ΚΕΦΑΛΑΙΟ 17

# ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

Στο κεφάλαιο αυτό θα αναφερθούμε σε ένα άλλο πρόβλημα της Στατιστικής που έχει κυρίως (αλλά όχι μόνο) σχέση με τις παραμέτρους ενός πληθυσμού (τις παραμέτρους της κατανομής ενός πληθυσμού).

Μέχρι τώρα, αντιμετωπίσαμε το πρόβλημα της εκτίμησης των παραμέτρων. Εδώ, θα ασχοληθούμε με τον έλεγχο υποθέσεων που αναφέρονται στις παραμέτρους. Αντί δηλαδή να ψάχνουμε για την καλύτερη εκτιμήτρια (είτε σημειακή εκτιμήτρια είτε κάποιο διάστημα εμπιστοσύνης για μια άγνωστη παράμετρο), θα προσπαθήσουμε να βρούμε τρόπους για να αποφασίζουμε για το κατά πόσο μια προκαθορισμένη τιμή μιας παραμέτρου είναι αποδεκτή με βάση κάποιες παρατηρήσεις.

Λογικά, θα μπορούσε κανείς να ισχυρισθεί ότι το πρόβλημα του ελέγχου υποθέσεων προηγείται αυτού που αναφέρεται στην εκτίμηση. Αν, για παράδειγμα, μας ενδιαφέρει να μελετήσουμε την διαφορά μεταξύ των μέσων δύο κανονικών πληθυσμών, ο φυσικός προβληματισμός είναι να εξετάσουμε αν οι παρατηρήσεις δίνουν κάποια ένδειξη ότι υπάρχει κάποια πραγματική διαφορά ανάμεσα στους μέσους. Θα πρέπει με άλλα λόγια να συγκρίνουμε την παρατηρηθείσα διαφορά (όπως αυτή εκφράζεται από τους δειγματικούς μέσους) με την υπόθεση που θα μπορούσε να είχε κάνει κάποιος ότι δεν υπάρχει πραγματικά διαφορά μεταξύ των μέσων, αλλά η όποια παρατηρούμενη, ή παρατηρηθείσα, διαφορά οφείλεται σε αποκλίσεις της τυχαίας δειγματοληψίας. Στην συνέχεια, εάν κατέληγε κανείς στο συμπέρασμα ότι υπάρχει πράγματι διαφορά μεταξύ των μέσων θα είχε έννοια να προχωρήσει στο επόμενο βήμα της εκτίμησης του μεγέθους της διαφοράς μεταξύ των μέσων των δύο πληθυσμών (δηλαδή στο πρόβλημα της εκτιμητικής).

Είναι προφανές ότι το πρόβλημα του ελέγχου υποθέσεων και της εκτιμητικής συνδέονται στενά. Παρ' όλα αυτά, μελετώνται χωριστά για λόγους, κυρίως, παρουσίασης των εννοιών.

## ΣΤΑΤΙΣΤΙΚΕΣ ΥΠΟΘΕΣΕΙΣ

Το είδος των υποθέσεων που θέλουμε να ελέγξουμε στην Στατιστική είναι περισσότερο περιορισμένο από τις γενικές επιστημονικές υποθέσεις. Μια επιστημονική υπόθεση είναι, για παράδειγμα, ότι κάθε σωματίδιο ύλης έλκει κάθε άλλο σωματίδιο.

Μια άλλη επιστημονική υπόθεση είναι ότι υπάρχει ζωή στον Άρη. Τέτοιας μορφής υποθέσεις δεν είναι δυνατόν να ελεγχθούν με στατιστικές μεθόδους. Οι στατιστικές υποθέσεις αναφέρονται στον τρόπο συμπεριφοράς τυχαίων μεταβλητών που είναι δυνατόν να παρατηρηθούν. Για να εξηγήσουμε το θέμα περισσότερο τεχνικά, έστω ότι έχουμε ένα σύνολο τυχαίων μεταβλητών $X_1, X_2, ..., X_n$ . Οι τυχαίες αυτές μεταβλητές μπορούν να θεωρηθούν ως συνιστώσες ενός διανύσματος $\mathbf{X}$, στον δειγματικό χώρο $\mathbf{n}$ διαστάσεων όπου κάθε ένας από τους άξονες του χώρου αυτού αντιστοιχεί σε μία από τις μεταβλητές. Δεδομένου ότι $\mathbf{X}$ είναι τυχαίο διάνυσμα, υπάρχει μία κατανομή πιθανότητας που αντιστοιχεί σ' αυτήν. Αν επομένως διαλέξουμε κάποια περιοχή w στον δειγματικό χώρο W, είναι δυνατόν (τουλάχιστον κατ' αρχήν) να υπολογίσουμε την πιθανότητα ότι το σημείο $\mathbf{X}$ του δειγματικού χώρου W είναι μέσα στην περιοχή w. Μπορούμε δηλαδή να μιλήσουμε για $P(\mathbf{X}\in W)$. Θα λέμε ότι οποιαδήποτε υπόθεση αναφέρεται στην πιθανότητα $P(\mathbf{X}\in W)$ είναι μια *στατιστική υπόθεση*. Με άλλα λόγια, *οποιαδήποτε υπόθεση αναφέρεται στην συμπεριφορά τυχαίων μεταβλητών για τις οποίες μπορούμε να έχουμε παρατηρήσεις είναι μια στατιστική υπόθεση.*

Για παράδειγμα:

α) Η υπόθεση ότι η κανονική κατανομή έχει μια συγκεκριμένη μέση τιμή και μια συγκεκριμένη διασπορά είναι στατιστική.

β) Η υπόθεση ότι μια κανονική κατανομή έχει μια δεδομένη μέση τιμή αλλά άγνωστη διασπορά είναι επίσης στατιστική.

γ) Η υπόθεση ότι η κατανομή είναι κάποιας κανονικής μορφής αλλά με άγνωστη μέση τιμή και άγνωστη διασπορά είναι επίσης στατιστική.

δ) Η υπόθεση ότι δύο άγνωστες κατανομές ταυτίζονται είναι επίσης στατιστική.

Καθένα από τα προηγηθέντα τέσσερα παραδείγματα συνεπάγεται (αναφέρεται σε) μια συγκεκριμένη συμπεριφορά των τυχαίων μεταβλητών στο δειγματικό χώρο η οποία είναι δυνατόν να ελεγχθεί με σύγκριση κάποιων παρατηρήσεων.

Όσα προηγήθηκαν μας επιτρέπουν να καταλήξουμε στον εξής ορισμό:

**Ορισμός:** *Στατιστική υπόθεση* είναι ένας ισχυρισμός που αναφέρεται στην κατανομή μιας ή περισσοτέρων τυχαίων μεταβλητών.

Για να ελέγξουμε μια υπόθεση την οποία συνήθως ονομάζουμε *μηδενική υπόθεση* (*null hypothesis*) και την οποία συμβολίζουμε με το $H_0$, χρειαζόμαστε μια *εναλλακτική υπόθεση* (*alternative hypothesis*) σε αντιπαράθεση προς την οποία ελέγχεται η $H_0$. Η εναλλακτική υπόθεση συμβολίζεται συνήθως με $H_1$ ή με $H_A$.

Θα επικεντρώσουμε την προσοχή μας σε ελέγχους υποθέσεων που αναφέρονται σε παραμέτρους πληθυσμών.

## ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ ΓΙΑ ΠΑΡΑΜΕΤΡΟΥΣ ΠΛΗΘΥΣΜΩΝ

Όπως έχουμε ήδη αναφέρει, στατιστική συμπερασματολογία για παραμέτρους πληθυσμών μπορεί να γίνει με δύο, κυρίως, τρόπους. Μπορούμε, όπως έχουμε ήδη περιγράψει, να εκτιμήσουμε τις τιμές των παραμέτρων ή μπορούμε να καταλήξουμε σε αποφάσεις σχετικά με τις παραμέτρους αυτές. Η διαδικασία που συνήθως ακολουθούμε εξαρτάται από το πρακτικό πρόβλημα που οδήγησε στην ανάγκη στατιστικής συμπερασματολογίας.

Για να δώσουμε ένα παράδειγμα, ας υποθέσουμε ότι μια μεγάλη ιδιωτική επιχείρηση ενδιαφέρεται να προσφέρει ένα πρόγραμμα πρόωρης συνταξιοδότησης στους εργαζομένους στην επιχείρηση. Προκειμένου να μελετήσει το πρόβλημα, η επιχείρηση θέλει να γνωρίζει το ποσοστό p των εργαζομένων στην επιχείρηση που είναι διατεθειμένοι να αποδεχθούν τη προσφορά. Σε μια τέτοια περίπτωση, η επιχείρηση θα ήθελε να διερευνήσει τις γνώμες ενός δείγματος από τους εργαζομένους και να εκτιμήσει την τιμή του ποσοστού p με κάποιο προκαθορισμένο λάθος εκτίμησης. Εναλλακτικά, ας υποθέσουμε ότι η ίδια η επιχείρηση έχει καταλήξει σε δύο προγράμματα πρόωρης συνταξιοδότησης και ενδιαφέρεται να δει την απήχηση που έχει το καθένα από αυτά στους εργαζομένους. Είναι προφανές ότι η μέθοδος που θα πρέπει να ακολουθήσει θα είναι να πάρει ένα δείγμα από τους εργαζομένους και από τις γνώμες που θα συγκεντρώσει να καταλήξει στο να επιλέξει το σχέδιο εκείνο που θα οδηγεί στο υψηλότερο ποσοστό αποδοχής από τους εργαζομένους. Στην περίπτωση αυτή δηλαδή, η επιχείρηση χρειάζεται να πάρει μια απόφαση που να αναφέρεται στη διαφορά μεταξύ των ποσοστών αποδοχής των δύο σχεδίων. Είναι φανερό ότι εξίσου ενδιαφέρον για την επιχείρηση είναι να γνωρίζει τους κινδύνους (ρίσκα) που θα αντιμετωπίσει αν πάρει μια λανθασμένη απόφαση και να προσπαθήσει να ελαχιστοποιήσει τους κινδύνους αυτούς.

## ΤΑ ΣΤΟΙΧΕΙΑ ΤΟΥ ΕΛΕΓΧΟΥ ΜΙΑΣ ΥΠΟΘΕΣΗΣ

### **Στατιστική Συνάρτηση Ελέγχου** *(Test Statistic)*

Η λογική που αναφέρεται και περιγράφει ένα στατιστικό έλεγχο μιας υπόθεσης μπορεί να περιγραφεί με ένα παράδειγμα. Ας υποθέσουμε ότι σε μιά έρευνα αγοράς έχουμε λόγους να πιστεύουμε ότι περισσότεροι από το 50% των καταναλωτών προτιμούν το προϊόν ενός κατασκευαστή Α. Για να καθορίσουμε και να αποφασίσουμε αν η υπόθεσή μας είναι σωστή, σχεδιάζουμε ένα πείραμα. Επιλέγουμε τυχαία ένα δείγμα από 100 καταναλωτές και ζητάμε την προτίμησή τους για το συγκεκριμένο προϊόν. Είναι φυσικό, ότι αν 99 από τους 100 καταναλωτές που συμπεριελήφθησαν στο δείγμα μας απαντήσουν ότι προτιμούν το προϊόν του κατασκευαστή Α, θα οδηγηθούμε στο συμπέρασμα ότι περισσότεροι από το 50% των καταναλωτών προτιμούν τον συγκεκριμένο κατασκευαστή. Βέβαια, είναι δυνατό να παρατηρήσουμε στο δείγμα μας 99 από τους 100 καταναλωτές να προτιμούν το προϊόν του κατασκευαστή Α, ενώ στην πραγματικότητα το ποσοστό στο σύνολο των καταναλωτών είναι μικρότερο από το 50%, κάτι τέτοιο όμως είναι εξαιρετικά σπάνιο.

Είναι προφανές ότι η απόφασή μας να απορρίψουμε ή να μην απορρίψουμε τη μηδενική υπόθεση βασίζεται σε πληροφορίες από παρατηρηθείσες τιμές μιας τυχαίας μεταβλητής. Η στατιστική συνάρτηση που χρησιμοποιείται στην διαδικασία της λήψης της απόφασης ονομάζεται *στατιστική συνάρτηση ελέγχου ή ελεγχοσυνάρτηση (test statistic)* και η διαδικασία που ακολουθείται ονομάζεται *έλεγχος της στατιστικής υπόθεσης* (*test of the statistical hypothesis*).

Η μέθοδος που χρησιμοποιήθηκε για να καταλήξουμε σε μια απόφαση στο προηγούμενο παράδειγμα είναι δυνατόν να περιγραφεί μέσα από μια τυπική διαδικασία στατιστικής συμπερασματολογίας στο πλαίσιο των εννοιών που έχουν μέχρι τώρα ορισθεί. Στο συγκεκριμένο μας παράδειγμα θέλουμε να ελέγξουμε την υπόθεση ότι η παράμετρος p μιας διωνυμικής κατανομής ξεπερνά το 0.5. Η υπόθεση αυτή μπορεί να θεωρηθεί ως η εναλλακτική υπόθεση $H_1$. Θα μπορέσουμε να καταλήξουμε στο συμπέρασμα αυτό δείχνοντας ότι μια άλλη υπόθεση, συμπληρωματική αυτής, η υπόθεση p=0.5 είναι λανθασμένη. Η δεύτερη αυτή υπόθεση είναι η μηδενική υπόθεση $H_0$. Έχουμε δηλαδή να ελέγξουμε την υπόθεση

$$H_0 : p = 0.5$$
$$H_1 : p > 0.5$$

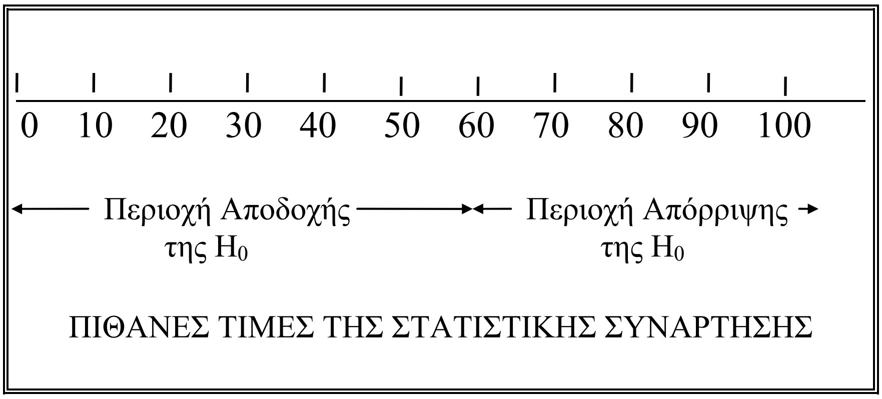
Ο τρόπος που θα καταλήξουμε σε ένα συμπέρασμα, ώστε να ισχυρισθούμε ότι η εναλλακτική υπόθεση ισχύει, είναι να δείξουμε ότι υπάρχουν ενδείξεις ότι η μηδενική υπόθεση $H_0$ δεν είναι δυνατόν να υποστηριχθεί και επομένως πρέπει να απορριφθεί.

Η απόφαση για να απορρίψουμε τη μηδενική υπόθεση και να υποστηρίξουμε την εναλλακτική στηρίζεται σε πληροφορίες που περιέχονται σε ένα δείγμα n μετρήσεων που έχει επιλεγεί από τον πληθυσμό. Στην περίπτωσή μας το δείγμα είναι n=100 παρατηρήσεις από ένα διωνυμικό πληθυσμό. Οι τιμές του δείγματος χρησιμοποιούνται για να υπολογισθεί ένας αριθμός που θα χρησιμοποιηθεί από αυτόν που θα πάρει την απόφαση για να αποφασίσει με τον ένα ή τον άλλο τρόπο. Η στατιστική αυτή συνάρτηση την τιμή της οποίας από ένα συγκεκριμένο δείγμα θα χρησιμοποιήσουμε για να αποφασίσουμε είναι *η στατιστική συνάρτηση ελέγχου (ελεγχοσυνάρτηση).*

*Η στατιστική συνάρτηση ελέγχου χρησιμοποιείται για την μέτρηση της διαφοράς των δεδομένων από αυτό που αναμένεται να συμβαίνει αν η μηδενική υπόθεση είναι ακριβής.*

## Περιοχή Απόρριψης και Κρίσιμο Σημείο

Το σύνολο των τιμών που η στατιστική αυτή συνάρτηση (η ελεγχοσυνάρτηση) μπορεί να πάρει για διαφορετικά δείγματα μπορεί να χωρισθεί σε δύο περιοχές. Μια από αυτές θα αντιστοιχεί στην *περιοχή απόρριψης* (*rejection region*) και η άλλη στην *περιοχή αποδοχής* της μηδενικής υπόθεσης (*acceptance region*). Ένα παράδειγμα τέτοιου χωρισμού του χώρου των τιμών της παραμέτρου που μας ενδιαφέρει για το συγκεκριμένο παράδειγμα που αναπτύξαμε εμφανίζεται στο σχήμα που ακολουθεί:




Εάν η τιμή της στατιστικής συνάρτησης που χρησιμοποιούμε για ένα συγκεκριμένο δείγμα βρίσκεται στην περιοχή απόρριψης, τότε η μηδενική υπόθεση απορρίπτεται και αποφασίζουμε υπέρ της εναλλακτικής υπόθεσης. Εάν, η τιμή της στατιστικής συνάρτησης "πέσει" στην περιοχή αποδοχής, η μηδενική υπόθεση δεν απορρίπτεται (πολλοί συγγραφείς χρησιμοποιούν τον όρο "αποδεχόμαστε την μηδενική υπόθεση" που όμως δεν είναι πάντα δόκιμος).

Η τιμή εκείνη της παραμέτρου η οποία διαχωρίζει την περιοχή αποδοχής από τη περιοχή απόρριψης λέγεται *κρίσιμο σημείο* (*critical point*) και συμβολίζεται με **c**. Στο παράδειγμά μας το κρίσιμο σημείο είναι η τιμή 60. Την τιμή αυτή την επιλέξαμε αυθαίρετα. Στην συνέχεια, θα δούμε τρόπους με τους οποίους καθορίζεται το κρίσιμο σημείο.

Στο πρόβλημά μας ελεγχοσυνάρτηση είναι το ποσοστό των καταναλωτών που προτιμούν το προϊόν του κατασκευαστού Α σε ένα δείγμα n=100. Το δείγμα αυτό προέρχεται από ένα διωνυμικό πληθυσμό. Η περιοχή απόρριψης περιλαμβάνει τιμές του Χ που υποστηρίζουν την εναλλακτική υπόθεση (ότι το ποσοστό p που προτιμά το προϊόν του κατασκευαστή Α είναι μεγαλύτερο από 0.5). Αυτό γιατί πολύ μεγάλες τιμές του Χ θα ήταν μάλλον απίθανο να παρατηρηθούν εάν στην πραγματικότητα το p ήταν 0.5 ή μικρότερο. Τιμές του Χ που υποστηρίζουν την μηδενική υπόθεση είναι αυτές που ανήκουν στην περιοχή αποδοχής. Όταν στην αρχή του παραδείγματος, αναφέραμε ότι με βάση το ποσοστό στο δείγμα των καταναλωτών που προτιμούσαν το προϊόν του κατασκευαστή Α (99 στους 100) θα μπορούσαμε να οδηγηθούμε στην υποστήριξη της εναλλακτικής υπόθεσης, το κάναμε για τον εξής λόγο:

Η τιμή Χ=99 θα ήταν εξαιρετικά απίθανο να παρατηρηθεί αν στη πραγματικότητα το 50% (ή και λιγότεροι) από όλους τους καταναλωτές προτιμούσαν το προϊόν του κατασκευαστή Α. Έτσι αυτόματα θεωρήσαμε την τιμή Χ=99 να ανήκει σ' αυτό που διαισθητικά είχαμε αποφασίσει να είναι η περιοχή απόρριψης της μηδενικής υπόθεσης.

## Τα Πιθανά Λάθη Αποφάσεων στους Ελέγχους Υποθέσεων

Ένα ερώτημα που προκύπτει αμέσως είναι ο καθορισμός του τρόπου με τον οποίο αποφασίζουμε αν κάποια ενδεχόμενη τιμή της ελεγχοσυνάρτησης θα πρέπει να τοποθετηθεί στην περιοχή απόρριψης ή στην περιοχή αποδοχής. Για παράδειγμα, αν στο πρόβλημά μας είχαμε παρατηρήσει 70 καταναλωτές να υποστηρίζουν το προϊόν του κατασκευαστή Α, θα τοποθετούσαμε την τιμή αυτή στην περιοχή απόρριψης ή στην περιοχή αποδοχής; Η απάντηση στο ερώτημα αυτό εξαρτάται από τους κινδύνους (ρίσκα) που είμαστε διατεθειμένοι να πάρουμε αν καταλήξουμε σε μια λάθος απόφαση. Λάθος αποφάσεις μπορεί να ληφθούν αν απορρίψουμε την μηδενική υπόθεση, ενώ αυτή ισχύει στην πραγματικότητα ή αν απορρίψουμε την εναλλακτική υπόθεση, ενώ στην πραγματικότητα η εναλλακτική υπόθεση είναι σωστή.

Τα λάθη αυτά που μπορούν να γίνουν ονομάζονται αντίστοιχα *λάθος τύπου I* και *λάθος τύπου II* (*type I error και type II error*) όσον αφορά τον στατιστικό έλεγχο.

Οι αποφάσεις που μπορεί να πάρει κανείς σε ένα στατιστικό έλεγχο υποθέσεων και τα λάθη στα οποία είναι ενδεχόμενο να υποπέσει εμφανίζονται παραστατικά στον πίνακα που ακολουθεί,

| | | ΠΡΑΓΜΑΤΙΚΗ ΚΑΤΑΣΤΑΣΗ | |
|---|---|---|---|
| | | Η $H_0$ είναι σωστή | Η $H_1$ είναι σωστή |
| ΑΠΟΦΑΣΗ | Δεν απορρίπτω την $H_0$ | σωστή απόφαση | λάθος τύπου ΙΙ |
| | Απορρίπτω την $H_0$ | λάθος τύπου Ι | σωστή απόφαση |

λάθος τύπου Ι : απορρίπτω την $H_0$ ενώ είναι σωστή

λάθος τύπου ΙΙ : δεν απορρίπτω την $H_0$ ενώ είναι σωστή η $H_1$

Φυσικά, αφού η οποιαδήποτε απόφασή μας θα στηρίζεται σε ένα μόνο δείγμα, δεν είμαστε εκ των προτέρων βέβαιοι αν η απόφαση αυτή θα είναι σωστή ή όχι και αν κάνουμε λάθος, τι μορφής λάθος θα κάνουμε. Υπάρχει επομένως κάποια πιθανότητα με την απόφασή μας να διαπράξουμε λάθος τύπου Ι και κάποια άλλη πιθανότητα να διαπράξουμε λάθος τύπου ΙΙ.

Οι πιθανότητες αυτές συμβολίζονται συνήθως με τα ελληνικά γράμματα α και β αντίστοιχα.

Έτσι έχουμε,

| |
|---|
| α = P (λάθος τύπου I ) = P (απορρίπτω την $H_0$ ενώ είναι σωστή) = P ($\cancel{H_0}$ \| $H_0$) |
| β = P (λάθος τύπου II) = P (δέχομαι την $H_0$ ενώ είναι σωστή η $H_1$) = P ($H_0$ \| $H_1$) |

**Ορισμός:** Η τιμή της πιθανότητας α ονομάζεται *επίπεδο σημαντικότητας* (*level of significance*).
Η τιμή αυτή δηλώνει τη μέγιστη πιθανότητα που ο ερευνητής επιτρέπει στον εαυτό του να κάνει λάθος τύπου I. Συνήθως, η τιμή του α επιλέγεται από τον ίδιο τον ερευνητή.

**Σημείωση:** Είναι προφανές ότι σε ένα στατιστικό έλεγχο υποθέσης τρεις είναι οι ποσότητες που καθορίζουν το πλαίσιο της απόφασης του ερευνητή, η πιθανότητα λάθους τύπου I, η πιθανότητα λάθους τύπου II και το κρίσιμο σημείο (δηλαδή τα α, β, c). Δοθέντος ότι έχουμε διαθέσιμες δύο εξισώσεις (αυτές για τα α και β), χρειάζεται να προσδιορίσουμε εκ των προτέρων κάποια από τις τρεις αυτές ποσότητες προκειμένου να υπολογίσουμε τις άλλες. Μια πρακτική που πολύ συχνά ακολουθείται είναι να καθορίσει ο ερευνητής το επίπεδο σημαντικότητας α, στην συνέχεια να υπολογίσει το κρίσιμο σημείο c ώστε να έχει μία μέθοδο απόφασης και στην συνέχεια να μελετά το β.

### Πιθανότητες Σωστής Απόφασης: Η Ισχύς ενός Ελέγχου

Το συμπλήρωμα της ποσότητας β, η πιθανότητα δηλαδή να απορριφθεί η μηδενική υπόθεση όταν πράγματι η μηδενική υπόθεση δεν ισχύει, είναι ένα μέτρο της ικανότητας του ελέγχου που έχουμε επιλέξει να λειτουργεί σωστά. Η συνάρτηση αυτή ονομάζεται *ισχύς (power) του στατιστικού ελέγχου.*

**Ορισμός:** Ως *ισχύς (power)* ενός στατιστικού ελέγχου ορίζεται η συνάρτηση,

$$1 - \beta = P(\cancel{H_0} \mid H_1)$$

Επειδή, συνήθως, η εναλλακτική υπόθεση είναι σύνθετη, περιέχει δηλαδή περισσότερες από μία πιθανές τιμές της παραμέτρου, δεν είναι δυνατόν να υπολογισθεί ένα μόνο $\beta$ που να αντιστοιχεί σε κάθε $\alpha$ αλλά υπάρχει μια τιμή του $\beta$ που αντιστοιχεί σε κάθε μιά από τις ενδεχόμενες τιμές της παραμέτρου που μελετάμε κάτω από την εναλλακτική υπόθεση. Για τον λόγο αυτό, πολύ συχνά μελετάμε την γραφική παράσταση της πιθανότητας του λάθους τύπου ΙΙ, δηλαδή του $\beta$, ως συνάρτησης της πραγματικής τιμής της παραμέτρου. Η γραφική αυτή παράσταση ονομάζεται *καμπύλη λειτουργικών χαρακτηριστικών* του στατιστικού ελέγχου (*operating characteristic curve*).

Αντίστοιχα, η γραφική παράσταση του 1-β, της ισχύος δηλαδή του στατιστικού ελέγχου, ονομάζεται *καμπύλη ισχύος* (*power curve*) του στατιστικού ελέγχου. Για σταθερές τιμές του δείγματος $n$ και του $\alpha$ η ισχύς ενός ελέγχου θα πρέπει να αυξάνει όσο μεγαλώνει η απόσταση μεταξύ της πραγματικής και της υποθετικής τιμής της παραμέτρου. Δοθέντος ότι η περιοχή απόρριψης καθορίζεται και παραμένει σταθερή για κάποιο δεδομένο έλεγχο, το $\alpha$ θα παραμένει επίσης σταθερό. Η καμπύλη των λειτουργικών χαρακτηριστικών εξάλλου θα περιγράφει τα χαρακτηριστικά του στατιστικού ελέγχου. Οποιαδήποτε αύξηση του μεγέθους $n$ του δείγματος θα ελαττώνει το $\beta$ και θα μειώνει την τιμή του για όλες τις εναλλακτικές τιμές της παραμέτρου η οποία βρίσκεται υπό έλεγχο. Υπάρχει επομένως για κάθε μέγεθος δείγματος μια αντίστοιχη καμπύλη λειτουργικών χαρακτηριστικών του δείγματος. (Στην μελέτη της ισχύος θα επανέλθουμε αργότερα).

Οι ιδιότητες των $\alpha$ και $\beta$ παρατίθενται στην συνέχεια.

## Σχέση των α και β

1. Η τιμή του α αποφασίζεται όταν καθορισθεί (επιλεγεί) η περιοχή απόρριψης.
2. Η τιμή του β εξαρτάται από την εναλλακτική υπόθεση που θα επιλέξουμε. Είναι ευκολότερο να διακρίνουμε μεγάλες αποκλίσεις από την υποτιθέμενη (μηδενική) τιμή της παραμέτρου παρά να

διακρίνουμε μικρές αποκλίσεις. Αν θέλουμε να διακρίνουμε μικρές αποκλίσεις από την υποτιθέμενη τιμή της παραμέτρου, το β θα είναι μεγάλο. Αν επιθυμούμε να διαπιστώνουμε μόνο τις μεγάλες διαφορές, το β θα είναι μικρό.

3. Για ένα δεδομένο μέγεθος δείγματος, όσο μεγαλώνουμε το μέγεθος της περιοχής απόρριψης (και επομένως το α) θα ελαττώνεται το μέγεθος του β. Αν ελαττώνουμε το α, το β θα μεγαλώνει.
4. Για δοθέν α, το β μπορεί να ελαττώνεται με αύξηση του μεγέθους n του δείγματος.

Σε μια ιδανική περίπτωση, ο ερευνητής θα έχει στο μυαλό του κάποιες τιμές του α και του β οι οποίες θα καθορίζουν το μέγεθος των αντιστοίχων λαθών που είναι διατεθειμένος να ανεχθεί. Επίσης, θα έχει στο μυαλό του κάποια απόκλιση από την υποτιθέμενη (τη μηδενική) τιμή της παραμέτρου που θα θεωρεί να είναι σημαντική από πρακτική άποψη και την οποία θα θέλει να διαγνώσει. Στην περίπτωση αυτή, η περιοχή απόρριψης θα καθορισθεί σύμφωνα με την προκαθορισμένη τιμή του α. Τέλος, θα επιλέξει ένα μέγεθος για το δείγμα για να πετύχει μια αποδεκτή τιμή για το β για την δεδομένη απόκλιση που θέλει να διαγνώσει. Η επιλογή μπορεί να γίνει με την μέλετη της καμπύλης των λειτουργικών χαρακτηριστικών που αντιστοιχεί σε διάφορα μεγέθη δείγματος για τον επιλεγέντα έλεγχο. Στην πράξη παρόλα αυτά, οι παραπάνω ιδέες δεν είναι εύκολο να εφαρμοσθούν. Στους περισσότερους στατιστικούς ελέγχους, είναι σχετικά εύκολο να καθορισθεί η τιμή του α για κάποια προκαθορισμένη περιοχή απόρριψης όπως θα δούμε σε κάποιο από τα παραδείγματα που ακολουθούν. Είναι όμως συνήθως δύσκολο να υπολογίσει κανείς την τιμή του β για διαφορετικές τιμές της εναλλακτικής υπόθεσης για την παραμέτρο που θέλουμε να ελέγξουμε. Για τον λόγο αυτό, συνήθως αποφασίζουμε και καθορίζουμε το ρίσκο όπου είμαστε διατεθειμένοι να αποδεχθούμε και στην συνέχεια επιλέγουμε την περιοχή απόρριψης. Ένα αντίστοιχο παράδειγμα θα δούμε στην συνέχεια. Αν η ελεγχοσυνάρτηση πάρει τιμή στην περιοχή απόρριψης θα ξέρουμε αμέσως το ρίσκο που παίρνουμε αν κάνουμε λάθος τύπου Ι. Παρ' όλα αυτά, αν η τιμή της ελεγχοσυνάρτησης δεν πέσει στην περιοχή απόρριψης, θα πρέπει να προχωρήσουμε με προσοχή. Δεν θα πρέπει να δεχθούμε αμέσως ότι αποδεχόμαστε την μηδενική υπόθεση αν δεν ξέρουμε την τιμή του β δηλαδή της πιθανότητας να κάνουμε λάθος δευτέρου είδους. Η καλύτερη ενέργεια στην περίπτωση αυτή είναι να μην αποφασίσουμε αμέσως, αλλά να συγκεντρώσουμε περισσότερα στοιχεία.

**Σημείωση:** Για ένα δεδομένο μέγεθος δείγματος, αυτός που παίρνει την απόφαση θα πρέπει να εξισορροπήσει τα δύο είδη των πιθανών λαθών. (Προφανώς δεν υπάρχει περίπτωση να γίνει κάποιο λάθος μέχρις ότου ληφθεί μια απόφαση. Όταν, εξάλλου, ληφθεί απόφαση, η απόφαση αυτή ή θα είναι σωστή ή θα έχει συμβεί ένα από τα δύο λάθη. Δεν είναι δυνατόν να συμβούν και τα δύο λάθη ταυτόχρονα). Αν το α μειωθεί τότε το β θα αυξηθεί. Αντίθετα, αν το β αυξηθεί το α θα ελαττωθεί. Οι τιμές για τα α και β εξαρτώνται από την σημασία του κάθε λάθους στο συγκεκριμένο πρόβλημα. Για παράδειγμα, αν με βάση την απόφαση που θα πάρουμε πρόκειται να κάνουμε αλλαγές στην κρατούσα κατάσταση, θα θέλαμε να είμαστε όσο το δυνατόν βέβαιοι ότι η οποιαδήποτε αλλαγή θα είναι αποδοτική. Στην περίπτωση αυτή, ο κίνδυνος λάθους τύπου I είναι ο πιο σημαντικός και το α θα πρέπει να κρατηθεί σε χαμηλά επίπεδα. Από το άλλο μέρος, αν θέλουμε να έχουμε μεγάλη βεβαιότητα ότι θα μπορέσουμε να διακρίνουμε μεταβολές από την υποθετική τιμή της παραμέτρου, πιο σημαντικός θα είναι ο κίνδυνος από ένα λάθος δευτέρου είδους οπότε ενδεχομένως να επιλέξουμε μια μεγάλη τιμή για το α. Βέβαια, είναι προφανές ότι αν αυξηθεί το μέγεθος του δείγματος, μπορούμε να ελέγξουμε και το α και το β. Αυτό όμως δεν είναι πάντα δυνατό γιατί αύξηση του μεγέθους του δείγματος σημαίνει και αύξηση του κόστους.

Στις περισσότερες πάντως περιπτώσεις, η διαδικασία που ακολουθούμε είναι τέτοια που στηρίζεται στον έλεγχο και καθορισμό του επιπέδου σημαντικότητας $\alpha$. Αυτό μπορεί καλύτερα να εξηγηθεί αν θεωρήσουμε ένα παράδειγμα ελέγχου στατιστικής υπόθεσης από την δικαστική λειτουργία. Στην διαδικασία απονομής δικαιοσύνης το δικαστήριο ξεκινά από την υπόθεση ότι ο κατηγορούμενος είναι αθώος ($H_0$: κατηγορούμενος αθώος). Η εναλλακτική υπόθεση (που η κατηγορούσα αρχή έχει ευθύνη να αποδείξει) είναι ότι ο κατηγορούμενος είναι ένοχος. Από το προηγούμενο σχεδιάγραμμα προκύπτει ότι λάθος πρώτου είδους είναι εκείνο που θα συμβεί αν το δικαστήριο καταδικάσει τον κατηγορούμενο, ενώ στην πραγματικότητα ο κατηγορούμενος είναι αθώος. Αντίστοιχα ένα λάθος δευτέρου είδους θα συμβεί αν το δικαστήριο απαλλάξει τον κατηγορούμενο, ενώ στην πραγματικότητα ο κατηγορούμενος είναι ένοχος. Όπως είναι γνωστό, και ισχύει παραδοσιακά, στην δικαστική διαδικασία είναι προτιμότερο να αθωωθεί ένας ένοχος παρά να καταδικασθεί ένας αθώος. Για τον λόγο αυτό, το λάθος τύπου Ι είναι το σημαντικότερο και γι' αυτό ενδείκνυται να ελέγχεται και να καθορίζεται από την αρχή.

## Είδη Στατιστικών Υποθέσεων

**Ορισμός:** Μια στατιστική υπόθεση ονομάζεται *απλή* (*simple*) αν η μηδενική της υπόθεση αναφέρεται σε μια μόνο συγκεκριμένη τιμή της παραμέτρου θ (π.χ. της μέσης τιμής του πληθυσμού). Δηλαδή

$$H_0 : \theta = \theta_0$$

**Ορισμός:** Μια στατιστική υπόθεση θα ονομάζεται *σύνθετη* (*composite*) αν περιλαμβάνει περισσότερες από μία ενδεχόμενες τιμές της υπό έλεγχο παραμέτρου. Αν δηλαδή είναι της μορφής,

$$H_0 : \theta \geq \theta_0 \quad \text{ή} \quad H_0 : \theta \leq \theta_0$$

**Ορισμός:** Ο έλεγχος μιας στατιστικής υπόθεσης θα ονομάζεται *μονόπλευρος (one sided)* αν η εναλλακτική είναι της μορφής,

$$H_1 : \theta < \theta_0 \quad \text{ή} \quad H_1 : \theta > \theta_0$$

Αντίστοιχα, ο έλεγχος μιας απλής στατιστικής υπόθεσης θα ονομάζεται *αμφίπλευρος (two sided)* αν η εναλλακτική είναι της μορφής,

$$H_1 : \theta \neq \theta_0 .$$

## Παρατηρούμενο Επίπεδο Σημαντικότητας
*(Observed level of significance ή p-value)*

Όλη η προηγηθείσα συζήτηση οδηγεί στο συμπέρασμα ότι δεν υπάρχει γενικά αποδεκτός κανόνας που να οδηγεί στην επιλογή του επιπέδου σημαντικότητας στα προβλήματα ελέγχου στατιστικών υποθέσεων. Οι δυσκολίες αυτές οδήγησαν τους επιστήμονες της Στατιστικής να ορίσουν ένα άλλο μέγεθος που να περιγράφει με καλύτερο τρόπο την κατάσταση που επικρατεί στον έλεγχο στατιστικών υποθέσεων. Η τιμή αυτή είναι το λεγόμενο *παρατηρούμενο επίπεδο σημαντικότητας* ή αλλιώς *τιμή πιθανότητας* ή *p-τιμή* (*observed level of significance ή probability value ή p-value*).

Στον έλεγχο στατιστικών υποθέσεων καθοριστικό ρόλο παίζει η τιμή της στατιστικής συνάρτησης ελέγχου για το συγκεκριμένο δείγμα που έχει παρατηρηθεί. Η πιθανότητα που αντιστοιχεί στην παρατηρηθείσα τιμή της στατιστικής συνάρτησης ελέγχου, ή κάποια άλλη τιμή πιθανότητας ακόμα πιο αντιφατική (αντίθετη) προς την μηδενική υπόθεση μετρά, κατά κάποιο τρόπο, το βάρος των ενδείξεων που υποστηρίζουν την απόρριψη της $H_0$.

**Ορισμός:** Ορίζουμε ως *παρατηρούμενο επίπεδο σημαντικότητας* ή *τιμή πιθανότητας* ή *p-τιμή* (*observed level of significance* ή *probability value* ή *p-value*) την πιθανότητα ή στατιστική συνάρτηση ελέγχου να πάρει μία τιμή τόσο ακραία ή περισσότερο ακραία από αυτήν που πήρε για το συγκεκριμένο δείγμα, κάτω από την μηδενική υπόθεση.

Η έννοια της p-τιμής θα γίνει καλύτερα κατανοητή στο επόμενο κεφάλαιο με την χρήση παραδειγμάτων.

**Σημείωση:** Η χρησιμοποίηση της p-τιμής αντί του επιπέδου σημαντικότητας α στην αντιμετώπιση ενός προβλήματος στατιστικού ελέγχου υπόθεσεως δεν μεταβάλλει την κλασσική στατιστική μεθοδολογία που αναλύθηκε νωρίτερα. Απλά εκείνο που συμβαίνει είναι ότι ο ερευνητής αναφέρει στον ενδιαφερόμενο την p-τιμή (το παρατηρούμενο επίπεδο σημαντικότητας) και αφήνει την επιλογή του κατά πόσο θα πρέπει να απορριφθεί ή όχι η μηδενική υπόθεση στον ενδιαφερόμενο. Μεταφέρεται δηλαδή η ευθύνη επιλογής του α από τον ερευνητή στον ενδιαφερόμενο.