανομές των υπό μελέτη πληθυσμών) παραβιάζονται, αποτελεί μια ιδιαίτερη περιοχή της στατιστικής που ονομάζεται *στατιστική ανθεκτικότητα (statistical robustness)*.

Αποδεικνύεται ότι ο έλεγχος  $t$  είναι *ανθεκτικός* (*robust*) με την έννοια ότι δεν είναι ιδιαίτερα ευαίσθητος στις αποκλίσεις από την κανονικότητα. Εφόσον τα μεγέθη των δειγμάτων δεν είναι εξαιρετικά μικρά, η υπόθεση της κανονικότητας μπορεί να παραβιασθεί χωρίς να επηρεασθεί σημαντικά η ισχύς του ελέγχου. Αντίθετα, οι κατανομές  $X^2$  και  $F$  είναι ιδιαίτερα ευαίσθητες στις αποκλίσεις από την υπόθεση της κανονικότητας. Έτσι, ακόμη και μικρές αποκλίσεις από την υπόθεση της κανονικότητας επηρεάζουν σοβαρά την συμπερασματολογία που βασίζεται στις κατανομές  $X^2$  και  $F$ . Ειδικότερα, ο έλεγχος της ισότητας των διακυμάνσεων δύο πληθυσμών είναι εξαιρετικά ευαίσθητος σε αποκλίσεις από την υπόθεση της κανονικότητας, ιδιαίτερα μάλιστα όταν τα μεγέθη των δύο δειγμάτων δεν είναι ίσα. (Για τον έλεγχο της υπόθεσης της κανονικότητας υπάρχουν διάφορες μέθοδοι μερικές από τις οποίες συναντήσαμε ήδη στις προηγούμενες ενότητες).

Και για τις περιπτώσεις όμως που η υπόθεση της κανονικότητας δεν ικανοποιείται, η στατιστική επιστήμη έχει βρει λύσεις με εναλλακτικές μεθόδους. Έτσι, όταν η υπόθεση της κανονικότητας παραβιάζεται ή ακόμη και όταν δεν υπάρχουν γνώσεις για την κατανομή του πληθυσμού, χρησιμοποιούνται οι λεγόμενες *μη παραμετρικές μέθοδοι* (*non parametric procedures*). Οι μέθοδοι αυτές δεν απαιτούν υποθέσεις για την κατανομή του υπό μελέτη πληθυσμού και εξετάζονται αναλυτικά στο μάθημα των μη Παραμετρικών Μεθόδων.

### **Επίπεδο Σημαντικότητας και $p$ -Τιμή**

Στους ελέγχους υποθέσεων, αναφερθήκαμε αναλυτικά στις δύο διαφορετικές προσεγγίσεις που μπορεί να ακολουθήσει ο ερευνητής. Αυτή του καθορισμού του επιπέδου σημαντικότητας  $\alpha$  από την αρχή και, στην συνέχεια, της απόρριψης ή όχι της μηδενικής υπόθεσης ανάλογα με τα στοιχεία του τυχαίου δείγματος (ή δειγμάτων) και στην διαδικασία του καθορισμού της  $p$ -τιμής, δηλαδή του παρατηρούμενου επιπέδου σημαντικότητας, σύμφωνα με την οποία, υπολογίζεται πρώτα η τιμή της στατιστικής συνάρτησης ελέγχου και, στην συνέχεια, το παρατηρούμενο επίπεδο σημαντικότητας. Όπως

εξηγήσαμε, η μέθοδος της  $p$ -τιμής είναι σαφώς προτιμότερη για την καλύτερη εξέταση ενός προβλήματος. Πέρα από αυτό, η υιοθέτηση της προσέγγισης αυτής είναι καλύτερη διότι δίνει στον στατιστικό την δυνατότητα να δώσει στον ενδιαφερόμενο μια πλήρη εικόνα του προβλήματος και να αφήσει σ' εκείνον τη λήψη της τελικής απόφασης. Ιδιαίτερα σήμερα, με την ευρεία χρησιμοποίηση των προσωπικών υπολογιστών και στατιστικών πακέτων, η χρησιμοποίηση της προσέγγισης με την  $p$ -τιμή είναι αναμφίβολα προτιμότερη.

**Σημείωση:** Στην στατιστική ορολογία, η προσέγγιση ελέγχου μέσω της χρήσης του επιπέδου σημαντικότητας ονομάζεται *έλεγχος σημαντικότητας (test of significance)*. Η προσέγγιση αυτή εναποθέτει την απόρριψη ή μη της μηδενικής υπόθεσης στον στατιστικό (αφού έχει προηγουμένως καθορισθεί το επίπεδο σημαντικότητας). Η προσέγγιση με την χρήση της  $p$ -τιμής ονομάζεται *έλεγχος υπόθεσης (hypothesis test)*. Εδώ, ο στατιστικός συνοψίζει τα δεδομένα, αναφέρει τον έλεγχο που χρησιμοποιεί, υπολογίζει την  $p$ -τιμή, αλλά αφήνει την τελική απόφαση για την υπό έλεγχο υπόθεση στον ενδιαφερόμενο.

**Ιστορική Σημείωση:** Πολλοί αναρωτιούνται γιατί στην προσέγγιση του ελέγχου σημαντικότητας χρησιμοποιούνται κυρίως οι τιμές 5% και 1%. Ο λόγος είναι ότι για τις τιμές αυτές έχουν κατασκευασθεί πίνακες που επιτρέπουν την λήψη απόφασης. Ο R. A. Fisher ήταν ο πρώτος που δημοσίευσε τέτοιους πίνακες και τους κατασκεύασε με αυτό τον τρόπο. Δεδομένου ότι υπάρχει περιορισμένος χώρος σε μια σελίδα χαρτιού, οι πίνακες θα έπρεπε να περιορισθούν σε ορισμένες μόνο τιμές του  $\alpha$ . Το 5% και 1% θεωρήθηκαν “καλές” τιμές και από τότε χρησιμοποιούνται παραδοσιακά, χωρίς να υπάρχει ειδικός λόγος. Με την διάδοση των υπολογιστών, τέτοιοι πίνακες είναι πλέον ελάχιστα χρήσιμοι. Το ίδιο συμβαίνει με το 5% και 1% ως τιμές του  $\alpha$  αλλά και με την προσέγγιση του ελέγχου σημαντικότητας. Η προσέγγιση της  $p$ -τιμής έχει πια κυριαρχήσει στην Στατιστική Συμπερασματολογία.

## Επιλογή των Υποθέσεων σε Σχέση με το $\alpha$ και $\beta$

Από τη συζήτηση που έχει προηγηθεί, γίνεται φανερό ότι ένα από τα πιο σημαντικά στάδια στον έλεγχο στατιστικών υποθέσεων είναι ο καθορισμός της μηδενικής και της εναλλακτικής υπόθεσης, όπως επίσης και του επιπέδου σημαντικότητας. Στο παράδειγμα συσκευασίας καφέ είδαμε ότι, ανάλογα με το τι ενδιαφέρεται να αποδείξει κανείς, είναι δυνατόν για το ίδιο πρόβλημα και με τα ίδια στοιχεία να χρησιμοποιήσει διαφορετικές εναλλακτικές υποθέσεις. Έτσι, ο μεν υπεύθυνος του ελέγχου ποιότητας της εταιρείας παρασκευής και συσκευασίας καφέ, ενδιαφέρεται να περιορίσει τις οποιεσδήποτε αποκλίσεις (θετικές ή αρνητικές) από την ποσότητα των 368gr, η δε εταιρεία προστασίας καταναλωτών ενδιαφέρεται να εξασφαλίσει ότι οι καταναλωτές δεν ζημιώνονται με την αγορά του συγκεκριμένου προϊόντος. Το παράδειγμα αυτό, αλλά και η όλη προβληματική που αναπτύχθηκε στον στατιστικό έλεγχο υποθέσεων, οδηγεί στο ερώτημα για το ποιά είναι η λογική με την οποία επιλέγουμε τη μηδενική υπόθεση, την εναλλακτική υπόθεση και αποφασίζουμε για το επίπεδο σημαντικότητας.

Σε πάρα πολλές περιπτώσεις, υπάρχει μια μεγάλη ποικιλία τιμών της παραμέτρου του πληθυσμού που μας ενδιαφέρει να μελετήσουμε (έστω  $\mu$ ) και αντίστοιχη πληθώρα για τις παραμέτρους  $\alpha$  και  $\beta$  που είναι δυνατόν να χρησιμοποιήσουμε. Όπως έχουμε ήδη πει, η συνήθης διαδικασία ελέγχου υποθέσεων είναι να καθορίζουμε το  $\alpha$ . Ως επί το πλείστον, χρησιμοποιείται η τιμή  $\alpha = 0.05$ . Επιλέγουμε δηλαδή κάποια συγκεκριμένη υπόθεση και διαμορφώνουμε έναν έλεγχο έτσι ώστε να υπάρχει μια μικρή μόνο πιθανότητα ότι θα απορρίψουμε την υπόθεση αυτή παρότι στην πραγματικότητα η υπόθεση αυτή είναι σωστή. Δίνεται δηλαδή μικρή προσοχή στις άλλες πιθανές τιμές της παραμέτρου  $\mu$  και στις διάφορες τιμές για το  $\alpha$  και το  $\beta$  που αντιστοιχούν στα διάφορα  $\mu$ . Η διαδικασία αυτή στηρίζεται στην λογική, όπως έχουμε ήδη πει, ότι η λανθασμένη απόρριψη της συγκεκριμένης μηδενικής υπόθεσης θα δημιουργούσε σημαντικά προβλήματα. Μερικοί από τους λόγους που επηρεάζουν την κρίση ή την απόφαση του ερευνητή για την υπόθεση που θα επιλέξει να χρησιμοποιήσει ως μηδενική και για τους λόγους που

δίνεται ιδιαίτερη βαρύτητα στον έλεγχο του λάθους τύπου I (του  $\alpha$ ) αναλύονται στη συνέχεια.

### Επιλογή της Μηδενικής Υπόθεσης

1. Μια από τις υποθέσεις αντιπροσωπεύει την κρατούσα κοινή λογική που προκύπτει από θεωρητικές γνώσεις και προηγούμενες μελέτες. Το βάρος αυτού του γεγονότος υποστηρίζει ότι δεν θα έπρεπε εύκολα να απορρίψουμε την υπόθεση αυτή με βάση μια απλή μελέτη (ένα πείραμα ουσιαστικά). Για τον λόγο αυτό, καθιστούμε την άποψη αυτή, που προφανώς είναι η προτιμητέα, μηδενική υπόθεση ώστε να είμαστε βέβαιοι ότι υπάρχει μικρή πιθανότητα να την απορρίψουμε λανθασμένα.
2. Σε άλλες περιπτώσεις, είναι ενδεχόμενο να μην έχουμε ισχυρούς λόγους να μας κάνουν να υποστηρίξουμε οποιαδήποτε από τις αντιπαρατιθέμενες υποθέσεις. Ξέρουμε όμως ότι θα υπάρχουν σοβαρές συνέπειες αν λανθασμένα απορρίψουμε μια συγκεκριμένη υπόθεση. Και στην περίπτωση αυτή, καθιστούμε την συγκεκριμένη αυτή υπόθεση μηδενική υπόθεση για να είμαστε βέβαιοι ότι θα υπάρξει μια μικρή μόνο πιθανότητα να την απορρίψουμε λανθασμένα. (Το παράδειγμα του δικαστικού συστήματος στο οποίο αναφερθήκαμε είναι χαρακτηριστικό).

Όπως εξηγήσαμε ήδη, η στατιστική απόρριψη μιας υπόθεσης είναι μια ισχυρότερη διατύπωση απ' ό,τι είναι η στατιστική παραδοχή (ουσιαστικά, η μη απόρριψη). Όλες οι τιμές της παραμέτρου που βρίσκονται ουσιαστικά μέσα σε ένα διάστημα εμπιστοσύνης είναι αποδεκτές υποθέσεις. Αποδεχόμενοι μια υπόθεση, ουσιαστικά την τοποθετούμε σε μια λίστα με άλλες παρόμοιες υποθέσεις οι οποίες δεν είναι απορριπτέες. Η περιοχή απόρριψης μιας υπόθεσης όμως είναι ένα αποφασιστικό συμπέρασμα. Για τον λόγο αυτό, οι επιστήμονες ίσως επιλέξουν ως μηδενική υπόθεση μια βοηθητική υπόθεση την οποία ελπίζουν να απορρίψουν. Στην πραγματικότητα, ο όρος μηδενική υπόθεση προέρχεται ακριβώς από την προσέγγιση αυτή. Για να αποδείξουμε ότι υπάρχει κάποια επίδραση προσπαθούμε να απορρίψουμε την υπόθεση ότι δεν υπάρχει επίδραση (μηδενική επίδραση).

Ας δώσουμε μερικά παραδείγματα: Ας υποθέσουμε ότι ένας καθηγητής ενδιαφέρεται να εξετάσει μια καινούργια μέθοδο διδασκαλίας ενός συγκεκριμένου μαθήματος Στατιστικής. Η πρακτική που θα μπορούσε να ακολουθήσει θα μπορούσε να είναι στο τέλος της διδασκαλίας του μαθήματος που το δίδαξε με την καινούργια μέθοδο να δώσει στους φοιτητές του ένα διαγώνισμα. Η μέση απόδοση των φοιτητών στο διαγώνισμα αυτό, θα μπορούσε να χρησιμοποιηθεί για να διερευνηθεί αν η νέα μέθοδος διδασκαλίας είναι αποδοτική. Για ένα τέτοιο πρόβλημα, η συνήθης άποψη είναι ότι μια τέτοια μέθοδος θα έχει λίγο ή καθόλου αποτέλεσμα που θα οδηγούσε για επιλογή ως μηδενικής υπόθεσης αυτής η οποία θα υποστηρίζει ακριβώς αυτό. Δεν είναι βέβαια στην περίπτωση αυτή σαφές τί είναι περισσότερο επικίνδυνο· να συστήσει κανείς μια καινούργια μέθοδο που δεν βελτιώνει σε τίποτα την κρατούσα κατάσταση ή να απορρίψει μια χρήσιμη μέθοδο; Αυτοί οι οποίοι θα ήθελαν να αποδείξουν ότι η μέθοδος είναι αποτελεσματική, είναι μάλλον υποχρεωμένοι να χρησιμοποιήσουν ως μηδενική υπόθεση εκείνη η οποία ισχυρίζεται ότι δεν υπάρχει κάποια ουσιαστική διαφοροποίηση με τη μιά μέθοδο ελπίζοντας να αποδείξουν την αποτελεσματικότητά της με την απόρριψη της μηδενικής υπόθεσης. Υπάρχει επομένως κάποιος λόγος για να θεωρήσει κανείς ακριβώς αυτή ως μηδενική υπόθεση.

Ας εξετάσουμε τώρα ένα πρόβλημα διαφορετικής μορφής. Έστω ότι μία βιομηχανία τροφίμων ενδιαφέρεται να χρησιμοποιήσει μια χρωστική ουσία η οποία κάνει τα φαγητά τα οποία παρασκευάζει περισσότερο γευστικά αλλά ενδέχεται να προκαλεί καρκίνο. Οι συνέπειες μιας λανθασμένης έγκρισης που θα επιτρέπει την χρησιμοποίηση καρκινογόνων προσθέτων είναι πολύ χειρότερη από τις συνέπειες να σταματήσει κανείς λανθασμένα ένα αβλαβές πρόσθετο. Εκτός και εάν υπάρχει βάσιμη επιστημονική γνώση η οποία θα μας πείθει ότι η πρόσθετη αυτή ουσία είναι ασφαλής με σιγουριά, η μηδενική υπόθεση θα πρέπει να είναι ότι η πρόσθετη ουσία είναι καρκινογόνος. Η ουσία αυτή δηλαδή θα πρέπει να θεωρείται επικίνδυνη μέχρις ότου αποδειχθεί το αντίθετο. Παρόμοια λογική ισχύει και στο παράδειγμα για το δικαστικό σύστημα.



Ας δούμε τώρα μια άλλη περίπτωση που τα πράγματα δεν είναι το ίδιο ξεκάθαρα. Ας υποθέσουμε ότι μας ενδιαφέρει να εξετάσουμε αν ένα καινούργιο φάρμακο είναι αποτελεσματικό ή όχι. Ίσως υπάρχουν κάποιες επιστημονικές γνώσεις για το κατά πόσο ένα τέτοιο φάρμακο έχει κάποια αξία. Ένας ερευνητής είναι ενδεχόμενο να επηρεασθεί από το γεγονός ότι υπάρχει το ενδεχόμενο της παρουσίασης στην αγορά κάποιου προϊόντος που γίνεται μόνο για να αποδώσει κέρδη στον παρασκευαστή του χωρίς να έχει κάποια ουσιαστική συνεισφορά. Από το άλλο μέρος, οι συνέπειες του να εμποδίσει κανείς τη χρήση ενός αποτελεσματικού φαρμάκου μπορεί να είναι περισσότερο σοβαρές από τις συνέπειες του να επιτρέψει την κυκλοφορία ενός άχρηστου φαρμάκου. Μια τέτοια λογική οδηγεί στην χρησιμοποίηση ως μηδενικής υπόθεσης εκείνης η οποία υποστηρίζει ότι το φάρμακο είναι αποτελεσματικό. Από το άλλο μέρος, η λογική της βοηθητικής υπόθεσης που αναπτύξαμε προηγουμένως υποστηρίζει την άποψη να αποδειχθεί η αξία του φαρμάκου με το να χρησιμοποιηθεί ως μηδενική υπόθεση εκείνη η οποία θεωρεί ότι το φάρμακο δεν είναι αποτελεσματικό. Εδώ, δηλαδή, έχουμε μία περίπτωση που μπορεί κανείς να καταλήξει στην χρησιμοποίηση δύο αντιθέτων υποθέσεων ως μηδενικών με σοβαρά επιχειρήματα.

### **Επιλογή του Επιπέδου Σημαντικότητας $\alpha$**

Παρόμοιος προβληματισμός με αυτόν που αναπτύχθηκε προηγουμένως υπάρχει στην επιλογή του επιπέδου σημαντικότητας  $\alpha$  για ένα δεδομένο πρόβλημα ελέγχου στατιστικής υπόθεσης. Ποιά είναι δηλαδή τα κριτήρια που μας οδηγούν στην επιλογή μιας συγκεκριμένης τιμής του  $\alpha$  και όχι κάποιας άλλης; Στην πράξη, όπως ήδη έχει λεχθεί, η επιλογή που έχει επικρατήσει είναι αυτή που χρησιμοποιεί ως τιμή του  $\alpha$  την τιμή 0.05. Ο πρώτος επιστήμονας που συνέστησε την χρησιμοποίηση του 0.05 ως επιπέδου σημαντικότητας ήταν ο μεγάλος άγγλος στατιστικός R. A. Fisher που στη δεκαετία του 1920 παρατήρησε: (R. A. Fisher (1926) “*The Arrangement of Field Experiments*” *Journal of the Ministry of Agriculture of Great Britain*, p. 504) “Διευκολύνει να σύρουμε την γραμμή στο σημείο

εκείνο (στο επίπεδο εκείνο) στο οποίο μπορούμε να πούμε: “ή υπάρχει κάτι σημαντικό στην επίδραση που μελετάμε ή έχει συμβεί κατά σύμπτωση κάτι που δεν περιμένει κανείς να συμβεί περισσότερο από μιά στις 20 φορές (δοκιμές)...”. Αν “μια στις 20 φορές” δεν φαίνεται αρκετά υψηλό ποσοστό, μπορούμε, αν κάτι τέτοιο κρίνουμε σκόπιμο, να σύρουμε τη γραμμή στο σημείο που αντιστοιχεί στην περίπτωση “μια σε 50 φορές” (δηλαδή στο σημείο 0.02) ή στην περίπτωση “μια στις 100” (δηλαδή στο σημείο 0.01). Προσωπικά, ο συγγραφέας προτιμά να καθορίσει ένα χαμηλό σημείο σημαντικότητας στο σημείο 5% και να αγνοήσει πλήρως όλα τα αποτελέσματα τα οποία αποτυγχάνουν να φθάσουν το επίπεδο αυτό”.

Παρότι βέβαια το  $\alpha=0.05$  έχει επικρατήσει να χρησιμοποιείται, τις περισσότερες φορές δεν υπάρχει αντικειμενικός λόγος που αυτό συμβαίνει. Στην πραγματικότητα, δεν υπάρχει κάποια συγκεκριμένη τιμή του  $\alpha$  που να μπορεί να θεωρηθεί ως σωστή.

Η λογική των ελέγχων υποθέσεων, όπως προαναφέραμε, είναι ότι επιλέγουμε ως μηδενική υπόθεση κάποια υπόθεση για την οποία θέλουμε να είμαστε ιδιαίτερα προσεκτικοί όσον αφορά την λανθασμένη απόρριψή της. Αν είναι δυνατόν να αποδώσει κανείς τις αποκλίσεις μεταξύ της υπόθεσης και των δεδομένων σε δειγματικό λάθος, τότε αυτή είναι και η εξήγηση που θα δοθεί. Εάν θεωρείται υπερβολικό, με κάποιο κριτήριο, να αποδώσουμε την απόκλιση αυτή σε δειγματικά λάθη, τότε θα απορρίψουμε την μηδενική υπόθεση. Επομένως, η ερώτηση σχετικά με τον καθορισμό του  $\alpha$  στην πραγματικότητα αναφέρεται στο εξής: πόσο δύσκολο θα πρέπει να γίνει το να αποδοθεί σε δειγματικά λάθη η απόκλιση της παρατηρηθείσας τιμής της παραμέτρου από την υποτιθέμενη με την μηδενική υπόθεση τιμή της, ώστε να εγκαταλείψουμε την υποστήριξη της μηδενικής αυτής υπόθεσης; Το 5% φαίνεται λογικό. Είναι αρκετά χαμηλή πιθανότητα, αλλά όχι τόσο χαμηλή, ώστε να μην επιτρέπει ποτέ την εγκατάλειψη της μηδενικής υπόθεσης.

Υπάρχουν βέβαια περιπτώσεις για τις οποίες ο καθορισμός της τιμής του  $\alpha$  χρειάζεται ιδιαίτερη προσοχή. Αν, για παράδειγμα, ελέγχουμε ένα φάρμακο που είναι ενδεχόμενο να αποδειχθεί επικίνδυνο για την υγεία, τότε θα πρέπει να ζητάμε πάρα πολύ

ισχυρές ενδείξεις πριν επιτραπεί η κυκλοφορία του. Ένα αυστηρό κριτήριο που μπορεί να επιβληθεί είναι να ορισθεί η μηδενική υπόθεση ότι το φάρμακο είναι επικίνδυνο για την υγεία και να χρησιμοποιηθεί ως τιμή του  $\alpha$  η τιμή 0.01 ή μια μικρότερη τιμή. Σε περιπτώσεις επίσης που η κρατούσα άποψη είναι πολύ ισχυρή, οι πειστικές ενδείξεις θα πρέπει να είναι ακόμη ισχυρότερες. Αν η μηδενική υπόθεση είναι μία κοινά αποδεκτή αλήθεια τότε το  $\alpha$  θα πρέπει να ορισθεί με κάποια πάρα πολύ μικρή τιμή, ώστε να ελαχιστοποιηθεί το ενδεχόμενο δειγματικού λάθους.

### **Η ΦΙΛΟΣΟΦΙΚΗ ΠΡΟΣΕΓΓΙΣΗ ΤΩΝ ΕΛΕΓΧΩΝ ΥΠΟΘΕΣΕΩΝ**

Όπως είναι γνωστό, στοιχεία από κάποιο δείγμα χρησιμοποιούνται συχνά για να επιβεβαιώσουν ή να απορρίψουν κάποιες θεωρίες. Θα ήταν βέβαια ιδιαίτερα ευχάριστο αν ήταν δυνατόν με στοιχεία που συγκεντρώνονται σε ένα δείγμα να καταλήξει κανείς σε ένα συμπέρασμα, πέρα από κάθε αμφιβολία ότι μια θεωρία είναι η σωστή ή λανθασμένη. Ελάχιστες όμως είναι οι περιπτώσεις που μπορεί να είναι κανείς 100% σίγουρος για κάποιο συγκεκριμένο πρόβλημα. Τα στοιχεία, συνήθως, προέρχονται από μια δειγματοληψία και επομένως υπόκεινται σε λάθη δειγματοληψίας. Ένα διαφορετικό δείγμα είναι ενδεχόμενο να διαφέρει σημαντικά από το προηγούμενο και να οδηγεί σε διαφορετικά συμπεράσματα. Αν προσπαθούμε, για παράδειγμα, να προβλέψουμε το αποτέλεσμα μιας επερχόμενης εκλογικής αναμέτρησης, διαφορετικά δείγματα είναι ενδεχόμενο να οδηγήσουν σε διαφορετικά συμπεράσματα. Δεν είναι επομένως δυνατόν να είμαστε βέβαιοι για το αποτέλεσμα μέχρις ότου πραγματοποιηθεί αυτή καθεαυτή η εκλογική αναμέτρηση και καταμετρηθούν όλες οι ψήφοι.

Υπάρχουν όμως και θεωρίες οι οποίες βρίσκονται σε μία συνεχή κατάσταση αβεβαιότητας επειδή δεν είναι δυνατό να γίνει κάτι ανάλογο με την εκλογική διαδικασία που κάποτε ολοκληρώνεται. Στις περιπτώσεις αυτές, νέα πειράματα και καινούργιες παρατηρήσεις του περιβάλλοντος οδηγούν σε νέα δεδομένα. Αυτό συμβαίνει αρκετά συχνά στις κοινωνικές επιστήμες. Για παράδειγμα, η συνάρτηση

κατανάλωσης του Keynes (*consumption function*) “επιβεβαιώθηκε” την δεκαετία του 40 αλλά θεωρήθηκε ακατάλληλη την δεκαετία του 50. Ακόμη και στις φυσικές επιστήμες, όπου υπάρχει πληθώρα στοιχείων που προέρχονται από εργαστηριακά πειράματα, πολλές θεωρίες παραμένουν σε αβεβαιότητα. Χαρακτηριστικό παράδειγμα είναι οι θεωρίες για την δημιουργία του Σύμπαντος. Όπως παρατηρεί ο καθηγητής Στατιστικής Edward Leamer (*American Economic Review, March 1983, B.36*), “Όλες οι γνώσεις είναι στην πραγματικότητα ανθρώπινες πεποιθήσεις ή καλύτερα ανθρώπινες γνώμες. Αυτό που συχνά συμβαίνει στις φυσικές επιστήμες είναι ότι υπάρχει ένας μεγάλος βαθμός σύγκλισης απόψεων. Όταν αυτό συμβεί, η άποψη την οποία συμερίζεται η πλειοψηφία καταξιώνεται ως αντικειμενικό γεγονός και αυτοί οι οποίοι την αμφισβητούν θεωρούνται “τρελλοί”. Η ιστορία όμως είναι γεμάτη με παραδείγματα από απόψεις που έχασαν το καθεστώς της “κρατούσας άποψης”, με περιπτώσεις όπου “αντικειμενικές αλήθειες” συρρικνώθηκαν στις σκοτεινές γωνιές των κοινωνικών σχέσεων”.

Επί πολλούς αιώνες πριν από τον Κοπέρνικο, οι άνθρωποι πίστευαν ότι ο ήλιος περιστρέφεται γύρω από την γη. Ειδικοί πίστευαν ότι τα βαριά αντικείμενα πέφτουν με μεγαλύτερη ταχύτητα από ό,τι τα ελαφρά μέχρις ότου ο Γαλιλαίος έκανε το γνωστό πείραμα από τον Πύργο της Πίζας. Η φυσική του Newton αντικαταστάθηκε από τη θεωρία του Einstein. Οι επιστήμονες βρίσκουν διαρκώς νέους τρόπους για να ελέγξουν παλιές θεωρίες ή καινούργιους τρόπους για να εξηγήσουν παλιά δεδομένα και έτσι οι αδυναμίες παλαιών θεωριών είναι δυνατό να αποκαλυφθούν. Στην περίπτωση που δεν μπορούμε να είμαστε ποτέ βέβαιοι αν μια θεωρία είναι αληθινή ή όχι, το επόμενο καλύτερο πράγμα είναι να καταλήξουμε σε συμπεράσματα που περιλαμβάνουν την έννοια της πιθανότητας. Μια τέτοια προσέγγιση μπορεί να εκφράζεται με την δήλωση “με βάση τα παρατηρηθέντα δεδομένα είμαστε 90% σίγουροι ότι η θεωρία είναι αληθής”. Αυτοί που υποστηρίζουν την Στατιστική Θεωρία κατά Bayes δεν θα είχαν κανένα ενδοιασμό να πουν κάτι τέτοιο. Αυτοί όμως που υποστηρίζουν τη κλασσική θεωρία, η οποία στηρίζεται στην αντικειμενική ερμηνεία πιθανοτήτων, απορρίπτουν τέτοια συμπεράσματα γιατί ισχυρίζονται ότι μια θεωρία ή είναι σωστή ή δεν

είναι σωστή. Κατ' αυτούς μια θεωρία δεν μπορεί να είναι σωστή 90% των περιπτώσεων με καθορισμό μάλιστα της αλήθειας ή του λάθους που μεταβάλλεται κάθε φορά ανάλογα με κάποιο τυχαίο πείραμα. Η κλασική Στατιστική έχει ακολουθήσει μια άλλη προσέγγιση. Αντί να αναφέρεται σε δεσμευμένη πιθανότητα της μορφής:

$$P(\text{μια θεωρία είναι σωστή} \mid \text{κάποιων στοιχείων}),$$

υπολογίζει την αντίστροφη πιθανότητα

$$P(\text{να έχουν παρατηρηθεί κάποια δεδομένα} \mid \text{η θεωρία είναι σωστή}).$$

Η πρώτη διατύπωση λέει ότι δοθέντων κάποιων στοιχείων που είναι διαθέσιμα, αυτή είναι η πιθανότητα ότι η θεωρία είναι σωστή. Η δεύτερη διατύπωση λέει ότι αν η θεωρία είναι σωστή αυτή είναι η πιθανότητα ότι τα δεδομένα που θα περίμενε κανείς να παρατηρήσει είναι όπως αυτά που στην πραγματικότητα παρατήρησε.

Είναι προφανές ότι οι δύο αυτές διατυπώσεις είναι τελείως διαφορετικές. Η αναγνώριση της διαφοράς αυτής είναι απαραίτητη για να κατανοήσει κανείς το νόημα και τους περιορισμούς της κλασικής προσέγγισης του ελέγχου υποθέσεων. Οι έλεγχοι υποθέσεων είναι στην πραγματικότητα μια προσπάθεια απόδειξης που βασίζεται στην *στατιστική αντίφαση* (*statistical contradiction*). Για οποιαδήποτε συγκεκριμένη θεωρία, οι στατιστικοί μπορούν να συμπεράνουν ότι αν η θεωρία είναι σωστή τότε ένα δείγμα δεδομένων είναι πιθανό να έχει την μιά μορφή και λιγότερο πιθανό να έχει μια άλλη μορφή. Στην συνέχεια, κάνουν το πείραμα και συλλέγουν κάποια στοιχεία. Αν τα στοιχεία αυτά ανήκουν στη κατηγορία εκείνων που θεωρούνται πιθανά, τότε τα δεδομένα είναι συνεπή με την θεωρία και επομένως την επιβεβαιώνουν. Αν, αντίθετα, είναι ελάχιστα πιθανά δοθέντος ότι η θεωρία είναι σωστή, τότε τα δεδομένα αυτά δεν συμβαδίζουν (δεν είναι συνεπή) με την συγκεκριμένη θεωρία και η θεωρία αυτή απορρίπτεται. Η λογική αυτή είναι εκείνο που ο Huxley ονομάζει “μεγάλη τραγωδία της επιστήμης. Την απόρριψη μιας όμορφης υπόθεσης από μια άσχημη πραγματικότητα”.

Μια θεωρία απορρίπτεται αν μπορεί "να αποδειχθεί" στατιστικά ότι τα δεδομένα που παρατηρήθηκαν είναι ελάχιστα πιθανό να συμβούν αν η θεωρία ήταν στην πραγματικότητα σωστή. Με την λογική αυτή, έχουμε μια απόδειξη που στηρίζεται σε στατιστική αντίφαση. Επειδή τα δεδομένα είναι ελάχιστα πιθανό να παρατηρηθούν αν η θεωρία ήταν σωστή, συμπεραίνουμε ότι η θεωρία δεν είναι σωστή. Βεβαίως, δε μπορεί να είμαστε ποτέ 100% βέβαιοι δεδομένου ότι και τα ενδεχόμενα (δεδομένα) που έχουν πολύ μικρή πιθανότητα να συμβούν μπορούν κάλλιστα να συμβούν (λάθη τύπου I και II).

Θα πρέπει επίσης να επισημανθεί ότι προκειμένου μια θεωρία να γίνει αποδεκτή δεν είναι αρκετό το ότι δεν την απορρίπτουμε με βάση τα δεδομένα που έχουμε παρατηρήσει. Αυτό θα ήταν ένα πολύ ασθενές συμπέρασμα δεδομένου ότι υπάρχουν πολλές διαφορετικές θεωρίες που θα μπορούσαν να συμβαδίζουν, να ήσαν συνεπείς, με τα συγκεκριμένα δεδομένα που παρατηρήθηκαν. Για τον λόγο αυτό, όπως είδαμε μέχρι τώρα, αν δεν απορρίψουμε την μηδενική υπόθεση δεν συμπεραίνουμε ότι ισχύει η μηδενική υπόθεση. Το γεγονός δηλαδή ότι δεν έχουμε αρκετές ενδείξεις για να απορρίψουμε μια μηδενική υπόθεση δεν σημαίνει απαραίτητα ότι θα πρέπει να καταλήξουμε στο συμπέρασμα ότι η υπόθεση αυτή είναι σωστή.

## **ΣΧΕΣΗ ΔΙΑΣΤΗΜΑΤΩΝ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΩΝ ΥΠΟΘΕΣΕΩΝ**

Από την ανάλυση που προηγήθηκε, είναι εμφανής η σχέση των μεθόδων που χρησιμοποιήσαμε στην κατασκευή διαστημάτων εμπιστοσύνης για παραμέτρους ενός πληθυσμού και στους ελέγχους υποθέσεων. Στην πραγματικότητα, οι δύο αυτές έννοιες και μεθοδολογίες αποτελούν τις δύο όψεις του αυτού νομίσματος. Οι έλεγχοι υποθέσεων είναι ουσιαστικά μια διαφορετική ερμηνεία και προσέγγιση της εκτίμησης με διαστήματα εμπιστοσύνης.

Η σχέση αυτή είναι ιδιαίτερα εμφανής στους αμφίπλευρους ελέγχους υποθέσεων. Όταν, για παράδειγμα, κατασκευάζουμε ένα 95% διάστημα εμπιστοσύνης για την μέση τιμή ενός πληθυσμού είναι σαν να θεωρούμε το διάστημα αυτό ως την περιοχή αποδοχής της

μηδενικής υπόθεσης. Στην πραγματικότητα, η λογική ισοδυναμία των ελέγχων υποθέσεων και των διαστημάτων εμπιστοσύνης είναι ότι μια υπόθεση απορρίπτεται εάν βρίσκεται έξω από το αντίστοιχο διάστημα εμπιστοσύνης. Ένα διάστημα εμπιστοσύνης για την μέση τιμή μ ενός πληθυσμού, περιλαμβάνει όλες τις δυνατές τιμές που μπορεί να έχει το  $\mu$  σε απόσταση δύο, περίπου, τυπικών αποκλίσεων από τον δειγματικό μέσο (στην περίπτωση του 95% διαστήματος εμπιστοσύνης). Αν ο δειγματικός μέσος απέχει περισσότερο από δύο τυπικές αποκλίσεις (για την ακρίβεια 1.96) από την τιμή η οποία καθορίζεται με την μηδενική υπόθεση, τότε η μηδενική υπόθεση βρίσκεται έξω από το διάστημα εμπιστοσύνης. Επομένως, ένας έλεγχος υποθέσεων μπορεί να γίνει απλώς με την κατασκευή ενός διαστήματος εμπιστοσύνης. Εξάλλου, όταν λέμε ότι έχουμε 95% εμπιστοσύνη (βεβαιότητα) ότι ένα διάστημα εμπιστοσύνης θα περιλαμβάνει την πραγματική τιμή  $\mu$ , είναι σαν να λέμε ότι αν το  $\mu$  ήταν ίσο με μια από τις τιμές που έχουν αποκλεισθεί (που βρίσκονται έξω από το διάστημα εμπιστοσύνης) δεν θα υπήρχε πιθανότητα μεγαλύτερη από 5% ότι ο δειγματικός μέσος θα βρισκόταν τόσο μακριά από την τιμή αυτή. Η μικρή πιθανότητα της αναντιστοιχίας αυτής είναι εκείνη που οδηγεί στον αποκλεισμό των τιμών που βρίσκονται έξω από το διάστημα εμπιστοσύνης.

Η προηγούμενη ανάλυση εξηγεί και τους λόγους για τους οποίους όταν δεν απορρίπτεται η μηδενική υπόθεση, στην περίπτωση που η υπόθεση αυτή είναι απλή, αποφεύγουμε να κάνουμε την δήλωση ότι δεχόμαστε τη μηδενική υπόθεση και λέμε ότι δεν έχουμε αρκετές ενδείξεις να απορρίψουμε τη μηδενική υπόθεση. Αυτό γιατί, από όσα είπαμε προηγουμένως, συνάγεται ότι η μη απόρριψη της μηδενικής υπόθεσης σημαίνει ότι η τιμή της παραμέτρου την οποία υποθέτουμε με τη μηδενική υπόθεση βρίσκεται μέσα στο αντίστοιχο διάστημα εμπιστοσύνης (δηλαδή στην περιοχή αποδοχής). Επομένως, η τιμή αυτή της παραμέτρου (η τιμή της παραμέτρου κάτω από τη μηδενική υπόθεση) είναι μία από τις άπειρες τιμές που είναι αποδεκτές για την παράμετρο και που βρίσκονται μέσα στο διάστημα εμπιστοσύνης. Προκειμένου ο ερευνητής να συμπεράνει ότι μη απόρριψη της μηδενικής υπόθεσης σημαίνει και αποδοχή της θα

πρέπει να έχει υπολογίσει την πιθανότητα λάθους δευτέρου είδους, δηλαδή το  $\beta$ . (Την πιθανότητα να αποδεχθεί την  $H_0$  παρότι η  $H_0$  δεν είναι σωστή). Για τον λόγο αυτό, πριν προχωρήσει κανείς σε αποδοχή της μηδενικής υπόθεσης, όταν η υπόθεση αυτή είναι απλή, θα πρέπει να έχει μελετήσει την ισχύ του ελέγχου. Αυτό δεν είναι πάντοτε εύκολο και αυτός είναι ο λόγος που αποφεύγουμε να αποδεχόμαστε την μηδενική υπόθεση σε περίπτωση απλών υποθέσεων όταν δεν απορρίπτεται η  $H_0$ , αλλά λέμε ότι δεν έχουμε αρκετές ενδείξεις για απόρριψή της. Η μελέτη της ισχύος των διαφόρων ελέγχων υποθέσεων γίνεται σε επόμενο κεφάλαιο.

Αντίθετα, όταν οι ενδείξεις του δείγματος οδηγούν σε απόρριψη της μηδενικής υπόθεσης, αυτό μπορούμε να το κάνουμε χωρίς προβληματισμούς δοθέντος ότι γνωρίζουμε εκ των προτέρων το  $\alpha$  (την πιθανότητα δηλαδή να απορρίψουμε την  $H_0$  ενώ στην παραγματικότητα αυτή ισχύει).

Η καλύτερη προσέγγιση στην διαδικασία ελέγχου υποθέσεων, όταν καταλήγουμε στο συμπέρασμα ότι η μηδενική υπόθεση δεν πρέπει να απορριφθεί, είναι να λέμε ακριβώς αυτό και στην συνέχεια να προχωρούμε σε εκτίμηση της παραμέτρου χρησιμοποιώντας ένα διάστημα εμπιστοσύνης. Το διάστημα αυτό θα μας δίνει μια σειρά πιθανών τιμών της υπό εξέταση παραμέτρου.

Οι δυσκολίες που προαναφέρθηκαν δεν υπάρχουν βέβαια όταν η μηδενική υπόθεση είναι σύνθετη.

Από τη μορφή των διαστημάτων εμπιστοσύνης, προκύπτει ότι όσο αυξάνει το μέγεθος του δείγματος το αντίστοιχο διάστημα εμπιστοσύνης συρρικνώνεται. Στην ακραία περίπτωση που το δείγμα είναι ολόκληρος ο πληθυσμός, ο δειγματικός μέσος θα συμπίπτει με την μέση τιμή του πληθυσμού και επομένως δεν θα υπάρχει καθόλου δειγματικό λάθος. Στην περίπτωση αυτή, οποιαδήποτε απλή μηδενική υπόθεση θα απορρίπτεται εκτός βέβαια από εκείνη που θα αναφέρεται στην ακριβή τιμή της μέσης τιμής του πληθυσμού. Όταν όμως το δείγμα υπολείπεται, έστω και λίγο, του πληθυσμού, η πιθανότητα να διαλέξει κανείς την ακριβή τιμή της παραμέτρου εκ των προτέρων είναι μηδενική. Από την σχέση διαστημάτων



εμπιστοσύνης και ελέγχων υποθέσεων προκύπτει ότι όσο αυξάνει το μέγεθος του δείγματος, όσο δηλαδή συρρικνώνεται το διάστημα εμπιστοσύνης, τόσο πιθανότερη γίνεται η απόρριψη μιας απλής μηδενικής υπόθεσης. Επομένως, οποιαδήποτε μηδενική υποθέση και αν χρησιμοποιήσουμε μπορούμε να την απορρίψουμε συγκεντρώνοντας απλώς αρκετά δεδομένα! Θα μπορούσε έτσι να πει κανείς με μια ακραία θέση ότι απλές υποθέσεις που δεν απορρίπτονται είναι μόνο αυτές για τις οποίες δεν έχουμε αρκετά δεδομένα για να τις απορρίψουμε!

Οι έλεγχοι υποθέσεων πολλές φορές, αν δεν αξιολογηθούν σωστά, είναι δυνατόν να είναι παραπλανητικοί. Αν για μια στατιστική υπόθεση, μετά τη συλλογή των δεδομένων, αποφασίσουμε να την απορρίψουμε, έχουμε μια μερική μόνο εικόνα των συμπερασμάτων που μπορεί να προκύψουν από τα δεδομένα. Για τον λόγο αυτό, όπως προναφέραμε, ένα διάστημα εμπιστοσύνης πολλές φορές δίνει καλύτερη εικόνα των δεδομένων από ό,τι ένας έλεγχος υποθέσεων. Αυτό γιατί, εκτός των άλλων, ένα διάστημα εμπιστοσύνης παρέχει το περιθώριο δειγματικού λάθους του οποίου η πρακτική σημασία μπορεί να αποτιμηθεί. Τέλος, ένα διάστημα εμπιστοσύνης δείχνει ποιές ακριβώς από τις ενδεχόμενες τιμές της υπό μελέτη παραμέτρου απορρίπτονται από τα δεδομένα και δίνει την δυνατότητα στον ερευνητή να εκτιμήσει αν κάποιες συγκεκριμένες από τις ενδεχόμενες τιμές διαφέρουν από την εκτίμηση τόσο πρακτικά όσο και με στατιστική έννοια.

Παρά τα μειονεκτήματα που αναφέραμε, οι έλεγχοι υποθέσεων αποτελούν τον συνήθη τρόπο εξαγωγής συμπερασμάτων ίσως γιατί, εμφανίζονται, ως μέθοδος, αποφασιστικοί. Αν όμως δεν ερμηνευθούν σωστά, μπορούν να οδηγήσουν σε εσφαλμένα συμπεράσματα.

Παραδείγματα κακής χρήσης της μεθόδου ελέγχου υποθέσεων υπάρχουν πολλά. Τέτοια παραδείγματα εμφανίζονται συχνά στις διαφημιστικές εκστρατείες για προώθηση συγκεκριμένων προϊόντων, στις οποίες προβάλλεται ο ισχυρισμός ότι το διαφημιζόμενο προϊόν έχει αποδειχθεί στατιστικά καλύτερο από άλλα παρόμοια προϊόντα. Εκείνο που στην πράξη πολλές φορές συμβαίνει είναι ότι σε μια

τυχαία δειγματοληψία, το διαφημιζόμενο προϊόν, ενδεχομένως, παρουσίασε καλύτερα αποτελέσματα όχι όμως σε διαφορά που να μπορεί να θεωρηθεί στατιστικά σημαντική.

Ένα άλλο χαρακτηριστικό παράδειγμα κακής ερμηνείας και εξαγωγής στατιστικών συμπερασμάτων αποτελεί μία έρευνα που έγινε από δύο οικονομολόγους που μελετούσαν την επίδραση του πληθωρισμού στα εκλογικά αποτελέσματα των Ηνωμένων Πολιτειών. (Arcelus, F. & Metlzer, A.H. “The effect of Aggregate Economic Variables on Congressional Elections”, *American Political Science Review*, 69 (1975), 1232-1239). Οι συγκεκριμένοι ερευνητές εκτίμησαν ότι, σε μια εκλογική αναμέτρηση, το πρόβλημα του πληθωρισμού πρόσθεσε επτά εκατοστιαίες μονάδες στις ψήφους των Ρεπουμπλικάνων. Μια τυπική απόκλιση 5 εκατοστιαίων σημείων οδήγησε σε ένα 95% διάστημα εμπιστοσύνης της τάξεως του  $0.07 \pm 0.1$ . Από αυτό το διάστημα εμπιστοσύνης, κατέληξαν στο συμπέρασμα ότι “στην πραγματικότητα και σε αντίθεση με την κοινά διαδεδομένη άποψη, ο πληθωρισμός δεν επηρέασε την συμπεριφορά των ψηφοφόρων”. Αυτό βέβαια δεν είναι ακριβώς ό,τι προκύπτει από τα δεδομένα. Το γεγονός ότι το ποσοστό 0% περιλαμβάνεται στο διάστημα εμπιστοσύνης (δηλαδή η υπόθεση  $H_0: P_X - P_Y = 0$  δεν απορρίπτεται με βάση τα στοιχεία) δεν αποδεικνύει ότι η διαφορά 0% είναι η πραγματική τιμή. Το διάστημα εμπιστοσύνης περιλαμβάνει πράγματι το 0%, αλλά περιλαμβάνει επίσης και όλες τις τιμές από το -3% ως το 17%. Η καλύτερη εκτίμηση για το ποσοστό αυτό είναι μια τιμή μέσα στο διάστημα  $0.07 \pm 0.1$  και, βέβαια, 7% είναι περισσότερο από αρκετό για να μεταβάλει ένα εκλογικό αποτέλεσμα προς την μια ή την άλλη κατεύθυνση.